



LISBON
SCHOOL OF
ECONOMICS &
MANAGEMENT
UNIVERSIDADE DE LISBOA

MESTRADO

ECONOMETRIA APLICADA E PREVISÃO

TRABALHO FINAL DE MESTRADO

DISSERTAÇÃO

Previsão de Séries Temporais com Períodos
Sazonais Complexos: Uma Aplicação Empírica

ANA CAROLINA SANTOS FERREIRA

OUTUBRO 2018

MESTRADO
ECONOMETRIA APLICADA E PREVISÃO

TRABALHO FINAL DE MESTRADO
DISSERTAÇÃO

Previsão de Séries Temporais com Períodos
Sazonais Complexos: Uma Aplicação Empírica

ANA CAROLINA SANTOS FERREIRA

ORIENTAÇÃO:
PROFESSOR DOUTOR NUNO SOBREIRA

COORIENTAÇÃO:
PROFESSOR DOUTOR MANUEL ESQUÍVEL

OUTUBRO 2018

Previsão de Séries Temporais com Períodos Sazonais Complexos: Uma Aplicação Empírica

Ana Carolina Santos Ferreira*

Orientação: Prof. Dr. Nuno Ricardo Martins Sobreira[†]

Coorientação: Prof: Dr. Manuel Leote Tavares Inglês Esquível[‡]

Resumo

O *call center* é, hoje em dia, um dos fatores estratégicos para o crescimento de uma empresa. Sendo a satisfação do cliente uma prioridade e o *call center* um dos primeiros pontos de contacto com o mesmo, garantir um bom serviço poderá melhorar a relação e gerar oportunidades de crescimento.

Cada contacto com o cliente traz mais conhecimento e este conhecimento é precioso para gerar aproximação e comunicação mais assertiva. Conhecer as suas características e os motivos que o fazem ligar são pontos-chave para antecipar as suas necessidades. Assim, um *call center* bem estruturado e preparado para atender de forma rápida mostra eficiência e preocupação da empresa.

A aplicação empírica mostra análises capazes de traçar o perfil do cliente ao nível sociodemográfico e ao nível de segmentações do negócio. Mostra também a interpretação de todas as interações entre motivos de chamada, capazes de criar *inputs* para a tomada de decisão ao nível operacional. Corrobora-se a presença de padrões sazonais através da análise espectral, que foi tida em conta na modelação da série. Conclui-se que o melhor método de previsão é o ARIMA cujo padrão sazonal é explicado por termos de uma série de Fourier.

Palavras-chave: Análise descritiva, Análise espectral, Diagrama de chord, ARIMA, Termos de uma série de Fourier, Dupla sazonalidade, HW, Previsão, Número de Chamadas, Call center.

*Contacto: carolina.s.ferreira@outlook.com

[†]ISEG Lisbon School of Economics & Management-Departamento de Matemática e CEMAPRE.

Contacto: nsobreira@iseg.ulisboa.pt

[‡]Faculdade de Ciências e Tecnologia - UNL - Departamento de Matemática e CEMA.

Contacto: mle@fct.unl.pt

Forecasting Time Series with Complex Seasonal: An Empirical Application

Ana Carolina Santos Ferreira^{*}

Advisor: Prof. Dr. Nuno Ricardo Martins Sobreira[‡]

Co-Advisor: Prof. Dr. Manuel Leote Tavares Inglês Esquível[¶]

Abstract

Call center are nowadays one of the strategic factors for the growth of a company. As customer satisfaction is priority, and since call centers are one of the first contact points with a company's clients, ensuring a good service can improve customer-relationship and generate growth opportunities.

Each contact with a client brings additional knowledge, and that knowledge is precious to generate proximity and increase communication assertivity. Knowing their characteristics and the motivations behind their call are key factors in the anticipation of their needs. Thus, a well-structured call center which is ready for quick call handling shows efficiency and concern from the company.

The empirical application introduces analytical methods capable of tracing the client profile at the socio-demographic level and at the business segmentation level. It also presents the interpretations of all interactions between call reasons capable of creating inputs for decision-making at the operational level. Spectral analysis validates the presence of seasonal patterns, which we controlled for when modelling the series. It is concluded that the best prediction method is the ARIMA model whose seasonal pattern is explained by terms of a Fourier series.

Keywords: Descriptive analytics, spectral analysis, Chord diagram, ARIMA, Terms of a Fourier series, double seasonality, HW, Forecasting, Number of calls, Call center.

^{*}Contact: carolina.s.ferreira@outlook.com

[‡]ISEG Lisbon School of Economics & Management-Departamento de Matemática and CEMAPRE.
Contact: nsobreira@iseg.ulisboa.pt

[¶]Faculdade de Ciências e Tecnologia - UNL - Departamento de Matemática and CEMA.
Contact: mle@fct.unl.pt

Dedicado aos meus amados Pais

Agradecimentos

Aos Professores Doutores Nuno Sobreira e Manuel Esquível, por terem aceitado orientar este estudo e partilhar o seu conhecimento. À minha manager, Vera Santos, por todo o apoio, disponibilidade e profissionalismo que tornaram possíveis a realização e conclusão desta dissertação. À minha família pelo apoio e paciência, aos meus amigos pelo companheirismo, ajuda e motivação. Ao meu namorado, Pedro Cardia, pelo incentivo diário e por todo o carinho partilhado em todos os momentos.

Índice

Resumo	i
Abstract	ii
Agradecimentos	iv
Lista de Figuras	vii
Lista de Tabelas	vii
Lista de Acrónimos	viii
1. Introdução	1
1.1 Motivação e Objetivos	1
1.2 Apresentação do Estudo e Contribuições	2
1.3 Apresentação da Empresa	3
1.4 Estrutura da Dissertação	4
2. Revisão de Literatura	5
2.1 Call Center: Breve História e Definição	5
2.2 Estudos sobre previsão das chamadas recolhidas num call center.....	6
3. Dados e Metodologia	10
3.1 Os Dados	10
3.1.1 Levantamento de Dados	10
3.1.2 Análise e Tratamento de Variáveis	11
3.1.3 Análise Exploratória das Variáveis	12
3.1.3.1 Variáveis Quantitativas	12
3.1.3.2 Variáveis Qualitativas	13
3.2 Metodologia	22
3.2.1 Métodos de Alisamento Exponencial	23
3.2.2 Modelos ARIMA e SARIMA	24
3.2.3 Modelos Sazonais Trigonométricos	25
3.2.3.1 Análise Espectral	26
4. Análise e Discussão dos Resultados	28

5. Conclusão	34
5.1 Contributos	36
5.2 Limitações	36
5.3 Investigações Futuras	36
6. Referências Bibliográficas	37
Anexos	40
A Análise e Tratamento de Variáveis	
A.1 Variáveis Quantitativas: Winsorização	41
A.2 Variáveis Qualitativas: Frequências	42
A.3 Correspondência Motivo Geral e Motivo Particular	43
A.4 Correlação entre Motivos e cada umas das Variáveis Explicativas	46
B Análise e discussão dos Resultados	
B.1 Análise Espectral	47
B.2 Boxplot: Análise Sazonal Semanal e Mensal.....	50
B.3 Comparação de Várias Regressões Harmónicas Dinâmicas com	
Períodos Sazonais Múltiplos	51
B.4 Valores Reais e Previstos pelo Método HW e pelo Método ARIMA	
para a Base de Validação	52

Lista de Figuras

1 - Matriz de correlações entre os 8 motivos	16
2 - Diagrama de chord - análise global	19
3 - Diagrama de chord - análise por ano	21
4 - Número de chamadas recebidas no <i>call center</i> , FAC e FACP	29
5 - Decomposição trigonométrica da série do número de chamadas recebidas no <i>call center</i>	30
6 - Diagnóstico aos resíduos da regressão harmónica dinâmica	33
7 - Comparação dos métodos aplicados para o horizonte de previsão	35

Lista de Tabelas

1 - Número total de chamadas efetuadas por ano	10
2 - Estatísticas descritivas das variáveis quantitativas	12
3 - Frequência por motivo de chamada	14
4 - Motivos das chamadas - Estatísticas descritivas	15
5 - Análise global vs análise ao longo da amostra: Peso de cada motivo sobre o total dos registos (em %)	22
6 - Estimativas do modelo HW	31
7 - Estimativas do modelo ARIMA(0, 1, 3) com $K = c(3, 3)$ termos de Fourier	32
8 - Valores dos MSE oriundos dos métodos HW e ARIMA	32

Lista de Acrónimos

AIC - Akaike Information Criterion

AICc - Akaike Information Criterion Adjusted

ARIMA - Autoregressive Integrated Moving Average

FAC - Função de Autocorrelação

FACP - Função de Autocorrelação Parcial

HW - Holt-Winters

IQR - Intervalo Inter-Quartis

IVR - Interactive Voice Response

MAE - Mean Absolute Error

MAPE - Mean Absolute Percent Error

MCMC - Monte Carlo via Cadeias de Markov

MSE - Mean Square Error

SARIMA - Seasonal Autoregressive Integrated Moving Average

SQE - Soma dos Quadrados Explicada

1. Introdução

1.1 Motivação e Objetivos

Os gestores do *call center* têm como desafio manter uma boa relação entre os níveis de serviço elevados e os baixos custos de operação, garantindo assim o sucesso financeiro do negócio; por conseguinte, para equilibrar estas 2 vertentes torna-se muito importante obter uma boa previsão do fluxo de chamadas. Este tipo de dados, tais como outros que podem ser observados nos estudos de Hyndman *et al.* (2011), apresentam padrões sazonais complexos difíceis de serem tratados. Ao longo do tempo, vários modelos de séries temporais foram desenvolvidos, mas muitos deles apenas são eficazes para tratar padrões sazonais simples.

Neste estudo, foi então usada uma metodologia capaz de incorporar padrões sazonais mais complexos: recorrendo aos dados das chamadas efetuadas para o *call center* do Oney, demonstrou-se a versatilidade da abordagem proposta para a decomposição da série e previsão da mesma. A aplicação empírica que vai ser apresentada ao longo deste estudo tem como grande objetivo a melhoria contínua do desempenho do *call center*, que pode ser subdividido em 3 etapas:

1. Traçar o perfil dos clientes que contactam o Oney;
2. Perceber os principais motivos de contacto do cliente com a empresa;
3. Definir o modelo de ajustamento para a variável de interesse e fazer a respetiva previsão.

Quanto ao primeiro objetivo pretende-se traçar o perfil sociodemográfico do cliente que efetua a chamada e fazer o cruzamento com segmentações da própria Instituição (perfil de utilização loja/ crédito). Características sociodemográficas como o agregado familiar, rendimento mensal, nº de filhos, idade, estado civil, região, categoria profissional, sexo, posse ou não de telemóvel e de *e-mail* são variáveis de estudo para, de certa forma, ficarmos a conhecer melhor o cliente que liga para o Oney. Através do cruzamento com as segmentações da empresa, é também possível perceber qual antiguidade do cliente, bem como a modalidade de compra preferencial nos últimos 12 meses.

Relativamente ao segundo objetivo, pretende-se agrupar/ organizar as ligações em categorias apropriadas. Mesmo com a organização definida, continuará a existir uma parte designada por “outros motivos” na base de dados que não se encaixam em nenhuma das

restantes categorias pré-determinadas.

Através destes 2 objetivos fica-se com um conhecimento bastante alargado acerca do cliente que efetua a chamada e qual o motivo que o leva a ligar. As vantagens de conhecer bem o cliente são inúmeras: oferecer produtos mais indicados; adaptar a linguagem; segmentar os clientes (ao fazê-lo, é possível identificar qual o grupo a que pertence, o que faz a comunicação tornar-se mais apropriada e melhorar assim o ritmo de interação, bem como antecipar as necessidades dos mesmos); identificar consumidores com maior valor; fidelizar cliente; entre outras.

Traçar o perfil dos clientes é importante para garantir que o comportamento inerente dos clientes que efetuam as chamadas é consistente com o comportamento da população da entidade. A análise dos motivos das chamadas permite observar distorções nos padrões de sazonalidade, sendo este fator muito importante para a construção do modelo.

Através do terceiro objetivo, ou seja, através dos resultados da previsão é então possível alocar os recursos de forma correta para atingir as metas estabelecidas com o menor custo possível.

1.2 Apresentação do Estudo e Contribuições

Para responder aos objetivos propostos, primeiramente, foi feita uma análise descritiva detalhada dos dados, recorrendo ao diagrama de chord e à análise espectral. De seguida, aplicaram-se modelos de séries temporais univariadas com dupla sazonalidade de forma a responder aos ciclos dominantes observados na análise espectral.

Através da análise do diagrama de chord foi possível retirar conclusões operacionais basilares ao negócio, como por exemplo: enviar informação com saldo devedor ao cliente após o mesmo contactar o Oney, independentemente do motivo que o leva a fazer o primeiro contacto. Esta conclusão é subsequente à evidência de que grande percentagem dos clientes, após efetuar o primeiro contacto, liga de volta a questionar sobre o montante a pagar. Ora, se for possível antecipar certos comportamentos do cliente, é provável que esse facto provoque uma diminuição no número de chamadas recebidas.

Os resultados obtidos na análise espectral realçam a presença de padrões sazonais, sendo estes fundamentais para a modelação e previsão da série. Tanto na série temporal global como em cada uma das séries dos motivos, o ciclo sazonal semanal (7 dias) e o ciclo sazonal

anual (360/365 dias) foram os que se destacaram como dominantes. Ainda assim, foi possível observar ciclos mais suaves de 3, de 2 e de 31 dias, ordenados de forma decrescente quanto à respetiva amplitude. Para a previsão, esta informação foi tida em conta e, após comparação de vários métodos, verificou-se que o modelo que produz melhores resultados é um ARIMA cujo padrão sazonal é modelado por séries de Fourier, que retratam esses ciclos dominantes.

1.3 Apresentação da Empresa

O Oney está em Portugal desde 1994, onde começou por oferecer soluções de financiamento associadas ao Cartão Jumbo¹ que, rapidamente, se tornou o maior cartão privativo do país. A história da empresa começou no Grupo Auchan, com o qual ainda se mantém relação através da Auchan Holding, com o objetivo de servir os clientes. Hoje o Oney está presente em 11 países, com mais de 8 milhões de clientes por todo o mundo.

Com a experiência ganha no decorrer do tempo, a empresa tornou-se especialista em produtos e serviços financeiros. Ampliaram as soluções de pagamento flexível a várias áreas de negócio, contribuindo para o crescimento do negócio dos parceiros e proporcionando apoios financeiros para muitos portugueses. Alargaram também a área de atuação, passando a oferecer seguros.

Atualmente o Oney tem 8 cartões (Jumbo Mais, Presente Escolha Feliz, Norauto, Leroy Merlin, AKI, MUST, Presente e OMB), 4 seguros (Proteção auto, casa e saúde; Proteção de meios de pagamento; Proteção de pessoas; Proteção de equipamentos), crédito pessoal e crédito conta corrente. Com a utilização dos cartões de crédito é possível, no ato da compra, escolher na caixa a modalidade de pagamento. Existem várias modalidades, nomeadamente: Débito Imediato - possibilidade de pagar as compras na totalidade por débito direto na conta bancária, a partir do 4º dia útil após ter realizado as mesmas; Com Juros - opções de financiamento (desde 3x até 30x com juros) de modo a escolher a modalidade que mais convém e que melhor se adapta às necessidades; Revolving - reembolso em prestações suaves que diminuem à medida que o montante em dívida diminui; Sem Juros - possibilidade de pagar em mensalidades iguais (sem juros) a partir de um determinado valor de compra.

¹Companhia Portuguesa de Hipermercados gerida pelo grupo Auchan.

1.4 Estrutura da Dissertação

A presente dissertação inicia com uma breve revisão bibliográfica, onde são introduzidos os tópicos relevantes ao estudo (capítulo 2): na primeira secção deste capítulo (2.1.) é apresentada de forma breve a história e definição dos *call centers*; na segunda secção (2.2.) são descritos vários estudos elaborados no âmbito da modelação e previsão do número de chamadas efetuadas para diversos *call centers*. O capítulo 3 é dividido em dois subcapítulos: no primeiro (3.1.) é feita uma análise descritiva e estatística dos dados e no segundo (3.2.) é descrita a metodologia a ser seguida. O capítulo 4 apresenta uma aplicação empírica dos dados descritos no subcapítulo 3.1. e emprega-se a metodologia apresentada no subcapítulo 3.2.. É também neste capítulo que são apresentados e discutidos os resultados do estudo. As conclusões e as limitações da abordagem são apresentadas no capítulo 5.

2. Revisão de Literatura

2.1. *Call Center*: Breve história e definição

Os *call centers* são unidades de prestação de serviços que incluem centros operacionais para automatizar uma quantidade volumosa de diferentes atividades e serviços a clientes, tipicamente utilizando o canal de contacto telefónico.

Os centros do tipo *inbound*, onde as chamadas são originadas pelos clientes, são caracterizados como um sistema de atendimento por vários operadores, que recebem ligações telefónicas, normalmente de clientes ou de potenciais clientes, que desejam obter alguma informação, contratar algum serviço, comprar algum produto, entre outras (referência também discutida por Grossman *et al*, 2001; Hawkins *et al*, 2001).

