

# **Análise de Dados**

## **Pureco - Serviços de Limpeza**

Aline Dias Nunes - RA 141724 - ME812

Ana Alice Scalet - RA 154577 - ME712

Marília Gabriela Rocha - RA 183881 - ME712

Taynan Candido Quitério - RA 148009 - ME812

Prof. Dr. Jesus Enrique Garcia

**Relatório técnico**

**11 de Junho de 2019**

## 1. INTRODUÇÃO

O Pureco, criado em Abril de 2018 por membros da Enactus Unicamp [4], é um aplicativo de celular que atua como plataforma de contratação de serviços de limpeza em Campinas, conectando pessoas de Barão Geraldo com três colaboradoras do Campo dos Amarais, comunidade na periferia da cidade. Visa principalmente o aumento na renda e independência financeira dessas três mulheres.

O aplicativo oferece faxinas com preços de acordo com a quantidade de banheiros e quartos da casa e pelos tipos categorizados em “comum”, “pós mudança” ou “pós festa”. Ao pedir uma faxina, o usuário informa seu nome, e-mail, CPF, celular, dados da residência (endereço, quantidade de quartos e banheiros) e seleciona um dos tipos. Após, escolhe data e hora dependendo da disponibilidade das faxineiras. Dado isso, uma das mulheres é alocada para o serviço.

As informações fornecidas são armazenadas, sendo de interesse dos membros do Pureco a análise destas para um maior entendimento do perfil dos clientes e faxinas realizadas. O interesse é obter um melhor direcionamento das ações de divulgação do aplicativo e aumento na renda das colaboradoras, verificando também se há demanda suficiente para a inserção de outra mulher, sem que isso reduza a renda das atuais.

## **2. OBJETIVOS**

Os objetivos do projeto são analisar os dados das faxinas realizadas; identificar descritivamente características referentes aos dias com maiores solicitações e perfil dos clientes; realizar inferências estatísticas para encontrar se as solicitações distribuem-se igualmente ao longo dos dias da semana e para cada uma das colaboradoras e analisar a demanda do aplicativo, por meio da quantidade de períodos não ocupados por faxinas dentre os disponíveis pelas mulheres.

## **3. BANCO DE DADOS**

Foi disponibilizada pelos membros do Pureco uma planilha com dados das faxinas que ocorreram no período de 12 de Abril de 2018 à 31 de Maio de 2019, contendo 266 observações e as seguintes variáveis: data da faxina (dia-mês-ano); nome e endereço da faxineira; tipo da faxina (K = 240 minutos, 80 reais; P = 300 minutos, 100 reais; M = 330 minutos, 130 reais; G = 420 minutos, 150 reais; X = 420 min, 170 reais); se é pós festa ou pós mudança (1 se sim, 0 caso contrário. Ser pós festa ou pós mudança acrescenta 60 reais e 120 minutos nos tipos de faxinas); nome e endereço do cliente.

Além destes, foram adicionados à planilha dados extraídos do GDE [5] (Plataforma de planejamento de grades curriculares para alunos da Unicamp), tais como: idade dos clientes; se é estudante da Unicamp ou

não; instituto ao qual o cliente pertence; se foi/é membro da Enactus Unicamp, além de aparições (1 se no dia o aplicativo apareceu nas mídias - jornal, televisão, sites e divulgações - e 0 caso contrário). Por segurança, os nomes dos clientes e colaboradoras foram anonimizados.

#### 4. METODOLOGIA

Nas análises inferenciais, foi utilizado Teste Qui-Quadrado de Pearson de Aderência [1] para verificar se as faxinas são igualmente distribuídas ao longo da semana. Considerando frequências multinomiais para as componentes da Tabela de Contingência, a hipótese nula especifica uma estrutura de mesma probabilidade de ocorrência para cada uma das categorias, sendo que a soma das probabilidades deve ser igual a 1. O primeiro passo é testar se o modelo dado na hipótese nula se ajusta aos dados. Um Teste Qui-Quadrado de Pearson de Aderência busca determinar se existe uma discrepância entre as frequências observadas e as esperadas, sob a hipótese nula. A medida para a discrepância é:

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - np_{i0})^2}{np_{i0}} = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

onde  $n_i$  é a frequência para a categoria  $i$  e  $p_{i0}$  a probabilidade de ocorrência, sob a hipótese nula, para a categoria  $i$ .  $O_i$  e  $E_i$  simbolizam a frequência observada e a respectiva frequência esperada e o valor de  $\chi^2$

pode ser encontrado com um quantil  $\alpha$  da distribuição  $\chi^2$  com  $k - 1$  graus de liberdade [6]. Como o teste é assintótico, como regra geral é adequado utilizá-lo para  $n$  grande, de modo que cada componente tenha frequência esperada de pelo menos 5.

Para a análise dos dados temporais, tentou-se tanto modelagem da quantidade de faxinas quanto dos períodos ocupados por faxinas e dos períodos que não foram ocupados por faxinas. Analisando-se os resultados para cada um dos casos, optou-se por modelar a quantidade de períodos não ocupados, ou seja, retirou-se os períodos ocupados por faxinas solicitadas dos períodos disponíveis das mulheres.

