

Previsão e Planeamento de Equipas- Workforce Management

Nelson Camões

Mestrado em Engenharia Matemática

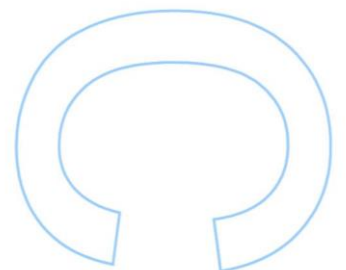
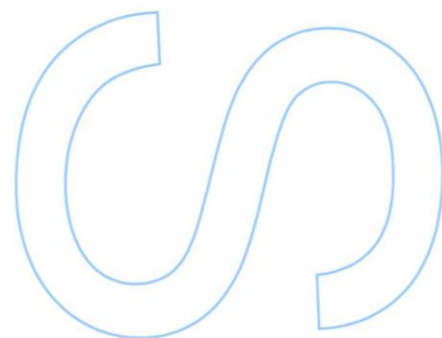
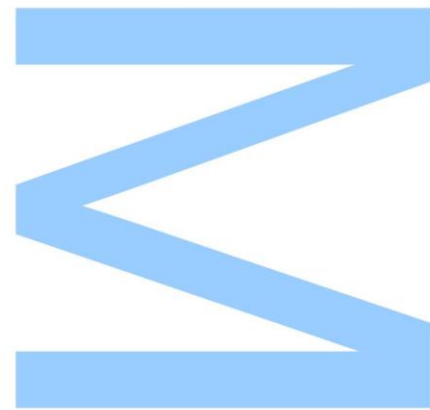
Departamento de Matemática da Faculdade de Ciências, Universidade do Porto
2022

Orientador

Prof. Dr. Teresa Mendonça, Faculdade de Ciências Universidade do Porto

Coorientador

Prof. Dra. Maria Eduarda Silva, Faculdade de Economia Universidade do Porto



U. PORTO

FC FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, ____/____/____

3

S

Q

UNIVERSIDADE DO PORTO

TESE DE MESTRADO

**Previsão e Planeamento de Equipas -
Workforce Management**

Author:

Nelson CAMÕES

Supervisor:

Joaquim SILVA

*A thesis submitted in fulfilment of the requirements
for the degree of Mestrado Engenharia Matemática*

at the

Faculdade de Ciências da Universidade do Porto
Departamento de Matemática

26 de novembro de 2022

Declaração de Honra

Eu, Nelson Miguel Brito Camões, inscrito no Mestrado em Engenharia Matemática da Faculdade de Ciências da Universidade do Porto declaro, nos termos do disposto na alínea a) do artigo 14.º do Código Ético de Conduta Académica da U.Porto, que o conteúdo da presente relatório de estágio reflete as perspetivas, o trabalho de investigação e as minhas interpretações no momento da sua entrega.

Ao entregar esta relatório de estágio, declaro, ainda, que a mesma é resultado do meu próprio trabalho de investigação e contém contributos que não foram utilizados previamente noutros trabalhos apresentados a esta ou outra instituição.

Mais declaro que todas as referências a outros autores respeitam escrupulosamente as regras da atribuição, encontrando-se devidamente citadas no corpo do texto e identificadas na secção de referências bibliográficas. Não são divulgados na presente relatório de estágio quaisquer conteúdos cuja reprodução esteja vedada por direitos de autor.

Tenho consciência de que a prática de plágio e auto-plágio constitui um ilícito académico.

Nelson Camões, 12 de Outubro de 2022

UNIVERSIDADE DO PORTO

Resumo

Faculdade de Ciências da Universidade do Porto

Departamento de Matemática

Mestrado em Engenharia Matemática

Previsão e Planeamento de Equipas - *Workforce Management*

por [Nelson CAMÕES](#)

Esta tese relata todo o trabalho realizado no âmbito de um estágio curricular do Mestrado em Engenharia Matemática na empresa FARFETCH.

O objetivo deste estágio é entender a complexidade do negócio da equipa de Operações para poder aplicar algoritmos de previsão a dados em forma de séries temporais e dessa forma prever o número de agentes que a empresa vai precisar para suportar todas as interações com o cliente (apoio ao cliente, envio de encomendas, pagamentos, etc) com o melhor desempenho possível.

Iremos recorrer a uma ampla variedade de modelos estatísticos como, por exemplo, as Redes Neurais, os *TBATS* e até ao *Prophet*, com o objetivo de abranger o máximo de abordagens de previsão e fazendo-as competir entre si utilizando 3 métricas de erro.

Numa fase final do relatório iremos explicar vários conceitos utilizados na indústria de *Workforce Management* de forma a construirmos um processo de cálculo de capacidade estruturado e transmissível dentro das várias equipas envolvidas na operação.

UNIVERSIDADE DO PORTO

Abstract

Faculdade de Ciências da Universidade do Porto

Departamento de Matemática

Mestrado Engenharia Matemática

Previsão e Planeamento de Equipas - Workforce Management

by [Nelson CAMÕES](#)

This thesis reports all the work carried out within the scope of a curricular internship of the Master in Mathematical Engineering at the company FARFETCH.

The purpose of this report is to understand the complexity of the Operations team's business in order to apply forecasting algorithms to data in the form of time series and thus predict the number of agents that the company will need to support all customer interactions (customer support, orders shipping, payments, etc.)

We will use a wide variety of statistical models, such as Neural Networks, TBATS and even Prophet, in order to cover as many prediction approaches as possible and make them compete between each others using 3 error metrics.

In a final phase of the report we will explain several concepts used in the Workforce Management industry in order to build a structured and transferable capacity calculation process within the various teams involved in the operation.

Conteúdo

Resumo	iv
Abstract	v
Conteúdo	vii
Lista de Figuras	ix
Glossário	xi
1 Introdução	1
1.1 Empresa e Equipa	1
1.2 Motivação e Objetivos	2
2 Métodos de Previsão de Séries Temporais	5
2.1 Decomposição de Séries Temporais	6
2.2 Modelos Matemáticos	7
2.2.1 <i>Naive</i> e <i>Seasonal Naive</i>	7
2.2.2 <i>Exponential Smoothing</i>	8
2.2.3 <i>ARIMA</i>	8
2.2.4 <i>TBATS</i>	9
2.2.5 <i>Prophet</i>	10
2.2.6 <i>Redes Neurais</i>	11
2.3 Modelo de Previsão Financeiro	11
2.4 Medidas de Erro	12
3 Pré-processamento dos Dados	15
3.1 Obtenção e Organização dos Dados	15
3.2 Tratamento de <i>Outliers</i>	16
4 Estrutura do processo de previsão	19
4.1 Apresentação da <i>framework</i> construída	19
4.2 Comparação de modelos e validação das métricas de erro apresentadas	21
4.3 Aplicação do processo de previsão a um exemplo real	26
5 Planeamento de Capacidade	29
5.1 Distribuições Horárias de Volumes	29

5.2	Pressupostos	30
5.3	Output de Capacidade Final	32
6	Conclusão	35
A	Apêndice geral	37
A.1	Explicação de termos e conceitos utilizados	37
	Bibliografia	39

Lista de Figuras

3.1	Evolução do Workload da linha de contacto BWP	18
3.2	Evolução do Workload da linha de contacto HPTI	18
4.1	<i>FPS Script Framework</i>	20
4.2	Erros obtidos utilizando o <i>MAE</i>	22
4.3	Erros obtidos utilizando o <i>MAPE</i>	23
4.4	Erros obtidos utilizando o <i>WAPE</i>	23
4.5	Resultados de Previsão (2022) da linha de contacto BWP	26

Glossário

WFM	<i>Workforce Management</i>
FPS	<i>FARFETCH Platform Solutions</i>
WL	<i>Workload</i> - carga de trabalho
BO	<i>Boutique Orders</i> - encomendas por boutique
WPO	<i>Workload per order</i> - carga de trabalho por encomenda
IQR	<i>Inter-quartile range</i> - intervalo interquartis
AHT	<i>Average handling time</i> - tempo médio por unidade de <i>workload</i>
OCP	Ocupação
SHRK	<i>Shrinkage In Office</i> - tempo despendido dentro do horário de trabalho
ABS	Absentismo
ATR	<i>Attrition</i> - saídas inesperadas de colaboradores
THR	<i>Thresholds</i> - férias a ser gozadas pelos colaboradores
MAE	Média dos erros absolutos
MAPE	Média dos erros absolutos em valores percentuais
WAPE	Média ponderada dos erros absolutos em valores percentuais
FTE	<i>Full Time Employee</i> - Colaborador a tempo inteiro (8 horas)

Capítulo 1

Introdução

1.1 Empresa e Equipa

Esta tese relata todo o trabalho desenvolvido no estágio realizado na empresa *FARFETCH* no âmbito do Mestrado em Engenharia Matemática na Faculdade de Ciências da Universidade do Porto.

Início com uma pequena apresentação da *FARFETCH*, dos seus ideais e da sua missão. A *FARFETCH* existe pela paixão pela moda desde 2007 e é neste momento rainha na sua indústria. Tem como missão ser uma plataforma global online de moda de luxo, conectando estilistas, curadores e clientes. Para além disso, a *Farfetch* também presta serviços a outras marcas de renome de forma a que estas mesmas possam também estar incluídas no *e-commerce* de luxo.

Apesar da sua jovem idade como empresa, a *FARFETCH* é já uma empresa bem estruturada com um conjunto de departamentos que trabalham em conjunto para atingir um objetivo: fornecer a melhor experiência aos clientes que visitam a nossa plataforma online. Um dos departamentos que luta por este objetivo é o departamento de Operações. Neste departamento somos apaixonados por excelência operacional, sempre com o foco nos nossos clientes e parceiros. Cobrindo todo o percurso de uma encomenda, criamos o nosso conteúdo online, pagamentos seguros e deteções de fraude, asseguramos o sucesso da encomenda e da sua respetiva entrega, tudo isto prestando sempre o melhor atendimento tanto ao cliente como ao parceiro. Em suma, o nosso objetivo é fornecer um experiência de luxo memorável a todos os nossos clientes por todo o mundo.