A indústria de *call center* tem assistido a um crescimento bastante elevado nos últimos anos, tornando-se assim necessário uma gestão de forma eficiente. Esta gestão consiste na previsão das chamadas recebidas e da consequente carga de trabalho associada, que precisam então ser estimadas de forma cuidada. Segundo vários autores, esta previsão é um pré-requisito para um nível de serviço consistente e requer que haja uma combinação entre a carga de trabalho e os recursos existentes. Um dos passos básicos para prever a carga de trabalho passa então por fazer uma previsão, o mais precisa possível, do volume de chamadas recebidas.

A existência de um bom modelo de previsão permite ajustar horários de funcionamento e recursos humanos necessários ao correto nível de performance. A previsão no *call center* deverá responder às várias motivações para a interação entre a empresa e o cliente, ou seja, além de prever o volume de chamadas, considera também o volume de *e-mails* e outro tipo de contacto que envolva os agentes.

É de notar ainda que, para além do histórico do volume de chamadas, outras características podem impactar esse número, tais como: próprio funcionamento da empresa (dias com maior frequência de chamadas), véspera de feriado, épocas festivas, entre outras. Assim, os resultados da previsão são considerados uns bons instrumentos, pois torna-se possível organizar as equipas de forma mais eficiente, estabelecendo um equilíbrio entre a elevada prestação de serviços e a redução de custos de uma operação.

2.2. Estudos sobre previsão das chamadas recebidas num *call center*

Na revisão de bibliografia realizada foi possível observar um conjunto de diferentes metodologias, nomeadamente:

Em Gans *et al.* (2003), a literatura de séries temporais para as chamadas recebidas num *call center* é bastante reduzida. No entanto, nos últimos anos foram feitos grandes avanços na modelação e previsão das chamadas recebidas num *call center*. Esses avanços podem ser observados em Steckley *et al.* (2004), Weinberg *et al.* (2007), Mehrotra *et al.* (2012), entre outros; que serão descritos mais à frente.

Um gestor de *call center*, geralmente, pode fazer dois tipos de estudos: prever o volume de chamadas dos diferentes dias e semanas seguintes e/ou, dado um determinado dia, ir atualizando as previsões com as informações disponíveis consoante as chamadas adicionais que chegam ao longo do dia. Grande parte dos trabalhos sobre este tema ignoram a dependência intra-diária das chamadas e focam-se na previsão do volume total de chamadas recebido diariamente, uma vez que os dados relevantes, que são necessários para a previsão intra-diária, nem sempre são fáceis de obter por parte das empresas. No entanto, com o decorrer do tempo, os *call centers* têm-se tornado cada vez mais sofisticados na identificação dos dados necessários para uma boa previsão e deve ter-se em consideração que a qualidade dos dados históricos tem vindo a aumentar, bem como a capacidade de armazenar essa informação.

Encontram-se poucos artigos que consideram o método de séries temporais para este tipo de aplicação, no entanto na maioria dos estudos existe um modelo em comum: *Autoregressive Integrated Moving Average* (ARIMA).

Taylor, (2008) comparou o desempenho de vários métodos de séries temporais univariadas para prever as chamadas recebidas intra-diárias. Extensão da suavização de Holt-Winters (HW) para o caso de dois ciclos sazonais, suavização exponencial robusta baseada na regressão dos mínimos dos desvios absolutos exponencialmente ponderados (ver Cipra, 1992) e a regressão harmónica dinâmica (forma do modelo de espaço de estado com a componente não observada) aplicada por Tych *et al.* (2002), são exemplos dos métodos considerados. A volumetria dos dados influencia de forma proporcional a performance do modelo.

Ao fazer uma análise do número de chamadas recebidas num *call center* de um serviço de urgência, Mabert (1985) usou vários métodos com séries cuja sazonalidade foi retirada e verificou que os resultados superaram os resultados obtidos num ARIMA simples.

Andrews & Cunningham (1995), para prever o volume de chamadas diárias no *call center* de uma empresa de retalho, incorporaram efeitos publicitários num modelo ARIMA através do uso de funções de transferência.

Bianchi *et al.* (1998) usaram também os modelos ARIMA para prever as chamadas diárias num *call center*, mais especificamente num serviço de *telemarketing*. Analisaram versões aditivas e multiplicativas do modelo de HW, incluindo a análise de *outliers*, e compararam com o modelo de Box-Jenkins. Concluíram que apesar de na literatura existir evidência de que “modelos simples” como o de HW apresentam melhores resultados que os “modelos mais complexos”, para a análise em questão os modelos ARIMA apresentam uma melhor performance para as séries temporais abordadas.

Com a facilidade de acesso aos dados das chamadas, Jongbloed & Koole (2001) verificaram que a variância era maior do que a média e, portanto, observaram que o processo de chegada segue um processo de Poisson não homogêneo. Dada a dependência intra-diária e a sobredispersão observada, propuseram um modelo de mistura de Poisson – um modelo de Poisson com uma taxa de chegada que não está fixa, verificando-se que esta é aleatória com um comportamento estocástico.

Com o intuito de estudar o número de aplicações diárias para empréstimos num *call center* de um serviço financeiro, Antipov & Meade (2002) concentraram-se em aplicações estimuladas pela publicidade, no entanto em estudos anteriores a análise passava pelos modelos ARIMA.

De modo a abordar os dados das chamadas intra-diárias, Tych *et al.* (2002) usaram o método de espaço de estados de componentes não observadas com base no modelo de regressão harmónica dinâmica e compararam com os modelos ARIMA sazonais.

Henderson (2003) assume uma taxa de chegada constante em diferentes intervalos de tempo, cujo comprimento vai “encolhendo” à medida que o volume de dados das respetivas chamadas aumenta, produzindo boas estimativas para a função da taxa de chegada.

O estudo de Avramidis *et al.* (2004) foi mais além, onde observaram que a taxa de chegada varia consideravelmente ao longo do dia, aparentando a existência de uma forte autocorrelação para o volume das chamadas em curtos períodos. Foram desenvolvidos modelos duplamente estocásticos – a variável aleatória é correlacionada entre os diferentes intervalos de tempo do mesmo dia - com quatro propriedades empiricamente observadas que definem os processos das chamadas recebidas:

Propriedade 1. A procura diária total (número de chamadas) revela uma sobredispersão em relação ao modelo de Poisson – a variância é maior do que o valor esperado. Note-se que o modelo de Poisson tem subjacente a igualdade entre a variância e o valor esperado.

Propriedade 2. A taxa de chamadas recebidas varia com a hora do dia.

Propriedade 3. Evidências de uma forte associação positiva entre períodos de tempo dentro de um determinado dia – dependência intra-diária.

Propriedade 4. Dependência significativa (correlação forte e positiva) entre o número de chamadas recebidas em dias sucessivos.

Outra conclusão observada é que o melhor modelo a ajustar também pode ser influenciado pela finalidade da chamada.

Steckley *et al.* (2004) e Mehrotra *et al.* (2012) analisam a correlação do número de chamadas recebidas em diferentes dias, com o intuito de atualizar os horários da carga de trabalho no início de cada dia. Para tentar aproximar os processos de Poisson não homogêneos, estimam a taxa de chegada dividindo o conjunto de dados em intervalos mais curtos.

Brown *et al.* (2005), ao analisarem os dados de outro *call center*, observaram também a existência de persistência dinâmica da taxa de chegada das chamadas do dia-a-dia e sugerem um processo de Poisson variável no tempo. Usaram um modelo multiplicativo de *mixed-effects* para prever a taxa de chegada de chamadas futura, onde foi incluído o volume de chamadas do dia anterior como termo autoregressivo.

O estudo de Weinberg *et al.* (2007) é uma extensão aos trabalhos de Avramidis *et al.* (2004) e de Brown *et al.* (2002) em que é considerado um modelo de Poisson não homogêneo, onde se assume que a taxa das chamadas recebidas incorpora fortes padrões intra-diários e a existência de dinâmicas do dia-a-dia para prever as taxas futuras da chegada das chamadas.

Como consequência, uma vez que a estrutura temporal desta taxa é não só caracterizada pela variação diária, como também pela intra-diária, Weinberg *et al.* (2007) modelam estes dois tipos de variação separadamente e só depois os combinam num modelo multiplicativo, através de um modelo Bayesiano multiplicativo e um algoritmo Monte Carlo via Cadeias de Markov (MCMC) – prevêem a taxa de chegada em curtos intervalos de tempo (15-60 minutos) de um dia de semana como produto da previsão de chamadas de um dia em proporção com as chamadas que chegam durante um intervalo de tempo.

Dada a escassa literatura de séries temporais para o volume de chamadas recebidas num *call center* não é surpreendente que os métodos mais simples dominem as previsões feitas

para este tema. Assim, métodos como o alisamento exponencial, os modelos ARIMA, a regressão linear e a decomposição das séries temporais são dos mais utilizados por parte das empresas e instituições.

3. Dados e Metodologia

3.1 Os Dados

Os dados foram fornecidos por uma Instituição Financeira - Oney Bank. Tratam-se de dados de chamadas *inbound*, correspondentes aos últimos quatro anos (desde janeiro de 2014 a dezembro de 2017), sendo consideradas as chamadas de segunda-feira a sábado, das 9h até às 21h. Quando a chamada chega ao *call center*, o cliente é recebido por um serviço de atendimento automático (*Interactive Voice Response*, IVR). Após receberem as informações solicitadas através do IVR alguns clientes desligam, outros avançam para a fila de serviço e aguardam até ao momento em que o próximo agente fica disponível.

3.1.1 Levantamento de Dados

A base de dados que vai ser utilizada para modelação pode ser dividida em dois grandes grupos:

- Dados da Chamada: colaborador que atende a chamada, motivo do contacto, data e hora do mesmo;
- Dados Pessoais (do cliente que efetua a chamada): agregado familiar, rendimento mensal do agregado, número de filhos, idade, estado civil, região, categoria profissional, segmento de compra, antiguidade do cliente, género, *e-mail*, telemóvel.

Ao longo dos quatro anos de análise (desde janeiro de 2014 até dezembro de 2017) foram tipificadas 1.183.883 chamadas pelos operadores de *inbound* do Oney, distribuídas da seguinte forma:

Tabela 1: Número total de chamadas efetuadas por ano

Ano	#Chamadas
2014	302.548
2015	305.577
2016	298.884
2017	276.874

Em média, verifica-se que, por mês, ocorrem cerca de 24.664 chamadas e, por dia, perto de 952 chamadas. Adicionalmente, verifica-se que as 1.183.883 chamadas foram efetuadas por um total de 320.673 contratos distintos. Ao analisarmos o universo de contratos que contactaram o Oney é possível verificar algumas diferenças no decorrer dos anos:

- Em 2014 verificou-se que o máximo de contactos foi de 77 vezes, com uma média de contactos de 2,20. 7% dos contratos (9.316 clientes) contactaram o Oney mais de 6 vezes por ano, representando 28% do total de chamadas;

- Em 2015 verificou-se que o máximo de contactos foi de 129 vezes, com uma média de contactos de 2,30. 7% dos contratos (9.987 clientes) contactaram o Oney mais de 6 vezes por ano, representando 31% do total de chamadas;

- Em 2016 verificou-se que o máximo de contactos foi de 153 vezes, com uma média de contactos de 2,29. 7% dos contratos (9.685 clientes) contactaram o Oney mais de 6 vezes por ano, representando 31% do total de chamadas;

- Em 2017 verificou-se que o máximo de contactos foi de 235 vezes, com uma média de contactos de 2,39. 8% dos contratos (9.401 clientes) contactaram o Oney mais de 6 vezes por ano, representando 34% do total de chamadas;

3.1.2 Análise e Tratamento das Variáveis

Em algumas variáveis deteta-se a presença de *outliers*² (observações extremas - observações que apresentam uma grande distância face às restantes observações da série). Para estes casos optou-se por utilizar a técnica de Winsorização das observações extremas (Winsor, 1947) e (Dixon, 1960). Esta técnica consiste na substituição de todas as observações extremas por um determinado valor.

É uma das maneiras de minimizar o impacto dos *outliers*, alterando o seu valor de modo a que fique mais próximo dos restantes. Poderá existir um *trade-off* entre eficiência e enviesamento pois, a substituição destes valores pode melhorar a eficiência estatística, mas, por outro lado, estamos a alterar os nossos dados. É de esperar que, ao aplicar esta técnica, os

²Para além de identificar os outliers é preciso tratá-los. Os procedimentos mais comuns são a eliminação dos mesmos e a Winsorização das observações extremas. A substituição dos valores dos outliers pela média da variável, analisar separadamente estes valores e utilizar métodos de clustering são outros procedimentos que podem ser utilizados.

resultados venham menos enviesados do que se os dados fossem excluídos da base; adicionalmente, se mantivéssemos os dados, estes poderiam distorcer drasticamente os resultados. Por não serem consideradas as observações extremas no cálculo do Intervalo Inter-Quartis (IQR), considera-se que esta medida de dispersão é robusta à presença deste tipo de observações. Consequentemente, nesta análise vamos substituir o valor observado pelo valor da barreira dos *outliers* severos.

O IQR é a diferença entre o terceiro e o primeiro quartis e compreende 50% das observações mais centrais da base de dados. Ou seja, são excluídas do seu cálculo 25% das observações de menor e de maior valor: $IQR=Q_3-Q_1$. Para esta análise vai ser usado um dos critérios mais utilizados, onde se considera que X_i é um *outlier* severo quando, $X_i \leq Q_1 - 3 \times IQR$ ou $X_i \geq Q_3 + 3 \times IQR$, o valor se afasta do primeiro ou terceiro quartis por mais de três vezes do intervalo inter-quartis.

Esta técnica foi aplicada às variáveis explicativas quantitativas (agregado familiar, rendimento mensal, número de filhos e idade), cuja transformação pode ser analisada nos gráficos presentes na secção dos anexos (A.1).

3.1.3 Análise Exploratória das Variáveis Explicativas

3.1.3.1 Variáveis Quantitativas³

Tabela 2: Estatísticas descritivas das variáveis quantitativas

	Média	Desvio Padrão	Mediana	Mínimo	Máximo
Agregado Familiar	1,98	1,01	2,00	1,00	9,00
Rendimento Mensal	1013,11	498,93	899,50	580,00	2899,50
Nº de filhos	0,56	0,88	0,00	0,00	4,00
Idade	47,34	14,61	46,00	18,00	91,00

Algumas observações podem ser feitas ao analisar as estatísticas destas variáveis:

- De uma forma geral, pode-se observar que o tipo de cliente que liga para o *call center* do Oney, em média, tem um agregado familiar reduzido (1,98 indivíduos). Esta estatística está ligeiramente abaixo do verificado com a dimensão do agregado familiar em Portugal, que se encontra na casa dos 2,5 indivíduos em 2017.

- Semelhantemente ao que se observa na variável acima descrita, o número de filhos do cliente que liga para o Oney é bastante reduzido - em média, os clientes têm 0,56 filhos.

³Os resultados apresentados na tabela 1 dizem respeito às variáveis quantitativas depois de aplicada a winsorização.

- A variável do rendimento mensal é referente ao rendimento do titular do contrato. O valor mínimo ronda o salário mínimo nacional e o rendimento médio mensal encontra-se perto dos mil euros. Ao analisar a remuneração base média ao longo dos anos em Portugal, verifica-se que esta tem evoluído positivamente; ainda assim encontra-se abaixo do valor quando comparamos com o típico cliente que contacta o Oney.

- A idade média do cliente que efetua a chamada é de, aproximadamente, 47 anos. É possível concluir também que a repartição do número de chamadas relativo a esta variável é semelhante à carteira global.

3.1.3.2 Variáveis Qualitativas

No anexo A.2 encontra-se uma breve análise obtida através do estudo da frequência de chamadas por cada uma das variáveis qualitativas, onde foi possível tirar algumas conclusões:

- O peso das chamadas dos clientes casados destaca-se face aos restantes estados civis (46%). Com um peso menor, apresentam-se as chamadas dos clientes solteiros (22%). Repare-se que as chamadas destes dois estados civis representam 68% das chamadas totais.

- Verifica-se grande concentração de chamadas na área metropolitana de Lisboa (Lisboa e Grande Lisboa, com um peso de 50%), Grande Porto (16%) e no Centro do País (13%). É possível verificar que as zonas do Algarve, Alentejo e Ilhas têm uma concentração de chamadas bastante reduzida.

- Mais de metade dos clientes que efetuam chamadas (55%) pertencem ao setor terciário - na tabela abaixo esta rubrica está categorizada como “Outras” (professores, advogados, hospedeira, etc.). Verifica-se também que apenas 10% se encontram inativos. Para os restantes não temos informação.

- São os clientes com comportamento de compras revolving ou sem juros que têm frequências de chamadas mais elevadas. É de notar que os clientes com comportamento de compra de Débito Imediato têm uma frequência de chamada muito reduzida (2%).

- A frequência de chamada aumenta com a antiguidade; ou seja, são os clientes mais antigos que ligam mais vezes - contratos com data de adesão superior ou igual a seis anos têm um peso de cerca de 70%.

- Clientes do sexo feminino, que têm telemóvel e *e-mail* contactam mais vezes o Oney.

A variável - motivo de chamada - pode influenciar o processo de modelização, pois quer a tendência, quer a sazonalidade, quer as características do cliente que efetua a chamada por um determinado motivo podem ser diferentes. Assim, esta variável carece de uma atenção especial, pelo que, de seguida, far-se-ão análises complementares.