Assim, ajustou-se aos dados de períodos não ocupados por faxinas um modelo ARIMA [2][8]. Seja  $d$  um inteiro não negativo, então  $\{y_t\}$  é um processo ARIMA( $p, d, q$ ) se  $x_t = (1 - B)^d y_t$  é um processo ARMA causal. Da definição,  $\{y_t\}$  é ARIMA( $p, d, q$ ) se satisfaz a equação em diferenças,

$$\varphi(B)(1 - B)^d y_t = \theta(B)\varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2)$$

onde  $\varphi(B)$  e  $\theta(B)$  são os polinômios autoregressivos (de grau  $p$ ) e de média móvel (de grau  $q$ ) respectivamente e  $d$  é conhecido como grau de diferenciação da série. Uma série é diferenciada em busca de torná-la estacionária para quando há tendência estocástica. Para satisfazer a

condição de causalidade, as raízes de  $\varphi(z) = 0$  deverão ser maiores do que 1 em módulo [2].

## **5. ANÁLISES E MODELAGEM**

É importante ressaltar que, além das análises apresentadas, o grupo preocupou-se em criar uma ferramenta que pode ser atualizada com novos dados e realiza as principais análises descritivas e temporais, para que o Pureco acompanhe com frequência seu desempenho. Tal ferramenta foi feita pensando em tornar o conhecimento acessível para as colaboradoras, portanto, os gráficos são de fácil interpretação e coloridos, para compreensão também das colaboradoras não alfabetizadas.

### **5.1. ANÁLISE DESCRITIVA**

As análises descritivas são feitas a partir dos dados apenas dos clientes que pediram faxinas. A Figura 5.1.1 representa a quantidade de faxinas realizadas por tipo, sendo que os tipos seguidos por “Pos” indicam que estas são “pós festa” ou “pós mudança”.

Observa-se que o tipo mais pedido é o G, seguido por M e K. Pela descrição dos tipos, dois dos três mais pedidos são os de maior duração. Como a duração é medida pela quantidade de quartos e banheiros da residência, essa análise indica a maior ocorrência de faxinas em grandes casas. Sabendo-se que Barão Geraldo é um distrito universitário, é possível que estas sejam repúblicas, moradias comuns de universitários.

Um fato interessante é a pouca quantidade de faxinas pós festa ou pós mudança, sendo que é frequente festas universitárias nesse local, podendo ser um tipo de faxina a ser explorada nas divulgações do aplicativo.

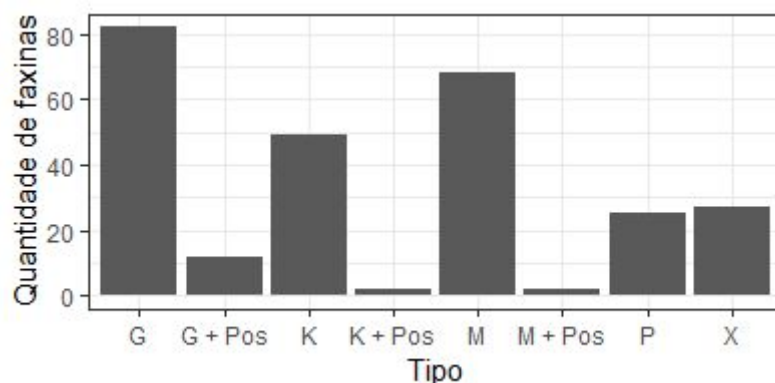


Figura 5.1.1 : Quantidade de faxinas por tipo de faxina.

Um ponto que influencia na quantidade de faxinas realizadas por mulher é a disponibilidade de cada uma (Figura 5.1.2). Segundo o Pureco, uma mulher, mesmo tendo o dia inteiro livre, realiza apenas uma faxina diária se esta for dos tipos X, G, M e P e duas faxinas caso ambas sejam do tipo K. Portanto, a disponibilidade é calculada em períodos (matutino e vespertino) e cada mulher tem somente dois disponíveis diariamente.

O Pureco conta com três colaboradoras, sendo que destas, apenas as Colaboradoras A e C iniciaram seus trabalhos juntamente com o aplicativo, em Abril de 2018. A Colaboradora B entrou para o projeto

apenas em Setembro de 2018, fato que explica a menor quantidade de faxinas, como apresentado na (Figura 5.1.3).

Da Figura 5.1.2, vê-se que a Colaboradora A possui a maior disponibilidade, o que tem impacto direto na grande quantidade de faxinas e maior renda (Figuras 5.1.4 e 5.1.5). Além disso, a Colaboradora B diminui sua disponibilidade ao longo dos meses, o que impactou negativamente sua renda.

Um ponto relevante é a maior quantidade de faxinas nos sábados ser da Colaboradora B, visto pelas Figuras 5.1.3 e 5.1.4. De acordo com o Pureco, esta possui um cliente fixo nesse dia da semana, o que gera boa parte de sua renda mensal. Ainda pela Figura 5.1.3, conclui-se que domingo é o dia menos procurado pelos clientes.

