

INSTITUTO TECNOLÓGICO DE AERONÁUTICA



Luis Victor Camilo das Chagas

**MODELAGEM COMPUTACIONAL E
ECONÔMICO FINANCEIRO DA EVOLUÇÃO
TEMPORAL DOS PREÇOS DE PASSAGENS
AÉREAS: UM ESTUDO SOB A PERSPECTIVA
DAS COMPANHIAS AÉREAS NACIONAIS**

Trabalho de Graduação

2018

Curso de Engenharia Civil-Aeronáutica

Luis Victor Camilo das Chagas

**MODELAGEM COMPUTACIONAL E ECONÔMICO
FINANCEIRO DA EVOLUÇÃO TEMPORAL DOS
PREÇOS DE PASSAGENS AÉREAS: UM ESTUDO
SOB A PERSPECTIVA DAS COMPANHIAS
AÉREAS NACIONAIS**

Orientador

Prof. Dr. Marcelo Xavier Guterres (ITA)

ENGENHARIA CIVIL-AERONÁUTICA

SÃO JOSÉ DOS CAMPOS
INSTITUTO TECNOLÓGICO DE AERONÁUTICA

2018

Dados Internacionais de Catalogação-na-Publicação (CIP)
Divisão de Informação e Documentação

das Chagas, Luis Victor Camilo

Modelagem computacional e econômico financeiro da evolução temporal dos preços de passagens aéreas: um estudo sob a perspectiva das companhias aéreas nacionais / Luis Victor Camilo das Chagas.

São José dos Campos, 2018.

Número de folhas no formato 116f.

Trabalho de Graduação – Engenharia Civil-Aeronáutica – Instituto Tecnológico de Aeronáutica, 2018. Orientador: Prof. Dr. Marcelo Xavier Guterres.

1. Gerenciamento de recursos. 2. *Web-scraping*. 3. Precificação de passagens. I. Instituto Tecnológico de Aeronáutica. II. Modelagem computacional e econômico financeiro da evolução temporal dos preços de passagens aéreas: um estudo sob a perspectiva das companhias aéreas nacionais.

REFERÊNCIA BIBLIOGRÁFICA

DAS CHAGAS, Luis Victor Camilo: **Modelagem computacional e econômico financeiro da evolução temporal dos preços de passagens aéreas: um estudo sob a perspectiva das companhias aéreas nacionais**. 2018. 116f. Trabalho de Conclusão de Curso. (Graduação em Engenharia Civil-Aeronáutica) – Instituto Tecnológico de Aeronáutica, São José dos Campos.

CESSÃO DE DIREITOS

NOME DO AUTOR: Luis Victor Camilo das Chagas

TÍTULO DO TRABALHO: Modelagem computacional e econômico financeiro da evolução temporal dos preços de passagens aéreas: um estudo sob a perspectiva das companhias aéreas nacionais.

TIPO DO TRABALHO/ANO: Graduação / 2018

É concedida ao Instituto Tecnológico de Aeronáutica permissão para reproduzir cópias deste trabalho de graduação e para emprestar ou vender cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte deste trabalho de graduação pode ser reproduzida sem a autorização do autor.

Luis Victor Camilo das Chagas

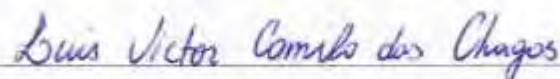
Luis Victor Camilo das Chagas

Rua vigésimo sexto batalhão, 291, cs 26

60712-130, Fortaleza - CE

MODELAGEM COMPUTACIONAL E ECONÔMICO FINANCEIRO DA EVOLUÇÃO TEMPORAL DOS PREÇOS DE PASSAGENS AÉREAS: UM ESTUDO SOB A PERSPECTIVA DAS COMPANHIAS AÉREAS NACIONAIS

Essa publicação foi aceita como Relatório Final de Trabalho de Graduação



Luis Victor Camilo das Chagas

Autor



Prof. Dr. Marcelo Xavier Guterres (ITA)

Orientador



Prof. Dr. Eliseu Lucena Neto

Coordenador do Curso de Engenharia Civil-Aeronáutica

São José dos Campos, 19 de novembro de 2018

Dedico esse trabalho a todos os meus amigos que me ajudaram ao longo de todo este árduo caminho e aos meus pais que sempre me incentivaram.

Agradecimentos

Gostaria de agradecer, primeiramente, aos meus familiares, pois o suporte e incentivo que me proporcionam foi fundamental para que hoje esse trabalho pudesse ser desenvolvido. Além deles, gostaria também de agradecer a todos os amigos que conheci ao longo dessa caminhada, pois os diversos desafios que tive que enfrentar foram minimizados devidos a sempre presente ajuda deles. Por fim, gostaria de agradecer ao professor Marcelo Xavier Guterres por sua orientação ao longo do desenvolvimento desse trabalho e ao professor Alessandor Oliveira pelos ensinamentos e pela ajuda em relação aos assuntos estudados ao longo desse trabalho.

Resumo

A compra de passagens para viagens aéreas vem se tornando cada vez mais fácil de ser realizada e de serem feitas comparações entre as empresas aéreas para se tentar obter o melhor preço de compra. Também fica evidente que os preços das passagens são variáveis e cada empresa tem sua própria política de precificação de passagens. Este trabalho tem como objetivo desempenhar um estudo para analisar e comparar o comportamento dos preços de passagens em alguns trechos do setor de aviação comercial brasileiro. A partir dessas análises, tentar notar padrões na política de precificação das empresas aéreas a fim de verificar o melhor período para se realizar compras de passagens. Além disso, será criado um modelo para previsão de compras de passagens para as diferentes rotas analisadas e será desenvolvido um estudo de economia em compras simuladas de passagens.

Abstract

The purchase of tickets to air travels are turning such easier to be done and to make comparisons among airlines to try achieve the better purchase price. Also it stays evident that the price of tickets are variables and each airline has theirs own ticket pricing policies. This work has as goal to develop a study to analyse and compare the behavior of ticket prices in some routes in the brazilian commercial aviation sector. From this analysis, it try to realize standards in airlines pricing policies in order to verify the best period to make airlines tickets purchase. Besides, it will be created a model to forecast airlines tickets purchases to differents analysed routes and it will be developed a economic study in simulated purchased tickets.

Lista de Figuras

3.1	Diagrama mostrando onde cada categoria do <i>web mining</i> influencia na rede	24
4.1	Fluxograma de explicação da metodologia	26
4.2	Tabela de interesse (Em azul) e atributo (Código à direita) buscados pelo programa em de <i>web scraping</i>	27
4.3	Diagrama de exemplo de coleta dados	30
4.4	Exemplo de dados referentes a uma busca	31
4.5	Série temporal de exemplo de coleta	32
4.6	Variação do preço ao longo do dia 30/05	33
4.7	Exemplo de utilização de gráfico <i>boxplot</i>	34
4.8	Leitura do <i>boxplot</i>	35
4.9	Exemplo de histograma	35
5.1	Relação entre antecedência e preços da empresa Azul no trecho GRU-FOR	42
5.2	Relação entre antecedência e preços da empresa Azul no trecho FOR-GRU	43
5.3	Relação entre antecedência e preços da empresa Avianca no trecho GRU-FOR	45
5.4	Relação entre antecedência e preços da empresa Avianca no trecho FOR-GRU	46
5.5	Relação entre antecedência e preços da empresa Gol no trecho GRU-FOR .	48
5.6	Relação entre antecedência e preços da empresa Gol no trecho FOR-GRU .	49
5.7	Relação entre antecedência e preços da empresa LATAM no trecho GRU-FOR	51
5.8	Relação entre antecedência e preços da empresa LATAM no trecho FOR-GRU	52
5.9	Relação entre atributo do dia e preços da empresa Azul no trecho GRU-FOR	53
5.10	Relação entre atributo do dia e preços da empresa Azul no trecho FOR-GRU	54
5.11	Relação entre atributo do dia e preços da empresa Avianca no trecho GRU-FOR	55

5.12	Relação entre atributo do dia e preços da empresa Avianca no trecho FOR-GRU	55
5.13	Relação entre atributo do dia e preços da empresa Gol no trecho GRU-FOR	56
5.14	Relação entre atributo do dia e preços da empresa Gol no trecho FOR-GRU	57
5.15	Relação entre atributo do dia e preços da empresa LATAM no trecho GRU-FOR	58
5.16	Relação entre atributo do dia e preços da empresa LATAM no trecho FOR-GRU	58
5.17	Histograma de precificação da empresa aérea Azul	60
5.18	Histograma de precificação da empresa aérea Avianca	61
5.19	Histograma de precificação da empresa aérea Gol	61
5.20	Histograma de precificação da empresa aérea LATAM	62
5.21	Relações entre as médias de preços para os trechos GRU-FOR e FOR-GRU	63
5.22	Relações entre as médias de preços para os trechos GRU-FOR e FOR-GRU	64
5.23	Intervalo dos preços mais prováveis x antecedência para a empresa Azul no trecho GRU-FOR	67
5.24	Intervalo dos preços mais prováveis x antecedência para a empresa Azul no trecho FOR-GRU	68
5.25	Intervalo dos preços mais prováveis x antecedência para a empresa Avianca no trecho GRU-FOR	69
5.26	Intervalo dos preços mais prováveis x antecedência para a empresa Avianca no trecho FOR-GRU	70
5.27	Intervalo dos preços mais prováveis x antecedência para a empresa Gol no trecho GRU-FOR	71
5.28	Intervalo dos preços mais prováveis x antecedência para a empresa Gol no trecho FOR-GRU	72
5.29	Intervalo dos preços mais prováveis x antecedência para a empresa LATAM no trecho GRU-FOR	73
5.30	Intervalo dos preços mais prováveis x antecedência para a empresa LATAM no trecho FOR-GRU	74
5.31	Evolução com erro com o aumento do n° de árvores no modelo da Azul . .	79
5.32	Importância de cada variável para o modelo de RF da empresa Azul	79

5.33	Evolução com erro com o aumento do n° de árvores no modelo da Avianca	80
5.34	Importância de cada variável para o modelo de RF da empresa Avianca . .	81
5.35	Evolução com erro com o aumento do n° de árvores no modelo da Gol . . .	82
5.36	Importância de cada variável para o modelo de RF da empresa Gol	82
5.37	Evolução com erro com o aumento do n° de árvores no modelo da LATAM	83
5.38	Importância de cada variável para o modelo de RF da empresa LATAM . .	84

Lista de Tabelas

4.1	Exemplo de coleta de dados	29
4.2	Exemplo de tabela final de coleta	31
4.3	Parâmetros estatísticos de exemplo a partir do dados coletados no dia 28/05, para o trecho GRU-FOR	34
4.4	Significados dos coeficientes da regressão	38
5.1	Sumário dos dados relativos ao trecho GRU-FOR	59
5.2	Sumário dos dados relativos ao trecho FOR-GRU	59
5.3	Correlações entre as médias das empresas aéreas no trecho GRU-FOR . . .	65
5.4	Correlações entre as médias das empresas aéreas no trecho FOR-GRU . . .	66
5.5	Correlações entre as médias dentro de cada empresa aérea nos trechos GRU- FOR e FOR-GRU	66
5.6	Categorias criadas para classificação de faixas de preços	76
5.7	Descrição das variáveis utilizadas	76
5.8	Coeficientes de contingência obtidos	77
5.9	Tabela de confusão obtida para o modelo da empresa Azul	78
5.10	Tabela de confusão obtida para o modelo da empresa Azul	80
5.11	Tabela de confusão obtida para o modelo da empresa Gol	81
5.12	Tabela de confusão obtida para o modelo da empresa LATAM	83
5.13	Matriz de confusão obtida após previsão com a base de dados completa . .	84
5.14	Coeficientes da regressão para categoria econômica A da empresa Azul . .	86
5.15	Coeficientes da regressão para categoria econômica B da empresa Azul . .	87
5.16	Coeficientes da regressão para categoria econômica C da empresa Azul . .	88
5.17	Coeficientes da regressão para categoria econômica D da empresa Azul . .	89
5.18	Coeficientes da regressão para categoria econômica E da empresa Azul . .	90

5.19	Coeficientes da regressão para categoria econômica F da empresa Azul . . .	91
5.20	Coeficientes da regressão para categoria econômica A da empresa Avianca .	92
5.21	Coeficientes da regressão para categoria econômica B da empresa Avianca .	93
5.22	Coeficientes da regressão para categoria econômica C da empresa Avianca .	94
5.23	Coeficientes da regressão para categoria econômica D da empresa Avianca .	95
5.24	Coeficientes da regressão para categoria econômica E da empresa Avianca .	96
5.25	Coeficientes da regressão para categoria econômica F da empresa Avianca .	97
5.26	Coeficientes da regressão para categoria econômica A da empresa Gol . . .	98
5.27	Coeficientes da regressão para categoria econômica B da empresa Gol . . .	99
5.28	Coeficientes da regressão para categoria econômica C da empresa Gol . . .	100
5.29	Coeficientes da regressão para categoria econômica D da empresa Gol . . .	101
5.30	Coeficientes da regressão para categoria econômica E da empresa Gol . . .	102
5.31	Coeficientes da regressão para categoria econômica F da empresa Gol . . .	103
5.32	Coeficientes da regressão para categoria econômica A da empresa LATAM	104
5.33	Coeficientes da regressão para categoria econômica B da empresa LATAM	105
5.34	Coeficientes da regressão para categoria econômica C da empresa LATAM	106
5.35	Coeficientes da regressão para categoria econômica D da empresa LATAM	107
5.36	Coeficientes da regressão para categoria econômica E da empresa LATAM .	108
5.37	Coeficientes da regressão para categoria econômica F da empresa LATAM .	109
5.38	Exemplo de previsões realizadas	110
5.39	Quantidade de vezes que cada empresa foi sugerida	111
5.40	Amostra de base de simulação	111

Sumário

1	Introdução	16
1.1	O problema	17
1.2	Objetivo geral	17
1.3	Objetivos específicos	18
1.4	Justificativa	18
1.5	Delimitação de tema	18
2	Gerenciamento de recursos	19
2.1	Introdução	19
3	Web Scraping	23
3.1	Introdução	23
3.2	Linguagem R	25
4	Metodologia	26
4.1	Revisão de literatura	26
4.2	Obtenção dos dados	27
4.3	Tratamento de dados	30
4.4	Análise de dados	31
4.4.1	<i>Boxplots</i>	34
4.4.2	Histogramas	35
4.4.3	Gráfico de linha	35
4.4.4	<i>Random Forest</i>	36
4.4.5	<i>Regressão linear</i>	36
4.5	Comparação entre previsões e histórico	39

5	Análises de resultados	40
5.1	Web-scraping em R	40
5.2	Estudo de variação de preços	41
5.2.1	Relações entre variáveis	41
5.2.2	Parâmetros estatísticos	59
5.2.2.1	Sumarização dos dados	59
5.3	Realizar previsões de preços	75
5.3.1	Descrição e preparação dos dados	75
5.3.2	<i>Random Forest</i>	77
5.3.3	Regressão	85
5.3.4	Resultados das previsões	109
5.3.5	Simulação	110
6	Considerações finais	112
6.1	Sugestões para trabalhos futuros	113

Capítulo 1

Introdução

Os preços das passagens aéreas são extremamente dinâmicos. As companhias aéreas variam os preços numa tentativa de se obter a máxima receita possível em cada um de seus voos [1]. Como o custo marginal devido a cada passageiro é relativamente baixo, quase todo o custo relativo ao voo é dado por custos fixos. Portanto, quanto maior o *yield* (receita obtida por assento vendido), maior será a receita que a empresa terá. Porém, o *load factor* (taxa de ocupação do voo) também é muito importante, pois cada assento vazio representa uma receita que deixou de ser obtida pela companhia. Por isso, as companhias aéreas buscam sempre, através do gerenciamento de recursos (*revenue management* ou *yield management*), encontrar os valores ótimos dos preços a fim de maximizar tanto o *yield*, como o *load factor*.

Pensando nisso, as empresas aéreas usam modelos de simulação e de previsões de procura de passagens para criar e estimar diferentes pacotes com precificações diferentes [2]. Esses modelos são fundamentados, principalmente, na premissa de diferenças de elasticidade de preço da demanda. Existem pessoas que são extremamente sensíveis ao preço, ou seja, compram passagens em ofertas e deixam de comprar passagens se o preço estiver um pouco mais alto que o esperado. Enquanto isso, existem pessoas que são insensíveis ao preço e estão dispostas a pagar altos preços por um conforto maior na viagem ou por uma viagem imediata à compra da passagem.

Porém, a diferença de elasticidade de preço da demanda não é o único fator que afeta o preço estipulado pela companhia aérea. Existem algumas outras variáveis que também o afetam, tais quais a sazonalidade, a disponibilidade dos assentos, o preço de concorrentes, horário do voo, entre outras [3]. Devido a todos esses fatores que os preços

de passagens aéreas se tornam tão dinâmicos, visto que alguns dos parâmetros podem mudar constantemente, *e.g.* disponibilidade de assentos.

Com o advento e crescimento da internet, ela passou a ser o principal canal de vendas de passagens aéreas. Com isso, o gerenciamento dos preços das passagens se tornou mais fácil, pois a dinâmica dos preços pode ser realizada quase que instantaneamente, de acordo com as variações dos parâmetros que definem os preços. Apesar de facilitar a dinâmica dos preços, ela facilita comparações de compras entre companhias aéreas e também permite o monitoramento dos preços das passagens ao longo do tempo. Assim, ela acaba possibilitando que consumidores busquem padrões de *pricing*¹ das companhias a fim de se ter uma previsão de qual período é melhor para comprar uma passagem aérea [3].

Esse é o foco desse trabalho, ou seja, coletar, observar e analisar dados de preços de companhias aéreas para alguns trechos aéreos nacionais em busca de padrões que ajudem os consumidores sobre qual o melhor momento para realizarem a compra de passagens em cada um desses trechos. Serão utilizados métodos de *web scraping* para obtenção de dados e de *data mining* com a finalidade de se detectar os padrões das estratégias de cada empresa do setor aéreo civil brasileiro.

1.1 O problema

Com a grande dinâmica existente no *pricing* das passagens aéreas, seria possível fazer o consumidor economizar na compra de passagens devido a descobertas de possíveis padrões no gerenciamento de recursos das companhias aéreas, fazendo assim, com que se houvesse uma previsão de quando seria o melhor dia para a compra?

1.2 Objetivo geral

Este trabalho tem como principal objetivo obter e realizar análises em dados relacionados à precificação estabelecida pelas companhias aéreas em alguns trechos a fim de verificar possíveis padrões na precificação desses trechos ao longo do tempo e proporcionar, com isso, maiores facilidades na compra de passagens aos consumidores.

¹É o ato de se decidir o quanto cobrar por algo.

1.3 Objetivos específicos

O presente trabalho tem como principais objetivos específicos:

1. Criar um programa que seja capaz de realizar *web scraping* de forma eficiente dos preços das companhias aéreas brasileiras.
2. Estudar a variação de preços ao longo do tempo para cada companhia aérea em cada um dos trechos estudados.
3. Verificar se as empresas aéreas brasileiras possuem padrões de *revenue management*.
4. Realizar previsões de preços a partir de análises de *data mining*.
5. Responder o quanto se pode economizar, em média, a partir da realização da análise dos resultados.

1.4 Justificativa

No Brasil não existem muitos estudos sobre como as companhias aéreas estipulam o valor de suas passagens. Enquanto isso, os algoritmos utilizados pelas empresas permanecem desconhecidos. Com isso, existe a possibilidade de se montar modelos visando-se prever os períodos nos quais os preços possuem seus menores patamares. Portanto, além de ser uma contribuição para o estudo do setor aéreo brasileiro, este trabalho também será importante para o consumidor brasileiro, que poderá entender um pouco melhor como funciona o sistema de precificação das companhias aéreas nacionais, e ainda poder economizar a partir desse estudo.

1.5 Delimitação de tema

O estudo realizado nesse trabalho foi realizado a partir dos preços coletados ao longo de dois meses, para cada uma das quatro companhias aéreas, levando em consideração os trechos Guarulhos - Fortaleza e Fortaleza - Guarulhos, devido a sua relevância a grande parte dos alunos de graduação do ITA.

Capítulo 2

Gerenciamento de recursos

2.1 Introdução

Segundo [1] “o setor aéreo é um dos mercados cuja volatilidade das posições financeiras e de lucratividade das empresas é alta, sobretudo por conta de suas características de necessidade de vultosos investimentos de capital, flutuação e sazonalidade da demanda, perecibilidade do produto, alta elasticidade-renda, vulnerabilidade a choques externos (sobretudo taxa de câmbio e preço do petróleo) e grau de contestabilidade do mercado¹ acima da média”. Além disso, [1] escreve que “o produto oferecido pelas empresas aéreas é altamente perecível, ou seja, deve ser comercializado em período estritamente definido no tempo.” Devido a esses motivos era fundamental que o setor aéreo desenvolvesse e realizasse um aprimoramento contínuo em um modelo de gerenciamento de receitas que o permitisse sobreviver com todas essas dificuldades.

De acordo com [5], “o gerenciamento de receitas surgiu das companhias aéreas, a partir da desregulação de tarifas que ocorreu na década de 70, nos EUA, conduzindo o mercado para uma pesada competição e gerando oportunidades para que modelos de gerenciamento de receitas pudessem ser reconhecidos, em um estágio inicial”. Uma das formas mais primitivas de gerenciamento de receitas utilizados aconteceu, conforme [6], no momento em que “algumas companhias aéreas começaram a oferecer passagens com preços diferentes que misturavam descontos e tarifas mais baixas para passageiros voando dentro de um mesmo avião, no mesmo compartimento”. Esse tipo de diferenciação de preços

¹O grau de contestabilidade dependerá do nível de obstáculos à entrada e à saída na indústria [4], quanto menor é o nível de obstáculos à entrada e à saída, maior é o grau de contestabilidade.

permitiu que assentos que antes voavam vazios, fossem preenchidos, gerando uma possível receita extra para a companhia aérea. Porém, percebeu-se que essa receita extra poderia ser maximizada de acordo com a quantidade de assentos que seriam disponibilizados para cada uma das tarifas estabelecidas. Portanto, tornou-se importante pesquisar e desenvolver técnicas de controle de estoque de assentos.

Segundo [5], “Littlewood (1972) foi o primeiro a propor um método como solução para o controle de estoque de assentos em um voo com duas classes de tarifas”. Ainda conforme [5], “ele sugeriu fechar a classe de menor tarifa quando o lucro devido a venda de outro assento com baixa tarifa é excedida pelo lucro esperado da venda do mesmo assento por uma tarifa mais alta”. Esse método ficou conhecido como *Littlewood’s rule* e marcou o início do que ficou popularmente conhecido, em um primeiro momento, como *yield management* e posteriormente veio a ser chamado de *revenue management*, ou gerenciamento de receitas [6].

[6] escreve que “nos Estados Unidos, o início do desenvolvimento de técnicas do *revenue management* iniciou-se a partir do *American Airlines’ Super Saver fares* em abril de 1977, um pouco antes da desregulação das companhias aéreas domésticas e internacionais dos Estados Unidos.” As *Super Saver fares* disponibilizadas pela American Airlines consistiam de passagens com um desconto médio de aproximadamente 50% [7]. Anos depois, em 1983, a American Airlines aplicou outro modelo de precificação, baseada na distância percorrida. De acordo com [7], isso foi criado numa tentativa de se simplificar os preços das passagens.

Com o advento e o desenvolvimento da tecnologia, o gerenciamento de receitas vem se tornando cada vez mais sofisticado e preciso, fazendo com que as receitas das companhias aéreas estejam sempre o mais próximo possível do seu máximo. Isso fez com que outros mercados também seguissem por esse caminho. Conforme [6], “o sucesso do *revenue management* foi largamente noticiado, e isso estimulou o desenvolvimento de sistemas de *revenue management* para outros setores de transporte e em outras áreas de serviços.”

O gerenciamento de receitas é definido por [5] como sendo “a arte de maximizar o lucro gerado a partir de uma capacidade limitada de um produto sobre um horizonte finito a partir da venda de cada produto para o consumidor certo, no tempo certo, pelo preço certo”. Para [1], “o objetivo central do *revenue management* é otimizar a receita total

das companhias aéreas, levando em consideração aspectos de *pricing*, controle de estoque de assentos, diferenciação de produto e condições da concorrência”. As duas definições apontam para que as empresas aéreas tenham uma boa segmentação de mercado a fim de se ter um *revenue management* ótimo. Enquanto que podem ser consideradas iguais no sentido de que para se encontrar essa segmentação de mercado, é necessário que as variáveis apresentadas por [1] estejam constantemente ajustadas ao longo do tempo.

De acordo com [8], a segmentação de mercado pode ser definida como “um grupo de consumidores que têm muito em comum para formar uma base apropriada para um produto, preço, distribuição e combinação de promoções”. Para se ter conhecimento sobre a segmentação ideal é fundamental que haja, nas companhias aéreas, uma ferramenta de previsão de demanda e oferta para cada um dos segmentos estabelecidos pelas empresas. Essa previsão tem como base, principalmente, as variáveis de *pricing* e controle de estoque de assentos.

A política de controle de reservas das companhias aéreas se tornou um fator importante para o gerenciamento de recursos delas. [5] escreve que “a alocação ótima do estoque de assentos foi então traduzida em uma política de controle de reservas, que determina se ou não aceita uma solicitação de reserva quando essa chega”. Com isso, é possível perceber que, devido a velocidade do processamento de dados nos sistemas de reservas, a escolha da quantidade de assentos a uma determinada tarifa acaba sendo definida em tempo real, fazendo com que as empresas acabem rejeitando reservas em prol da ocupação daquela mesma reserva por um preço maior.

O controle de estoque vem sendo estudado desde 1972, quando Littlewood formulou a primeira teoria nesse sentido. Segundo [6], “existe uma progressão a partir da regra de Littlewood para duas tarifas, para o controle do lucro marginal esperado por assento para múltiplas classes, para os limites ótimos de reservas para um voo direto, para o controle do segmento e, mais recentemente, para o controle de tarifas de origem-destino (*ODF Control*)”. Mais informações sobre esses métodos podem ser observadas em [6].

As técnicas de controle de estoque atuais tentam calcular os limites ótimos de reserva de uma dada viagem fazendo com que as empresas aéreas permitam a reserva de mais assentos do que o disponível na aeronave. Essa prática é chamada de *overbooking*, e as companhias aéreas a praticam devido ao frequente número de assentos vazios observados em viagens nas quais todos os assentos foram reservados. [1] expõe que “isso ocorre

devido a dois tipos de comportamentos dos passageiros: 1. o comportamento conhecido como *no-show*, definidos como a situação em que o passageiro que compra a passagem, mas não comparece ao embarque por conta de algum fator de desistência pessoal; e 2. o comportamento de cancelamentos de reservas feitos com pouca antecedência em relação ao horário do voo, situação em que o passageiro que comprou a passagem comunica à companhias que não irá comparecer a um voo cuja partida está programada para um futuro bem próximo”. Ambas situações acabam não permitindo que haja reposições nesses assentos.

A prática do *overbooking* permite que as empresas aéreas consigam maximizar seus lucros, desde que seja feita de maneira otimizada, pois em caso de excesso de assentos em *overbooking*, a empresa terá que lidar com muitos problemas, tanto de imagem, quanto financeiros. Confirmando tal afirmação, [1] escreve que “para que esse controle seja feito de forma eficaz, é necessário conhecer como ocorre a chegada das reservas nos Sistemas de Reservas das companhias aéreas, isto é, como se comportam os passageiros em relação ao período em que efetuam a reserva/compra de assentos para determinado voo”.

Outro fator importante para gerenciamento de receitas é a discriminação de preços. Segundo [9], “o objetivo básico de toda estratégia de preços é a ‘captação do excedente do consumidor e sua conversão em lucros adicionais para a empresa’”. Esse excedente de consumidor pode ser captado através da exploração da pouca sensibilidade ao preço de alguns de seus consumidores. Geralmente, passageiros que realizam viagens a lazer são sensíveis a variação de preços em passagens, por isso, acabam comprando suas passagens com muita antecedência, buscando melhores preços. Enquanto isso, consumidores que viajam a trabalho tendem a ser menos sensíveis à variação de preços e, muitas vezes compram suas passagens com pouca antecedência até a decolagem de sua aeronave.

Portanto, é possível perceber que as técnicas de *revenue management* são de fundamental importância para as empresas aéreas, sem esse controle no *pricing*, dificilmente elas conseguiriam sobreviver as complexidades do setor aéreo.

Capítulo 3

Web Scraping

3.1 Introdução

O grande crescimento de dados disponíveis na rede mundial de computadores fez com que ela se tornasse o maior banco de dados públicos do mundo [10]. Se antigamente um dos maiores problemas relacionados a estudos era a falta de informação, hoje ela existe de forma abundante. Empresas, instituições públicas e privadas geram todo tipo imaginável de informação e disponibilizam-nas em seus espaços virtuais [11]. Parte da informação disponível na internet é de fácil obtenção, porém, frequentemente a aplicação de técnicas simples para coleta de dados tornam-se inviáveis devido ao enorme volume de dados que muitas vezes se tem que extrair. Nesses casos, é muito comum o uso de técnicas de *web scraping* para obtenção de dados.

Web scraping é parte de um campo científico que visa descobrir conhecimentos úteis a partir de informações que estão disponíveis na internet nominado como *web mining*. Ele pode ser categorizado de três maneiras: *web structure mining*, *web usage mining* e *web content mining*. A primeira extrai padrões de estruturas ligadas entre páginas, e apresenta a rede como um gráfico direcionado no qual os nós representam páginas e fronteiras diretas representam *links*. A segunda minera padrões de atividades coletadas a partir de análises de registros da web, com o objetivo de se entender o comportamento do usuário durante suas visitas em páginas da internet. A última extrai e minera informações valiosas do conteúdo da internet [10]. É nesse contexto que a técnica de *web scraping* está inserida.

A Figura 3.1 ilustra um diagrama no qual mostra aonde cada categoria influencia em páginas da internet.

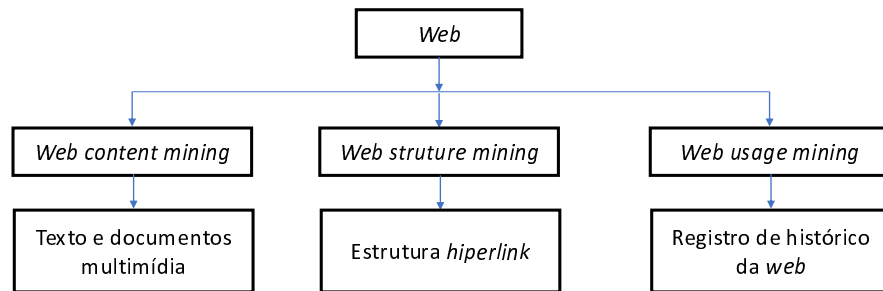


Figura 3.1: Diagrama mostrando onde cada categoria do *web mining* influencia na rede

A utilização do *web scraping* permite que as páginas de interesse na *internet* sejam acessadas e, a partir delas, as informações desejadas sejam extraídas, transformadas caso seja necessário, e salvas em um banco de dados como um conjunto de dados estruturados [12], para posterior análise. Além da obtenção de dados, a utilização dessa técnica faz com que muitos processos de coleta de dados, em diversas páginas da *internet*, acabem sendo automatizados e trazendo informações de maneira mais organizadas. Existem diversos softwares criados exclusivamente para realizar *web scraping* de páginas e algumas linguagens de programação de permitem a criação de *scripts* que também a realizam, tais como a *linguagem Python* e a *linguagem R*. No desenvolvimento desse trabalho foi criado e utilizado um *script* de *web scriping* na *linguagem R*.

A técnica de *web scraping* é bastante utilizada para desenvolvimento de trabalhos acadêmicos. Alguns exemplos disso são os artigos [12] e [13].

O primeiro utiliza a técnica para se obter valores do mercado de aluguéis nos Estados Unidos. Foram obtidos milhões de dados de aluguéis extraídos de classificados especializados em aluguéis de casas. A partir desses dados, os autores tentaram refletir a atividade recente do mercado e o escopo total do mercado de aluguéis dos EUA.

Enquanto que o segundo artigo mencionado utilizada a técnica para se obter dados da *internet* para estudo na área da psicologia. O autor comenta sobre diversas teorias sobre como se obter dados de qualidade para os estudos por meio do *web scraping*, como o estudo dos dados deve ser realizado, além de discutir as preocupações legais, práticas e éticas encaradas na condução de projetos envolvendo *web scraping*.

3.2 Linguagem R

R é uma linguagem e um ambiente para a computação estatística e gráfica. Ela foi inicialmente desenvolvida por Ihaka e Gentleman (1996), ambos da Universidade de Auckland, Nova Zelândia. O atual desenvolvimento da linguagem é realizado por um time principal composto por algumas pessoas de diferentes instituições ao redor do planeta [14].

R foi escolhida para a implementação das rotinas de *web scraping* e para as análises dos dados obtidos devido ao fato de ser um programa livre e facilmente acessível, que não necessita da compra de licenças para se ter acesso as suas ferramentas [11].

Além disso, R é código aberto e conta com uma comunidade que frequentemente contribui para desenvolvimento da linguagem [12]. Esse fato contribui para que exista diversos tutoriais disponíveis na internet para consulta, facilitando bastante o trabalho de quem não tem domínio completo do programa.

Para a implementação do código utilizado para a obtenção de dados foi necessário a utilização de mais dois programas, o *Docker* e o *VNC Viewer*. O *Docker* permite criar aplicações e “*containers*” que isolam o sistema operacional base e toda pilha de dependências de seu *app* de forma leve em espaço e performance. Enquanto isso, o *VNC Viewer* é um visualizador do sistema VNC de conexão remota, que permite se conectar a outro computador, ou “*container*”. Ele permite ver o passo a passo do código em R faz em browser aberto dentro do *Docker*, permitindo que se façam *debugs* na rotina escrita.

O método utilizado para a obter os dados de preços dos sites é fundamentado em *HTML*. Ela é a linguagem responsável por exibir a informação dos sites e pode ser visualizada com facilidade ao se clicar com o botão direito em alguma página e escolher a opção “inspecionar elemento” (O *browser* utilizado foi *Google Chrome*). A página na *web* é composta por vários elementos e é através desses elementos que conseguimos fazer a extração das informações contidas nos sites. Para isso, basta descobrir qual é o elemento que contém a informação desejada.

A busca pelo elemento desejado no site, neste trabalho, foi realizada através da busca por atributo. Atributo são informações que são passadas para que o código se comporte da maneira esperada. Na rotina criada, a busca foi realizada através do atributo do tipo “*Class*” com o de nome ‘*matrix-container*’ e transformar os valores obtidos em texto.

Capítulo 4

Metodologia

O trabalho proposto foi realizado de acordo com os passos mostrados no diagrama mostrado na Figura 4.1.

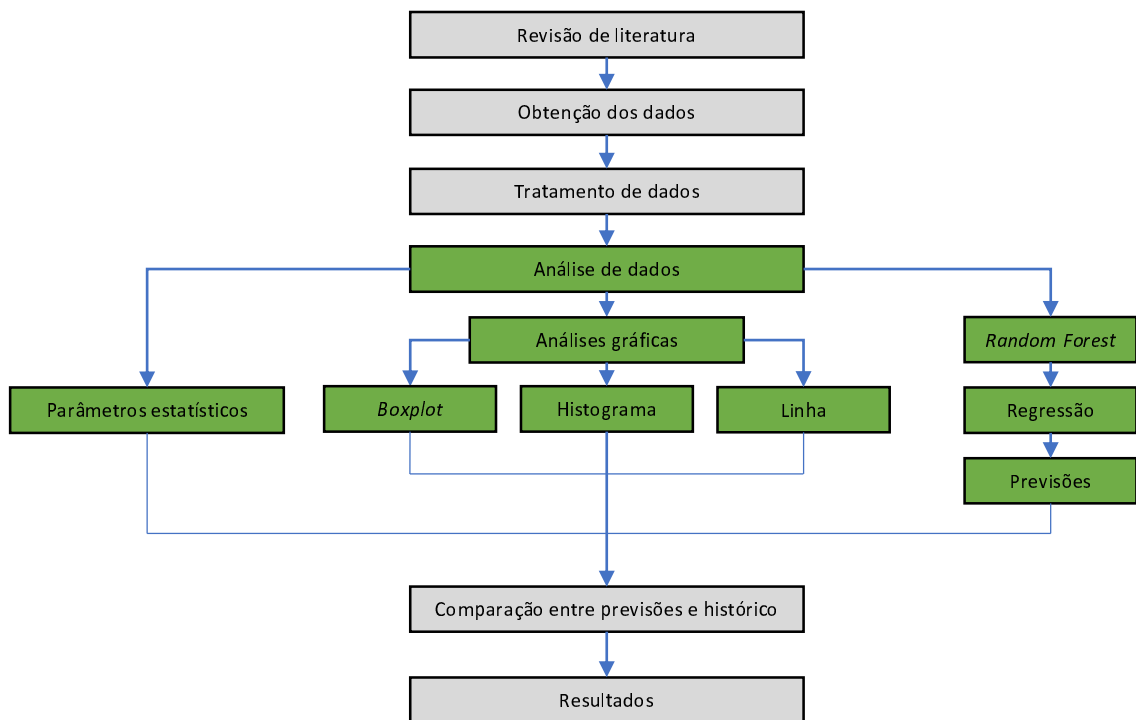


Figura 4.1: Fluxograma de explicação da metodologia

4.1 Revisão de literatura

O trabalho apresentado foi iniciado a partir da revisão de literatura em artigos que levam em consideração análises da dinâmica de preços observados na reserva de passagens

em viagens de companhias aéreas. Os métodos utilizados nesse trabalho são similares aos utilizados no artigo de [3], porém restringindo-se um pouco na quantidade de vezes que os dados foram obtidos para passagens distintas do dia de coleta, que no artigo citado são realizadas três obtenções e no trabalho proposto foram realizadas doze, ocorrendo de duas em duas horas.