Dentro do departamento de Operações, encontramos novamente uma organização bem estruturada com os objetivos já referidos. Mais especificamente, um dos mais importantes ramos deste departamento é *Workforce Management (WFM)*, no qual eu estou incluído.

WFM é um conjunto de processos integrados que são usados para otimizar a produtividade, eficiência e consistência a um nível individual e departamental. Sendo uma equipa de suporte, o nosso objetivo é suportar o lado operacional do negócio, incluindo Atendimento ao Cliente, Atendimento ao Parceiro, Operação de Encomendas e outras áreas de negócio dentro da *FARFETCH*. Somos uma equipa global que trabalha 24 horas, 7 dias por semana por várias regiões no mundo incluindo Porto, Lisboa, Londres, Los Angeles, Tokyo, São Paulo, Shanghai e Hong Kong.

Se formos ainda mais a fundo dentro dos diferentes conjuntos de departamentos que a *FARFETCH* apresenta, encontramos a equipa onde estou incluído: Forecast & Planning - Previsão e Planeamento de Equipas. As nossas principais responsabilidades são prever a evolução de toda a operação envolvida no negócio de forma a podermos gerir a capacidade das várias equipas desde o atendimento ao cliente até ao processamento das encomendas.

1.2 Motivação e Objetivos

O meu projeto de estágio centrou-se na previsão e planeamento de várias equipas que constituem uma área de serviço da empresa - *FARFETCH Platform Solutions (FPS)* - que até então não tinha um processo de previsão estabelecido, devido à sua elevada complexidade em termos de negócio e escassez de dados. Esta área de serviço engloba todos os serviços prestados pela *FARFETCH* a marcas de renome que nos entregam a responsabilidade de as colocarmos no coração do *e-commerce* de luxo. Ou seja, *FPS* engloba tanto serviços de atendimento ao cliente como processamento de encomendas, entre outros, podendo ser visto como uma micro representação de todos os serviços já existentes na empresa.

Assim sendo, o objetivo do projeto foi construir um algoritmo robusto que englobaria tanto uma fase de pré-processamento dos dados existentes, como uma fase de previsão onde iremos percorrer várias métodos e ainda uma fase final que engloba o cálculo das

necessidades de capacidade previstas para estas equipas de forma a dimensionar corretamente estas mesmas com base no volume de trabalho previsto (unidades de *Workload* - WL).

As unidades de *workload* serão vistas como inserções de atividade numa plataforma de trabalho. Por exemplo, se um cliente liga com uma questão, o agente que recebeu a chamada contacta por email uma equipa de suporte para esclarecer uma dúvida, recebe o esclarecimento e depois liga ao cliente para dar a resposta final, neste caso são geradas 4 unidades de workload.

Apesar da explicação ser simples, o processo foge dessa realidade. Isto porque a quantidade e a qualidade de dados que temos disponíveis tornam os métodos de previsão difíceis de ser implementados e, por vezes, as suas previsões serão pouco fundamentadas estatisticamente. O grande objetivo deste relatório será investigar que tipo de procedimentos devemos adotar em situações de escassez e má qualidade de dados mensais na forma de séries temporais e ainda percorrer todo o processo de cálculo de capacidade, muito utilizado em WFM.

O processo de previsão pretendido tem como objetivo ser percorrido por várias séries temporais, cada uma contendo os volumes mensais de WL (ou seja, o volume de trabalho registado ao longo dos meses) de todas as marcas que estão associadas a FPS. Na realidade, o processo de previsão é ainda mais aprofundado, sendo feito para cada marca, canais de contacto associados (telefone, email ou chat) e ainda língua utilizada (Inglês, Francês, etc...). No final, podemos organizar cada série temporal como estando associada a uma linha de contacto. Por exemplo, a linha de contacto por telefone em Inglês da marca A terá a sua correspondente série temporal que representará o volume de trabalho que essa linha gerou ao longo dos meses.

Outro grande desafio com que me deparei é o facto de trabalharmos com diferentes séries temporais no que toca a comprimento, dimensão e comportamento. Isto é, tanto podemos encontrar uma série temporal com 3 anos de dados, com volumes significativos e com comportamentos sazonais vincados (o correspondente a 36 pontos dispostos ao longo do tempo com, por exemplo, uma média de 1000 unidades de WL por mês e correspondente a uma linha de uma marca já bem estabelecida no mercado), como poderemos encontrar uma série temporal sem sequer ter um ano de informação, com volumes mensais quase residuais e sem qualquer comportamento sazonal até à data (por exemplo,

uma linha ter apenas um histórico de 6 meses e uma média de 10 unidades de WL geradas por mês - o que pode facilmente corresponder a apenas 5 chamadas por mês). Esta grande discrepância no tipo de dados que temos disponíveis leva a uma maior cautela na escolha dos modelos a utilizar e leva-nos a uma separação na lógica do algoritmo que iremos ver mais à frente.

Por fim, o objetivo é transformar os volumes mensais de trabalho previstos em volumes horários (aplicando, para cada linha/série temporal as respetivas distribuições horárias históricas) e ainda, com base em algumas suposições predefinidas e recorrendo a várias fórmulas utilizadas em *WFM* em contexto de *call-center*, calcular o número de pessoas que cada equipa deverá ter para suportar toda a carga de trabalho prevista. Na impossibilidade de sabermos a 100% se as nossas previsões de capacidade estão corretas, a empresa utiliza varias métricas de atendimento ao cliente (*Response Rate*, *Full Resolution Time*, etc) que nos ajudam a perceber se a equipa está bem dimensionada para as expectativas estipuladas.

Capítulo 2

Métodos de Previsão de Séries Temporais

Visto que queremos trabalhar os nossos dados na forma de séries temporais, este capítulo será dedicado a explorar os conceitos essenciais aos vários métodos/modelos de previsão. Queremos explorar desde modelos mais simples e comuns na indústria - como por exemplo os modelos *naive*, *ARIMA* (correspondente a : *AutoRegressive Integrated Moving Average*) e *Exponential Smoothing*- até modelos mais complexos que possam cobrir algumas especificidades como por exemplo múltiplas sazonalidades - onde iremos investigar o modelo *TBATS* (correspondente a : *Trigonometric seasonality; Box-cox transformation; ARIMA erros; Trend and Seasonal componentes*) e ainda o *Prophet*.

Iremos ainda explorar um novo método proposto neste trabalho: o Modelo Financeiro. Como o nome indica, este modelo terá informações da intenção de crescimento financeiro da própria empresa, mas será complementado com as previsões efetuadas pelos modelos matemáticos. Desta forma, teremos uma previsão que tanto terá em conta a perspetiva económica da empresa como também o desenvolvimento natural dos dados, mostrando-se assim uma solução adequada ao negócio e não puramente matemática.

No final do capítulo iremos considerar algumas métricas de erro, as suas vantagens e desvantagens e a escolha final da métrica de comparação entre modelos. Nesta secção estarão incluídas o *MAE* (média dos erros absolutos), o *MAPE* (média dos erros absolutos em valores percentuais) e ainda, uma alternativa não muito usual, o *WAPE* (média ponderada dos erros absolutos em valores percentuais).

2.1 Decomposição de Séries Temporais

Uma série temporal pode ser decomposta em 3 componentes:

- A componente *trend-cycle* que representa a direção geral do desenvolvimento da série temporal, descrita como T_t ;
- A componente sazonal que representa a repetição de alguns fenómenos em certos períodos de tempo, descrita como S_t ;
- A componente remanescente que representa tudo o resto presente na série temporal, descrita como R_t .

Assim sendo, qualquer série temporal pode ser descrita como a soma ou a multiplicação destas 3 componentes.

$$Y_t = T_t + S_t + R_t, t \in \mathbb{N} \quad (2.1)$$

ou

$$Y_t = T_t \times S_t \times R_t, t \in \mathbb{N} \quad (2.2)$$

Tal decomposição torna-se bastante útil para analisar uma série temporal, mas também pode trazer benefícios aquando do processo de previsão. Seguindo uma decomposição *STL* [1] ("*Seasonal and Trend decomposition using Loess*"), Hyndman [2] propõe efetuar dois processos de previsão - um para a componente sazonal ajustada ($A_t = T_t + R_t$) onde temos alguma flexibilidade na escolha do método de previsão e outro para a componente sazonal (S_t) onde Hyndman sugere o uso do método de previsão *seasonal naive*. Desta forma, ao tratarmos cada componente de forma separada estaremos a ajustar o melhor processo de previsão para cada uma com o objetivo de no final obter uma melhor previsão.

Como o objetivo do trabalho é construir um algoritmo robusto que consiga obter previsões justificadas para qualquer série temporal, justifica-se a escolha pela decomposição *STL*. Isto porque esta decomposição suporta várias tipos de sazonalidade e permite mudanças ao longo do tempo na componente sazonal, tornando-se assim versátil para um grande leque de séries temporais. No entanto, de forma a que a decomposição *STL* possa ser executada precisamos de estar perante uma série temporal com mais do que 24 meses de histórico, para que possa reconhecer alguma componente sazonal. Assim sendo, aqui temos a primeira grande separação no algoritmo proposto:

- Para séries temporais com mais do que 24 meses de histórico os métodos de previsão serão executados tanto para a série temporal original como para a série temporal decomposta;
- Para séries temporais com mais do que 12 meses de histórico (e menos do que 24), os métodos de previsão serão executados apenas para a série temporal original (não sendo efetuada qualquer decomposição);
- Para séries temporais com menos do que 12 meses o único método de previsão a ser executado é o Modelo Financeiro.