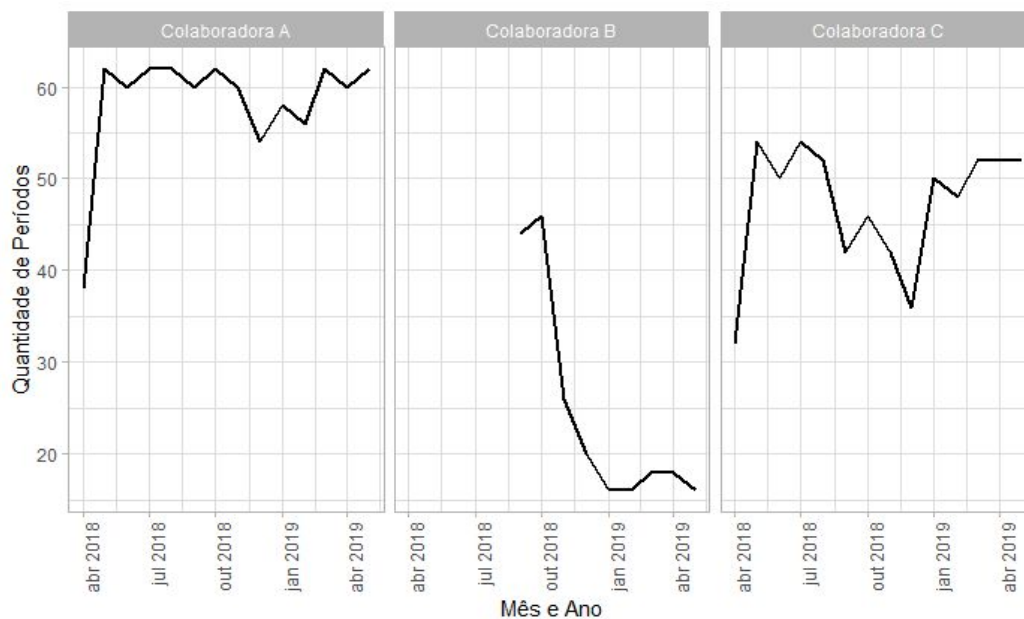


Figura 5.1.2: Disponibilidade das mulheres, dada em períodos



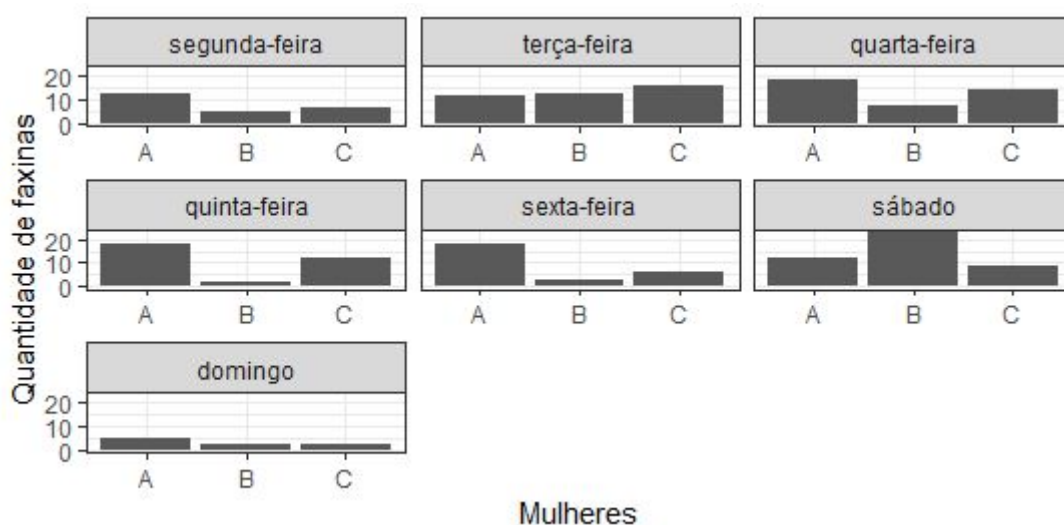


Figura 5.1.3: Quantidade de faxinas por dia da semana em relação às colaboradoras.

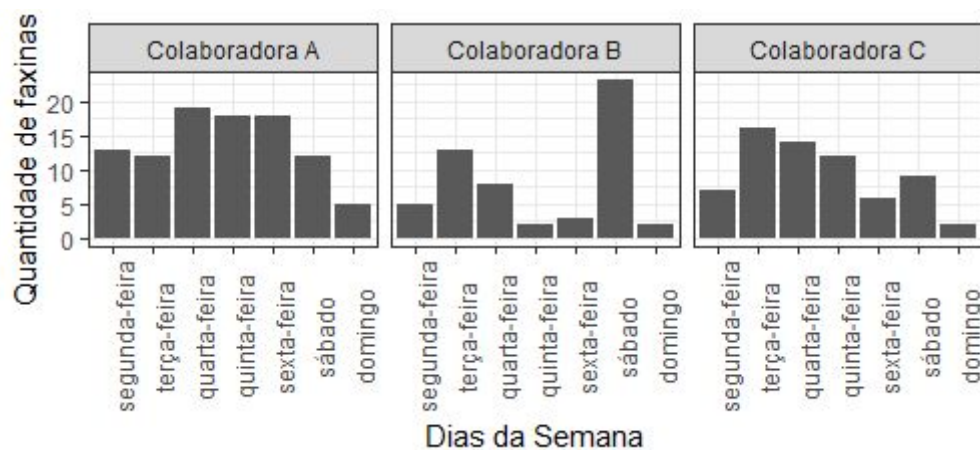


Figura 5.1.4: Quantidade de faxinas por mulher em relação aos dias da semana.