4.2 Obtenção dos dados

O programa de *web scraping*, cujo *script* criado em linguagem R pode ser visualizado no Anexo A. Ele foi utilizado para se fazer a obtenção dos preços das passagens ao longo dos dias, atingindo, assim, o objetivo específico 1.

A coleta de dados foi feita a partir do site da empresa de vendas de passagens aéreas “Decolar.com” [15], empresa de comercio de passagens aéreas que possuía *market share* de 11% em 2016, segundo o site euromonitor.com [16]. A aquisição de dados foi realizada pelo programa por meio de um atributo de classe existente no código *HTML* do site, nomeado de *matrix-container*. É nesse atributo que estão contidos os valores da tabela mostrada na Figura 4.2. Além disso, a Figura 4.2 também mostra a parte do código *HTML* que foi útil para a realização desse trabalho.

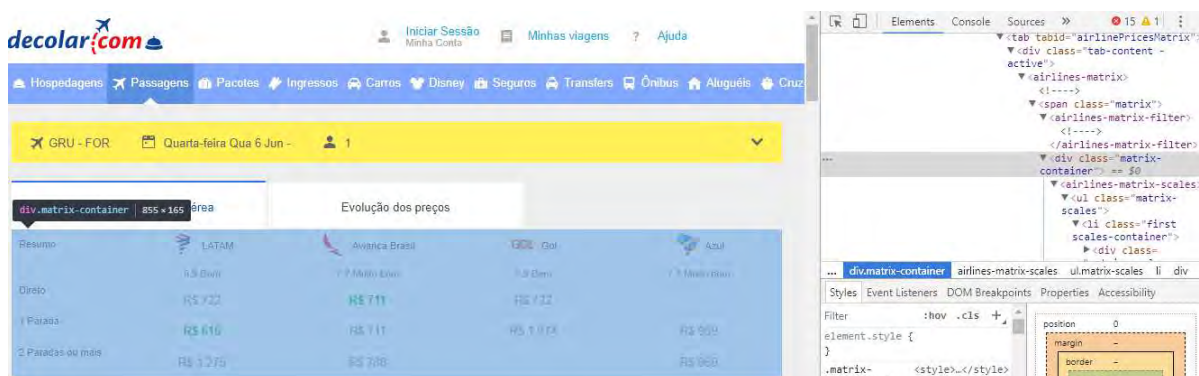


Figura 4.2: Tabela de interesse (Em azul) e atributo (Código à direita) buscados pelo programa em de *web scraping*

No site foi possível encontrar as menores tarifas para cada uma das companhias aéreas nacionais nos trechos pesquisados. Os trechos escolhidos para a análise foram o Guarulhos - Fortaleza e Fortaleza - Guarulhos, devido a sua relevância aos alunos de graduação do ITA.

A coleta dos preços foi realizada diariamente ao longo de dois meses, iniciando-se

no dia 01/08/2018. Ela foi iniciada às 08:00 horas da manhã e, desde então, uma nova coleta era realizada de duas em duas horas. Em cada uma das obtenções de dados, eram extraídas as seguintes variáveis:

- Tipo de voo
- Preço da Avianca
- Preço da Azul
- Preço da Gol
- Preço da LATAM
- Data da extração
- Data da viagem
- Trecho
- Horário

A cada rotina do código de *web scraping* (a cada duas horas), a variável “Data da viagem” variava desde a “Data de Extração” até noventa dias após a mesma, coletando cada uma das outras variáveis supracitadas a cada variação na “Data da Viagem”, gerando, assim, a extração do horário. Um exemplo de coleta é mostrada na Tabela 4.1.

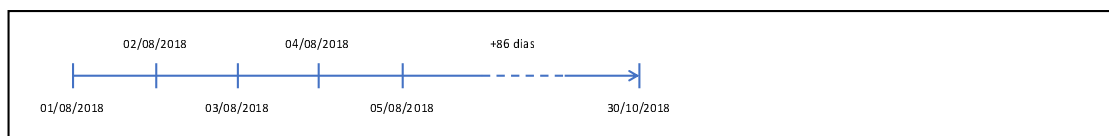
Tipo de Voo	Avianca Brasil	Azul	Gol	LATAM	Data de Extração	Data da Viagem	Trecho	Horário
Direto	469	567	1074	834	5/29/2018	5/29/2018	GRU-FOR	29-05-18_07-00
1 Parada	584	1327	834	1139	5/29/2018	5/29/2018	GRU-FOR	29-05-18_07-00
2 Paradas ou mais	#N/A	#N/A	#N/A	1368	5/29/2018	5/29/2018	GRU-FOR	29-05-18_07-00
Direto	469	1327	834	834	5/29/2018	5/30/2018	GRU-FOR	29-05-18_07-00
1 Parada	777	1678	834	720	5/29/2018	5/30/2018	GRU-FOR	29-05-18_07-00
2 Paradas ou mais	#N/A	#N/A	#N/A	1401	5/29/2018	5/30/2018	GRU-FOR	29-05-18_07-00
Direto	469	567	834	834	5/29/2018	5/31/2018	GRU-FOR	29-05-18_07-00
1 Parada	777	567	834	1139	5/29/2018	5/31/2018	GRU-FOR	29-05-18_07-00
2 Paradas ou mais	1184	#N/A	#N/A	1237	5/29/2018	5/31/2018	GRU-FOR	29-05-18_07-00
Direto	469	1327	834	834	5/29/2018	6/1/2018	GRU-FOR	29-05-18_07-00
1 Parada	788	1327	834	1139	5/29/2018	6/1/2018	GRU-FOR	29-05-18_07-00
2 Paradas ou mais	#N/A	#N/A	#N/A	1259	5/29/2018	6/1/2018	GRU-FOR	29-05-18_07-00
Direto	469	567	834	834	5/29/2018	6/2/2018	GRU-FOR	29-05-18_07-00
1 Parada	584	567	834	1058	5/29/2018	6/2/2018	GRU-FOR	29-05-18_07-00
2 Paradas ou mais	931	#N/A	#N/A	1183	5/29/2018	6/2/2018	GRU-FOR	29-05-18_07-00
Direto	469	1074	1981	874	5/29/2018	6/3/2018	GRU-FOR	29-05-18_07-00
1 Parada	1184	1197	834	1456	5/29/2018	6/3/2018	GRU-FOR	29-05-18_07-00
2 Paradas ou mais	1184	#N/A	#N/A	1150	5/29/2018	6/3/2018	GRU-FOR	29-05-18_07-00
Direto	469	1074	834	834	5/29/2018	6/4/2018	GRU-FOR	29-05-18_07-00
1 Parada	584	1191	834	616	5/29/2018	6/4/2018	GRU-FOR	29-05-18_07-00
2 Paradas ou mais	1184	#N/A	#N/A	1150	5/29/2018	6/4/2018	GRU-FOR	29-05-18_07-00
Direto	469	567	780	722	5/29/2018	6/5/2018	GRU-FOR	29-05-18_07-00
1 Parada	711	714	780	874	5/29/2018	6/5/2018	GRU-FOR	29-05-18_07-00
2 Paradas ou mais	2229	#N/A	#N/A	1246	5/29/2018	6/5/2018	GRU-FOR	29-05-18_07-00
Direto	711	567	780	722	5/29/2018	6/6/2018	GRU-FOR	29-05-18_07-00
1 Parada	711	567	1074	616	5/29/2018	6/6/2018	GRU-FOR	29-05-18_07-00
2 Paradas ou mais	931	#N/A	#N/A	1160	5/29/2018	6/6/2018	GRU-FOR	29-05-18_07-00

Tabela 4.1: Exemplo de coleta de dados

A Figura 4.3 mostra diagramas exemplificando como acontecem as coletas.

Data da coleta: 01/08/2018 Horário da coleta: 08:00

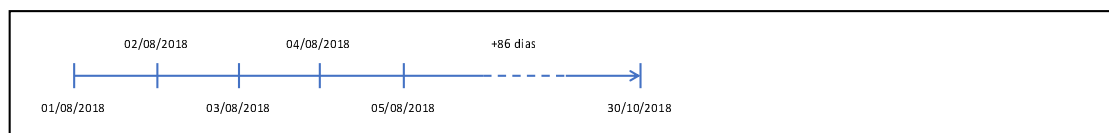
Data nos quais os preços foram coletados:



Próxima coleta:

Data da coleta: 01/08/2018 Horário da coleta: 10:00

Data nos quais os preços foram coletados:

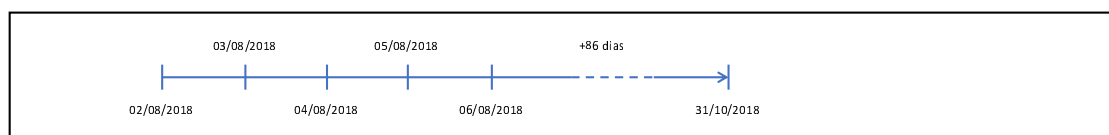


Sucessivamente de duas em duas horas...

Exemplo de coleta do próximo dia

Data da coleta: 02/08/2018 Horário da coleta: 08:00

Data nos quais os preços foram coletados:



Outro exemplo de coleta

Data da coleta: 03/08/2018 Horário da coleta: 10:00

Data nos quais os preços foram coletados:

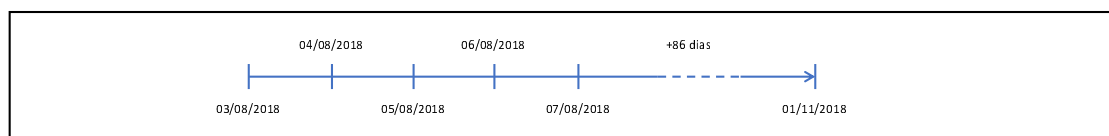


Figura 4.3: Diagrama de exemplo de coleta dados

4.3 Tratamento de dados

O tratamento dos dados é necessário por dois motivos: existem algumas inconsistências obtidas a partir da coleta de dados que devem ser retiradas, pois elas atrapalham as análises e, além disso, os dados obtidos ainda não estão no formato ideal para análise, após obtenção.

Nas bases obtida através do *web scraping*, a data de extração, a data de viagem, o trecho e o horário definem uma coleta, ou seja, todos os preços obtidos que apresentarem esses parâmetros iguais estão relacionados a mesma busca. Um exemplo pode ser visto na Figura 4.4

Tipo de Voo	Avianca Brasil	Azul	Gol	LATAM	Data de Extração	Data da Viagem	Trecho	Horário
Direto	469	567	1074	834	5/29/2018	5/29/2018	GRU-FOR	29-05-18_07-00
1 Parada	584	1327	834	1139	5/29/2018	5/29/2018	GRU-FOR	29-05-18_07-00
2 Paradas ou mais	#N/A	#N/A	#N/A	1368	5/29/2018	5/29/2018	GRU-FOR	29-05-18_07-00
Direto	469	1327	834	834	5/29/2018	5/30/2018	GRU-FOR	29-05-18_07-00
1 Parada	777	1678	834	720	5/29/2018	5/30/2018	GRU-FOR	29-05-18_07-00
2 Paradas ou mais	#N/A	#N/A	#N/A	1401	5/29/2018	5/30/2018	GRU-FOR	29-05-18_07-00
Direto	469	567	834	834	5/29/2018	5/31/2018	GRU-FOR	29-05-18_07-00
1 Parada	777	567	834	1139	5/29/2018	5/31/2018	GRU-FOR	29-05-18_07-00
2 Paradas ou mais	1184	#N/A	#N/A	1237	5/29/2018	5/31/2018	GRU-FOR	29-05-18_07-00

Figura 4.4: Exemplo de dados referentes a uma busca

A análise dos dados foi realizada com base nos preços mínimos de cada empresa aérea, de tal forma que o “Tipo de voo” não foi levado em consideração na continuidade das análises. Portanto, foi necessário se ajustar a base de tal forma a se desprezar o “Tipo de voo” e se obter os valores mínimos para cada empresa aérea, a cada dia coletado.

Além disso, algumas variáveis foram criadas e outras reorganizadas para efeito de se ter uma otimização no momento da análise. A Tabela 4.2 mostra um exemplo da tabela final utilizada nas análises dos dados.

key	Preço	Data de extração	Data de Viagem	Trecho	Horário	Antecedência	Mês_Viagem	Nome_Dia	Fds_Sem	Atributo	Cat. Economica	Empresa
2018-07-05_2018-07-04_GRU-FOR_10-00	695	04/07/2018	05/07/2018	GRU-FOR	10-00	1	julho	quinta-feira	Semana	Normal	C	Azul
2018-07-06_2018-07-04_GRU-FOR_10-00	1250	04/07/2018	06/07/2018	GRU-FOR	10-00	2	julho	sexta-feira	Semana	Normal	F	Azul
2018-07-07_2018-07-04_GRU-FOR_10-00	695	04/07/2018	07/07/2018	GRU-FOR	10-00	3	julho	sábado	Fds	Normal	C	Azul
2018-07-08_2018-07-04_GRU-FOR_10-00	1153	04/07/2018	08/07/2018	GRU-FOR	10-00	4	julho	domingo	Fds	Normal	E	Azul
2018-07-09_2018-07-04_GRU-FOR_10-00	695	04/07/2018	09/07/2018	GRU-FOR	10-00	5	julho	segunda-feira	Semana	Normal	C	Azul
2018-07-10_2018-07-04_GRU-FOR_10-00	695	04/07/2018	10/07/2018	GRU-FOR	10-00	6	julho	terça-feira	Semana	Normal	C	Azul
2018-07-11_2018-07-04_GRU-FOR_10-00	629	04/07/2018	11/07/2018	GRU-FOR	10-00	7	julho	quarta-feira	Semana	Normal	C	Azul
2018-07-12_2018-07-04_GRU-FOR_10-00	1031	04/07/2018	12/07/2018	GRU-FOR	10-00	8	julho	quinta-feira	Semana	Normal	E	Azul

Tabela 4.2: Exemplo de tabela final de coleta

4.4 Análise de dados

Nessa etapa foram realizadas análises dos dados obtidos com a coleta ao longo dos meses. Fundamentada nessas análises, é possível se obter respostas para os objetivos específicos 2 e 3. Elas foram realizadas primeiramente por meio de gráficos montados e medidas estatísticas para a busca por padrões na precificação de passagens das empresas aéreas analisadas. O gráfico mostrado na Figura 4.5 mostra um exemplo de visualização de uma série obtida.

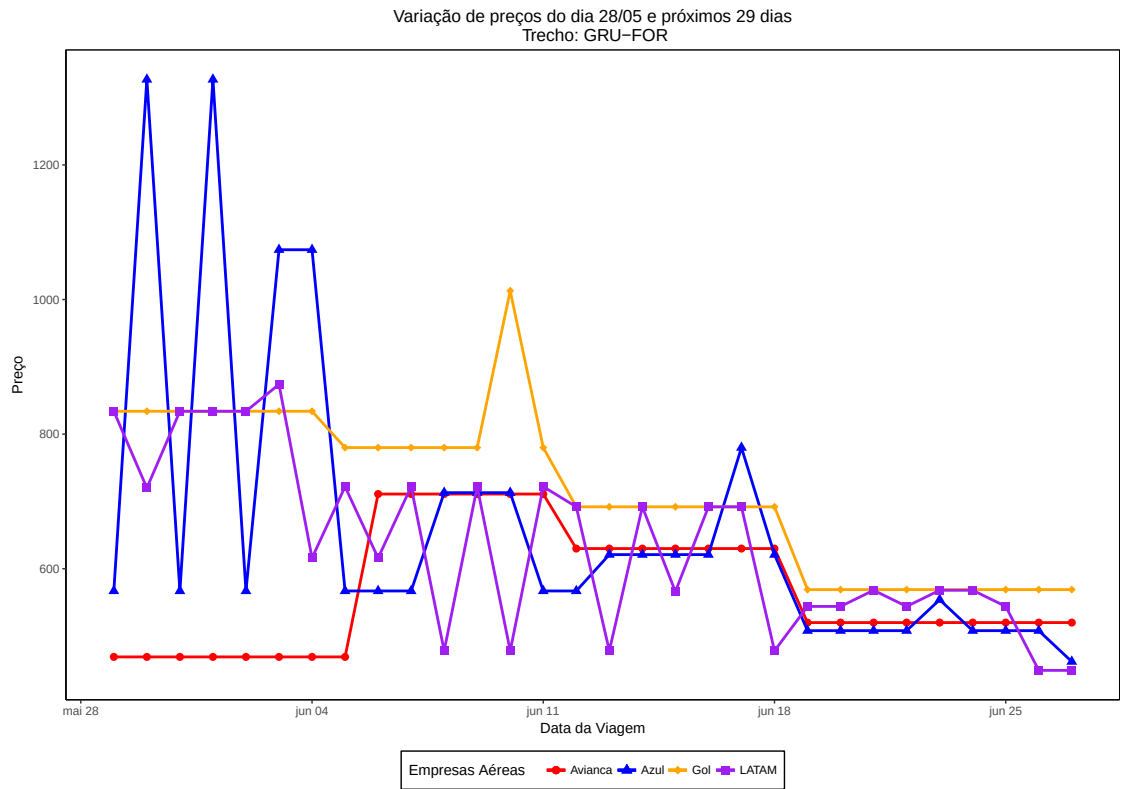


Figura 4.5: Série temporal de exemplo de coleta

O gráfico mostrado na Figura 4.6 mostra a variação do preço mínimo no trecho GRU-FOR ao longo do dia 30/05.

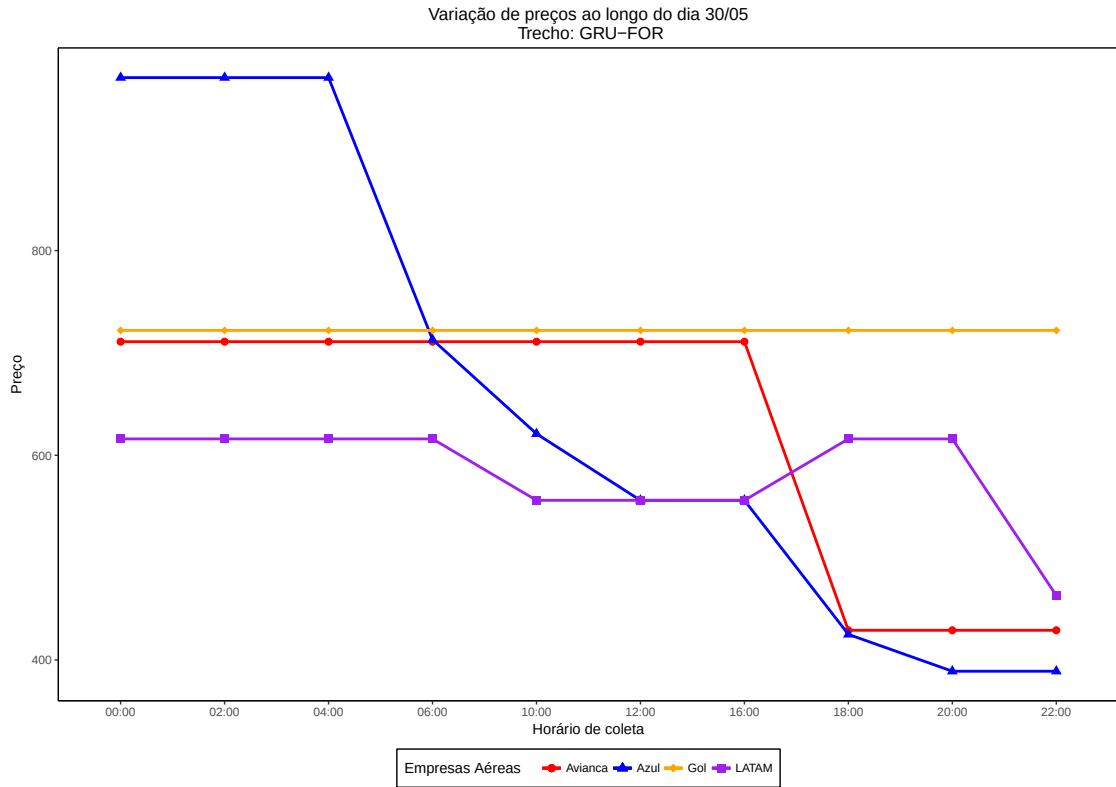


Figura 4.6: Variação do preço ao longo do dia 30/05

As medidas utilizadas serão a média, desvio padrão, máximo e mínimo. A média pode ser calculada com base na equação 4.1.

$$\mu = \frac{x_1 + x_2 + \dots + x_i + \dots + x_n}{n} \quad (4.1)$$

Nos quais, μ é a média obtida para a série analisada, x_i são valores da série e n é a quantidade de elementos da série. O desvio padrão é calculado a partir da equação 4.2.

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2} \quad (4.2)$$

Onde σ é o desvio padrão calculado, x_i são valores da série e n é a quantidade de elementos da série. Enquanto que o máximo e o mínimo podem ser obtidos por meio das equações 4.3 e 4.4, respectivamente.

$$x_j = \max(x_1; x_2; \dots; x_i; \dots; x_n) \quad (4.3)$$

$$x_k = \min(x_1; x_2; \dots; x_i; \dots; x_n) \quad (4.4)$$

Onde x_j é o maior valor contido na série observada, x_k é o menor valor contido na série observada, x_i são valores da série e n é a quantidade de elementos da série.

Um exemplo de utilização dos parâmetros citados nas equações 4.1, 4.2, 4.3 e 4.4 pode ser observado na tabela 4.3, na qual foi construída com base nos mesmos dados utilizados para a criação do gráfico mostrado na Figura 4.5.

Medida	Preço (R\$)			
	Avianca Brasil	Azul	Gol	LATAM
Mínimo	469.0	462.0	569.0	449.0
Máximo	711.0	1 327.0	1 013.0	874.0
Média	570.3	664.0	716.5	635.8
Desvio Padrão	92.6	230.7	118.7	129.4

Tabela 4.3: Parâmetros estatísticos de exemplo a partir do dados coletados no dia 28/05, para o trecho GRU-FOR

Para a análise gráfica dos resultados obtidos, foram utilizados gráficos do tipo *boxplot*, histogramas e de linha.

4.4.1 *Boxplots*

Gráficos do tipo *boxplots* são bastante comuns em análises estatísticas, pois, com eles, é possível se ter uma percepção do nível de variação dos valores obtidos na base de dados trabalhada, além de se obter os *outliers*, e de se poder realizar comparações de alguns parâmetros estatísticos (mediana e quartis) entre classes estudadas. O gráfico 4.7 mostra um exemplo de *boxplot* criado.

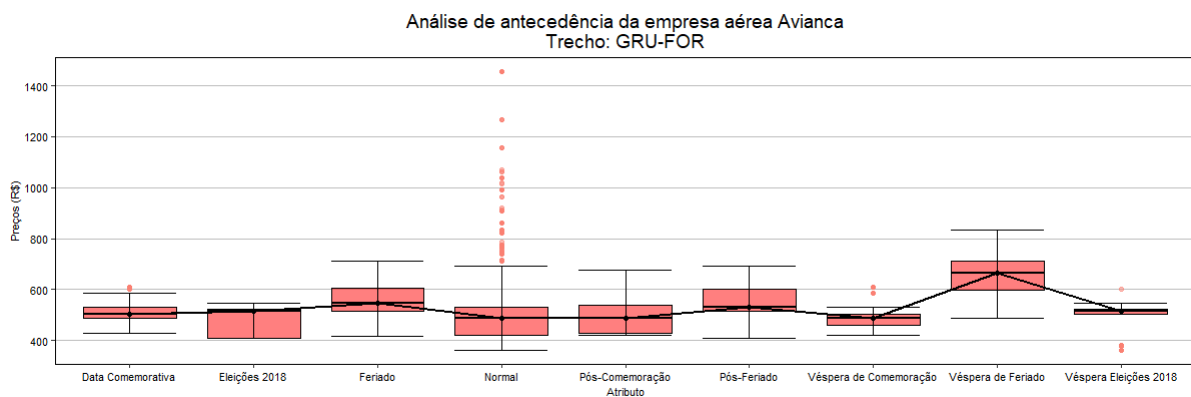
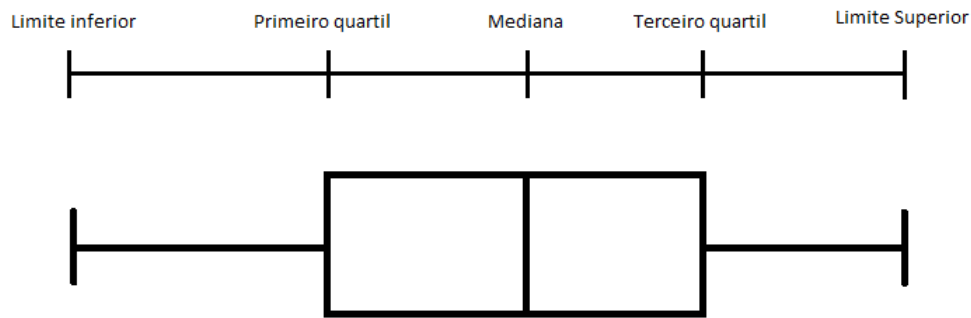


Figura 4.7: Exemplo de utilização de gráfico *boxplot*

Em gráficos do tipo *boxplot* a análise é realizada de acordo com a Figura 4.8.

Figura 4.8: Leitura do *boxplot*

Valores estudados que não pertencem ao intervalo entre o limite inferior e o limite superior são *outliers*.

4.4.2 Histogramas

Histogramas também são muito utilizados em análises estatísticas. Eles são uma ferramenta essencial quando se deseja estudar a distribuição dos dados existentes. A Figura 4.9 mostra um exemplo de histograma criado ao longo das análises.

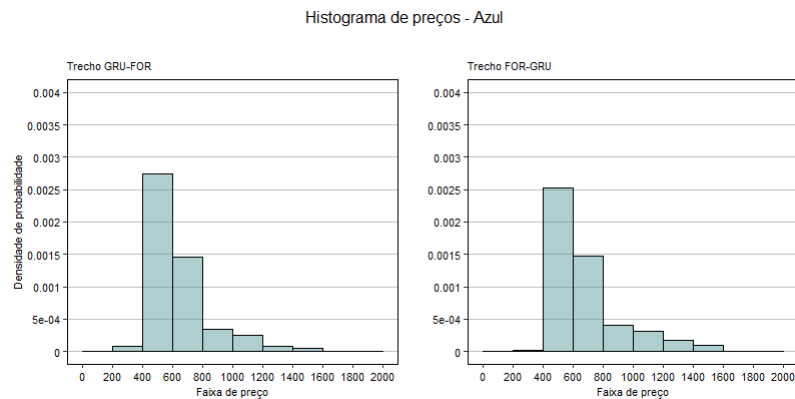


Figura 4.9: Exemplo de histograma

4.4.3 Gráfico de linha

Gráficos de linha são muito recorrentes para análises temporais. Nesse tipo de gráfico é possível se observar a variação de uma dada variável em função do tempo estudado. Nas análises deste trabalho esse tipo de gráfico foi criado relacionando a média dos preços das companhias com a antecedência da compra de passagem. O gráfico 4.5 mostra um exemplo de gráfico de linha.

Além das análises de gráficos, foram realizadas, a partir dos dados coletados, previsões de quando é o melhor momento para se comprar uma passagem, atingindo-se, assim, o objetivo específico 4. Para isso, foi utilizada a combinação de dois métodos de *data mining*: o *Random Forest* e a Regressão linear.

4.4.4 *Random Forest*

O *Random Forest* é um dos algoritmos de treinamento supervisionado mais utilizados tanto para tarefas de classificação, quanto para tarefas de predição [17]. O funcionamento desse algoritmo se dá com a criação de “floresta” aleatória de árvores de decisão. A partir das árvores criadas, realiza-se uma combinação dos resultados trazidos por cada árvore para, então, se obter um resultado final. Essa combinação entre as árvores é o que faz com que o método tenha, em média, previsões com maior acurácia e maior estabilidade que a maioria dos métodos existentes em *machine learning*.

O *Random Forest* foi utilizado para se realizar a previsão de qual faixa de preço uma dada configuração de parâmetros traria. Além disso, foi elaborado um *random forest* para cada companhia aérea. Os campos utilizados para essa classificação foram:

- Antecedência
- Horário da busca
- Nome do dia da semana na data da viagem buscada
- Trecho
- Mês da viagem
- Se o dia da viagem é um dia normal, feriado etc

4.4.5 *Regressão linear*

O outro método utilizado na previsão dos preços de passagem foi o da regressão linear. A regressão linear tenta estimar o valor de uma variável dependente em função variáveis independentes. No caso das estimativas realizadas nesse trabalho, quase todas as variáveis são categóricas, ou seja, não são numéricas. Por isso, foi necessário criar *dummies* para cada uma dessas variáveis. A equação geral para a regressão é:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \dots + \beta_n \cdot x_n \quad (4.5)$$

no qual y é a variável independente; β_0 é o coeficiente linear da equação, β_i , com $i \in [1, n]$ e $n \in \mathbb{Z}$, são coeficientes que são calculados representando cada uma das variáveis e x_i , com $i \in [1, n]$ e $n \in \mathbb{Z}$, representa o valor das variáveis do modelo. A Tabela 4.4 mostra o que o y e cada um dos β_i representam na equação.

Classe A	Variável
y	Preço
β_0	(Intercept)
β_1	TrechoGRU-FOR
β_2	Horário02-00
β_3	Horário04-00
β_4	Horário06-00
β_5	Horário08-00
β_6	Horário10-00
β_7	Horário12-00
β_8	Horário14-00
β_9	Horário16-00
β_{10}	Horário18-00
β_{11}	Horário20-00
β_{12}	Horário22-00
β_{13}	Antecedencia
β_{14}	Mês_Viagemdezembro
β_{15}	Mês_Viagemjulho
β_{16}	Mês_Viagemnovembro
β_{17}	Mês_Viagemoutubro
β_{18}	Mês_Viagemsetembro
β_{19}	Nome_Diasegunda-feira
β_{20}	Nome_Diaterça-feira
β_{21}	Nome_Diaquarta-feira
β_{22}	Nome_Diaquinta-feira
β_{23}	Nome_Diasexta-feira
β_{24}	Nome_Diasábado
β_{25}	AtributoEleições
β_{26}	AtributoFeriado
β_{27}	AtributoNormal
β_{28}	AtributoPós-Comemoração
β_{29}	AtributoPós-Feriado
β_{30}	AtributoVéspera de comemoração
β_{31}	AtributoVéspera de Feriado
β_{32}	AtributoVéspera de Eleições 2018

Tabela 4.4: Significados dos coeficientes da regressão

As variáveis “Trecho”, “Horário”, “Mês_Viagem”, “Nome_Dia” e “Atributo” são variáveis categóricas e, por isso, são caracterizadas na regressão por meio de *dummies*, ou seja, cria-se $n - 1$ coeficientes, nas quais, n é o número de categorias que cada variável possui. Para representar um valor categórico, as variáveis *dummy* possuem a propriedade de que seus x_i só podem assumir valores de 0 ou 1 e, além disso, a soma de seus x_i não

pode ser maior que 1. Ou seja, por exemplo, para se representar o trecho GRU-FOR, a *dummy* “trechoGRU-FOR” deve ter seu $x_1 = 1$, enquanto que, para se representar o trecho FOR-GRU deve-se ter $x_1 = 0$ (Considerando-se os índices da Tabela 4.4).

4.5 Comparação entre previsões e histórico

Foram realizadas várias previsões e comparações com casos simulados. A partir disso foi verificado o quanto de economia seria obtido em cada previsão. Com base na economia em cada caso, calculou-se a economia média geral nas compras simuladas a partir das análises. A partir disso, foi possível atingir o objetivo específico 5.

Capítulo 5

Análises de resultados

Este capítulo foi destinado para a análise dos resultados obtidos na tentativa de se alcançar os objetivos listados no item 1.3 deste trabalho.

5.1 Web-scraping em R

O *script* desenvolvido para o início dos estudos sobre os preços dos voos em cada um dos trechos (GRU-FOR e FOR-GRU) foi executado por dois meses (julho e agosto), a cada duas horas, sempre coletando os menores preços de cada uma das principais companhias aéreas que atuam no setor aéreo brasileiro nos próximos 90 dias desde a data da busca. Com isso, foi obtida uma base de dados, que, após diversos tratamentos em seus dados, contém 281876 linhas com as seguintes informações:

- Preço do voo
- Data da extração do conjunto de dados
- Data de uma suposta viagem
- Trecho buscado
- Horário da busca
- Antecedência da busca
- Mês de uma suposta viagem
- Nome do dia buscado

- Se a data de busca é no fim de semana ou na semana
- O atributo do dia, ou seja, se é um dia normal, um feriado etc.
- Empresa aérea

A Tabela 4.1 mostra um exemplo de como base de dados foi extraída.

5.2 Estudo de variação de preços

5.2.1 Relações entre variáveis

Para se iniciar os estudos de variações de preços das companhias aéreas foram construídos alguns gráficos relacionando variáveis obtidas com os preços, para cada uma das empresas analisadas. O objetivo desses gráficos é analisar se existe alguma relação explícita entre as variáveis e o preço. A tabela do Anexo C, contida no CD que faz parte desse trabalho, é a tabela de onde todas as análises desta seção são retiradas.

Foram criados diversos gráficos relacionando as diversas variáveis com os preços coletados e as variáveis cuja dependência com o preço se mostrou mais visível com os gráficos criados foram a “antecedência da busca” e o “atributo do dia”. Por esse motivo, os gráficos criados com essas variáveis serão expostos no decorrer deste item, para cada empresa aérea e para cada um dos trechos estudados.

Os primeiros gráficos expostos serão *boxplots* dos preços em relação as variáveis supracitadas. Esse tipo de gráfico serve para se ter uma noção estatística entre variáveis, tendo-se uma visão da quantidade de *outliers*, da localização da mediana dos valores de preços e a sua variação de acordo com a variável estudada e a variação dos mesmos dentro de cada valor de variável.

As Figuras 5.1 e 5.2 são referentes aos preços das taxas da empresa aérea Azul relacionadas com a antecedência de busca.

