



Instituto Brasileiro de Geografia e Estatística
Escola Nacional de Ciências Estatísticas
Bacharelado em Estatística



Bike Sharing: previsão da quantidade de bicicletas alugadas na cidade de Montreal, Canadá

GUILHERME GONÇALVES DE LIMA

Rio de Janeiro

2018

GUILHERME GONÇALVES DE LIMA

Bike Sharing: previsão da quantidade de bicicletas alugadas na cidade de Montreal, Canadá

Monografia apresentada à Escola Nacional de Ciências Estatísticas do Instituto Brasileiro de Geografia e Estatística, como requisito parcial à obtenção do título de Bacharel em Estatística.

Orientador(a): Professora Sandra Canton Cardoso

Coorientador(a): Professora Denise Britz do Nascimento Silva

Rio de Janeiro

2018

Autorizo a reprodução e divulgação total ou parcial desse trabalho, por parte da Escola Nacional de Ciências Estatísticas, através dos seus recursos eletrônicos, para fins de estudo e pesquisa, desde que citada a fonte.

L732b Lima, Guilherme Gonçalves de

Bike Sharing: previsão da quantidade de bicicletas alugadas na cidade de Montreal, Canadá / Guilherme Gonçalves de Lima. – Rio de Janeiro, 2018.

62 f.

Inclui referências, apêndices e anexos

Orientador: Prof. MSc. Sandra Canton Cardoso.

Coorientador: Prof. Dr. Denise Britz do Nascimento Silva.

Monografia (Graduação em Ciências estatísticas) – Escola Nacional de Ciências Estatísticas.

1. Estatística matemática. I. Cardoso, Sandra Canton. II. Silva, Denise Britz do Nascimento. III. Escola Nacional de Ciências Estatísticas. IV. IBGE. V. Título.

CDU: 519.2

GUILHERME GONÇALVES DE LIMA

BIKE SHARING: PREVISÃO DA QUANTIDADE DE BICICLETAS ALUGADAS NA
CIDADE DE MONTREAL, CANADÁ

Monografia apresentada como parte das atividades para a obtenção do título de Bacharel em Estatística no curso de Estatística da Escola Nacional de Ciências Estatísticas do Instituto Brasileiro de Geografia e Estatística.

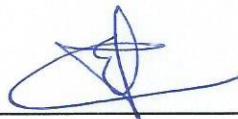
BANCA EXAMINADORA

Sandra Canton Cardoso

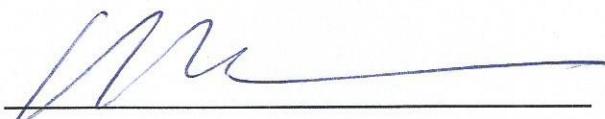
Profª. Sandra Canton Cardoso
Orientadora – ENCE/IBGE



Denise Britz
Profª. Denise Britz do Nascimento Silva
Coorientadora – ENCE/IBGE



Prof. Eduardo Lima Campos
ENCE/IBGE



Prof. Cristiano Roberto dos Santos
DPE/IBGE

Rio de Janeiro, 13 de Julho de 2018

AGRADECIMENTOS

Às professoras Sandra Canton e Denise Britz pela orientação e motivação durante a realização deste trabalho.

Agradeço à minha mãe, que sempre foi meu exemplo de força e fonte de inspiração. Ao meu pai e irmãos, por todo apoio e amor. Ton, sem o seu suporte não teria concluído o curso, muito obrigado. Sou grato também ao Dustin Bone pelo carinho, conversas e por sempre me fazer rir durante essa jornada.

Agradeço a todos os professores da Escola Nacional de Ciências Estatísticas, principalmente aos professores Alexandre Federici e Cristiano Santos e a professora Luisa La Croix que me aconselharam e me deram forças para concluir este curso.

RESUMO

Nos últimos anos, os sistemas de compartilhamento de bicicletas foram se tornando mais comuns nas grandes cidades. No ano de 2009 foi instalado na cidade de Montreal, no Canadá, o seu primeiro sistema de compartilhamento de bicicletas, BIXI Montreal. Estudos apontam que estes sistemas geram benefícios e são uma ótima opção de transporte público para as cidades. Este trabalho, visando contribuir com uma melhor gestão dos sistemas de compartilhamento de bicicletas, aplica técnicas de séries temporais com variáveis exógenas para previsão dos fluxos de chegada e saída de bicicletas nas estações da BIXI Montreal. A previsão de chegadas e saídas para o mês de novembro de 2017 é realizada com base nos dados sobre o uso do sistema e nas informações sobre clima e uso do espaço geográfico entre abril e outubro de 2017, considerando-se as estações com maior demanda e fluxo. Na seleção das estações para realizar as previsões foi elaborada uma análise de agrupamento na qual foi possível classificar as estações em dois tipos de perfis de uso diário. Para realizar a previsão dos fluxos de bicicletas que chegam e saem, o modelo utilizado foi o Auto-Regressivo Integrado de Médias Móveis com variáveis Exógenas (ARIMAX), o qual apresentou valores previstos próximos dos valores observados.

Palavras-chave: Sistema de Compartilhamento de Bicicletas. BIXI Montreal. Séries temporais. ARIMAX.

ABSTRACT

In recent years, bike sharing systems have become more popular in large cities. In 2009, BIXI Montreal was installed, which was Montreal's first bike sharing system. Studies show that these systems generate many benefits and are a great option for public transportation within cities. This work, aimed at contributing to better management of bike sharing systems, applies time series techniques with exogenous variables to predict the arrival and departure flows of bikes at BIXI Montreal stations. The forecast of arrivals and departures for November 2017 is made based on the data on the use of the system and on the information on climate and use of the geographic space between April and October 2017, considering stations with higher demand and flow. In the selection of the stations for forecasting, a clustering analysis was performed where it was possible to classify the stations into two types of daily use profiles. The Autoregressive Integrated Moving Average with Exogenous variables model (ARIMAX) was used to predict the flow of incoming and outgoing bicycles, on which it presented predicted values close to observed values.

Key-words: Bicycle-sharing systems. BIXI Montreal. Time series. ARIMAX.

LISTA DE ILUSTRAÇÕES

Figura 1 – Estações de bicicletas da BIXI Montreal	28
Figura 2 – Gráfico da Silhueta para o método KMeans	36
Figura 3 – Perfil do fluxo diário nas estações: (a) Chegada (b) Saída	36
Figura 4 – Distribuição geográfica das estações de cada perfil segundo o fluxo: (a) Chegada (b) Saída	37
Figura 5 – Matriz de correlação entre o fluxos de chegada e saída e as variáveis sobre o clima por dia	38
Figura 6 – Média diária de bicicletas em todo o sistema por temperatura máxima segundo o fluxo: (a) Chegada (b) Saída	39
Figura 7 – Média diária de bicicletas em todo o sistema por precipitação total segundo o fluxo: (a) Chegada (b) Saída	40
Figura 8 – Média diária de bicicletas em todo o sistema por umidade relativa segundo o fluxo: (a) Chegada (b) Saída	40
Figura 9 – Séries das variáveis exógenas temperatura máxima, precipitação total e umidade relativa	41
Figura 10 – Fluxo de chegada de bicicletas por dia, funções de autocorrelação e correlação parcial da série da estação 6015	42
Figura 11 – Fluxo de chegada de bicicletas por dia, funções de autocorrelação e correlação parcial da série diferenciada da estação 6015	42
Figura 12 – Análise dos resíduos do modelo ARIMAX(7, 1, 1, 3) incompleto para o Perfil 1 do Fluxo de Chegada	43
Figura 13 – Teste de Ljung-Box dos resíduos do modelo ARIMAX(7, 1, 1, 3) incompleto para o Perfil 1 do Fluxo de Chegada	44
Figura 14 – Análise dos resíduos do modelo ARIMAX(7, 1, 1, 3) incompleto para o Perfil 1 do Fluxo de Chegada com transformação logarítmica	45
Figura 15 – Teste de Ljung-Box dos resíduos do modelo ARIMAX(7, 1, 1, 3) incompleto para o Perfil 1 do Fluxo de Chegada com transformação logarítmica	45
Figura 16 – Previsões do modelo ARIMAX(7, 1, 1, 3) incompleto para o Perfil 1 do Fluxo de Chegada com transformação logarítmica	46
Figura 17 – Fluxo de chegada de bicicletas por dia, funções de autocorrelação e correlação parcial da série da estação 6012	47
Figura 18 – Fluxo de chegada de bicicletas por dia, funções de autocorrelação e correlação parcial da série diferenciada da estação 6012	47

Figura 19 – Análise dos resíduos do modelo ARIMAX(9, 1, 1, 3) incompleto para o Perfil 2 do Fluxo de Chegada	48
Figura 20 – Teste de Ljung-Box dos resíduos do modelo ARIMAX(9, 1, 1, 3) incompleto para o Perfil 2 do Fluxo de Chegada	49
Figura 21 – Análise dos resíduos do modelo ARIMAX(11, 1, 1, 3) incompleto para o Perfil 2 do Fluxo de Chegada com transformação logarítmica	50
Figura 22 – Teste de Ljung-Box dos resíduos do modelo ARIMAX(11, 1, 1, 3) incompleto para o Perfil 2 do Fluxo de Chegada com transformação logarítmica	50
Figura 23 – Previsões do modelo ARIMAX(11, 1, 1, 3) incompleto para o Perfil 2 do Fluxo de Chegada com transformação logarítmica	51
Figura 24 – Fluxo de saída de bicicletas por dia, funções de autocorrelação e correlação parcial da série da estação 6184	52
Figura 25 – Fluxo de saída de bicicletas por dia, funções de autocorrelação e correlação parcial da série diferenciada da estação 6184	52
Figura 26 – Análise dos resíduos do modelo ARIMAX(7, 1, 1, 3) incompleto para o Perfil 1 do Fluxo de Saída	53
Figura 27 – Teste de Ljung-Box dos resíduos do modelo ARIMAX(7, 1, 1, 3) incompleto para o Perfil 1 do Fluxo de Saída	54
Figura 28 – Previsões do modelo ARIMAX(7, 1, 1, 3) incompleto para o Perfil 1 do Fluxo de Saída	54
Figura 29 – Fluxo de saída de bicicletas por dia, funções de autocorrelação e correlação parcial da série da estação 6100	55
Figura 30 – Fluxo de saída de bicicletas por dia, funções de autocorrelação e correlação parcial da série diferenciada da estação 6100	55
Figura 31 – Análise dos resíduos do modelo ARIMAX(8, 1, 3, 3) incompleto para o Perfil 2 do Fluxo de Saída	56
Figura 32 – Teste de Ljung-Box dos resíduos do modelo ARIMAX(8, 1, 3, 3) incompleto para o Perfil 2 do Fluxo de Saída	57
Figura 33 – Previsões do modelo ARIMAX(8, 1, 3, 3) incompleto para o Perfil 2 do Fluxo de Saída	57

LISTA DE TABELAS

Tabela 1 – Sistemas de compartilhamento de bicicletas em operação no Brasil	20
Tabela 2 – Principais diferenças das gerações dos sistemas de compartilhamento de bicicletas	22
Tabela 3 – Estrutura dos dados sobre o início e término das viagens	27
Tabela 4 – Estrutura dos dados sobre a geolocalização das estações	27
Tabela 5 – Estrutura dos dados sobre o clima e as características urbanas	29
Tabela 6 – Variáveis utilizadas para o agrupamento das estações	35
Tabela 7 – Estações com o maior volume de Chegadas e Partidas segundo o perfil	38
Tabela 8 – Parâmetros e estimativas do modelo ARIMAX(7, 1, 1, 3) incompleto, Equação 5.1, para o Perfil 1 do Fluxo de Chegada	43
Tabela 9 – Parâmetros e estimativas do modelo ARIMAX(7, 1, 1, 3) incompleto, Equação 5.2, para o Perfil 1 do Fluxo de Chegada com transformação logarítmica	44
Tabela 10 – Estatísticas comparativas entre os modelos ajustados para o Perfil 1 do Fluxo de Chegada	46
Tabela 11 – Parâmetros e estimativas do modelo ARIMAX(9, 1, 1, 3) incompleto, Equação 5.3, para o Perfil 2 do Fluxo de Chegada	48
Tabela 12 – Parâmetros e estimativas do modelo ARIMAX(11, 1, 1, 3) incompleto, Equação 5.4, para o Perfil 2 do Fluxo de Chegada com transformação logarítmica	49
Tabela 13 – Estatísticas comparativas entre os modelos ajustados para o Perfil 2 do Fluxo de Chegada	51
Tabela 14 – Parâmetros e estimativas do modelo ARIMAX(7, 1, 1, 3) incompleto, Equação 5.5, para o Perfil 1 do Fluxo de Saída	53
Tabela 15 – Parâmetros e estimativas do modelo ARIMAX(8, 1, 3, 3) incompleto, Equação 5.6, para o Perfil 2 do Fluxo de Saída	56

SUMÁRIO

1	INTRODUÇÃO	17
1.1	Motivação	17
1.2	Objetivo	18
2	CARACTERIZAÇÃO DO OBJETO DE ESTUDO	19
2.1	Sistema de Compartilhamento de Bicicletas	19
2.2	História do Sistema de Compartilhamento de Bicicletas	19
2.3	Sistema de compartilhamento de bicicletas em Montreal	21
3	REVISÃO DA LITERATURA	23
4	MATERIAIS E MÉTODOS	27
4.1	Base de Dados	27
4.2	Metodologia Estatística	30
4.2.1	Séries Temporais	30
4.2.2	Modelos ARIMAX	30
4.2.3	A Função de Autocorrelação	31
4.2.4	A Função de Autocorrelação Parcial	31
4.2.5	O Teste de Ljung-Box	32
4.2.6	O Teste de Jarque-Bera	32
4.2.7	Critério de Informação de Akaike	33
4.2.8	Medidas de Erro de Previsão	33
4.2.9	Agrupamento KMeans	34
4.2.10	Método da Silhueta	34
5	ANÁLISE DE RESULTADOS	35
5.1	Seleção das Estações para Modelagem	35
5.2	Variáveis Exógenas	38
5.3	Fluxo de Chegada	42
5.4	Fluxo de Saída	52
6	CONCLUSÃO	59
	REFERÊNCIAS	61

1 INTRODUÇÃO

1.1 MOTIVAÇÃO

Os programas de compartilhamento de bicicletas são sistemas de serviço de transporte público que foram projetados especificamente para aumentar a oferta deste tipo de serviço, com baixo custo e facilidade de uso, com grande concentração de bicicletas nas áreas de atuação do programa e disponibilidade de 24 horas, além da velocidade de implementação da infraestrutura quando comparado com outros tipos de serviços de transporte público (BURDEN et al., 2009).

Segundo Zhang et al. (2016), nos sistemas de compartilhamento de bicicletas, o usuário pode pegar uma bicicleta numa determinada estação e retornar esta bicicleta em outra estação na cidade, podendo torna-se um complemento para deslocamento de curta distância (por exemplo, entre ônibus e trem). O sistema contém diversas estações espalhadas pela cidade, onde cada bicicleta pode ser usada várias vezes por dia. Segundo este mesmo autor, tais sistemas têm se tornado cada vez mais populares e já estão implementados em grandes cidades como Chicago (*Divvy Bike*), Nova York (*Citi Bike*), São Francisco (*Bay Area Bike Share*), Washington, D.C. (*Capital Bikeshare*).

Diversas melhorias podem ser observadas nas cidades que adotaram os sistemas de compartilhamento de bicicletas. Tais como, (a) o aumento de opção de transporte para população (moradores e turistas), (b) a criação de novos postos de trabalho, (c) a melhoria na saúde (possibilidade de redução das taxas de obesidade), (d) a possibilidade de diminuir o uso de carros (reduzindo as emissões de gases do efeito estufa e congestionamentos), (e) a maior consciência ambiental, (f) a integração com outros sistemas de serviço de transporte público (por exemplo, ônibus, metro, trem, etc.), (g) o baixo custo de implementação e operação comparados com outros modais de transporte (BURDEN et al., 2009; DEMAIO, 2009; SHAHEEN; GUZMAN; ZHANG, 2010).

Diante destes benefícios, o sistema de compartilhamento de bicicletas constitui uma ótima opção de transporte público. O foco deste trabalho é o desenvolvimento de um modelo de predição que seja capaz de estimar o fluxo de chegada e saída de bicicletas nas estações com base em dados sobre o uso do sistema, variáveis climáticas e de uso do espaço geográfico. Como produto, o modelo permitirá calcular a quantidade de bicicletas que chegam e que saem das estações, auxiliando os gestores dos sistemas no reposicionamento e dimensionamento da quantidade de bicicletas.

1.2 OBJETIVO

Para resolver o problema do reposicionamento e dimensionamento da quantidade de bicicletas no sistema de compartilhamento, o principal objetivo deste trabalho é realizar a previsão dos fluxos de chegada e saída de bicicletas nas estações, através de modelos de séries temporais, considerando os dados do ano de 2017 sobre o uso do sistema de compartilhamento de bicicletas e as variáveis sobre o clima e uso do espaço geográfico na cidade de Montreal, Quebec, Canadá.

O trabalho segue apresentado na seguinte ordem. O Capítulo 2 contém a caracterização do sistema de compartilhamento de bicicletas. Uma breve revisão da literatura é o assunto do Capítulo 3, no qual são descritos os trabalhos sobre a previsão da demanda de bicicletas nos sistemas. No Capítulo 4, são apresentados os dados e a metodologia estatística de análise de séries temporais usada para modelagem e previsão. O Capítulo 5 descreve as análises e resultados da metodologia aplicada. Finalmente, o Capítulo 6 relata as conclusões do trabalho.

2 CARACTERIZAÇÃO DO OBJETO DE ESTUDO

2.1 SISTEMA DE COMPARTILHAMENTO DE BICICLETAS

O compartilhamento de bicicleta é um conceito simples sob a qual as pessoas podem usar bicicletas sem a necessidade de serem proprietárias da bicicleta. Os sistemas fornecem diversas estações nas áreas urbanas das cidades onde o usuário pode retirar e devolver a bicicleta, para uso de baixo custo e em determinado período de tempo (SHAHEEN; GUZMAN; ZHANG, 2010).