Tabela 3: Frequência por motivo de chamada

Motivos	Frequência
Esclarecimento Extrato	55.195
Info Contratual	151.292
Info Pagamentos	53.382
Info Saldos	376.184
Info/ Pedidos Diversos	158.541
Pedido Financiamento	174.882
Reclamações	3.113
Outro	211.294

Nas análises apresentadas acima trabalhou-se com a informação do registo da chamada, a partir daqui trabalhar-se-á com a informação do contrato (cliente) que a efetua.

Note-se que as chamadas são tipificadas em dois momentos: um mais geral, que atribui a tipologia da chamada (se é um Pedido, Esclarecimento, Incidente, etc); e um mais específico, que caracteriza a tipologia de chamada (aumento de *plafond*, bloqueio de contrato, débitos em duplicado, etc). A base de dados apresenta 15 motivos gerais e 96 motivos particulares, ou seja, quando estes são interligados ficamos com um total de 110 motivos para tratar (ver anexo A.3). Para simplificar agruparam-se os dois grupos de tipificação do seguinte modo:

- Motivo 1) Esclarecimento Extrato - quando o cliente pretende obter esclarecimentos acerca do extrato (ciclo de extrato: 21M a 20M+1, com pagamento até 5M+2);
- Motivo 2) Informações Contratuais - quando o cliente pretende obter informações contratuais (ex: modalidades de pagamento, anulação de seguro/ contrato, etc);
- Motivo 3) Informação Pagamentos - quando o cliente pretende obter informações da entidade/ referência para efetuar o pagamento, fazer amortizações, etc;
- Motivo 4) Saldos - quando o cliente pretende saber o valor que tem em dívida ou o valor que tem disponível;
- Motivo 5) Info/Pedidos Diversos - quando o cliente pretende obter informações/ fazer pedidos diversos (ex: esclarecimentos clube de descontos, autorização de compra, etc);
- Motivo 6) Pedido Financiamento - quando o cliente pretende obter um financiamento;

- Motivo 7) Reclamações - quando o cliente efetua uma reclamação;
- Motivo 8) Outros - restantes motivos cuja tipificação não é nenhuma das acima descritas.

Neste ponto fez-se a análise da visão diária tendo como variável de referência o motivo de chamada; ou seja, realizou-se uma análise do perfil das chamadas por dia. Neste sentido, transforma-se cada motivo numa série temporal (são criadas oito séries temporais correspondentes aos oito motivos). Assim, para cada motivo fez-se a contagem do número de registos por dia, ao longo do período da amostra.

Tabela 4: Motivos das Chamadas - Estatísticas descritivas

	Nº Obs.	Média	Desvio Padrão	Mediana	Mínimo	Máximo	Assimetria	1º Quantil	3º Quantil
Motivo 1	1.235	44,69	29,54	38	1	205	1,27	25	58
Motivo 2	1.243	121,72	68,23	107	8	428	0,82	71	167
Motivo 3	1.174	45,47	38,27	43,50	1	202	0,92	8	65
Motivo 4	1.243	302,64	151,47	276	1	1.185	0,99	188	389
Motivo 5	1.243	127,55	57,28	129	7	419	0,26	81	169
Motivo 6	1.243	140,69	66,50	142	1	337	0,12	87	188
Motivo 7	983	3,17	2,38	3	1	20	2,18	1	4
Motivo 8	1.242	170,12	88,06	163	2	575	0,82	111	219

Ao analisarmos a tabela acima, é possível observar algumas conclusões, como:

- Os motivos que registam maior número de observações são: motivo 2, motivo 4, motivo 5, motivo 6 e motivo 8. Por outro lado, o motivo 7 é o que regista o menor número.

- Note-se que as médias dos registos de chamadas por dia são bastante diferentes devido às diferenças entre os motivos de chamada. Em média, o motivo 4 é o que regista mais chamadas, seguido do motivo 8 e do motivo 6. Já o motivo 7 apenas regista em média 3 chamadas por dia.

- Para distinguir as distribuições assimétricas apresenta-se o coeficiente de assimetria, onde um valor negativo indica que a cauda do lado esquerdo da função densidade de probabilidade é maior que a do lado direito e um valor positivo indica que a cauda do lado direito é maior que a do lado esquerdo. Verifica-se assim que todos os motivos têm a cauda do lado direito maior que a do lado esquerdo.

Uma análise importante passa por construir um gráfico que apresente as correlações entre os diversos motivos.

Apresenta-se de seguida uma matriz de correlações entre os oito motivos acima descritos. Esta matriz mostra os valores de correlação de Pearson (que podem compreender

valores entre -1 e 1), que mede o grau de relação linear entre cada par de motivos de chamada.

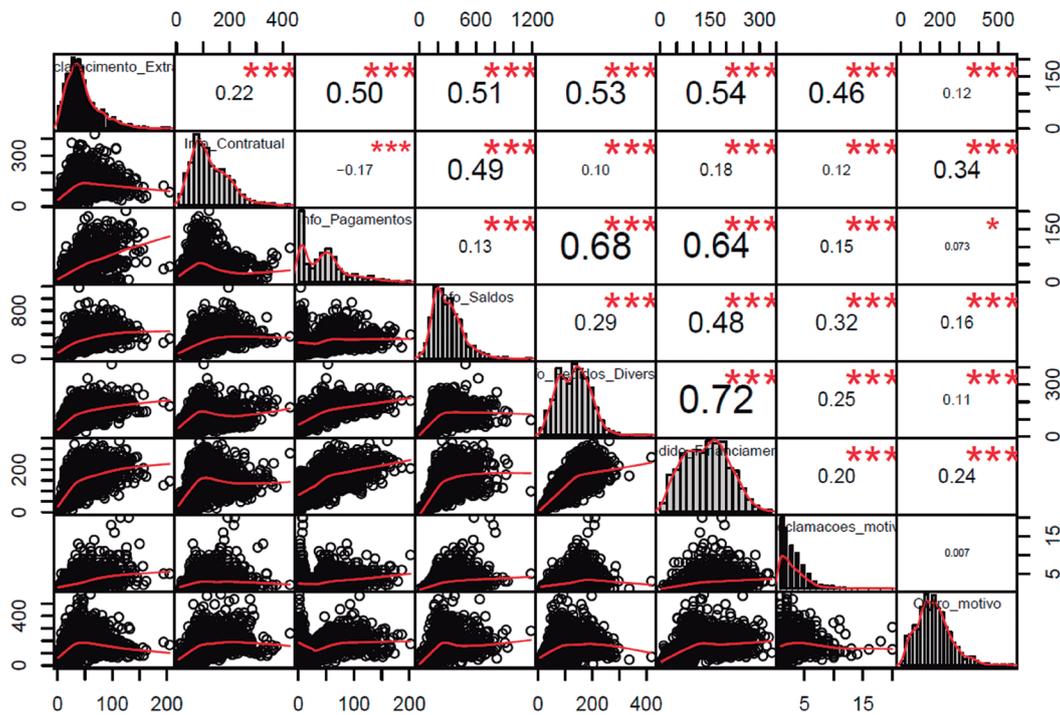


Figura 1: Matriz de Correlações entre os 8 Motivos

Vemos, inicialmente, que as correlações entre todos os motivos são positivas, à exceção da correlação entre o motivo de pedido de informações contratuais e pedido de informações sobre pagamentos. Admitindo que o intervalo entre 0 e 0,3 (positivo ou negativo) indica uma correlação desprezível; 0,3 e 0,5 (positivo ou negativo) uma correlação fraca; 0,5 a 0,7 (positivo ou negativo) uma correlação moderada; 0,7 a 0,9 (positivo ou negativo) uma correlação forte; e acima de 0,9 uma correlação muito forte é possível obterem-se algumas conclusões⁴:

- A maioria das correlações entre motivos são não significativas;
- Apenas existe uma correlação forte entre o motivo pedidos diversos com o motivo pedido de financiamento;
- O motivo esclarecimento extrato apresenta correlações moderadas com os motivos informações pagamentos, informações saldos, pedidos diversos e pedido de financiamento; e apresenta uma correlação fraca com o motivo reclamações;
- O motivo informações contratuais apresenta uma correlação baixa com outros motivos e uma correlação muito próxima da moderada com o motivo informações sobre saldos;

⁴ Método de classificação de intervalos de correlação usado por Mukaka (2012) em “A guide to appropriate use of Correlation coefficient in medical research” publicado em Malawai Medical Journal.

- O motivo informações sobre pagamentos apresenta correlações moderadas (fortes) com os motivos pedidos diversos e pedido de financiamento;
- O motivo informação sobre saldos apresenta uma correlação moderada muito próxima da correlação alta com o motivo pedido de financiamento e uma correlação moderada com o motivo reclamações.

Após análise da correlação entre motivos de chamada, torna-se interessante analisar estes mesmos motivos por variáveis de negócio e variáveis sociodemográficas (ver anexo A.4).

Quer na relação com as variáveis de negócio, quer na relação com as variáveis sociodemográficas verifica-se que o motivo com maior peso é o motivo 4 - informação sobre saldos. No entanto é ainda possível tirar algumas conclusões:

- *Antiguidade*: Para os clientes com menos de 1 ano, o peso das chamadas para o motivo informação sobre saldos e para outros motivos é bastante semelhante; já nos restantes escalões de antiguidade o peso do motivo 4 destaca-se face aos restantes. Para os clientes com mais de 6 anos o peso do motivo pedido de financiamento aumenta, ficando perto do peso dos outros motivos.

- *Idade*: O peso das chamadas dos motivos pedido de financiamento e esclarecimento de extrato vai aumentando com a idade. Verifica-se maior peso dos motivos informações sobre pagamentos e informação/ pedidos diversos junto da população mais jovem.

- *Estado civil*: O comportamento do peso das chamadas dos motivos informações contratuais, informações sobre saldos, pedido de financiamento e outros motivos é ligeiramente diferente no estado civil categorizado como “outros”. O peso das chamadas do motivo informação/ pedidos diversos dos clientes viúvos é mais reduzido do que para os restantes estados civis.

- *Número de filhos* : Quanto à variável número de filhos, o peso das chamadas para os diferentes motivos não altera significativamente. No entanto, é possível verificar certas diferenças ligeiras. O peso das chamadas dos motivos informações contratuais, informação sobre saldos e outros motivos decresce tenuemente quanto mais filhos o cliente tiver; já os motivos informação sobre pagamentos, informações/ pedidos diversos e pedido de financiamento comportam-se de forma inversa.

- *Rendimento mensal*: O peso das chamadas dos motivos informações sobre pagamentos e informações sobre saldos vai diminuindo à medida que o salário aumenta. O peso das chamadas dos outros motivos e dos pedidos de financiamento é bastante semelhante ao longo dos diferentes escalões de rendimento, sendo que, no escalão de rendimento mais baixo, o peso das chamadas dos outros motivos é superior e, no escalão de rendimento dos 1.250€ aos 1.500€, o peso das chamadas do motivo pedido de financiamento é superior.

- *Região*: O peso das chamadas do motivo informações sobre saldos é bastante considerável em Lisboa e Grande Lisboa, Grande Porto e Algarve. As chamadas efetuadas para o motivo informações/ pedidos diversos têm um peso maior na Região Autónoma dos Açores. É na região do Alentejo onde se verifica um peso maior de chamadas do motivo pedido de financiamento.

- *Modalidade de compra*⁵: O comportamento da tipificação das chamadas é bastante diferente consoante a modalidade de compra preferencial nos últimos 12 meses dos clientes. O peso das chamadas do motivo informações contratuais e dos outros motivos é superior para aqueles que não têm uma modalidade preferencial (inativos). A rúbrica de informação sobre saldos tem um peso maior nos clientes que preferem compras com juros.

- *Categoria profissional*: O peso das chamadas do motivo informações sobre saldos é superior na categoria profissional desempregados e domésticas. Verifica-se também que os desempregados e os estudantes são os que menos ligam para fazer pedidos de financiamento e para obter informações contratuais. O cruzamento entre estas duas variáveis permite-nos discutir sobre o aspeto socioeconómico desta conclusão: em geral, estas profissões descritas acima têm rendimentos mais baixos e conseqüentemente maiores dificuldades em pagar. Assim, o limite de crédito atribuído é mais reduzido e faz com que liguem com maior frequência.

Interessa-nos ainda perceber a interação entre as diferentes tipologias de chamada (exemplo: se o cliente liga apenas uma vez e não volta a ligar; ou se liga uma vez e volta a ligar por um outro motivo). Apresenta-se de seguida um diagrama - *chord diagram* - onde é possível visualizar as inter-relações entre motivos de chamada.

As relações são traçadas através das curvas de Bézier ou arcos, cuja espessura traduz a proporção de conexão. O círculo é dividido em oito motivos, cujo comprimento do arco externo do diagrama de cada motivo corresponde ao peso da tipificação no registo das chamadas. Ora, é possível verificar que o motivo 4 (informação saldos) é o que tem maior peso (21%),

⁵ Para definir a modalidade de compra preferencial calculou-se para cada cliente em qual das modalidades se verificou um montante mais elevado de compra.

seguido do motivo 8 (outros motivos) com 18% e, em terceiro lugar o motivo 5 (informações ou pedidos diversos) com 16%.

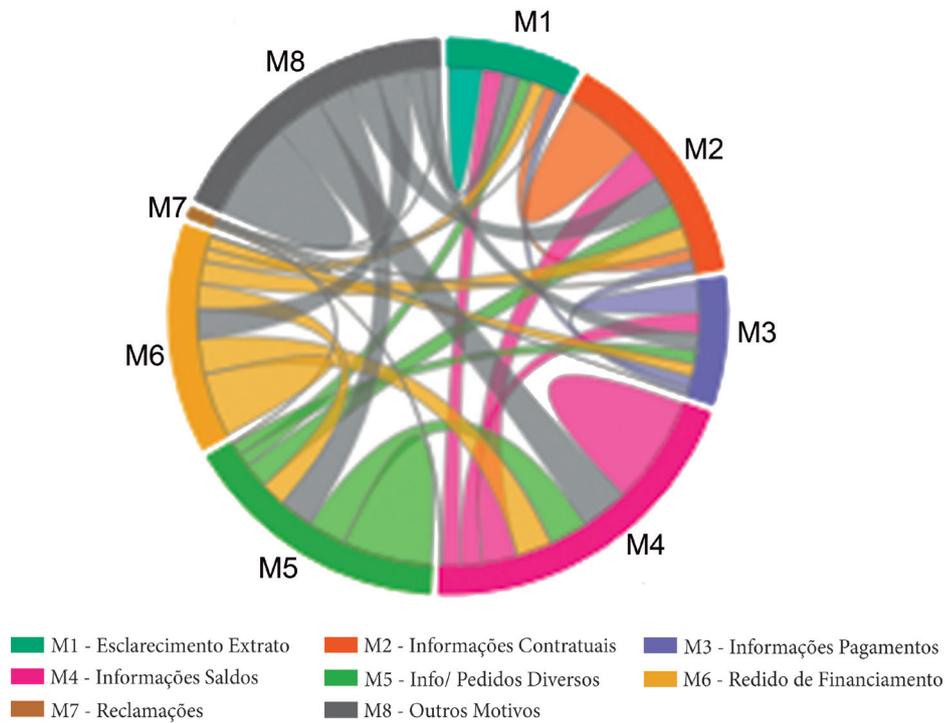


Figura 2: Diagrama de chord - análise global

Os arcos correspondentes a cada motivo têm a cor do motivo que obteve mais ligações. Com isso em mente, repare-se que todos os acordes dos arcos do motivo 7 e do motivo 8 têm a cor do motivo correspondente; ora, verifica-se assim que são mais os clientes que ligam para fazer alguma reclamação ou que ligam para um outro motivo e voltam a ligar para os restantes, do que o contrário. Existem também arcos de cada motivo que conectam o motivo entre si, ou seja, os arcos em forma de “colina” representam os clientes que só ligaram para um determinado motivo e não voltaram a ligar para um outro diferente. Algumas considerações a ter em conta:

- Cerca de 17% dos clientes que ligaram para obter esclarecimentos acerca do extrato, ligaram também para questionar acerca dos saldos; esta percentagem é mais de metade do que a correspondente aos clientes que ligam para obter esclarecimentos acerca do extrato. Por outro lado, apenas cerca de 7% dos clientes que ligaram para questionar acerca dos saldos, ligaram também para obter esclarecimentos acerca do extrato. Verifica-se ainda que 30% dos clientes ligam apenas para obter esclarecimentos acerca do extrato.

- Dos clientes que ligaram para obter informações contratuais, observa-se que cerca de 16% ligaram também para questionar acerca dos saldos. Por outro lado, cerca de 11% dos clientes que ligaram para questionar acerca dos saldos, ligaram também para obter informações contratuais. Observa-se ainda que 37% dos clientes ligam apenas para obter informações contratuais.

- Cerca de 18% dos clientes que ligaram para pedir informações sobre pagamentos, ligaram também para questionar acerca dos saldos; esta percentagem é mais de metade do que a correspondente aos clientes que ligam para obter informações sobre pagamentos. Por outro lado, apenas cerca de 7% dos clientes que ligaram para questionar acerca dos saldos, ligaram também para pedir informações sobre pagamentos. Verifica-se ainda que 37% dos clientes ligam apenas para pedir informações sobre pagamentos.

- Observa-se que 15% dos clientes que ligaram para questionar acerca dos saldos, ligaram também por outros motivos. Já 17% dos clientes que ligaram por outros motivos, ligaram também para fazer questões sobre os saldos. É possível observar que cerca de 36% dos clientes ligam apenas para questionar acerca dos saldos.