Esta separação permitirá a construção de um algoritmo automático e robusto para vários tipos de séries temporais.

2.2 Modelos Matemáticos

Nesta secção iremos abordar vários métodos de previsão, enumerando quais as maiores vantagens e desvantagens de cada um.

2.2.1 *Naive e Seasonal Naive*

Começando com o modelo *Naive*, este modelo assume como previsão para o período de tempo futuro $h \in \mathbb{N}$ o valor da observação anterior, isto é:

$$\hat{Y}_{t+h,t} = Y_t \quad (2.3)$$

Apesar da simplicidade do método, por vezes acaba por ser das melhores opções, especialmente para previsões a curto prazo. No entanto, é um método que ignora qualquer tendência ou sazonalidade presente nos dados, resultando numa previsão pobre.

De seguida, teremos o modelo *Seasonal Naive*, que assume como previsão para o período de tempo futuro $h \in \mathbb{N}$ o valor da observação do período anterior, sendo T o período sazonal da série temporal.

Assim sendo, teremos a seguinte equação:

$$\hat{Y}_{t+h,t} = Y_{t+h-T} \quad (2.4)$$

Este modelo apresenta-se como uma boa opção para conjuntos de dados vincadamente sazonais e estáveis ano após ano. No entanto, também ignora qualquer tendência

presente nos dados, não sendo a solução ideal para conjuntos de dados em claro crescimento/decrescimento.

2.2.2 *Exponential Smoothing*

Existem vários modelos de *Exponential Smoothing*, desde modelos mais simples como o *Simple Exponential Smoothing* - que assume como previsões a média das observações passadas - até modelos ligeiramente mais complexos como os *Holt-Winters* - que dão uma grande importância à *trend* presente nos dados. Todos estes modelos assumem como previsão para o período de tempo futuro $h \in \mathbb{N}$ uma média ponderada de todas as observações passadas, dando mais peso/importância a observações mais recentes e, por consequência, menos peso a observações mais distantes no passado.

De forma a poder percorrer os vários tipos de modelos de *Exponential Smoothing* iremos recorrer a uma aplicação apresentada em [2]: modelos de espaço de estado (*state space models*). Desta forma, todos os modelos serão descritos por uma equação que descreve os dados observados e outra que descreve componentes não observáveis (como a *trend* e a sazonalidade). Para além disso, não produzem apenas previsões isoladas, mas também intervalos de confiança que limitam o intervalo de valores para as previsões efetuadas. Assim sendo, poderemos descrever os vários tipos de modelos como $ETS(R, T, S)$, onde $R \in \{A, M\}$ representa uma componente aditiva ou multiplicativa de erros, $T \in \{N, A, A_d\}$ representa uma componente *trend* inexistente, aditiva ou aditiva amortecida e $S \in \{N, A, M\}$ representa uma componente sazonal inexistente, aditiva ou multiplicativa.

Para podermos automatizar o processo, recorreremos à função *ets()* presente no *software R (package forecast)*, que percorre os vários modelos $ETS(R, T, S)$ possíveis para a série temporal em análise fazendo uma seleção dos mesmos usando o AIC_c (*Akaike's Information Criterion* corrigido para pequenas amostras).

2.2.3 *ARIMA*

Enquanto que os modelos *Exponential Smoothing* são mais baseados nas componentes de sazonalidade e de *trend*, os modelos *ARIMA* tentam descrever a auto-correlação presente nos dados.

Os modelos *ARIMA* são uma combinação de modelos auto-regressivos (AR) com modelos de médias móveis (MA), podendo ser descritos como *ARIMA*(p,d,q) onde p representa a ordem da componente auto-regressiva e q representa a ordem da componente de médias móveis. Para além disso, estes modelos têm como requisito a estacionariedade da série temporal, ou seja, a média e a covariância das observações têm de ser independentes do tempo. Para atingir este propósito, os modelos *ARIMA* recorrem ao operador de diferenciação ∇ que é aplicado o número de vezes que forem necessárias para atingir a estacionariedade. Ou seja, considerando d o número de diferenciações a realizar na série temporal Y_t , a resultante série temporal será descrita por $Z_t = \nabla^d Y_t$, onde $\nabla^d = (1 - B)^d$ e B é um operador que desloca cada observação para observação imediatamente anterior, isto é, $BY_t = Y_{t-1}$. Por exemplo, se tivermos perante um modelo *ARIMA*($p,1,q$), com $Z_t = Y_t - Y_{t-1}$ teremos a seguinte equação:

$$Z_t = \alpha_1 Z_{t-1} + \alpha_2 Z_{t-2} + \dots + \alpha_p Z_{t-p} + e_t - \beta_1 e_{t-1} - \beta_2 e_{t-2} - \dots - \beta_q e_{t-q} \quad (2.5)$$

Mais uma vez, de forma automatizar o processo de escolha das ordens do modelo *ARIMA*, iremos recorrer à função *auto.arima()* presente no *software* R (*package forecast*). Esta função também pode considerar uma componente sazonal (ajustando o modelo nesse sentido) e seleciona o melhor modelo para a série temporal em questão minimizando o valor *AIC*.

2.2.4 *TBATS*

Proposto em 2011, o modelo *TBATS* [3] revela-se uma boa alternativa na modelação de séries temporais com componentes sazonais mais complexas. Por exemplo, poderemos ter uma série temporal com períodos mensais que, por exemplo, em vez de ter uma sazonalidade fixa de período inteiro (por exemplo um fenómeno que é recorrente a cada 12 meses) apresenta uma sazonalidade que muda ligeiramente com o tempo e com um período não inteiro. Nesta situação, os modelos já vistos anteriormente não seriam capazes de capturar corretamente a sazonalidade resultando em previsões pobres. Com o modelo *TBATS* teremos a flexibilidade de capturar múltiplas sazonalidades com períodos não inteiros, visto que este modelo recorre a uma representação trigonométrica da sazonalidade baseada em séries de *Fourier*. Para além disso, este modelo incorpora uma componente auto-regressiva de erros e ainda uma possível transformação *Box-Cox* para

estabilizar a covariância. Os modelos *TBATS* apresentam uma grande complexidade, podendo ser descritos pelo seguinte conjunto de equações:

$$Y_t^{(\lambda)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^T S_{t-m_i}^{(i)} + d_t \quad (2.6)$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t \quad (2.7)$$

$$b_t = \phi b_{t-1} + \beta d_t \quad (2.8)$$

$$d_t = \sum_{i=1}^p \varphi_i d_{t-i} + \sum_{i=1}^q \theta_i e_{t-i} + e_t \quad (2.9)$$

Neste conjunto de equações, $Y_t^{(\lambda)}$ representa a série temporal no instante t , após uma transformação Box-Cox com o parâmetro λ , $S_t^{(i)}$ é a i -ésima componente sazonal com possíveis períodos m_i , l_t representa o nível local da série, b_t representa a *trend* presente no instante t e, por fim, d_t representa um processo *ARMA*(p,q) e e_t é um processo Gaussiano de ruído branco (*white noise*) com média 0 e variância constante igual a σ^2 .

Este modelo está implementado de uma forma automatizada no *software* R (*package forecast*) pela função *tbats()*, apresentando-se assim como uma alternativa viável para um grande leque de séries temporais. No entanto, a estimação do modelo tem um maior custo computacional relativamente aos métodos já apresentados e, como todos os processos automatizados, pode nem sempre efetuar boas previsões.

2.2.5 Prophet

O *Prophet* é um algoritmo automatizado de previsão criado pela equipa de *Data Science* do *Facebook* [4]. O facto de ser um algoritmo de previsão "à escala" facilita bastante o uso deste modelo. Isto significa que podemos fornecer ao modelo os nossos dados sem uma fase de pré-processamento visto que o modelo funciona com valores em falta, múltiplas sazonalidades e ainda possíveis *outliers* presentes (no caso em estudo não é muito relevante porque existirá uma fase tratamento de *outliers*). Podemos descrever o modelo pela seguinte equação:

$$Y_t = T_t + S_t + H_t + e_t \quad (2.10)$$

onde Y_t representa a observação no instante t , T_t representa a componente de *trend*, S_t representa a componente sazonal da série temporal, H_t representa a componente de feriados e, por fim, e_t representa a componente de erro. Apesar de ser um modelo bastante

promissor, o *Prophet* funciona melhor com séries temporais com um componente sazonal bastante vincada, o que pode nem sempre ser o caso.

2.2.6 Redes Neurais

As Redes Neurais -*Neural Networks* são um algoritmo que tenta imitar o comportamento dos nossos neurónios. Quando "alimentadas" com um dado *input*, trabalham essa informação nas possíveis várias camadas escondidas compostas por uma ou várias funções de ativação e retornam uma resposta/previsão. As funções de ativação e o peso de cada conexão entre nós de diferentes camadas, são treinados com registos pré-existent da série temporal que tentamos prever.

No relatório em questão iremos recorrer a redes neurais *feed-forward* com uma só camada de escondida correspondente a valores passados consecutivos (ou seja, de *lag* 1, idêntico ao processo usado nos modelos *ARIMA*). Desta forma, a cada valor previsto estaremos a utilizar uma função de ativação aplicada ao valor imediatamente anterior.

Este modelo está implementado de uma forma automatizada no *software* R (*package forecast*) pela função *nnetar()*.

2.3 Modelo de Previsão Financeiro

O Modelo Financeiro tem como objetivo a previsão da quantidade de *workload* futura tendo em conta não só o crescimento natural do *workload* como também as expectativas financeiras da empresa. Desta forma, estaremos a dimensionar cada equipa de uma forma mais ajustada às expectativas de crescimento do negócio.