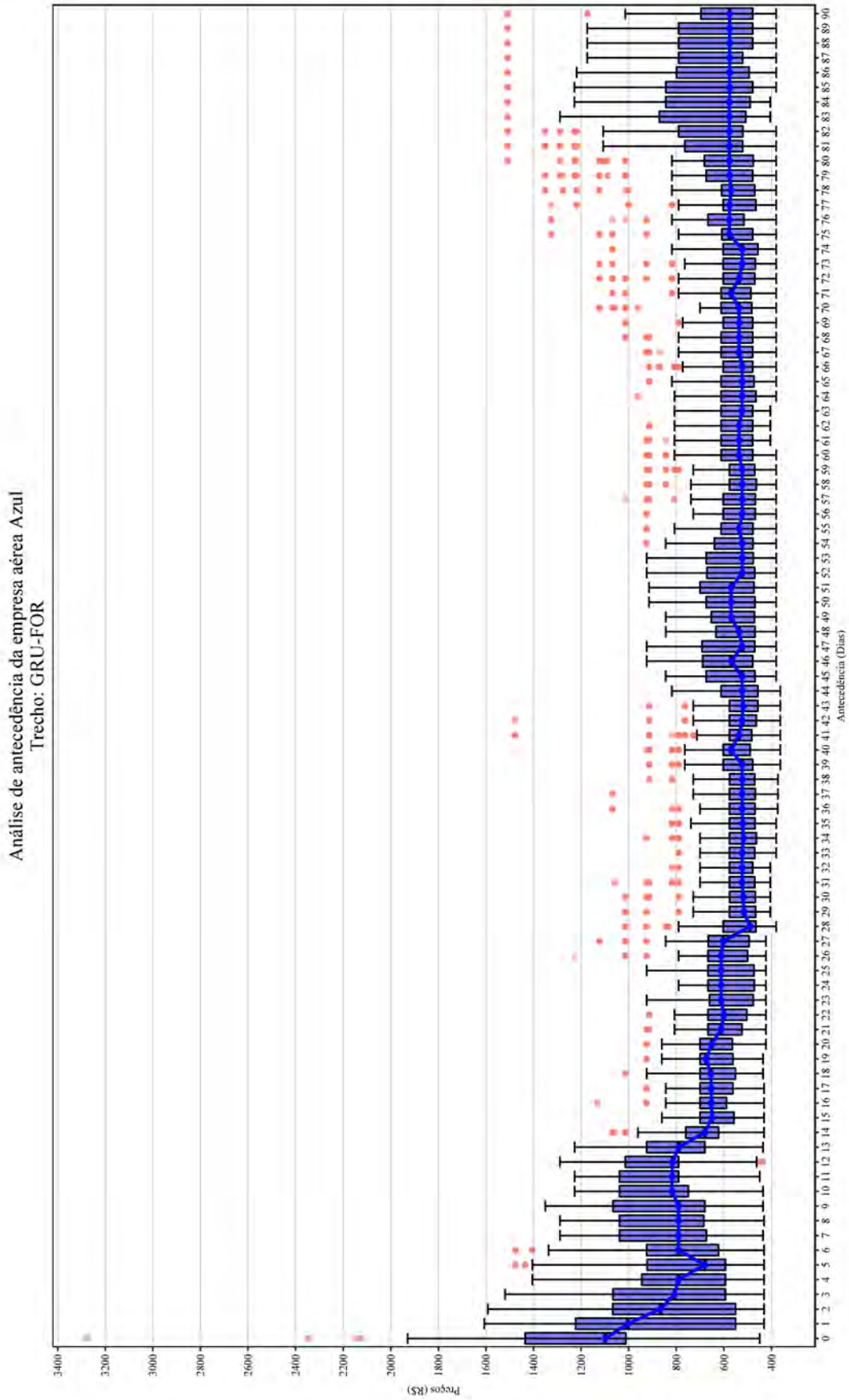


Figura 5.1: Relação entre antecedência e preços da empresa Azul no trecho GRU-FOR

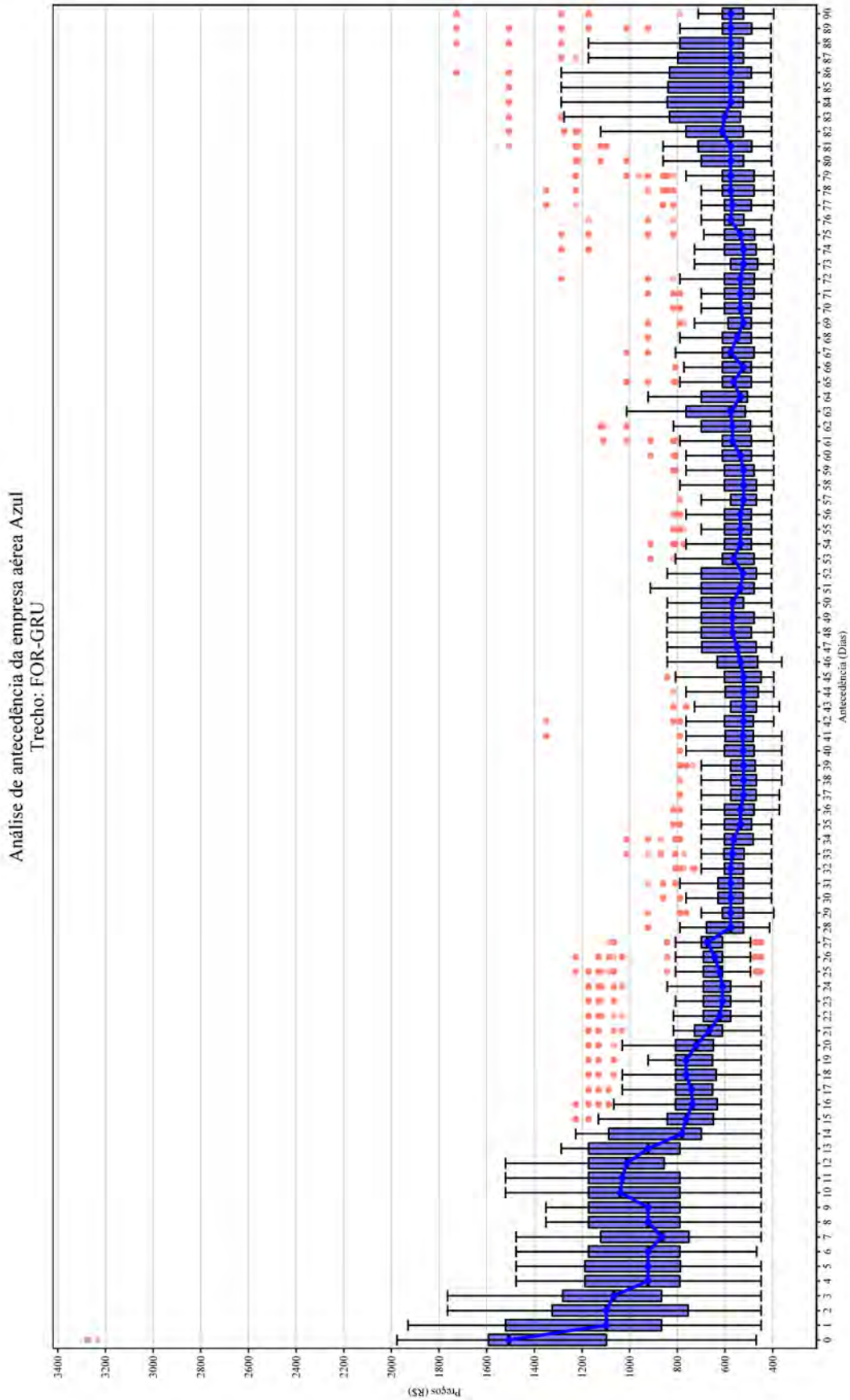


Figura 5.2: Relação entre antecedência e preços da empresa Azul no trecho FOR-GRU

Por meio da análise dos gráficos de *boxplots* referentes a Azul é possível se ter as seguintes conclusões:

1. As diferenças entre os trechos são mínimas, ou seja, os trechos podem ser considerados iguais para o período estudado.
2. Passagens observadas dentro de uma antecedência de até 13 dias da compra ou com uma antecedência entre 80 e 90 dias têm uma variância muito alta, ou seja, são mais difíceis de se prever e apresentam uma amplitude muito grande.
3. Enquanto isso, passagens com 14 e 80 dias de antecedência, em geral, possuem uma variância reduzida, ou seja, não possuem uma grande amplitude, e podem ser previstas com uma probabilidade menor de erro.
4. Além disso, também é possível perceber que, passagens observadas com uma antecedência dentro do intervalo de 28 a 45 dias tendem a ter um valor reduzido em seu valor, ou seja, teoricamente, para essa empresa, seus menores preços estarão em passagens compradas dentro desse período de antecedência.
5. Por outro lado, passagens observadas com uma antecedência de zero a treze dias tendem a ter seus preços valores mais elevados que o geral.

As Figuras 5.3 e 5.4 abaixo mostram os gráficos com a análise de *boxplots* para a empresa aérea Avianca.

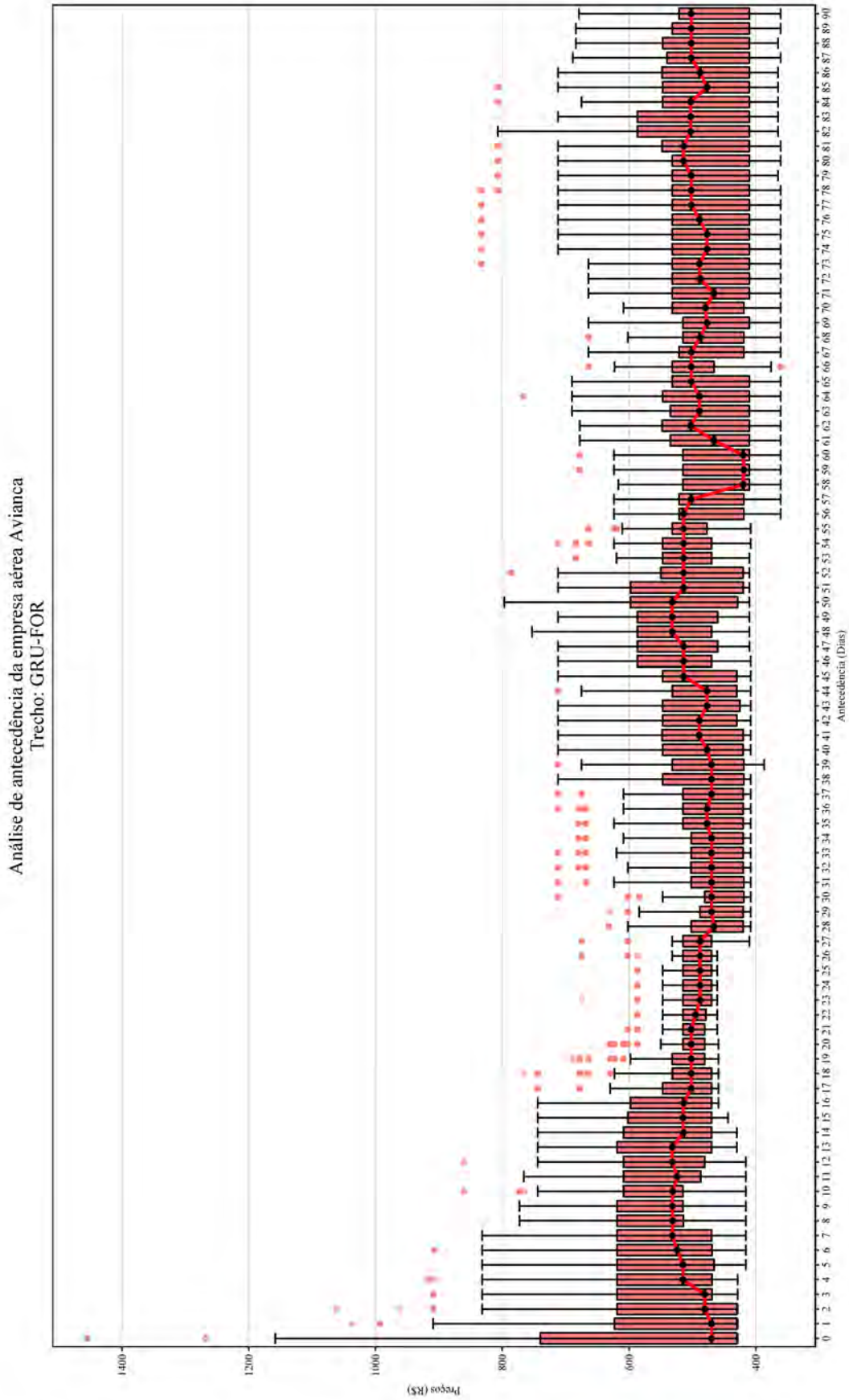


Figura 5.3: Relação entre antecedência e preços da empresa Avianca no trecho GRU-FOR

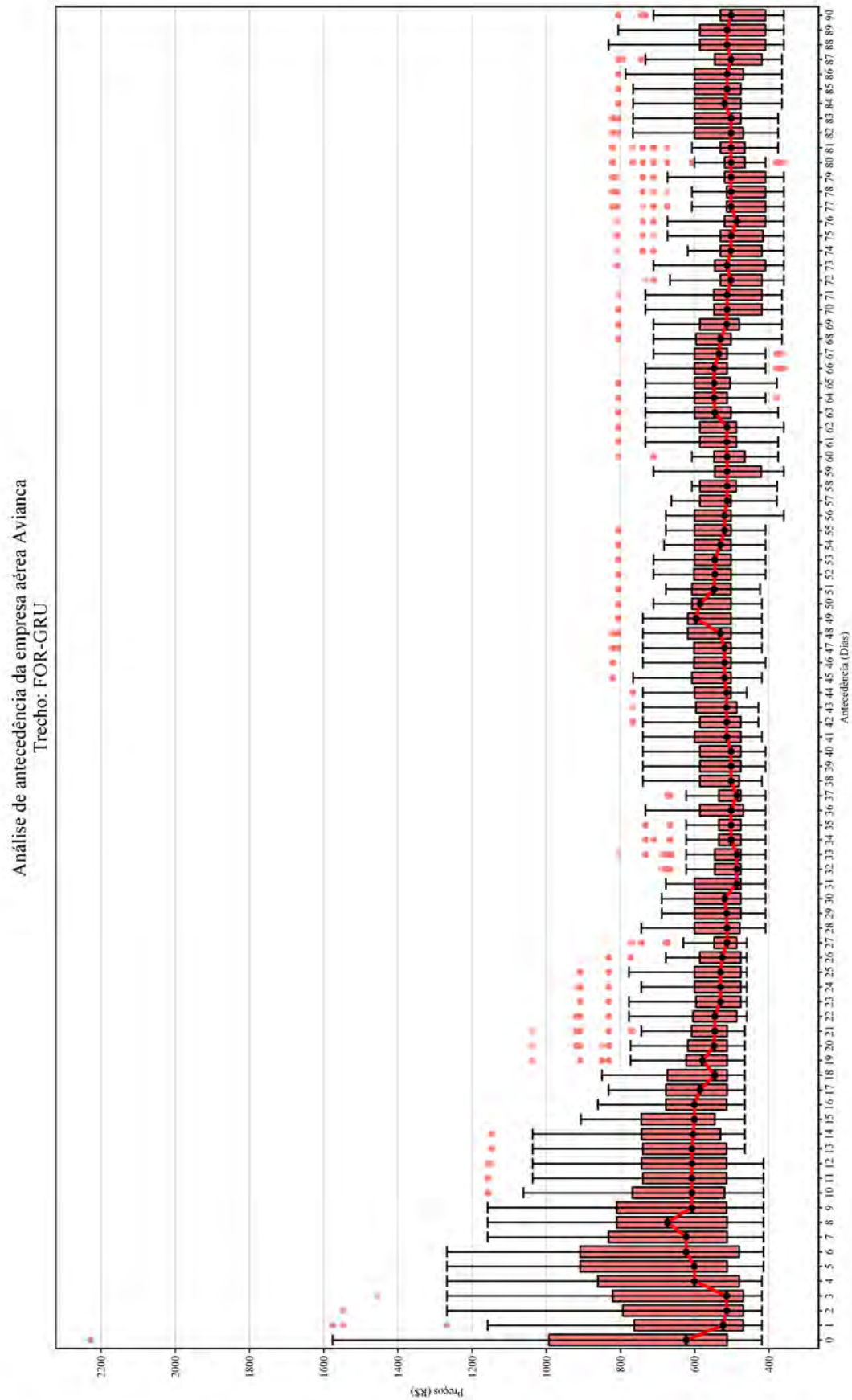


Figura 5.4: Relação entre antecedência e preços da empresa Avianca no trecho FOR-GRU

Na análise dos gráficos percebe-se as possíveis proposições:

1. Os trechos possuem uma leve diferença entre si: o trecho GRU-FOR possui uma variação média alta ao longo dos dias de antecedência, enquanto que o trecho FOR-GRU essa tem um intervalo em que a variação dos valores dos preços é maior que o valor médio para a empresa.
2. Para o trecho FOR-GRU, o intervalo de 0 a 15 dias possui uma variação média mais elevada que o restante do período analisado, enquanto que para um período de antecedência maior a variação se torna menor.
3. No trecho GRU-FOR o período entre 20 e 27 dias de antecedência é o período com menor variação nos preços, portanto um período com uma maior assertividade.

As Figuras 5.5 e 5.6 mostram as análises para a empresa Gol.

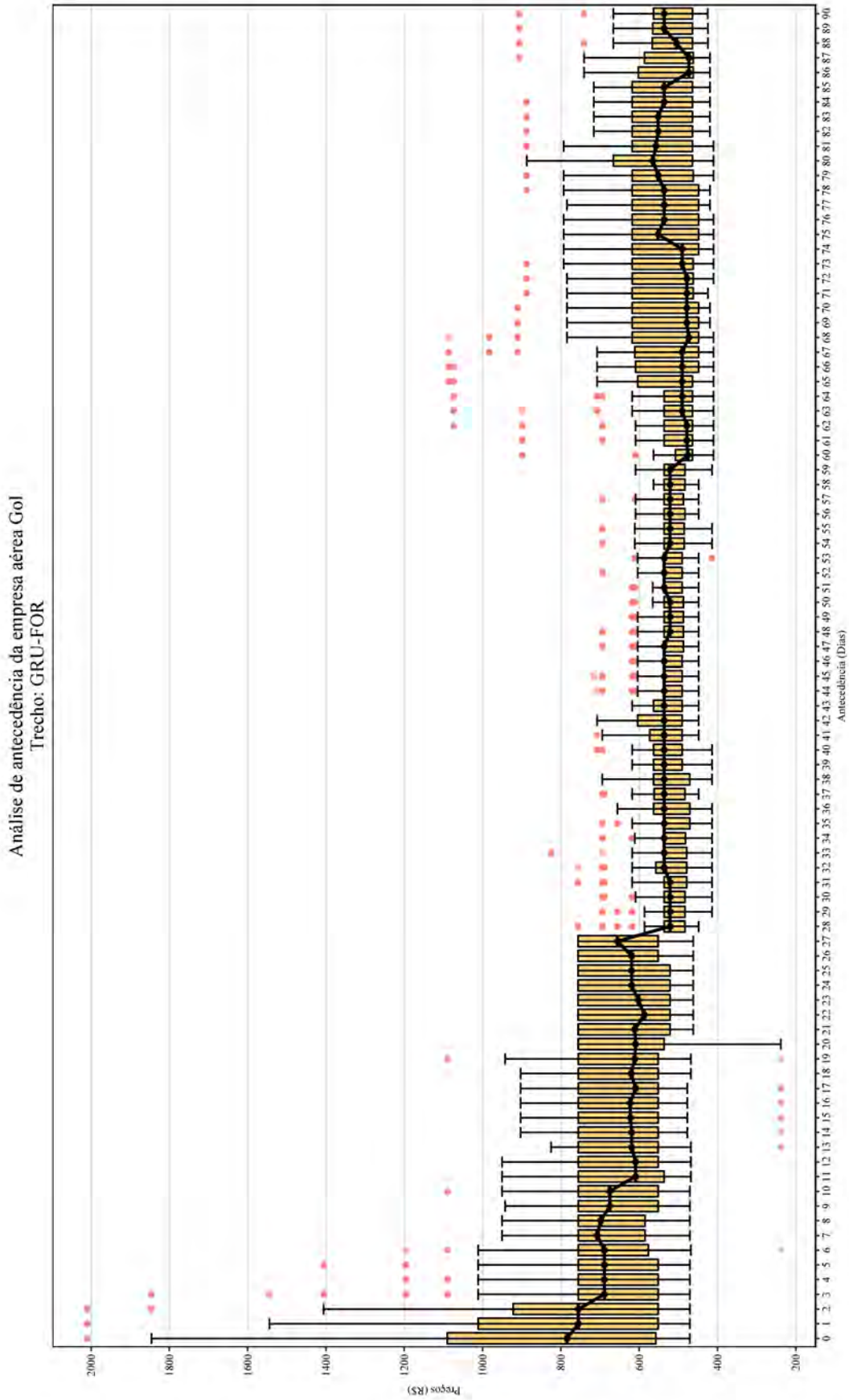


Figura 5.5: Relação entre antecedência e preços da empresa Gol no trecho GRU-FOR

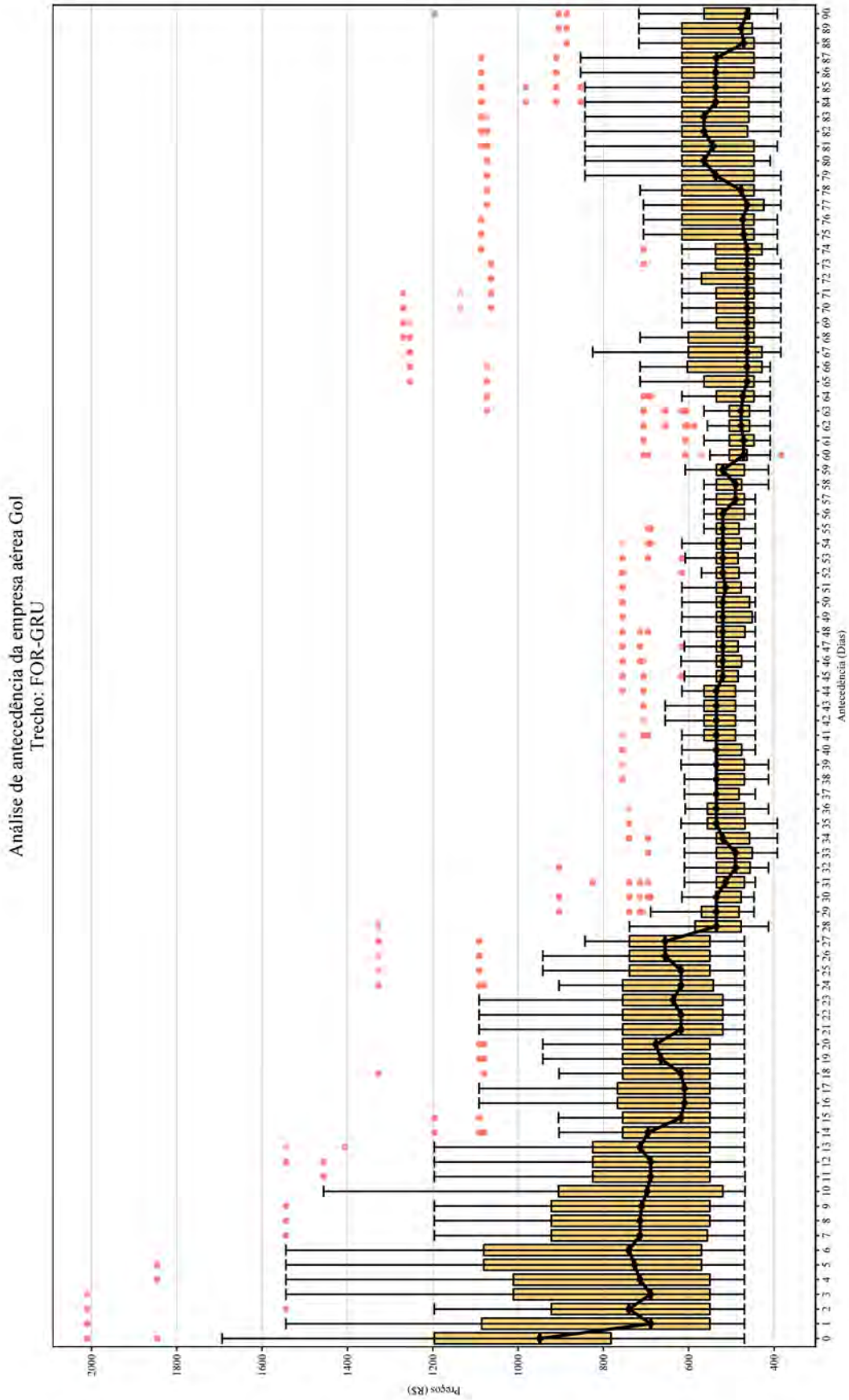


Figura 5.6: Relação entre antecedência e preços da empresa Gol no trecho FOR-GRU

As análises dos Gráficos 5.5 e 5.6 permite as seguintes proposições:

1. Os trechos não apresentam muitas diferenças entre si. Os períodos de grandes variações nos preços das passagens são aproximadamente os mesmos.
2. Para um intervalo entre 0 e 27 dias as variações dos preços são bastante alto.
3. Enquanto isso, os preços das passagens apresentam uma variação reduzida no intervalo entre 28 e 65 dias de antecedência.

A conclusão das análises dos *boxplots* dos preços das empresas aéreas em relação à antecedência da compra se faz com a companhia aérea LATAM.

As Figuras 5.7 e 5.8 ilustram os valores obtidos.

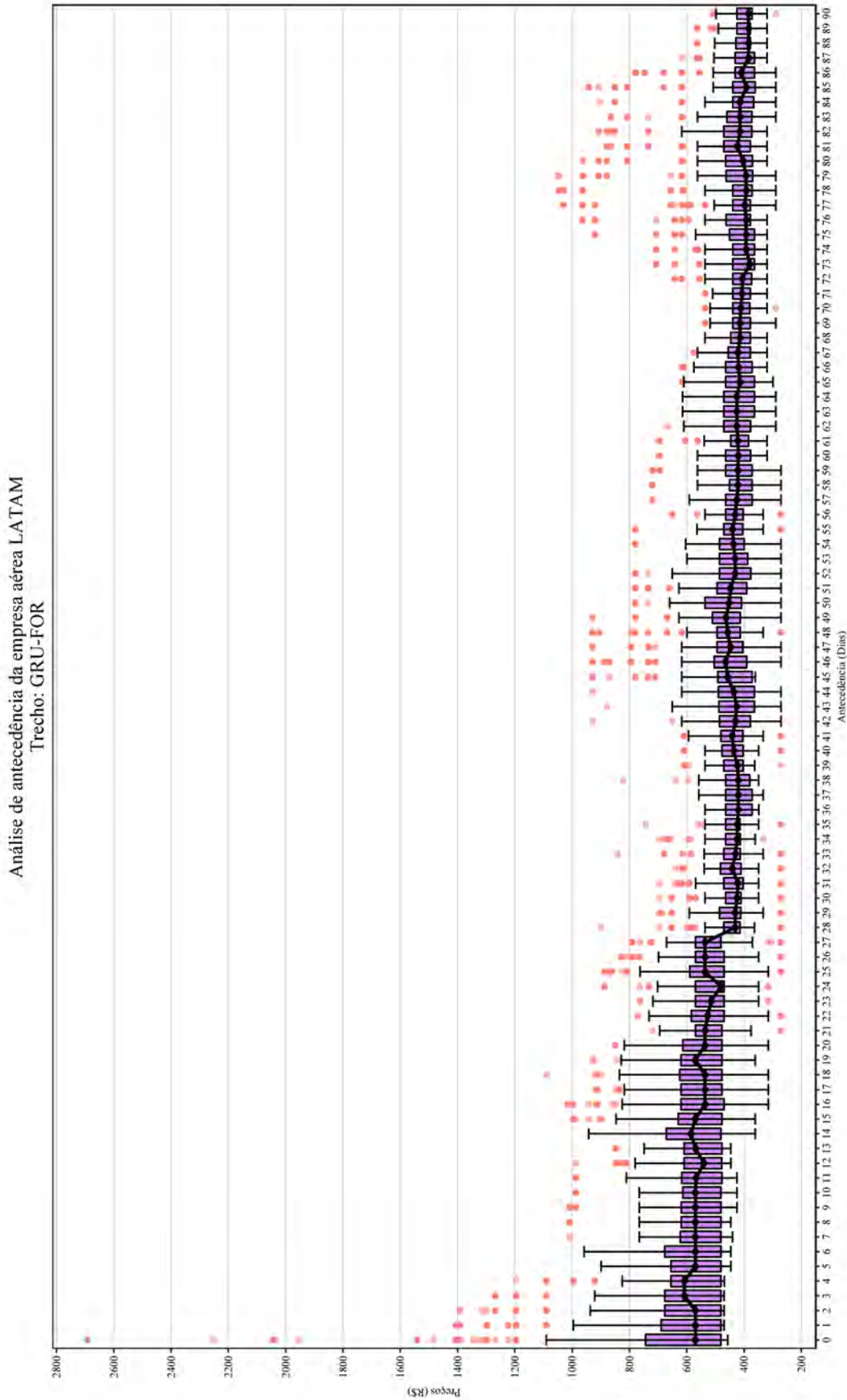


Figura 5.7: Relação entre antecedência e preços da empresa LATAM no trecho GRU-FOR

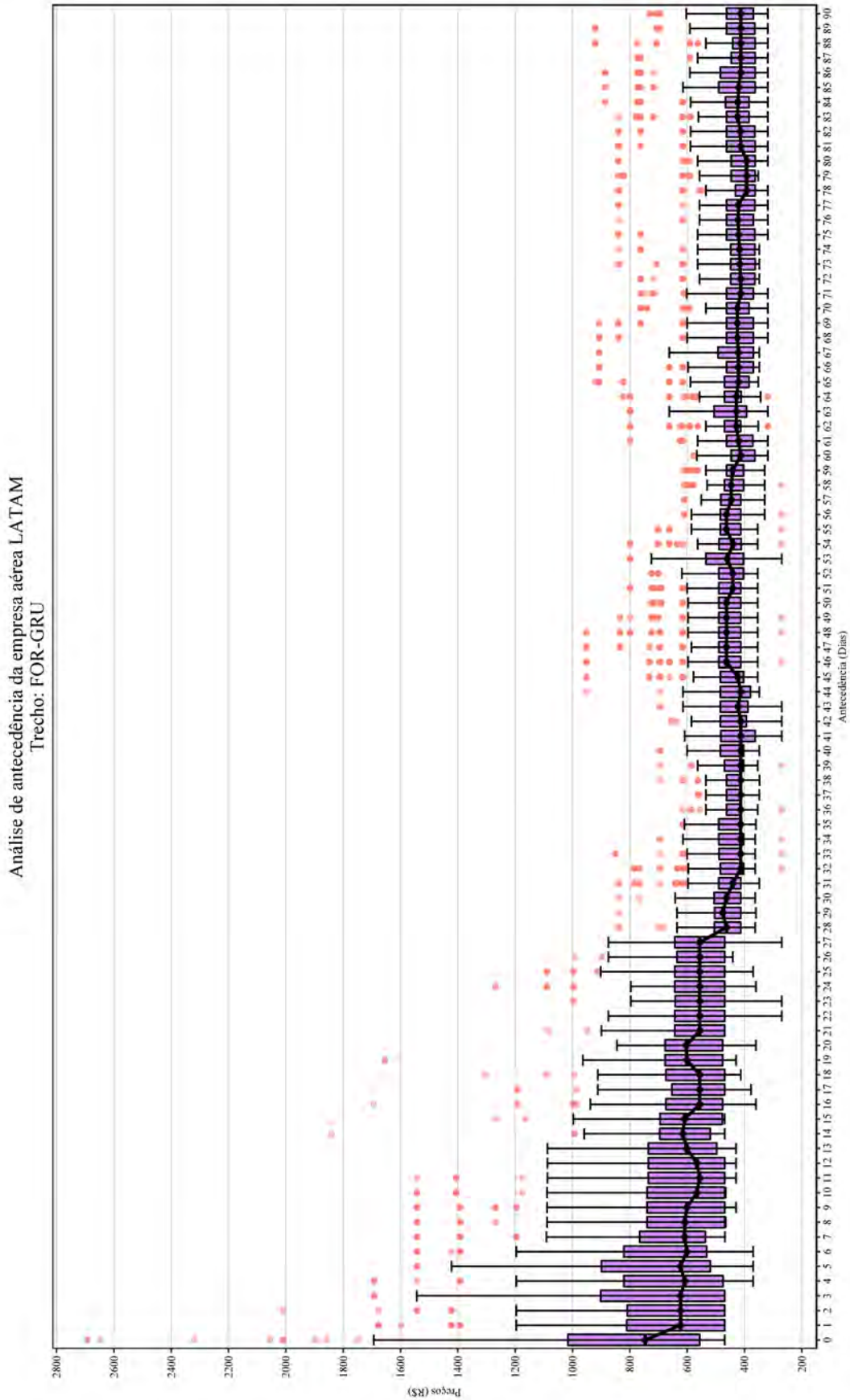


Figura 5.8: Relação entre antecedência e preços da empresa LATAM no trecho FOR-GRU

As análises realizadas a partir dos Gráficos 5.7 e 5.8 permitem concluir que:

1. Os trechos estudados não apresentam grandes diferenças entre si.
2. O intervalo de 0 a 27 dias de antecedência apresentam as maiores variações.
3. O intervalo de 28 a 90 dias apresentam variações muito pequenas.

Concluindo-se a análise dos *boxplots* das empresas aéreas, é possível se dizer que todas as empresas se comportam de forma parecida nos trechos analisados, excetuando-se a Avianca. Além disso, pode-se dizer que a Azul é a empresa aérea que tem uma maior variação no patamar da mediana nos primeiros 15 dias de antecedência. A Avianca e a LATAM são as empresas que possuem mediana menos variável ao longo do tempo.

A partir dessa análise é possível se ter ideia de que a antecedência é uma variável que influencia bastante o preço das passagens. Ademais, para se ter maior precisão na compra de passagens, deve se pesquisar com no mínimo 27 dias de antecedência, em média.

A próxima variável a ser relacionada com os preços será o atributo do dia. Os Diagramas 5.9 e 5.10 ilustram as análises relativas a companhia aérea Azul.

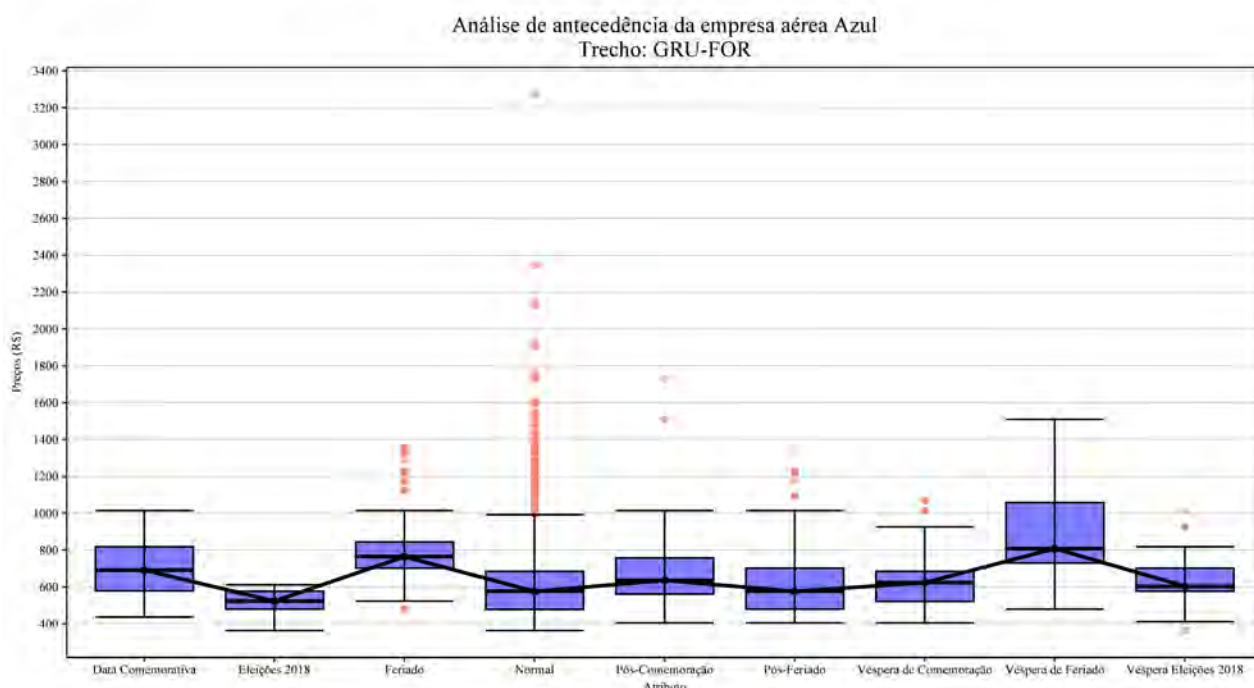


Figura 5.9: Relação entre atributo do dia e preços da empresa Azul no trecho GRU-FOR.

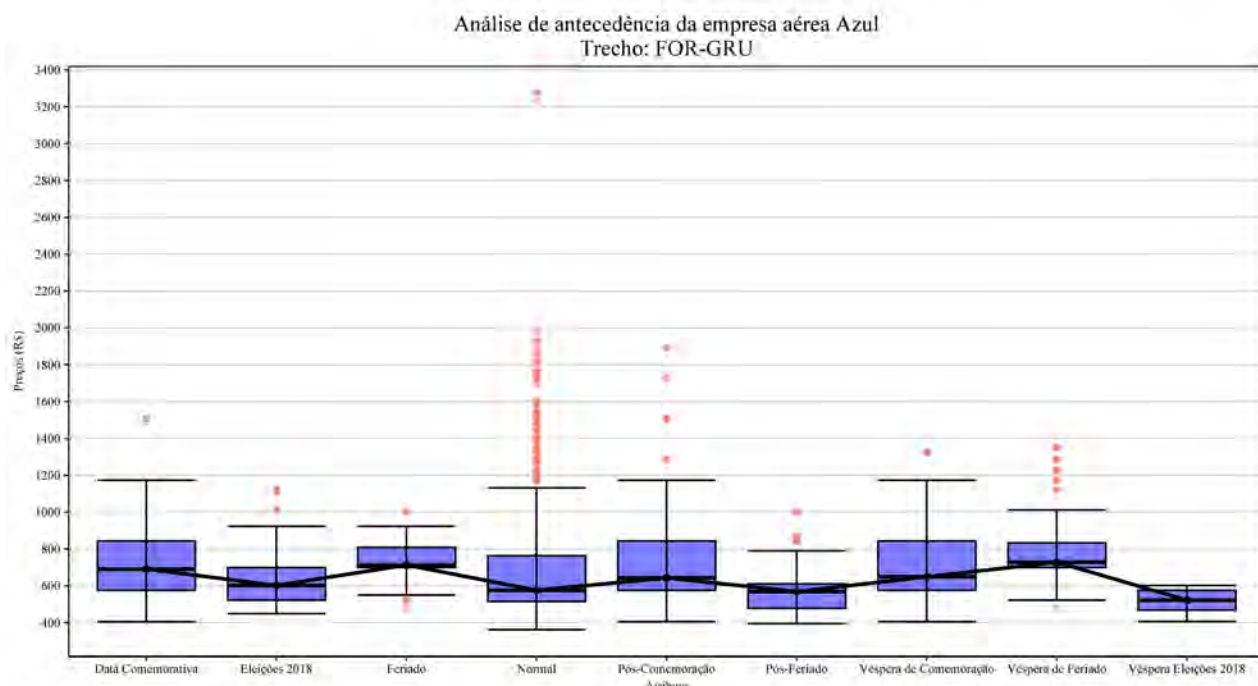


Figura 5.10: Relação entre atributo do dia e preços da empresa Azul no trecho FOR-GRU

É possível observar por meio dos Gráficos 5.9 e 5.10 que, em ambos os trechos, os preços parecem se comportar de maneira semelhante. Também se percebe que a maioria dos *outliers* acontecem quando o “atributo” é classificado com “normal”, o que ocorre, principalmente, pela enorme quantidade de dados que existem classificados dessa maneira na base de dados obtida.