- Ao analisar as interações do motivo 5 (Informações/ Pedidos Diversos), verifica-se que 16% dos clientes que ligaram para este motivo, ligaram também para obter informações sobre os saldos; de forma inversa, 13% dos clientes que ligaram para se informarem sobre os saldos, ligaram também para obter informações ou fazer pedidos diversos. 37% dos clientes ligam unicamente para este motivo.

- Passando para os clientes que ligam para fazer pedidos financiamento, verifica-se que 32% ligam somente para este motivo. Cerca de 18% ligam também para se informarem sobre os saldos; e, dos clientes que ligam por este último motivo, perto de 12% ligam para fazer pedidos financiamento.

- A percentagem de clientes que liga para fazer reclamações é a mais reduzida - face ao total de registos de chamadas, a proporção correspondente ao motivo 7 é de apenas 1%. Ainda assim, é possível visualizar uma interação de 13% com o motivo 4 (informação sobre saldos); inversamente, a percentagem dos clientes que ligam para obter informações sobre os saldos e voltam a ligar para fazer alguma reclamação, não chega a 1%. Verifica-se também que cerca de 28% dos clientes ligam unicamente para fazer alguma reclamação e não ligam de novo por um outro motivo.

- Para analisar o motivo 8 veja-se a abordagem feita no ponto 4, onde a maior interação é feita com o motivo 4. Verifica-se ainda que cerca de 37% dos clientes ligam apenas para a tipificação outros motivos.

Paralelamente torna-se interessante fazer uma análise mais detalhada, por ano, o comportamento do cliente que efetua a chamada. Em geral, o comportamento do cliente por ano é bastante semelhante ao que foi detetado de forma global.

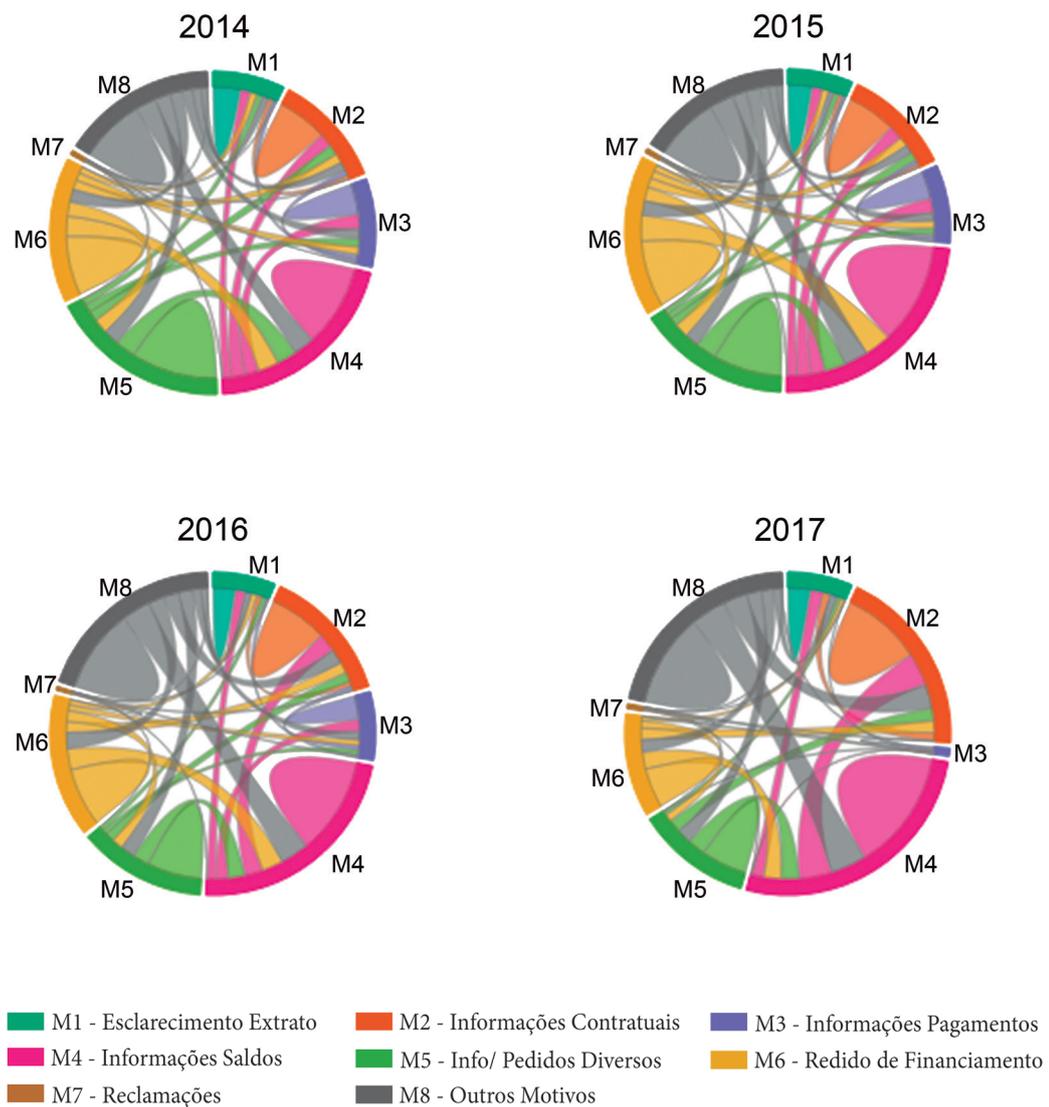


Figura 3: Diagrama de chord - Análise por Ano

Analisando o quadro abaixo, podem-se tirar conclusões mais detalhadas:

Tabela 5: Análise global vs análise ao longo da amostra:

Peso de cada motivo sobre o total dos registos (em %)

	Global	Ano 2014	Ano 2015	Ano 2016	Ano 2017
Motivo 1	8	7	7	7	7
Motivo 2	15	12	11	14	20
Motivo 3	8	9	8	7	1
Motivo 4	21	21	24	24	28
Motivo 5	16	19	16	13	12
Motivo 6	14	15	17	15	11
Motivo 7	1	0	0	0	0
Motivo 8	18	16	16	20	22

- Para o motivo 1 e para o motivo 7 as alterações ao longo do período de análise são praticamente nulas e o peso destes motivos mantém-se coerente com o verificado na análise global;

- O motivo 2, o motivo 4 e o motivo 8 vão ganhando peso ao longo do período de análise. Respetivamente, em 2014 têm um peso de 12%, 21% e 16%; já em 2017 aumentam para 20%, 28% e 22% - valores ligeiramente diferentes do observado no primeiro diagrama apresentado, onde estes motivos tinham um peso de 15%, 21% e 18%;

- À exceção do ano 2017, cujo peso do motivo 3 diminui para 1%, para os restantes anos, mantém-se nos valores do que se observa de forma global;

- O motivo 5 vai perdendo peso no decorrer dos anos: em 2014 apresenta um peso de 19%, que vai diminuindo até chegar aos 12% em 2017. Se olharmos para a análise global verificamos que este motivo tem um peso de 16%.

- Os pesos observados para o motivo 6, ao longo dos três primeiros anos, estão acima do observado quando se faz a análise junta (14%). No ano 2017, verifica-se que o peso está abaixo (11%).

3.2 Metodologia

Após leitura das revisões bibliográficas já existentes, optámos por seguir a metodologia usada por Taylor (2003) – extensão da suavização exponencial de HW que acomoda dois ciclos sazonais, através da introdução de um novo índice sazonal e uma equação extra de suavização para esse novo índice – e a metodologia usada por Hyndman & Athanasopoulos (2014) – regressão harmónica dinâmica, sendo esta considerada uma extensão aos modelos ARIMA

para permitir que outras informações sejam incluídas nos modelos.

3.2.1 Métodos de Alisamento Exponencial

Esta técnica, introduzida por Robert G. Brown, embora simples é muito utilizada para a previsão de séries temporais (Snyder *et al.* 2002, Makridakis *et al.* 1982, Makridakis & Hibon 2000). Surge no seguimento da metodologia de médias móveis, onde se pretendia generalizar o processo de médias móveis ponderadas a todo o intervalo da amostra. Para obter as estimativas dos valores futuros da série, este método utiliza a previsão do último instante ajustada pelo respetivo erro de previsão, através da seguinte expressão: $P_{t+1} = P_t + \alpha(Y_t - P_t)$, onde Y_t é o valor da série no instante t , P_t a previsão obtida no mesmo instante e α a constante de amortecimento ou alisamento que varia entre 0 e 1. Veja-se que quanto mais perto de 1 estiver esta constante maior será o efeito de ajustamento do erro de previsão; inversamente, quanto menor for o valor da constante, maior será a intensidade do alisamento. Segundo Caiado (2011), o termo “alisamento exponencial” significa que os ponderadores são exponencialmente decrescentes com a antiguidade das observações, ou seja, ao desenvolver de forma recursiva a expressão acima, facilmente se nota que esta ponderação dá maior importância às observações mais recentes. Assim, a ponderação atribuída a cada observação vai decaindo exponencialmente à medida que as observações se vão afastando do período t .

Consoante as características das séries, nomeadamente a tendência linear (nula, aditiva, aditiva amortecida, multiplicativa ou multiplicativa amortecida) e a sazonalidade (nula, aditiva ou multiplicativa) outras abordagens a este método são empregues (ver Gardner, Jr. 2006). Para séries temporais que apresentam tendência linear utiliza-se o método de alisamento exponencial duplo, que consiste na aplicação do método de alisamento exponencial simples duas vezes, utilizando a mesma constante de alisamento. O método de Holt é adequado a séries com tendência linear e sem movimentos de carácter sazonal, constituindo uma alternativa ao método de alisamento exponencial duplo para estimar o nível e o declive da tendência da série. Para séries que apresentam tendência linear e movimentos sazonais surge o método de HW na forma multiplicativa e na forma aditiva. Taylor (2003) ampliou o método de HW de modo a conseguir incorporar a segunda componente sazonal da seguinte forma:

$$S_t = \alpha(\gamma_t/(D_{t-s_1} W_{t-s_2})) + (1 + \alpha)(S_{t-1} + T_{t-1}) \quad (1a)$$

$$T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma) T_{t-1} \quad (1b)$$

$$D_t = \delta(\gamma_t/(S_t W_{t-s_2})) + (1 - \delta) D_{t-s_1} \quad (1c)$$

$$W_t = \omega(\gamma_t/(S_t D_{t-s_1})) + (1 - \omega) W_{t-s_2} \quad (1d)$$

$$\hat{y}_t(k) = (S_t + kT_t) D_{t-s_1+k} W_{t-s_2+k} + \phi^k (\gamma_t - ((S_{t-1} + T_{t-1}) D_{t-s_1} W_{t-s_2})) \quad (1e)$$

onde S_t e T_t representam, respetivamente, as componentes em nível e a tendência da série; D_t e W_t são os índices sazonais para o ciclo semanal e anual respetivamente; α , γ , δ e ω são os parâmetros de suavização; e $\hat{y}_t(k)$ é a previsão de k passos à frente feita a partir da previsão com origem em t . O termo que envolve o parâmetro ϕ na expressão (1e) é um simples ajustamento para a autocorrelação de primeira ordem.

Os modelos usados para este tipo de estimação assumem de d_t é um processo não-autocorrelacionado; no entanto, essa hipótese pode não se verificar. Chatfield (1978), ao usar o método de HW na forma multiplicativa mostrou que o erro pode ser descrito por um processo AR(1). Taylor (2003), Gardner (1985), Reid (1975) e Gilchrist (1976) também encontraram problemas semelhantes em estudos realizados neste âmbito.

Para esta aplicação empírica vamos implementar uma versão do método de HW (Taylor, 2003) com 2 períodos sazonais: um ciclo de 7 dias ($s_1=7$), que retrata o ciclo dentro da semana, e um ciclo de 365,25 dias ($s_2^6=365,25$), que retrata o ciclo dentro do ano.

3.2.2 Modelos ARIMA e SARIMA

Os processos ARIMA foram extensivamente estudados por George Box e Gwilym Jenkins no início dos anos 1970; como resultado estes processos são conhecidos como modelos de Box-Jenkins. Estes são modelos mais sofisticados, cuja identificação é feita através da análise do correlograma da série temporal. Assim, segundo Hyndman (2001) a abordagem ARIMA para previsão baseia-se nas seguintes ideias:

⁶ Segundo Hyndman *et al.* (2011) o período de 365,25 dias é a duração média de um ano que permite anos bissextos.

- As previsões são baseadas em funções lineares das observações da amostra;
- O objetivo é encontrar os modelos mais simples que fornecem uma descrição adequada dos dados observados - princípio da parcimónia.

Segundo Caiado (2011), a metodologia de Box-Jenkins mostra que a modelação de uma série temporal consiste na identificação do modelo, estimação dos parâmetros, avaliação do diagnóstico, seleção dos modelos e previsão.

Para a primeira etapa tem-se então como objetivo a identificação de um modelo SARIMA (p, d, q)(P,D,Q)_s - p descreve a parte autorregressiva AR, d a parte do integrado, q a parte das médias móveis (MA) e s a componente/ período sazonal. Na figura 5 é apresentado o cronograma da série original, FAC e FACP e são apresentados os mesmos gráficos para a série após diferenciação sazonal semanal. Após identificação do ARIMA “provisório”, os parâmetros foram estimados através da função auto.arima do software R, sendo esta uma função uma implementação do algoritmo que compara todos os modelos apropriados à série e seleciona o melhor dos modelos de acordo com o critério de informação de Akaike, 1973 (AIC). Para a terceira etapa, foi feita a avaliação da qualidade estatística das estimativas obtidas (teste de significância individual) e a avaliação da qualidade do ajustamento do modelo ARIMA (através da análise dos resíduos). Para a seleção de modelos foi usado o AIC.

3.2.3 Modelos sazonais trigonométricos

Neste tópico são introduzidas componentes sazonais com representação trigonométrica (senos e cossenos) baseadas nas séries de Fourier⁷ (ver West & Harrison 1997, Harvey 1989):

$$s_t^{(i)} = \sum_{j=1}^{K_i} s_{j,t}^{(i)} \quad (2a)$$

$$s_{j,t}^{(i)} = s_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \sin \lambda_j^{(i)} + \gamma_1^{(i)} d_t \quad (2b)$$

$$s_{j,t}^{*(i)} = -s_{j,t-1}^{(i)} \sin \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} d_t \quad (2c)$$

onde, $\gamma_1^{(i)}$ e $\gamma_2^{(i)}$ são parâmetros de suavização, com $\lambda_j^{(i)} = 2\pi j/m_i$. O número de parâmetros harmónicos necessários para a i-ésima componente sazonal é denotado por K_i . Para o estudo empírico apresentado estamos a considerar $m_1 = 7$ e $m_2 = 365,25$. Repare-se que o argumento k especifica quantos pares de termos senos e cossenos se devem incluir; ou seja, é o número

⁷ Em R, para definir a componente sazonal através de funções trigonométricas, usou-se a função msts (*Multi-Seasonal Time Series*), uma vez que estamos a trabalhar com dois períodos sazonais $m_1 = 365,25$ e $m_2 = 7$. Consequentemente, como usámos a função msts, para posterior aplicação das séries Fourier, o K (máximo de termos da série de Fourier) deve ser um vetor, que especifica o número de termos seno e cosseno para cada um dos períodos sazonais.

total de termos da série de Fourier para cada período sazonal e $k = m/2$ é o valor máximo que k pode tomar. Testaram-se vários valores para k e considerou-se $k = 3$, pois é o valor que minimiza o AIC:

$$\frac{\sin 2\pi t}{365,25}, \frac{\cos 2\pi t}{365,25}, \frac{\sin 2\pi t}{7}, \frac{\cos 2\pi t}{7}, \text{ para } k = 1, 2, 3. \quad (2d)$$

Em vez de usarmos os termos da série de Fourier, poder-se-ia trabalhar com variáveis *dummy* sazonais, no entanto para m grande são menos os termos da série de Fourier que têm de se ajustar face às primeiras. Inversamente, quando se trata de períodos sazonais curtos, existe pouca vantagem em usar as séries de Fourier.

Segundo Hyndman & Athanasopoulos (2014), quando existem períodos sazonais extensos, uma regressão dinâmica com termos de uma série de Fourier apresenta frequentemente melhores resultados do que outros modelos. Assim, fez-se uma regressão harmónica em que o padrão sazonal é modelado por termos de uma série de Fourier com dinâmicas de curto prazo a serem tratadas com um erro tipo ARIMA.

3.2.3.1 Análise Espectral

As componentes sazonais consideradas nos parâmetros harmónicos foram definidas com base na análise espectral apresentada no anexo (B.1), que é uma das várias técnicas estatísticas que permite caracterizar e analisar dados observados num dado período de tempo. Neste caso, utilizou-se esta análise para caracterizar o número de chamadas efetuadas para o *call center* do Oney. Segundo J.N. Rayner (2001), a análise espectral refere-se à decomposição de uma sequência em oscilações de diferentes comprimentos. Ou seja, passa-se do domínio de dados para o domínio espectral, pelos seguintes motivos:

1. Algumas formas de manipulação são mais fáceis no domínio espectral;
2. Os comprimentos das oscilações são dados estatísticos necessários para concluir possíveis fatores que afetam/ produzem os dados que nos propomos analisar.