Analisando a Figura 5.1.5, vê-se que um ponto positivo para a Colaboradora B é o fato de sua renda, até Fevereiro de 2019 acompanhar a das outras mulheres, mesmo tendo iniciado os trabalhos meses após a criação do aplicativo. Após esse mês, sua renda diminuiu, mas isso pode ter relação com sua pouca disponibilidade, como mostra a Figura 5.1.2.

Apesar da renda das Colaboradoras A e C ter aumentado mesmo com a entrada da Colaboradora B, não podem ser feitas afirmações de que a entrada da Colaboradora B não prejudicou a renda das Colaboradoras A e C, já que sua disponibilidade não era semelhante às das outras e os pedidos de faxinas aumentaram ao longo dos meses (Figura 5.1.6).

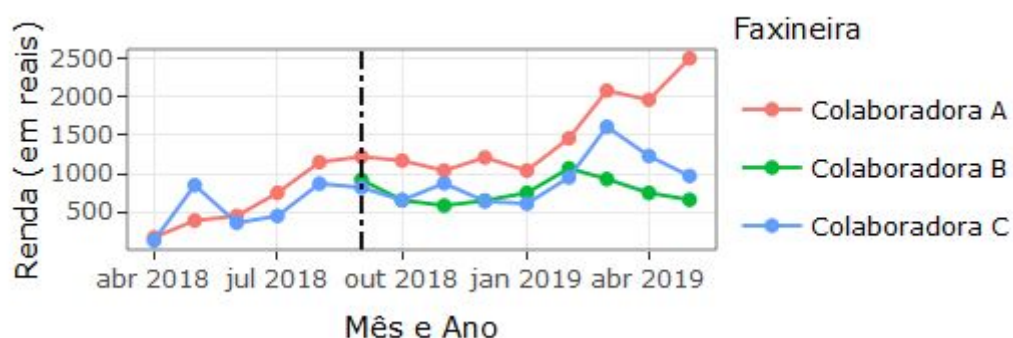


Figura 5.1.5: Renda mensal de cada mulher.



Figura 5.1.6: Quantidade de faxinas por mês.

Pelo local de atuação do Pureco ser apenas Barão Geraldo - Campinas, é esperado que o público seja majoritariamente constituído por alunos da Unicamp, o que pode ser observado na porcentagem de

clientes que possuem vínculo com a Universidade, sendo maior do que 67%.

Pelas faixas etárias dos clientes que pediram faxinas e que possuem informação de sua idade pelo GDE [5], a maioria possui entre 18 e 25 anos. Os dados são bem equilibrados em relação ao sexo dos clientes, sendo 34 mulheres e 32 homens.

A Figura 5.1.7 mostra o número de clientes novos no aplicativo desde seu início até Maio de 2019. Clientes novos, neste caso, denotam os clientes que pediram faxina pela primeira vez no mês indicado. Observa-se que o número de clientes novos aumenta de Abril de 2018 a Setembro de 2018, apesar de não possuir nenhum em Junho de 2018. Além disso, houve uma queda no número a partir de Outubro de 2018.

É importante ressaltar que nos meses de Maio e Junho de 2018, o Pureco fez divulgações nos jornais da Unicamp, Metro, RAC e PUC e no Site G1, aparições na rede Band de televisão e na rádio G1. Tais divulgações podem ter relação com o aumento no número de clientes novos em Julho e Agosto de 2018.

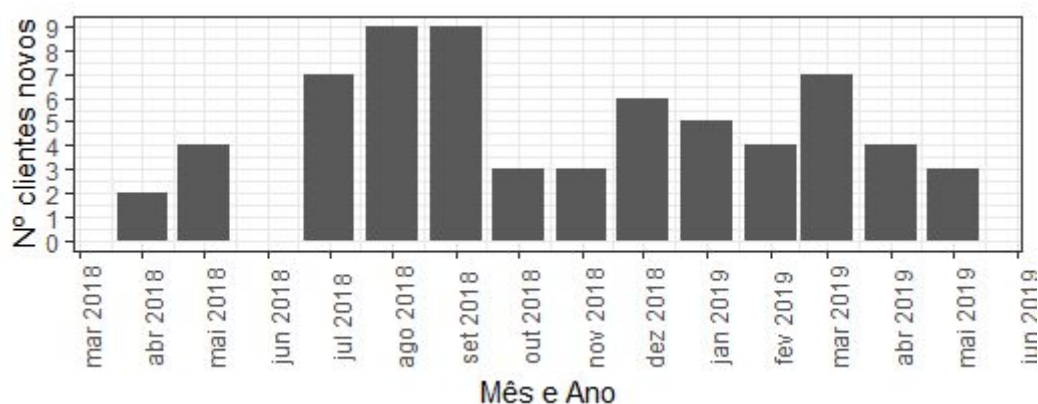


Figura 5.1.7: Clientes novos por mês.

Observando-se a Figura 5.1.8, é possível perceber que a maior parte dos clientes não se fidelizam ao aplicativo, pedindo uma única faxina. Portanto, é interessante que o Pureco pense em mais estratégias para reter clientes.