Primeiro começamos por obter as previsões financeiras da empresa em termos de encomendas de *boutiques* (*BO- boutique orders*) para cada marca associada a *FPS*. Depois, para cada linha de contacto, obtemos a informação histórica das *BO* e do *workload* gerados em cada mês, conseguindo assim obter um rácio de $\frac{Workload}{BO}$ a que iremos chamar *WPO* (*workload per order*).

Queremos sempre ter cuidado com o *WPO* a ser usado de forma a não efetuarmos previsões nem muito altas nem muito baixas e queremos ter sempre em conta os desempenhos mais recentes da equipa que está encarregue pela linha de contacto em questão. Por exemplo, se usarmos o *WPO* de um mês onde a equipa em questão apresentou grandes níveis de eficiência de processos então teremos um *WPO* baixo o que implicaria uma

previsão enganadora de valores baixos de *workload*. Assim sendo, a forma como calculamos o *WPO* depende do número de meses de histórico da linha de contacto em questão. Se tivermos perante uma linha de contacto com menos do que 12 meses de histórico, então o nosso *WPO* a ser aplicado será a média dos *WPO* registados nos meses existentes. Por outro lado, se tivermos uma linha de contacto com 12 ou mais meses de histórico iremos recorrer tanto ao mês homólogo do ano anterior como também à média do *WPO* dos últimos 3 meses (desta forma teremos em conta a altura do ano para o qual queremos aplicar o *WPO* e ainda o desempenho dos últimos 3 meses da equipa em questão). Neste último caso o processo a seguir é o seguinte:

1. Calcular o *WPO* do mês homólogo do ano anterior - A ;
2. Calcular a média do *WPO* dos últimos 3 meses - B ;
3. Se o *WPO* do mês homólogo do ano anterior for mais do dobro (inclusive) da média do *WPO* dos últimos 3 meses, isto é, $\frac{A}{B} \geq 2$ então o *WPO* a ser usado será a média dos últimos 3 meses;
4. Caso contrário, o *WPO* a ser usado será uma média dos dois, isto é $\frac{A+B}{2}$.

Em suma, para obtermos as previsões de *workload* para cada linha de contacto basta multiplicar o *WPO* calculado pelas previsões de *BO* associadas à respectiva marca. No entanto, este processo estará sempre depende das previsões financeiras da empresa e, portanto, sujeito a erros que não podem ser controlados por nós.

2.4 Medidas de Erro

Nesta secção iremos percorrer 3 medidas de erro e explicitar algumas vantagens ou desvantagens da utilização de cada uma.

Começando com o *MAE* (média dos erros absolutos), esta medida apresenta-se como uma medida simples, mas eficaz. No entanto, o facto de ser dependente da escala torna a medida pouco interpretativa sem termos o conhecimento da ordem dos valores da série temporal em estudo. Esta medida é definida da seguinte forma:

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t|, e_t = \hat{Y}_t - Y_t, t \in \mathbb{N} \quad (2.11)$$

Por outro lado temos o *MAPE*, uma medida de erro percentual que vem combater o problema da interpretação levantado pelo *MAE*. Por isso apresenta-se como uma métrica muito utilizada e podemos ver em baixo como é definida:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|e_t|}{|Y_t|} * 100, e_t = \hat{Y}_t - Y_t, t \in \mathbb{N} \quad (2.12)$$

Apesar da sua grande vantagem interpretativa, torna-se uma medida que pode produzir previsões enviesadas. Isto acontece porque em cada instante t o erro absoluto nesse instante é dividido pelo volume, d_t , registado, levando a que grandes erros registados em instantes de pouco volume inflacionem o *MAPE*. Como no caso em estudo, é comum termos linhas de contacto com volumes bastante oscilatórios e de pequena grandeza, à primeira vista o *MAPE* não se apresenta como um bom candidato para medida de erro final (que irá ser usada para comparar as previsões de vários modelos).

Por fim, apresentamos o *WAPE* que se apresenta como uma alternativa ao *MAPE*, mas mais robusta. O *WAPE* é calculado pela soma dos erros absolutos normalizada pelo total de volumes registados. Assim, *WAPE* apresenta a mesma capacidade interpretativa do *MAPE*, mas penaliza de igual forma previsões demasiado altas ou demasiado baixas. Em baixo vemos a fórmula que descreve o *WAPE*:

$$WAPE = \frac{\sum_{t=1}^n |e_t|}{\sum_{t=1}^n |Y_t|} * 100, e_t = \hat{Y}_t - Y_t, t \in \mathbb{N} \quad (2.13)$$

Assim sendo, estas serão as medidas de erro que nos irão acompanhar nos processos de previsão tendo como objetivo comparar as várias métricas num exemplo teste que iremos ver mais à frente.

Capítulo 3

Pré-processamento dos Dados

3.1 Obtenção e Organização dos Dados

Na área de WFM, um tema que geralmente gera bastante discussão é qual a melhor forma de medir o trabalho gerado pelos funcionários. Na *FARFETCH* usamos uma plataforma a partir da qual todos os contactos são estabelecidos, quer os contactos externos (como uma chamada recebida de um cliente) quer os contactos internos (comunicação entre equipas para a resolução de problemas levantados pelos clientes). A partir desta plataforma conseguimos medir as já referidas unidades de *workload*.

Dentro desta plataforma os contactos são separados por linhas de contacto (por exemplo, uma linha de chamadas em inglês para responder a questões relacionadas com a marca A), podendo existir múltiplas linhas de contacto que são da responsabilidade de uma só equipa. Assim sendo, para cada linha de contacto, teremos o histórico de *workload* gerado por mês, podendo facilmente ser associado a uma série temporal Y_t , onde $t \in \mathbb{N}$ representa o mês em análise e $Y_t \geq 0, \forall t \in \mathbb{N}$ representa a quantidade *workload* registada nesse mês. Cada linha de contacto tem também um conjunto de 3/4 letras que a identifica de forma única.

Associadas a *FPS* teremos em análise 82 linhas de contacto, ou seja, 82 séries temporais com diferentes tamanhos, existindo algumas séries temporais com 3 anos/36 meses de informação e outras com menos de 12 meses de informação.

Devido à pequena dimensão de grande maioria das séries temporais, é de grande importância a qualidade de cada série temporal, isto é, queremos ao máximo ter um conjunto de dados coerente e isento de possíveis falhas existentes no processamento dos dados. No

entanto, e devido a *FPS* ser ainda um serviço recente e de alguma forma instável em algumas das suas componentes, há uma grande possibilidade de nem todos os dados referentes a uma série temporal serem relevantes para o desenvolvimento natural da mesma. Mais especificamente, esta área de serviço por vezes apresenta períodos de teste em algumas das marcas existentes, nomeadamente no início de vida destas marcas no contexto de *FPS*. Por exemplo, poderemos ter uma série temporal onde os primeiros 2 meses de histórico de *workload* são residuais, comparando com os restantes meses. Neste caso estaremos claramente perante 2 meses de teste da marca ou linha de contacto em questão, a partir dos quais a marca/linha de contacto realmente entrou em "ação". Na tabela a baixo vemos uma representação real do exemplo referido e mais à frente iremos percorrer uma solução para este problema.

	2021				
	Agosto	Setembro	Outubro	Novembro	Dezembro
NTE	6	7	97	121	149

Para além disso, outro problema que encontramos em algumas das séries temporais é não existência de volumes em alguns meses. Isto não se deve a nenhum fator sazonal, podendo apenas surgir de falhas no sistema de recolha de dados ou até por desistência temporária de algumas marcas do serviço prestado por *FPS*. Estas falhas nos dados vetoriais resultariam numa inconsistência temporal quando transformados para série temporal. Para contornar este problema, a abordagem adotada foi de substituir estes valores não existentes pela média dos valores do mesmo ano em que a observação está em falta. Com isto, o objetivo é afetar o mínimo possível a evolução da série temporal para conseguir correr o processo de previsão para todas as linhas de contacto.

3.2 Tratamento de *Outliers*

O tratamento de *outliers* é uma fase importante no pré-processamento dos dados e ainda mais importante no processo de previsão proposto neste relatório, que tenta criar um algoritmo robusto de previsão num conjuntos de dados instável e curto.

Inicialmente, para ultrapassar a possibilidade de existirem meses iniciais residuais em termos de volumes como já foi acima referido, e no intuito de aprimorar a qualidade dos dados, decidimos optar por uma abordagem de remoção das observações iniciais não significativas, desde que estas sejam consecutivas. Para tal iremos recorrer à

fórmula mais usual de cálculo de *outliers* - o método de *Tukey* (1977) [5] . Este método considera como valores atípicos (*outliers*) todos os valores que estejam fora do intervalo $[Q1 - 1.5 \times IQR, Q3 + 1.5 \times IQR]$ e como valores atípicos extremos valores que estejam fora do intervalo $[Q1 - 3 \times IQR, Q3 + 3 \times IQR]$, onde $Q1$ e $Q3$ representam o primeiro e terceiro quartil dos dados analisados na forma vetorial e IQR representa a diferença do terceiro quartil para o primeiro, isto é, $IQR = Q3 - Q1$. A escolha deste método deve-se ao facto de ser um método independente de qualquer suposição da distribuição dos dados. Para além disso, colocamos os dados numa escala logarítmica para limitar o mínimo e o máximo dos intervalos em cima apresentados a apenas valores positivos (desta forma garantimos que valores muito baixos - como, por exemplo, valores próximos de 0 - sejam considerados *outliers*)

Desta forma, se as primeiras observações consecutivas da série temporal forem consideradas como sendo possíveis *outliers* (fora do intervalo $1.5 \times IQR$) então teremos motivos para assumir que não são relevantes no conjunto de dados presente e por isso poderão representar meses de teste, devendo assim ser removidas. Por exemplo, se tivermos a 1^a, 2^a e 3^a observação como possíveis *outliers*, visto que são observações iniciais e consecutivas devem as 3 ser removidas. No entanto, se tivermos a 1^a e a 3^a observação como possíveis *outliers* apenas iremos remover a 1^a observação.