Esta análise é baseada no teorema de Fourier⁸, que afirma que qualquer forma de onda pode ser decomposta na soma de ondas senoidais com diferentes frequências e amplitudes. Assim, quando somadas, essas ondas reconstituem então a forma da onda original. Diz-se então que toda a série de tempo pode ser expressa a partir da soma de senos e cossenos que oscilam de

⁸ Introduzida por Jean Baptiste Joseph Fourier (1807), é uma forma trigonométrica usada para representar funções infinitas e periódicas complexas, na forma de funções trigonométricas simples de senos e cossenos.

acordo com uma determinada frequência. O desafio é poder identificar essas frequências, e nesta análise, será feito via estimativa do periodograma - conhecido como densidade espectral e relaciona as variabilidades do conjunto de dados com as frequências (veja-se que na análise temporal, as variabilidades são relacionadas com o domínio do tempo). A transformação de Fourier não é mais do que uma operação matemática que fornece os parâmetros de frequência, amplitude e períodos associados.

Esta análise requer que a série temporal seja estacionária: as propriedades estatísticas da série temporal não podem variar durante o período da amostra. Caso a série seja não-estacionária, as suas propriedades irão influenciar as estimativas da amplitude, frequência e períodos associados. Um método a aplicar para o caso de séries não estacionárias é usar um modelo de regressão linear para a tendência. Esta alternativa deve ser usada com cautela, visto que os coeficientes da regressão linear são sensíveis a *outliers*. Uma outra alternativa mais dinâmica passa por “prejudicar” as diferenças de primeira ordem das séries temporais.

Após identificação da densidade espectral podemos então avaliar os picos de frequência e os períodos associados a este. Ora, supondo que temos uma série mensal e que na análise espectral identificámos um pico na frequência “a”; então o período associado a esse ciclo, medido em meses, é dado por $1/a$ (nota: quanto maior for o valor de a, menor será o período, em tempo, associado a essa frequência e, na maioria das vezes, é comum termos mais de um pico na densidade espectral).

4. Análise e Discussão dos Resultados

A figura 4 mostra o número de chamadas recebidas no *call center* do Oney por dia desde janeiro 2014 até dezembro de 2017. Os 1.461 dados são observados diariamente e mostram um forte padrão sazonal semanal e anual. Esta afirmação pode ser verificada no anexo (B.1), onde é possível observar os gráficos do periodograma global e dos periodogramas de cada um dos motivos (repare-se que apenas pela observação ao gráfico não seria possível identificar as frequências); no entanto, é também apresentada uma tabela com 3 colunas, onde a primeira coluna corresponde ao eixo vertical e permite identificar os picos da respectiva densidade espectral. A segunda coluna corresponde ao eixo horizontal do gráfico do periodograma, a terceira coluna mostra-nos a relação tempo/ ciclo – ou seja, o período de ocorrência do ciclo, dado por $1/\text{frequência}$. Para o motivo 1 – esclarecimento extrato – por exemplo, a primeira frequência com valor 0,14 corresponde ao maior pico encontrado. A terceira coluna mostra-nos então a relação tempo/ ciclo, ou seja, o período de ocorrência do ciclo, dado por $1/\text{frequência}$. Ora, para a frequência de 0,14 temos um período cíclico de 7 dias. Note também que há um pico no periodograma associado a uma frequência de 0,03 e que gera um período de 31 dias. Existem ainda outros picos que geram períodos cíclicos de 3 dias e 2 dias com frequências de 0,29 e 0,43 respetivamente; revelando assim a existência de sazonalidade nos nossos dados. Para os restantes motivos, a interpretação é feita da mesma forma, realçando os períodos cíclicos semanais (7 dias), anuais (360/365), $2/3$ dias e mensais (aproximadamente 31 dias) que ocorrem para todos os motivos. Assim as correspondentes componentes sazonais devem ser refletidas no modelo com frequências $m_1=7$ e $m_2=365,25$.

Verifica-se maior número de chamadas nos dias de inverno do que nos dias de verão; a análise sazonal semanal mostra que o número das chamadas é mais elevado no início da semana (2ª feira) e vai decrescendo até sábado. Conforme podemos ver no anexo (B.2), na análise sazonal mensal é possível verificar um elevado número de chamadas por volta do dia 5 e do dia 20 de cada mês, que coincide com as datas dos extrato apresentadas na descrição dos motivos do sub-capítulo 3.1.3.2.

Como a tendência parece estar próxima de zero, a taxa de crescimento foi omitida dos modelos. A figura 4 mostra também as autocorrelações e as autocorrelações parciais para a amostra de treino. A FAC e a FACP do lado esquerdo são relativas à série original e mostram o elevado grau de persistência e padrões acentuados de sazonalidade em múltiplos períodos

sazonais de 7 e 365 dias. Isto sugere então uma diferença sazonal semanal ($1 - B^7$) e uma diferença sazonal anual ($1 - B^{365}$), para tornar a série estacionária. Do lado direito, é apresentada a série duplamente diferenciada, $(1 - B^7)(1 - B^{365})Y_t$, bem como a respetiva FAC e FACP.

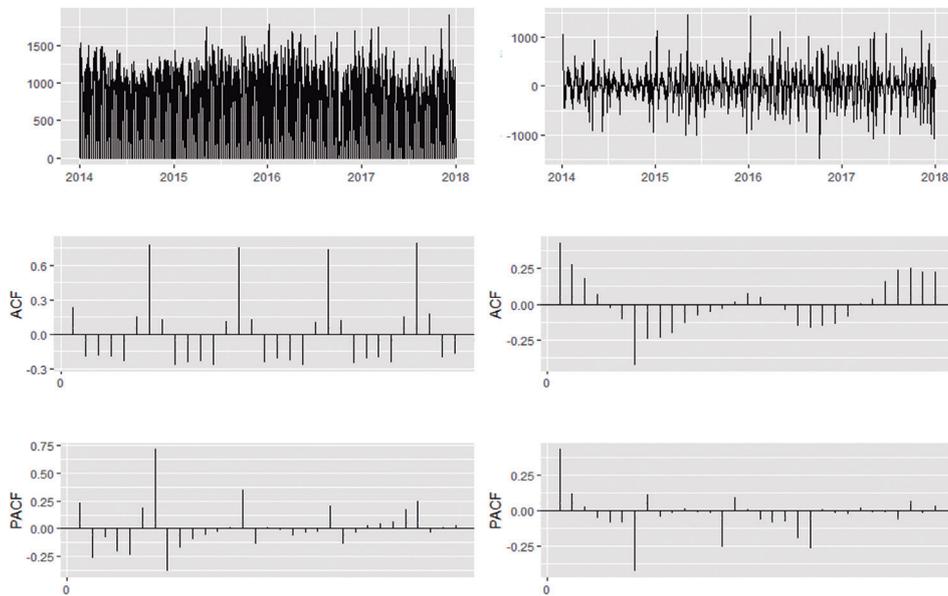


Figura 4: Número de Chamadas Recebidas no *Call Center*, FAC e FACP

Definiu-se como base de treino do modelo a base com dados entre janeiro 2014 e dezembro de 2017 (1.461 observações); e como base de validação do modelo a base com dados entre janeiro 2018 e julho 2018 (212 observações). Foram então obtidas as estimativas de máxima verossimilhança para seleccionar o número apropriado de termos da série de Fourier e o número de ordens do modelo ARIMA.

Existem vários critérios para fazer a seleção do modelo: Erro Quadrático Médio (*Mean Square Error*, MSE), o Erro Absoluto Médio (*Mean Absolute Error*, MAE) e o Erro Percentual Absoluto Médio (*Mean Absolute Percent Error*, MAPE), AIC, entre outros. Neste estudo vou usar o AIC $AIC_p = -2\log(L_p) + 2[(P+1) + 1]$, onde p é o número de variáveis explicativas consideradas no modelo. O AIC aumenta conforme SQE (Soma dos Quadrados Explicada) aumenta; ora, facilmente se percebe que este critério penaliza modelos com muitas variáveis e valores menores de AIC são preferíveis. Num estudo empírico, Billah *et al.* (2005) indicou que abordagens que usam como critério o AIC fornecem a melhor seleção de modelos em relação a outros métodos como validação de previsão. Ainda assim, existem outros critérios de informação alternativos como o AICc (Burnham & Anderson 2002).

Conforme podemos ver no anexo (B.3), o número de termos da série de Fourier $K = c(3, 3)$ - 3 pares de senos e cossenos para cada período sazonal - e a ordem ARIMA (0,1,3) são os resultados que produzem o menor valor do AIC. A decomposição da série temporal nas componentes sazonais (semanal e anual), tendência e restantes é mostrada na figura 5. As componentes sazonais são estimadas iterativamente usando as previsões são obtidas através da aplicação de um método de previsão não sazonal aos dados depois de retirada a sazonalidade. Os eixos verticais do lado direito têm todos a mesma altura, no entanto encontram-se em diferentes escalas, fornecendo assim uma comparação ao nível do tamanho de cada componente. Como podemos ver na mesma figura, o padrão sazonal anual evolui consideravelmente no tempo, enquanto que o padrão semanal é relativamente estável. A componente da tendência é bastante pequena em magnitude em comparação com as componentes sazonais.

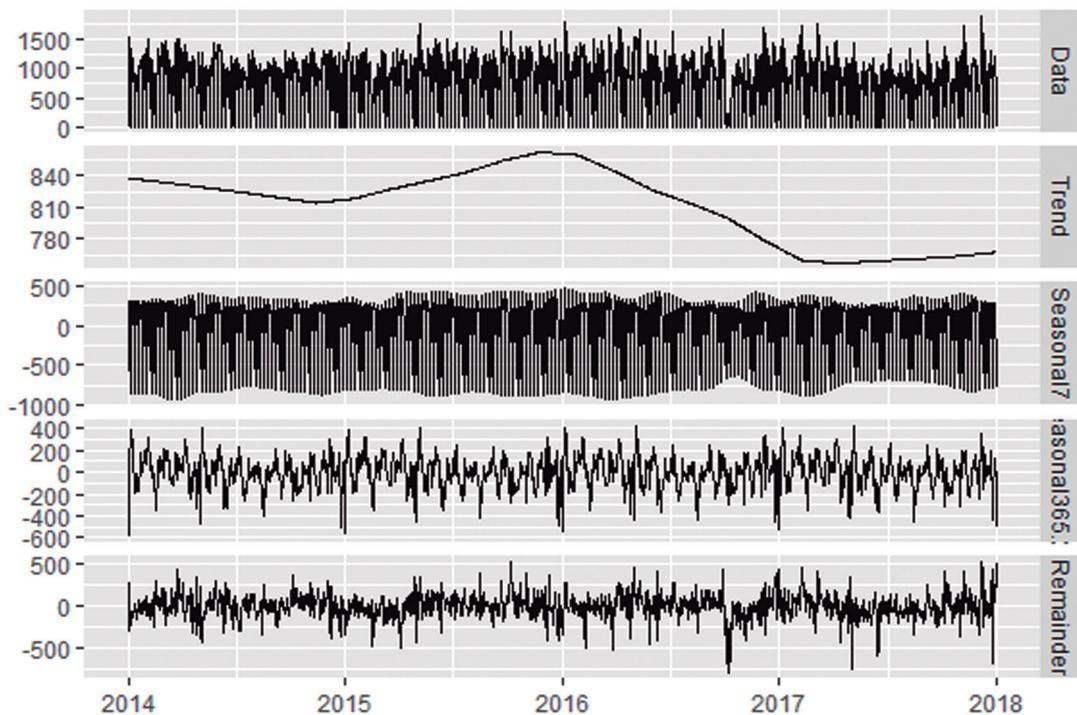


Figura 5: Decomposição trigonométrica da série do número de chamadas recebidas no *Call Center*

Foram usados dois métodos para as previsões desde janeiro 2018 até julho de 2018 (7 meses com 212 observações): Holt-Winters (HW) e regressão harmónica cujo padrão sazonal é modelado por séries de Fourier com dinâmicas de curto prazo a serem tratadas com um erro tipo ARIMA.

Um ponto interessante a ser observado em relação ao primeiro método de previsão

escolhido (abordagem de suavização exponencial sazonal dupla) face à modelação ARIMA é que não existe especificação do modelo envolvida. Segundo Taylor (2006), este facto dá ao método um forte apelo em termos de simplicidade e robustez. Quanto ao segundo método de previsão, este foi escolhido em detrimento da versão sazonal do modelo ARIMA, uma vez que, segundo Hyndman & Athanasopoulos (2014), quando há longos períodos sazonais uma regressão dinâmica com termos de uma série de Fourier é frequentemente melhor⁹. Assim, as vantagens do método escolhido são:

- Permite qualquer sazonalidade (mesmo que períodos longos);
- Para dados com mais de um período sazonal, os termos da série de Fourier de diferentes frequências podem ser incluídos;
- A suavidade do padrão sazonal pode ser controlada por K (pares de senos e cossenos para cada período sazonal) – o padrão sazonal é mais suave para valores menores de K ;
- As dinâmicas de curto prazo são facilmente manipuladas com um erro ARMA simples.

A única desvantagem, quando comparamos com um modelo SARIMA, é que a sazonalidade é considerada fixa – o padrão sazonal não pode mudar com o tempo.

A tabela 6 mostra os parâmetros que estimámos através do modelo de HW. ω apresenta-se com um valor alto e α com um valor baixo, sugerindo que o ajuste para a autocorrelação de primeira ordem tornou, em grande parte, redundante a equação de suavização em nível. Os baixos valores para o parâmetro de suavização de tendência, γ , foram acompanhados por valores muito pequenos para a tendência suavizada, T_t . Esta conclusão parece-nos razoável, uma vez que a variação nos períodos de estimação de 3 anos (base de treino) é dominada pela sazonalidade.

Tabela 6: Estimativas do modelo HW.

	Level α	Trend γ	Weekly seasonality δ	Annual seasonality ω	AR λ
Calls	0,01	0,00	0,39	0,15	0,59

⁹ As funções `Arima()` e `auto.arima()` permitem um período sazonal até $m=350$, mas, na prática, estas funções perdem memória quando o período sazonal é superior a 200.

A tabela 7 mostra os parâmetros que estimámos através do modelo ARIMA (0, 1, 3) com $K = c(3, 3)$ termos de Fourier. Os valores obtidos nas estimativas dos parâmetros são consistentes com as análises anteriormente feitas: os coeficientes que apresentam valores mais elevados dizem respeito aos valores que dominam os padrões sazonais de variação, nomeadamente o padrão semanal. Se comparamos os coeficientes dos parâmetros harmónicos do padrão anual verificamos que estes são bastante mais baixos, refletindo que o padrão sazonal tem uma amplitude mais elevada tal como visualizado nos periodogramas em anexo.

Tabela 7: Estimativas do modelo ARIMA(0, 1, 3) com $K = c(3, 3)$ termos de Fourier.

Parameter	Estimate	Standard error	Parameter	Estimate	Standard error
θ_1	-0,598	0,026	$C_7^{(3)}$	-174,795	5,592
θ_2	-0,227	0,032	$S_{365,25}^{(1)}$	31,973	15,187
θ_3	-0,155	0,024	$C_{365,25}^{(1)}$	26,956	14,490
$S_7^{(1)}$	323,927	9,627	$S_{365,25}^{(2)}$	1,557	12,542
$C_7^{(1)}$	657,117	9,631	$C_{365,25}^{(2)}$	-5,956	12,329
$S_7^{(2)}$	-285,573	6,209	$S_{365,25}^{(3)}$	-1,568	11,883
$C_7^{(2)}$	156,055	6,209	$C_{365,25}^{(3)}$	-3,222	11,883
$S_7^{(3)}$	-81,508	5,585			

As previsões são apresentadas no anexo (B.4) e o desempenho dos métodos univariados estimados foi avaliado através do cálculo do MSE para previsões em várias etapas: mês 1 (janeiro 2018) até ao mês 7 (julho 2018).

Tabela 8: Valores dos MSE oriundos dos métodos HW e ARIMA

Horizonte	MSE_HW	MSE_A
1	1037827	698221
2	1038628	648817
3	945520	590625
4	1207129	583446
5	853403	437284
6	719364	469824
7	591665	330968

A interpretação inicial dos resultados sugere que a capacidade de prever o número de chamadas recebidas no *call center* não diminui com o horizonte de previsão. O modelo ARIMA, cuja parte sazonal é explicada por termos da série de Fourier, supera o modelo HW

em todos os horizontes. Dado este resultado, as previsões a ter em conta são então as mostradas pelo modelo que apresentou os melhores resultados.

O teste de Box-Ljung (1978) foi utilizado como ferramenta de diagnóstico à capacidade de ajuste do modelo. Este teste é aplicado aos resíduos da série temporal, depois de ter sido ajustado o modelo que obteve melhores resultados (ARIMA, cuja parte sazonal é explicada por termos da série de Fourier). Através da análise à presença de autocorrelação dos resíduos, os resultados obtidos por este teste permitem então perceber se o modelo é apropriado.