Atualmente os sistemas de compartilhamento de bicicletas estão presentes em 63 países diferentes, somando aproximadamente 1.175 cidades (LOZANO et al., 2018). Os sistemas de compartilhamento de bicicletas foram se tornando mais comuns no mundo nos últimos anos, e estudos sobre os fatores que influenciam o seu uso foram surgindo, com finalidade de melhor compreender e maximizar a eficácia destes programas para que sejam bem sucedidos (FAGHIH-IMANI et al., 2014).

A Tabela 1 apresenta os sistemas de compartilhamento de bicicletas em operação no Brasil. Ao todo são 20 sistemas com um total de 10.085 bicicletas e 1.016 estações. São Paulo e Rio de Janeiro são as duas cidades com o maior número de bicicletas e estações em operação atualmente.

2.2 HISTÓRIA DO SISTEMA DE COMPARTILHAMENTO DE BICICLETTAS

Os sistemas de compartilhamento de bicicletas surgiram na década de 60, na Europa, e a sua história pode ser dividida em quatro gerações. A primeira geração começou na cidade de Amsterdam, Holanda, como solução para o transporte no centro da cidade. O sistema colocava as bicicletas no centro da cidade para uso sem limite de tempo, gratuito, e sem identificação do usuário. Os usuários poderiam pegar a bicicleta e usar o quanto necessitassem e depois deixar disponível para outro usuário. Porém, devido aos problemas de roubos das bicicletas e danos causados às mesmas, o sistema não teve sucesso (DEMAIO, 2009; SHAHEEN; GUZMAN; ZHANG, 2010).

Tabela 1 – Sistemas de compartilhamento de bicicletas em operação no Brasil

Nome	Cidade	Quantidade de Bicicletas	Quantidade de Estações
Bike Sampa	São Paulo, SP	2.600	260
Bike Rio	Rio de Janeiro, RJ	2.600	260
Bike PE	Recife, PE	800	80
Biciletar	Fortaleza, CE	800	80
Bike Salvador	Salvador, BA	500	50
+Bike	Brasília, DF	450	50
Bike POA	Porto Alegre, RS	410	41
Bike BH	Belo Horizonte, MG	400	40
Bike Santos	Santos, SP	370	37
CycloSampa	São Paulo, SP	200	17
Bike VV	Vila Velha, ES	200	20
Debike	Goiânia, GO	150	15
Manôbike	Manaus, AM	110	11
Vai de Bici	Passo Fundo, RS	100	10
Petro Bike	Petrolina, PE	100	10
Integra Bike	Sorocaba, SP	80	10
Bike Belém	Belém, PA	80	9
Caju Bike	Aracaju, SE	50	5
Rivi Bike	Bertioga, SP	45	7
Udi Bike	Uberlândia, MG	40	4
Total		10.085	1.016

Fonte – Adaptado de DeMaio e Meddin (2018).

No começo da década de 90, ainda na Europa, a segunda geração surgiu como resultado da primeira, com a incorporação de estações em lugares determinados na cidade onde as bicicletas ficavam presas. O usuário poderia retirar a bicicleta mediante um depósito de moeda e tinha que retornar a bicicleta numa das estações. O uso de estações foi um conceito introduzido ao sistema para evitar os roubos das bicicletas. Os sistemas ainda não identificavam os usuários e por isso ainda enfrentavam os mesmos problemas de roubo das bicicletas e do limite de tempo de uso que também ainda não era controlado, fazendo com que os usuários permanecessem muito tempo com as bicicletas e diminuindo a oferta para as pessoas que desejavam usar o sistema (SHAHEEN; GUZMAN; ZHANG, 2010).

Diante desses problemas, na terceira geração foi adicionada tecnologia baseada em sistemas de informação. Com isso, os programas passaram a ter um melhor rastreamento das retiradas e devoluções das bicicletas. Os usuários passaram a ser identificados e o uso cobrado por intervalos de tempo definidos. Nesta geração, a tecnologia trouxe mais segurança para o sistema e melhor fluxo de informação em tempo real. Diminuindo os

problemas com roubos e provendo dados aos usuários e operadores, ajudando na otimização e melhoria nos sistemas. A partir dessa geração, os sistemas de compartilhamento de bicicletas se tornaram populares no mundo (SHAHEEN; GUZMAN; ZHANG, 2010).

Com o sucesso da terceira geração, surgiu a quarta geração com a proposta de ser um sistema multimodal adaptável à demanda. Nesta geração, as estações tiveram melhorias como funcionamento por painel solar e melhores mecanismos para prender as bicicletas. Passou também a incluir bicicletas elétricas, sistema de distribuição e redistribuição de bicicletas nas estações durante o funcionamento do sistema, tecnologia de rastreamento através de *Global Positioning System* (GPS), quiosques com painéis *touch-screen* e uso de *smart-cards* para conexões com outros modais de transporte (DEMAIO, 2009; SHAHEEN; GUZMAN; ZHANG, 2010).

Na Tabela 2, podemos ver as principais diferenças entre as gerações. O uso da tecnologia é uma das principais diferenças entre a primeira e segunda geração para a terceira e quarta geração. O papel da tecnologia é muito importante para fornecer dados em tempo real, auxiliando no desafio de como dimensionar a quantidade de bicicletas em cada estação para atender à demanda do sistema.

2.3 SISTEMA DE COMPARTILHAMENTO DE BICICLETAS EM MONTREAL

BIXI (uma contração das palavras inglesas '*bicycle*' e '*taxi*') é o nome dado ao primeiro sistema de compartilhamento de bicicletas da cidade de Montreal, que foi implementado em maio de 2009, sendo um dos primeiros sistemas da quarta geração de compartilhamento de bicicletas (SHAHEEN; GUZMAN; ZHANG, 2010). Em 2014, a cidade de Montreal criou a **BIXI Montreal**, uma organização sem fins-lucrativos, para gerenciar o seu sistema de compartilhamento de bicicletas, que tem como missão contribuir com a saúde e bem-estar dos moradores de Montreal através de um meio de transporte ativo (BIXI MONTREAL, 2018b).

Segundo Faghih-Imani et al. (2014), **BIXI Montreal** já é um sistema maduro que possibilita o estudo e a compreensão dos fatores que influenciam no uso de sistemas de compartilhamento de bicicletas. No início, o sistema tinha 3.000 bicicletas e 300 estações. No ano de 2012, contava com 4.000 bicicletas e 400 estações. Hoje o sistema tem 6.200 bicicletas e 540 estações distribuídas em áreas de Montreal, Longueuil e Westmount (BIXI MONTREAL, 2018b).

Tabela 2 – Principais diferenças das gerações dos sistemas de compartilhamento de bicicletas

1 ^a Geração	2 ^a Geração	3 ^a Geração	4 ^a Geração
Componentes: • Bicicletas.	Componentes: • Bicicletas; • Estações.	Componentes: • Bicicletas; • Estações.	Componentes: • Bicicletas; • Bicicletas elétricas; • Estações.

Características: • Bicicletas livres para uso sem custo e limite de tempo.	Características: • Pagamento de taxas para uso; • Uso sem limite de tempo.	Características: • Identificação dos usuários e rastreamento; • Custo por tempo; • Tecnologia para retirar e devolver bicicleta.	Características: • Identificação dos usuários e rastreamento (GPS); • Custo por tempo; • Tecnologia para retirar e devolver bicicleta; • Integração com outros modais de transporte; • Sistemas de redistribuição de bicicletas.
--------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Fonte – Adaptado de DeMaio (2009), Shaleen, Guzman e Zhang (2010).

3 REVISÃO DA LITERATURA

Este capítulo apresenta a literatura sobre a previsão da demanda do sistema de compartilhamento de bicicletas a partir de fatores que influenciam no fluxo de chegada e saídas nas estações.

Faghih-Imani et al. (2014) propõe uma abordagem de modelo linear misto baseado no modelo de regressão linear para quantificar a influência de fatores climáticos (temperatura e umidade), temporais (dia da semana e parte do dia em que a viagem iniciou) e características urbanas (número estabelecimentos comerciais, universidades, densidade demográfica) no fluxo de chegada e saída nas estações do sistema de compartilhamento de bicicletas da cidade de Montreal, Canadá. Para cada estação, os autores observaram o fluxo de chegada e saída para cada hora e desenvolveram um modelo para cada um dos fluxos. Os autores mostram que as pessoas estão mais propensas a usar o sistema quando existe boa condição climática. Concluem também que o fluxo de bicicletas reduz em finais de semanas e em estações distantes do centro administrativo da cidade, e que as taxas de chegada e saída de cada estação são influenciadas expressivamente por estabelecimentos comerciais e universidades nas proximidades de cada estação.

Rudloff e Lackner (2014) utilizou os modelos de contagem Poisson, Binomial Negativa e Hurdle para analisar a demanda do sistema de compartilhamento de bicicletas da cidade de Viena, Áustria. Segundo os autores, estes modelos foram usados por representarem bem a estrutura de contagem para a demanda de cada estação. Neste caso, foram utilizadas variáveis sobre o clima (temperatura, velocidade do vento e chuva), bem como informações temporais (dia da semana, hora e período do ano) e de infraestrutura do sistema (disponibilidade de bicicletas nas estações mais próximas). O artigo demonstra que as variáveis climáticas têm impacto no uso do sistema, e a disponibilidade de bicicletas nas estações próximas influenciam na demanda de cada estação. Os autores também testaram modelos por agrupamento de estações segundo localização e verificaram melhores resultados quando modelaram os dados das diversas estações em separado. Quanto à performance dos modelos, os autores concluíram que o modelo com melhor performance foi o Hurdle, seguido (em ordem) pelos modelos Binomial Negativa e Poisson.

Parikh e Ukkusuri (2015) propõe modelar a demanda de bicicletas como um processo de Markov para cada estação do sistema de compartilhamento de bicicletas da cidade da Antuérpia, Bélgica. Para o modelo, os autores usam o retorno líquido, que é a diferença entre o número de bicicletas devolvidas e o número de bicicletas retiradas, em intervalo de tempo de 30 minutos. Segundo os autores, para previsão da demanda é importante que o número de bicicletas permaneça constante em todo o sistema. Para

isso, eles usam um modelo de Programação Inteira Mista para manter constante o número de bicicletas. Todas as estações foram consideradas independentes uma das outras. Os autores concluíram que a proposta tem bons resultados para previsão da demanda.

Faghih-Imani e Eluru (2015) sugere um modelo de regressão logística multinomial para examinar o comportamento da preferência de qual estação de destino os usuários vão entregar as bicicletas. Para isso, os autores usaram os dados do sistema de compartilhamento de bicicletas da cidade de Chicago, Estados Unidos e desenvolveram dois modelos, um para os usuários com assinatura anual e outro para usuários que compraram o passe diário de uso do sistema. Em cada modelo, foram considerados fatores associados à infraestrutura próxima da estação (número de estações vizinhas, capacidade das estações, tamanho das ruas e ciclovias), características urbanas (número estabelecimentos comerciais, universidades, densidade demográfica), temporal (em qual período do dia iniciou a viagem), e atributos sobre a viagem e do usuário (distância das estações próximas e gênero). A pesquisa demonstra que as pessoas escolhem as estações próximas de ciclovias mais extensas e com mais capacidade para receber bicicletas. Destaca ainda que, de manhã e de tarde, no período de deslocamento para o trabalho, estações em áreas com maior densidade populacional e estabelecimentos comerciais são mais propensas a serem escolhidas. Quanto ao tipo de usuário, aqueles que são membros anuais preferem locais com mais estações com pouca capacidade e que os usuários do plano diário preferem locais com poucas estações com muita capacidade.

Chen et al. (2017) utiliza um único modelo para estimar a quantidade de bicicletas alugadas e devolvidas para cada estação baseado em *Recurrent Neural Network* (RNN), para o sistema da cidade de Nova York, Estados Unidos. Segundo os autores, a estrutura da RNN é boa para séries temporais por conseguir identificar padrões temporais, e para lidar com muitas variáveis, tanto de entrada quanto de saída para o modelo. Assim, um único modelo pode prever a quantidade dos alugueis e retornos das bicicletas para todas as estações imediatamente. Como dados de entrada no modelo, são consideradas variáveis temporais (ano, mês, dia da semana e hora), climáticas por hora (clima, temperatura, umidade, precipitação, velocidade do vento e visibilidade), número de aluguéis e devoluções no intervalo de 5 minutos por tipo de usuário (usuário membro anual ou diário). Para a saída do modelo, considera-se o número de aluguéis e devoluções por hora para cada tipo de usuário. Os autores usam o modelo com a estrutura de uma *bi-directional RNN* e comparam os resultados com os modelos: *Ordinary Least-Squares Regression*, *Random Forest* e *Feedforward Neural Network*. O trabalho evidência que o modelo proposto apresenta melhores estimativas que os outros três modelos utilizados como base de comparação, tanto para cada estação individualmente, quanto para todo o sistema.

O presente trabalho está interessado em prever a quantidade de bicicletas que chegam e que saem das estações no sistema da BIXI Montreal, a revisão da literatura evidenciou que variáveis sobre o clima e características urbanas influenciam no uso dos sistemas de compartilhamento de bicicletas, por isso, serão utilizadas para o desenvolvimento da modelagem e previsão. Diversas opções de modelagem para a demanda de bicicletas alugadas foram apresentadas neste capítulo, e neste trabalho será usado modelos de séries temporais com variáveis exógenas, método não utilizado nos trabalhos citados.

4 MATERIAIS E MÉTODOS

Este capítulo apresenta uma breve descrição da base de dados, dos conceitos teóricos sobre séries temporais e da metodologia estatística utilizados neste trabalho.

4.1 BASE DE DADOS

Para elaboração deste trabalho são utilizadas quatro bases de dados diferentes para o ano de 2017: (a) uma contendo informações sobre o início e o término da viagem, (b) uma base identificando a geolocalização das estações, (c) uma com dados sobre o clima na região de estudo e (d) outra base com as características urbanas da área de localização das estações.

Os dados sobre o inicio e término das viagens e geolocalização das estações do sistema de compartilhamento de bicicletas na cidade de Montreal no Canadá estão disponíveis no site da BIXI Montreal (2018a). A Tabela 3 contém a lista das variáveis da base de dados sobre as viagens de 2017 totalizando 4.740.357 observações.

Tabela 3 – Estrutura dos dados sobre o início e término das viagens

Variável	Descrição
<i>start_date</i>	Data e hora do início da viagem
<i>start_station_code</i>	Código da estação de início da viagem
<i>end_date</i>	Data e hora do término da viagem
<i>end_station_code</i>	Código da estação de término da viagem
<i>duration_sec</i>	Duração da viagem em segundos
<i>is_member</i>	Indicação se o usuário é associado ou ocasional

Fonte – Adaptado de BIXI Montreal (2018a).

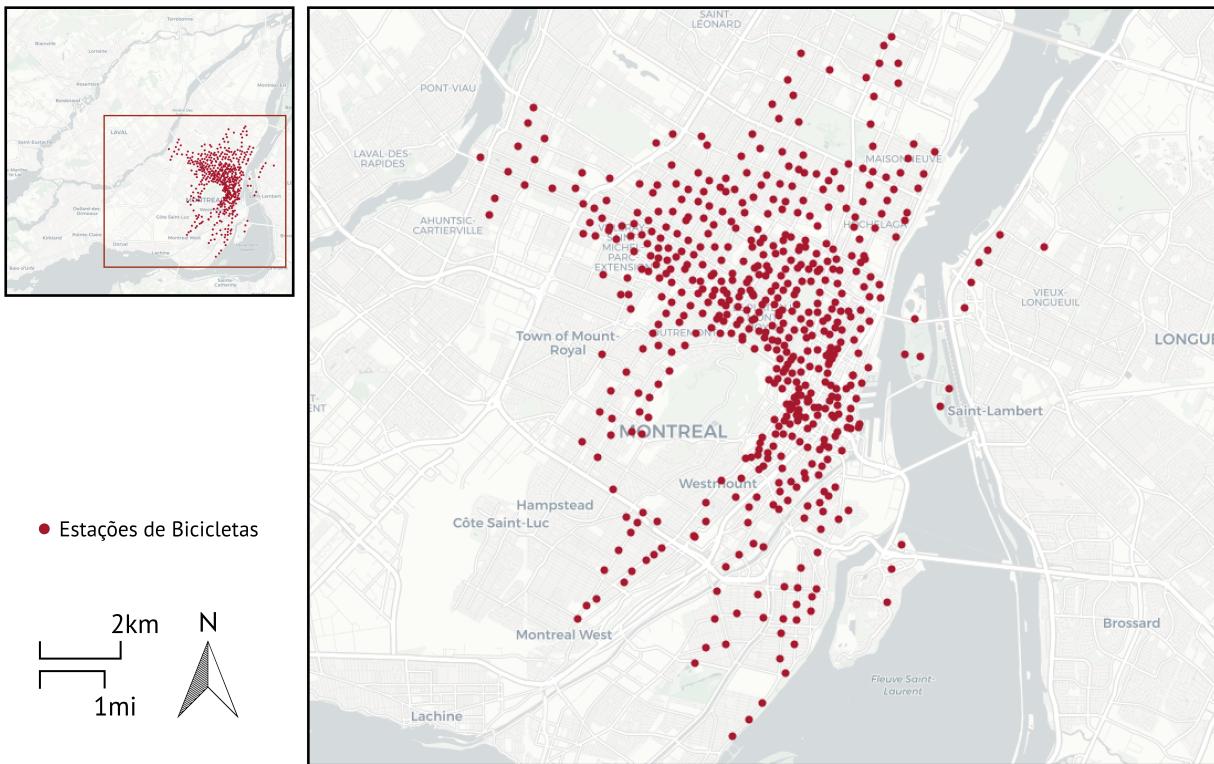
Na Tabela 4 encontra-se a lista das variáveis sobre a geolocalização das estações. No ano de 2017, o sistema totalizava 546 estações que podem ser observadas na Figura 1.

Tabela 4 – Estrutura dos dados sobre a geolocalização das estações

Variável	Descrição
<i>code</i>	Código da estação
<i>name</i>	Nome da estação
<i>latitude</i>	Latitude da estação
<i>longitude</i>	Longitude da estação

Fonte – Adaptado de BIXI Montreal (2018a).