Figura 5.1.8: Quantidade de clientes por quantidade de faxinas.

## 5.2.INFERÊNCIAS

O Teste Qui-Quadrado de Pearson de Aderência [1] foi utilizado para verificar se as faxinas são igualmente distribuídas ao longo da semana. Para as frequências de faxinas realizadas, como mostrado na Tabela 5.2.1, as hipóteses consideradas são:

$H_0 : p_i = p_j = \frac{1}{7}$ , para todo  $i, j = 1, \dots, 7$  x  $H_1 : p_i \neq p_j$  para algum  $i, j = 1, \dots, 7$

ao nível de significância  $\alpha = 0.05$ , sendo  $p_i$  e  $p_j$  as probabilidades das faxinas ocorrerem nos dias da semana (domingo = 1 a sábado = 7).

Tabela 5.2.1: Frequência de faxinas realizadas para os dias da semana.

Dias	Freq. Observada
segunda-feira	31
terça-feira	44
quarta-feira	51
quinta-feira	43
sexta-feira	36
sábado	50
domingo	11

O *p-valor* obtido do teste foi  $3,269.10^{-5}$ , portanto tem-se evidências para rejeitar a hipótese de que as faxinas distribuem-se igualmente para os dias da semana.

Observando a Figura 5.1.3 e com a informação dada pelos membros do Pureco de que domingo é um dia que difere muito dos demais em relação às solicitações de faxinas, foi testado se a quantidade de faxinas distribui-se igualmente de segunda-feira à sábado. Assim, com:

$H_0 : p_i = p_j = \frac{1}{6}$ , para todos  $i, j = 1, \dots, 6$  x  $H_1 : p_i \neq p_j$ , para algum  $i, j = 1, \dots, 6$

e nível de significância  $\alpha = 0.05$ , sendo  $p_i$  e  $p_j$  as probabilidades das faxinas ocorrerem nos dias da semana (segunda = 1 a sábado = 6), obteve-se um *p-valor* de 0,207. Portanto, de segunda-feira a sábado não

há evidências para rejeitar a hipótese de igualdade de faxinas entre os dias da semana.

Nos testes acima é importante ressaltar que como varia a disponibilidade das colaboradoras, não se pode considerar que está sendo analisado unicamente a demanda de faxinas. Além disso, não há informações sobre as faxinas que foram solicitadas e não havia disponibilidade de faxineiras para realizá-las. Portanto, o teste perde um pouco de sua interpretabilidade para a questão da demanda para os dias da semana. Uma sugestão é que o aplicativo comece a armazenar também as datas das faxinas que os clientes tentaram solicitar mas não havia faxineira disponível para o dia.

Apesar do aplicativo possuir um mecanismo para escolha das faxineiras de acordo com as últimas faxinas realizadas e quantidade realizada por cada mulher, os clientes podem solicitar por uma mulher em específico pela aba de conversa com o administrador.

Assim, uma questão de interesse dos membros do projeto Pureco é se os clientes possuem preferência por uma faxineira. Para responder a essa questão, pensou-se em realizar um Teste Cochran-Mantel-Haenszel [1] de homogeneidade para a quantidade de faxinas realizadas por cada mulher em cada dia da semana.

Porém, como a disponibilidade é diferente entre as mulheres, não teria interpretabilidade analisar a quantidade de faxinas realizadas por cada mulher em cada dia da semana e desconsiderar que essa quantidade está relacionada com a disponibilidade. Pensou-se, então, na realização do teste apenas para os dias da semana em que as mulheres tinham igual disponibilidade, porém isso varia mensalmente e a realização de um teste para cada mês de funcionamento do aplicativo não seria efetivo para responder a questão de preferência dos clientes, principalmente porque é ao longo dos meses que os clientes vão se fidelizando e podem preferir ou não uma faxineira.

### **5.3. ANÁLISE TEMPORAL**

Para a análise temporal, foram utilizados os dados das faxinas, além da disponibilidade das mulheres. Como o projeto é recente, enfrenta-se o problema da pouca quantidade de observações, muitos valores nulos, além da não detecção de componente sazonal na série. Decidiu-se, portanto, agregar a série para contornar esse problema.

Da decisão do nível de agregação, leva-se em conta a interpretação e utilidade por parte do Pureco das previsões feitas. Como são poucos dados, as previsões não possuem muita confiança, quando feitas para muitos passos a frente. Logo, dados semanais são os mais adequados.

Pensando em encontrar uma resposta para o objetivo do Pureco de estimar um bom momento para a contratação de uma nova colaboradora, a série foi feita considerando prever os períodos disponíveis das mulheres não ocupados por faxinas. Ou seja, quanto mais este for próximo de 0, maior está sendo a demanda para a disponibilidade das mulheres e poderia ser um indicativo de qual período ideal para inclusão de uma nova colaboradora. Na Figura 5.3.1, são mostrados estes dados já agregados por semana.

Visualmente pela Figura 5.3.1, os dados possuem leve tendência decrescente, além de pontos distantes uns dos outros, mostrando que a série não é estacionária. Dado isso, esta é diferenciada e obtêm-se as autocorrelações e autocorrelações parciais mostradas na Figura 5.3.2.