Outro ponto que de atenção que pode ser notado por intermédio dos diagramas é que passagens buscadas para dias que são feriados ou vésperas de feriados tendem a ser, realmente, mais caras que dias normais. Por outro lado, passagens buscadas para dias que são pós-feriados não apresentam nenhuma elevação aparente nos preços.

Outra observação relacionada às vésperas de feriados, é que o trecho de GRU-FOR apresenta uma variação consideravelmente maior que o trecho FOR-GRU, tendo-se, assim, uma possibilidade maior de se encontrarem preços ainda mais elevados no momento da busca. Isso deve acontecer pelo fato que que Fortaleza é uma cidade turística e que, devido a isso, acaba sendo muito procurada em períodos de feriados.

As Figuras 5.11 e 5.12 expõem a análises realizadas para a Avianca.

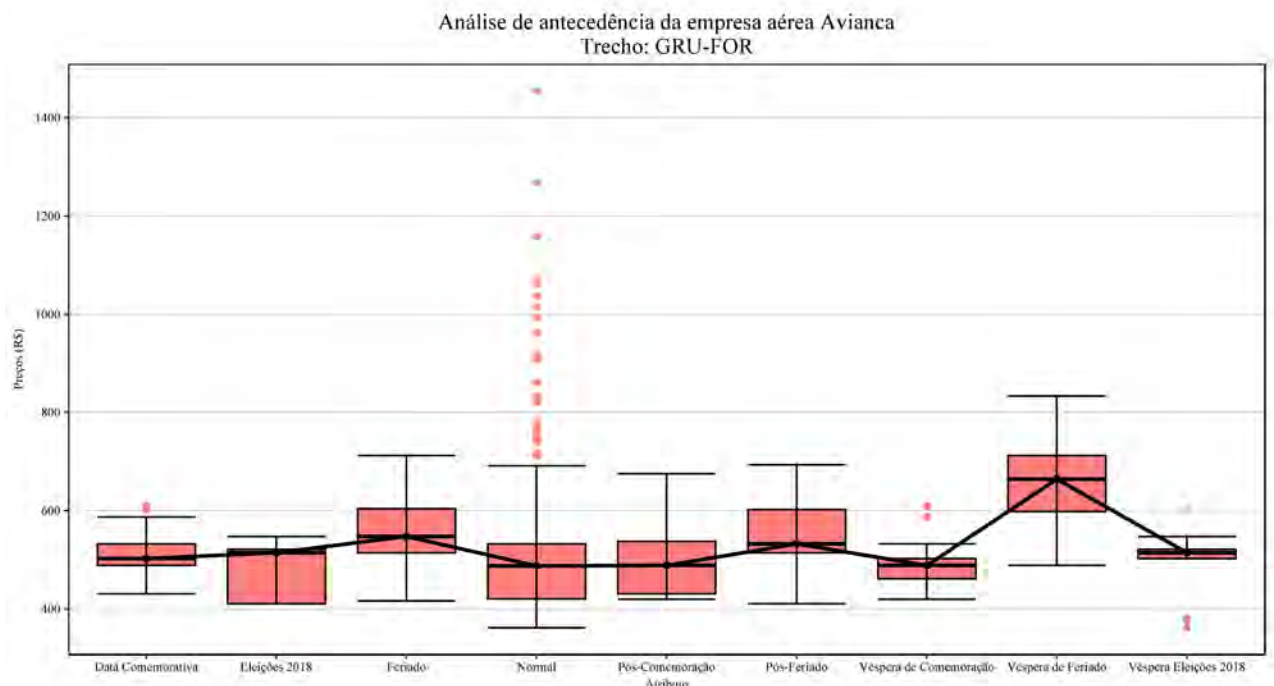


Figura 5.11: Relação entre atributo do dia e preços da empresa Avianca no trecho GRU-FOR

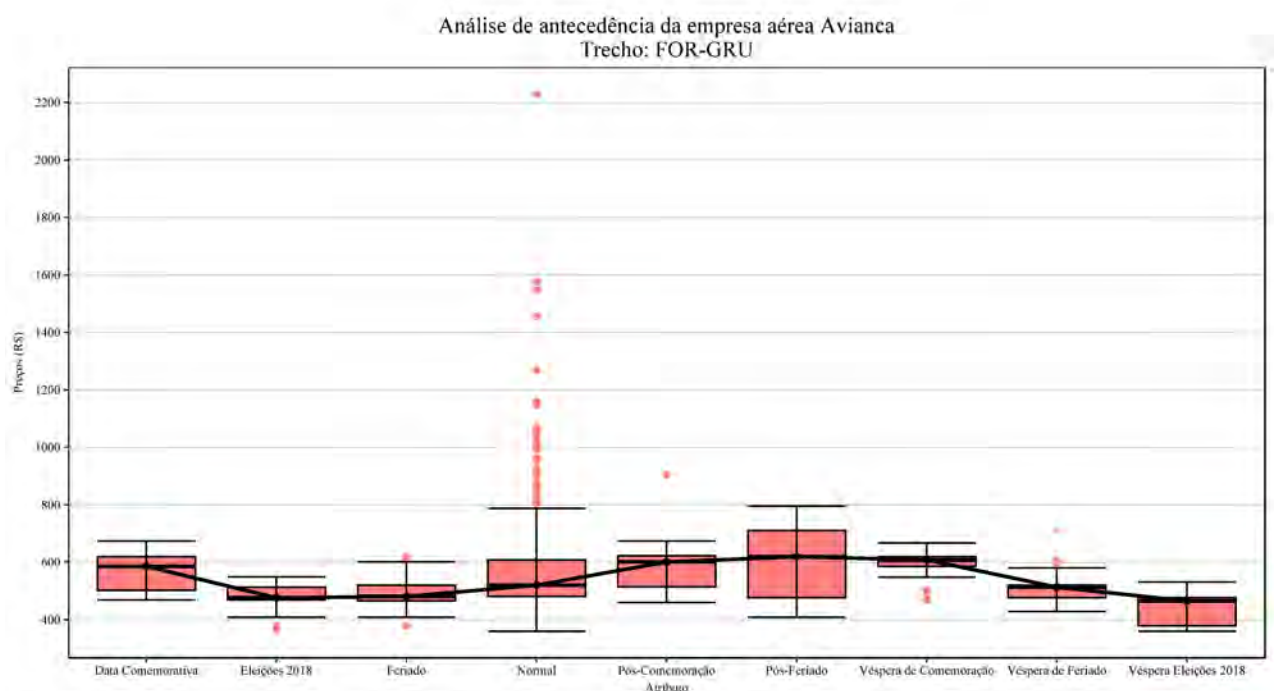


Figura 5.12: Relação entre atributo do dia e preços da empresa Avianca no trecho FOR-GRU

Para o caso da Avianca, percebe-se de forma mais explícita a diferença no patamar dos preços em vésperas de feriado dos preços em dias normais, no trecho de GRU-FOR. Porém, no trecho FOR-GRU, não existe muita diferença entre dias normais e vésperas de

feriado. Isso é mais um indício de que o fator turístico de Fortaleza seja um fator que é levado em conta no gerenciamento de recursos das empresas aéreas.

Outro ponto importante que se pode obter a partir dos gráficos é o de que, para a Avianca, passagens com data para pós-feriados apresentam um valor maior em passagens no trecho FOR-GRU, mostrando que a empresa tenta aproveitar o feriado no trecho GRU-FOR e o pós-feriado no trecho FOR-GRU aumentando seus preços nesses períodos em que os clientes estão mais dispostos a gastar.

Os Gráficos 5.13 e 5.14 mostram as relações para a empresa aérea Gol.

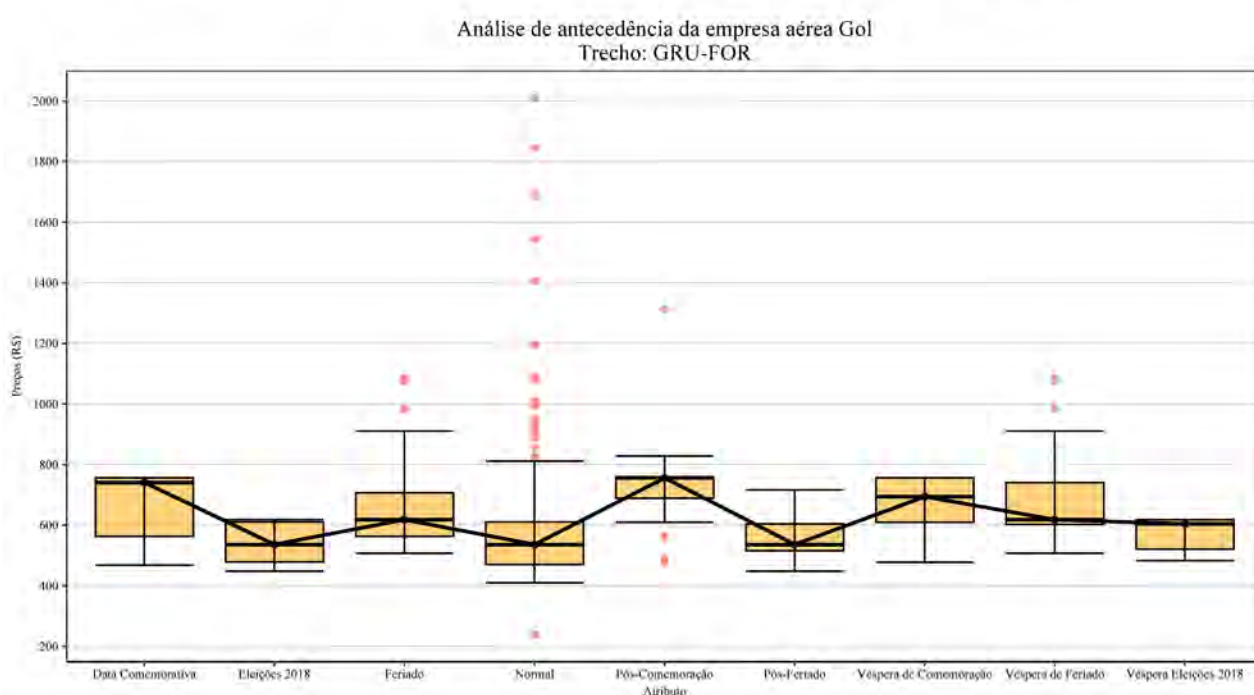


Figura 5.13: Relação entre atributo do dia e preços da empresa Gol no trecho GRU-FOR

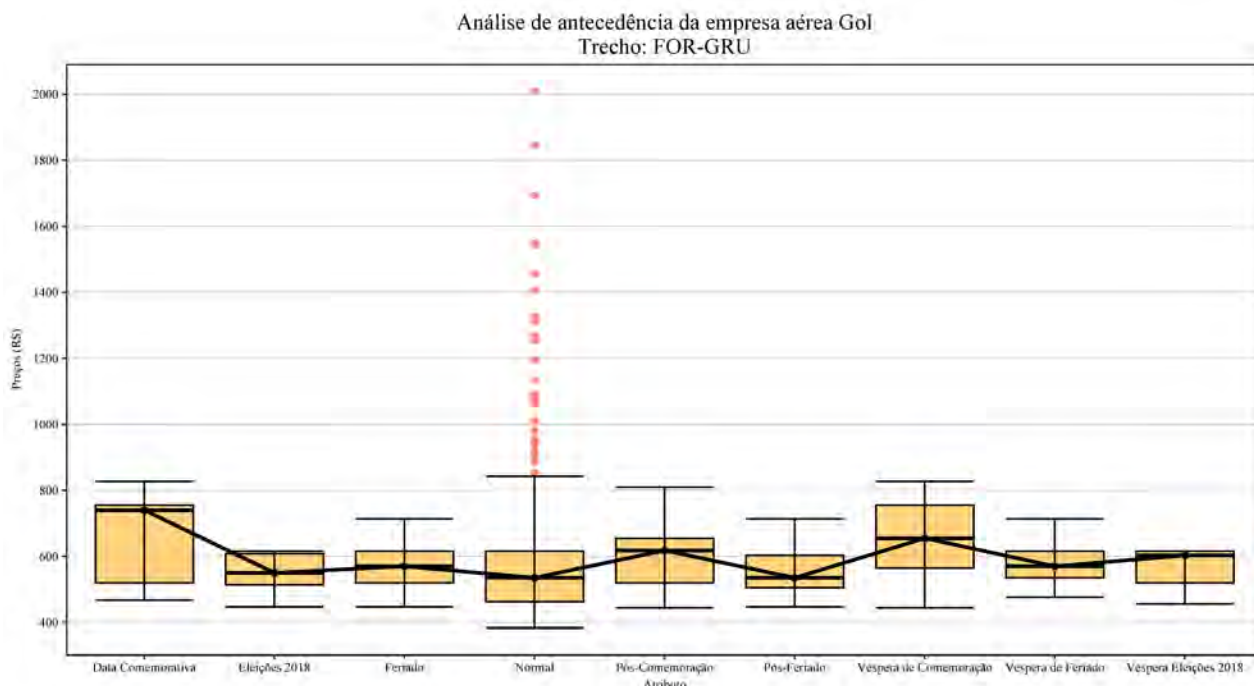


Figura 5.14: Relação entre atributo do dia e preços da empresa Gol no trecho FOR-GRU

Percebe-se que, diferentemente das empresas analisadas anteriormente, feriados e vésperas de feriados parecem não se destacar nos valores de seus preços, quando comparados com as outras categorizações de atributos, para ambos os trechos estudados. Porém, pode-se observar que, para o trecho de GRU-FOR quase todos os atributos de dia de viagem são maiores que datas com atributos normais, quando se analisa a mediana dos preços e o seu nível de variação. Enquanto isso, no trecho FOR-GRU, o atributo do dia não parece ser um fator decisivo na precificação da passagem, em geral, excetuando-se datas comemorativas e vésperas de datas comemorativas.

Por fim, completando as análises entre preço e atributo do dia, os gráficos 5.15 e 5.16 mostram os resultados obtidos para a empresa aérea LATAM.

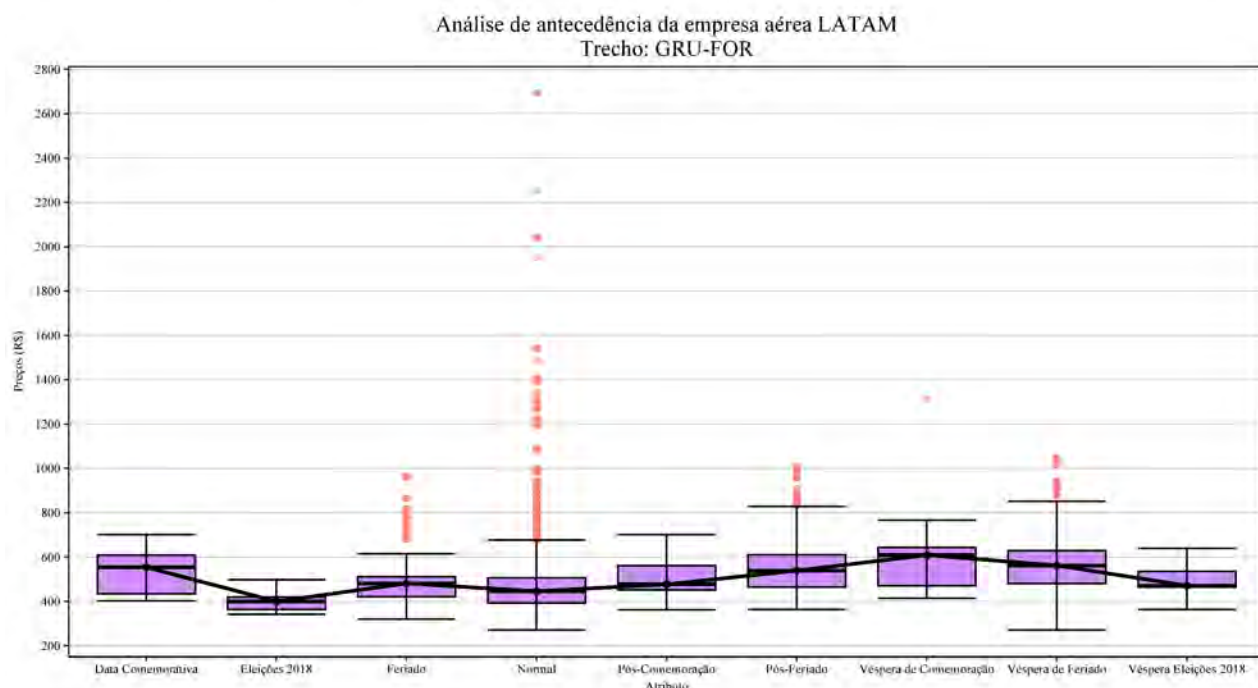


Figura 5.15: Relação entre atributo do dia e preços da empresa LATAM no trecho GRU-FOR

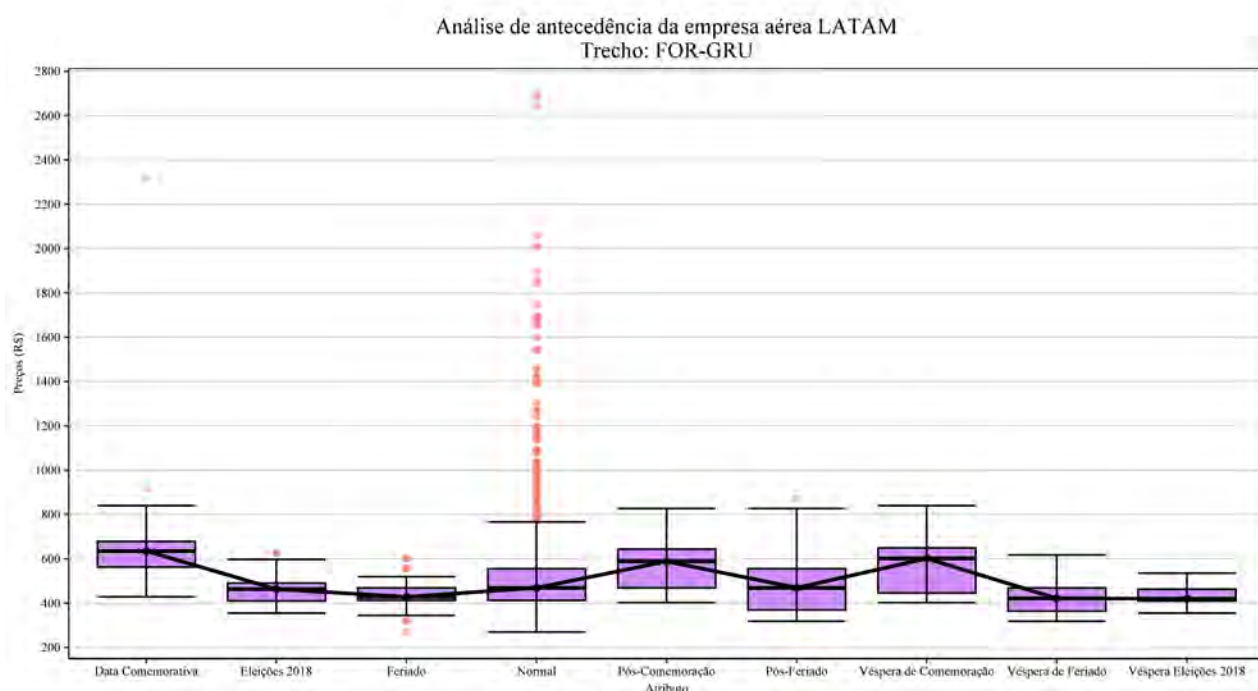


Figura 5.16: Relação entre atributo do dia e preços da empresa LATAM no trecho FOR-GRU

Para a LATAM, observa-se que não há grandes variações entre os dois trechos estudados. Também se nota que feriados e vésperas de feriados parecem não ter grande influência no gerenciamento de receitas da empresa. Além disso, pode-se dizer que a

empresa apresenta um comportamento bem parecido para em todos os atributos categorizados, pois a mediana não varia muito de uma categoria para outra e a variação nos *boxplots* não é muito grande.

5.2.2 Parâmetros estatísticos

Neste tópico serão apresentadas a sumarização dos preços obtidos para cada empresa aérea, também serão analisadas se as médias de preços das empresas são estatisticamente iguais, além de comparações entre as médias de preços das empresas em função da antecedência de compra. Todos os dados utilizados foram sumarizados a partir da tabela contida no Anexo C.

5.2.2.1 Sumarização dos dados

A sumarização dos dados obtidos gerou os parâmetros estatísticos explicitados nas Tabelas 5.1 e 5.2 ilustradas a seguir, com valores para os trechos GRU-FOR e FOR-GRU, respectivamente.

Parâmetro	GRU-FOR			
	Avianca	Azul	Gol	LATAM
Mínimo	361.0	363.0	238.0	271.0
Mediana	502.0	577.0	536.0	463.0
Média	504.0	628.2	572.1	476.0
Máximo	1 455.0	3 276.0	2 010.0	2 693.0
Desvio Padrão	88.0	204.6	134.1	122.0

Tabela 5.1: Sumário dos dados relativos ao trecho GRU-FOR

Parâmetro	FOR-GRU			
	Avianca	Azul	Gol	LATAM
Mínimo	360.0	362.0	384.0	270.0
Mediana	520.0	599.0	535.0	469.0
Média	560.6	670.3	591.3	504.5
Máximo	2 229.0	3 276.0	2 010.0	2 693.0
Desvio Padrão	132.0	245.8	193.5	164.6

Tabela 5.2: Sumário dos dados relativos ao trecho FOR-GRU

Pode-se inferir, por meio dos parâmetros obtidos, que o trecho FOR-GRU é mais caro que o trecho GRU-FOR em torno de 7%, em média. Esse cálculo foi realizado de

acordo com a equação 6.

$$\mu_{trecho} = \frac{\mu_{Avianca} + \mu_{Azul} + \mu_{Gol} + \mu_{LATAM}}{4} \quad (5.1)$$

Outra inferência geral que se pode tomar pela análise das tabelas é a de que, aparentemente, as empresas possuem o mesmo piso e o mesmo teto para a precificação das passagens nos dois trechos.

Comparativamente, pode-se inferir que a Avianca é a empresa as menores taxas de variação de preços de passagens ($\sigma_{GRU-FOR} = 88.0$ e $\sigma_{FOR-GRU} = 132.0$); a Azul possui a maior média de preços, além de possuir também a maior taxa observada (R\$ 3 276.00) e as maiores taxa de variação de preço ($\sigma_{GRU-FOR} = 204.66$ e $\sigma_{FOR-GRU} = 245.8$) ao longo do período analisado; A Gol possui a menor taxa observada ao longo do período analisado no trecho GRU-FOR e a LATAM apresenta a menor média de preços ($\mu_{GRU-FOR} = 476.0$ e $\mu_{FOR-GRU} = 504.5$).

A fim de se analisar a distribuição dos preços das passagens aéreas, foram criados histogramas para cada uma das empresas e cada um dos trechos estudados. Com os histogramas é possível se ter noção de quais faixas de preços são as mais utilizadas por cada empresa. Os Gráficos 5.20, 5.17, 5.18, 5.19 e a seguir mostram os histogramas relativos as empresas Azul, Avianca, Gol e LATAM, respectivamente.

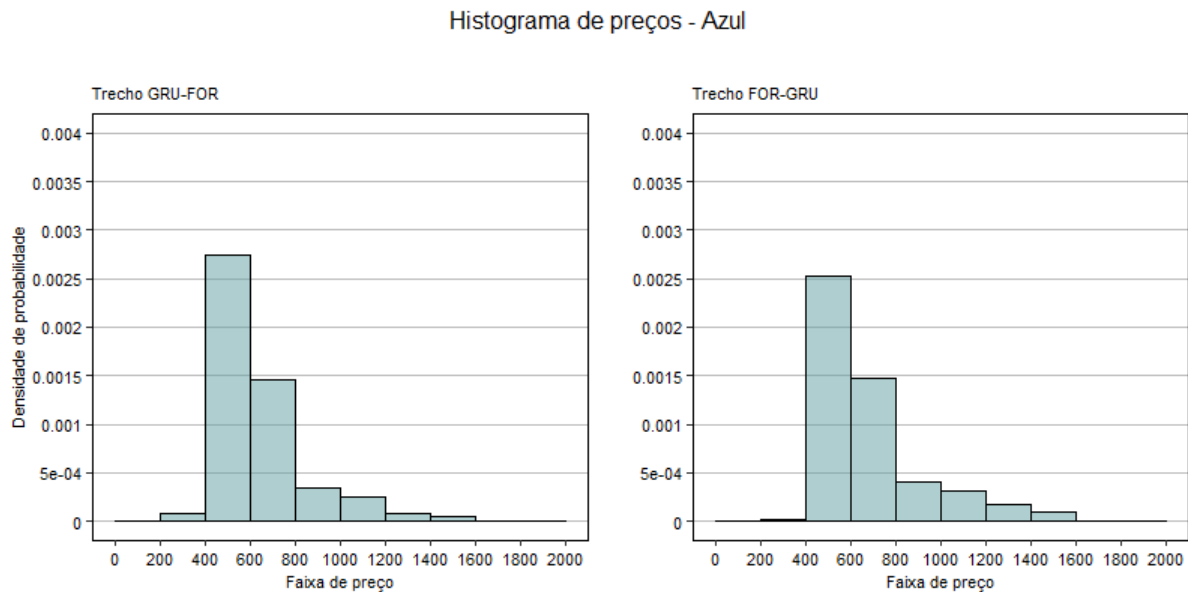


Figura 5.17: Histograma de precificação da empresa aérea Azul

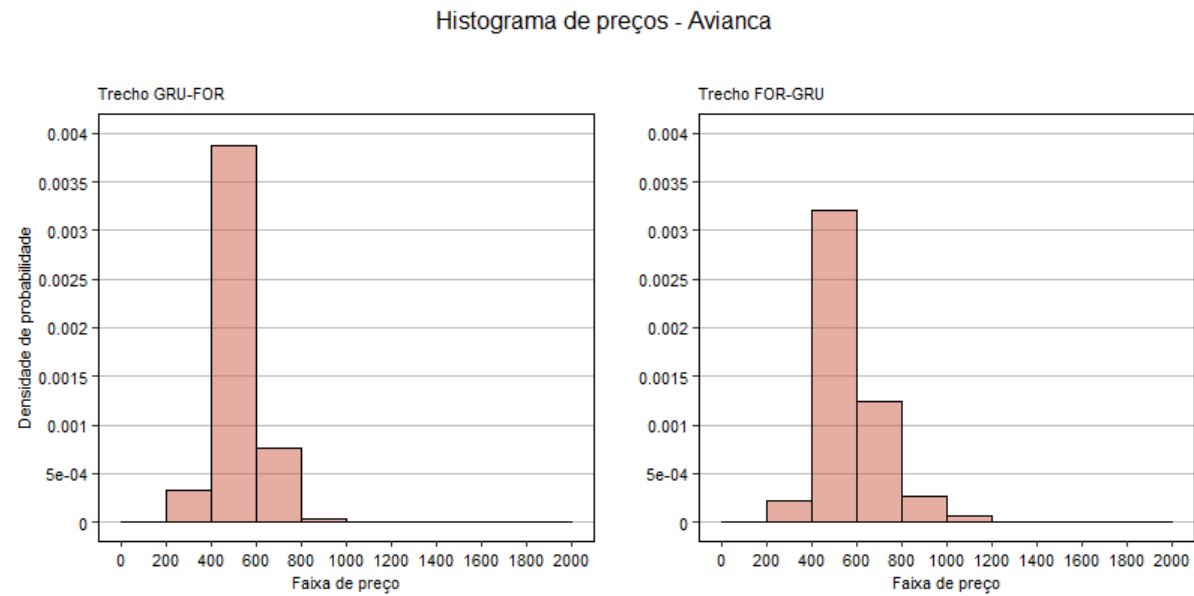


Figura 5.18: Histograma de precificação da empresa aérea Avianca

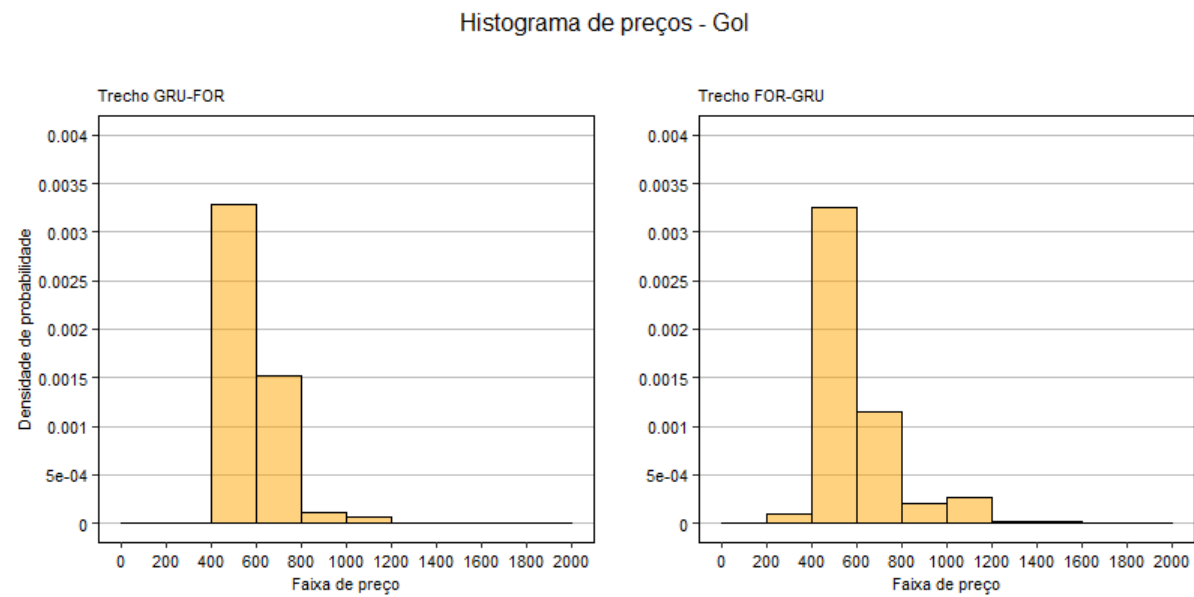


Figura 5.19: Histograma de precificação da empresa aérea Gol

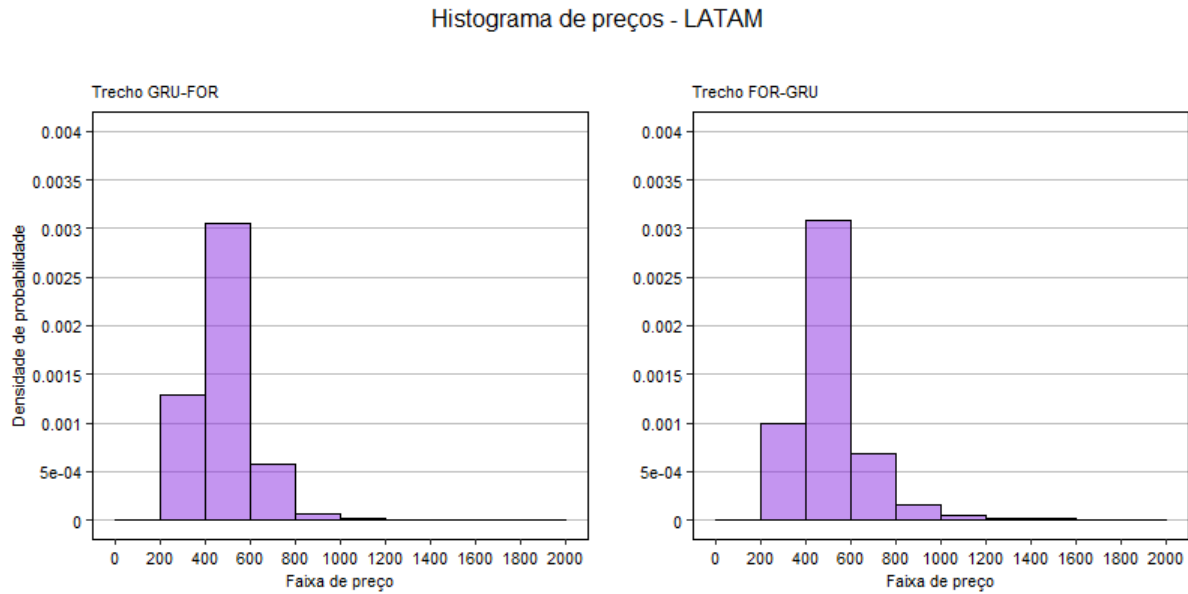


Figura 5.20: Histograma de precificação da empresa aérea LATAM

Nesses histogramas, preços maiores que R\$ 2000 foram considerados *outliers*. Observando os histogramas é perceptível que o trecho FOR-GRU possui colunas maiores em faixas de preços maiores quando relacionados com o trecho de GRU-FOR. Também se consegue notar que, realmente, a Avianca possui uma variação de preços bem menor que as outras empresas aéreas, sendo sua precificação, em grande parte, na faixa de R\$ 400.00 a R\$ 600.00. Enquanto isso, nota-se nos histogramas da Azul que ela apresenta uma variação maior, com suas colunas sendo representativas até R\$ 1600.00.

As Figura 5.21 e Figura 5.22 mostram as relações entre as médias das empresas aéreas e a antecedência de compra.

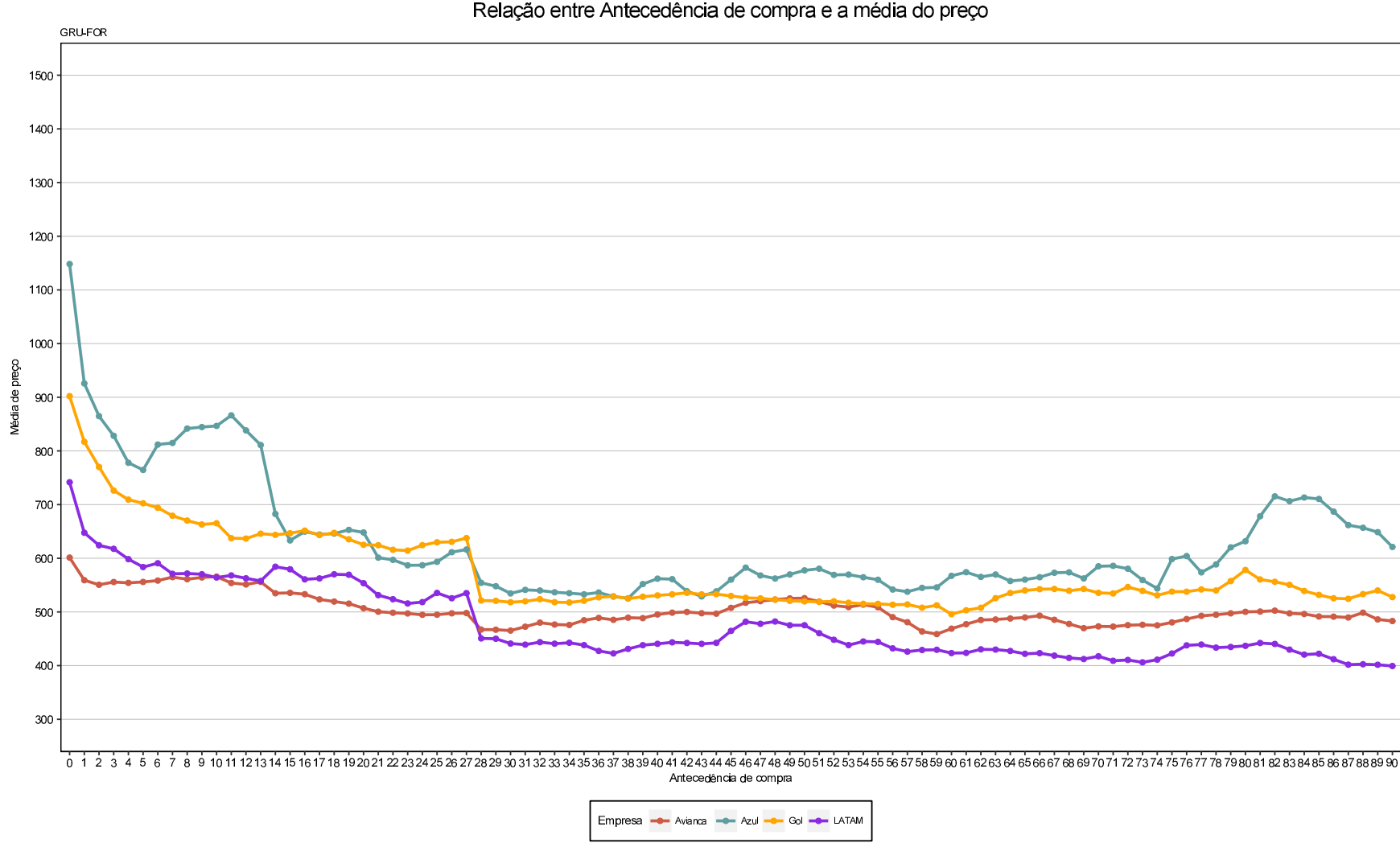


Figura 5.21: Relações entre as médias de preços para os trechos GRU-FOR e FOR-GRU

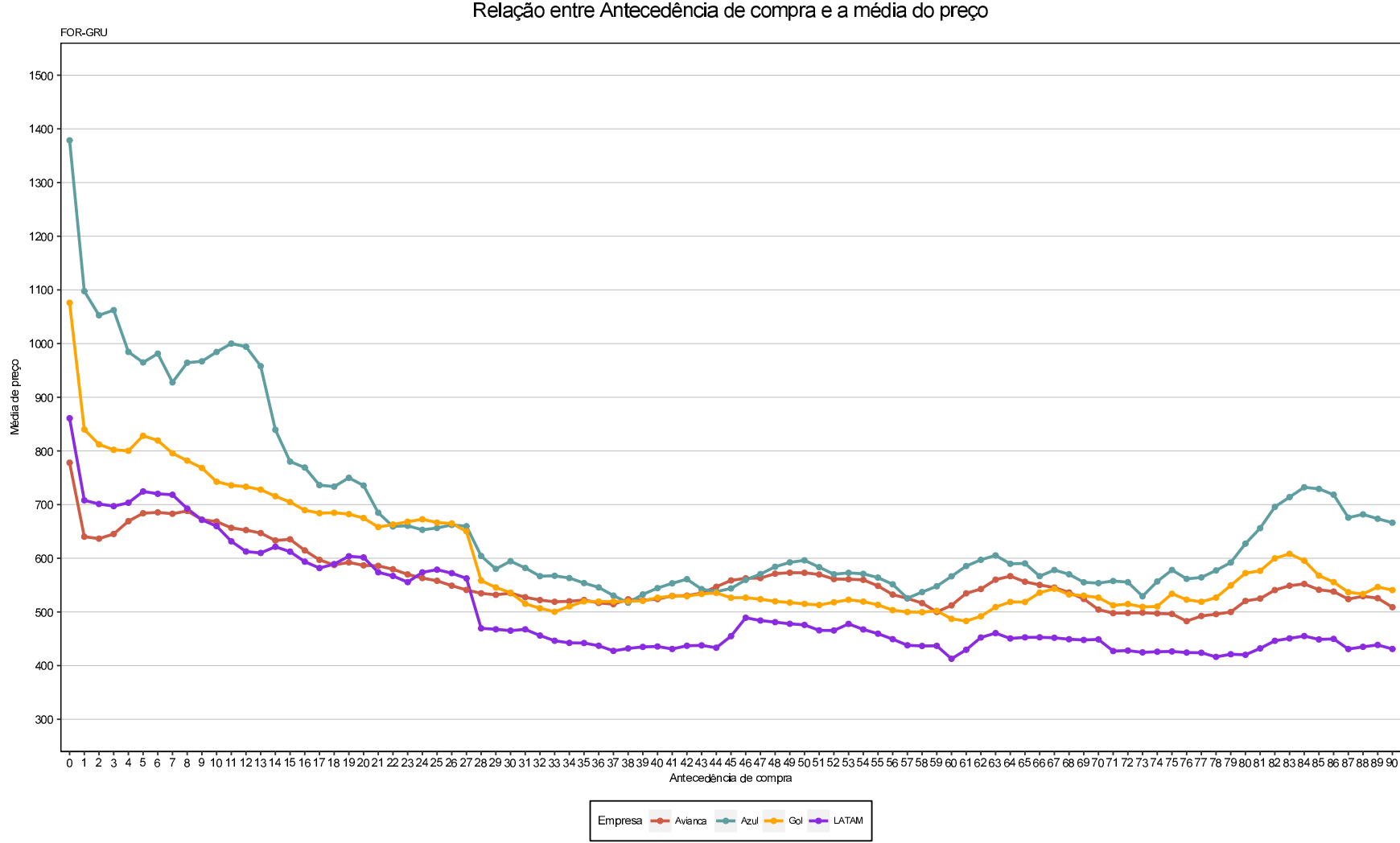


Figura 5.22: Relações entre as médias de preços para os trechos GRU-FOR e FOR-GRU

Com isso, pode-se observar que a LATAM é a empresa que mantém os preços mais baixos por quase todo o período de antecedência, a Avianca não parece fazer uma mudança de precificação muito significativa, permanecendo quase sempre próximo do patamar dos R\$ 600.00. Enquanto isso, a Azul é a detentora das maiores médias entre as empresas observadas.