Figura 1 – Estações de bicicletas da BIXI Montreal



Fonte – BIXI Montreal (2018a).

Adicionalmente foram coletados dados do governo do Canadá (2018) sobre o clima, e da *Application Programming Interface* (API) do Google (2018) as variáveis sobre as características urbanas, tais como: geolocalização dos estabelecimentos comerciais, restaurantes, transporte público, lugares de lazer e prédios do sistema educacional, órgãos públicos e locais religiosos (Tabela 5).

A partir dos dados coletados sobre o início e o término da viagem, para cada estação, foi obtido o total de bicicletas que chegam e saem por dia para realizar a previsão dos fluxos de chegada e saída de bicicletas através do modelo de séries temporais. As variáveis sobre o clima serão avaliadas e incluídas nos modelos como variáveis exógenas. Usando os dados sobre as características urbanas e o uso do sistema, será elaborada uma análise do perfil das estações através da método de agrupamento KMeans, para selecionar as estações com maior demanda para a modelagem.

Tabela 5 – Estrutura dos dados sobre o clima e as características urbanas

Base de Dados	Variável	Descrição
Dados sobre o clima	$date$ $Max\ Temp\ ({}^{\circ}\text{C})$ $Min\ Temp\ ({}^{\circ}\text{C})$ $Mean\ Temp\ ({}^{\circ}\text{C})$ $Total\ Rain\ (\text{mm})$ $Total\ Snow\ (\text{cm})$ $Total\ Precip\ (\text{mm})$ $Mean\ Dew\ Point\ Temp\ ({}^{\circ}\text{C})$ $Mean\ Rel\ Hum\ (\%)$	<p>Dia do ano</p> <p>Temperatura máxima do dia em graus Celsius</p> <p>Temperatura mínima do dia em graus Celsius</p> <p>Temperatura média do dia em graus Celsius</p> <p>Quantidade total de precipitação líquida (mm) do dia</p> <p>Quantidade total de precipitação congelada (cm) do dia</p> <p>Soma da precipitação líquida e equivalente da água da precipitação congelada (mm) do dia</p> <p>Temperatura média do ponto de orvalho do dia (temperatura na qual o ar teria que ser resfriado para que o vapor de água no ar comece a se condensar)</p> <p>Umidade relativa do ar média do dia</p>
Dados sobre o clima	$type_place$ $place_name$ $place_lat$ $place_lng$	<p>Tipo de local estabelecimentos comerciais, restaurantes, transporte público, lugares de lazer e prédios do sistema educacional, órgãos públicos e locais religiosos</p> <p>Nome do local</p> <p>Latitude do local</p> <p>Longitude do local</p>

Fonte – Adaptado de Canadá (2018), Google (2018).

4.2 METODOLOGIA ESTATÍSTICA

4.2.1 SÉRIES TEMPORAIS

Uma série temporal é definida como um conjunto de observações ordenadas no tempo, seja contínuo ou discreto. Os fluxos de chegada e saídas de bicicletas nas estações dos sistemas de compartilhamento são consideradas como séries temporais, pois podemos obter observações a partir de intervalo unitário de tempo, como horas ou dias.

Em geral, uma série temporal é uma realização de um processo estocástico, ou seja, uma série temporal é uma família de variáveis aleatórias definidas num mesmo espaço de probabilidade que está associado a um conjunto de números inteiros ou de números reais (MORETTIN; TOLOI, 2006).

Os modelos disponíveis na literatura decompõem a série temporal como uma soma de três componentes não-observáveis: (a) tendência, (b) sazonalidade e (c) componente aleatória. A tendência é a componente que indica a existência de um direcionamento crescente ou decrescente no conjunto de dados. Já a sazonalidade indica um padrão relacionado a um intervalo de tempo fixo conhecido (por exemplo, trimestre do ano ou dia da semana). Por fim, a componente aleatória corresponde às influências aleatórias características do processo estocástico e que não podem ser caracterizadas como tendência ou sazonalidade (MORETTIN; TOLOI, 2006).

4.2.2 MODELOS ARIMAX

A metodologia conhecida como abordagem de Box e Jenkins que consiste em ajustar modelos auto-regressivos integrados de médias móveis, ARIMA(p, d, q), é bastante utilizada para análise de séries temporais (MORETTIN; TOLOI, 2006).

Neste trabalho, será considerado o modelo ARIMAX que é uma extensão do modelo ARIMA com a inclusão de variáveis exógenas. O modelo é composto por uma combinação dos modelos Auto-Regressivo AR(p), Integrado (d), Média-Móvel MA(q) e Exógeno X(r). O modelo ARIMAX(p, d, q, r), onde p é o número de lags auto-regressivos, d é número de vezes que a série temporal é diferenciada, q é o número de lags das médias móveis, e r é o número de variáveis exógenas, é definido por:

$$\Delta^d y_t = \sum_{k=1}^p \phi_k \Delta^d y_{t-k} + \sum_{k=1}^q \theta_k \epsilon_{t-k} + \sum_{m=1}^r \beta_m X_{m,t} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2) \quad (4.1)$$

em que y_t é variável dependente no tempo t , y_{t-k} é a variável dependente defasada por k passos de tempo, ϕ_k é o coeficiente de y_{t-k} , ϵ_{t-k} é o erro no tempo t defasado de k , θ_k é o coeficiente do termo de ϵ_{t-k} , β_m representa o coeficiente das variáveis exógenas $X_{m,t}$ no tempo t , e ϵ_t é a componente de erro do modelo (BENNETT; STEWART; LU, 2014).

4.2.3 A FUNÇÃO DE AUTOCORRELAÇÃO

A função de autocorrelação (FAC) é uma medida da correlação entre observações de uma série temporal que são separadas por k unidades de tempo (BOX et al., 2015). A FAC é utilizada para detectar não-aleatoriedade nos dados e identificar um modelo de série temporal apropriado. A FAC do k -ésimo lag pode ser estimada por:

$$\tau_k = \frac{\sum_{t=1}^{N-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^N (y_t - \bar{y})^2} \quad (4.2)$$

onde N é o número de observações da série temporal, y_t é o t -ésimo valor observado da série temporal, \bar{y} é a média da série temporal observada. Para amostras finitas, a Equação 4.2 precisa de um termo de correção. Assim, a autocorrelação estimada será dada por:

$$\rho_k = \frac{N}{N - k} \tau_k \quad (4.3)$$

4.2.4 A FUNÇÃO DE AUTOCORRELAÇÃO PARCIAL

A função de autocorrelação parcial (FACP) é a medida da correlação parcial de uma série temporal com seus próprios valores defasados. A FACP pode auxiliar na definição das defasagens p apropriadas para um modelo AR(p) (BOX et al., 2015).

Seja ρ_k a autocorrelação no k -ésimo lag dado pela FAC. A FACP pode ser estimada como:

$$\begin{aligned} \phi_{11} &= \rho_1 \\ \phi_{22} &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \\ \phi_{33} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} \end{aligned}$$

e, em geral,

$$\phi_{kk} = \frac{|P_k^*|}{|P_k|} \quad (4.4)$$

onde P_k é a matrix de autocorrelações até o k -ésimo lag, P_k^* é a matrix P_k com a última coluna contendo o k -ésimo vetor de autocorrelações.

4.2.5 O TESTE DE LJUNG-BOX

O teste de Ljung-Box é utilizado para verificar a independência dos resíduos do modelo ajustado, ou seja, se não apresentam correlação em série. As hipóteses do teste são:

H_0 : os dados são distribuídos de forma independente.

H_a : os dados não são distribuídos de forma independente.

A estatística de teste, que sob a hipótese nula tem distribuição qui-quadrado com h graus de liberdade, é dada por:

$$Q = N(N+2) \sum_{k=1}^h \frac{\rho_k^2}{N-k} \quad (4.5)$$

onde N é o número de observações da série, ρ_k^2 é a autocorrelação no lag k , e h é o número de lags a ser testado. Assim, rejeita-se a hipótese nula se:

$$P(\chi^2 \geq Q | H_0) \leq \alpha \quad (4.6)$$

onde α é o nível de significância atribuído (LJUNG; BOX, 1978).

4.2.6 O TESTE DE JARQUE-BERA

O teste de Jarque-Bera é um teste de adequação para avaliar se um conjunto de dados tem uma distribuição Normal. As hipóteses do teste são:

H_0 : os dados são normalmente distribuídos.

H_a : os dados seguem alguma outra distribuição.

A estatística de teste é baseada nos coeficientes de assimetria e curtose dos dados. Sob a hipótese nula, tem distribuição assintótica qui-quadrado com 2 graus de liberdade, definida por:

$$JB = N \left(\frac{\mu_1^2}{6} + \frac{(\mu_2 - 3)^2}{24} \right) \quad (4.7)$$

onde N é o número de observações no conjunto de dados, μ_1 e μ_2 são, respectivamente, os coeficientes de assimetria e curtose. Assim, rejeita-se a hipótese nula se:

$$P(\chi^2 \geq JB|H_0) \leq \alpha \quad (4.8)$$

onde α é o nível de significância atribuído (JARQUE; BERA, 1980).

4.2.7 CRITÉRIO DE INFORMAÇÃO DE AKAIKE

O critério de informação de Akaike é uma estatística da qualidade relativa do modelo para um conjunto de dados, considerando a log-verossimilhança e o número de parâmetros estimados. O AIC, do inglês *Akaike Information Criterion*, penaliza a verossimilhança, considerando o número de parâmetros do modelo, para que seja selecionado um modelo mais parcimonioso. O AIC é definido por:

$$AIC = 2g - 2\mathcal{L}(Y; \theta) \quad (4.9)$$

onde Y é o vetor de variáveis aleatórias de uma distribuição de probabilidade, $\mathcal{L}(Y; \theta)$ é a log-verossimilhança, e g é o número de parâmetros estimado pelo modelo (AKAIKE, 1974).

4.2.8 MEDIDAS DE ERRO DE PREVISÃO

O Erro Quadrático Médio (EQM) é uma estatística que mede a média da distância entre os valores observados e previstos, definido por:

$$EQM = \frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2 \quad (4.10)$$

onde N é o tamanho da série, y_t e \hat{y}_t são, respectivamente, o t -ésimo valor observado e seu correspondente valor estimado (DEGROOT; SCHERVISH, 2012). Pode-se também calcular a Raiz do EQM que representa o desvio médio dos erros da previsão, definido por:

$$\text{Raiz do EQM} = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2} \quad (4.11)$$

Adicionalmente, o Erro Absoluto Médio (EAM) é uma estatística que mede a média da diferença absoluta entre os valores observados e previstos da forma (DEGROOT; SCHERVISH, 2012):

$$EAM = \frac{1}{N} \sum_{t=1}^N |y_t - \hat{y}_t| \quad (4.12)$$

4.2.9 AGRUPAMENTO KMEANS

O agrupamento KMeans é o método que particiona n observações em c grupos (MCQUEEN, 1967). O método é iterativo no qual, inicialmente, é selecionado um centro de forma aleatória para cada um dos c grupos e, em seguida, repetem-se os passos:

1. Para cada observação, calcula-se a distância Euclidiana para os c centros, sendo a observação classificada como pertencente ao grupo do centro mais próximo.
2. Cada centro é atualizado com a média das observações que compõem o grupo.

A convergência do algoritmo ocorre quando não há mais nenhuma mudança de grupo para cada observação.

4.2.10 MÉTODO DA SILHUETA

O método da Silhueta fornece uma visualização gráfica de quão bem cada observação está classificada dentro do seu grupo. Este método pode ser utilizado para validar métodos de agrupamentos, tais como o KMeans. O valor da silhueta é baseado na medida de quão semelhante a observação é ao seu próprio grupo (coesão) comparado a outros grupos (separação). Para cada observação i , o valor da silhueta é definido por:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \quad -1 \leq s(i) \leq 1 \quad (4.13)$$

onde $a(i)$ é a média da distância entre i e todas as demais observações do mesmo grupo de i , $b(i)$ é a menor distância média entre i e todos os pontos dos grupos aos quais i não pertence. Um valor alto de $s(i)$ indica que a observação i é compatível com o grupo ao qual pertence (ROUSSEEUW, 1987).

5 ANÁLISE DE RESULTADOS

São analisados os fluxos das bicicletas do sistema BIXI Montreal que chegam e que saem das estações por dia, considerando os dados do período de 15 de abril até 15 de novembro de 2017, totalizando 215 dias. É importante ressaltar que a BIXI Montreal não tem operação durante inverno.

Considera-se o período de 15 de abril até 16 de outubro de 2017 para estimar os modelos de predição, totalizando 185 dias. Os últimos 30 dias de operação do sistema são utilizados para projeção das séries e validação dos modelos. O sistema BIXI Montreal operou com 546 estações no ano de 2017 e para este trabalho foram selecionadas as estações com as maiores demanda para modelar os fluxos de chegadas e saídas.

5.1 SELEÇÃO DAS ESTAÇÕES PARA MODELAGEM

Para selecionar as estações, foi necessário conhecer os perfis da demanda do fluxo de chegada e saída de bicicletas nas estações durante o dia. A abordagem adotada neste trabalho para identificar os perfis foi o uso do método de agrupamento KMeans, descrito na subseção 4.2.9.

As variáveis utilizadas no agrupamento das estações foram construídas a partir das bases de dados sobre o início e término das viagens, geolocalização das estações e características urbanas, conforme apresentado no Capítulo 4. A Tabela 6 lista as variáveis elaboradas e as respectivas bases de dados para a sua construção.

Tabela 6 – Variáveis utilizadas para o agrupamento das estações

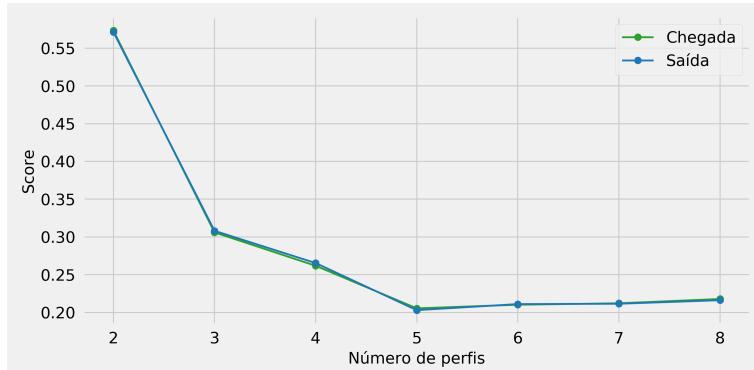
Base de Dados	Variáveis
Dados de início e término da viagem	Percentual da quantidade de chegada por hora durante o dia Percentual da quantidade de saída por hora durante o dia
Dados de geolocalização das estações e características urbanas	Número de restaurantes localizados a 200m da estação Número de estabelecimentos comerciais a 200 m da estação Número de transporte público a 200m da estação Número de lugares de lazer a 200m da estação Número de prédios do sistema educacional a 200m da estação Número de órgãos públicos a 200m da estação Número de locais religiosos a 200m da estação

Fonte – Produzido pelo autor.

Na análise de agrupamento das estações, para considerar situações semelhantes e descrever os perfis de chegada e saída de cada estação durante o dia, os dados foram agregados sem incluir os dias de finais de semana e feriados. Segundo Faghih-Imani et al.

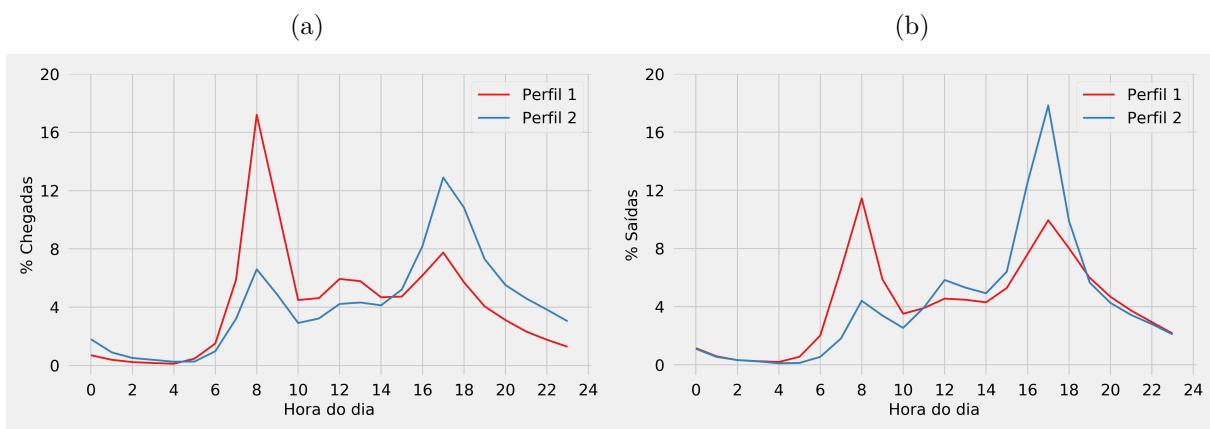
(2014), BIXI Montreal tem perfil de uso diferente ao longo do dia ao comparar finais de semanas e dias de trabalhos.

Figura 2 – Gráfico da Silhueta para o método KMeans



Fonte – Produzido pelo autor.