Da Figura 5.3.2, a função de autocorrelação (FAC) possui apenas a primeira defasagem estatisticamente diferente de 0. Ainda, pela função de autocorrelação parcial (FACP), as primeiras defasagens são significativas.

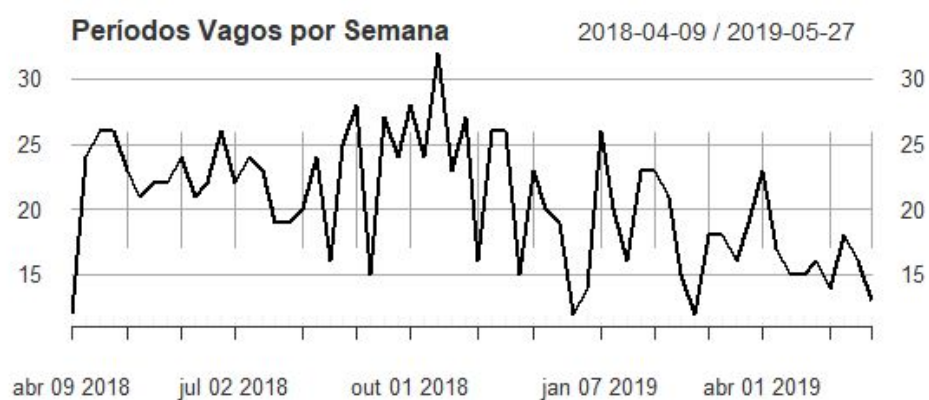


Figura 5.3.1: Série temporal do horário vago das mulheres.



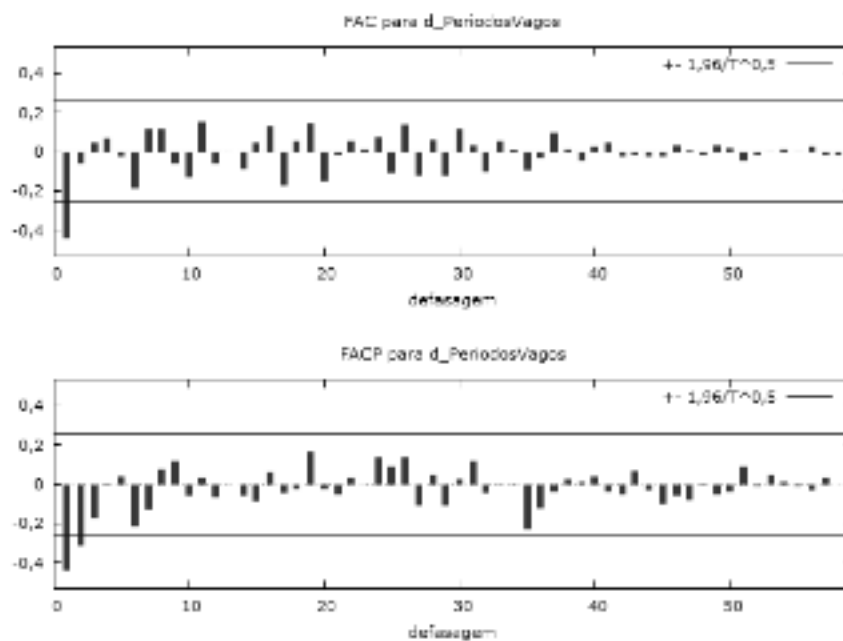


Figura 5.3.2: FAC e FACP das primeiras diferenças dos períodos vagos das mulheres.

Assim, pode se ajustar os dados por um ARIMA(0,0,1) após a diferenciação da série. Portanto, um ARIMA(0,1,1) é razoável para o ajuste da série da Figura 5.3.1.

Pelo teste de *Shapiro-Wilk* [7], tem-se que, com hipótese nula de normalidade dos resíduos, o *p-valor* resultante é de 0.7243, portanto não há evidências para rejeitar que os resíduos seguem distribuição normal.

Pela Figura 5.3.3, o teste *Box-Ljung* [3] para independência dos resíduos resulta em p-valores maiores do que 0.4, portanto não se pode rejeitar a hipótese de independência. As FAC e FACP dos resíduos possuem todas as defasagens estatisticamente iguais a zero, portanto não são correlacionados. Logo, podem ser feitas previsões com o ajuste da Figura 5.3.4.

As previsões para Maio de 2019, vistas na Figura 5.3.5, foram geradas como forma de comparar os valores preditos com os observados. A partir da Semana 60 (27-05-2019), são vistas as previsões para o mês de julho de 2019.

Por estas, não se pode ter uma resposta concreta sobre o momento de contratação, porém de acordo com a Figura 5.3.5, percebe-se que valores possíveis da quantidade de períodos vagos das mulheres estão entre 25 e 5. Um momento adequado (mais conservador) é quando o valor 0 se encontra dentro do intervalo de confiança das previsões.