Outra observação interessante do gráfico é a de que, é perceptível que as empresas parecem variar a precificação de suas passagens aéreas praticamente ao mesmo tempo. Quando há uma diminuição significativa de uma, as outras tendem a acompanhá-la. Da mesma forma, quando há um aumento no preço de uma empresa, as outras a acompanham. Com isso, aparentemente, o preço estabelecido pelos concorrentes é um fator considerado pelas empresas aéreas na precificação de suas passagens aéreas.

Além disso, os gráficos mostram que, não necessariamente, quanto maior a antecedência de compra de uma passagem, menor será o preço pago pelo cliente, pois a partir de 80 dias de antecedência na compra observa-se que as médias de preços podem ser maiores que buscas com antecedência de 28 a 60 dias.

Outro fator que sustenta a ideia de que o preço de uma empresa aérea influencia no preço das outras é a análise das correlações entre as médias dos preços das empresas aéreas. A Tabela 5.3 mostra as relações de correlação entre cada uma das empresas aéreas, duas a duas, para o trecho GRU-FOR.

Empresa1	Empresa2	Correlação
Avianca	Azul	0.83
Avianca	Gol	0.81
Avianca	LATAM	0.85
Azul	Gol	0.85
Azul	LATAM	0.77
Gol	LATAM	0.93

Tabela 5.3: Correlações entre as médias das empresas aéreas no trecho GRU-FOR

A Tabela 5.4 mostra as correlações entre as empresas no trecho FOR-GRU.

Empresa1	Empresa2	Correlação
Avianca	Azul	0.90
Avianca	Gol	0.90
Avianca	LATAM	0.94
Azul	Gol	0.95
Azul	LATAM	0.91
Gol	LATAM	0.97

Tabela 5.4: Correlações entre as médias das empresas aéreas no trecho FOR-GRU

A partir das Tabela 5.3 e Tabela 5.4 pode-se dizer que existe uma alta correlação entre as médias dos preços das passagens aéreas das empresas, o que sugere que esses preços estão relacionados de alguma forma. Percebe-se que existe uma maior correlação no trecho FOR-GRU que no GRU-FOR, porém ambos são bastante elevados.

Outra análise realizada foi a verificação da correlação entre os preços médios das empresas aéreas nos dois trechos estudados. A Tabela 5.5 mostra os resultados obtidos dessa análise.

Empresa	Correlação
Avianca	0.93
Azul	0.97
Gol	0.97
LATAM	0.97

Tabela 5.5: Correlações entre as médias dentro de cada empresa aérea nos trechos GRU-FOR e FOR-GRU

Essa análise mostra que o nível de correlação entre os trechos dentro de cada empresa aérea é muito elevado, levando a supor que os preços em cada trecho são muito dependentes entre si.

Para concluir as análises objetivadas no tópico presente, serão ilustrados os gráficos das médias de preços das empresas acrescentando-se o duas vezes o desvio padrão, ou seja, os gráficos mostram o intervalo de valores mais prováveis de preço (95% de confiabilidade, admitindo distribuição normal) para cada empresa aérea ao longo da antecedência de busca. Os Gráficos 5.23 e 5.24 mostram os intervalos de preços mais prováveis para a empresa Azul. Enquanto isso, os Gráficos 5.25 e 5.26 mostram os intervalos para a empresa Avianca, os Gráficos 5.27 e 5.28 mostram os intervalos para a empresa Gol e os Gráficos 5.29 e 5.30 mostram os intervalos para a empresa LATAM.

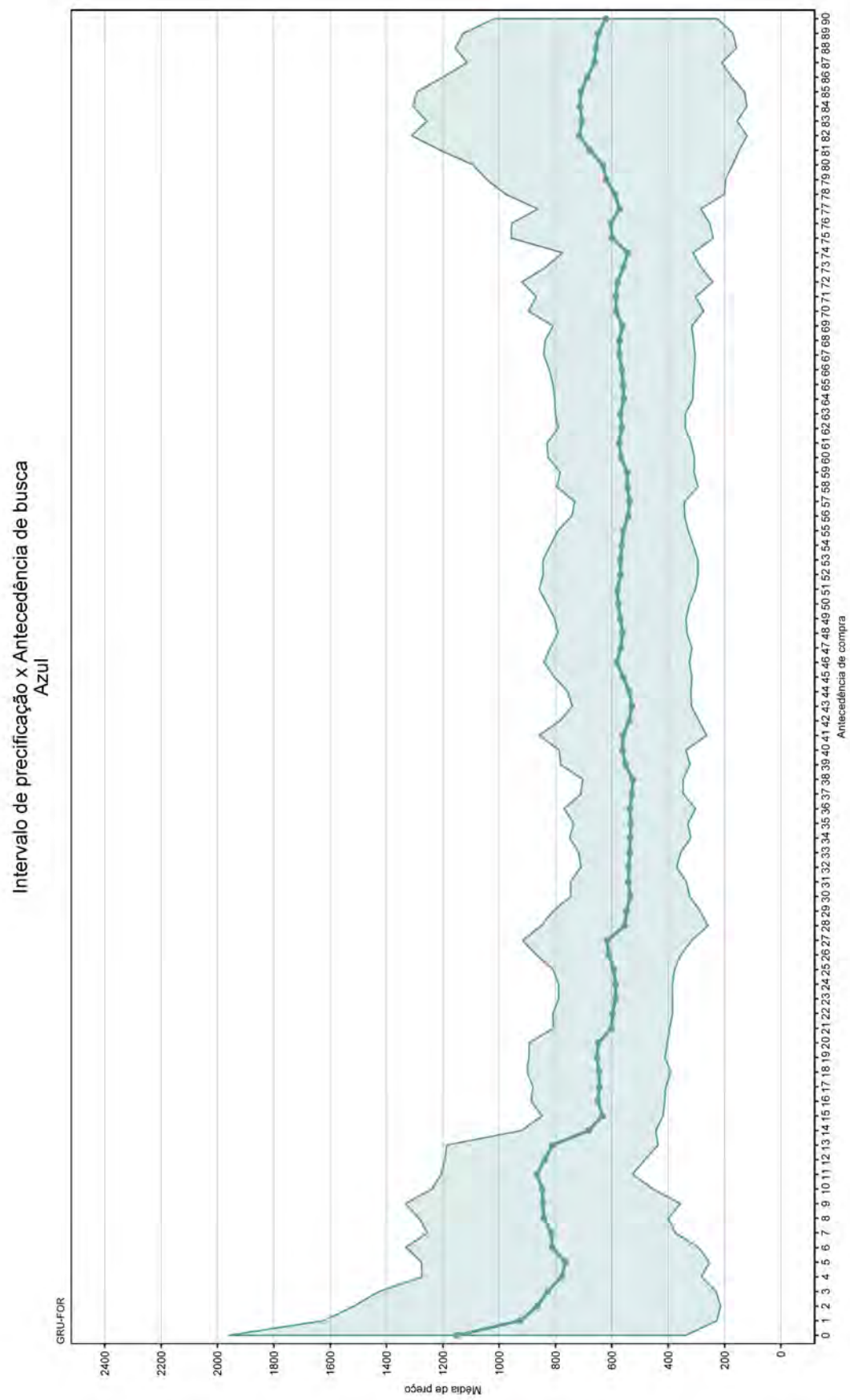


Figura 5.23: Intervalo dos preços mais prováveis x antecedência para a empresa Azul no trecho GRU-FOR

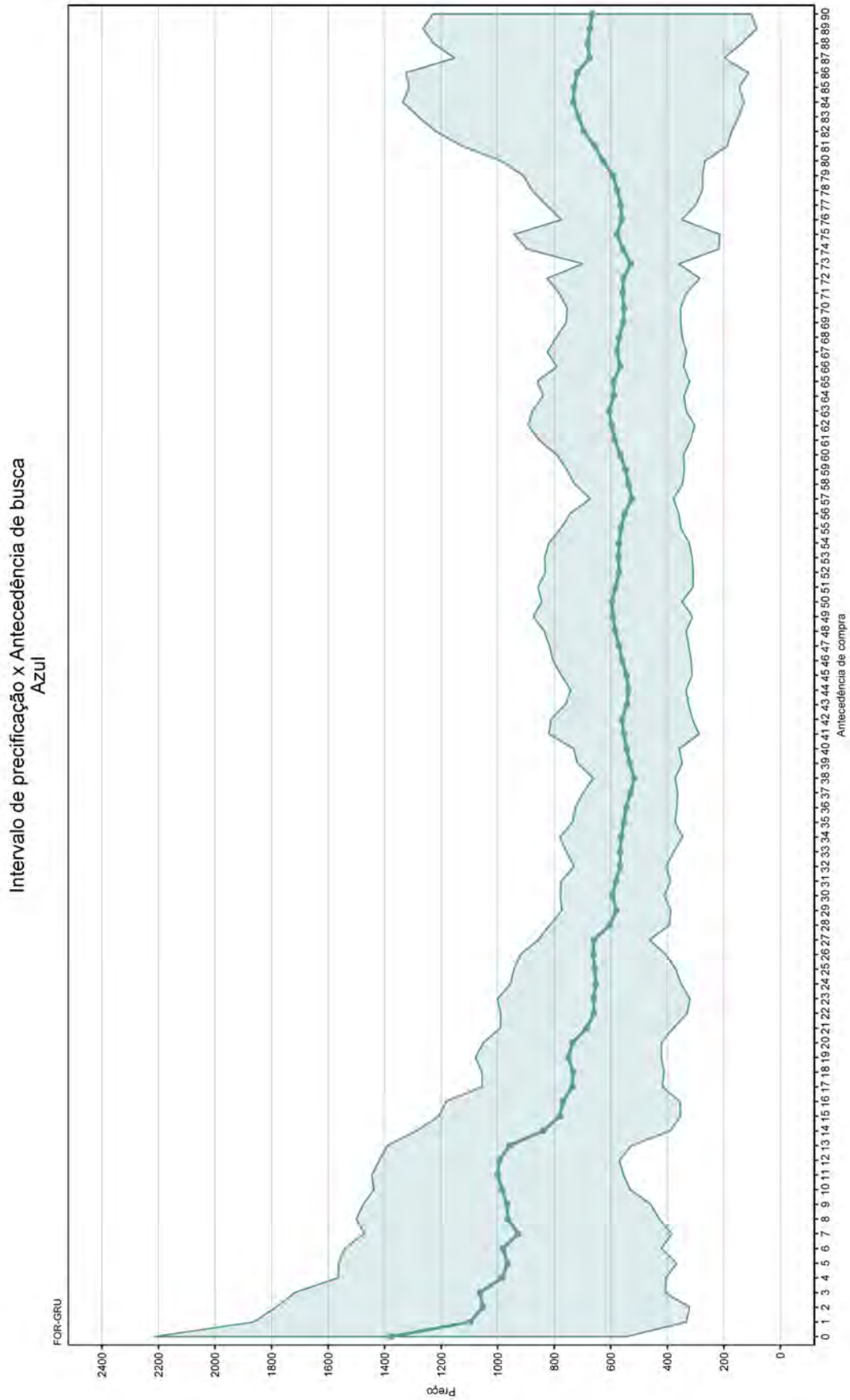


Figura 5.24: Intervalo dos preços mais prováveis x antecedência para a empresa Azul no trecho FOR-GRU

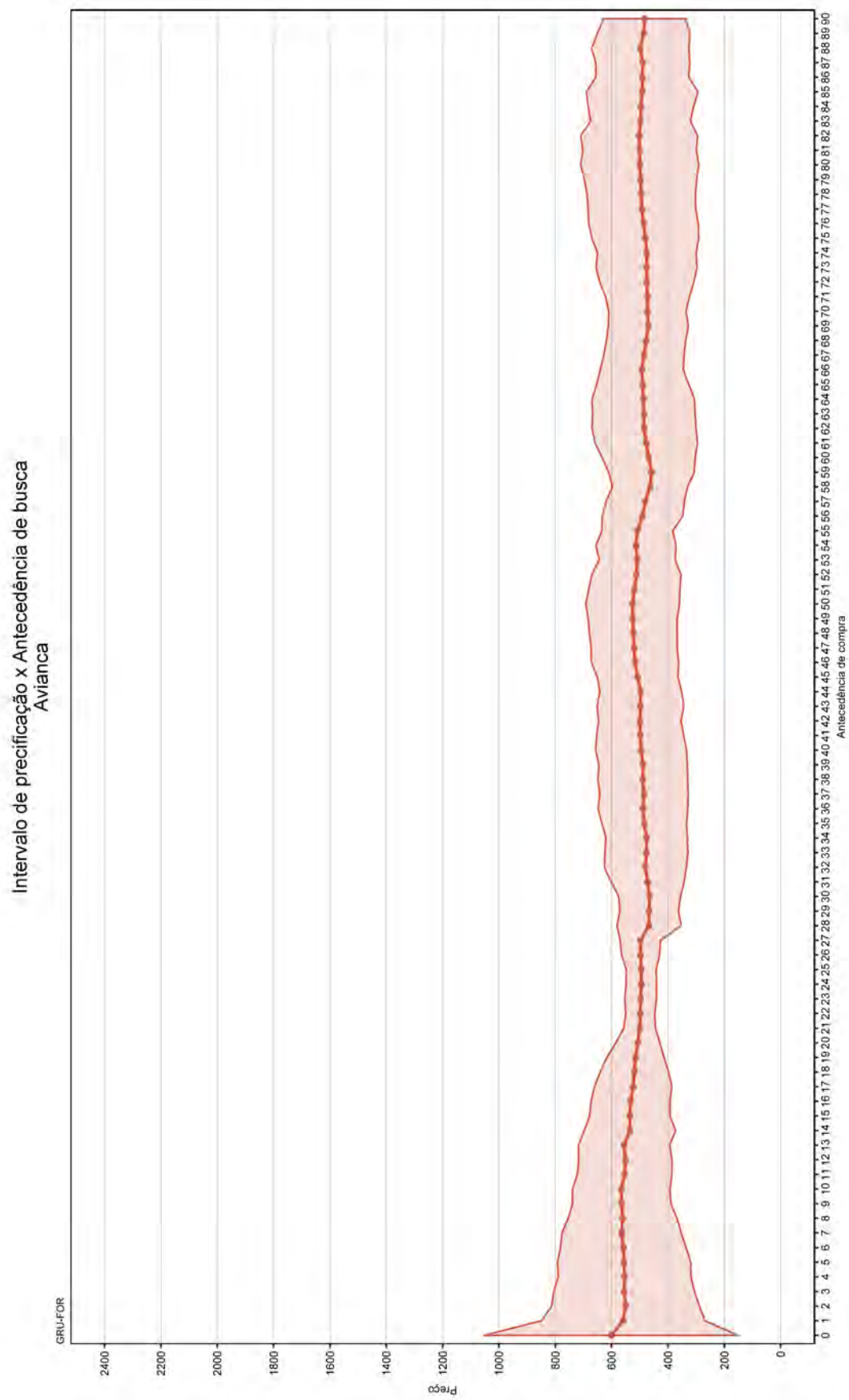


Figura 5.25: Intervalo dos preços mais prováveis x antecedência para a empresa Avianca no trecho GRU-FOR

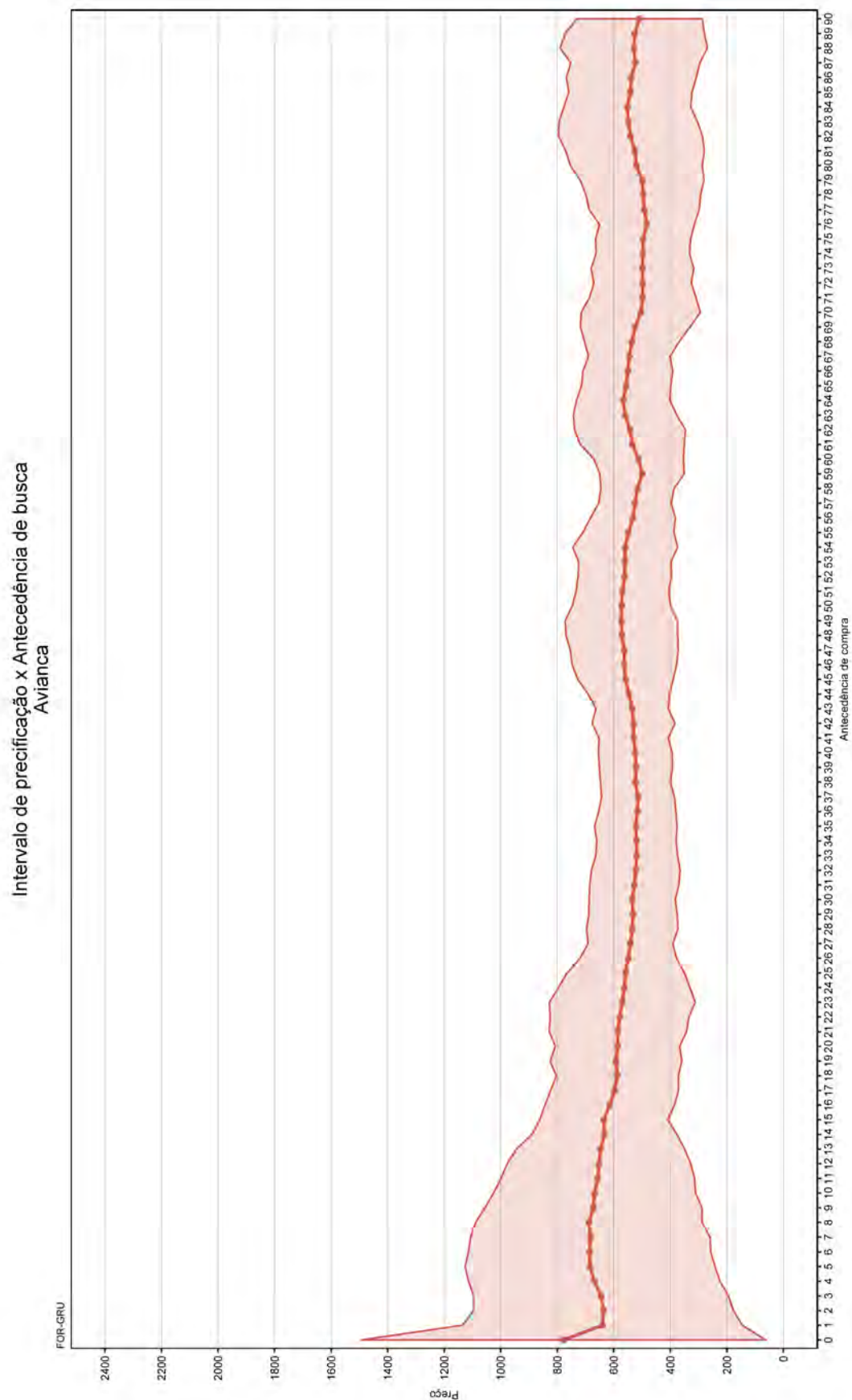


Figura 5.26: Intervalo dos preços mais prováveis x antecedência para a empresa Avianca no trecho FOR-GRU

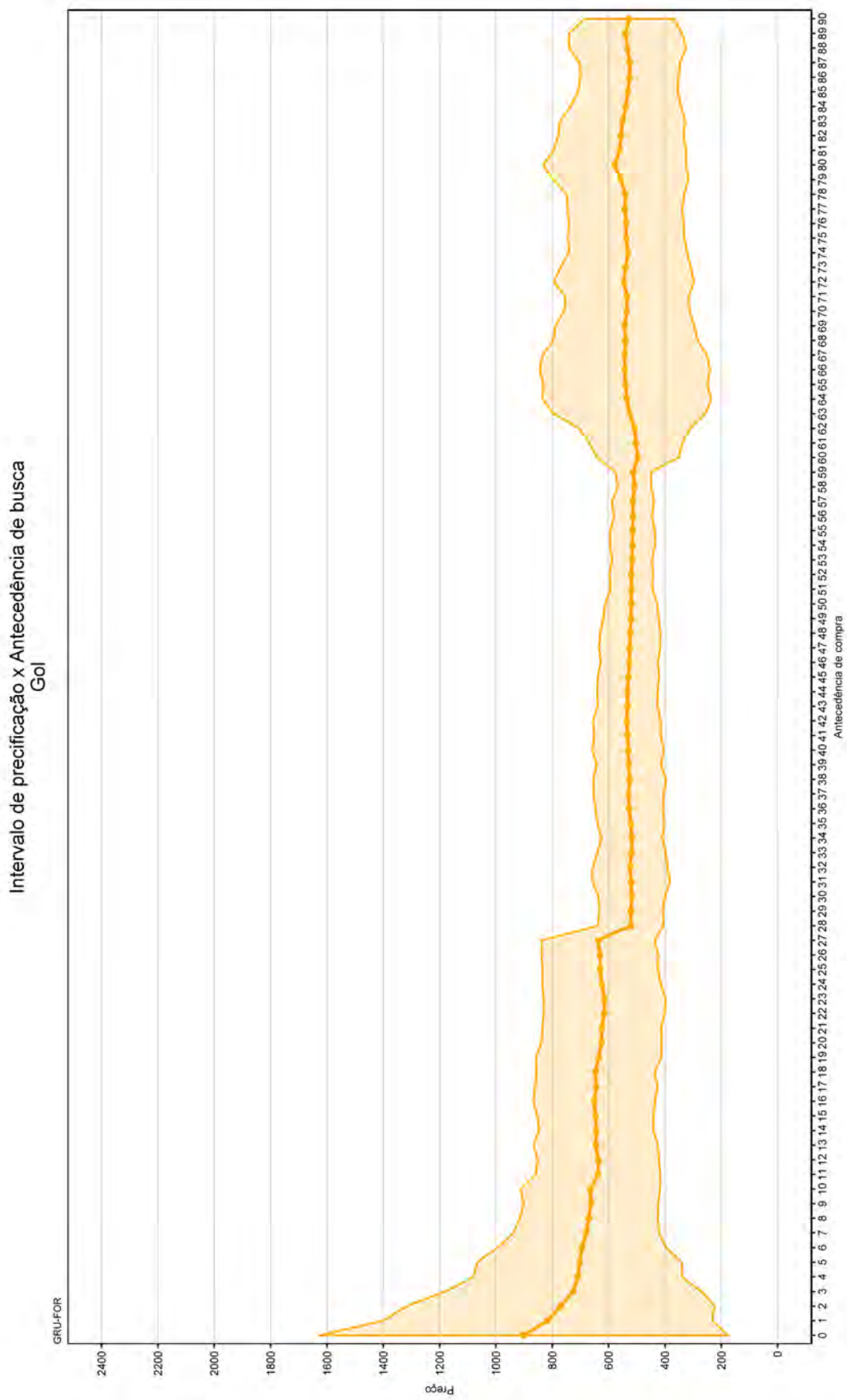


Figura 5.27: Intervalo dos preços mais prováveis x antecedência para a empresa Gol no trecho GRU-FOR

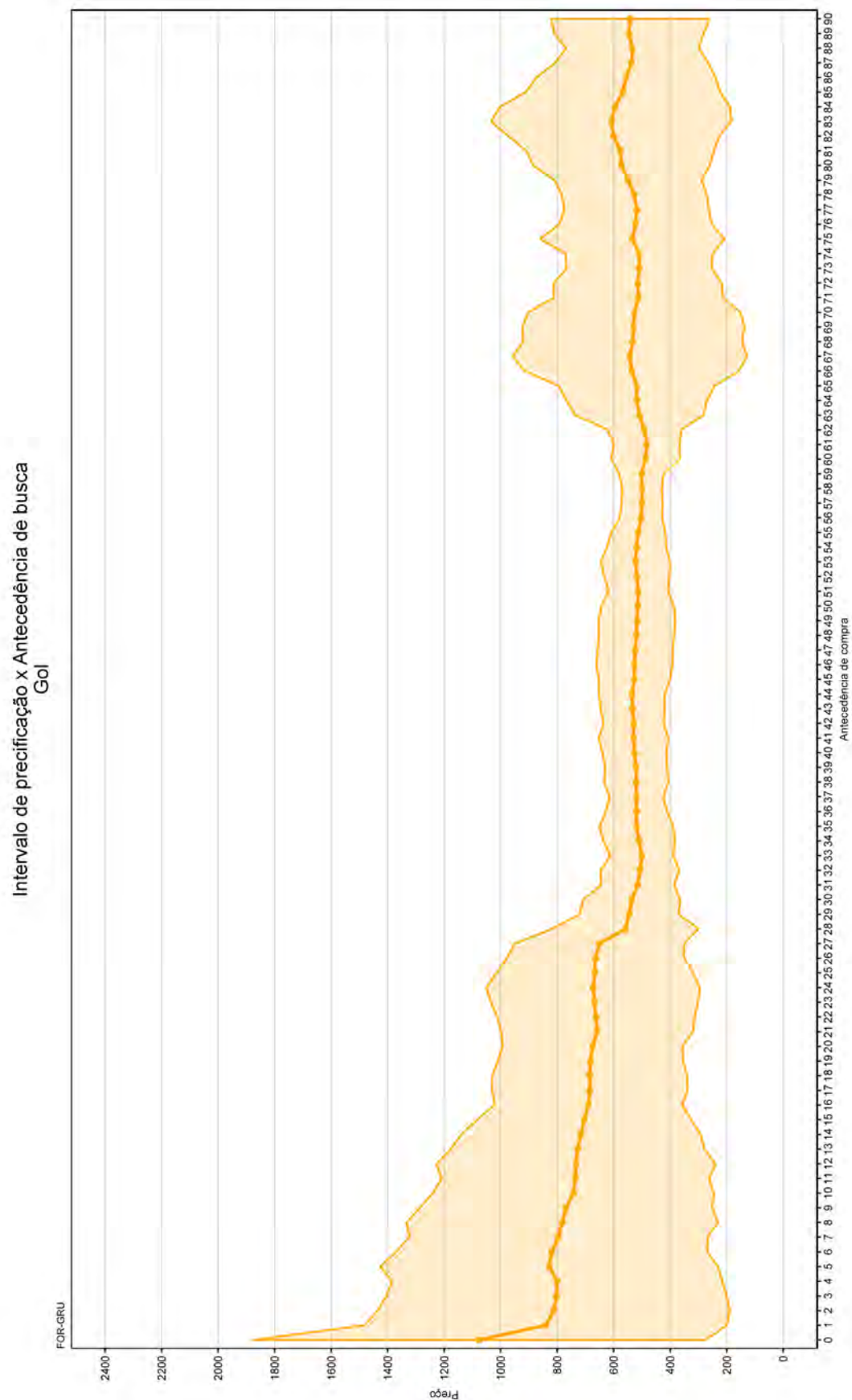


Figura 5.28: Intervalo dos preços mais prováveis x antecedência para a empresa Gol no trecho FOR-GRU

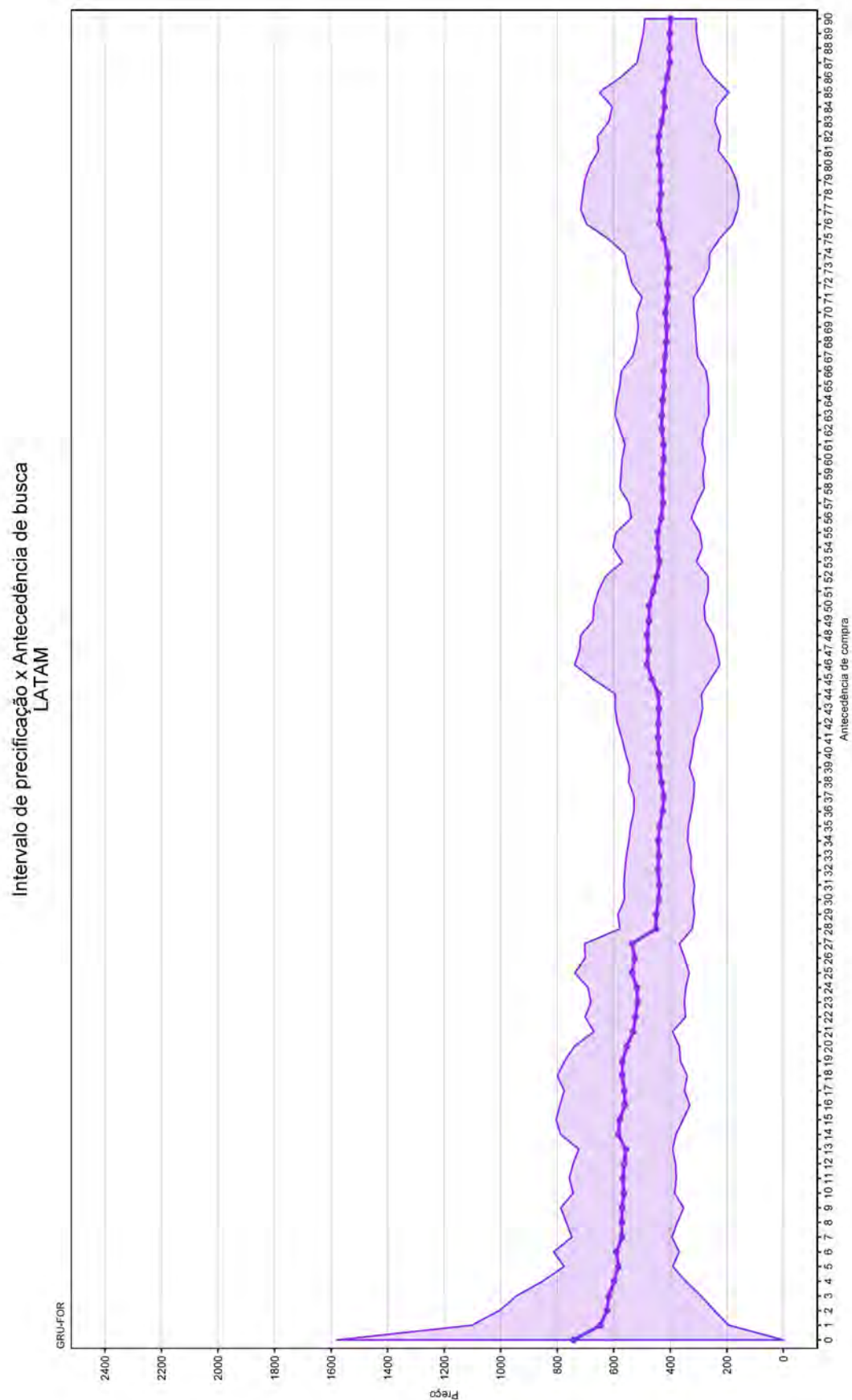


Figura 5.29: Intervalo dos preços mais prováveis x antecedência para a empresa LATAM no trecho GRU-FOR

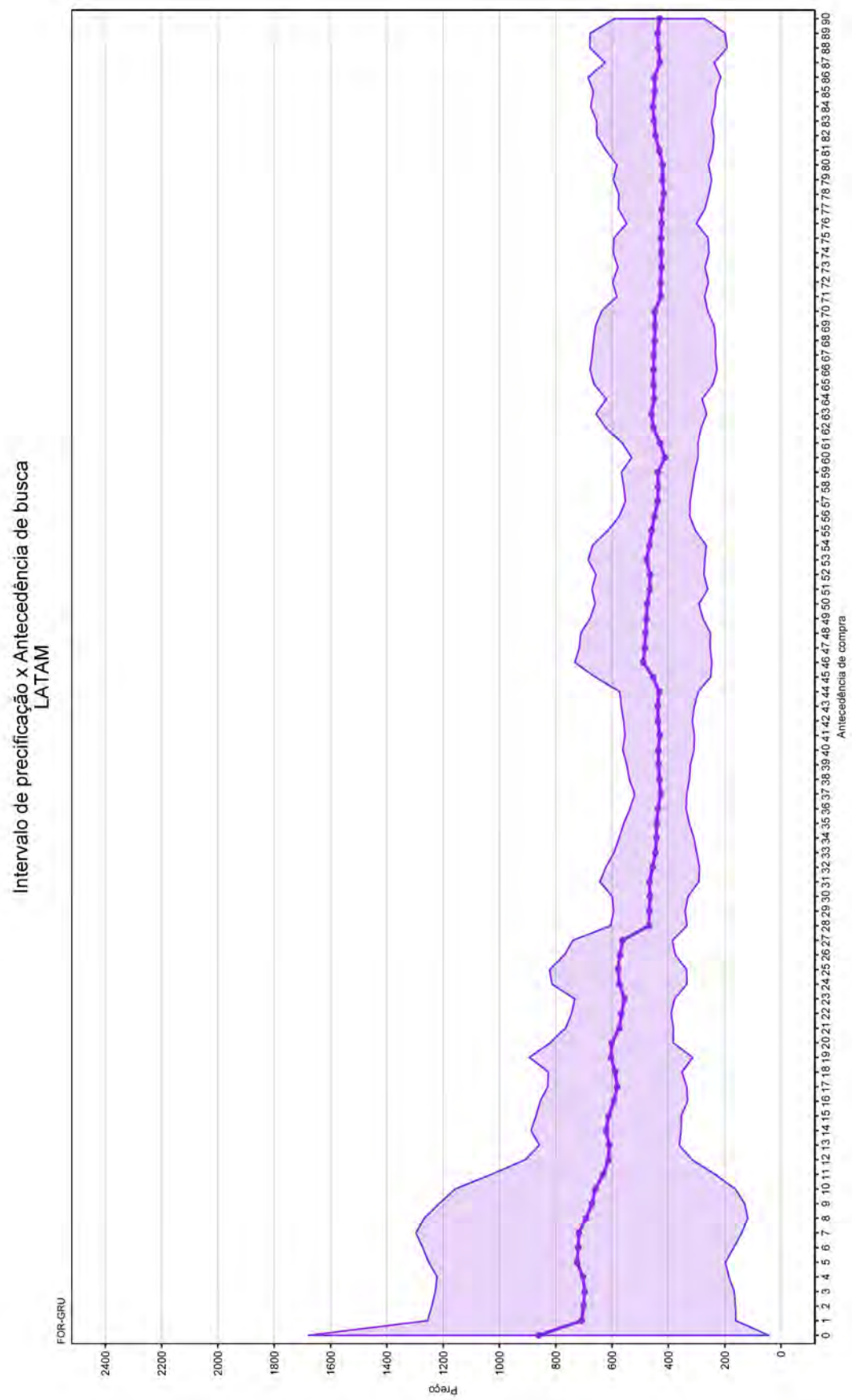


Figura 5.30: Intervalo dos preços mais prováveis x antecedência para a empresa LATAM no trecho FOR-GRU

A análise observada nos gráficos define de maneira mais detalhada como funciona a variação dos preços em cada uma das empresas aéreas brasileiras. Ela nos permite observar em momentos as empresas estão mais susceptíveis a variação de preços e em quais desses momentos de variação será possível se ter passagens mais caras. Os gráficos mostram que, praticamente todas as empresas possuem um período no qual o intervalo de precificação é relativamente alto, que pode ser observado entre 0 e 27 dias, aproximadamente, para todas as empresas estudadas.