O passo seguinte é analisar também possíveis valores atípicos no meio da série temporal para não influenciarem negativamente as previsões. Para isso, recorreremos novamente ao método de *Tukey*, mas desta vez usando o intervalo $3 \times IQR$. Assim, aumentando o intervalo de valores aceitáveis, iremos identificar apenas *outliers* extremos. Desta forma, a abordagem usada foi a de substituir qualquer valor abaixo do limite mínimo do intervalo $[Q1 - 3 \times IQR, Q3 + 3 \times IQR]$ pelo mínimo do intervalo, isto é, por $Q1 - 3 \times IQR$ e qualquer valor acima do limite máximo do intervalo pelo limite máximo do intervalo. O objetivo é suavizar a série temporal no sentido de substituir valores extremamente altos por valores não tão altos e valores extremamente baixos por valores não tão baixos.

Por fim, poderemos ter uma primeira visão das séries temporais com que iremos trabalhar. A linha de contacto identificada pelas letras BWP (figura 3.1) será usada daqui em diante como exemplo para entendermos todas as fases de previsão e de planeamento de uma forma mais objetiva. A linha de contacto BWP existe desde Janeiro de 2019, mostrando bastante sazonalidade ao longo dos anos e uma tendência claramente crescente.

Por outro lado, na figura 3.2 temos uma linha de contacto com apenas 2 anos de histórico e uma tendência de crescimento quase nula.

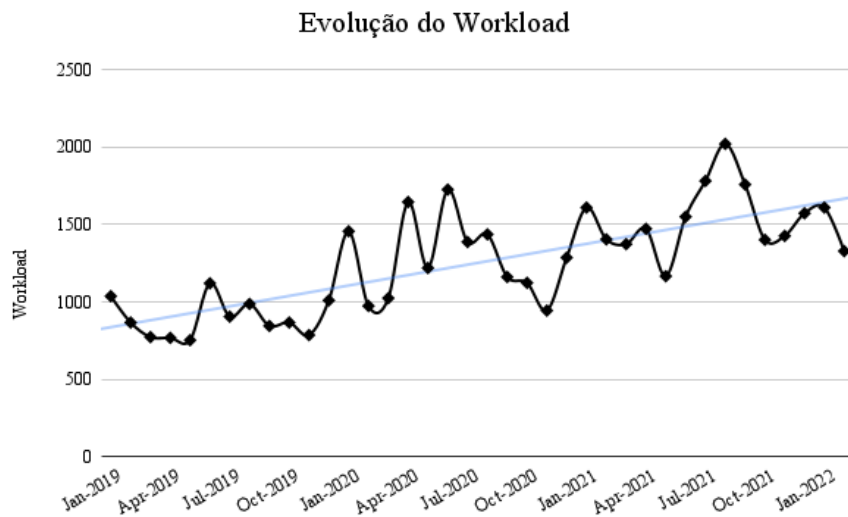


FIGURA 3.1: Evolução do Workload da linha de contacto BWP

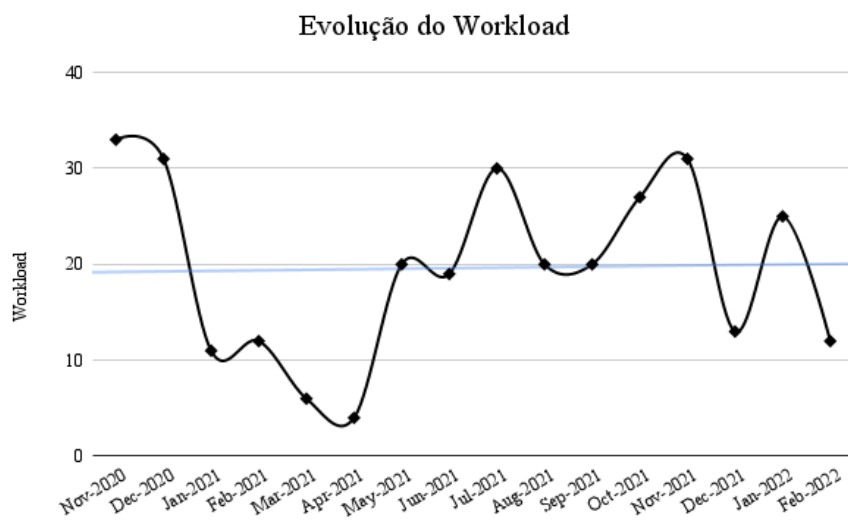


FIGURA 3.2: Evolução do Workload da linha de contacto HPTI

Capítulo 4

Estrutura do processo de previsão

Nesta secção iremos aplicar os modelos introduzidos no capítulo 2 ao conjunto de dados descrito no capítulo 3. Temos como objetivo analisar quais os modelos escolhidos com mais frequência e comparar os resultados obtidos utilizando várias medidas de erro. Para tal iremos utilizar Janeiro, Fevereiro e Março de 2022 como meses de teste para todas as linhas de contacto.

4.1 Apresentação da *framework* construída

Na figura 4.4 ilustramos a *framework* do *script* construída de forma a podermos resumir todos os passos do processo de previsão. De realçar que a fase de obtenção e pré-processamento dos dados é referente às secções 3.1 e 3.2. do presente relatório, as fases de previsão usando os modelos estatísticos e o modelo financeiro estão detalhadas nas secções 2.2 e 2.3, respetivamente e, por fim, a última fase de distribuições horárias será detalhada mais à frente na secção 5.1. As secções 5.2 e 5.3 não estão presentes na *framework* apresentada visto que são fases acopladas ao processo de previsão usadas para conectar as previsões meramente matemáticas com o objetivo inerente ao negócio: apurar a capacidade futura para as várias equipas.



FIGURA 4.1: FPS Script Framework

Como vemos na *framework* apresentada, para cada linhas de contacto, para além dos métodos estatísticos, iremos também correr o modelo financeiro. Com isto, no final de cada iteração da *framework*, dependendo do comprimento da série temporal em questão, poderemos ter 3 *outputs* de previsão:

1. Um *output* obtido exclusivamente pelos métodos estatísticos;
2. Um *output* obtido exclusivamente pelo modelo financeiro;
3. Um *output* que resulta da **média** dos dois últimos *outputs*, a que chamaremos o Modelo Financeiro Ajustado.

Tanto o *output* do modelo financeiro como o do modelo financeiro ajustado não estarão sujeitos a qualquer métrica de erro visto que apenas produzem previsões e não possuem qualquer processo de ajuste aos dados históricos. Ambos serão retornados no final de cada iteração da *framework* juntamente com o *output* dos métodos estatísticos para, numa fase final, compararmos os 3 *outputs* e vermos qual se adequa melhor ao negócio. Esta última fase será sempre bastante subjetiva e dependente de vários fatores externos e internos do negócio, mas é essencial para um melhor enquadramento dos valores de previsão com a realidade complexa da empresa.

4.2 Comparação de modelos e validação das métricas de erro apresentadas

No que respeita aos modelos estatísticos teremos vários resultados provenientes das diferentes métricas de erro usadas. A primeira análise que iremos fazer é qual a frequência de escolha dos vários métodos estatísticos, dependendo da métrica de erro usada para comparação entre modelos na fase de treino dos mesmos. Isto é, se, por exemplo, usarmos o *MAE* como métrica de erro na competição dos modelos para cada série temporal, teremos no final os métodos estatísticos que foram utilizados para produzir as previsões finais. Se, por outro lado, utilizarmos o *MAPE* então teremos certamente outro conjunto de escolhas nos modelos utilizados. Posto isto, na tabela 4.1 temos um resumo das escolhas dos modelos de cada linha de contacto, utilizando as diferentes métricas de erro.

	<i>MAE</i>	<i>MAPE</i>	<i>WAPE</i>	TOTAL
<i>Seasonal Naive</i>	3	2	3	8
<i>Naive</i> com decomposição	2	0	3	5
<i>ARIMA</i>	5	4	6	15
<i>ARIMA</i> com decomposição	4	3	0	7
<i>Exponential Smoothing</i>	7	2	9	18
<i>Exponential Smoothing</i> com decomposição	5	1	7	13
<i>TBATS</i>	4	5	4	13
<i>TBATS</i> com decomposição	6	2	0	8
<i>Neural Networks</i>	31	44	28	103
<i>Prophet</i>	3	7	10	20

TABELA 4.1: Frequência absoluta da escolha de cada modelo para cada métrica de erro

Podemos ver que as redes neuronais (*Neural Networks*) foram largamente escolhidas (ainda mais quando a métrica de erro usada é o *MAPE*), seguidas pelo *Prophet* como segundo modelo mais escolhido. Comparando entre as 3 métricas, todas tiveram as escolhas dos modelos relativamente distribuídas, realçando que o *MAE* foi a única métrica onde todos os modelos foram escolhidos pelo menos uma vez.

Para além desta primeira análise, foram extraídas as previsões finais para os 3 meses de teste (janeiro, fevereiro e março) para podermos analisar a distribuição dos erros obtidos para cada linha de contacto utilizando as 3 métricas de erro. O erro "real" para cada linha de contacto é obtido dividindo o **valor real** de cada mês pelo o **valor previsto**. Esta forma de cálculo traz um valor interpretativo para cada erro obtido. Isto é, por exemplo, um erro de 110% significa que a linha de contacto em questão recebeu 10% acima do previsto e um erro de 90% significa que recebeu 10% abaixo. Em baixo temos a distribuição dos erros para cada métrica de erro e na tabela 4.2 temos um resumo relativamente a todos os *boxplots* apresentados.