Figura 3 – Perfil do fluxo diário nas estações: (a) Chegada (b) Saída

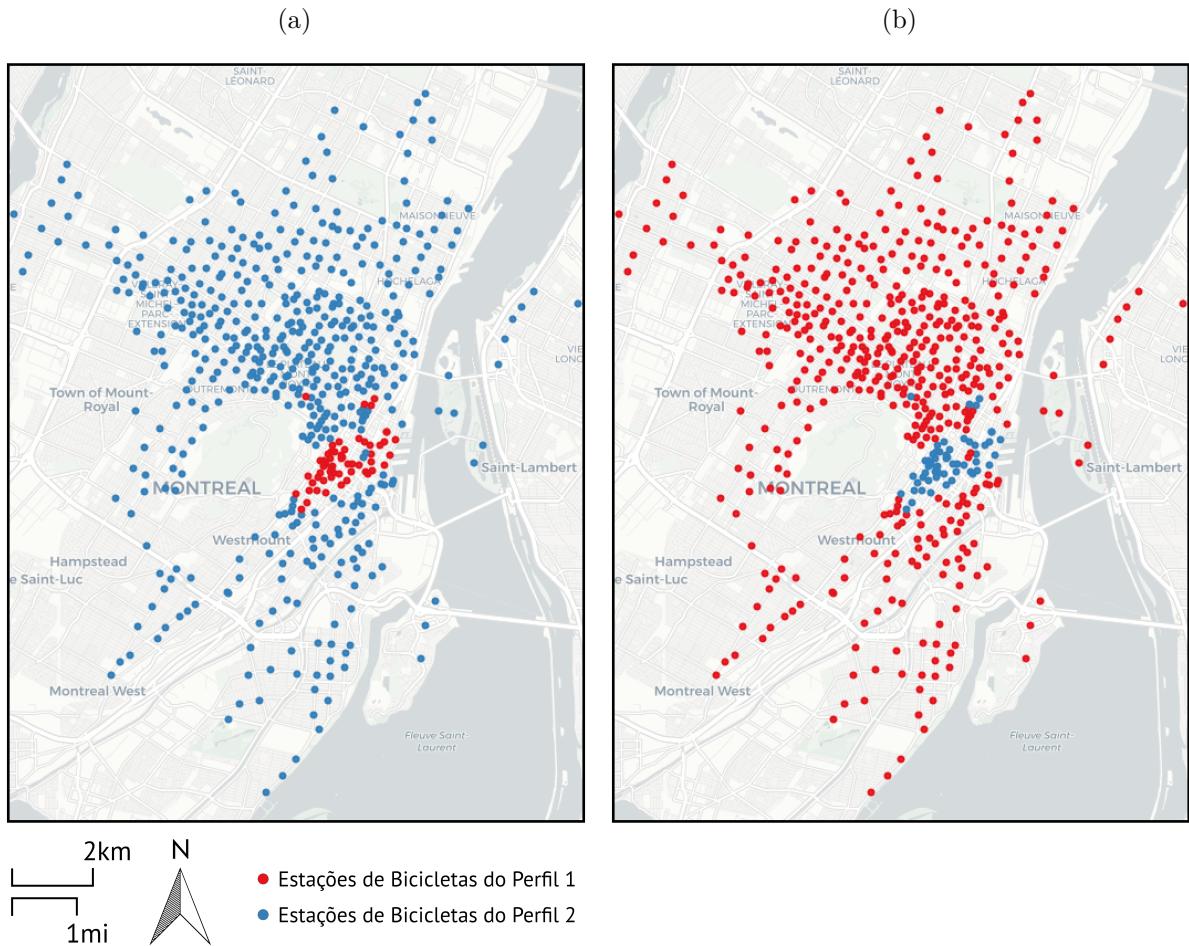


Fonte – Produzido pelo autor.

A Figura 2 apresenta o resultado do método da silhueta para o agrupamento KMeans. O maior valor para $s(i)$ ocorre com dois perfis e fica menor a partir de três, tanto para o fluxo de chegada quanto para o fluxo de saída, sugerindo a existência de dois perfis. A Figura 3 evidencia os perfis de cada um dos fluxos. Os padrões dos perfis parecem comparáveis entre os fluxos de chegada e saída:

- **Perfil 1:** agrupa as estações nas quais a maior parte das chegadas/saídas ocorrem por volta das 8h da manhã e 17h da tarde.
- **Perfil 2:** agrupa as estações nas quais a maior parte das chegadas/saídas ocorrem no período da tarde, por volta das 17h da tarde.

Figura 4 – Distribuição geográfica das estações de cada perfil segundo o fluxo: (a) Chegada
(b) Saída



Fonte – Produzido pelo autor.

A Figura 4 exibe a visualização da distribuição geográfica das estações de cada perfil segundo o fluxo. Nota-se que as estações localizadas no Distrito Central de Negócios da cidade de Montreal seguem o **Perfil 1**, para o fluxo de chegada, e o **Perfil 2** para o fluxo de saída. Ou seja, percebe-se que os usuários utilizam o sistema para chegar no distrito na parte da manhã e sair do distrito na parte da tarde.

Diante da análise de agrupamento, foram selecionadas as 4 estações com o maior quantidade de viagens. Para o fluxo de chegada, as estacões selecionadas foram: **Perfil 1** - estação 6015 e **Perfil 2** - estação 6012. Para o fluxo de saída: **Perfil 1** - estação 6184 e **Perfil 2** - estação 6100. A Tabela 7 apresenta as 5 estações com o maior volume de chegadas e partidas segundo o perfil.

Tabela 7 – Estações com o maior volume de Chegadas e Partidas segundo o perfil

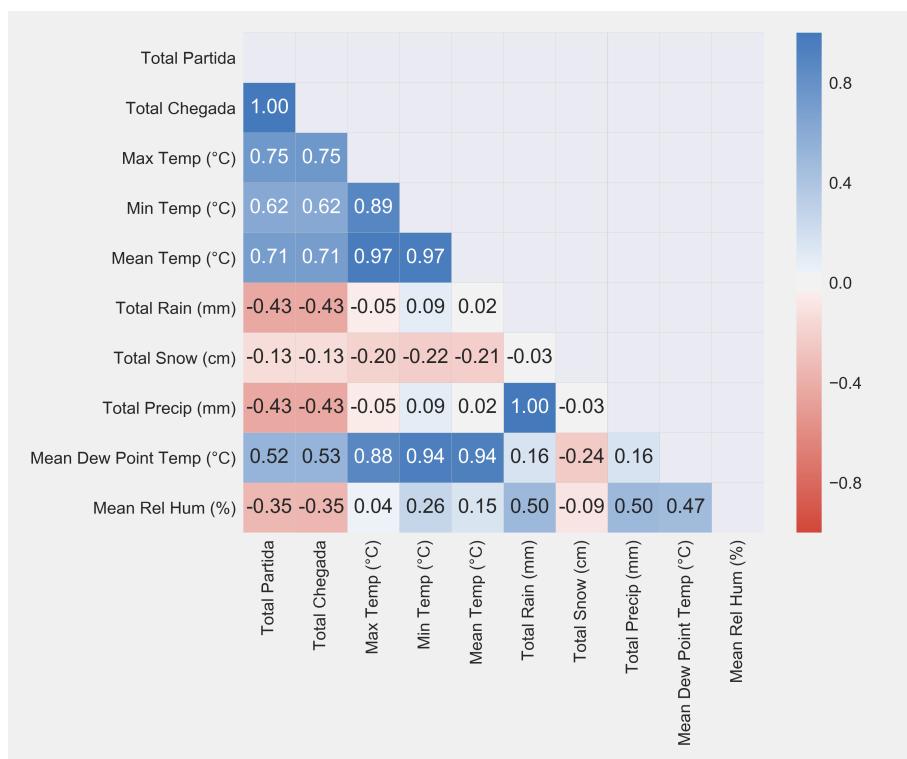
Fluxo de Chegada				Fluxo de Saída			
Perfil 1		Perfil 2		Perfil 1		Perfil 2	
Estação	Total	Estação	Total	Estação	Total	Estação	Total
6015	43.741	6012	33.774	6184	28.837	6100	43.112
6100	43.325	6184	27.853	6136	28.608	6078	30.963
6078	37.070	6052	27.364	6221	26.469	6064	28.567
6064	29.536	6070	24.664	6214	23.628	6015	27.861
6043	27.995	6136	22.213	6070	23.289	6067	17.533

Fonte – Produzido pelo autor.

5.2 VARIÁVEIS EXÓGENAS

Segundo a literatura, fatores climáticos influenciam no uso de sistemas de compartilhamento de bicicletas. Assim, as variáveis sobre o clima serão consideradas como exógenas nos modelos de predição deste trabalho. A Figura 5 mostra a matriz de correlação entre o fluxo de chegadas e saídas e as variáveis climáticas.

Figura 5 – Matriz de correlação entre o fluxos de chegada e saída e as variáveis sobre o clima por dia



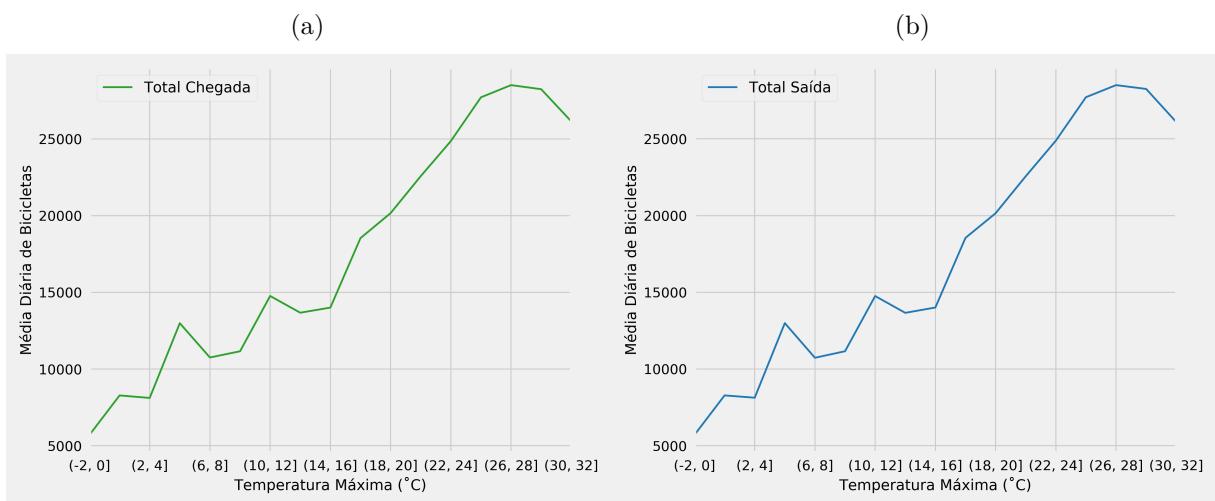
Fonte – Produzido pelo autor.

Existe uma correlação positiva forte entre as variáveis de temperatura. O ponto de orvalho tem correlação positiva forte com temperatura máxima (0,88), mínima (0,94) e média (0,94). Assim, esta variável não será considerada como exógena para o modelo. Já a temperatura máxima tem correlação positiva forte com o total de bicicletas que chegam (0,75) e que saem (0,75) das estações, e também está fortemente correlacionada com a temperatura mínima (0,89) e média (0,97). Então, dentre as variáveis sobre temperatura, somente a temperatura máxima será selecionada para a inclusão nos modelos.

Dentre as variáveis de precipitação, foi considerada nos modelos somente a precipitação total pois é altamente correlacionada com a precipitação de chuva (1,00), e a precipitação de neve tem correlação desprezível com todas as variáveis. A precipitação total tem correlação negativa com o total de bicicletas que chegam (0,43) e que saem (0,43). Já a umidade relativa tem correlação negativa com o total de bicicletas que chegam (0,35) e que saem (0,35), esta também será considerada nos modelos de predição.

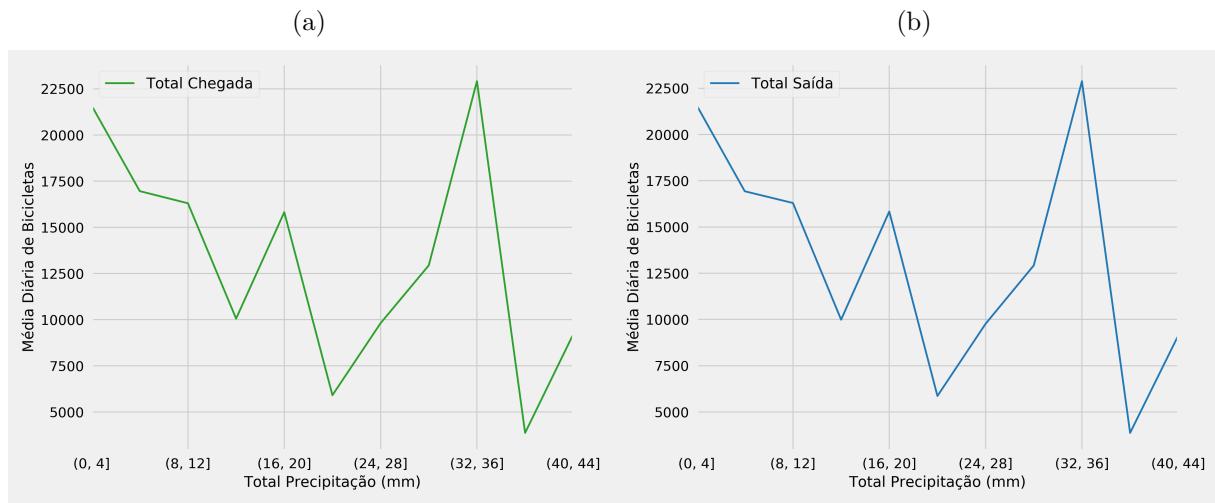
As Figuras 6, 7 e 8 mostram as relações com a quantidade média de bicicletas por dia e as variáveis temperatura máxima, precipitação e umidade relativa, respectivamente. Nota-se que quanto maior a temperatura, maior o número médio de bicicletas em cada fluxo. Já a relação com precipitação não apresenta uma relação linear e nem uma tendência clara pois há números altos e baixos para a quantidade de bicicletas quando a precipitação é alta. Para umidade relativa maior que 70%, o número médio de bicicletas por dia tende a reduzir para cada fluxo.

Figura 6 – Média diária de bicicletas em todo o sistema por temperatura máxima segundo o fluxo: (a) Chegada (b) Saída



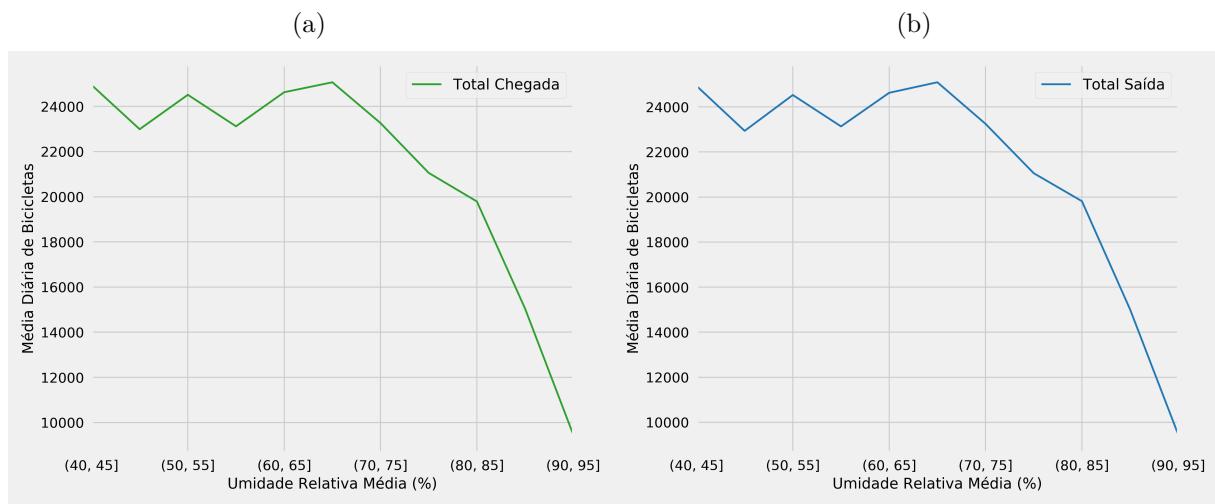
Fonte – Produzido pelo autor.

Figura 7 – Média diária de bicicletas em todo o sistema por precipitação total segundo o fluxo: (a) Chegada (b) Saída



Fonte – Produzido pelo autor.

Figura 8 – Média diária de bicicletas em todo o sistema por umidade relativa segundo o fluxo: (a) Chegada (b) Saída



Fonte – Produzido pelo autor.

As séries das variáveis exógenas temperatura máxima, precipitação total e umidade relativa consideradas nos modelos de predição são apresentadas na Figura 9.

Figura 9 – Séries das variáveis exógenas temperatura máxima, precipitação total e umidade relativa

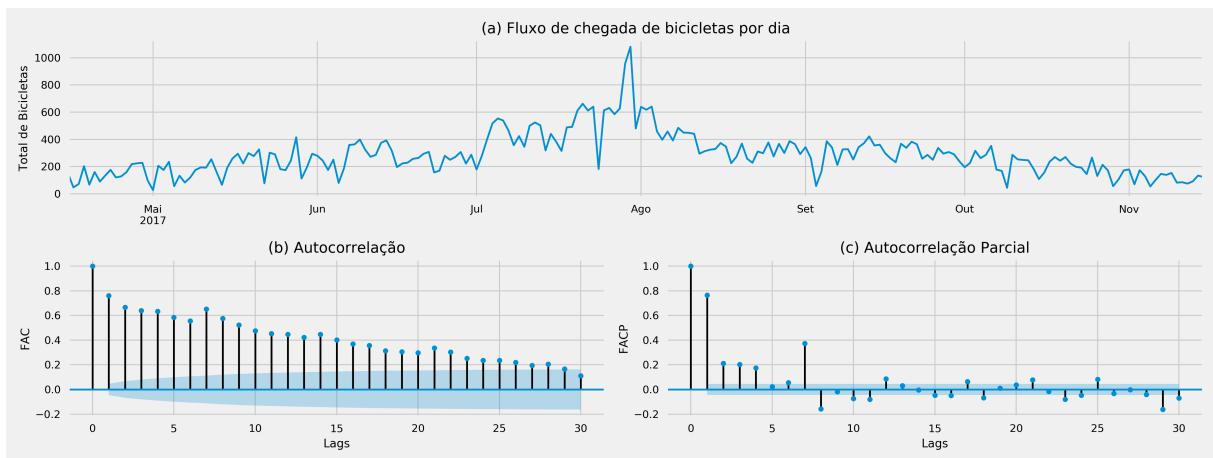


Fonte – Produzido pelo autor.

5.3 FLUXO DE CHEGADA

A primeira série analisada é a da estação 6015 que representa o Perfil 1 do Fluxo de Chegada de bicicletas. Observa-se pelo gráfico da série, Figura 10, um aumento no fluxo de bicicletas que chegam entre os meses de julho a setembro, indicando que o fluxo não é constante ao longo do tempo. A FAC e FACP para os primeiros 30 *lags*, indicam a presença de sazonalidade pois há um aumento nos *lags* múltiplos de 7 e o decaimento lento na FAC pode indicar a necessidade de diferenciação na série.