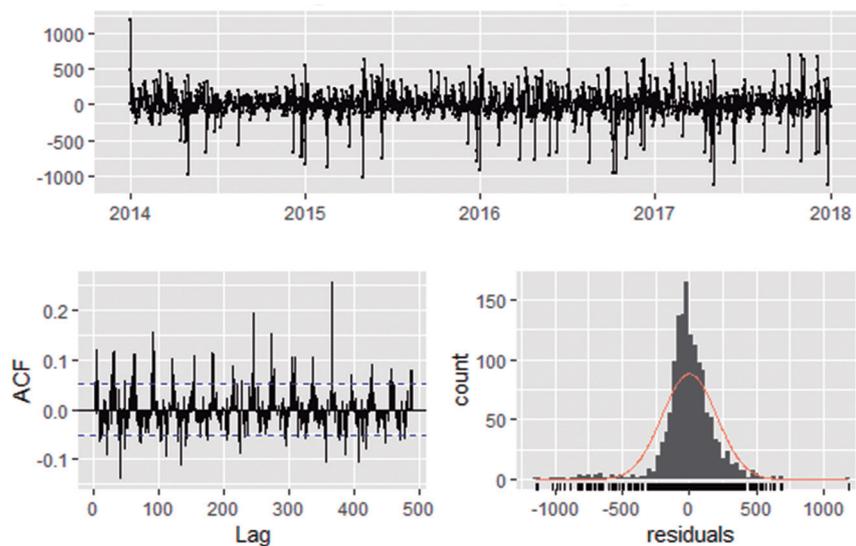


Figura 6: Diagnóstico aos Resíduos da Regressão Harmónica Dinâmica

A estatística de Ljung-Box, $Q=1089,90$ com um valor-p perto de 0, sugere que o modelo poderá não ser apropriado para fazer a modelação da série. Este resultado (rejeição de H_0) demonstra que há muita informação que não foi capturada com este modelo, ainda que produza previsões razoáveis.

5. Conclusão

Como abordado em capítulos anteriores, as conclusões obtidas serão divididas pelos objetivos propostos no presente estudo.

Quanto ao primeiro objetivo, o estudo realizado veio confirmar algumas ideias já pensadas, também estas obtidas em estudos anteriores, e trouxe outras conclusões relevantes. Com o passar dos anos, o número máximo de contactos tem vindo a aumentar e uma pequena parte dos clientes chega a ligar mais do que 6 vezes por ano.

Na caracterização dos clientes que ligam para o Oney verificou-se que, em média, têm um agregado familiar reduzido e, conseqüentemente, o número de filhos é também reduzido. O rendimento mensal médio destes clientes encontra-se na casa dos 1.000€. A idade média dos clientes que efetuam as chamadas é semelhante à idade média da carteira global do Oney, que se encontra perto dos 47 anos. Observa-se que quase metade das ligações (46%) são efetuadas por clientes cujo estado civil é casado. Ao nível demográfico observa-se uma elevada percentagem de chamadas nas áreas da Grande Lisboa/Grande Porto e Lisboa. A frequência de chamada aumenta com a antiguidade do cliente e são os clientes com comportamento de compra revolving ou sem juros que efetuam mais chamadas.

Para responder ao segundo objetivo, recorreremos ao auxílio do diagrama de chord que permite obter conclusões muito importantes. Através desta análise foi possível verificar que o motivo informação sobre saldos é que o tem maior peso (21%), seguido dos outros motivos (18%) e informações/ pedidos diversos (16%). A interpretação deste diagrama é essencial para adquirir conclusões ao nível operacional, tais como: sempre que um cliente liga para o Oney por qualquer motivo, dever-se-ia, por exemplo, enviar uma mensagem com o saldo correspondente; isto porque verificou-se que todos os clientes que ligam para um dos motivos voltam sempre a ligar, e com uma percentagem considerável, a pedir informações sobre saldos. Quando se faz a mesma análise repartida pelos anos, é possível verificar que o motivo 2 (informações contratuais) ganhou bastante peso no ano 2017.

Para o último objetivo analisou-se e discutiu-se a contribuição dos modelos de HW e ARIMA, com dupla sazonalidade para traduzir os dois efeitos sazonais – um ciclo de 365,25 dias e um ciclo de 7 dias que retrata os ciclos dentro da semana, - para previsão do número de chamadas efetuadas para o *call center* do Oney para os primeiros 7 meses do ano de 2018. A comparação dos resultados dos modelos fez-se através do MSE e escolheu-se aquele que possuía um valor menor. Como se pode ver na figura abaixo o modelo ARIMA, cujo padrão sazonal foi modelado com termos da série de Fourier é o que produz, em todos os momentos do período de previsão, um menor MSE.

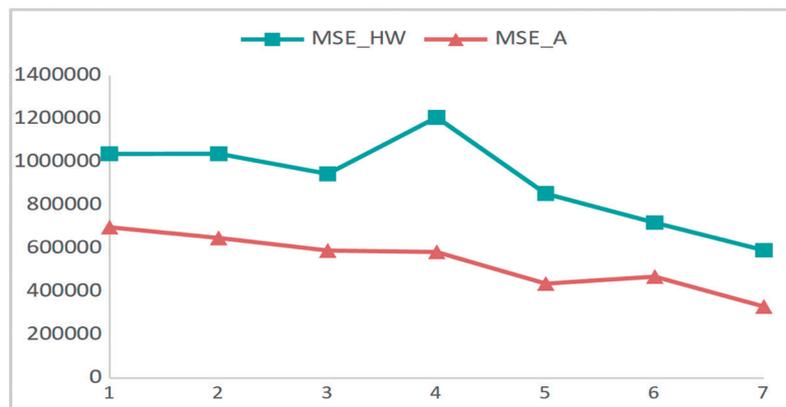


Figura 7: Comparação dos métodos aplicados para o horizonte de previsão.

Aplicação do modelo ao nível operacional

Para além do resultado estatístico apresentado, este modelo poderá também permitir a implementação de uma previsão para o fluxo de chamadas diárias. Ao nível operacional, este resultado permite avaliar a carga de trabalho das equipas de *call center* e, conseqüentemente, contribuir para o sucesso do negócio.

A sua implementação permite prever para o mês seguinte a necessidade de recursos a ter disponível promovendo a eficácia operacional.

Outro fator importante é que a integração consecutiva de novos dados no modelo, poderá vir a melhorar a performance de previsão do mesmo pois vamos enriquecendo o ajuste dos dados. Por outro lado, caso existam novos padrões de comportamento no sistema de chamadas (por exemplo: novos produtos, diferentes tipologias de cliente, campanhas de marketing específicas, entre outras), poderá existir necessidade de rever o modelo e fazer o acompanhamento do desvio entre a previsão e os dados reais.

5.1 Contributos

A concretização do primeiro objetivo traz a possibilidade de traçar os perfis de clientes e por seguinte antecipar com maior facilidade o volume de chamadas, visto que estas se encontram num leque de clientes mais reduzido. O segundo objetivo – perceber os principais motivos de contacto do cliente com a empresa - veio complementar o primeiro, no sentido em que traz um maior conhecimento do cliente e as vantagens de conhecer detalhadamente o cliente já foram aprofundadas neste estudo. A realização do terceiro objetivo – obtenção de resultados de previsão – são fundamentais para a tomada de decisão da empresa.

5.2 Limitações

Neste estudo verificou-se uma grande limitação em trabalhar com dados de empresas: a maioria dos dados são gerados através do desenvolvimento de vários sistemas de informação e alimentados por diversas fontes. Em outubro 2016 a plataforma de registo de chamadas alterou e perdeu-se parte dos registos nesse mês, em certos dias ao longo do período de análise a plataforma teve incidentes; ora, estes são acontecimentos que influenciam a modelação e previsão.

5.3 Investigação Futura

Em pesquisas futuras, seria interessante investigar a dependência intra-diária, fornecendo então uma alternativa ao vetor de alisamento exponencial. Seria também interessante alargar a investigação de métodos univariados e multivariados que incorporassem possíveis fatores (como por exemplo, o motivo da chamada) que possam influenciar o número de chamadas recebidas no *call center* e ajudassem a melhorar a precisão dos modelos. Outro ponto de interesse passa por obter um melhor ajustamento, de modo a aprimorar os resultados da estatística de Ljung-Box e, ao mesmo tempo, obter previsões mais fiáveis.

6. Referências Bibliográficas

Avramidis, PL, Deslauriers, ANA (2004). *Modeling Daily Arrivals to a Telephone Call Center*. Management Science 50 (7): 896–908.

Bianchi, RCH & Jarrett, LJ (1998). *Improving forecasting for telemarketing centers by ARIMA modelling with intervention*. International Journal of Forecasting, 14(4): 497–504.

Billah, B, Hyndman, RJ and Koehler, AB (2005). *Empirical information criteria for time series forecasting model selection*. Journal of Statistical Computation and Simulation 522 (75). 831–840.

Box, GEP & Jenkins, GM (1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Burnham, KP & Anderson DR (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. 2nd Edition. Springer-Verlag.

Caiado, J (2011). *Métodos de Previsão em Gestão com Aplicações em Excel*. 1ª Ed. Lisboa: Edições Sílabo.

Chatfield, C (1978). *The Holt-Winters forecasting procedures*. Applied Statistics 27. 264–279.

De Livera, AM, Hyndman, RJ and Snyder, RD (2011). *Forecasting time series with complex seasonal patterns using exponential smoothing*. Journal of the American Statistical Association 106 (496). 1513-1527.

Dixon, WJ (1960). *Simplified Estimation from Censored Normal Samples*. The Annals of Mathematical Statistics, 31(2): 385-391.

Gans, N, Koole, G and Mandelbaum A (2003). *Telephone call centers: Tutorial, review and research prospects*. Manufacturing & Service Operations Management 5 (2). 79–141.

Gardner, ES (2006). *Exponential smoothing: The state of the art - Part II*. International Journal of Forecasting 22. 637–666.

Gardner, ES & McKenzie, E (1985). *Forecasting Trends in Time Series*. Management Science 31. 1237-1246.

Gilchrist, W (1976). *Statistical Forecasting*. Wiley: Chichester.

Harvey, A (1989). *Forecasting structural time series models and the Kalman filter*. Cambridge University Press.

Hyndman, RJ & Athanasopoulos G (2014). *Optimally reconciling forecasts in a hierarchy*. Foresight 35 (Fall, 2014). 42-48.

Hyndman, RJ, Koehler, AB, Snyder, RD and Grose, S (2002). *A state space framework for automatic forecasting using exponential smoothing methods*. International Journal of Forecasting 18 (3). 439–454.

Jongbloed, G & Koole, G (2001). *Managing uncertainty in call centers using Poisson mixtures*. Applied Stochastic Models in Business and Industry, 17: 1–12 e 307–318.

Makridakis, S and Winkler, R (1983). *Average of forecasts: Some empirical results*. Management Science 29(9). 987-996.

Mehrotra, V, Ross, K, Ryder, G and Zhou, YP (2012). *Routing to Manage Resolution and Waiting Time in Call Centers with Heterogeneous Servers*. Manufacturing & Service Operations Management 14 (1). 66–81.

- Rayner, JN (2001). *International Encyclopedia of the Social & Behavioral Sciences*.
- Reid, DJ (1975). A review of short-term projection techniques. Gordon HA (ed). Practical aspects of forecasting. *Operational Research Society: London*. 825.
- Steckley, SG & Henderson, SG (2004). *Service system planning in the presence of a random arrival rate*. Technical Report, Cornell University, Ithaca: 1–32.
- Taylor, JW (2008). *A Comparison of Univariate Time Series Methods for Forecasting Intraday Arrivals at a Call Center*. *Management Science* 54(2): 253–265.
- Taylor, JW (2003). *Short-term electricity demand forecasting using double seasonal exponential smoothing*. *Journal of the Operational Research Society* 54 (8). 799-805.
- Weinberg, J, Brown, LD and Stroud, JR (2006). *Bayesian Forecasting of an Inhomogeneous Poisson Process with Applications to Call Center Data*. *Jour. Amer. Statist. Assoc.*: 1–29.
- West, M & Harrison, J (1997). *Bayesian forecasting and dynamic models*. 2nd Edition. Springer-Verlag, New York.
- Winsor, JCP & Hastings (1947). *Low Moments for Small Samples: A Comparative Study of Order Statistics*. *Annals of Mathematical Statistics*, 18: 413-426.

Anexos

Anexo A

Análise e tratamento das variáveis

A.1. Variáveis quantitativas: Winsorização

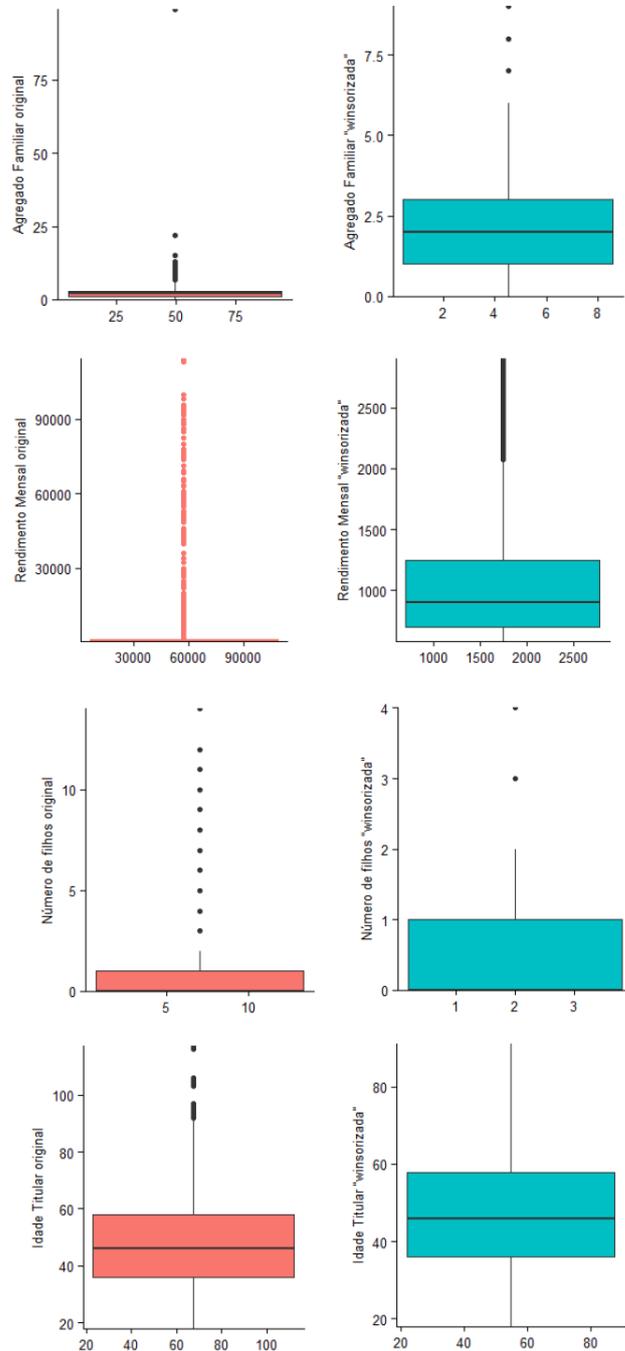


Figura A.1.1. Winsorização das variáveis do agregado familiar, rendimento mensal, número de filhos e idade.

A.2. Variáveis Qualitativas: Frequências

Tabela A.2.1 - Frequência por estado civil

	Frequência
Casado	543.430
Solteiro	265.954
Viúvo	43.530
Divorciado	134.884
Outros	196.085

Tabela A.2.2 - Frequência por antiguidade cliente

	Frequência
<1ano	855
1a2anos	18.953
2a4anos	121.362
4a6anos	221.702
>=6anos	821.011

Tabela A.2.3 - Frequência por região

	Frequência
Desconhecida	83.621
Alentejo	31.197
Algarve	61.180
Centro	154.850
Grande Lisboa	362.307
Grande Porto	188.653
Lisboa	230.107
Norte	69.985
RA Madeira	1.383
RA Açores	600

Tabela A.2.4 - Frequência por categoria profissional

	Frequência
Desconhecida	408.244
Desempregado	9.848
Doméstica	2.896
Estudante	5.753
Reformado	101.450
Outras	655.692

Tabela A.2.5 - Frequência por posse de telemóvel

	Frequência
Desconhecido	83.134
Tem telemóvel	1.090.610
Não tem telemóvel	10.139

Tabela A.2.6 - Frequência por modalidade de compra dos últimos 12M

	Frequência
Com Juros	179.314
DImediato	21.911
Inactivo	176.308
Revolving	462.772
Sem Juros	343.578