O Gráfico 5.23 mostra que a Azul tem um intervalo de precificação bem abrangente ao longo de todo o período analisado, ratificando que os seus preços costumam ter muitas variações em função do tempo de antecedência. Enquanto isso, pode-se observar no Gráfico 5.25 que a Avianca possui uma curva de intervalo de preços delgada, confirmando que o preço imposto pela empresa não costuma variar de forma abrupta em função do tempo de antecedência. Por outro lado, nota-se análise do Gráfico 5.27 que a empresa aérea Gol possui três momentos bem definidos quanto ao intervalo de precificação: dois com altas variações relativas, um entre 0 e 30 dias de antecedência e outro entre 60 e 90 dias de antecedência, e um com baixas variações entre 30 e 60 dias.

5.3 Realizar previsões de preços

5.3.1 Descrição e preparação dos dados

Neste tópico serão mostradas previsões de preços por meio das análises dos dados obtidos e o erro médio obtido nessas previsões em comparação com os preços reais coletados. A partir do erro obtido, pode-se ter ideia do nível de confiança das previsões realizadas e pode-se gerar simulações para se calcular o quanto se pode economizar, em média, na compra de passagens.

A previsão de preços foi realizada em dois passos: o primeiro foi classificar em que categoria de preço cada linha da base estaria dado o preço obtido e, após a classificação, prever o preço daquela passagem mediante a análises de regressão linear. A base de dados utilizada pode ser encontrada no Anexo C, contido no CD deste trabalho.

De início, os preços foram categorizados de acordo com faixas de preços pré-definidas. As categorias criadas e suas respectivas faixas de preço são expostas na Tabela 5.6.

Categoria	Faixa de preço (R\$)
A	0 - 400
B	400 - 600
C	600 - 800
D	800 - 1000
E	1000 - 1200
F	1200 - 4000

Tabela 5.6: Categorias criadas para classificação de faixas de preços

As variáveis que seriam utilizadas no modelo são mostradas na Tabela 5.7, além dos seus possíveis valores.

Variável	Descrição	Possíveis valores
Horário	Horário da busca	00:00, 02:00, 04:00 ...
Antecedência	Antecedência da busca	de 0 a 90 dias
Mês_Viagem	Mês em que se buscando a viagem	De julho a dezembro
Nome_Dia	Nome do dia da semana	Domingo, segunda-feira, terça-feira ...
Fds_Sem	Se a data em que se busca é Fim de semana ou semana	Fds ou Sem
Atributo	Se é um dia normal, feriado, véspera de feriado etc	Normal, Feriado, Data comemorativa, Véspera de feriado, Véspera de comemoração ...
Empresa	Empresa aérea buscada	Azul, Avianca, Gol ou LATAM

Tabela 5.7: Descrição das variáveis utilizadas

Porém, percebeu-se que as variáveis “Nome_Dia” e “Fds_Sem” são linearmente dependentes e, por isso, acabam prejudicando a previsão do preço da passagem no momento em que os modelos de regressão são montados. Por isso, utilizou-se o coeficiente de contingência proposto por Pearson para medida de associação entre duas variáveis a fim de se saber qual das variáveis seria menos relevante para o modelo.

O coeficiente de contingência tem um valor mínimo de 0 e um máximo de 1. Quanto mais próximo de uma o coeficiente calculado para uma variável, mais associada ela está ao parâmetro estudado. Ele é dado pela seguinte equação:

$$P = \sqrt{\frac{Q_{obs}^2}{n + Q_{obs}^2}} \quad (5.2)$$

onde n é a quantidade de observações e Q_{obs}^2 calcula-se da seguinte forma:

$$Q_{obs}^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{O_{ij} - E_{ij}}{E_{ij}} \quad (5.3)$$

com O_{ij} sendo as frequências observadas e E_{ij} sendo as frequências esperadas no teste qui-quadrado de independência.

A Tabela 5.8 mostra os valores calculados para cada uma das variáveis estudadas.

Variável	Coefficiente de contingência
Antecedência	0.48
Atributo	0.16
Empresa	0.39
Fds_Sem	0.11
Horário	0.07
Nome_Dia	0.15
Trecho	0.10

Tabela 5.8: Coeficientes de contingência obtidos

A partir da Tabela 5.8 percebemos que a variável “Antecedência” e “Empresa” possuem uma grande associação com a categoria de preço, enquanto que as variáveis “Horário”, “Fds_Sem” e “Trecho” praticamente não estão associadas.

Na análise de regressão percebe-se que as variáveis “Fds_Sem” e “Nome_Dia” são linearmente dependentes, o que prejudicaria os modelos de regressão criados, por isso, uma delas deveria sair do modelo. Verificando-se que a variável “Fds_Sem” possui uma associação muito pequena com a variável “Categoria de preço”, ela foi escolhida para sair do modelo.

Portanto, para a construção do modelo, as variáveis “Horário”, “Antecedência”, “Mês da viagem”, “Nome do dia”, “Atributo do dia” e “Trecho” foram as variáveis utilizadas na descrição dos modelos criados para a descrição do problema.

5.3.2 *Random Forest*

Para se iniciar a previsão dos preços das passagens aéreas tornou-se necessário, primeiramente, criar-se um algoritmo para classificar em que categoria de preços uma dada busca estaria de acordo com as variáveis supracitadas. O algoritmo escolhido para a se realizar a classificação da base de dados foi o *Random Forest*.

Foram criados quatro algoritmos de classificação desse tipo, um para cada empresa aérea. Os algoritmos criados para cada uma das empresas possuem os mesmos parâmetros de definição, pois, devido a semelhança entre as bases de cada uma delas, a *Random Forest* de cada uma fica com a máxima eficiência com os mesmos parâmetros. Cada *Random Forest* é definido pelo número de árvores que são criadas em cada modelo e pelo número de variáveis utilizadas para predição nos nós de cada árvore. No modelo mais eficiente

encontrado, encontrou-se que o número de árvores utilizadas seria de 200 e o número de variáveis seria de 2 por cada nó.

Para a empresa aérea Azul, foi obtido um resultado com uma precisão de 76.48%.

A matriz de confusão é matriz na qual pode ser observada a relação entre os valores previstos pelo algoritmo utilizado e os valores reais utilizados no treinamento do mesmo. É a partir dela que é calculada a precisão do algoritmo, fazendo-se

$$Precisão = \frac{tr(Matriz\ de\ confusão)}{Soma(Matriz\ de\ confusão)} \quad (5.4)$$

A Tabela 5.9 mostra a matriz de confusão obtida como resultado do modelo para a Azul.

Previsão	Referência					
	A	B	C	D	E	F
A	1	0	0	0	0	0
B	534	28 545	6 477	1 028	357	273
C	1	1 336	9 726	1 110	621	255
D	0	48	90	1 631	93	28
E	0	63	155	204	1 944	173
F	0	69	158	165	95	1 509

Tabela 5.9: Tabela de confusão obtida para o modelo da empresa Azul

A Figura 5.31 mostra a evolução do erro na previsão de cada categoria, enquanto que a Figura 5.32 mostra o nível de importância de cada variável para esse modelo. O nível de importância é medido por meio do coeficiente de Gini, no qual, quanto maior, mais importante a variável é na descrição daquele modelo.

Relação entre os erros e a quantidade de árvores criadas no modelo RF - Azul

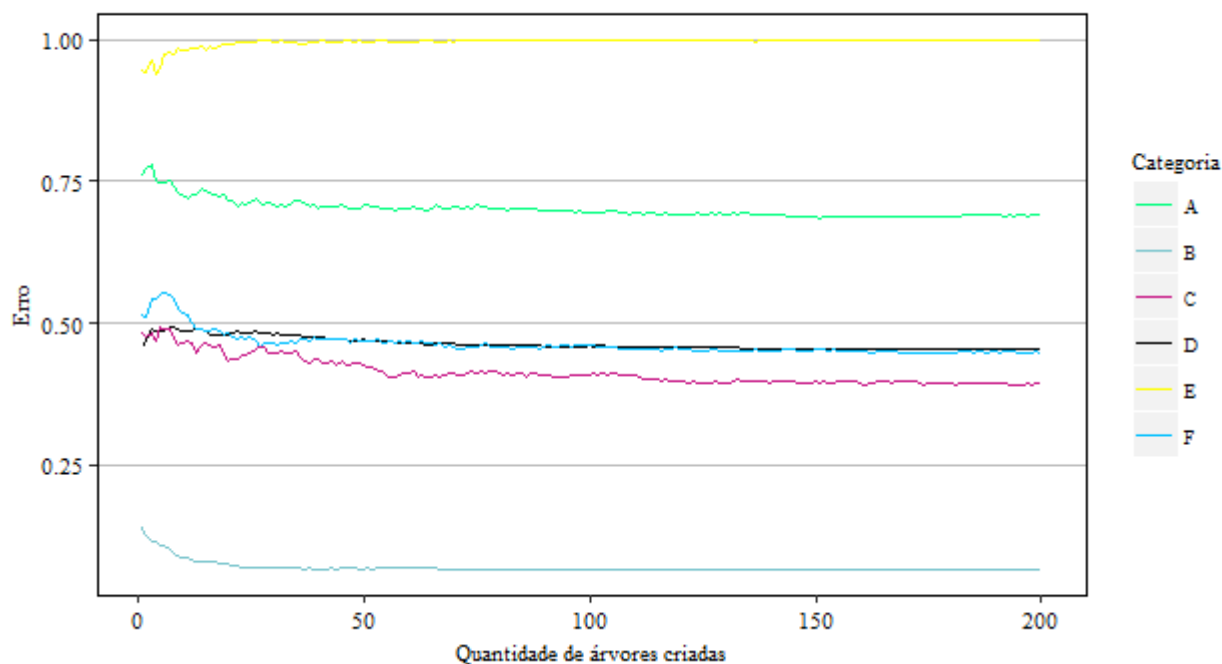


Figura 5.31: Evolução com erro com o aumento do n° de árvores no modelo da Azul

Importância de cada variável para o modelo - Azul

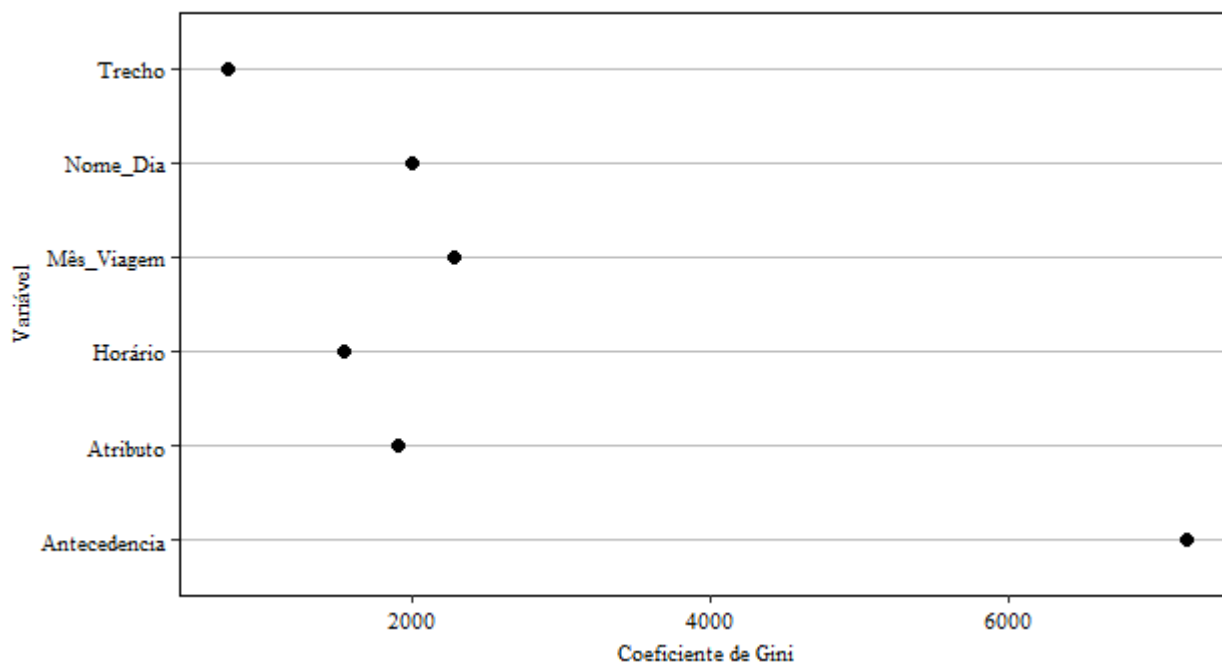


Figura 5.32: Importância de cada variável para o modelo de RF da empresa Azul

Para a empresa aérea Avianca, foi obtido um resultado com uma precisão de 82.09%. A Tabela 5.10 mostra a matriz de confusão obtida como resultado do modelo para a Avianca.

Previsão	Referência					
	A	B	C	D	E	F
A	1355	461	2	0	0	0
B	1 732	38 932	5 744	464	17	3
C	6	656	5 129	366	152	16
D	0	146	172	768	104	32
E	0	0	12	3	44	0
F	0	0	0	0	0	0

Tabela 5.10: Tabela de confusão obtida para o modelo da empresa Azul

A Figura 5.33 mostra a evolução do erro na previsão de cada categoria, enquanto que a Figura 5.34 mostra o nível de importância de cada variável para esse modelo.

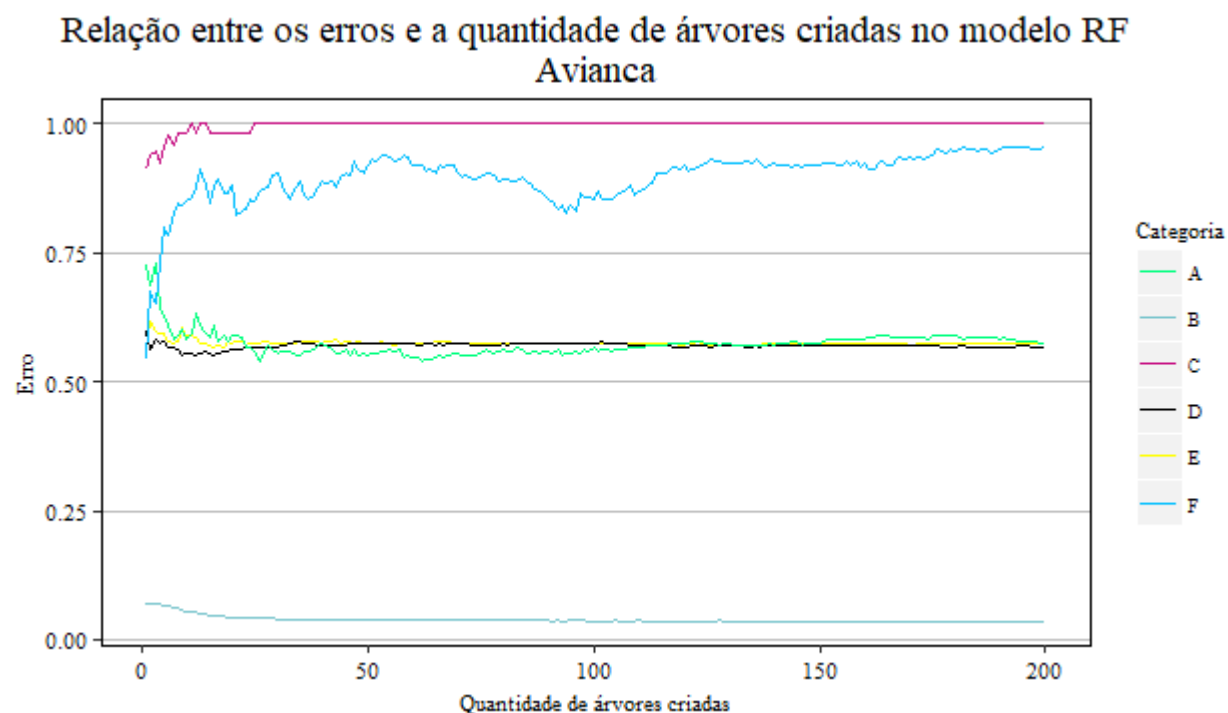


Figura 5.33: Evolução com erro com o aumento do n° de árvores no modelo da Avianca

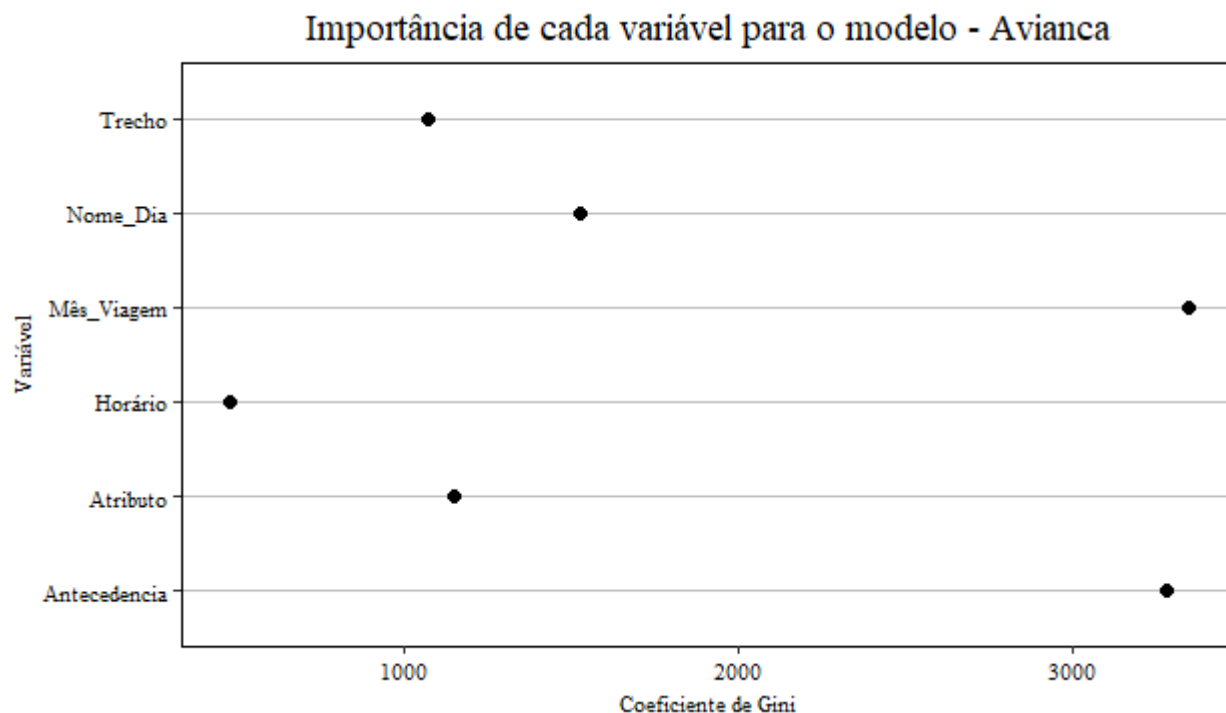


Figura 5.34: Importância de cada variável para o modelo de RF da empresa Avianca

Já para a empresa aérea Gol, foi obtido um resultado com uma precisão de 82.60%. A Tabela 5.11 mostra a matriz de confusão obtida como resultado do modelo para a Gol.

Previsão	Referência					
	A	B	C	D	E	F
A	3	0	0	0	0	0
B	448	34 612	5 166	320	234	23
C	9	1 634	9 367	413	345	45
D	1	46	28	597	19	3
E	0	43	313	365	1 150	194
F	0	0	0	0	2	66

Tabela 5.11: Tabela de confusão obtida para o modelo da empresa Gol

A Figura 5.35 mostra a evolução do erro na previsão de cada categoria, enquanto que a Figura 5.36 mostra o nível de importância de cada variável para esse modelo.

Relação entre os erros e a quantidade de árvores criadas no modelo RF - Gol

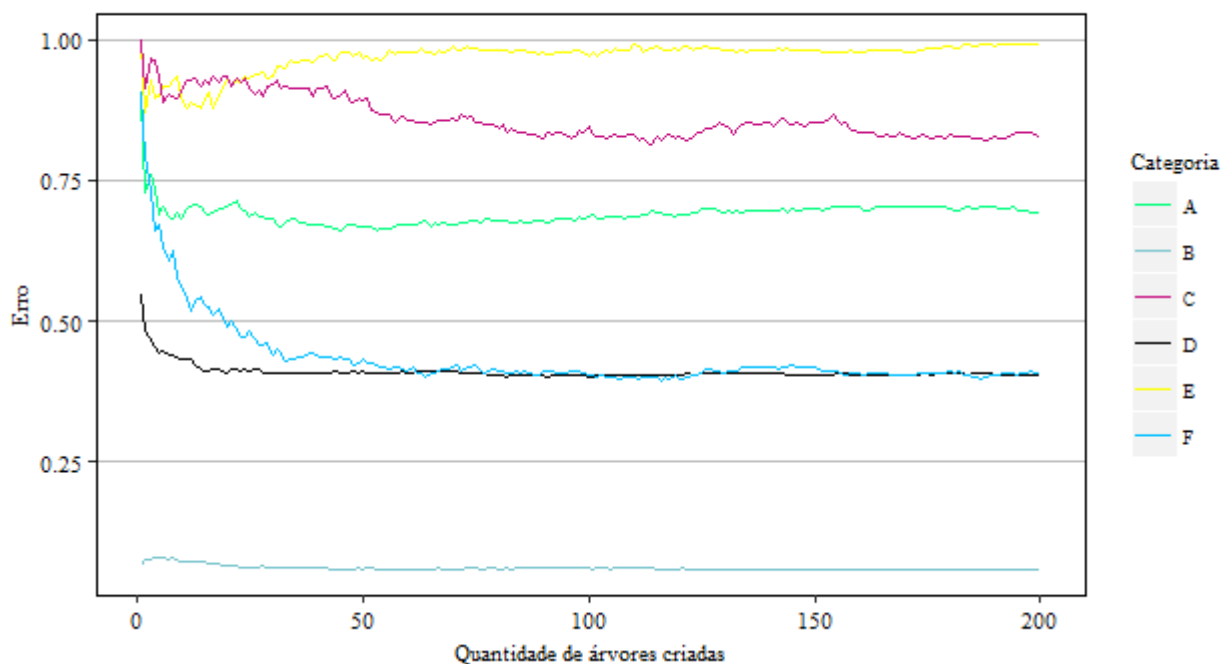


Figura 5.35: Evolução com erro com o aumento do n° de árvores no modelo da Gol

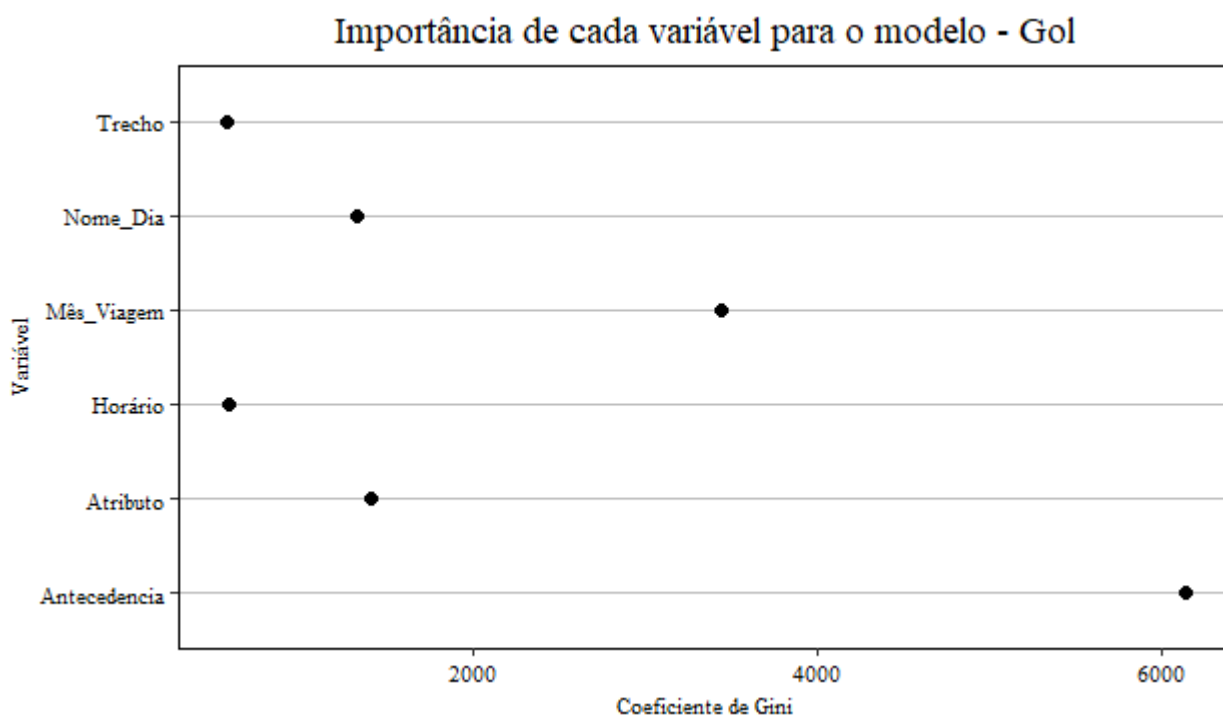


Figura 5.36: Importância de cada variável para o modelo de RF da empresa Gol

Para a empresa aérea LATAM, foi obtido um resultado com uma precisão de 75.92%. A Tabela 5.12 mostra a matriz de confusão obtida como resultado do modelo para a LATAM.

Previsão	Referência					
	A	B	C	D	E	F
A	7725	2437	82	3	0	0
B	5 240	31 428	3 370	377	4	12
C	24	886	3 662	578	316	234
D	0	15	42	215	25	1
E	0	0	0	0	0	0
F	0	0	0	0	0	0

Tabela 5.12: Tabela de confusão obtida para o modelo da empresa LATAM

A Figura 5.37 mostra a evolução do erro na previsão de cada categoria, enquanto que a Figura 5.38 mostra o nível de importância de cada variável para esse modelo.

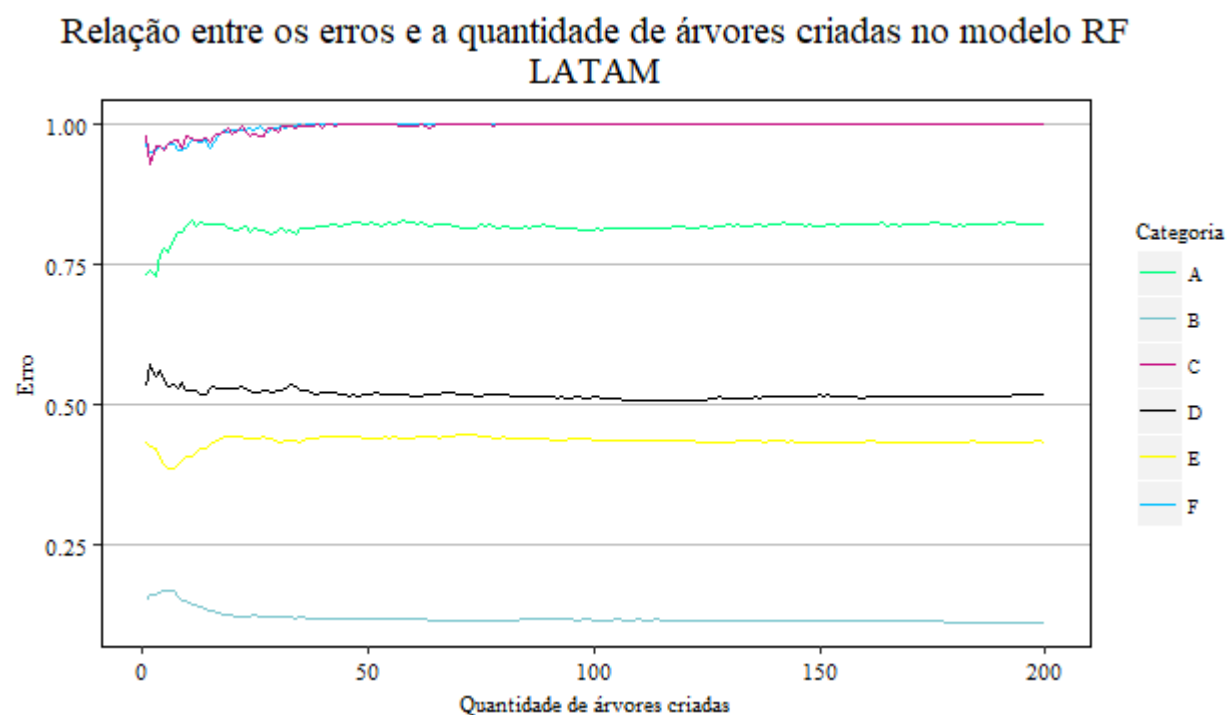


Figura 5.37: Evolução com erro com o aumento do n° de árvores no modelo da LATAM

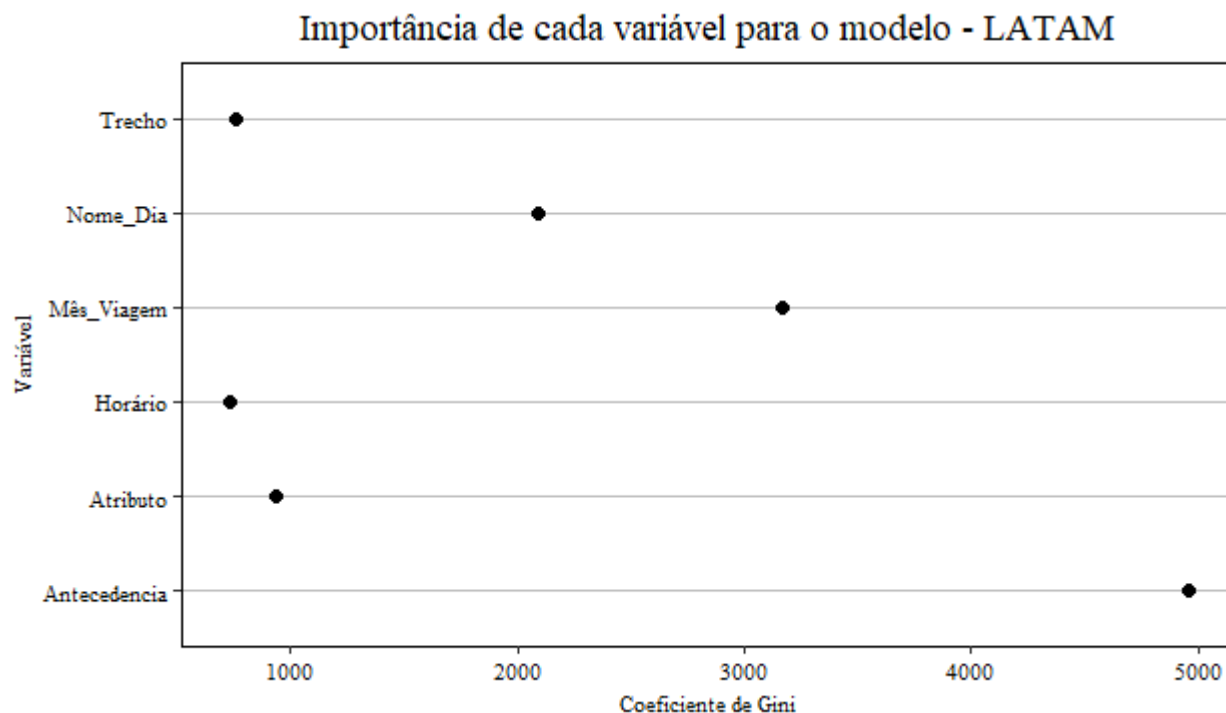


Figura 5.38: Importância de cada variável para o modelo de RF da empresa LATAM

Percebe-se desses modelos que a precisão para cada um deles pertence ao intervalo entre 75% e 83%, que não é o ideal, porém é um intervalo razoável. Além disso, pode-se perceber, por meio do gráfico de importância, que a antecedência e o mês da viagem são de fundamental importância para a classificação nos modelos, enquanto que o Trecho e o Horário não são fundamentais para essa classificação.

Com os modelos de classificação criados, realizou-se a previsão dos preços para cada linha existente na base a fim de ser ter e comparar a categoria de preço prevista e a categoria de preço real. Os valores obtidos são mostrados na Tabela 5.13.

Previsão	Referência					
	A	B	C	D	E	F
A	11263	3700	107	3	0	0
B	10079	166700	26353	2793	784	401
C	53	5878	34457	3197	1843	682
D	2	319	474	3959	321	86
E	0	147	622	759	3851	486
F	0	86	200	215	129	1927

Tabela 5.13: Matriz de confusão obtida após previsão com a base de dados completa

A Tabela 5.13 mostra que o modelo de classificação possui uma precisão de 78.81%

para a classificação da base de dados. O próximo passo foi se criar um modelo de previsão dos preços, o que foi obtido por meio de análises de regressões.

5.3.3 Regressão

Para se ter um modelo de previsão das passagens aéreas foi utilizado análises de regressões. A estratégia utilizada para a utilização das regressões foi criar uma para cada categoria de preço e empresa aérea, ou seja, foram criadas $6 \text{ categorias} \times 4 \text{ empresas} = 24$ regressões. Essa ação foi tomada por dois motivos:

1. O primeiro é que a faixa de preços existentes na base de dados tem uma amplitude muito grande, porém está concentrada abaixo dos R\$ 800.00, o que faz com que a regressão tenha uma tendência de encontrar valores muito mais baixos quando os preços originais forem altos. Logo, o primeiro motivo será para tentar ter uma melhor precisão em todas as categorias de preços.
2. O segundo é que cada empresa, teoricamente, possui uma estratégia de precificação diferente e, por isso, foi criado um grupo para cada uma delas.

A seguir serão ilustradas todos os coeficientes obtidos para cada uma das regressões calculadas. As Tabelas 5.14, 5.15, 5.16, 5.17, 5.18 e 5.19 ilustram os valores obtidos para os coeficientes de cada regressão proposta para cada categoria para a empresa aérea Azul. As regressões foram criadas transformando as variáveis categóricas “Trecho”, “Horário”, “Mês_Viagem”, “Nome_Dia” e “Atributo” em *dummies*.