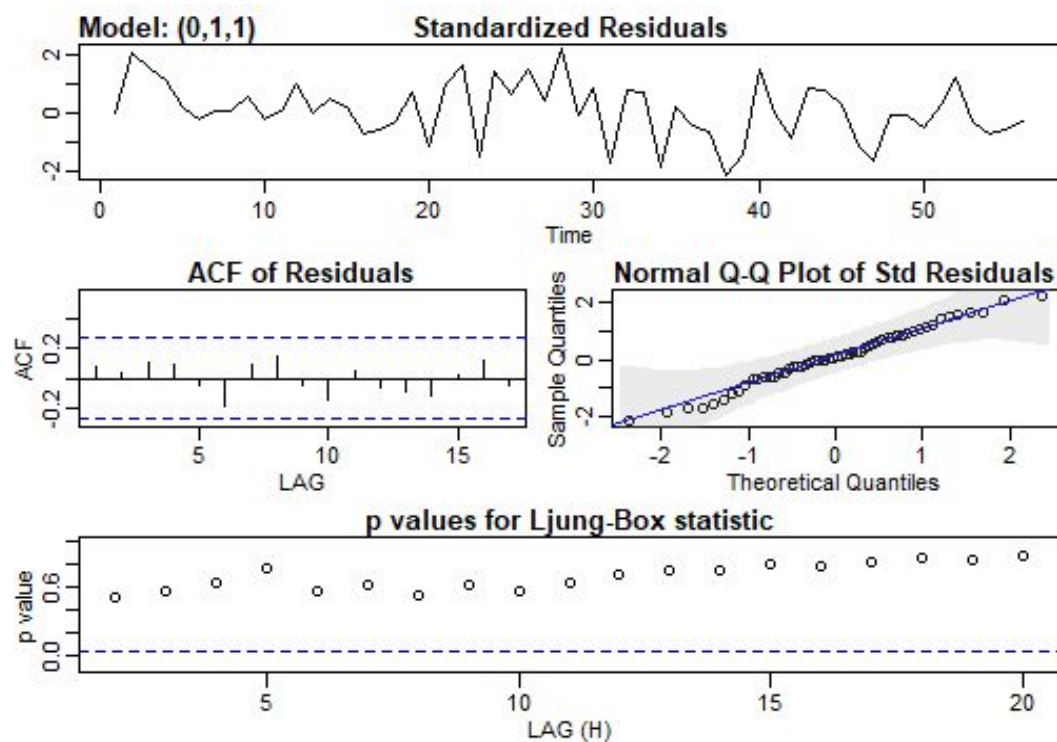


Figura 5.3.3: Teste para suposição dos resíduos ARIMA(0,1,1) dos períodos vagos das mulheres.

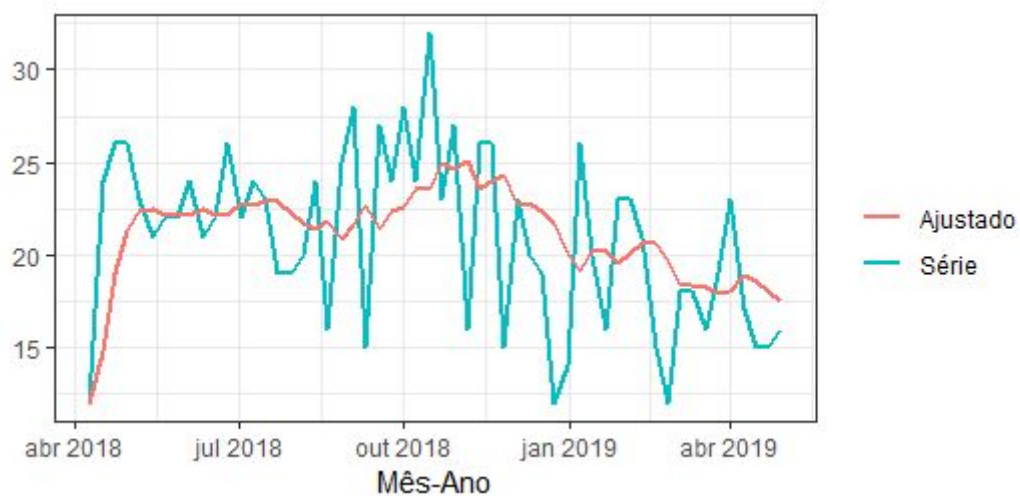


Figura 5.3.4: Ajuste com modelo ARIMA(0,1,1) dos períodos vagos das mulheres.

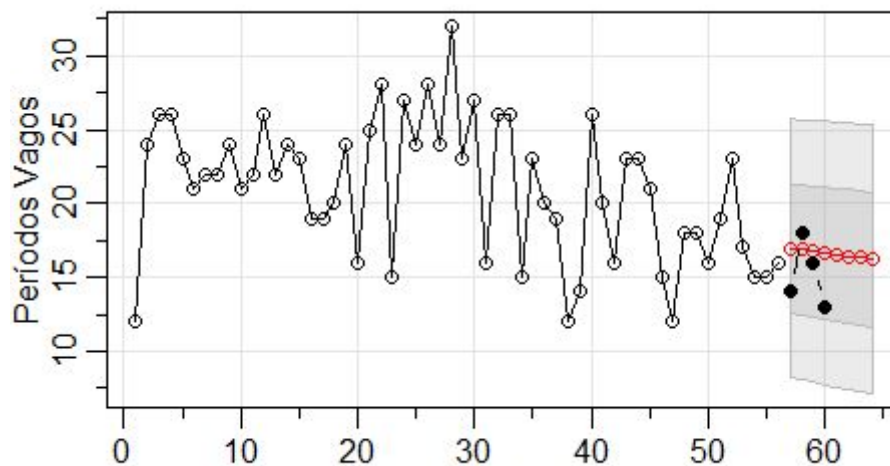


Figura 5.3.5: Previsões ARIMA(0,1,1) dos períodos vagos das mulheres para Maio e Junho de 2019, onde os valores preditos estão em vermelho e os observados em preto.