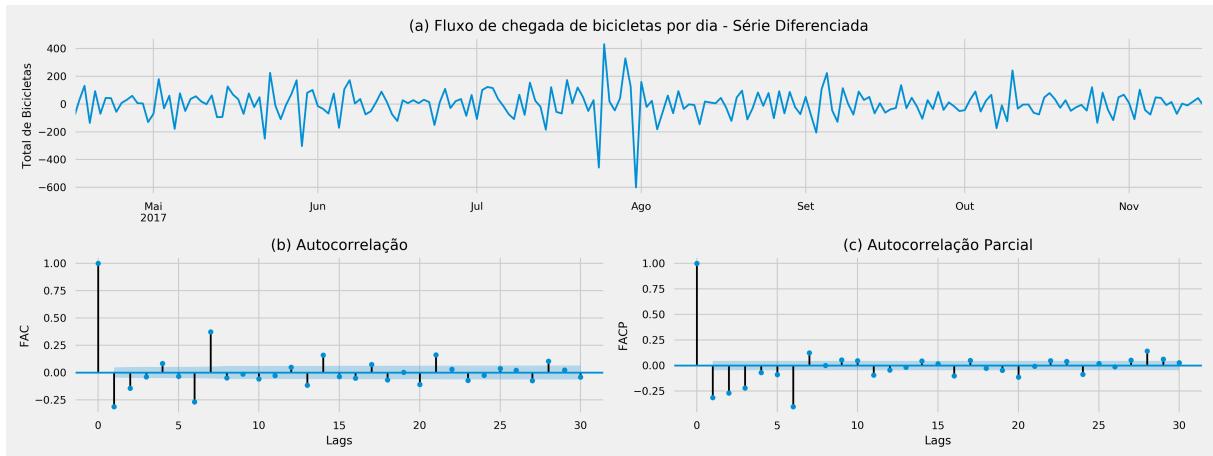
Figura 10 – Fluxo de chegada de bicicletas por dia, funções de autocorrelação e correlação parcial da série da estação 6015



Fonte – Produzido pelo autor.

Com a série diferenciada, Figura 11, a FAC e FACP apresentam significância no primeiro *lag* e alguns *lags* futuros, sugerindo um modelo ARMA.

Figura 11 – Fluxo de chegada de bicicletas por dia, funções de autocorrelação e correlação parcial da série diferenciada da estação 6015



Fonte – Produzido pelo autor.

Após testes com diversos modelos, verificou-se que um modelo ARIMAX(7, 1, 1, 3) incompleto, Equação 5.1, gera resíduos não correlacionados. Porém, os resíduos não seguem uma distribuição normal (teste Jarque-Bera: p-valor $\ll 0,001$). As Figuras 12 e 13 apresentam evidências para a análise dos resíduos e o teste de Ljung-Box, respectivamente. A Tabela 8 mostra os parâmetros e estimativas, todos significativos ao nível de 5%.

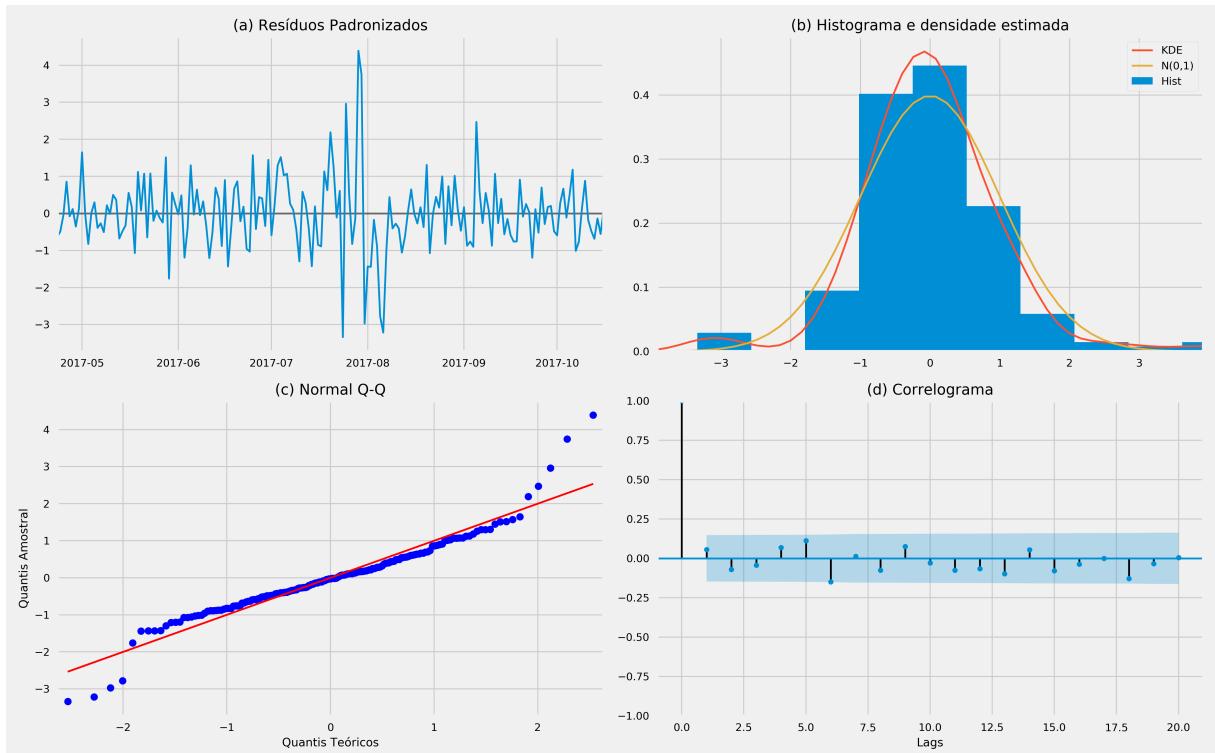
$$\Delta y_t = \phi_1 \Delta y_{t-1} + \phi_7 \Delta y_{t-7} + \theta_1 \epsilon_{t-1} + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \beta_3 X_{3,t} + \epsilon_t \quad (5.1)$$

Tabela 8 – Parâmetros e estimativas do modelo ARIMAX(7, 1, 1, 3) incompleto, Equação 5.1, para o Perfil 1 do Fluxo de Chegada

Parâmetro	Estimativa	Erro Padrão	Estatística Z	p-valor	IC 95%
β_1 : Max Temp (°C)	5,4280	1,934	2,807	0,005	(1,638, 9,218)
β_2 : Total Precip (mm)	-3,8506	0,672	-5,729	0,000	(-5,168, -2,533)
β_3 : Mean Rel Hum (%)	-2,3801	0,643	-3,703	0,000	(-3,640, -1,120)
ϕ_1 : AR.L1	0,3264	0,074	4,414	0,000	(0,181, 0,471)
ϕ_7 : AR.L7	0,3094	0,048	6,475	0,000	(0,216, 0,403)
θ_1 : MA.L1	-0,7550	0,070	-10,846	0,000	(-0,891, -0,619)

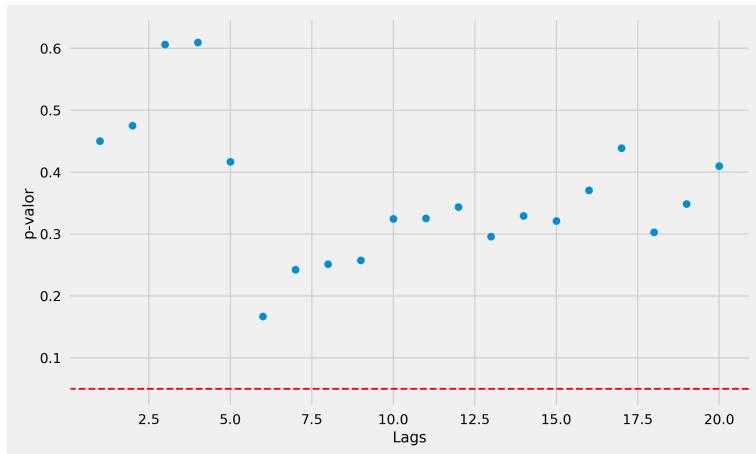
Fonte – Produzido pelo autor.

Figura 12 – Análise dos resíduos do modelo ARIMAX(7, 1, 1, 3) incompleto para o Perfil 1 do Fluxo de Chegada



Fonte – Produzido pelo autor.

Figura 13 – Teste de Ljung-Box dos resíduos do modelo ARIMAX(7, 1, 1, 3) incompleto para o Perfil 1 do Fluxo de Chegada



Fonte – Produzido pelo autor.

Opta-se por aplicar a transformação logarítmica na série original da estação 6015 para obter resíduos normalmente distribuídos. Após testar alguns modelos, verificou-se que um modelo ARIMAX(7, 1, 1, 3) incompleto, Equação 5.2, gerava resíduos não correlacionados e normalmente distribuídos (teste Jarque-Bera: p -valor = 0,97). A Figura 14 apresenta a análise dos resíduos e a Figura 15 o resultado do teste de Ljung-Box. A Tabela 9 mostra os parâmetros e estimativas, todos significativos ao nível de 5%.

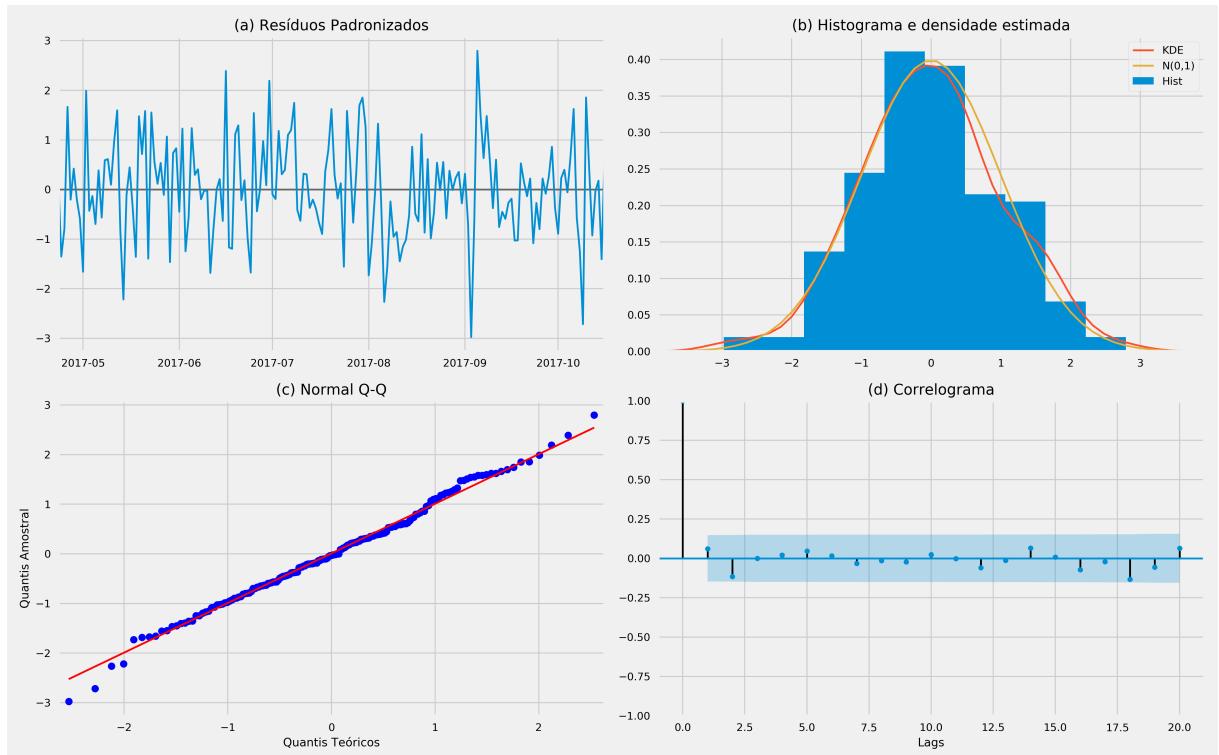
$$\Delta \ln y_t = \phi_3 \Delta \ln y_{t-3} + \phi_7 \Delta \ln y_{t-7} + \theta_1 \epsilon_{t-1} + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \beta_3 X_{3,t} + \epsilon_t \quad (5.2)$$

Tabela 9 – Parâmetros e estimativas do modelo ARIMAX(7, 1, 1, 3) incompleto, Equação 5.2, para o Perfil 1 do Fluxo de Chegada com transformação logarítmica

Parâmetro	Estimativa	Erro Padrão	Estatística Z	p-valor	IC 95%
β_1 : Max Temp (°C)	0,0286	0,005	5,837	0,000	(0,019, 0,038)
β_2 : Total Precip (mm)	-0,0201	0,002	-9,809	0,000	(-0,024, -0,016)
β_3 : Mean Rel Hum (%)	-0,0124	0,002	-6,704	0,000	(-0,016, -0,009)
ϕ_3 : AR.L3	-0,2420	0,079	-3,052	0,002	(-0,397, -0,087)
ϕ_7 : AR.L7	0,3022	0,076	3,985	0,000	(0,154, 0,451)
θ_1 : MA.L1	-0,6822	0,056	-12,245	0,000	(-0,791, -0,573)

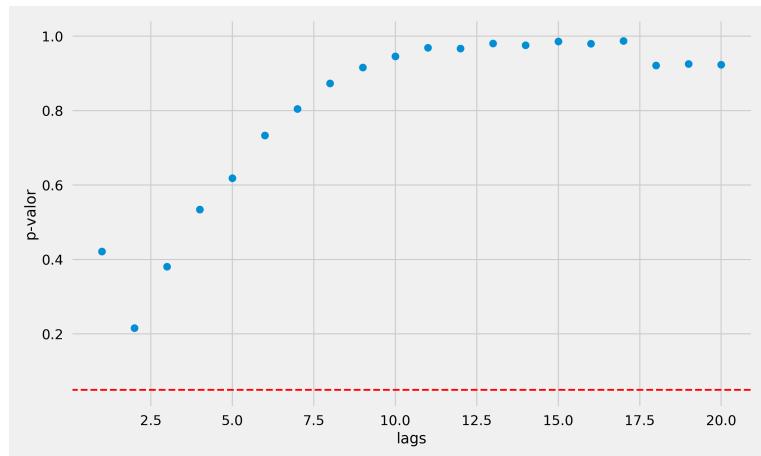
Fonte – Produzido pelo autor.

Figura 14 – Análise dos resíduos do modelo ARIMAX(7, 1, 1, 3) incompleto para o Perfil 1 do Fluxo de Chegada com transformação logarítmica



Fonte – Produzido pelo autor.

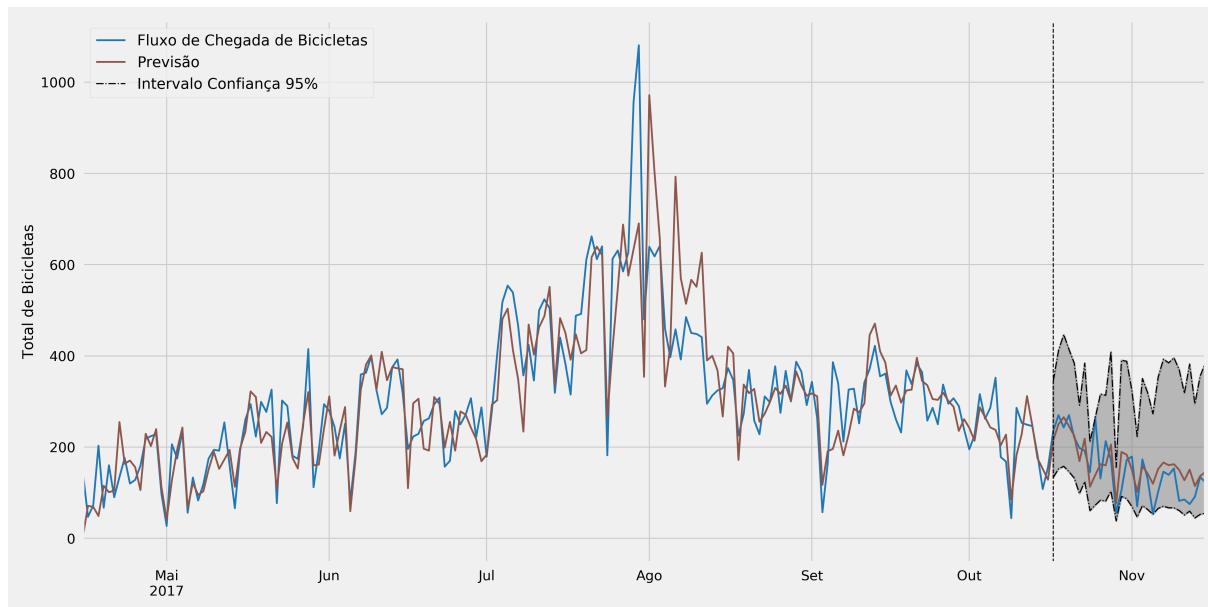
Figura 15 – Teste de Ljung-Box dos resíduos do modelo ARIMAX(7, 1, 1, 3) incompleto para o Perfil 1 do Fluxo de Chegada com transformação logarítmica



Fonte – Produzido pelo autor.

A Figura 16 apresenta as previsões para os últimos 30 dias de operação de 2017 com base no modelo ARIMAX(7, 1, 1, 3) incompleto com a série da estação 6015 usando transformação logarítmica, Equação 5.2, já com a transformação inversa. A raiz do EQM é igual a 43,39, o EAM tem valor 34,28. Apesar do intervalo de confiança de 95% das previsões serem largos, nota-se que os valores previstos são razoavelmente próximos dos valores observados com EAM de 1,42 bicicletas por hora. O modelo ajustado com a transformação apresenta o maior valor para o EAM (34,28), comparado ao EAM (31,55) do modelo sem a transformação, porém opta-se pelo modelo com a transformação por apresentar resíduos não correlacionados e normalmente distribuídos (Tabela 10). Outro ponto a ressaltar é o aumento do fluxo entre os meses de julho e agosto, que ocorreu devido ao não funcionamento de 5 das 12 estações a 500m de distância da estação 6015.