Os códigos referentes aos níveis de significância são:

- ‘***’ < 0.001
- ‘**’ < 0.01
- ‘*’ < 0.05
- ‘.’ < 0.1
- ‘ ’ < 1

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	362.25657	10.95047	33.081	< 2e-16	***
TrechoGRU-FOR	-1.439476	1.311751	-1.097	0.27299	
Horário10-00	-1.19968	4.640953	-0.258	0.79613	
Horário12-00	-0.491753	4.468917	-0.11	0.91242	
Horário14-00	-1.411788	4.489052	-0.314	0.75327	
Horário16-00	-0.032366	4.467013	-0.007	0.99422	
Horário18-00	1.797198	4.293948	0.419	0.67572	
Horário20-00	2.218631	4.29082	0.517	0.60533	
Horário22-00	1.12618	4.296262	0.262	0.79333	
Antecedencia	0.026819	0.038726	0.693	0.48892	
Mês_Viagemnovembro	6.004873	6.252296	0.96	0.33729	
Mês_Viagemoutubro	-0.794334	5.977794	-0.133	0.89434	
Mês_Viagemsetembro	-0.009975	5.890197	-0.002	0.99865	
Nome_Diasegunda-feira	-4.584649	3.177668	-1.443	0.14969	
Nome_Diaterça-feira	-4.501651	3.121669	-1.442	0.14989	
Nome_Diaquarta-feira	-3.680821	3.010443	-1.223	0.22201	
Nome_Diaquinta-feira	-8.253419	3.076627	-2.683	0.00754	**
Nome_Diasexta-feira	6.83259	4.665702	1.464	0.14369	
Nome_Diasábado	-3.330073	3.419697	-0.974	0.33062	
AtributoNormal	24.677473	8.84056	2.791	0.00544	**
AtributoPós-Feriado	27.089512	10.228425	2.648	0.00833	**
AtributoVéspera Eleições	3.172995	10.600141	0.299	0.7648	

Residual standard error: 8.183 on 516 degrees of freedom

Multiple R-squared: 0.2316, Adjusted R-squared: 0.2004

F-statistic: 7.407 on 21 and 516 DF, p-value: < 2.2e-16

Tabela 5.14: Coeficientes da regressão para categoria econômica A da empresa Azul

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	571.32915	5.19155	110.05	< 2e-16	***
TrechoGRU-FOR	-6.33621	0.60563	-10.462	< 2e-16	***
Horário02-00	-1.49488	1.59718	-0.936	0.3493	
Horário04-00	-1.08635	1.59822	-0.68	0.4967	
Horário06-00	-39.61929	1.52109	-26.047	< 2e-16	***
Horário08-00	-41.59599	1.53801	-27.045	< 2e-16	***
Horário10-00	-49.17796	1.51382	-32.486	< 2e-16	***
Horário12-00	-47.39925	1.54041	-30.77	< 2e-16	***
Horário14-00	-47.72854	1.55307	-30.732	< 2e-16	***
Horário16-00	-48.44677	1.60471	-30.19	< 2e-16	***
Horário18-00	-52.46798	1.56817	-33.458	< 2e-16	***
Horário20-00	-51.66537	1.54234	-33.498	< 2e-16	***
Horário22-00	-50.3778	1.54929	-32.517	< 2e-16	***
Antecedencia	0.04575	0.01871	2.445	0.0145	*
Mês_Viagemdezembro	44.80407	18.28965	2.45	0.0143	*
Mês_Viagemjulho	14.44028	2.45489	5.882	4.09E-09	***
Mês_Viagemnovembro	-8.74041	1.51411	-5.773	7.88E-09	***
Mês_Viagemoutubro	-6.44791	1.08024	-5.969	2.41E-09	***
Mês_Viagemsetembro	-6.0905	0.89717	-6.789	1.15E-11	***
Nome_Diassegunda-feira	-16.04935	1.39785	-11.481	< 2e-16	***
Nome_Diaterça-feira	-27.20889	1.33112	-20.441	< 2e-16	***
Nome_Diaquarta-feira	-28.22997	1.33798	-21.099	< 2e-16	***
Nome_Diaquinta-feira	-33.05791	1.37375	-24.064	< 2e-16	***
Nome_Diasexta-feira	-17.15838	1.38169	-12.418	< 2e-16	***
Nome_Diasábado	-19.79986	1.44088	-13.742	< 2e-16	***
AtributoEleições 2018	-3.8026	5.47704	-0.694	0.4875	
AtributoFeriado	39.20234	6.88482	5.694	1.25E-08	***
AtributoNormal	-2.9378	5.15869	-0.569	0.569	
AtributoPós-Comemoração	-5.10468	6.6604	-0.766	0.4434	
AtributoPós-Feriado	-2.97064	5.50352	-0.54	0.5894	
AtributoVéspera de Comemoração	-2.47237	6.87521	-0.36	0.7191	
AtributoVéspera de Feriado	60.31983	7.35667	8.199	2.51E-16	***
AtributoVéspera Eleições 2018	9.14848	5.92322	1.545	0.1225	

Residual standard error: 51.54 on 30032 degrees of freedom

Multiple R-squared: 0.1566, Adjusted R-squared: 0.1557

F-statistic: 174.2 on 32 and 30032 DF, p-value: < 2.2e-16

Tabela 5.15: Coeficientes da regressão para categoria econômica B da empresa Azul

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	718.58077	4.81431	149.259	< 2e-16	***
TrechoGRU-FOR	3.64279	0.8769	4.154	3.28E-05	***
Horário02-00	-0.53682	1.78487	-0.301	0.763601	
Horário04-00	-0.54328	1.79777	-0.302	0.762504	
Horário06-00	-0.5658	1.88554	-0.3	0.764124	
Horário08-00	-0.38146	1.94252	-0.196	0.84432	
Horário10-00	0.05717	2.08485	0.027	0.978126	
Horário12-00	-0.19621	2.11752	-0.093	0.926175	
Horário14-00	2.7645	2.13132	1.297	0.19462	
Horário16-00	7.84236	2.22111	3.531	0.000415	***
Horário18-00	10.56846	2.21354	4.774	1.82E-06	***
Horário20-00	14.26035	2.17191	6.566	5.33E-11	***
Horário22-00	9.28692	2.18102	4.258	2.07E-05	***
Antecedencia	-1.79647	0.0341	-52.677	< 2e-16	***
Mês_Viagemdezembro	192.16954	21.28483	9.028	< 2e-16	***
Mês_Viagemjulho	-23.06288	2.16345	-10.66	< 2e-16	***
Mês_Viagemnovembro	130.18534	2.73576	47.587	< 2e-16	***
Mês_Viagemoutubro	47.89687	2.10541	22.749	< 2e-16	***
Mês_Viagemsetembro	14.156	1.3647	10.373	< 2e-16	***
Nome_Diassegunda-feira	-15.81047	1.60536	-9.849	< 2e-16	***
Nome_Diaterça-feira	-31.35548	1.71041	-18.332	< 2e-16	***
Nome_Diaquarta-feira	-29.43836	1.69268	-17.392	< 2e-16	***
Nome_Diaquinta-feira	-33.98611	1.78379	-19.053	< 2e-16	***
Nome_Diasexta-feira	-23.47703	1.7255	-13.606	< 2e-16	***
Nome_Diasábado	-10.74347	1.77752	-6.044	1.53E-09	***
AtributoEleições 2018	-4.86585	6.12527	-0.794	0.42698	
AtributoFeriado	53.55375	5.33619	10.036	< 2e-16	***
AtributoNormal	21.91956	4.76482	4.6	4.25E-06	***
AtributoPós-Comemoração	10.95948	6.37791	1.718	0.085751	.
AtributoPós-Feriado	10.87426	5.54893	1.96	0.050047	.
AtributoVéspera de Comemoração	-17.90611	6.94974	-2.577	0.009989	**
AtributoVéspera de Feriado	52.46721	5.42714	9.668	< 2e-16	***
AtributoVéspera Eleições 2018	7.76727	6.84271	1.135	0.256343	

Residual standard error: 55.79 on 16602 degrees of freedom

Multiple R-squared: 0.2709, Adjusted R-squared: 0.2695

F-statistic: 192.8 on 32 and 16602 DF, p-value: < 2.2e-16

Tabela 5.16: Coeficientes da regressão para categoria econômica C da empresa Azul

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	867.26529	4.41708	196.344	< 2e-16	***
TrechoGRU-FOR	4.68647	1.19121	3.934	8.48E-05	***
Horário02-00	-1.26604	2.38814	-0.53	0.596046	
Horário04-00	-2.01859	2.40695	-0.839	0.401713	
Horário06-00	-1.59024	2.53294	-0.628	0.530153	
Horário08-00	0.15997	2.54637	0.063	0.94991	
Horário10-00	1.52296	2.68258	0.568	0.570254	
Horário12-00	1.53012	2.74525	0.557	0.577303	
Horário14-00	-1.02097	2.78368	-0.367	0.713809	
Horário16-00	4.14944	2.93482	1.414	0.157476	
Horário18-00	3.21574	2.82857	1.137	0.255654	
Horário20-00	4.19235	2.80289	1.496	0.134801	
Horário22-00	0.12922	2.8479	0.045	0.963811	
Antecedencia	-1.53419	0.03954	-38.801	< 2e-16	***
Mês_Viagemjulho	-30.53369	2.00628	-15.219	< 2e-16	***
Mês_Viagemnovembro	138.05324	3.5522	38.864	< 2e-16	***
Mês_Viagemoutubro	84.00536	3.34962	25.079	< 2e-16	***
Mês_Viagemsetembro	15.60084	1.87214	8.333	< 2e-16	***
Nome_Diassegunda-feira	1.56362	1.94708	0.803	0.421984	
Nome_Diaterça-feira	-6.10519	2.56543	-2.38	0.017367	*
Nome_Diaquarta-feira	-9.24058	2.30591	-4.007	6.25E-05	***
Nome_Diaquinta-feira	5.93456	2.62518	2.261	0.023833	*
Nome_Diasexta-feira	2.73321	2.62659	1.041	0.298124	
Nome_Diasábado	10.55494	2.11068	5.001	5.94E-07	***
AtributoEleições 2018	49.0291	7.71238	6.357	2.27E-10	***
AtributoFeriado	0.26973	5.69008	0.047	0.962194	
AtributoNormal	15.03047	4.4275	3.395	0.000693	***
AtributoPós-Comemoração	-10.13897	7.48776	-1.354	0.175786	
AtributoPós-Feriado	-12.15179	6.7154	-1.81	0.070438	.
AtributoVéspera de Comemoração	-9.13473	6.8234	-1.339	0.180729	
AtributoVéspera de Feriado	13.08817	5.66864	2.309	0.020999	*
AtributoVéspera Eleições 2018	36.4948	7.41799	4.92	8.99E-07	***

Residual standard error: 36.55 on 4223 degrees of freedom

Multiple R-squared: 0.4015, Adjusted R-squared: 0.3971

F-statistic: 91.39 on 31 and 4223 DF, p-value: < 2.2e-16

Tabela 5.17: Coeficientes da regressão para categoria econômica D da empresa Azul

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	1045.3744	9.1266	114.542	< 2e-16	***
TrechoGRU-FOR	-24.6949	1.9755	-12.501	< 2e-16	***
Horário02-00	3.6234	4.0787	0.888	0.374411	
Horário04-00	7.8416	4.068	1.928	0.053993	.
Horário06-00	-2.7217	4.2531	-0.64	0.522267	
Horário08-00	-2.1553	4.3874	-0.491	0.623281	
Horário10-00	-12.7084	4.4193	-2.876	0.004059	**
Horário12-00	-12.6493	4.525	-2.795	0.005215	**
Horário14-00	-10.4517	4.6333	-2.256	0.024154	*
Horário16-00	-15.2699	4.6095	-3.313	0.000934	***
Horário18-00	-12.3995	4.7788	-2.595	0.009512	**
Horário20-00	-11.7883	4.6744	-2.522	0.011721	*
Horário22-00	-6.3262	4.687	-1.35	0.177203	
Antecedencia	2.0178	0.1542	13.082	< 2e-16	***
Mês_Viagemjulho	18.0735	2.3727	7.617	3.41E-14	***
Mês_Viagemnovembro	-174.857	12.7189	-13.748	< 2e-16	***
Mês_Viagemoutubro	-110.3534	12.6772	-8.705	< 2e-16	***
Mês_Viagemsetembro	-34.9729	3.853	-9.077	< 2e-16	***
Nome_Diasegunda-feira	0.395	3.3141	0.119	0.905122	
Nome_Diaterça-feira	-11.1414	3.9045	-2.854	0.004352	**
Nome_Diaquarta-feira	-19.7363	3.512	-5.62	2.08E-08	***
Nome_Diaquinta-feira	2.5201	3.8964	0.647	0.517827	
Nome_Diasexta-feira	17.3698	3.478	4.994	6.23E-07	***
Nome_Diasábado	-1.3967	3.7434	-0.373	0.709099	
AtributoEleições 2018	-9.6196	13.7396	-0.7	0.483891	
AtributoFeriado	85.171	12.4725	6.829	1.02E-11	***
AtributoNormal	36.5539	9.0062	4.059	5.05E-05	***
AtributoPós-Comemoração	34.4243	12.7119	2.708	0.006805	**
AtributoPós-Feriado	-11.6191	13.5445	-0.858	0.391043	
AtributoVéspera de Comemoração	-15.348	13.3669	-1.148	0.250971	
AtributoVéspera de Feriado	36.04	10.8299	3.328	0.000885	***
AtributoVéspera Eleições 2018	5.3238	52.517	0.101	0.919261	

Residual standard error: 51.24 on 3131 degrees of freedom

Multiple R-squared: 0.2873, Adjusted R-squared: 0.2803

F-statistic: 40.72 on 31 and 3131 DF, p-value: < 2.2e-16

Tabela 5.18: Coeficientes da regressão para categoria econômica E da empresa Azul

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	1522.9629	92.8811	16.397	< 2e-16	***
TrechoGRU-FOR	-30.9351	7.7295	-4.002	6.49E-05	***
Horário02-00	5.6649	13.7958	0.411	0.6814	
Horário04-00	2.7072	13.6619	0.198	0.8429	
Horário06-00	-22.9327	15.2043	-1.508	0.1316	
Horário08-00	-28.6661	14.9989	-1.911	0.0561	.
Horário10-00	-21.9298	16.63	-1.319	0.1874	
Horário12-00	-5.0872	17.2321	-0.295	0.7679	
Horário14-00	3.0887	17.5741	0.176	0.8605	
Horário16-00	28.6476	17.5556	1.632	0.1029	
Horário18-00	20.8333	17.8924	1.164	0.2444	
Horário20-00	-41.6214	17.6918	-2.353	0.0187	*
Horário22-00	-34.7374	16.5579	-2.098	0.036	*
Antecedencia	-10.6713	0.8932	-11.947	< 2e-16	***
Mês_Viagemjulho	57.6523	11.0635	5.211	2.06E-07	***
Mês_Viagemnovembro	934.7097	69.182	13.511	< 2e-16	***
Mês_Viagemoutubro	722.0538	67.6012	10.681	< 2e-16	***
Mês_Viagemsetembro	237.535	39.2501	6.052	1.68E-09	***
Nome_Diassegunda-feira	-6.2837	11.0613	-0.568	0.57	
Nome_Diaterça-feira	-19.0862	14.0848	-1.355	0.1755	
Nome_Diaquarta-feira	9.2002	15.1236	0.608	0.543	
Nome_Diaquinta-feira	7.598	14.3332	0.53	0.5961	
Nome_Diasexta-feira	-25.688	12.9544	-1.983	0.0475	*
Nome_Diasábado	1.7054	13.1593	0.13	0.8969	
AtributoFeriado	-65.0496	98.1729	-0.663	0.5077	
AtributoNormal	-111.1475	93.3555	-1.191	0.2339	
AtributoPós-Comemoração	26.3892	101.7149	0.259	0.7953	
AtributoPós-Feriado	-115.1151	99.3596	-1.159	0.2468	
AtributoVéspera de Comemoração	-75.047	113.8422	-0.659	0.5098	
AtributoVéspera de Feriado	-111.2675	95.9874	-1.159	0.2465	

Residual standard error: 159.5 on 2170 degrees of freedom

Multiple R-squared: 0.2178, Adjusted R-squared: 0.2073

F-statistic: 20.83 on 29 and 2170 DF, p-value: < 2.2e-16

Tabela 5.19: Coeficientes da regressão para categoria econômica F da empresa Azul

As Tabelas 5.20, 5.21, 5.22, 5.23, 5.24 e 5.25 mostram os coeficientes das regressões criadas para cada faixa de preço da empresa Avianca.

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	388.16849	3.73005	104.065	< 2e-16	***
TrechoGRU-FOR	3.31497	0.37383	8.868	< 2e-16	***
Horário02-00	-0.31024	0.80189	-0.387	0.698863	
Horário04-00	-1.62786	0.79653	-2.044	0.041069	*
Horário06-00	-1.69343	0.78704	-2.152	0.031504	*
Horário08-00	-1.63385	0.79121	-2.065	0.039006	*
Horário10-00	-2.79388	0.81026	-3.448	0.000572	***
Horário12-00	-4.45631	0.82144	-5.425	6.25E-08	***
Horário14-00	-4.45778	0.84223	-5.293	1.29E-07	***
Horário16-00	-4.10405	0.854	-4.806	1.62E-06	***
Horário18-00	-1.84038	0.8681	-2.12	0.034086	*
Horário20-00	-2.14984	0.83788	-2.566	0.010341	*
Horário22-00	-1.86222	0.81013	-2.299	0.021591	*
Antecedencia	-0.39146	0.02435	-16.076	< 2e-16	***
Mês_Viagemoutubro	21.69842	3.03371	7.152	1.06E-12	***
Mês_Viagemsetembro	11.45692	2.95736	3.874	0.000109	***
Nome_Diassegunda-feira	-2.81534	0.88361	-3.186	0.001456	**
Nome_Diaterça-feira	-4.84808	0.80439	-6.027	1.87E-09	***
Nome_Diaquarta-feira	-2.25551	0.81178	-2.778	0.005495	**
Nome_Diaquinta-feira	-1.85351	0.81172	-2.283	0.022473	*
Nome_Diasexta-feira	-3.2964	0.80751	-4.082	4.58E-05	***
Nome_Diasábado	-1.36588	0.86335	-1.582	0.113735	
AtributoFeriado	8.62955	3.31353	2.604	0.00925	**
AtributoNormal	2.3104	1.96155	1.178	0.23895	
AtributoVéspera Eleições	-2.4564	2.15621	-1.139	0.2547	

Residual standard error: 9.202 on 3068 degrees of freedom

Multiple R-squared: 0.1561, Adjusted R-squared: 0.1495

F-statistic: 23.64 on 24 and 3068 DF, p-value: < 2.2e-16

Tabela 5.20: Coeficientes da regressão para categoria econômica A da empresa Avianca

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	521.26814	2.845083	183.217	< 2e-16	***
TrechoGRU-FOR	-17.07136	0.481195	-35.477	< 2e-16	***
Horário02-00	-1.316469	1.093807	-1.204	0.228764	
Horário04-00	-1.338479	1.092997	-1.225	0.220735	
Horário06-00	0.153016	1.096609	0.14	0.889027	
Horário08-00	0.05435	1.11975	0.049	0.961288	
Horário10-00	-5.157593	1.133683	-4.549	5.40E-06	***
Horário12-00	-4.856501	1.153102	-4.212	2.54E-05	***
Horário14-00	-5.588966	1.165913	-4.794	1.64E-06	***
Horário16-00	-7.033198	1.20123	-5.855	4.81E-09	***
Horário18-00	-6.030123	1.169794	-5.155	2.55E-07	***
Horário20-00	-4.905437	1.1457	-4.282	1.86E-05	***
Horário22-00	-1.586994	1.146958	-1.384	0.166472	
Antecedencia	-0.005129	0.014406	-0.356	0.721792	
Mês_Viagemdezembro	-17.22149	19.081601	-0.903	0.366787	
Mês_Viagemjulho	21.134956	1.788568	11.817	< 2e-16	***
Mês_Viagemnovembro	0.691152	1.198436	0.577	0.564138	
Mês_Viagemoutubro	15.454105	0.901312	17.146	< 2e-16	***
Mês_Viagemsetembro	9.139673	0.675105	13.538	< 2e-16	***
Nome_Diassegunda-feira	-6.429331	0.976428	-6.585	4.62E-11	***
Nome_Diaterça-feira	-13.75523	0.947032	-14.525	< 2e-16	***
Nome_Diaquarta-feira	-17.77195	0.958804	-18.536	< 2e-16	***
Nome_Diaquinta-feira	-18.12013	0.981742	-18.457	< 2e-16	***
Nome_Diasexta-feira	-23.16115	0.984779	-23.519	< 2e-16	***
Nome_Diasábado	-6.905787	1.009884	-6.838	8.13E-12	***
AtributoEleições 2018	-41.05248	3.205604	-12.806	< 2e-16	***
AtributoFeriado	12.063352	3.21449	3.753	0.000175	***
AtributoNormal	-13.50488	2.833683	-4.766	1.89E-06	***
AtributoPós-Comemoração	-19.11214	4.13822	-4.618	3.88E-06	***
AtributoPós-Feriado	-0.411078	3.282848	-0.125	0.90035	
AtributoVéspera de Comemoração	-0.142203	4.220341	-0.034	0.973121	
AtributoVéspera de Feriado	7.044962	3.301877	2.134	0.03288	*
AtributoVéspera Eleições 2018	-16.26718	3.733264	-4.357	1.32E-05	***

Residual standard error: 46.63 on 40217 degrees of freedom

Multiple R-squared: 0.08207, Adjusted R-squared: 0.08134

F-statistic: 112.4 on 32 and 40217 DF, p-value: < 2.2e-16

Tabela 5.21: Coeficientes da regressão para categoria econômica B da empresa Avianca

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	626.1169	4.53471	138.072	< 2e-16	***
TrechoGRU-FOR	-24.30881	0.98656	-24.64	< 2e-16	***
Horário02-00	-0.33463	1.97416	-0.17	0.865401	
Horário04-00	-1.0072	1.98711	-0.507	0.612258	
Horário06-00	2.03637	1.98034	1.028	0.303836	
Horário08-00	-0.78171	2.01168	-0.389	0.69759	
Horário10-00	0.26271	2.00512	0.131	0.895763	
Horário12-00	0.85729	2.07475	0.413	0.679467	
Horário14-00	2.24729	2.13362	1.053	0.292238	
Horário16-00	2.48526	2.22952	1.115	0.264999	
Horário18-00	-1.19987	2.1569	-0.556	0.578023	
Horário20-00	0.72484	2.12926	0.34	0.733549	
Horário22-00	1.23562	2.14792	0.575	0.565127	
Antecedencia	0.2665	0.03472	7.677	1.77E-14	***
Mês_Viagemdezembro	-36.66634	18.16585	-2.018	0.043572	*
Mês_Viagemjulho	52.86782	1.45453	36.347	< 2e-16	***
Mês_Viagemnovembro	-9.3828	2.80609	-3.344	0.000829	***
Mês_Viagemoutubro	-21.52422	2.05418	-10.478	< 2e-16	***
Mês_Viagemsetembro	-25.3957	1.59466	-15.926	< 2e-16	***
Nome_Diassegunda-feira	-4.87087	1.49621	-3.255	0.001135	**
Nome_Diaterça-feira	-14.42976	1.71058	-8.436	< 2e-16	***
Nome_Diaquarta-feira	-19.83407	1.53721	-12.903	< 2e-16	***
Nome_Diaquinta-feira	-18.32342	1.86716	-9.814	< 2e-16	***
Nome_Diasexta-feira	-28.47532	1.77921	-16.004	< 2e-16	***
Nome_Diasábado	-18.7616	1.83081	-10.248	< 2e-16	***
AtributoFeriado	72.22577	5.38698	13.407	< 2e-16	***
AtributoNormal	31.2951	4.493	6.965	3.47E-12	***
AtributoPós-Comemoração	15.98995	5.80248	2.756	0.005866	**
AtributoPós-Feriado	74.14693	5.157	14.378	< 2e-16	***
AtributoVéspera de Comemoração	8.81201	5.8689	1.501	0.133261	
AtributoVéspera de Feriado	74.59828	5.22869	14.267	< 2e-16	***
AtributoVéspera Eleições 2018	24.85127	44.13638	0.563	0.573408	

Residual standard error: 43.83 on 10911 degrees of freedom

Multiple R-squared: 0.2395, Adjusted R-squared: 0.2373

F-statistic: 110.8 on 31 and 10911 DF, p-value: < 2.2e-16

Tabela 5.22: Coeficientes da regressão para categoria econômica C da empresa Avianca

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	875.99974	5.80721	150.847	< 2e-16	***
TrechoGRU-FOR	-42.44772	4.51158	-9.409	< 2e-16	***
Horário02-00	-2.54009	5.3723	-0.473	0.63641	
Horário04-00	-2.22383	5.44621	-0.408	0.68309	
Horário06-00	-5.67721	5.21002	-1.09	0.27603	
Horário08-00	-3.15058	5.33747	-0.59	0.55509	
Horário10-00	-4.80699	5.50393	-0.873	0.38259	
Horário12-00	-3.32243	5.4822	-0.606	0.54458	
Horário14-00	-0.1571	5.62939	-0.028	0.97774	
Horário16-00	-0.05565	5.92769	-0.009	0.99251	
Horário18-00	-1.92321	5.86483	-0.328	0.74301	
Horário20-00	-4.38505	5.96552	-0.735	0.46241	
Horário22-00	-3.06854	5.82131	-0.527	0.59818	
Antecedencia	-1.8738	0.16597	-11.29	< 2e-16	***
Mês_Viagemjulho	27.10696	3.75117	7.226	7.73E-13	***
Mês_Viagemnovembro	96.11479	13.50866	7.115	1.70E-12	***
Mês_Viagemoutubro	42.20091	9.35962	4.509	7.00E-06	***
Mês_Viagemsetembro	-50.93503	31.2785	-1.628	0.10363	
Nome_Diassegunda-feira	-2.21318	3.49809	-0.633	0.52703	
Nome_Diaterça-feira	2.93661	3.94031	0.745	0.45622	
Nome_Diaquarta-feira	14.3481	4.89822	2.929	0.00345	**
Nome_Diaquinta-feira	20.76572	5.15396	4.029	5.87E-05	***
Nome_Diasexta-feira	-6.49128	4.48442	-1.448	0.14795	
Nome_Diasábado	-3.96013	4.19054	-0.945	0.3448	
AtributoPós-Comemoração	33.69548	31.31169	1.076	0.28203	
AtributoVéspera de Feriado	24.74189	8.89997	2.78	0.0055	**

Residual standard error: 43.67 on 1564 degrees of freedom

Multiple R-squared: 0.3691, Adjusted R-squared: 0.359

F-statistic: 36.6 on 25 and 1564 DF, p-value: < 2.2e-16

Tabela 5.23: Coeficientes da regressão para categoria econômica D da empresa Avianca

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	1204.9943	17.8422	67.536	< 2e-16	***
Horário02-00	-11.1759	13.1403	-0.851	0.395743	
Horário04-00	-9.7444	13.0199	-0.748	0.454808	
Horário06-00	-6.0493	13.3251	-0.454	0.650185	
Horário08-00	-6.2665	12.8605	-0.487	0.626437	
Horário10-00	-19.0695	14.4404	-1.321	0.187682	
Horário12-00	-11.717	14.5284	-0.806	0.420619	
Horário14-00	-22.4363	14.2521	-1.574	0.116516	
Horário16-00	-30.613	15.0568	-2.033	0.042942	*
Horário18-00	-24.5	18.4405	-1.329	0.185023	
Horário20-00	-26.9067	16.1734	-1.664	0.097261	.
Horário22-00	-0.205	14.645	-0.014	0.988839	
Antecedencia	-8.9985	0.9185	-9.797	< 2e-16	***
Mês_Viagemjulho	-2.2583	15.2446	-0.148	0.882336	
Nome_Diasegunda-feira	-29.0072	7.9018	-3.671	0.000287	***
Nome_Diaterça-feira	-9.0953	12.8013	-0.71	0.477965	
Nome_Diaquarta-feira	-62.8056	17.8617	-3.516	0.000508	***
Nome_Diaquinta-feira	-45.3985	14.1167	-3.216	0.001447	**
Nome_Diasexta-feira	-52.8675	10.0781	-5.246	3.00E-07	***
Nome_Diasábado	-20.1559	8.9378	-2.255	0.024868	*
TrechoGRU-FOR	-109.1351	18.379	-5.938	8.21E-09	***

Residual standard error: 50.24 on 291 degrees of freedom

Multiple R-squared: 0.4296, Adjusted R-squared: 0.3904

F-statistic: 10.96 on 20 and 291 DF, p-value: < 2.2e-16

Tabela 5.24: Coeficientes da regressão para categoria econômica E da empresa Avianca

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	2017.59	257.57	7.833	4.06E-09	***
Horário02-00	101.15	158.94	0.636	0.52879	
Horário04-00	117.46	153.47	0.765	0.44933	
Horário06-00	42.05	148.34	0.283	0.77853	
Horário08-00	3.14	162.78	0.019	0.98472	
Horário10-00	-335.76	219.91	-1.527	0.13606	
Horário12-00	-120.66	180.99	-0.667	0.50948	
Horário14-00	-117.01	176.41	-0.663	0.51162	
Horário16-00	-44.27	240.17	-0.184	0.85487	
Horário18-00	68.98	212.59	0.324	0.74757	
Horário20-00	124.54	283.08	0.44	0.66276	
Horário22-00	133.44	231.96	0.575	0.56889	
Antecedencia	-86.93	27.22	-3.193	0.00303	**
Mês_Viagemjulho	-370.36	217.1	-1.706	0.09713	.
Nome_Diasegunda-feira	-69.12	85.33	-0.81	0.42357	
Nome_Diaterça-feira	-304.17	115.15	-2.642	0.01238	*
Nome_Diaquarta-feira	45.78	276.8	0.165	0.86962	
Nome_Diaquinta-feira	-466.12	164.79	-2.828	0.00779	**
Nome_Diasexta-feira	-214.66	144.5	-1.485	0.14663	
TrechoGRU-FOR	-115.7	156.76	-0.738	0.4655	

Residual standard error: 242 on 34 degrees of freedom

Multiple R-squared: 0.4251, Adjusted R-squared: 0.1038

F-statistic: 1.323 on 19 and 34 DF, p-value: 0.2325

Tabela 5.25: Coeficientes da regressão para categoria econômica F da empresa Avianca

As Tabelas 5.26, 5.27, 5.28, 5.29, 5.30 e 5.31 mostram os coeficientes das regressões criadas para cada faixa de preço da empresa Gol.

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	240.76353	0.882657	272.771	< 2e-16	***
Horário02-00	0.040173	0.296716	0.135	0.89237	
Horário04-00	0.020828	0.28985	0.072	0.94275	
Horário06-00	-0.11256	0.307248	-0.366	0.71428	
Horário08-00	-0.20758	0.293313	-0.708	0.47951	
Horário10-00	-0.123006	0.2877	-0.428	0.66919	
Horário12-00	0.339245	0.302878	1.12	0.26331	
Horário14-00	0.336978	0.296818	1.135	0.25688	
Horário16-00	-0.003543	0.303645	-0.012	0.9907	
Horário18-00	0.226475	0.290444	0.78	0.43596	
Horário20-00	0.353317	0.292887	1.206	0.22835	
Horário22-00	-0.089435	0.308113	-0.29	0.77175	
Antecedencia	-0.15927	0.008106	-19.649	< 2e-16	***
Mês_Viagemjulho	-0.837756	0.890361	-0.941	0.34728	
Mês_Viagemnovembro	164.08625	0.911418	180.034	< 2e-16	***
Mês_Viagemoutubro	157.98875	0.927762	170.29	< 2e-16	***
Mês_Viagemsetembro	154.7235	0.835711	185.14	< 2e-16	***
Nome_Diassegunda-feira	0.345963	0.343315	1.008	0.31416	
Nome_Diaterça-feira	-0.19972	0.377886	-0.529	0.59741	
Nome_Diaquarta-feira	-0.021788	0.406336	-0.054	0.95726	
Nome_Diaquinta-feira	0.347032	0.416069	0.834	0.4047	
Nome_Diasexta-feira	0.292424	0.341239	0.857	0.39195	
Nome_Diasábado	0.764717	0.285451	2.679	0.00767	**

Residual standard error: 1.273 on 430 degrees of freedom

Multiple R-squared: 0.9969, Adjusted R-squared: 0.9968

F-statistic: 6384 on 22 and 430 DF, p-value: < 2.2e-16

Tabela 5.26: Coeficientes da regressão para categoria econômica A da empresa Gol

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	518.31959	3.75549	138.016	< 2e-16	***
TrechoGRU-FOR	5.09483	0.41919	12.154	< 2e-16	***
Horário02-00	1.33958	0.99404	1.348	0.177791	
Horário04-00	1.98777	1.00375	1.98	0.047672	*
Horário06-00	2.08476	1.00114	2.082	0.037315	*
Horário08-00	2.33759	1.00897	2.317	0.020519	*
Horário10-00	1.39562	1.00984	1.382	0.166973	
Horário12-00	1.13515	1.02407	1.108	0.267667	
Horário14-00	1.90586	1.03486	1.842	0.065533	.
Horário16-00	0.51888	1.05958	0.49	0.624343	
Horário18-00	-1.2447	1.03718	-1.2	0.230115	
Horário20-00	-0.17403	1.02615	-0.17	0.865329	
Horário22-00	0.44869	1.04562	0.429	0.66784	
Antecedencia	-0.72696	0.01242	-58.548	< 2e-16	***
Mês_Viagemdezembro	75.79927	12.52253	6.053	1.44E-09	***
Mês_Viagemjulho	-7.29646	1.91462	-3.811	0.000139	***
Mês_Viagemnovembro	14.74252	1.10656	13.323	< 2e-16	***
Mês_Viagemoutubro	-6.8476	0.78737	-8.697	< 2e-16	***
Mês_Viagemsetembro	-5.88984	0.62438	-9.433	< 2e-16	***
Nome_Diassegunda-feira	-4.91602	0.84241	-5.836	5.40E-09	***
Nome_Diaterça-feira	-11.2777	0.81908	-13.769	< 2e-16	***
Nome_Diaquarta-feira	-14.34909	0.82251	-17.445	< 2e-16	***
Nome_Diaquinta-feira	-12.96104	0.85323	-15.191	< 2e-16	***
Nome_Diasexta-feira	-13.95489	0.85271	-16.365	< 2e-16	***
Nome_Diasábado	-11.35322	0.86855	-13.071	< 2e-16	***
AtributoEleições 2018	36.53161	4.07776	8.959	< 2e-16	***
AtributoFeriado	60.33338	4.09163	14.746	< 2e-16	***
AtributoNormal	22.68434	3.74193	6.062	1.36E-09	***
AtributoPós-Comemoração	7.03331	5.5091	1.277	0.201727	
AtributoPós-Feriado	47.52969	3.98331	11.932	< 2e-16	***
AtributoVéspera de Comemoração	24.92731	5.28614	4.716	2.42E-06	***
AtributoVéspera de Feriado	69.31326	4.12147	16.818	< 2e-16	***
AtributoVéspera Eleições 2018	65.84827	4.56162	14.435	< 2e-16	***

Residual standard error: 39.45 on 36371 degrees of freedom

Multiple R-squared: 0.1942, Adjusted R-squared: 0.1935

F-statistic: 273.9 on 32 and 36371 DF, p-value: < 2.2e-16

Tabela 5.27: Coeficientes da regressão para categoria econômica B da empresa Gol

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	735.83806	3.43548	214.188	< 2e-16	***
TrechoGRU-FOR	5.80448	0.82151	7.066	1.67E-12	***
Horário02-00	-0.67965	1.83156	-0.371	0.710586	
Horário04-00	1.10697	1.82831	0.605	0.544881	
Horário06-00	-0.39532	1.81151	-0.218	0.827257	
Horário08-00	-0.76393	1.8628	-0.41	0.68174	
Horário10-00	-5.47146	1.9132	-2.86	0.004244	**
Horário12-00	-1.62645	2.00235	-0.812	0.416648	
Horário14-00	-2.22598	2.01619	-1.104	0.26959	
Horário16-00	-3.87449	2.09455	-1.85	0.064363	.
Horário18-00	-4.63228	2.02796	-2.284	0.022373	*
Horário20-00	-3.11475	1.97484	-1.577	0.114768	
Horário22-00	-1.12107	1.96476	-0.571	0.568287	
Antecedencia	-0.31653	0.03106	-10.19	< 2e-16	***
Mês_Viagemjulho	-22.37534	1.52383	-14.684	< 2e-16	***
Mês_Viagemnovembro	1.88974	2.39964	0.788	0.430998	
Mês_Viagemoutubro	-74.89224	1.86458	-40.166	< 2e-16	***
Mês_Viagemsetembro	-50.23899	1.35375	-37.111	< 2e-16	***
Nome_Diassegunda-feira	-14.54083	1.64394	-8.845	< 2e-16	***
Nome_Diaterça-feira	-8.78619	1.64329	-5.347	9.09E-08	***
Nome_Diaquarta-feira	-2.67984	1.64085	-1.633	0.102448	
Nome_Diaquinta-feira	-8.31969	1.72535	-4.822	1.44E-06	***
Nome_Diasexta-feira	-6.25719	1.72537	-3.627	0.000288	***
Nome_Diasábado	-6.24455	1.91274	-3.265	0.001098	**
AtributoEleições 2018	-30.20375	4.32351	-6.986	2.95E-12	***
AtributoFeriado	-21.83524	4.13419	-5.282	1.30E-07	***
AtributoNormal	-17.25218	3.40806	-5.062	4.19E-07	***
AtributoPós-Comemoração	-8.95685	4.58386	-1.954	0.05072	.
AtributoPós-Feriado	-34.31112	4.43335	-7.739	1.06E-14	***
AtributoVéspera de Comemoração	-14.15865	4.73308	-2.991	0.002781	**
AtributoVéspera de Feriado	-19.04434	4.1007	-4.644	3.44E-06	***
AtributoVéspera Eleições 2018	-24.43557	4.76119	-5.132	2.90E-07	***

Residual standard error: 48.59 on 14939 degrees of freedom

Multiple R-squared: 0.4065, Adjusted R-squared: 0.4053

F-statistic: 330.1 on 31 and 14939 DF, p-value: < 2.2e-16

Tabela 5.28: Coeficientes da regressão para categoria econômica C da empresa Gol

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	829.8607	26.2755	31.583	< 2e-16	***
TrechoGRU-FOR	-20.7068	2.3913	-8.659	< 2e-16	***
Horário02-00	-9.0665	4.5836	-1.978	0.048088	*
Horário04-00	-5.7426	4.4996	-1.276	0.202036	
Horário06-00	-7.2167	4.5996	-1.569	0.116842	
Horário08-00	-4.1151	4.5871	-0.897	0.369792	
Horário10-00	-2.5998	4.4394	-0.586	0.558209	
Horário12-00	-3.1215	4.5214	-0.69	0.49004	
Horário14-00	0.8078	4.6136	0.175	0.861025	
Horário16-00	-1.0642	4.597	-0.232	0.816952	
Horário18-00	2.6575	4.6096	0.577	0.564345	
Horário20-00	6.5367	4.6765	1.398	0.162362	
Horário22-00	3.6655	4.9025	0.748	0.454759	
Antecedencia	-1.37	0.1362	-10.06	< 2e-16	***
Mês_Viagemjulho	14.903	2.7438	5.432	6.40E-08	***
Mês_Viagemnovembro	90.715	10.7449	8.443	< 2e-16	***
Mês_Viagemsetembro	42.4632	6.2303	6.816	1.30E-11	***
Nome_Diassegunda-feira	7.5313	3.1406	2.398	0.016589	*
Nome_Diaterça-feira	-14.3411	3.4555	-4.15	3.49E-05	***
Nome_Diaquarta-feira	-12.1525	3.4003	-3.574	0.000362	***
Nome_Diaquinta-feira	-10.8671	3.8486	-2.824	0.004804	**
Nome_Diasexta-feira	-6.6094	3.3972	-1.946	0.051877	.
Nome_Diasábado	11.6356	3.5584	3.27	0.001098	**
AtributoFeriado	113.5633	27.1101	4.189	2.95E-05	***
AtributoNormal	76.7915	26.2866	2.921	0.003532	**
AtributoPós-Comemoração	1.1556	27.2008	0.042	0.966119	
AtributoVéspera de Comemoração	-18.463	37.0234	-0.499	0.618067	
AtributoVéspera de Feriado	116.206	27.1876	4.274	2.03E-05	***

Residual standard error: 36.73 on 1688 degrees of freedom

Multiple R-squared: 0.204, Adjusted R-squared: 0.1913

F-statistic: 16.02 on 27 and 1688 DF, p-value: < 2.2e-16

Tabela 5.29: Coeficientes da regressão para categoria econômica D da empresa Gol

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	1213.0026	13.3864	90.614	< 2e-16	***
TrechoGRU-FOR	-77.2354	3.892	-19.845	< 2e-16	***
Horário02-00	-6.7855	6.3625	-1.066	0.28636	
Horário04-00	-2.2477	6.2738	-0.358	0.72019	
Horário06-00	5.2207	6.2689	0.833	0.40507	
Horário08-00	6.0585	6.2568	0.968	0.33303	
Horário10-00	-5.3407	6.3108	-0.846	0.39751	
Horário12-00	-10.3521	6.4305	-1.61	0.10761	
Horário14-00	-9.4831	6.6978	-1.416	0.157	
Horário16-00	-1.4357	6.6844	-0.215	0.82996	
Horário18-00	-3.3847	6.933	-0.488	0.62547	
Horário20-00	-2.1045	7.1163	-0.296	0.76747	
Horário22-00	7.8819	7.0893	1.112	0.26638	
Antecedencia	-2.5443	0.2063	-12.332	< 2e-16	***
Mês_Viagemjulho	15.3864	3.897	3.948	8.19E-05	***
Mês_Viagemnovembro	125.5698	15.0296	8.355	< 2e-16	***
Mês_Viagemsetembro	-84.0937	54.1298	-1.554	0.12047	
Nome_Diassegunda-feira	-20.9136	3.8983	-5.365	9.21E-08	***
Nome_Diaterça-feira	-11.3072	5.1233	-2.207	0.02745	*
Nome_Diaquarta-feira	-1.5208	5.0881	-0.299	0.76506	
Nome_Diaquinta-feira	-12.2178	5.5018	-2.221	0.0265	*
Nome_Diasexta-feira	-14.3497	5.3233	-2.696	0.00709	**
Nome_Diasábado	-30.1196	4.9532	-6.081	1.47E-09	***
AtributoNormal	-45.5567	10.9721	-4.152	3.46E-05	***
AtributoVéspera de Feriado	-5.0749	13.3332	-0.381	0.70353	

Residual standard error: 53.74 on 1715 degrees of freedom

Multiple R-squared: 0.3148, Adjusted R-squared: 0.3052

F-statistic: 32.84 on 24 and 1715 DF, p-value: < 2.2e-16

Tabela 5.30: Coeficientes da regressão para categoria econômica E da empresa Gol

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	1535.987	46.157	33.278	< 2e-16	***
TrechoGRU-FOR	-15.911	27.019	-0.589	0.55638	
Horário02-00	-14.395	43.275	-0.333	0.73962	
Horário04-00	9.292	44.007	0.211	0.83291	
Horário06-00	35.786	41.893	0.854	0.39364	
Horário08-00	30.512	41.077	0.743	0.45816	
Horário10-00	20.396	45.918	0.444	0.65722	
Horário12-00	34.264	43.126	0.795	0.4275	
Horário14-00	7.651	46.604	0.164	0.8697	
Horário16-00	-40.926	47.735	-0.857	0.3919	
Horário18-00	-24.785	45.405	-0.546	0.58555	
Horário20-00	-40.601	46.688	-0.87	0.38518	
Horário22-00	3.449	48.445	0.071	0.94329	
Antecedência	-14.185	1.657	-8.562	5.15E-16	***
Mês_Viagemjulho	76.761	34.771	2.208	0.028	*
Mês_Viagemnovembro	628.524	111.684	5.628	4.07E-08	***
Nome_Diassegunda-feira	125.528	21.365	5.876	1.08E-08	***
Nome_Diaterça-feira	32.487	33.906	0.958	0.33873	
Nome_Diaquarta-feira	72.665	43.042	1.688	0.09236	.
Nome_Diaquinta-feira	105.413	71.738	1.469	0.14273	
Nome_Diasexta-feira	57.84	55.714	1.038	0.3	
Nome_Diasábado	-42.608	33.299	-1.28	0.20166	
AtributoPós-Comemoração	-315.968	117.536	-2.688	0.00757	**

Residual standard error: 151.9 on 312 degrees of freedom

Multiple R-squared: 0.6007, Adjusted R-squared: 0.5726

F-statistic: 21.34 on 22 and 312 DF, p-value: < 2.2e-16

Tabela 5.31: Coeficientes da regressão para categoria econômica F da empresa Gol

As Tabelas 5.32, 5.33, 5.34, 5.35, 5.36 e 5.37 mostram os coeficientes das regressões criadas para cada faixa de preço da empresa LATAM.