## 6. CONCLUSÃO

Observando todas as informações descritivas apresentadas, em conjunto com a série temporal dos períodos de disponibilidade das mulheres não ocupados por faxinas, é possível notar que o Pureco está tendo um aumento no número de faxinas agendadas. No entanto, o

principal público atualmente são os clientes que já se fidelizaram (pediram mais de uma faxina), tendo diminuição de clientes novos. Além disso, é grande a quantidade de clientes que usaram o aplicativo uma única vez. Deve ser uma estratégia de divulgação tanto atrair novos clientes quanto fidelizá-los após a primeira faxina.

Com os dados disponibilizados, não foi possível encontrar um momento ideal para inserir uma nova colaboradora, sendo necessário dados referentes a um período maior de uso do aplicativo, além dos dados dos pedidos que não foram agendados devido à falta de disponibilidade de mulheres no dia solicitado. Entretanto, sugere-se, com o modelo  $ARIMA(0,1,1)$  ajustado aos dados de períodos não ocupados de faxinas, que seja inserido uma nova colaboradora quando as previsões para esses dados estejam próximas de zero, ou seja, não estão sobrando disponibilidade das mulheres para realização das faxinas.

No Apêndice A, são dispostas algumas sugestões ao aplicativo.

## REFERÊNCIAS

- [1] AGRESTI, A. Categorical data analysis. john wiley & sons. Inc., Publication, 2002.
- [2] BROCKWELL,P.J.;DAVIS,R.A.;CALDER,M.V. Introduction to time series and forecasting. [S.l.]: Springer, 2002. v.
- [3] BURNS, PATRICK. Robustness of the Ljung-Box test and its rank equivalent, 2002.
- [4] ENACTUS UNICAMP. Informações Pureco. Disponível em: <<https://www.enactusunicamp.org/pureco-1>> Acesso: 15 de abril de 2019.
- [5] GDE. Disponível em: <<https://grade.daconline.unicamp.br>>. Acesso: 07 de maio de 2019.
- [6] PINHEIRO, H.P.; ALVES, R.; VARGAS, E. *Notas de Aula: Tabelas de Contingência*. 2013.
- [7] RAZALI, N. M.; WAH, Y. B. et al. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. Journal of statistical modeling and analytics, v. 2, n. 1, p. 21–33, 2011.
- [8] SHUMWAY, R. H.; STOFFER, D. S. Time Series Analysis and Its Applications: With R Examples. 3rd ed. New York: Springer, 2011.

## **Apêndice A**

### **Sugestões para o aplicativo Pureco - Serviços de Limpeza**

- Dados cadastrais:
  - Nome;
  - Idade;
  - É aluno(a) da Unicamp?
  - Qual seu curso?
  - Como conheceu o aplicativo? (Colocar opções, exemplo:  
Facebook, recomendação de amigo(a), etc)
- Dados para solicitação de faxina:
  - Quantidade de banheiros;
  - Quantidade de quartos;
  - Mora em República?
  - Endereço e ponto de referência;
  - Tem preferência por alguma faxineira?
- Se sim:
  - Qual?
  - Mostrar disponibilidade apenas dessa colaboradora e dar a opção de selecionar outro dia, estando ciente que será com outra faxineira.
  - Lista com produtos que a faxineira precisa para realização da faxina;

- Opções de forma de pagamento que não seja em dinheiro físico.
- Após a faxina:
  - Ter uma opção para a colaboradora informar que a faxina foi finalizada e, assim que ela selecionar essa opção, disponibilizar um *feedback* a ser preenchido tanto pelo cliente quanto pela faxineira. O *feedback* deve ser obrigatório caso cliente deseje pedir outra faxina pelo aplicativo. Evitar perguntas dissertativas no *feedback*.
- Dados importantes de serem armazenados pelo aplicativo:
  - Dias em que o cliente tentou solicitar por faxina mas não havia mulher disponível;
  - Faxinas canceladas (disponibilizar ao clientes opções para selecionar o motivo do cancelamento);
  - Disponibilidade semanal das colaboradoras;
  - Clientes que tentaram pedir faxina mas não eram da região de Barão Geraldo.

#### **Sugestões de divulgações:**

Considerando que pelas análises descritivas foi observado que o aplicativo não está tendo adesão de novos clientes e que mais de 50% pediram somente uma faxina, uma sugestão seria investir em cartões fidelidade ou promoções que incentivassem os clientes fidelizados a

trazerem um novo cliente para o aplicativo, de modo que este entraria já com um bom *feedback* do aplicativo dado à ele pelo cliente fidelizado.

Além disso, como a maior parte do público do Pureco são alunos da Unicamp e as faxinas pós festa são pouco solicitadas, poderia ser investido em divulgações para faxinas pós festas em repúblicas de Barão Geraldo, buscando, por exemplo, parcerias com Atléticas e Centros Acadêmicos.