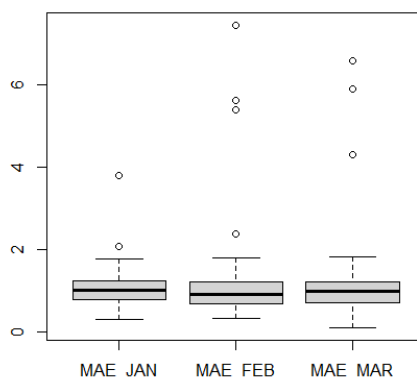
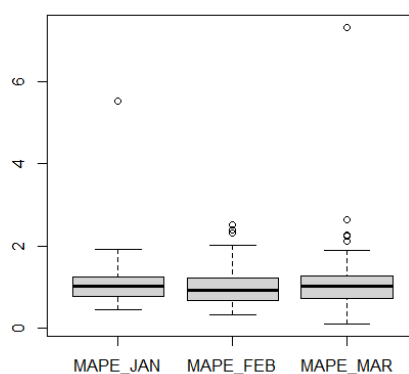
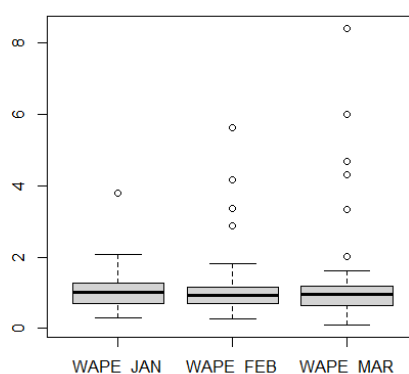


FIGURA 4.2: Erros obtidos utilizando o *MAE*

FIGURA 4.3: Erros obtidos utilizando o *MAPE*FIGURA 4.4: Erros obtidos utilizando o *WAPE*

	Janeiro	Fevereiro	Março	Média
Média dos erros utilizando o <i>MAE</i>	105%	116%	120%	114%
Desvio padrão dos erros utilizando o <i>MAE</i>	46%	115%	109%	87%
Média dos erros utilizando o <i>MAPE</i>	110%	100%	112%	108%
Desvio padrão dos erros utilizando o <i>MAPE</i>	63%	50%	90%	68%
Média dos erros utilizando o <i>WAPE</i>	105%	110%	123%	112%
Desvio padrão dos erros utilizando o <i>WAPE</i>	48%	85%	129%	85%

TABELA 4.2: Média e desvio padrão dos erros obtidos por mês para cada medida de erro

Com base nestes valores, a métrica de erro que levou a um melhor resultado foi o *MAPE* com uma média de erros e desvio padrão mais baixos.

No entanto, estes valores podem ser enganadores (principalmente as médias) porque temos grandezas de valores nos volumes muito diferentes entre as várias linhas de contacto. Este facto também é visível na presença de vários *outliers* nos *boxplots* apresentados.

Todos estes valores atípicos devem-se a erros elevadíssimos em linhas de contacto com muito pouco volume. Por exemplo, se tivermos uma linha de contacto com um volume real de 25 unidades de *workload* para o mês de Janeiro e a previsão for de 7 unidades de *workload*, facilmente temos um erro de 357% que irá impactar o valor da média global do mês de Janeiro apesar de ser um erro pouco significativo devido ao baixo volume apresentado.

Sendo assim, é importante ponderarmos os erros de cada linha de contacto consoante a sua importância no planeamento de equipas, ou seja, consoante a sua grandeza de volumes. Dito isto, na tabela em baixo vemos as médias dos erros já ponderados para cada métrica.

	Janeiro	Fevereiro	Março	Média Ponderada
Média ponderada dos erros utilizando o <i>MAE</i>	98%	99%	92%	96%
Média ponderada dos erros utilizando o <i>MAPE</i>	96%	92%	91%	93%
Média ponderada dos erros utilizando o <i>WAPE</i>	118%	116%	110%	115%

TABELA 4.3: Média ponderada dos erros obtidos por mês para cada medida de erro

Desta forma, vemos que a métrica que leva a um erro ponderado melhor é o *MAE*, ainda que apenas por uma diferença de 3 pontos percentuais comparativamente ao *MAPE*.

Na análise da frequência de escolha dos vários modelos estatísticos realçamos a grande superioridade das redes neuronais comparativamente aos restantes modelos. Esta superioridade foi motivo de alarme com receio de que as redes neuronais, apesar de demonstrarem grande maioria das vezes um ótimo ajuste aos dados históricos e ao conjunto de teste, pudessem produzir previsões desajustadas (ainda para mais sendo modelos conhecidos pela pouca precisão em previsões a longo prazo) e ofuscassem a utilização de outros modelos que poderiam (ou não) ter um melhor desempenho no que toca às previsões finais. Assim sendo, decidimos a título de curiosidade testar a *framework* apresentada, mas sem a presença das redes neuronais no conjunto de modelos estatísticos disponíveis. Na tabela 4.4 vemos a frequência de escolha dos restantes modelos, agora excluindo as redes neuronais.

Podemos constatar que a frequência absoluta da escolha dos modelos para cada métrica de erro encontra-se mais distribuída, não havendo um modelo com uma maioria indiscutível. Vemos que os modelos mais escolhidos foram os *ARIMA*, logo seguidos dos *Exponential Smoothing*, *TABTS* e *Prophet*. Nos 3 meses de teste obtivemos os resultados expostos na tabela 4.5.

	MAE	MAPE	WAPE	TOTAL
<i>Seasonal Naive</i>	4	4	3	11
<i>Naive com decomposição</i>	4	3	4	11
<i>ARIMA</i>	11	10	14	35
<i>ARIMA com decomposição</i>	7	8	2	17
<i>Exponential Smoothing</i>	10	9	15	34
<i>Exponential Smoothing com decomposição</i>	7	2	11	20
<i>TBATS</i>	11	16	7	34
<i>TBATS com decomposição</i>	8	6	0	14
<i>Prophet</i>	8	12	14	34

TABELA 4.4: Frequência absoluta da escolha de cada modelo para cada métrica de erro sem utilizar as Redes Neurais

	Janeiro	Fevereiro	Março	Média
Média dos erros utilizando o <i>MAE</i>	103%	117%	119%	112%
Desvio padrão dos erros utilizando o <i>MAE</i>	45%	91%	85%	72%
Média dos erros utilizando o <i>MAPE</i>	108%	117%	123%	115%
Desvio padrão dos erros utilizando o <i>MAPE</i>	64%	94%	114%	90%
Média dos erros utilizando o <i>WAPE</i>	106%	112%	127%	114%
Desvio padrão dos erros utilizando o <i>WAPE</i>	52%	88%	144%	92%

TABELA 4.5: Média e desvio padrão dos erros obtidos por mês para cada medida de erro sem utilizar as Redes Neurais

Obtivemos valores mais desviados da média desejada (100%) e mais dispersos, isto é, com um maior desvio padrão e a presença dos outliers manteve-se. Neste caso a medida de erro escolhida seria o *MAE* com média e desvio padrão mais baixos que as restantes, mas maiores comparativamente aos erros obtidos com o *MAPE* quando não excluimos as redes neurais.

Fomos também calcular a média ponderada para cada mês, obtendo os resultados presentes na tabela 4.6.

	Janeiro	Fevereiro	Março	Média ponderada
Média ponderada dos erros utilizando o <i>MAE</i>	114%	111%	101%	109%
Média ponderada dos erros utilizando o <i>MAPE</i>	105%	100%	102%	102%
Média ponderada dos erros utilizando o <i>WAPE</i>	117%	114%	119%	117%

TABELA 4.6: Média ponderada dos erros obtidos por mês para cada medida de erro sem utilizar as Redes Neurais

Neste caso, a medida com melhor desempenho foi o *MAPE*, sendo melhor apenas 2 pontos percentuais do que a média ponderada usando o *MAE* e as redes neurais.

Com isto concluímos que, apesar da quase hegemonia das redes neurais, estas mesmas aparentam conduzir a melhores resultados a curto prazo e, por isso mesmo, deverão

ser utilizadas na *framework*. No que toca à escolha da métrica a utilizar, o *MAE* leva vantagem (ainda que curta) relativamente ao *MAPE* na média ponderada dos erros nos meses de teste, mas o *MAPE* apresenta melhor precisão dos erros obtidos para as várias linhas de contacto, visto que apresenta um desvio padrão significativamente mais baixo do que o *MAE* (tudo isto utilizando as redes neuronais). Concluimos assim, que a melhor métrica a utilizar seria o *MAPE* e não deveremos excluir as redes neuronais do conjunto de modelos estatísticos a serem utilizados. No entanto, fica a dúvida da veracidade das afirmações anteriores em previsões a longo prazo.

4.3 Aplicação do processo de previsão a um exemplo real

Depois de descrito o processo de previsão na secção anterior, vamos agora aplicar o mesmo à linha de contacto BWP de forma a podermos ter maior visibilidade sobre os possíveis *outputs* que o processo retorna. Visto que é uma linha de contacto com mais do que 24 meses de histórico tanto estará sujeita aos métodos estatísticos como ao modelo financeiro. Este processo demora menos de 1 minuto a correr sendo por isso um processo rápido e eficiente. Na figura 4.5 temos em conjunto o *output* dos métodos estatísticos (onde o modelo escolhido foi o *Prophet* com um *MAPE* de 9.78%), o *output* do modelo financeiro e ainda o *output* do modelo financeiro ajustado que pondera de igual forma os dois *outputs* anteriores.