Figura 16 – Previsões do modelo ARIMAX(7, 1, 1, 3) incompleto para o Perfil 1 do Fluxo de Chegada com transformação logarítmica



Fonte – Produzido pelo autor.

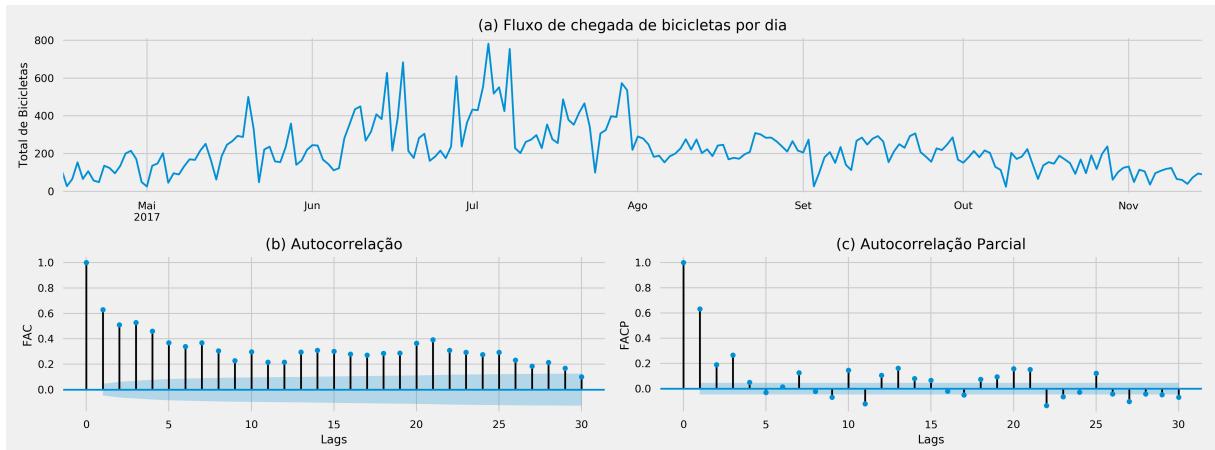
Tabela 10 – Estatísticas comparativas entre os modelos ajustados para o Perfil 1 do Fluxo de Chegada

Modelo	Raiz do EQM	EAM
ARIMAX(7, 1, 1, 3) incompleto	44,08	31,55
ARIMAX(7, 1, 1, 3) incompleto com transformação logarítmica	43,39	34,28

Fonte – Produzido pelo autor.

A segunda série analisada que representa o Perfil 2 do Fluxo de Chegada de bicicletas é a da estação 6012. No gráfico da série, Figura 17, observa-se que existe um aumento no fluxo de bicicletas que chegam entre os meses de julho e agosto, o que indica que o fluxo não é constante ao longo do tempo. Percebe-se também um decaimento lento na FAC indicando a necessidade de diferenciação na série.

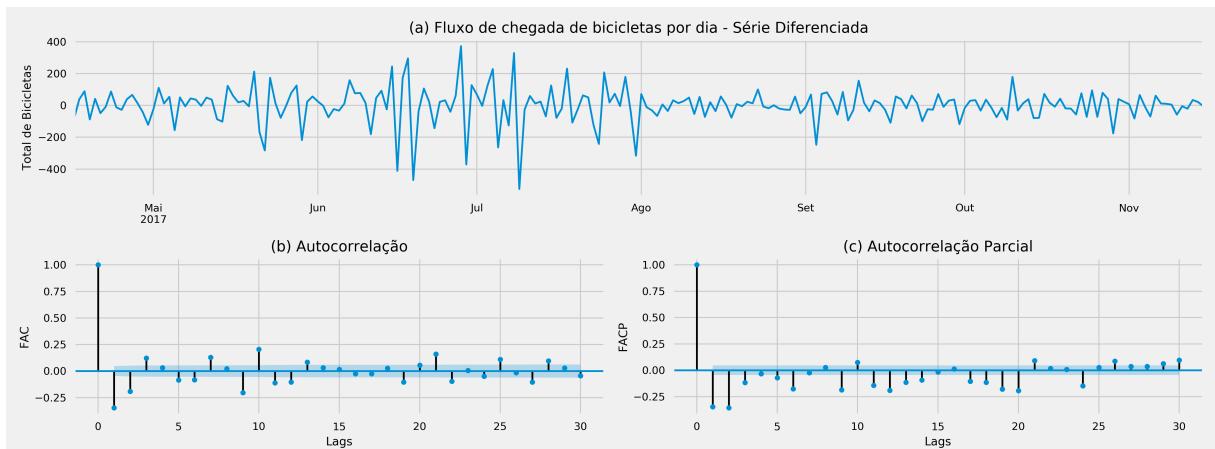
Figura 17 – Fluxo de chegada de bicicletas por dia, funções de autocorrelação e correlação parcial da série da estação 6012



Fonte – Produzido pelo autor.

A Figura 18 mostra a série diferenciada e os gráficos da função da FAC e FACP para os primeiros 30 lags. Verifica-se também que a FAC e FACP apresentam significância no primeiro *lag* e alguns *lags* futuros, o que sugere um modelo ARMA.

Figura 18 – Fluxo de chegada de bicicletas por dia, funções de autocorrelação e correlação parcial da série diferenciada da estação 6012



Fonte – Produzido pelo autor.

Também foram realizados testes com vários modelos e verificou-se que o ARIMAX(9, 1, 1, 3) incompleto, Equação 5.3, apresentava resíduos não correlacionados. Mas, os resíduos não são normalmente distribuídos (teste Jarque-Bera: p-valor $\ll 0,001$). As Figuras 19 e 20 apresentam a análise dos resíduos e o teste de Ljung-Box, respectivamente. A Tabela 11 mostra os parâmetros e estimativas, todos significativos ao nível de 5%.

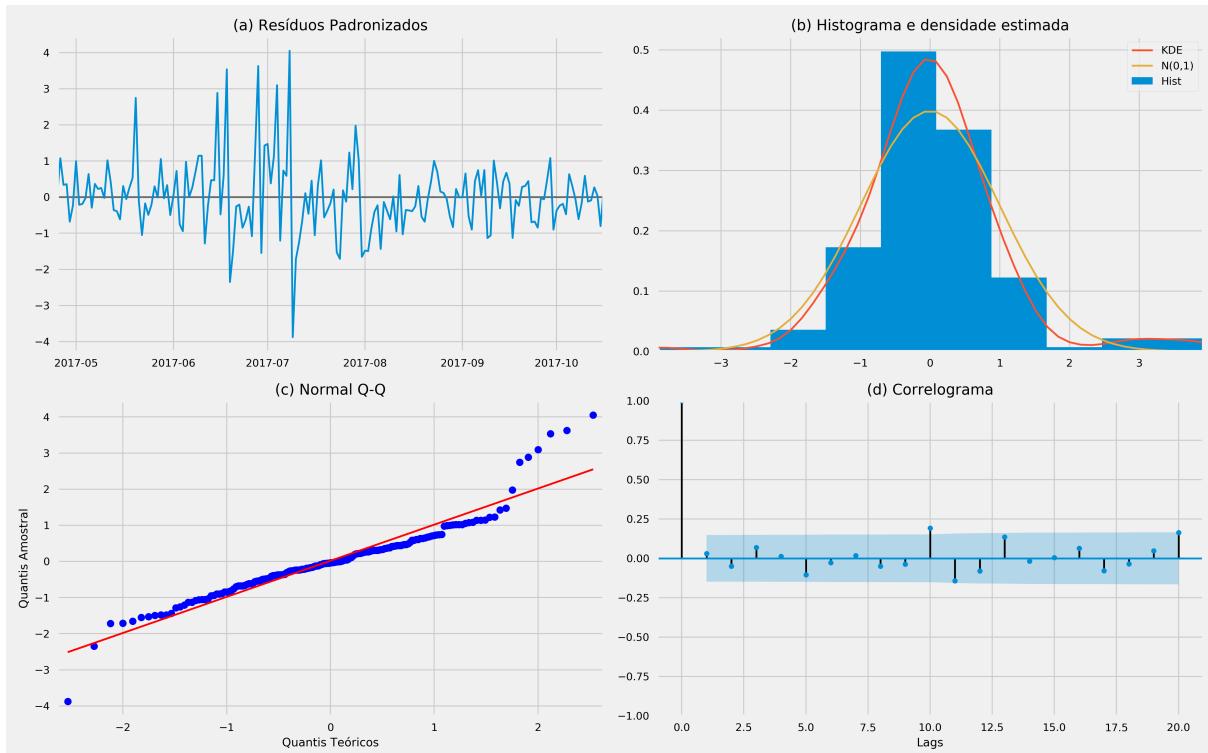
$$\Delta y_t = \phi_1 \Delta y_{t-1} + \phi_9 \Delta y_{t-9} + \theta_1 \epsilon_{t-1} + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \beta_3 X_{3,t} + \epsilon_t \quad (5.3)$$

Tabela 11 – Parâmetros e estimativas do modelo ARIMAX(9, 1, 1, 3) incompleto, Equação 5.3, para o Perfil 2 do Fluxo de Chegada

Parâmetro	Estimativa	Erro Padrão	Estatística Z	p-valor	IC 95%
β_1 : Max Temp (°C)	6,0154	2,142	2,808	0,005	(1,816, 10,214)
β_2 : Total Precip (mm)	-2,1345	0,924	-2,310	0,021	(-3,945, -0,324)
β_3 : Mean Rel Hum (%)	-2,9113	0,727	-4,005	0,000	(-4,336, -1,486)
ϕ_1 : AR.L1	0,2028	0,075	2,690	0,007	(0,055, 0,351)
ϕ_9 : AR.L9	-0,2362	0,076	-3,118	0,002	(-0,385, -0,088)
θ_1 : MA.L1	-0,8284	0,041	-20,252	0,000	(-0,909, -0,748)

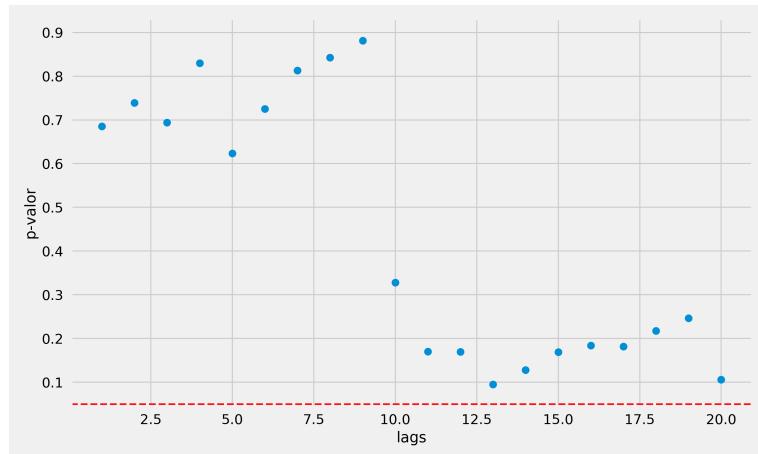
Fonte – Produzido pelo autor.

Figura 19 – Análise dos resíduos do modelo ARIMAX(9, 1, 1, 3) incompleto para o Perfil 2 do Fluxo de Chegada



Fonte – Produzido pelo autor.

Figura 20 – Teste de Ljung-Box dos resíduos do modelo ARIMAX(9, 1, 1, 3) incompleto para o Perfil 2 do Fluxo de Chegada



Fonte – Produzido pelo autor.

Também, para obter resíduos normalmente distribuídos, aplica-se a transformação logarítmica à série original da estação 6012. Depois de testar alguns modelos, verificou-se que um modelo ARIMAX(11, 1, 1, 3) incompleto, Equação 5.4, apresentava resíduos não correlacionados e, ao nível de 5%, os resíduos não são normalmente distribuídos (teste Jarque-Bera: $p\text{-valor} = 0,03$). Porém, opta-se por esse modelo pois foi o único que apresentou todos os parâmetros significativos ao nível de 5% e resíduos não correlacionados (Tabela 12). A Figura 21 mostra a análise dos resíduos e a Figura 22 o resultado do teste de Ljung-Box.

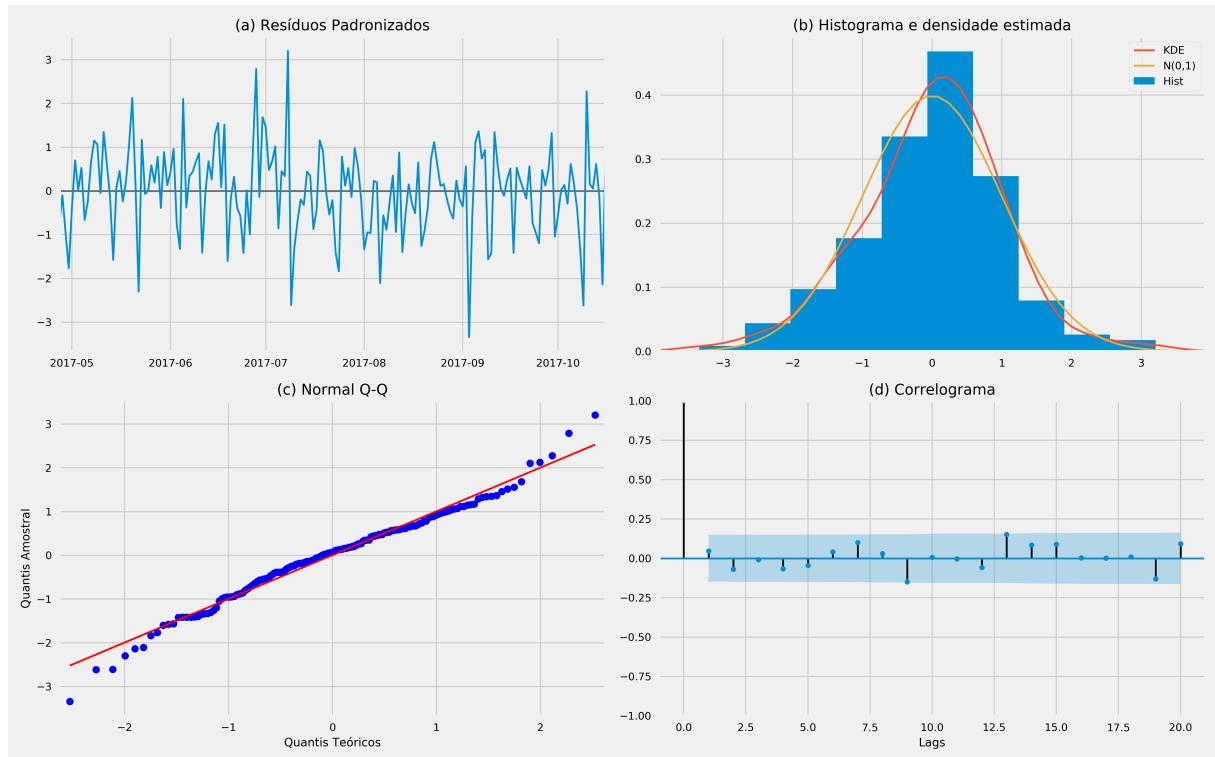
$$\Delta \ln y_t = \phi_1 \Delta \ln y_{t-1} + \phi_{11} \Delta \ln y_{t-11} + \theta_1 \epsilon_{t-1} + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \beta_3 X_{3,t} + \epsilon_t \quad (5.4)$$

Tabela 12 – Parâmetros e estimativas do modelo ARIMAX(11, 1, 1, 3) incompleto, Equação 5.4, para o Perfil 2 do Fluxo de Chegada com transformação logarítmica

Parâmetro	Estimativa	Erro Padrão	Estatística Z	p-valor	IC 95%
β_1 : Max Temp (°C)	0,0344	0,007	4,934	0,000	(0,021, 0,048)
β_2 : Total Precip (mm)	-0,0187	0,003	-6,700	0,000	(-0,024, -0,013)
β_3 : Mean Rel Hum (%)	-0,0134	0,003	-5,112	0,000	(-0,018, -0,008)
ϕ_1 : AR.L1	0,2579	0,089	2,901	0,004	(0,084, 0,432)
ϕ_{11} : AR.L11	-0,1575	0,080	-1,979	0,048	(-0,314, -0,002)
θ_1 : MA.L1	-0,8767	0,048	-18,077	0,000	(-0,972, -0,782)

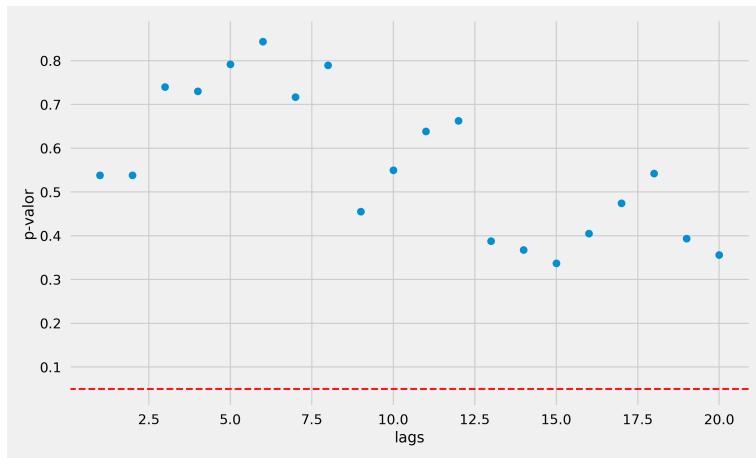
Fonte – Produzido pelo autor.

Figura 21 – Análise dos resíduos do modelo ARIMAX(11, 1, 1, 3) incompleto para o Perfil 2 do Fluxo de Chegada com transformação logarítmica



Fonte – Produzido pelo autor.