Tabela A.2.7 - Frequência por género

	Frequência
Desconhecido	83.134
Feminino	661.550
Masculino	439.199

Tabela A.2.8 - Frequência por posse de e-mail

	Frequência
Desconhecido	83.134
Tem email	786.418
Não tem email	314.331

A.3. Correspondência Motivo Geral e Motivo Particular

Motivo.geral	Motivo.particular	Motivo.agrupado	Total_Registos
Anomalia	Anomalia do Contrato/Validação	Outro	2619
Anomalia	Anomalia do Sistema	Outro	295
Anomalia	Saldos Creedores	Info Saldos	269
Automatric	Abastecimento	Info/ Pedidos Diversos	138
Automatric	Website Automatric	Info/ Pedidos Diversos	171
Cartão Clube Descontos	Esclarecimentos	Info/ Pedidos Diversos	2
Cartão Clube Descontos	Pedido Cartão	Info/ Pedidos Diversos	3
Cartão Decathlon encerramento	NA	Outro	155
Esclarecimento	Alteração TAEG 10-2013	Info/ Pedidos Diversos	8
Esclarecimento	Alteração Taxas	Info/ Pedidos Diversos	128
Esclarecimento	Bloqueio Contrato	Info/ Pedidos Diversos	4319
Esclarecimento	Bloqueio de Cartão	Info/ Pedidos Diversos	5006
Esclarecimento	Comissão Inactividade	Info Contratual	2825
Esclarecimento	Conta Corrente	Pedido Financiamento	52765
Esclarecimento	Extracto	Esclarecimento Extrato	51066
Esclarecimento	IRS	Info/ Pedidos Diversos	404
Esclarecimento	Pagamento Efectuado	Info Pagamentos	15676
Esclarecimento	Site	Info/ Pedidos Diversos	3145
Esclarecimento	Sucursal	Info Contratual	445
Incidente	Débitos diferidos em duplicado	Outro	132
Incidente	Ocorrência SEPA Março 2014	Outro	8174
Informação	NA	Info/ Pedidos Diversos	2
Informação	Alteração IBAN	Info/ Pedidos Diversos	487
Informação	Alteração PIN	Info/ Pedidos Diversos	168
Informação	Alteração de Dados Obrigatórios	Info Contratual	653
Informação	Carregamento Conta Mais	Info Contratual	51
Informação	Comercial	Info/ Pedidos Diversos	17581
Informação	Como fazer pagamentos	Info Pagamentos	8399
Informação	Comprovativo de pagamento	Info Pagamentos	121
Informação	Contratual	Info Contratual	75677
Informação	Devolução de Artigo/Anulação de Compra	Info/ Pedidos Diversos	480
Informação	IBAN	Info/ Pedidos Diversos	1757
Informação	Modalidades Compra	Info/ Pedidos Diversos	2311
Informação	Modalidades Pagamento	Info Pagamentos	1808
Informação	Oney Contacto	Info/ Pedidos Diversos	3866
Informação	Pagamento-Entidade e Referência	Info Pagamentos	27636
Informação	Participação Sinistro	Info/ Pedidos Diversos	536
Informação	Recepção do Extracto	Esclarecimento Extrato	447
Informação	Roubo, perda, extravio do cartão	Info/ Pedidos Diversos	980
Informação	Valor TB	Info Saldos	191251
Informação	Valor disponível	Info Saldos	78560
Informação	Valor em dívida	Info Saldos	112059
Informação	Área privada Site	Info/ Pedidos Diversos	13525
Inquérito	Boas-vindas	Outro	380
Outro	NA	Outro	87398
Pedido	2ª via de documento	Info/ Pedidos Diversos	1004
Pedido	Acordo - Cliente normal	Info Contratual	65
Pedido	Activação cartão visa	Info/ Pedidos Diversos	1175
Pedido	Adesão	Info/ Pedidos Diversos	3924
Pedido	Adesão Seguro	Info/ Pedidos Diversos	902

Motivo.geral	Motivo.particular	Motivo.agrupado	Total_Registos
Pedido	Adesão Site	Info/ Pedidos Diversos	600
Pedido	Alteração FM-CP	Info Contratual	41554
Pedido	Alteração Dados Pessoais	Info/ Pedidos Diversos	5991
Pedido	Alteração IBAN	Info/ Pedidos Diversos	4331
Pedido	Alteração Modalidade Compra	Info/ Pedidos Diversos	699
Pedido	Alteração Modalidade Reembolso	Info/ Pedidos Diversos	47
Pedido	Alteração Modo Pagamento	Info Contratual	1185
Pedido	Alteração StopMail	Info/ Pedidos Diversos	297
Pedido	Alteração de Seguro	Info/ Pedidos Diversos	121
Pedido	Amortização total ou Pacial	Info/ Pedidos Diversos	1741
Pedido	Anexação 2º Titular	Info Contratual	206
Pedido	Anulação de Contrato	Info Contratual	17271
Pedido	Anulação de Seguro	Info Contratual	15055
Pedido	Anulação de Serviço	Info Contratual	62
Pedido	Ativação Cliente	Info/ Pedidos Diversos	3
Pedido	Atribuição Cartão - 2º Titular	Info/ Pedidos Diversos	20
Pedido	Aumento Plafond	Info/ Pedidos Diversos	10547
Pedido	Autorização de compra	Info/ Pedidos Diversos	15278
Pedido	Bloqueio/desbloqueio cartão	Info/ Pedidos Diversos	13232
Pedido	Declaração	Info/ Pedidos Diversos	4207
Pedido	Desanexação 2º titular	Info Contratual	485
Pedido	Documentos Plafond	Info/ Pedidos Diversos	40
Pedido	Financiamento	Pedido Financiamento	124716
Pedido	Oney Contacto	Info/ Pedidos Diversos	19946
Pedido	Redução Plafond	Info/ Pedidos Diversos	49
Pedido	Reembolso	Info/ Pedidos Diversos	15209
Pedido	Reemissão PIN Jumbo visa	Info/ Pedidos Diversos	427
Pedido	Retenção	Info/ Pedidos Diversos	52
Pedido	Retenção Seguro	Info/ Pedidos Diversos	8
Pedido	Substituição Cartão	Info/ Pedidos Diversos	9386
Pedido	Transformação J.Mais->J. Visa	Info Contratual	96
Pedido	Transformação J.Oxi->J. Mais	Info Contratual	287
Pedido	NA	Info/ Pedidos Diversos	4
Reclamação/ Sugestão	NA	Reclamações	2915
Resposta a Contacto	Campanha	Outro	809
Resposta a Contacto	Carta SEPA	Outro	24
Resposta a Contacto	Conversão extracto digital	Esclarecimento Extrato	4410
Resposta a Contacto	NPAI	Outro	2059
Resposta a Contacto	Outros	Outro	3101
Resposta a Contacto	Redução Plafond	Info/ Pedidos Diversos	25
Seguimento Pedido	2ª via de documento	Info/ Pedidos Diversos	16
Seguimento Pedido	Alteração de modo de pagamento	Info Contratual	62
Seguimento Pedido	Anulação de Compra	Info/ Pedidos Diversos	8
Seguimento Pedido	Anulação de Contrato	Info Contratual	308
Seguimento Pedido	Anulação de Seguro	Info Contratual	102
Seguimento Pedido	Carta ou Declaração	Outro	518
Seguimento Pedido	Consulta/Pesquisa de Valores	Info Saldos	60
Seguimento Pedido	Reclamação	Reclamações	74
Seguimento Pedido	Reclamação Site	Reclamações	173
Seguimento Pedido	Reembolso	Info/ Pedidos Diversos	521
Trabalho Administrativo	Callback	Outro	2582
Trabalho Administrativo	SMS	Outro	8983

Motivo.geral	Motivo.particular	Motivo.agrupado	Total_Registos
Transferência Contacto	Frágeis	Outro	1641
Transferência Contacto	Mediação Legal-Contencioso	Outro	72
Transferência Contacto	Mediação Legal-PréContencioso	Outro	2002
Transferência Contacto	Outros	Outro	309
Transferência Contacto	Recuperação Amigável	Outro	3738
Transferência Contacto	Recuperação Comercial	Outro	89935
Transferência Contacto	Telesales	Outro	737
Transferência Contacto	NA	Outro	1

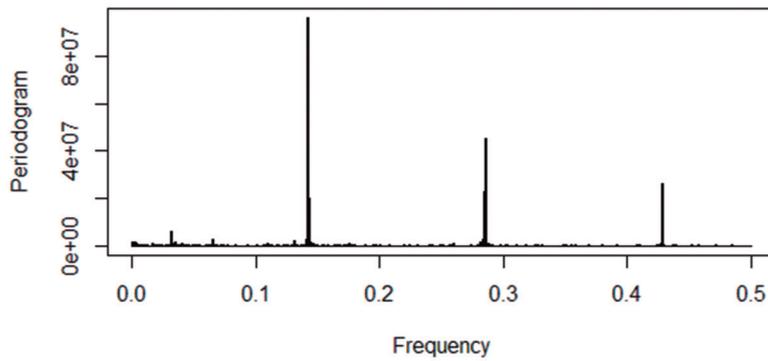
A.4. Correlação entre Motivos e cada umas das Variáveis Explicativas



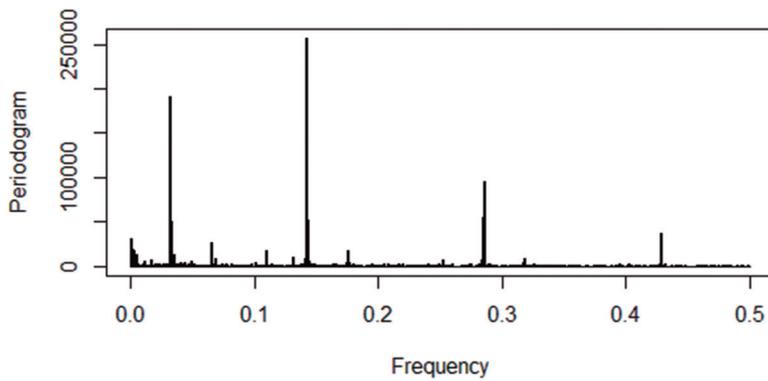
Anexo B

Análise e discussão dos Resultados

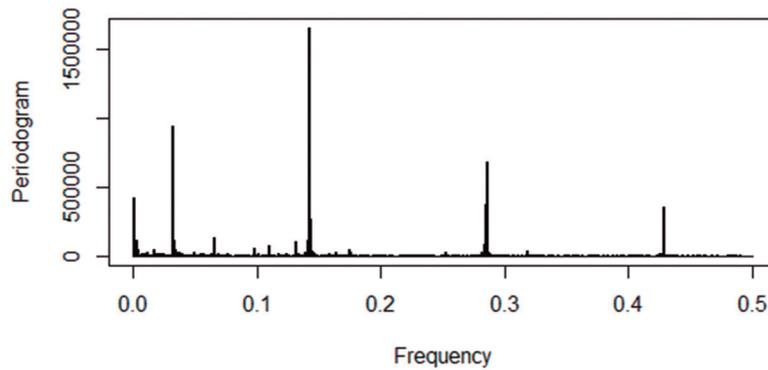
B.1. Análise Espectral



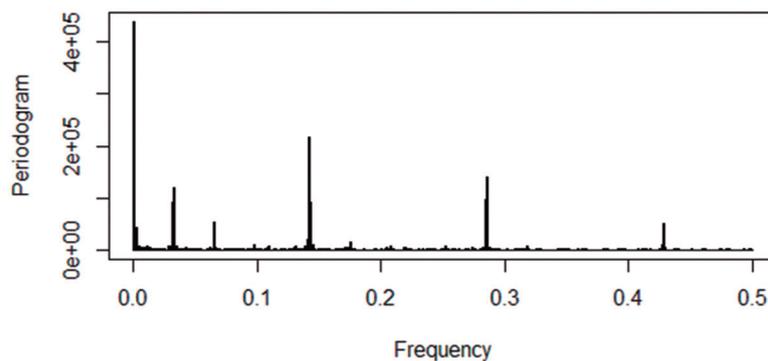
periodograma	frequência	ciclo
96092938	0,14	7
45453490	0,29	360
25920363	0,43	2
22969889	0,29	3
19968748	0,14	7
6224756	0,03	31
2806611	0,14	7



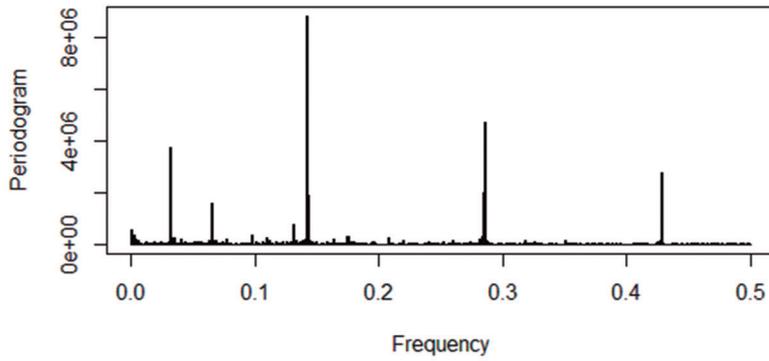
periodograma	frequência	ciclo
256639	0,14	7
190671	0,03	31
95602	0,29	3
55617	0,29	4
52413	0,14	7
50345	0,03	30
36828	0,43	2



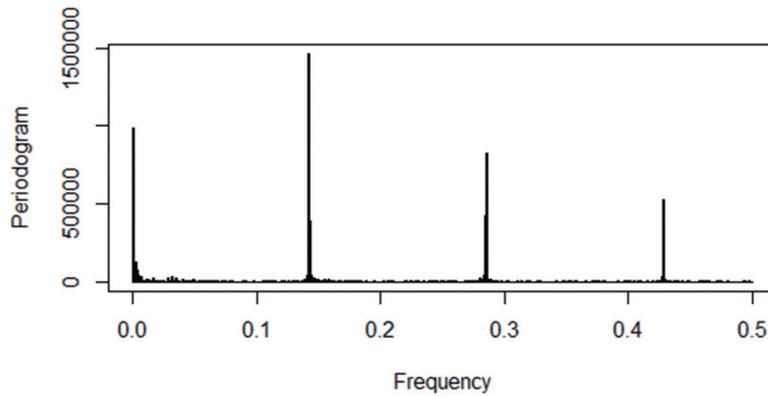
periodograma	frequência	ciclo
1650550	0,14	7
943305	0,03	31
678219	0,29	3
416557	0,00	1500
374431	0,29	4
354546	0,43	2
267325	0,14	7



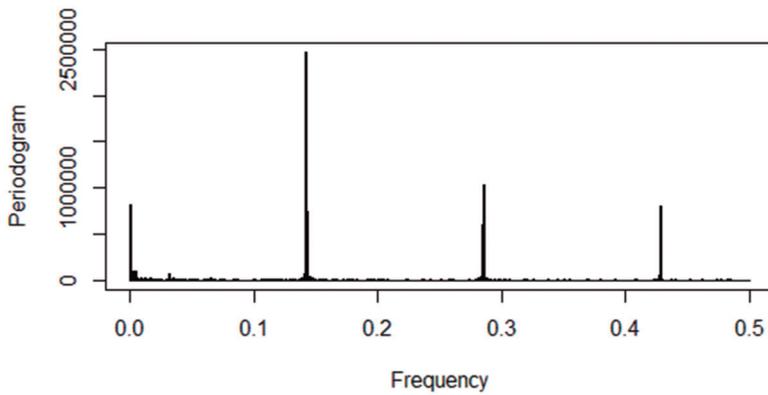
periodograma	frequência	ciclo
438946	0,00	1500
225972	0,00	365
217222	0,14	7
138870	0,29	3
118657	0,03	30
97128	0,29	4
91572	0,14	7



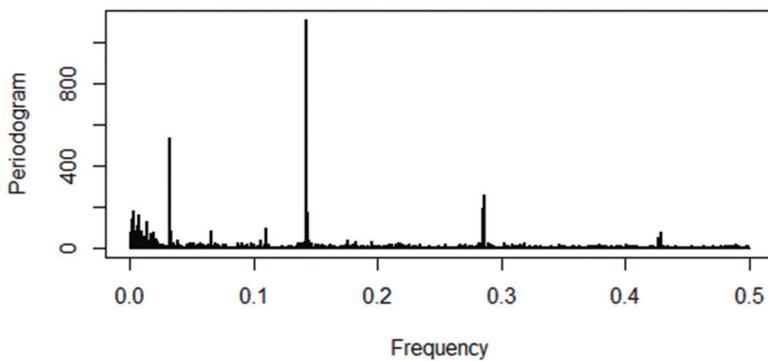
periodograma	frequência	ciclo
8796008	0,14	7
4701229	0,29	3
3722514	0,03	31
2747767	0,43	2
2005245	0,29	4
1889058	0,14	7
1549662	0,07	15



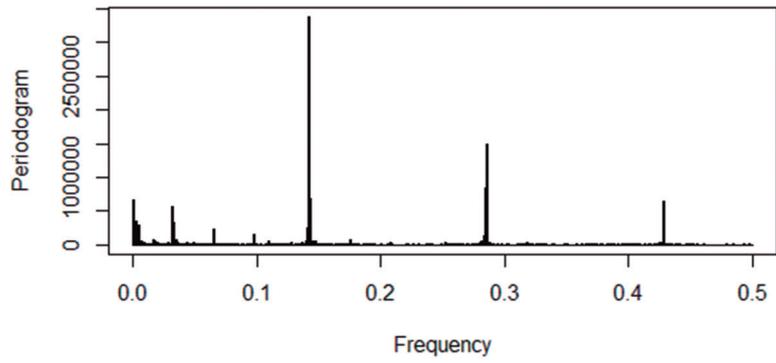
periodograma	frequência	ciclo
1462608	0,14	7
988655	0,00	1500
822097	0,29	358
528654	0,43	2
427185	0,29	4
393648	0,14	7
344956	0,00	3



periodograma	frequência	ciclo
2470118	0,14	7
1037266	0,29	3
812509	0,00	1500
795938	0,43	2
741743	0,14	7
603619	0,29	4
266128	0,00	750