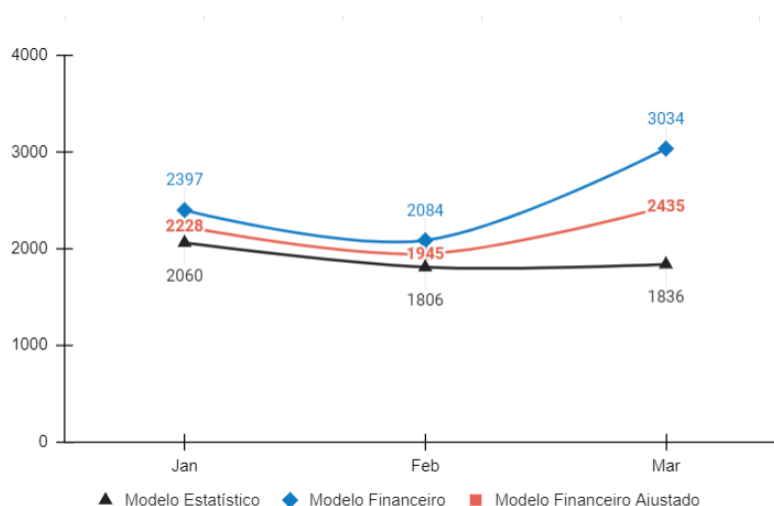


FIGURA 4.5: Resultados de Previsão (2022) da linha de contacto BWP

Vemos que o modelo estatístico prevê uma tendência ligeiramente decrescente até março, enquanto que o modelo financeiro prevê uma subida de volumes no mês de março (sendo indicador de uma possível iniciativa de marketing que possa vir a acontecer nesse mês). De forma a não ignorar por completo nem as expectativas financeiras nem o desenvolvimento natural dos dados, o mais sensato seria escolher o modelo financeiro ajustado, mas poderíamos ter escolhido outro modelo com base em informações internas do negócio. Os valores reais registados para os meses de janeiro, fevereiro e março foram 2293, 1669 e 1664, respetivamente, ou seja, na realidade o modelo estatístico teria melhor desempenho comparativamente aos restantes e será o *output* do modelo estatístico que iremos usar daqui em diante.

Capítulo 5

Planeamento de Capacidade

Estando concluído o processo de previsão, segue-se o processo de planeamento. Este capítulo será dedicado a dissecar todo o processo do cálculo da capacidade das várias equipas envolvidas na unidade de trabalho em estudo. Iremos começar por explicar as distribuições horárias aplicadas aos *outputs* de volumes mensais, seguido de uma explicação dos vários pressupostos usados para transformarmos volumes de *workload* em número de agentes necessários. Por fim, iremos apresentar uma agregação do *output* final de forma a construirmos uma narrativa acessível e consumível por parte do negócio.

5.1 Distribuições Horárias de Volumes

Numa perspetiva de planeamento, é extremamente importante conhecermos a capacidade necessária por hora de forma a garantirmos um bom serviço de atendimento ao cliente. No caso de estudo que temos em mãos teremos em conta tanto linhas de contacto via telefone como linhas de contacto via email. Em linhas de contacto telefónicas, um planeamento horário torna-se um exercício crucial de forma a garantirmos um atendimento pronto e eficiente. Por outro lado, para linhas de contacto via email, a relevância do planeamento horário continua a existir, mas deixa de ser crucial (o facto de ser um canal de contacto "offline" não traz tanta urgência no atendimento dos vários casos que vão chegando à *queue*) sendo por isso, grande maioria das vezes, remetido a um planeamento diário.

Nesta secção iremos discutir o processo usado na transformação de volumes mensais em volumes horários, utilizando como base distribuições horárias históricas, referentes ao ano passado, divididas por *isoweek*, dia da semana e hora. Desta forma, iremos capturar

possíveis sazonalidades anuais, intra semanais e intra diárias. Começamos por calcular a percentagem de volume que cada hora representou no mês em que esteve inserida para, de seguida, multiplicar esse mesmo valor pelo valor mensal que estamos a prever, obtendo assim o volume horário previsto por hora para cada linha de contacto.

Agora aplicando este processo para a linha de contacto BWP, por exemplo para o mês de Janeiro de 2022 temos um valor previsto de 2060 unidades de *workload*. Para calcularmos o volume previsto para a hora 1 de uma sexta feira da *isoweek* 2 teremos primeiro de calcular a percentagem que essa mesma hora, no mesmo dia da semana e na mesma *isoweek* de 2021, representou no mês de janeiro de 2021. De seguida multiplicamos esta mesma percentagem obtida pelo valor previsto mensal para janeiro de 2022 e obtemos o volume previsto para a hora em questão. Neste caso em concreto obtivemos uma percentagem de 0.06%, multiplicamos esta percentagem pelas 2060 unidades de *workload* previstas para janeiro 2022 e obtivemos uma previsão de 1.236 unidades de *workload* para essa mesma hora. Percorremos este processo para todas as horas presentes nos meses de janeiro, fevereiro e março de 2022 obtendo os volumes horários previstos para todos estes meses. É essencial construirmos o *output* neste formato de forma a conectarmos as nossas previsões com as restantes equipas de *Workforce Management* (esta informação poderá ser útil, por exemplo, para a criação de horários com base nos volumes previstos). Assim conseguimos construir um planeamento de capacidade estruturado na granularidade que o negócio precisa.

5.2 Pressupostos

Quando finalmente conseguimos extrair as previsões de volumes horários temos de transformar esta informação em número de horas de trabalho necessárias para suportar as várias linhas de contacto e atingir os níveis de serviço desejados. Para tal, e visto que nunca assumimos 100% de produtividade por parte dos agentes, teremos de aplicar vários pressupostos que aproximam as nossas previsões da realidade presente no dia a dia dos agentes. Sempre que possível (ou sempre que os dados nos permitam) iremos utilizar registos históricos dos pressupostos para suportar os valores usados no cálculo de planeamento.

O primeiro pressuposto que aplicamos aos volumes horários previstos é o *Average Handling Time* (AHT), que representa o tempo médio que cada agente despende por cada

unidade de *workload* (normalmente medido em segundos). Desta forma, se multiplicarmos *workload* pelo AHT em horas, obtemos rapidamente o número de horas necessárias para suportar a quantidade de *workload* previsto. Os valores de AHT que iremos utilizar terão como base o registo histórico tanto anual, como intra semanal e ainda intra diário de cada linha de contacto. Nesta secção iremos ver progressivamente a construção da equação de capacidade, tendo até ao momento a estrutura da equação 5.1 .

$$Hourly_Needs = \frac{workload * AHT}{3600} \quad (5.1)$$

Das 8h de trabalho estabelecidas (já sem contar com a hora de almoço) teremos de assumir que os agentes não estão constantemente a atender chamadas ou a responder a emails. Sendo assim, introduzimos o segundo pressuposto: a ocupação (OCP). Se assumirmos, por exemplo, uma ocupação de 80% estaremos a assumir apenas 6.4h produtivas por agente. Este valor garante o bem estar no ambiente de trabalho dos agentes, sendo uma prioridade na *FARFETCH*. Aplicando este valor à quantidade de horas obtidas pelo pressuposto anterior teremos o que chamamos de *Net Needs* (quantidade de horas brutas produtivas necessárias), ou seja teremos o nosso cálculo de capacidade até ao momento estruturado como podemos ver na equação 5.2 .

$$Net_Needs = \frac{workload * AHT}{3600 * OCP} \quad (5.2)$$

Teremos ainda pressupostos de redução das horas brutas produtivas, que serão aplicados apenas a volumes diários por razões de coerência. O mais comum é o *Shrinkage In Office* (SHRK), que corresponde ao tempo despendido, em horário de trabalho, em reuniões, intervalos e desenvolvimentos pessoais. Teremos também em consideração possíveis faltas ou atrasos por parte dos agentes, a que iremos chamar de absentismo (ABS). Por fim, teremos em conta saídas inesperadas de agentes que optam por seguir outro caminho profissional, a que chamaremos *attrition* (ATR). Estes 3 últimos pressupostos serão descritos em forma percentual e serão aplicados indiretamente ao número de horas brutas produtivas. Isto é, se tivermos um SHRK de 15% num total de 10 horas brutas produtivas então estaremos a dizer que, das 6.4h horas produtivas que cada agente possui, 15% destas mesmas serão usadas para reuniões, intervalos, desenvolvimentos etc. Ou seja, se utilizarmos FTEs (*Full Time Employee*) como unidade de cálculo, teríamos na equação 5.3 um espelho nas necessidades com o SHRK aplicado.

$$Net_Needs_With_SHRK = \frac{workload * AHT}{3600 * OCP * 8 * (1 - SHRK)} \quad (5.3)$$

Voltando às necessidades horárias e acrescentando o ABS e o ATR, podemos ver na equação 5.4 as horas necessárias para suportar todo o volume de trabalho nas realidade do dia a dia dos agentes.

$$Total_Needs = \frac{workload * AHT}{3600 * OCP * (1 - (SHRK + ABS + ATR))} \quad (5.4)$$

Teremos também de incluir as férias que os agentes precisam de gozar por lei, introduzindo assim o último pressuposto: *Thresholds* (THR), que representa o número de pessoas recomendamos a tirar férias por cada dia, ou então a percentagem de equipa recomendada a tirar férias por cada mês. Este pressuposto é calculado com base nos volumes previstos ao longo do ano, garantindo tanto algumas regras impostas por lei (como, por exemplo, cada agente tem de ter disponíveis 10 dias úteis consecutivos de férias por ano) como também algumas regras "humanas" que enquadram os *thresholds* à realidade de marcação de férias (mais frequente em meses de verão). Assim, chegamos finalmente ao número de horas finais necessárias (equação 5.5).

$$Final_Hour_Needs = \frac{workload * AHT}{3600 * OCP * (1 - (SHRK + ABS + ATR)) * (1 - THR)} \quad (5.5)$$

Por fim, de forma a termos o número de FTEs necessários (que corresponde ao número de agentes a *full-time* necessários) basta dividir o resultado da equação 5.5 por 8.