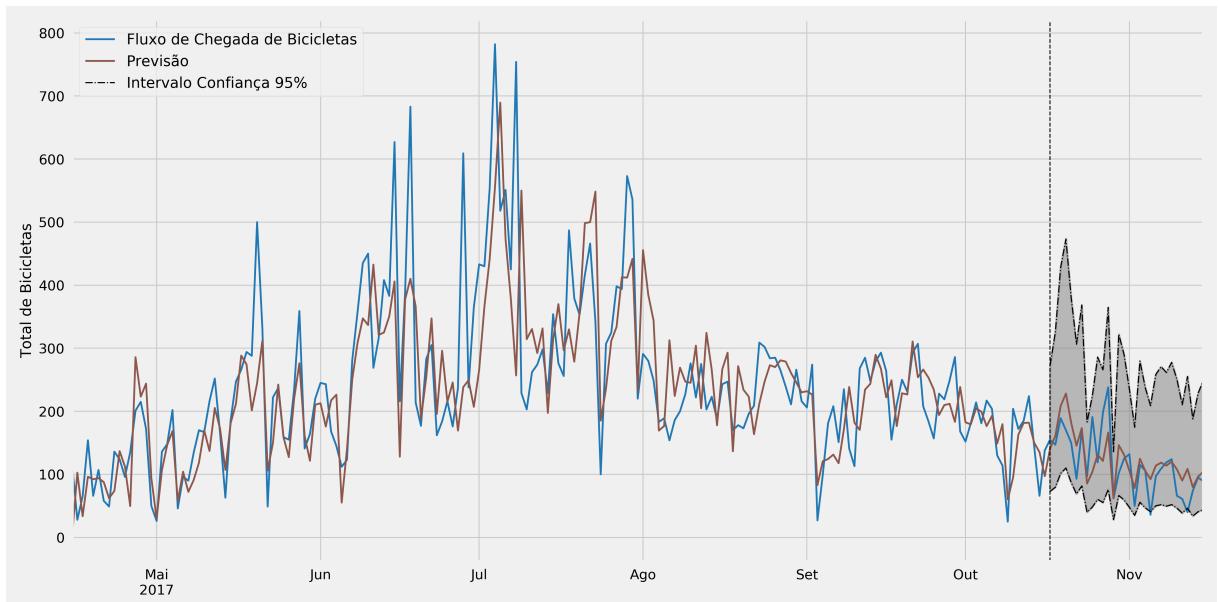
Figura 22 – Teste de Ljung-Box dos resíduos do modelo ARIMAX(11, 1, 1, 3) incompleto para o Perfil 2 do Fluxo de Chegada com transformação logarítmica



Fonte – Produzido pelo autor.

A Figura 23 exibe as previsões para os últimos 30 dias de operação de 2017 com base no modelo ARIMAX(11, 1, 1, 3) incompleto com a série da estação 6012 usando transformação logarítmica, Equação 5.4, já com a transformação inversa. A raiz do EQM é igual a 37,22 e o EAM tem valor 27,41. Apesar do intervalo de confiança de 95% das previsões serem largos, as previsões também acompanham bem os valores observados, com erro absoluto médio de 1,14 bicicletas por hora. O modelo ajustado com a transformação apresenta o menor valor para o EAM (27,41), comparado ao EAM (29,89) do modelo sem a transformação, indicando um melhor ajuste com a transformação (Tabela 13). Outro ponto a ressaltar é o aumento do fluxo entre os meses de junho e agosto, que ocorreu devido ao não funcionamento de 4 das 13 estações a 500m de distância da estação 6012.

Figura 23 – Previsões do modelo ARIMAX(11, 1, 1, 3) incompleto para o Perfil 2 do Fluxo de Chegada com transformação logarítmica



Fonte – Produzido pelo autor.

Tabela 13 – Estatísticas comparativas entre os modelos ajustados para o Perfil 2 do Fluxo de Chegada

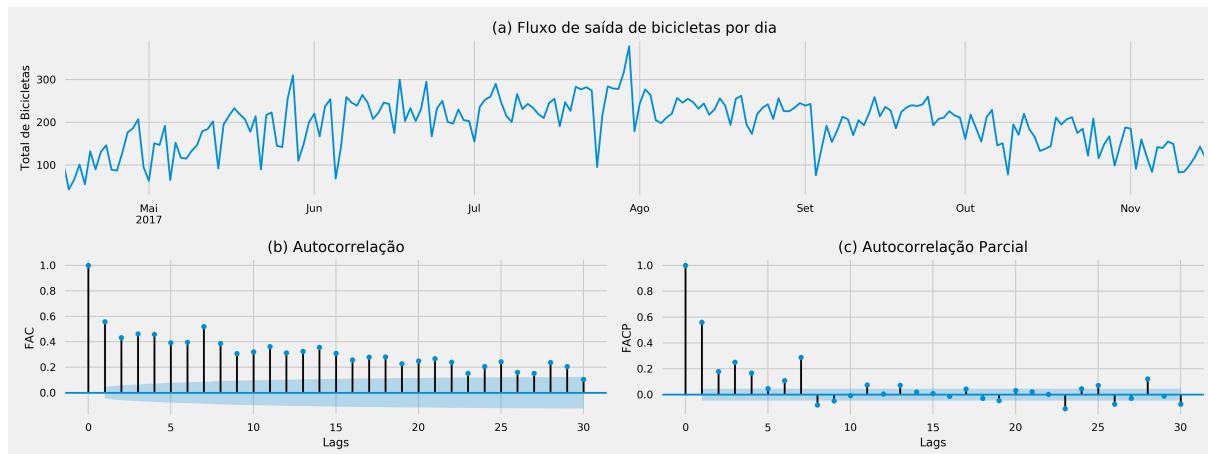
Modelo	Raiz do EQM	EAM
ARIMAX(9, 1, 1, 3) incompleto	36,40	29,89
ARIMAX(11, 1, 1, 3) incompleto com transformação logarítmica	37,22	27,41

Fonte – Produzido pelo autor.

5.4 FLUXO DE SAÍDA

A terceira série analisada é da estação 6184 que representa o Perfil 1 do Fluxo de Saída de bicicletas. Na Figura 24, observa-se pelo gráfico da série um aumento no fluxo de bicicletas que saem entre os meses de junho e outubro, indicando que ao longo do tempo o fluxo não é constante. O decaimento lento na FAC indica a necessidade de diferenciação na série.

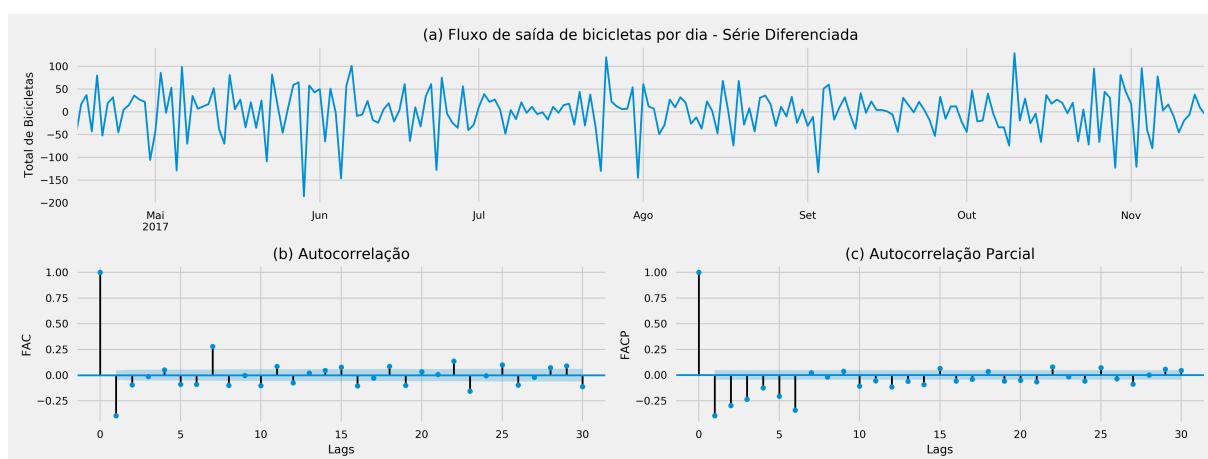
Figura 24 – Fluxo de saída de bicicletas por dia, funções de autocorrelação e correlação parcial da série da estação 6184



Fonte – Produzido pelo autor.

Encontram-se na Figura 25 a série diferenciada e os gráficos da função da FAC e FACP para os primeiros 30 *lags*. A FAC e FACP apresentam significância no primeiro *lag* e alguns *lags* futuros, sugerindo um modelo ARMA.

Figura 25 – Fluxo de saída de bicicletas por dia, funções de autocorrelação e correlação parcial da série diferenciada da estação 6184



Fonte – Produzido pelo autor.

Diante disto, diversos modelos foram testados e verificou-se que um modelo ARIMAX(7, 1, 1, 3) incompleto, Equação 5.5, apresentou resíduos não correlacionados e, ao nível de 5%, os resíduos não seguem uma distribuição normal (teste Jarque-Bera: p-valor = 0,02). Foram testados diversos modelos com transformação logarítmica e quadrática cujos resíduos não eram normalmente distribuídos. Assim, opta-se pelo modelo da Equação 5.5 por ter todos os parâmetros significativos ao nível de 5%, Tabela 14. As Figuras 26 e 27 apresentam a análise dos resíduos e o teste de Ljung-Box, respectivamente.

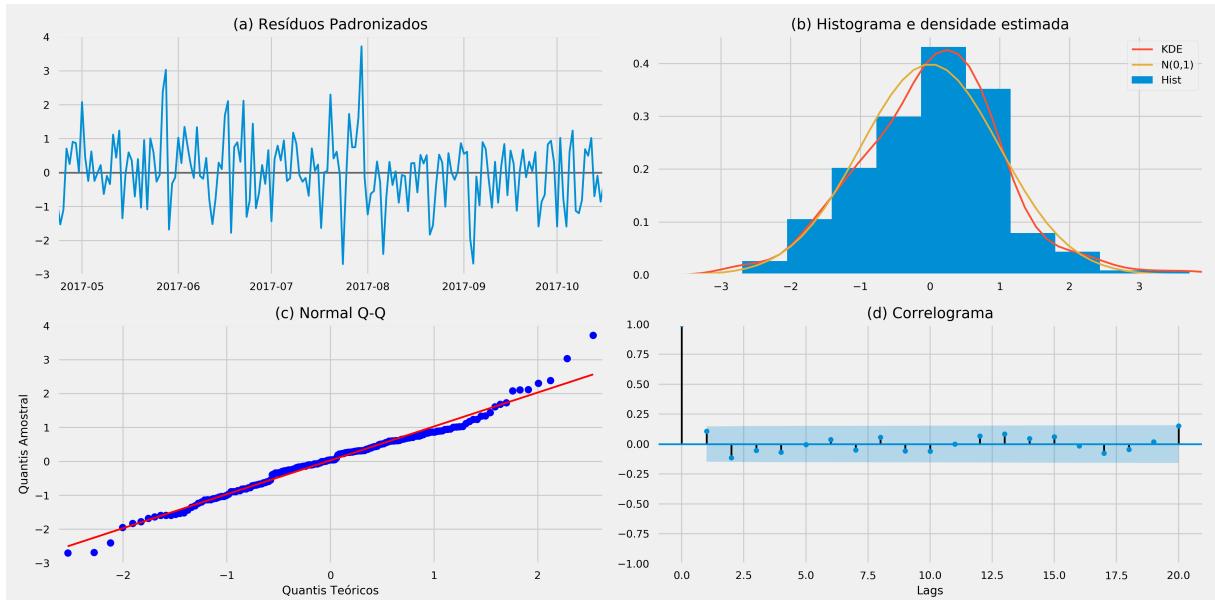
$$\Delta y_t = \phi_7 \Delta y_{t-7} + \theta_1 \epsilon_{t-1} + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \beta_3 X_{3,t} + \epsilon_t \quad (5.5)$$

Tabela 14 – Parâmetros e estimativas do modelo ARIMAX(7, 1, 1, 3) incompleto, Equação 5.5, para o Perfil 1 do Fluxo de Saída

Parâmetro	Estimativa	Erro Padrão	Estatística Z	p-valor	IC 95%
β_1 : Max Temp (°C)	3,9212	0,615	6,380	0,000	(2,717, 5,126)
β_2 : Total Precip (mm)	-2,2359	0,358	-6,240	0,000	(-2,938, -1,534)
β_3 : Mean Rel Hum (%)	-0,9335	0,239	-3,913	0,000	(-1,401, -0,466)
ϕ_7 : A.R.L7	0,2587	0,074	3,481	0,000	(0,113, 0,404)
θ_1 : M.A.L1	-0,9053	0,035	-26,124	0,000	(-0,973, -0,837)

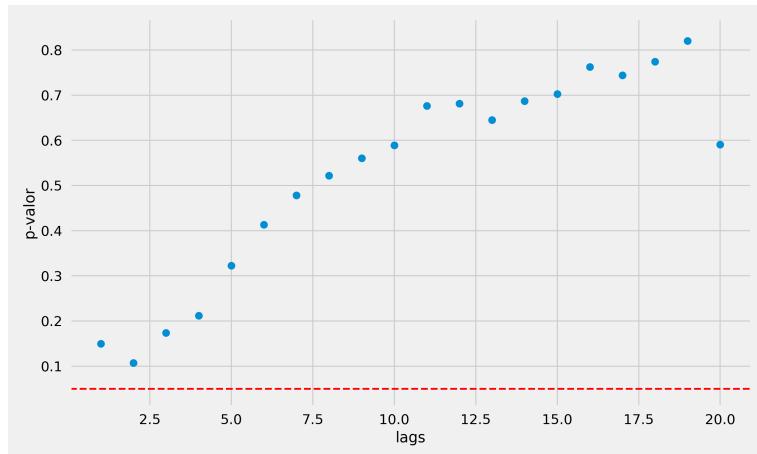
Fonte – Produzido pelo autor.

Figura 26 – Análise dos resíduos do modelo ARIMAX(7, 1, 1, 3) incompleto para o Perfil 1 do Fluxo de Saída



Fonte – Produzido pelo autor.

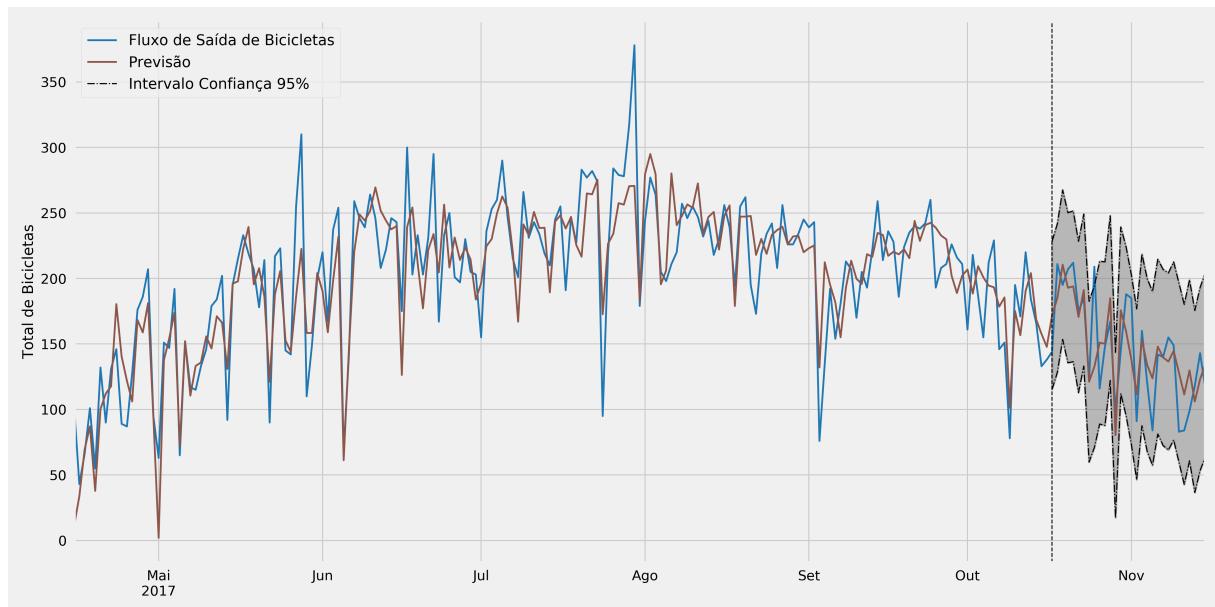
Figura 27 – Teste de Ljung-Box dos resíduos do modelo ARIMAX(7, 1, 1, 3) incompleto para o Perfil 1 do Fluxo de Saída



Fonte – Produzido pelo autor.

A Figura 28 apresenta as previsões para os últimos 30 dias de operação de 2017 com base no modelo ARIMAX(7, 1, 1, 3) incompleto, Equação 5.5, para a série da estação 6184. A raiz do EQM é igual a 26,57 e o EAM obtido é 21,07. Apesar do intervalo de confiança de 95% das previsões serem largos, nota-se que as previsões são próximas dos valores observados com erro absoluto médio de 1,10 bicicletas por hora.

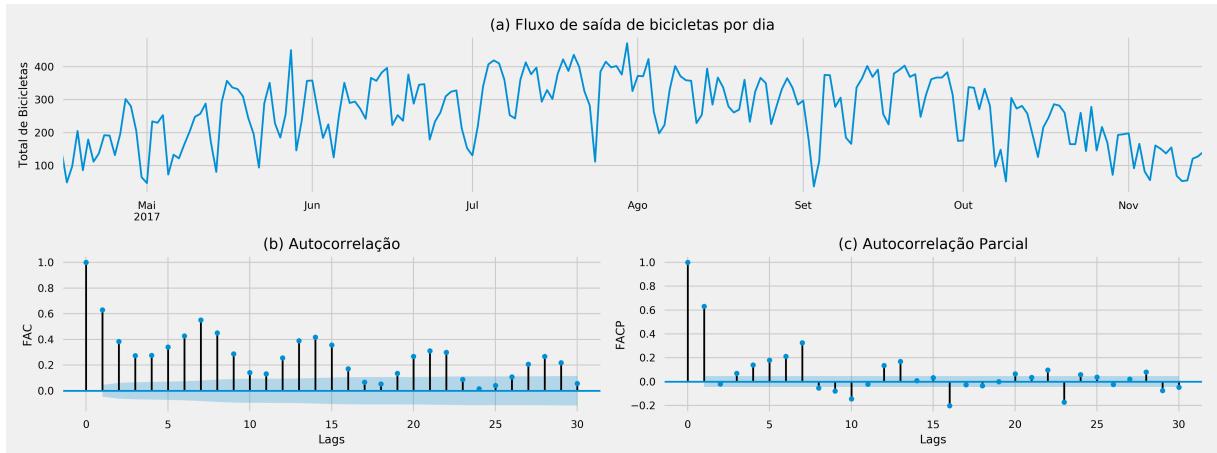
Figura 28 – Previsões do modelo ARIMAX(7, 1, 1, 3) incompleto para o Perfil 1 do Fluxo de Saída



Fonte – Produzido pelo autor.