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	381.15727	1.57071	242.666	< 2e-16	***
TrechoGRU-FOR	-1.99227	0.37512	-5.311	1.11E-07	***
Horário02-00	-0.11288	0.83166	-0.136	0.892038	
Horário04-00	0.50232	0.8418	0.597	0.550704	
Horário06-00	1.60107	0.85093	1.882	0.059919	.
Horário08-00	2.05697	0.87736	2.345	0.019067	*
Horário10-00	-2.54664	0.85897	-2.965	0.003035	**
Horário12-00	-1.36561	0.87559	-1.56	0.118869	
Horário14-00	-0.93963	0.85866	-1.094	0.273846	
Horário16-00	-0.81387	0.89232	-0.912	0.361738	
Horário18-00	-2.93292	0.84073	-3.489	0.000487	***
Horário20-00	-1.43621	0.84666	-1.696	0.089847	.
Horário22-00	-0.65494	0.86521	-0.757	0.449086	
Antecedencia	0.16542	0.01288	12.84	< 2e-16	***
Mês_Viagemdezembro	3.63907	11.68392	0.311	0.755456	
Mês_Viagemjulho	15.21147	2.96624	5.128	2.97E-07	***
Mês_Viagemnovembro	-17.99117	1.09251	-16.468	< 2e-16	***
Mês_Viagemoutubro	-21.65373	0.93112	-23.256	< 2e-16	***
Mês_Viagemsetembro	-8.58049	0.89599	-9.577	< 2e-16	***
Nome_Diasegunda-feira	1.8612	1.02834	1.81	0.070333	.
Nome_Diaterça-feira	-1.94382	1.00004	-1.944	0.051947	.
Nome_Diaquarta-feira	-5.05832	0.9979	-5.069	4.06E-07	***
Nome_Diaquinta-feira	-5.5294	1.01143	-5.467	4.66E-08	***
Nome_Diasexta-feira	-5.2247	1.02729	-5.086	3.71E-07	***
Nome_Diasábado	-8.35125	1.26512	-6.601	4.24E-11	***
AtributoFeriado	-7.04406	1.9587	-3.596	0.000324	***
AtributoNormal	-6.8534	1.4405	-4.758	1.98E-06	***
AtributoPós-Comemoração	-10.92917	5.30957	-2.058	0.039573	*
AtributoPós-Feriado	-4.83306	1.94247	-2.488	0.012855	*
AtributoVéspera de Feriado	-4.18098	1.90908	-2.19	0.028539	*
AtributoVéspera Eleições 2018	6.41686	2.95474	2.172	0.029895	*

Residual standard error: 20.08 on 12937 degrees of freedom

Multiple R-squared: 0.09274, Adjusted R-squared: 0.09063

F-statistic: 44.08 on 30 and 12937 DF, p-value: < 2.2e-16

Tabela 5.32: Coeficientes da regressão para categoria econômica A da empresa LATAM

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	495.32943	3.55138	139.475	< 2e-16	***
TrechoGRU-FOR	-1.03493	0.49068	-2.109	0.034938	*
Horário02-00	-1.01142	1.13799	-0.889	0.374131	
Horário04-00	0.21934	1.13693	0.193	0.847018	
Horário06-00	2.70247	1.14273	2.365	0.018039	*
Horário08-00	2.00237	1.15545	1.733	0.083108	.
Horário10-00	1.13937	1.16758	0.976	0.329151	
Horário12-00	-0.03734	1.19032	-0.031	0.974973	
Horário14-00	-0.48214	1.20504	-0.4	0.689086	
Horário16-00	-2.38978	1.25624	-1.902	0.057136	.
Horário18-00	-5.32327	1.22667	-4.34	1.43E-05	***
Horário20-00	-3.87921	1.20349	-3.223	0.001268	**
Horário22-00	-0.99266	1.20392	-0.825	0.409647	
Antecedencia	-0.80917	0.01453	-55.696	< 2e-16	***
Mês_Viagemdezembro	57.06261	15.9363	3.581	0.000343	***
Mês_Viagemjulho	38.14891	1.4674	25.998	< 2e-16	***
Mês_Viagemnovembro	8.29603	1.30006	6.381	1.78E-10	***
Mês_Viagemoutubro	2.70337	0.94668	2.856	0.004298	**
Mês_Viagemsetembro	8.65489	0.67544	12.814	< 2e-16	***
Nome_Diassegunda-feira	-13.21152	0.95474	-13.838	< 2e-16	***
Nome_Diaterça-feira	-21.40523	0.96412	-22.202	< 2e-16	***
Nome_Diaquarta-feira	-25.25167	0.97651	-25.859	< 2e-16	***
Nome_Diaquinta-feira	-25.71449	1.02885	-24.993	< 2e-16	***
Nome_Diasexta-feira	-21.91764	0.99956	-21.927	< 2e-16	***
Nome_Diasábado	-2.08014	0.9781	-2.127	0.033451	*
AtributoEleições 2018	7.77412	4.01168	1.938	0.052647	.
AtributoFeriado	27.69281	3.84967	7.194	6.44E-13	***
AtributoNormal	20.98133	3.52743	5.948	2.74E-09	***
AtributoPós-Comemoração	8.90192	4.48663	1.984	0.047253	*
AtributoPós-Feriado	58.89894	3.8672	15.23	< 2e-16	***
AtributoVéspera de Comemoração	9.25406	5.05644	1.83	0.067236	.
AtributoVéspera de Feriado	45.3851	3.92318	11.568	< 2e-16	***
AtributoVéspera Eleições 2018	23.20333	4.19981	5.525	3.32E-08	***

Residual standard error: 44.94 on 34622 degrees of freedom

Multiple R-squared: 0.209, Adjusted R-squared: 0.2083

F-statistic: 285.9 on 32 and 34622 DF, p-value: < 2.2e-16

Tabela 5.33: Coeficientes da regressão para categoria econômica B da empresa LATAM

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	654.77129	4.30414	152.126	< 2e-16	***
TrechoGRU-FOR	-18.53351	1.30085	-14.247	< 2e-16	***
Horário02-00	0.63252	2.82857	0.224	0.823061	
Horário04-00	3.12129	2.74984	1.135	0.256379	
Horário06-00	11.49129	2.70045	4.255	2.11E-05	***
Horário08-00	6.01123	2.77949	2.163	0.030597	*
Horário10-00	6.59171	2.88537	2.285	0.02237	*
Horário12-00	-0.22936	2.99641	-0.077	0.938987	
Horário14-00	-0.18294	3.1119	-0.059	0.953124	
Horário16-00	3.3249	3.12771	1.063	0.287796	
Horário18-00	1.11243	3.17133	0.351	0.725766	
Horário20-00	0.97943	3.06736	0.319	0.749503	
Horário22-00	-0.5048	2.99497	-0.169	0.866155	
Antecedencia	-0.4005	0.05901	-6.787	1.24E-11	***
Mês_Viagemjulho	24.46585	1.58881	15.399	< 2e-16	***
Mês_Viagemnovembro	60.69321	5.10468	11.89	< 2e-16	***
Mês_Viagemoutubro	65.308	3.56807	18.303	< 2e-16	***
Mês_Viagemsetembro	2.74568	2.14026	1.283	0.199579	
Nome_Diassegunda-feira	-9.74751	2.2264	-4.378	1.21E-05	***
Nome_Diaterça-feira	-16.61862	2.42701	-6.847	8.15E-12	***
Nome_Diaquarta-feira	-18.23109	2.37958	-7.661	2.08E-14	***
Nome_Diaquinta-feira	-14.54658	2.43696	-5.969	2.50E-09	***
Nome_Diasexta-feira	-13.78703	2.46562	-5.592	2.33E-08	***
Nome_Diasábado	-4.14503	2.27258	-1.824	0.068204	.
AtributoEleições 2018	-74.28631	18.22305	-4.077	4.62E-05	***
AtributoFeriado	0.66266	8.55521	0.077	0.938262	
AtributoNormal	19.32865	4.0012	4.831	1.39E-06	***
AtributoPós-Comemoração	2.87399	6.09448	0.472	0.637246	
AtributoPós-Feriado	-4.78301	6.32113	-0.757	0.449272	
AtributoVéspera de Comemoração	14.5615	5.52209	2.637	0.008384	**
AtributoVéspera de Feriado	9.4838	5.82562	1.628	0.103581	
AtributoVéspera Eleições 2018	-53.34301	14.85294	-3.591	0.000331	***

Residual standard error: 49.96 on 7111 degrees of freedom

Multiple R-squared: 0.1494, Adjusted R-squared: 0.1456

F-statistic: 40.27 on 31 and 7111 DF, p-value: < 2.2e-16

Tabela 5.34: Coeficientes da regressão para categoria econômica C da empresa LATAM

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	841.4954	15.6521	53.763	< 2e-16	***
TrechoGRU-FOR	-18.437	5.4686	-3.371	0.000772	***
Horário02-00	0.5977	8.5704	0.07	0.94441	
Horário04-00	8.1882	8.2208	0.996	0.319443	
Horário06-00	5.5513	7.9521	0.698	0.485255	
Horário08-00	10.6338	8.2281	1.292	0.196484	
Horário10-00	0.9767	8.6258	0.113	0.909865	
Horário12-00	3.2275	8.9684	0.36	0.71901	
Horário14-00	5.2931	9.2076	0.575	0.5655	
Horário16-00	0.9239	9.2152	0.1	0.920158	
Horário18-00	0.7668	9.3147	0.082	0.934409	
Horário20-00	5.8122	9.2582	0.628	0.530268	
Horário22-00	12.0561	9.5791	1.259	0.208432	
Antecedencia	-2.6902	0.2381	-11.297	< 2e-16	***
Mês_Viagemjulho	65.7432	5.0912	12.913	< 2e-16	***
Mês_Viagemnovembro	213.2928	18.1046	11.781	< 2e-16	***
Mês_Viagemoutubro	150.9953	13.9071	10.857	< 2e-16	***
Mês_Viagemsetembro	37.7362	8.1199	4.647	3.75E-06	***
Nome_Diassegunda-feira	-15.9719	6.2813	-2.543	0.011127	*
Nome_Diaterça-feira	3.5807	6.7799	0.528	0.597504	
Nome_Diaquarta-feira	-11.8701	7.1409	-1.662	0.096731	.
Nome_Diaquinta-feira	-9.7488	8.9416	-1.09	0.27582	
Nome_Diasexta-feira	-20.7486	9.275	-2.237	0.025473	*
Nome_Diasábado	-57.0276	6.7447	-8.455	< 2e-16	***
AtributoFeriado	73.2881	21.2758	3.445	0.000592	***
AtributoNormal	54.0028	14.9856	3.604	0.000327	***
AtributoPós-Comemoração	-11.1746	30.5831	-0.365	0.71489	
AtributoPós-Feriado	107.0767	18.3518	5.835	6.99E-09	***
AtributoVéspera de Comemoração	95.8306	37.3688	2.564	0.010459	*
AtributoVéspera de Feriado	76.5515	20.9137	3.66	0.000263	***

Residual standard error: 58.52 on 1160 degrees of freedom

Multiple R-squared: 0.3218, Adjusted R-squared: 0.3048

F-statistic: 18.98 on 29 and 1160 DF, p-value: < 2.2e-16

Tabela 5.35: Coeficientes da regressão para categoria econômica D da empresa LATAM

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	1113.3021	15.1675	73.401	< 2e-16	***
TrechoGRU-FOR	-28.6654	7.4403	-3.853	0.000141	***
Horário02-00	-3.5271	11.6175	-0.304	0.761621	
Horário04-00	3.7226	11.6961	0.318	0.750476	
Horário06-00	9.4097	11.2764	0.834	0.404633	
Horário08-00	6.5501	11.4629	0.571	0.56811	
Horário10-00	-0.1453	12.6839	-0.011	0.990866	
Horário12-00	-3.2831	12.1622	-0.27	0.787377	
Horário14-00	-19.6997	12.0245	-1.638	0.102322	
Horário16-00	-22.1692	11.8388	-1.873	0.062019	.
Horário18-00	-42.7089	13.6703	-3.124	0.001942	**
Horário20-00	-6.844	12.7717	-0.536	0.592413	
Horário22-00	-10.4324	12.7019	-0.821	0.412059	
Antecedencia	-2.113	0.6563	-3.22	0.001412	**
Mês_Viagemjulho	46.6614	12.9812	3.595	0.000375	***
Mês_Viagemnovembro	140.1653	56.5331	2.479	0.013666	*
Mês_Viagemsetembro	-33.6109	39.1065	-0.859	0.390711	
Nome_Diassegunda-feira	-30.0665	8.5761	-3.506	0.000519	***
Nome_Diaterça-feira	2.792	10.0617	0.277	0.78158	
Nome_Diaquarta-feira	-20.5309	11.1119	-1.848	0.065556	.
Nome_Diaquinta-feira	-5.8494	10.5486	-0.555	0.579601	
Nome_Diasexta-feira	20.134	9.9195	2.03	0.043192	*
Nome_Diasábado	15.1017	8.7023	1.735	0.083619	.
AtributoPós-Feriado	-30.6518	37.6275	-0.815	0.415888	

Residual standard error: 46.1 on 327 degrees of freedom

Multiple R-squared: 0.4365, Adjusted R-squared: 0.3968

F-statistic: 11.01 on 23 and 327 DF, p-value: < 2.2e-16

Tabela 5.36: Coeficientes da regressão para categoria econômica E da empresa LATAM

Coeficientes	Estimativa	Erro Padrão	t value	Pr(> t)	
(Intercept)	2289.3622	254.02863	9.012	< 2e-16	***
TrechoGRU-FOR	-32.52986	47.69658	-0.682	0.49593	
Horário02-00	-6.85821	74.1119	-0.093	0.92635	
Horário04-00	33.40207	74.66605	0.447	0.65505	
Horário06-00	-0.03779	71.9742	-0.001	0.99958	
Horário08-00	-37.36263	71.48399	-0.523	0.60172	
Horário10-00	-5.26459	78.87254	-0.067	0.94684	
Horário12-00	42.98816	78.7073	0.546	0.58549	
Horário14-00	30.63785	77.89124	0.393	0.69444	
Horário16-00	-64.03998	81.97641	-0.781	0.43551	
Horário18-00	268.90765	78.25144	3.436	0.0007	***
Horário20-00	-2.85074	88.42288	-0.032	0.97431	
Horário22-00	12.30046	83.63091	0.147	0.8832	
Antecedencia	-21.0521	5.74446	-3.665	0.00031	***
Mês_Viagemjulho	-15.7054	109.40112	-0.144	0.88598	
Mês_Viagemsetembro	325.18275	174.86168	1.86	0.06425	.
Nome_Diasegunda-feira	-168.9335	54.03912	-3.126	0.00201	**
Nome_Diaterça-feira	-183.2294	97.12952	-1.886	0.06053	.
Nome_Diaquarta-feira	-197.3496	92.04534	-2.144	0.03311	*
Nome_Diasexta-feira	-100.0177	53.33343	-1.875	0.06205	.
Nome_Diasábado	-78.46284	38.42372	-2.042	0.04232	*
AtributoNormal	-615.9833	266.31808	-2.313	0.02163	*

Residual standard error: 241.8 on 224 degrees of freedom

Multiple R-squared: 0.2217, Adjusted R-squared: 0.1488

F-statistic: 3.039 on 21 and 224 DF, p-value: 2.173e-05

Tabela 5.37: Coeficientes da regressão para categoria econômica F da empresa LATAM

Com as regressões estimadas, tornou-se possível realizar as previsões de preços. Para tal, primeiro se faz a classificação com as *Random Forest* calculadas e, a partir das classificações, calcula-se a estimativa de preço usando a regressão que convier ao caso.

5.3.4 Resultados das previsões

Foi realizada uma previsão com base nos dados já contidos na base de dados estudada e comparada a fim de se ter uma noção do nível de erro que se obtém a partir do modelo de análise proposto. Uma parte da tabela obtida é exibida na Tabela 5.38. A Tabela completa pode ser encontrada no Anexo D, no CD deste trabalho.

Horário	Antecedência	Mês_Viagem	Nome_Dia	Atributo	Empresa	Busca	Trecho	Categoria.Original	Categoria.Estimada	Preco	Preco.Estimado	Diferença	Erro.Absoluto	Erro.Relativo
12-00	39	setembro	terça-feira	Normal	Avianca	27/07/2018	GRU-FOR	B	B	532	480.54	51.46	51.46	0.10
00-00	37	setembro	quinta-feira	Normal	Gol	07/08/2018	GRU-FOR	B	B	477	499.11	-22.11	22.11	0.05
00-00	70	outubro	terça-feira	Normal	Avianca	14/08/2018	FOR-GRU	B	B	514	508.77	5.23	5.23	0.01
18-00	26	agosto	sexta-feira	Normal	Azul	22/07/2018	FOR-GRU	C	C	622	680.88	-58.88	58.88	0.09
14-00	66	setembro	quinta-feira	Normal	Azul	16/07/2018	GRU-FOR	B	B	428	477.60	-49.60	49.60	0.12
06-00	82	novembro	segunda-feira	Normal	Avianca	29/08/2018	FOR-GRU	D	B	821	501.29	319.71	319.71	0.39
12-00	23	agosto	terça-feira	Normal	Gol	05/08/2018	GRU-FOR	C	C	609	707.67	-98.67	98.67	0.16
14-00	30	agosto	sábado	Normal	Gol	05/07/2018	FOR-GRU	C	B	695	510.14	184.86	184.86	0.27
14-00	5	julho	terça-feira	Normal	Gol	05/07/2018	FOR-GRU	D	E	922	1 150.32	-228.32	228.32	0.25
14-00	44	outubro	sexta-feira	Feriado	Avianca	29/08/2018	GRU-FOR	B	B	547	503.10	43.90	43.90	0.08
16-00	73	novembro	quarta-feira	Normal	Azul	26/08/2018	GRU-FOR	A	B	397	479.40	-82.40	82.40	0.21
10-00	27	agosto	segunda-feira	Pós-Comemoração	Azul	17/07/2018	GRU-FOR	C	C	624	668.92	-44.92	44.92	0.07
18-00	50	outubro	sábado	Pós-Feriado	Gol	24/08/2018	GRU-FOR	B	B	515	514.79	0.21	0.21	0.00
04-00	74	outubro	terça-feira	Normal	Azul	20/07/2018	GRU-FOR	C	B	603	530.33	72.67	72.67	0.12
04-00	62	setembro	domingo	Normal	Azul	23/07/2018	FOR-GRU	C	C	602	642.73	-40.73	40.73	0.07
10-00	27	agosto	terça-feira	Normal	Azul	01/08/2018	GRU-FOR	B	B	477	488.12	-11.12	11.12	0.02
22-00	48	setembro	sexta-feira	Normal	Avianca	11/08/2018	GRU-FOR	B	B	470	474.75	-4.75	4.75	0.01
10-00	28	setembro	terça-feira	Normal	Avianca	07/08/2018	FOR-GRU	C	B	601	498.40	102.60	102.60	0.17
12-00	55	outubro	terça-feira	Normal	Gol	22/08/2018	GRU-FOR	B	B	521	489.41	31.59	31.59	0.06
04-00	80	setembro	quarta-feira	Normal	Azul	08/07/2018	GRU-FOR	B	B	576	529.88	46.12	46.12	0.08

Tabela 5.38: Exemplo de previsões realizadas

Calculando-se a média do erro relativo para toda a base de dados obtemos

$$Erro_{relativo\,médio} = 11.54\%$$

que pode ser considerado uma margem de erro aceitável. O erro absoluto médio foi estimado em

$$Erro_{absoluto\,médio} = R\$ 69.00$$

5.3.5 Simulação

Foi realizada uma simulação a fim de se verificar o quanto se economizaria com base nesse modelo de previsão. A simulação foi realizada escolhendo-se, aleatoriamente, “passageiros” que comprariam suas passagens com uma antecedência pré-definida, podendo ser com 1, 7, 14, 30, 60 ou 90 dias de antecedência, numa tentativa de simular passageiros que comprem passagens com antecedência longa (viagens de férias, por exemplo), passageiros que comprem passagens com antecedência média e os que comprem com pouca antecedência (viagens de negócios, por exemplo), a Tabela 5.39 mostra a quantidade de vezes em que cada empresa aérea foi sugerida na simulação para cada tipo de perfil. A simulação foi realizada com 201 amostras da base.

Ant.Base.Simulação	Azul	Avianca	Gol	LATAM
1	2	20	0	10
7	0	29	1	5
14	2	24	2	6
30	0	28	0	7
60	0	6	0	27
90	0	0	0	32

Tabela 5.39: Quantidade de vezes que cada empresa foi sugerida

Nessa simulação, o horário de compra permaneceu fixo para a busca de passagens mais baratas na simulação e essa busca foi realizada verificando-se as empresas aéreas dentro de uma antecedência menor que a preestabelecida. As comparações realizadas nesse exercício foram feitas com base no valor do preço estimado no modelo e do preço original de compra que o passageiro estaria comprando. A Tabela 5.40 mostra parte da base de simulação.

Horário	Antecedencia	Mês_Viagem	Nome_Dia	Atributo	Empresa	Busca	Trecho	Preco.Estimado	Categoria.Estimada	Preco.Cliente	Diferenca	Econ.Relativa
12-00	1	julho	quinta-feira	Normal	LATAM	04/07/2018	GRU-FOR	667.89	C	744	-76.11	-0.10
12-00	8	julho	quinta-feira	Normal	Gol	10/07/2018	GRU-FOR	520.64	B	630	-109.36	-0.17
20-00	57	agosto	quinta-feira	Normal	LATAM	07/07/2018	GRU-FOR	439.78	B	419	20.78	0.05
12-00	6	julho	terça-feira	Normal	LATAM	05/07/2018	GRU-FOR	664.12	C	711	-46.88	-0.07
20-00	0	julho	quinta-feira	Normal	Avianca	06/07/2018	GRU-FOR	667.06	C	808	-140.94	-0.17
10-00	8	julho	sexta-feira	Normal	Gol	11/07/2018	GRU-FOR	520.98	B	630	-109.02	-0.17
14-00	28	agosto	quinta-feira	Normal	Avianca	07/07/2018	GRU-FOR	467.50	B	419	48.50	0.12
22-00	56	agosto	quinta-feira	Normal	LATAM	09/07/2018	FOR-GRU	444.75	B	384	60.75	0.16
10-00	7	julho	quinta-feira	Normal	Gol	05/07/2018	GRU-FOR	522.27	B	689	-166.73	-0.24
16-00	90	outubro	quarta-feira	Normal	LATAM	05/07/2018	FOR-GRU	361.43	A	365	-3.57	-0.01
04-00	0	julho	sexta-feira	Normal	LATAM	07/07/2018	GRU-FOR	669.90	C	744	-74.10	-0.10
00-00	14	julho	sexta-feira	Normal	LATAM	06/07/2018	GRU-FOR	519.18	B	622	-102.82	-0.17

Tabela 5.40: Amostra de base de simulação

Os resultados obtidos por meio da relação entre os preços estimados e os preços pagos pelos passageiros é de que, aproximadamente, 115 passageiros economizariam, ou seja, 57.2% da base teriam economia se seguissem o modelo. A economia que eles obteriam seria de, em média, R\$ 63.53. Quando se observa a toda a base (tanto os que economizariam, quanto os que não economizariam), tem-se que a economia relativa seria de, aproximadamente, 2%.

Portanto, pode-se dizer que, é possível se ter economia fazendo-se simulações de preços. Além disso, essa economia pode ser maximizada com o aumento da eficiência dos algoritmos utilizados fazendo com que as previsões sejam mais certas e, desta forma, permita ao consumidor que se tenha mais segurança na simulação de preço de passagens.

Capítulo 6

Considerações finais

Os objetivos do estudo foram alcançados. Primeiramente, o programa de obtenção de dados funcionou ao longo dos dois meses de coleta e obteve uma base de dados suficientemente grande para que o estudo fosse realizado. Em posse da base de dados, foi possível se extrair gráficos e gerar modelos de predição de valores de passagens aéreas que permitiu que algumas conclusões fossem tomadas.

A partir das análises dos gráficos, é possível se dizer que as empresas tendem a seguir um padrão de precificação. As médias de preço de todas as empresas são muito próximas ao longo de todo período de antecedência analisado, para os dois trechos analisados, sendo talvez um reflexo da “mão invisível do mercado”.

A Azul é a empresa que apresenta padrões de precificação mais divergente entre as empresas analisadas. Ela possui preços, em média, mais elevado que qualquer outra empresa ao longo de quase todo o período de antecedência analisado. Além disso, ela também possui a maior variação de preços notada na análise realizada. Por outro lado, a Avianca é a empresa aérea que possui os preços mais constantes, em média, no estudo e, também, possui a menor variação de preços ao longo do tempo de antecedência. Portanto, isso infere que a Azul possui um *revenue management* bastante elaborado, sempre procurando adequar o valor do preço de suas passagens ao máximo que seu cliente consegue comprar, enquanto que a Avianca parece não aplicar medidas de *revenue management* com grande frequência na precificação de suas passagens.

Também é possível perceber que, os períodos entre 0 e 27 dias e 50 e 90 dias de antecedência costuma-se encontrar preços mais elevados quando comparados com o período entre 28 e 50 dias, além desses períodos possuírem as maiores variações de preços. É no

primeiro intervalo que é possível se encontrar os maiores preços constatados para cada empresa aérea.

Porém, vale salientar que, o estudo foi realizado com a coleta de dados em apenas dois meses (agosto e setembro), de tal forma que datas com as maiores antecedências pertenciam aos meses de outubro, novembro e dezembro. Esses dois últimos meses, principalmente, costumam ter preços mais elevados devido ao final de ano. Portanto, esse fator pode ter enviesado as análises finais obtidas com relação a antecedência de compra. Por isso, o estudo com dados de todos os meses é fundamental para um entendimento melhor das estratégias de preços das empresas aéreas.

Os modelos de *Random Forest* calculados nos permitem ter uma ideia da importância de cada variável no modelo, o que permite inferir que as variáveis mais importantes no modelo também são as variáveis mais importantes no *revenue management* de cada uma das empresas. É possível notar que a antecedência é a variável mais importante para a precificação da Azul, enquanto que nas outras, além da antecedência, também têm-se o mês da viagem como variável fundamental. Na modelagem da precificação da Avianca percebe-se que essa variável é até mais fundamental que a antecedência da compra.

A partir da simulação realizada, obteve-se uma economia geral de 2% comparando-se os valores estimados e os valores reais. Porém, quando se observa apenas os valores em a previsão obteve sucesso em economizar, tem-se uma taxa de 57.2% de economia. Portanto, é possível se obter economia através de modelos de predição de passagens.

Observando-se os perfis simulados, ou seja, clientes com os perfis diferentes de antecedência na compra pode-se observar que a Avianca tende a ser economicamente viável para consumidores que estejam em busca de passagens aéreas com até 30 dias de antecedência, enquanto que a LATAM tende a ser a empresa com melhores preços para antecedências melhores.

6.1 Sugestões para trabalhos futuros

Seria interessante se realizar essa análise com dados de mais meses do ano, isso certamente é uma limitação do atual trabalho. Outra sugestão seria a de mudar ou implementar outros modelos de *machine learning* para tentar melhorar a eficiência tanto da classificação, quanto da predição. Modelos que podem ser testados em algum trabalho

futuro seria a árvore de decisão para a classificação e redes neurais para a predição dos valores dos preços.

Referências Bibliográficas

- [1] OLIVEIRA, A. V. M. de; FERRAZ, R. B. Overbooking, gerenciamento de receitas e previsão de demanda: Estudo empírico das posições em sistemas de reservas de companhias aéreas. SciELO Brasil, 2008.
- [2] OLIVEIRA, A. Os impactos das estratégias de gerenciamento de receitas de ativos perecíveis na ligação rio de janeiro-são paulo. *Unpublished MSc dissertation*, 2000.
- [3] ETZIONI, O. et al. To buy or not to buy: mining airfare data to minimize ticket purchase price. In: ACM. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. [S.l.], 2003. p. 119–128.
- [4] FAGUNDES, J.; PONDE, J. Barreiras à entrada e defesa da concorrência: notas introdutórias. *Texto para discussão*, n. 1, 1998.
- [5] PAK, K.; PIERSMA, N. airline revenue management. 2002.
- [6] MCGILL, J. I.; RYZIN, G. J. V. Revenue management: Research overview and prospects. *Transportation science*, Informs, v. 33, n. 2, p. 233–256, 1999.
- [7] SALPUKAS, A. American to base fares on mileage. *The New York Times*, Print, p. D00001, 1983.
- [8] MAK, C. Y. *Revenue impacts of airline yield management*. Tese (Doutorado) — Massachusetts Institute of Technology, 1992.
- [9] PINDYCK R. E RUBINFELD, D. Microeconomia. 1994.
- [10] KHALIL, S.; FAKIR, M. Rcrawler: An r package for parallel web crawling and scraping. *SoftwareX*, Elsevier, v. 6, p. 98–106, 2017.

- [11] MUNZERT, S. et al. *Automated data collection with R: A practical guide to web scraping and text mining*. [S.l.]: John Wiley & Sons, 2014.
- [12] BOEING, G.; WADDELL, P. New insights into rental housing markets across the united states: web scraping and analyzing craigslist rental listings. *Journal of Planning Education and Research*, SAGE Publications Sage CA: Los Angeles, CA, v. 37, n. 4, p. 457–476, 2017.
- [13] LANDERS, R. N. et al. A primer on theory-driven web scraping: Automatic extraction of big data from the internet for use in psychological research. *Psychological methods*, American Psychological Association, v. 21, n. 4, p. 475, 2016.
- [14] TORGO, L. *Data mining with R: learning with case studies*. [S.l.]: CRC press, 2016.
- [15] PASSAGENS aéreas, hotéis e pacotes de viagens | Decolar.com. 2018. Disponível em: <www.decolar.com>.
- [16] EUROMONITOR International | Strategic Market Research, Data & Analysis. 2018. Disponível em: <<http://www.euromonitor.com>>.
- [17] O Algoritmo da Floresta Aleatória | Machina Sapiens | Medium. 2018. Disponível em: <<https://medium.com/machina-sapiens/o-algoritmo-da-floresta-aleatoria-3545f6babdf8>>.

FOLHA DE REGISTRO DO DOCUMENTO

1. CLASSIFICAÇÃO/TIPO TC	2. DATA 20 de novembro de 2018	3. REGISTRO Nº DCTA/ITA/TC-076/2018	4. Nº DE PÁGINAS 116
5. TÍTULO E SUBTÍTULO: Modelagem computacional e econômico financeiro da evolução temporal dos preços de passagens aéreas: um estudo sob a perspectiva das companhias aéreas nacionais.			
6. AUTOR(ES): Luis Victor Camilo das Chagas			
7. INSTITUIÇÃO(ÕES)/ÓRGÃO(S) INTERNO(S)/DIVISÃO(ÕES): Instituto Tecnológico de Aeronáutica – ITA			
8. PALAVRAS-CHAVE SUGERIDAS PELO AUTOR: Gerenciamento de recursos; Web-scraping; Precificação de passagens.			
9. PALAVRAS-CHAVE RESULTANTES DE INDEXAÇÃO: Administração de recursos; Fixação de preços; Aquisição de dados; ; Operações de linhas aéreas; Administração.			
10. APRESENTAÇÃO: X Nacional Internacional ITA, São José dos Campos. Curso de Graduação em Engenharia Civil-Aeronáutica. Orientador: Prof. Dr. Marcelo Xavier Guterres. Publicado em 2018.			
11. RESUMO: A compra de passagens para viagens aéreas vem se tornando cada vez mais fácil de ser realizada e de serem feitas comparações entre as empresas aéreas para se tentar obter o melhor preço de compra. Também fica evidente que os preços das passagens são variáveis e cada empresa tem sua própria política de precificação de passagens. Este trabalho tem como objetivo desempenhar um estudo para analisar e comparar o comportamento dos preços de passagens em alguns trechos do setor de aviação comercial brasileiro. A partir dessas análises, tentar notar padrões na política de precificação das empresas aéreas a fim de verificar o melhor período para se realizar compras de passagens. Além disso, será criado um modelo para previsão de compras de passagens para as diferentes rotas analisadas e será desenvolvido um estudo de economia em compras simuladas de passagens.			
12. GRAU DE SIGILO: (X) OSTENSIVO () RESERVADO () SECRETO			