É importante de referir que o único pressuposto aplicado aos volumes horários é apenas o AHT, enquanto que os restantes são aplicados aos volumes diários. No final, podemos agregar ao mês as necessidades calculadas, retornando o número de agentes necessários por mês para suportar o volume de trabalho para equipa linha de contacto.

5.3 Output de Capacidade Final

Agora que temos os volumes mensais previstos para a linha de contacto BWP e o procedimento de cálculo de capacidade resumido, basta apenas encontrar qual o valor que iremos utilizar para cada pressuposto. De forma a simplificar a demonstração do *output* final iremos aplicar todos os pressupostos ao mês para Janeiro, Fevereiro e Março de 2022 (os meses de teste que estamos a utilizar).

Todos os pressupostos, com exceção da AHT (que será proveniente de dados históricos da própria linha de contacto BWP), serão provenientes da equipa de capacidade em que a linha de contacto BWP está inserida. Isto é, por uma questão de eficiência e de estrutura, foram criadas várias equipas de capacidade onde foram alocadas as várias linhas de contacto com base na língua de atendimento ao cliente e na localização geográfica da correspondente marca. Assim, na tabela 5.1 podemos fazer a construção do cálculo de necessidade para os 3 meses de teste.

	Janeiro	Fevereiro	Março
<i>Volumes Workload</i>	2293	1669	1664
AHT (segundos)	228	207	190
OCP	80%	80%	80%
<i>Net Hour Needs</i>	182	120	110
SHRK	14.78%	14.77%	14.89%
ABS	5%	5%	5%
ATR	3%	3%	3%
<i>Total Hour Needs</i>	235	155	142
THR	4.13%	8.68%	8.26%
<i>Final Hour Needs</i>	245	170	155
<i>Final FTE Needs</i>	1.4	1.0	0.9

TABELA 5.1: Construção do cálculo de capacidade para os meses de janeiro, fevereiro e março da linha de contacto BWP

Na tabela 5.1 podemos ver que, por exemplo para Janeiro 2021, a linha de contacto tinha uma necessidade de aproximadamente 1.4 FTEs. Se a linha de contacto fosse atribuída de forma isolada a uma equipa de capacidade então precisaríamos de 2 agentes para fazer o trabalho de 1.4 FTEs (visto que não é possível contratar 0.4 agentes). No entanto, como a linha de contacto BWP está atribuída a uma equipa de capacidade com mais linhas de contacto atribuídas, somámos as necessidades de FTEs de cada linha de contacto e só procedemos ao arredondamento de agentes necessários no final do cálculo de capacidade.

Capítulo 6

Conclusão

Todos os processos inseridos em WFM são de extrema importância para o desenvolvimento de qualquer negócio. O processo de previsão é mais uma peça fundamental numa máquina que trabalha 24/7 para garantir um bom serviço ao cliente.

Para que o processo de previsão possa ser percorrido com sucesso e tenha usabilidade por parte de todo o negócio é importante garantir métodos de previsão adequados à realidade existente, ao mesmo tempo que garantimos uma grande diversidade no tipo de modelos utilizados. É importante inovar ao longo dos tempos na perspetiva que usamos no processo de previsão, pois o negócio e os próprios dados disponíveis nunca serão imutáveis e devemos ser capazes de nos adequarmos a novas realidades. Neste relatório, para o conjunto de dados em questão, as redes neuronais aplicadas em séries temporais tiveram sem sombra de dúvida bastante sucesso na fase de modelação dos dados, logo seguidas do *Prophet*, dos *TBATS* e dos *Exponential Smoothing*. Obtivemos erros de previsão satisfatórios para os meses de teste, mas com espaço para melhorar no desvio padrão dos mesmos. O facto de termos em estudo várias séries temporais com volumes quase residuais potencia erros de previsão elevados. Das 3 métricas de erro utilizadas, não foi registada uma grande diferença, tendo o MAPE uma ligeira vantagem sobre as restantes.

Para além disso, neste relatório percorremos uma fase de processamento e organização de dados existentes de forma a podermos extrair o máximo de informação valiosa possível e fiável. Apresentamos um procedimento de remoção e/ou substituição de *outliers* com o objetivo em cima referido. Recorremos ao método de *Tukey* que, apesar da sua simplicidade, teve resultados bastante satisfatórios para o tipo de dados que tivemos em estudo.

Por fim, percorremos o processo de cálculo de capacidade que tem como objetivo estimar o número de agentes necessários para aguentar toda a carga de trabalho esperada.

Neste processo é crucial conhecermos a fundo todos os pressupostos que estamos a utilizar para entendermos a interligações que possam existir entre os mesmos e para conseguirmos interpretar os vários *inputs* que a operação possa fornecer. Estes pressupostos são por si mesmo uma previsão do comportamento das várias equipas. Esta previsão é feita com base em dados históricos, mas é importante entender as várias componentes humanas que podem intervir nas previsões para entender possíveis desvios.

Apesar das muitas variáveis envolvidas em todo o processo tanto de previsão como de capacidade, estamos apenas perante a ponta do *icebergue* do mundo de variáveis que podemos incluir, todas elas podendo enriquecer o *output* final. Como trabalho futuro podemos incluir, por exemplo, informações de possíveis iniciativas de marketing (que certamente terão impacto no volume de trabalho esperado) e transformar a nossa perspetiva bidimensional de previsão numa perspetiva multidimensional com múltiplas variáveis correlacionadas com o *workload*.

Apêndice A

Apêndice geral

A.1 Explicação de termos e conceitos utilizados

- *Response Rate*: Percentagem de clientes que são atendidos nos primeiros X segundos da chamada. Por exemplo, assumindo que o intervalo a ser utilizado é de 60 segundos, então todos os clientes que são atendidos nos primeiros 60 segundos de espera contribuem para o aumento da percentagem e todos os clientes que são atendidos depois dos primeiros 60 segundos de espera já não contribuem para a percentagem. O desejado seria atingir uma *Response Rate* de 100%, mas devido a limitações de serviço e processos normalmente é estipulado um objetivo entre 80% e 95%
- *Full Resolution Time*: Tempo médio em horas que cada caso leva a ser completamente resolvido. Normalmente o objetivo é estipulado entre 24h e 72h.
- *Isoweek*: Sistema de organização das semanas ao longo de um ano onde cada semana começa a uma segunda-feira e acaba a um domingo
- *Queue*: Lista de chegada de *e-mails* ou chamadas.

Bibliografia

- [1] R. Cleveland, W. Cleveland, J. McRae, and I. Terpenning, "Stl: A seasonal-trend decomposition procedure based on loess," *Journal of Official Statistics*, vol. 6, no. 1, pp. 3–73, 1990. [Cited on page 6.]
- [2] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed. <https://otexts.com/fpp3/>. [Cited on pages 6 and 8.]
- [3] A. M. D. Livera, R. J. Hyndman, and R. D. Snyder, "Forecasting time series with complex seasonal patterns using exponential smoothing," *Journal of the American Statistical Association*, vol. 106, no. 496, pp. 1513–1527, 2011. [Cited on page 9.]
- [4] S. J. Taylor, B. Letham, and G. I. Webb, "Forecasting at scale," <https://doi.org/10.7287/peerj.preprints.3190v2>, 2017. [Cited on page 10.]
- [5] S. Seo, "A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets," *Kyunghee University*, 2002. [Cited on page 17.]
- [6] S. Chaudhary. Knuth: Computers and typesetting. [Online]. Available: <http://www-cs-faculty.stanford.edu/~uno/abcde.html>
- [7] R. Godahewa, C. Bergmeir, G. I. Webb, and P. Montero-Manso, "A strong baseline for weekly time series forecasting," *arXiv:2010.08158*, 2020.
- [8] S. Makridakis, S. C. Wheelwright, and R. J. Hyndman, "Forecasting methods and applications," *Journal of the American Statistical Association*, vol. 94, no. 445, 1998.
- [9] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, 2006.
- [10] A. Adineh, Z. Narimani, and S. Satapathy, "Importance of data preprocessing in time series prediction using sarima: A case study," *International Journal of Knowledge-based and Intelligent Engineering Systems*, vol. 24, no. 4, pp. 897–902, 2020.

- [11] G. S. and Andreas Gützlaff, F. Sauer mann, and T. Theunissen, "Application of time series data mining for the prediction of transition times in production," *Procedia CIRP*, vol. 93, pp. 897–902, 2020.
- [12] R. J. Hyndman and Y. Khandakar, "Automatic time series forecasting: the forecast package for r," *Journal of Statistical Software*, vol. 27, no. 3, 2008.
- [13] A. Trapletti, K. Hornik, and B. LeBaron, *Time Series Analysis and Computational Finance*. CRAN, 2022.
- [14] P. S. Cowpertwait and A. V. Metcalfe, *Introductory Time Series with R (Use R!)*. Springer, 1981.
- [15] G. Koole, *Call Center Optimization*. MG Books Amestardam, 2013.
- [16] R. Sparkman, *Strategic Workforce Planning*. Kogan Page, 2009.
- [17] C. N. Knaflic, *Storytelling With Data Visualization Guide for Business Professionals*. John Wiley Sons, Inc., Hoboken, New Jersey, 2015.