A quarta série analisada representa o Perfil 2 do Fluxo de Saída de bicicletas, que é a da estação 6100. A Figura 29 apresenta o gráfico da série e observa-se um aumento no fluxo de bicicletas que saem entre os meses de julho e outubro, mostrando que o fluxo não é constante ao longo do tempo. Percebe-se também um decaimento lento na FAC indicando a necessidade de diferenciação na série. Adicionalmente observa-se um padrão sazonal pois há um pico na autocorrelação nos *lags* múltiplos de 7.

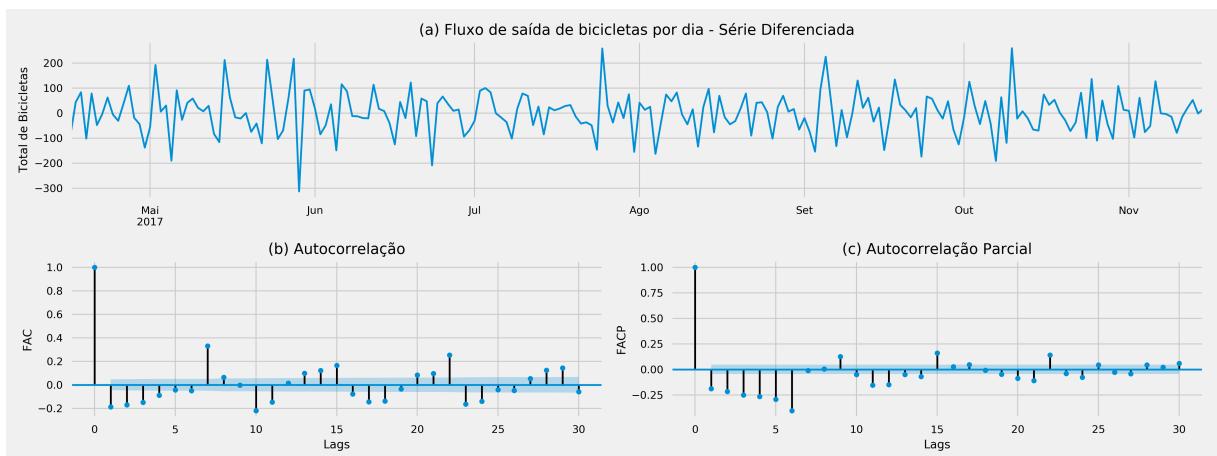
Figura 29 – Fluxo de saída de bicicletas por dia, funções de autocorrelação e correlação parcial da série da estação 6100



Fonte – Produzido pelo autor.

A Figura 30 mostra a série diferenciada e os gráficos da função da FAC e FACP para os primeiros 30 *lags*. Verifica-se também que a FAC e FACP apresentam significância no primeiro *lag* e alguns *lags* futuros, o que indica um modelo ARMA.

Figura 30 – Fluxo de saída de bicicletas por dia, funções de autocorrelação e correlação parcial da série diferenciada da estação 6100



Fonte – Produzido pelo autor.

Foram testados diversos modelos, e obteve-se que um modelo ARIMAX(8, 1, 3, 3) incompleto, Equação 5.6, apresentava resíduos não correlacionados e normalmente distribuídos (teste Jarque-Bera: p-valor = 0,20). As Figuras 31 e 32 apresentam a análise dos resíduos e o teste de Ljung-Box, respectivamente. A Tabela 15 mostra os parâmetros e estimativas, todos significativos ao nível de 5%.

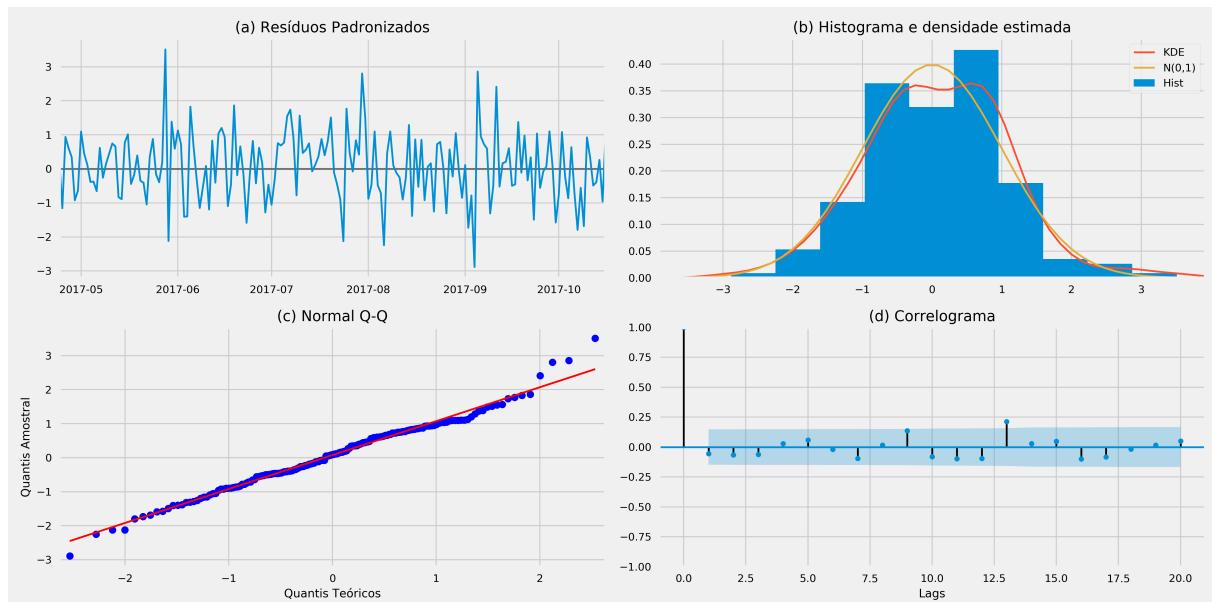
$$\Delta y_t = \phi_1 \Delta y_{t-1} + \phi_7 \Delta y_{t-7} + \phi_8 \Delta y_{t-8} + \theta_2 \epsilon_{t-2} + \theta_3 \epsilon_{t-3} + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \beta_3 X_{3,t} + \epsilon_t \quad (5.6)$$

Tabela 15 – Parâmetros e estimativas do modelo ARIMAX(8, 1, 3, 3) incompleto, Equação 5.6, para o Perfil 2 do Fluxo de Saída

Parâmetro	Estimativa	Erro Padrão	Estatística Z	p-valor	IC 95%
β_1 : Max Temp (°C)	5,9405	1,169	5,080	0,000	(3,649, 8,232)
β_2 : Total Precip (mm)	-3,2903	0,529	-6,222	0,000	(-4,327, -2,254)
β_3 : Mean Rel Hum (%)	-1,3127	0,467	-2,809	0,005	(-2,229, -0,397)
ϕ_1 : AR.L1	-0,5434	0,083	-6,582	0,000	(-0,705, -0,382)
ϕ_7 : AR.L7	0,4270	0,066	6,497	0,000	(0,298, 0,556)
ϕ_8 : AR.L8	0,3693	0,093	3,992	0,000	(0,188, 0,551)
θ_2 : MA.L2	-0,6852	0,084	-8,139	0,000	(-0,850, -0,520)
θ_3 : MA.L3	-0,3009	0,067	-4,482	0,000	(-0,432, -0,169)

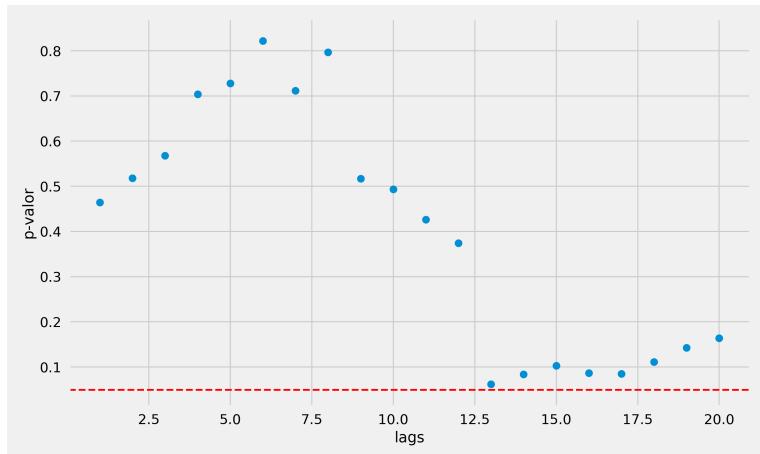
Fonte – Produzido pelo autor.

Figura 31 – Análise dos resíduos do modelo ARIMAX(8, 1, 3, 3) incompleto para o Perfil 2 do Fluxo de Saída



Fonte – Produzido pelo autor.

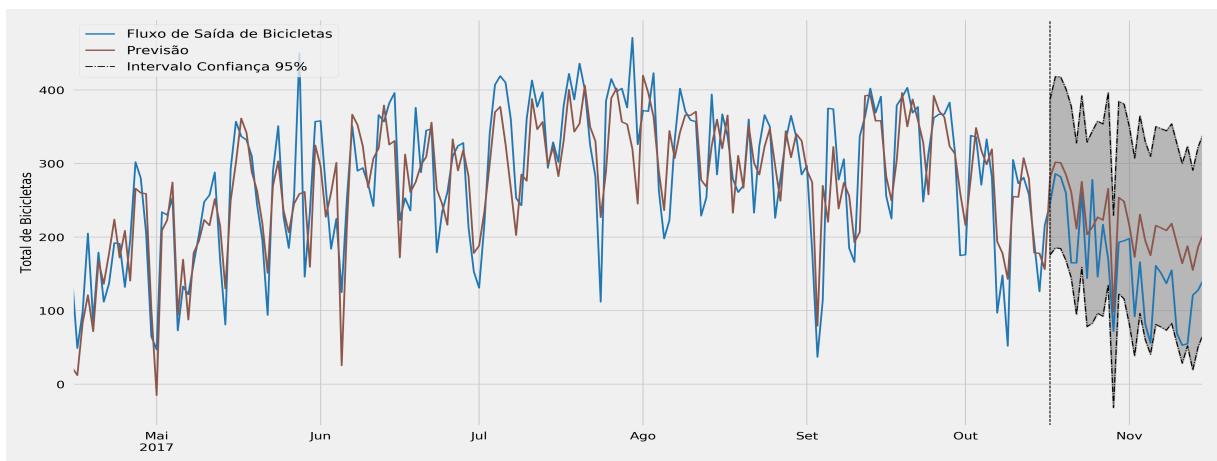
Figura 32 – Teste de Ljung-Box dos resíduos do modelo ARIMAX(8, 1, 3, 3) incompleto para o Perfil 2 do Fluxo de Saída



Fonte – Produzido pelo autor.

A Figura 33 mostra as previsões para os últimos 30 dias de operação de 2017 com base no modelo ARIMAX(8, 1, 3, 3) incompleto, Equação 5.6, para a série da estação 6100. A raiz do EQM é igual a 71,28 e o EAM estimado é 62,33. Apesar do intervalo de confiança de 95% das previsões serem largos e que a previsão superestima consistentemente a série observada, percebe-se que os valores previstos também são próximos dos valores observados com erro absoluto médio de 2,59 bicicletas por hora.

Figura 33 – Previsões do modelo ARIMAX(8, 1, 3, 3) incompleto para o Perfil 2 do Fluxo de Saída



Fonte – Produzido pelo autor.

6 CONCLUSÃO

Devido às características dos dados de uso do sistema de compartilhamento de bicicletas, este trabalho contribui para a literatura propondo o uso de modelos de séries temporais com variáveis exógenas para a previsão do fluxo diário de chegada e saída de bicicletas nas estações.

Neste trabalho foi observado que cada fluxo pode ser classificado em dois perfis de uso: (a) o Perfil 1 no qual a maior parte das chegadas/saídas ocorre por volta das 8h e 17h; e (b) o Perfil 2 no qual a maior parte das chegadas/saídas ocorrem por volta das 17h. Para estimar modelos de previsão, foram analisadas o comportamento das quatro estações com o maior demanda para cada perfil de cada fluxo.

O uso dos modelos ARIMAX (com as variáveis exógenas temperatura máxima, precipitação total e umidade relativa) propostos para realizar as previsões dos fluxos de chegada e saída de bicicletas coloca-se como uma boa abordagem para o problema, pois os fluxos de chegadas e saídas previstos podem ser considerados próximos dos valores observados, com erro absoluto médio de menos de 3 bicicletas por hora.

Este trabalho poderá ser útil como possível ferramenta para os gestores de sistemas de compartilhamento de bicicletas pois com o uso de modelos ARIMAX será possível prever a quantidade de bicicletas que chegam e saem das estações, auxiliando na otimização do uso do sistema através da redistribuição das bicicletas.

Para trabalhos futuros, propõe-se investir em estudos para aprimorar as previsões com o uso de uma abordagem multivariada, com base em modelos dos tipos: Vetorial Auto-Regressivo e Redes Neurais Recorrentes. Recomenda-se considerar os fluxos de chegadas e saídas como variáveis endógenas do modelo, lembrando que a quantidade de bicicletas que saem das estações dependem da quantidade de bicicletas que chegam nas estações.

REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. *IEEE transactions on automatic control*, Ieee, v. 19, n. 6, p. 716–723, 1974. Citado na página 33.
- BENNETT, C.; STEWART, R. A.; LU, J. Autoregressive with exogenous variables and neural network short-term load forecast models for residential low voltage distribution networks. *Energies*, Multidisciplinary Digital Publishing Institute, v. 7, n. 5, p. 2938–2960, 2014. Citado na página 30.
- BIXI MONTREAL. *Open Data*. 2018. Disponível em: <<https://www.bixi.com/en/open-data>>. Citado 2 vezes nas páginas 27 e 28.
- BIXI MONTREAL. *Who we are*. 2018. Disponível em: <<https://www.bixi.com/en/who-we-are>>. Citado na página 21.
- BOX, G. E. et al. *Time series analysis: forecasting and control*. [S.l.]: John Wiley & Sons, 2015. Citado na página 31.
- BURDEN, A. M. et al. Bike-share opportunities in new york city. *New York: Department of City Planning*, 2009. Citado na página 17.
- CHEN, P.-C. et al. Prediction of station level demand in a bike sharing system using recurrent neural networks. In: IEEE. *Vehicular Technology Conference (VTC Spring), 2017 IEEE 85th*. [S.l.], 2017. p. 1–5. Citado na página 24.
- DEGROOT, M. H.; SCHERVISH, M. J. *Probability and statistics*. [S.l.]: Pearson Education, 2012. Citado na página 33.
- DEMAIO, P. Bike-sharing: History, impacts, models of provision, and future. *Journal of public transportation*, v. 12, n. 4, p. 3, 2009. Citado 4 vezes nas páginas 17, 19, 21 e 22.
- DEMAIO, P.; MEDDIN, R. *The bike-sharing world map*. 2018. Disponível em: <<https://www.bikesharingmap.com>>. Citado na página 20.
- FAGHIH-IMANI, A.; ELURU, N. Analysing bicycle-sharing system user destination choice preferences: Chicago's divvy system. *Journal of Transport Geography*, Elsevier, v. 44, p. 53–64, 2015. Citado na página 24.
- FAGHIH-IMANI, A. et al. How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (bixi) in montreal. *Journal of Transport Geography*, Elsevier, v. 41, p. 306–314, 2014. Citado 4 vezes nas páginas 19, 21, 23 e 36.
- GOOGLE. *Google Places API Web Service*. 2018. Disponível em: <<https://developers.google.com/places/web-service/intro>>. Citado 2 vezes nas páginas 28 e 29.
- GOVERNMENT OF CANADA. *Historical Climate Data*. 2018. Disponível em: <<http://climate.weather.gc.ca/>>. Citado 2 vezes nas páginas 28 e 29.
- JARQUE, C. M.; BERA, A. K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, Elsevier, v. 6, n. 3, p. 255–259, 1980. Citado na página 33.

- LJUNG, G. M.; BOX, G. E. On a measure of lack of fit in time series models. *Biometrika*, Oxford University Press, v. 65, n. 2, p. 297–303, 1978. Citado na página 32.
- LOZANO, Á. et al. Multi-agent system for demand prediction and trip visualization in bike sharing systems. *Applied Sciences*, v. 8, n. 1, p. 67, 2018. Citado na página 19.
- MCQUEEN, J. Some methods for classification and analysis of multivariate observations. In: *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability, 1967*. Berkeley, Calif.: University of California Press, 1967. p. 281–297. Citado na página 34.
- MORETTIN, P. A.; TOLOI, C. Análise de séries temporais. In: *Análise de séries temporais*. [S.l.: s.n.], 2006. Citado na página 30.
- PARIKH, P.; UKKUSURI, S. V. Estimation of optimal inventory levels at stations of a bicycle sharing system. In: *Transportation Research Board Annual Meeting*. [S.l.: s.n.], 2015. v. 15. Citado na página 23.
- ROUSSEEUW, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, Elsevier, v. 20, p. 53–65, 1987. Citado na página 34.
- RUDLOFF, C.; LACKNER, B. Modeling demand for bikesharing systems: neighboring stations as source for demand and reason for structural breaks. *Transportation Research Record: Journal of the Transportation Research Board*, Transportation Research Board of the National Academies, n. 2430, p. 1–11, 2014. Citado na página 23.
- SHAHEEN, S.; GUZMAN, S.; ZHANG, H. Bikesharing in europe, the americas, and asia: past, present, and future. *Transportation Research Record: Journal of the Transportation Research Board*, Transportation Research Board of the National Academies, n. 2143, p. 159–167, 2010. Citado 5 vezes nas páginas 17, 19, 20, 21 e 22.
- ZHANG, J. et al. Bicycle-sharing system analysis and trip prediction. In: *Mobile Data Management (MDM), 2016 17th IEEE International Conference on*. [S.l.]: IEEE, 2016. p. 174–179. Citado na página 17.