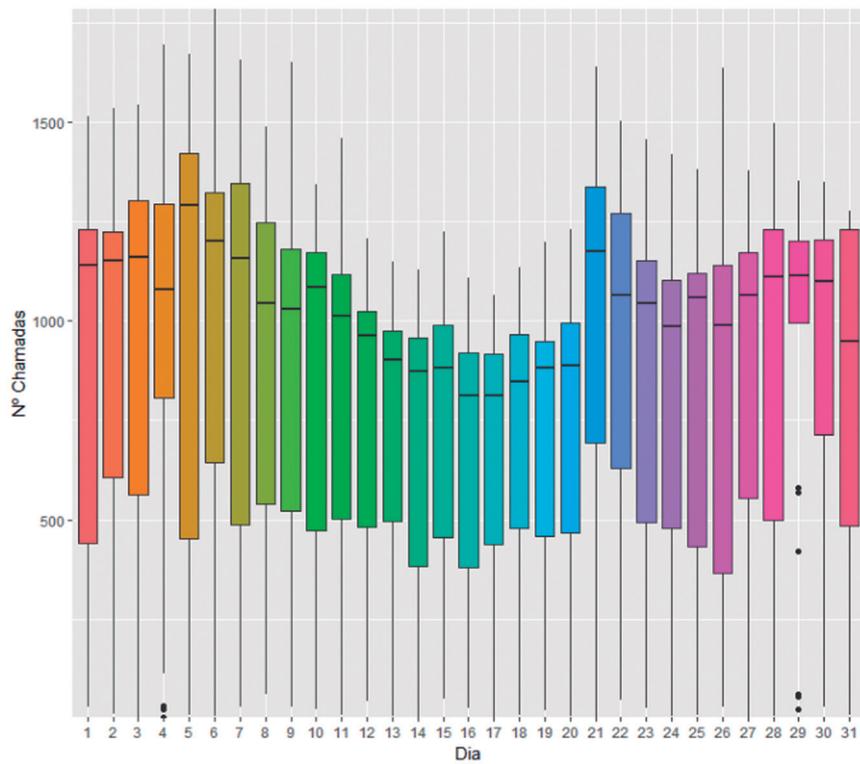
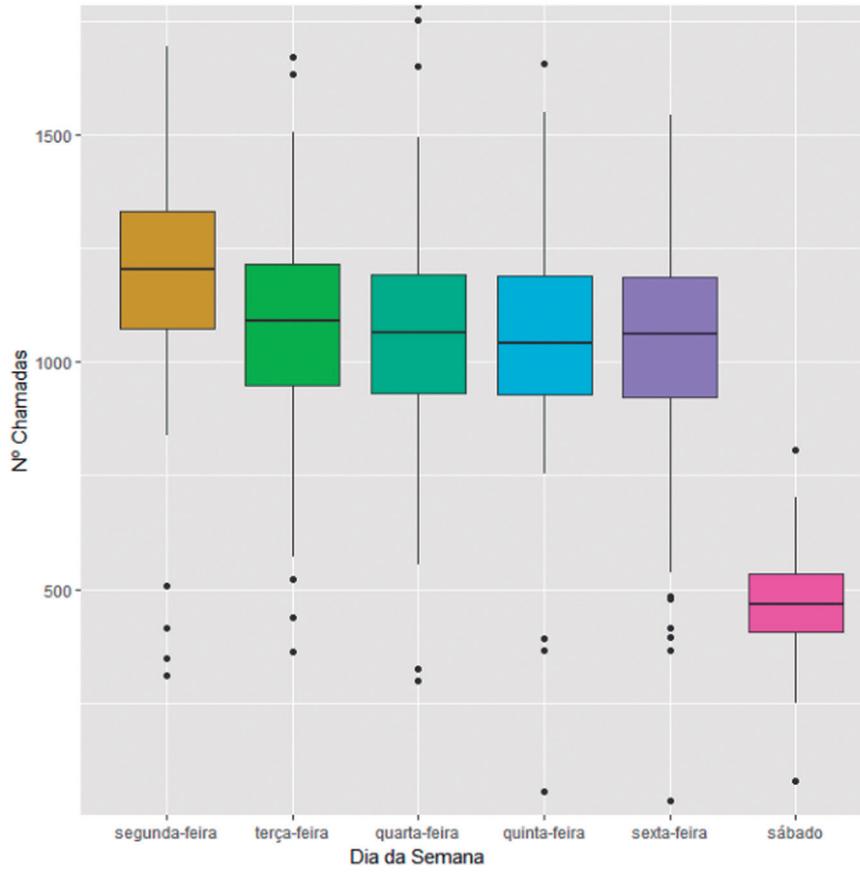


periodograma	frequência	ciclo
1112	0,14	7
533	0,03	31
257	0,29	3
192	0,29	4
178	0,00	300
172	0,14	7
161	0,01	125

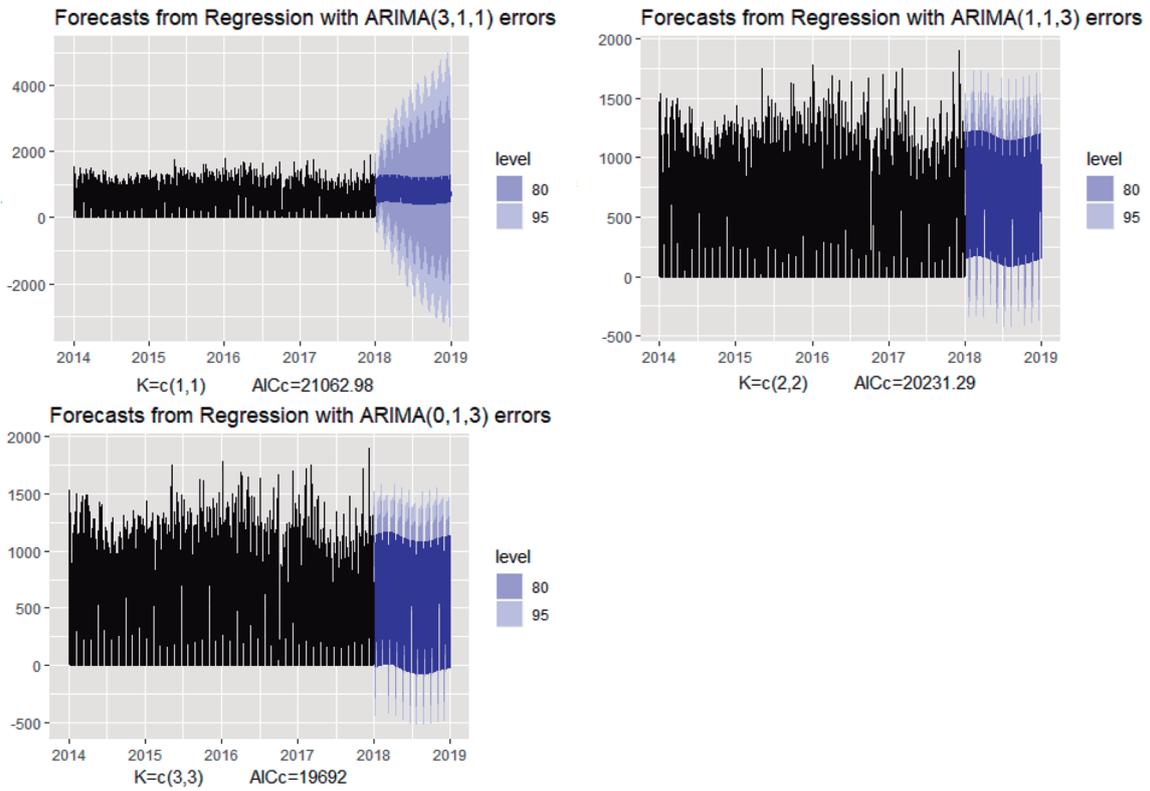


periodograma	frequência	ciclo
3382621	0,14	7
1488288	0,29	365
846627	0,29	4
686927	0,14	7
672046	0,00	1500
645849	0,43	2
568579	0,03	31

B.2 - Boxplot: Análise Sazonal Semanal e Mensal



B.3 - Comparação de Várias Regressões Harmônicas Dinâmicas com Períodos Sazonais Múltiplos



B.4 - Valores Reais e Previstos pelo Método HW e pelo Método ARIMA para a Base de Validação

Horizonte	Dia_semana	Observado	HW	A
2018-01-02	3 ^a feira	2536	736	1030
2018-01-03	4 ^a feira	2563	1238	1018
2018-01-04	5 ^a feira	2752	1370	993
2018-01-05	6 ^a feira	3033	1504	986
2018-01-06	Sabado	730	1239	418
2018-01-08	2 ^a feira	2567	229	1142
2018-01-09	3 ^a feira	1621	914	1038
2018-01-10	4 ^a feira	1667	979	1022
2018-01-11	5 ^a feira	1324	949	997
2018-01-12	6 ^a feira	1029	859	990
2018-01-13	Sabado	422	830	422
2018-01-15	2 ^a feira	1294	162	1146
2018-01-16	3 ^a feira	1106	594	1042
2018-01-17	4 ^a feira	1301	732	1027
2018-01-18	5 ^a feira	1160	821	1001
2018-01-19	6 ^a feira	988	819	994
2018-01-20	Sabado	472	1008	426
2018-01-22	2 ^a feira	2108	218	1151
2018-01-23	3 ^a feira	1602	766	1046
2018-01-24	4 ^a feira	1323	1090	1031
2018-01-25	5 ^a feira	1446	1150	1006
2018-01-26	6 ^a feira	1511	1137	999
2018-01-27	Sabado	601	1160	431
2018-01-29	2 ^a feira	1833	246	1155
2018-01-30	3 ^a feira	1768	893	1051
2018-01-31	4 ^a feira	1605	1097	1035
2018-02-01	5 ^a feira	1591	1193	1010
2018-02-02	6 ^a feira	2398	1442	1003
2018-02-03	Sabado	777	1454	435
2018-02-05	2 ^a feira	3100	218	1159
2018-02-06	3 ^a feira	3008	1173	1055
2018-02-07	4 ^a feira	2269	1277	1040
2018-02-08	5 ^a feira	1672	1101	1014
2018-02-09	6 ^a feira	1353	1157	1007
2018-02-10	Sabado	537	1114	439
2018-02-12	2 ^a feira	1585	153	1163
2018-02-13	3 ^a feira	54	732	1058
2018-02-14	4 ^a feira	1391	780	1043
2018-02-15	5 ^a feira	1083	733	1018
2018-02-16	6 ^a feira	1052	672	1010
2018-02-17	Sabado	418	840	442
2018-02-19	2 ^a feira	1201	146	1166
2018-02-20	3 ^a feira	1121	745	1061
2018-02-21	4 ^a feira	1646	1144	1046
2018-02-22	5 ^a feira	1751	1057	1020
2018-02-23	6 ^a feira	1490	1023	1013

Horizonte	Dia_semana	Observado	HW	A
2018-02-24	Sabado	606	973	444
2018-02-26	2 ^a feira	1876	202	1168
2018-02-27	3 ^a feira	1542	750	1063
2018-02-28	4 ^a feira	1612	641	1047
2018-03-01	5 ^a feira	2191	1071	1022
2018-03-02	6 ^a feira	1755	1158	1014
2018-03-03	Sabado	560	1171	445
2018-03-05	2 ^a feira	2903	202	1169
2018-03-06	3 ^a feira	3316	1171	1064
2018-03-07	4 ^a feira	2671	1202	1048
2018-03-08	5 ^a feira	2117	1243	1022
2018-03-09	6 ^a feira	1491	1105	1014
2018-03-10	Sabado	587	1051	445
2018-03-12	2 ^a feira	1536	120	1168
2018-03-13	3 ^a feira	1130	714	1063
2018-03-14	4 ^a feira	1070	762	1047
2018-03-15	5 ^a feira	1092	898	1021
2018-03-16	6 ^a feira	1028	875	1013
2018-03-17	Sabado	373	849	444
2018-03-19	2 ^a feira	1229	153	1167
2018-03-20	3 ^a feira	1138	793	1061
2018-03-21	4 ^a feira	1715	1211	1045
2018-03-22	5 ^a feira	1347	1272	1018
2018-03-23	6 ^a feira	1282	1079	1010
2018-03-24	Sabado	496	959	441
2018-03-26	2 ^a feira	1623	131	1163
2018-03-27	3 ^a feira	1443	685	1058
2018-03-28	4 ^a feira	1495	853	1041
2018-03-29	5 ^a feira	1410	957	1015
2018-03-30	6 ^a feira	421	1014	1006
2018-03-31	Sabado	444	934	437
2018-04-02	2 ^a feira	2022	123	1159
2018-04-03	3 ^a feira	1754	843	1053
2018-04-04	4 ^a feira	2168	1253	1037
2018-04-05	5 ^a feira	NA	1415	1010
2018-04-06	6 ^a feira	3010	1250	1001
2018-04-07	Sabado	746	1185	432
2018-04-09	2 ^a feira	2557	173	1154
2018-04-10	3 ^a feira	1876	770	1048
2018-04-11	4 ^a feira	1487	850	1031
2018-04-12	5 ^a feira	1209	835	1004
2018-04-13	6 ^a feira	1111	733	996
2018-04-14	Sabado	385	405	426
2018-04-16	2 ^a feira	1311	134	1148
2018-04-17	3 ^a feira	974	546	1041
2018-04-18	4 ^a feira	1055	641	1025
2018-04-19	5 ^a feira	1051	776	998
2018-04-20	6 ^a feira	1093	826	989
2018-04-21	Sabado	818	1129	420
2018-04-23	2 ^a feira	2014	175	1141
2018-04-24	3 ^a feira	1629	701	1035
2018-04-25	4 ^a feira	416	187	1018

Horizonte	Dia_semana	Observado	HW	A
2018-04-26	5 ^a feira	2006	1010	991
2018-04-27	6 ^a feira	1557	1011	982
2018-04-28	Sabado	514	1041	413
2018-04-30	2 ^a feira	2106	64	1134
2018-05-02	4 ^a feira	1999	1080	1011
2018-05-03	5 ^a feira	2059	1195	984
2018-05-04	6 ^a feira	1858	1241	976
2018-05-05	Sabado	593	1452	406
2018-05-07	2 ^a feira	2814	264	1128
2018-05-08	3 ^a feira	2453	752	1021
2018-05-09	4 ^a feira	1646	1057	1005
2018-05-10	5 ^a feira	1397	1096	978
2018-05-11	6 ^a feira	1152	885	969
2018-05-12	Sabado	372	716	400
2018-05-14	2 ^a feira	1273	177	1121
2018-05-15	3 ^a feira	1093	500	1015
2018-05-16	4 ^a feira	1018	627	998
2018-05-17	5 ^a feira	942	802	972
2018-05-18	6 ^a feira	914	718	963
2018-05-19	Sabado	314	916	393
2018-05-21	2 ^a feira	1995	273	1115
2018-05-22	3 ^a feira	1592	763	1009
2018-05-23	4 ^a feira	1507	942	993
2018-05-24	5 ^a feira	1599	1040	966
2018-05-25	6 ^a feira	1278	945	957
2018-05-26	Sabado	480	720	388
2018-05-28	2 ^a feira	1405	232	1110
2018-05-29	3 ^a feira	1423	668	1004
2018-05-30	4 ^a feira	1456	954	987
2018-05-31	5 ^a feira	255	1079	961
2018-06-01	6 ^a feira	1739	1071	952
2018-06-02	Sabado	461	1089	383
2018-06-04	2 ^a feira	NA	257	1105
2018-06-05	3 ^a feira	2673	842	999
2018-06-06	4 ^a feira	2675	1058	982
2018-06-07	5 ^a feira	2039	1015	956
2018-06-08	6 ^a feira	1658	1001	948
2018-06-09	Sabado	504	871	378
2018-06-11	2 ^a feira	1875	205	1100
2018-06-12	3 ^a feira	1318	613	995
2018-06-13	4 ^a feira	1175	667	978
2018-06-14	5 ^a feira	1154	327	952
2018-06-15	6 ^a feira	924	314	944
2018-06-16	Sabado	311	747	374
2018-06-18	2 ^a feira	1037	190	1097
2018-06-19	3 ^a feira	948	598	991
2018-06-20	4 ^a feira	989	780	975
2018-06-21	5 ^a feira	1667	1157	948
2018-06-22	6 ^a feira	1534	1078	940
2018-06-23	Sabado	492	923	371
2018-06-25	2 ^a feira	1667	232	1093
2018-06-26	3 ^a feira	1551	667	988

Horizonte	Dia_semana	Observado	HW	A
2018-06-27	4ªfeira	1522	914	971
2018-06-28	5ªfeira	1481	944	945
2018-06-29	6ªfeira	1239	1016	937
2018-06-30	Sabado	370	952	368
2018-07-02	2ªfeira	1630	260	1090
2018-07-03	3ªfeira	1748	766	985
2018-07-04	4ªfeira	1735	1156	968
2018-07-05	5ªfeira	2066	1347	942
2018-07-06	6ªfeira	1520	1262	934
2018-07-07	Sabado	476	1095	365
2018-07-09	2ªfeira	1730	243	1087
2018-07-10	3ªfeira	1263	746	982
2018-07-11	4ªfeira	1178	811	966
2018-07-12	5ªfeira	1010	801	939
2018-07-13	6ªfeira	937	824	931
2018-07-14	Sabado	301	731	362
2018-07-16	2ªfeira	NA	170	1085
2018-07-17	3ªfeira	949	633	979
2018-07-18	4ªfeira	993	683	963
2018-07-19	5ªfeira	909	713	937
2018-07-20	6ªfeira	941	880	929
2018-07-21	Sabado	708	1124	360
2018-07-23	2ªfeira	1658	218	1083
2018-07-24	3ªfeira	1698	700	977
2018-07-25	4ªfeira	1240	885	961
2018-07-26	5ªfeira	NA	894	935
2018-07-27	6ªfeira	1346	931	927
2018-07-28	Sabado	439	901	358
2018-07-30	2ªfeira	1941	254	1081
2018-07-31	3ªfeira	NA	848	976