

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL



**Ciências
ULisboa**

**Estratégias de realocização de bicicletas em serviços de
*bikesharing***

Débora Batista Mendes

Mestrado em Matemática Aplicada à Economia e Gestão

Trabalho de Projeto orientado por:
João Miguel Paixão Telhada

Agradecimentos

O término deste projeto significa a conclusão de mais um objetivo entre tantos que tenho e com toda a certeza irei concluir. Quero começar por agradecer à minha família pois tudo o que sou, todo o meu caminho, deve-se a eles. Quero agradecer à minha mãe por mais do que estar ao meu lado nas conquistas, estar sempre ao meu lado nos obstáculos pelos quais passei até alcançar as mesmas, fazendo sempre tudo o que pode para que eu seja feliz. Quero agradecer ao meu pai por esperar sempre o melhor de mim, pois isso dá-me motivação para ser mais e melhor. Quero agradecer à minha avó por sempre ter acreditado em mim, ajudando-me a seguir os meus sonhos.

Agradeço ao meu orientador, João Telhada por toda a ajuda e constante partilha de conhecimento. Pela facilidade com que desde o início se apresentou em ajudar-me sempre que preciso.

Agradeço aos meus amigos por todo o suporte dado ao longo da minha vida. Um gigantesco obrigada à Anastasiya Brykalova e Joana Costa, não só pela ajuda que sempre me deram com toda a vontade do mundo, mas também pelos momentos maravilhosos que tive o prazer de viver com vocês até aqui. Às minhas companheiras de longa data Raksana Udagedara e Jéssica Pascoal, obrigada por tudo.

Quero ainda agradecer a todos os docentes com quem tive o prazer de me cruzar ao longo da vida, pois cada passo dado na minha trajetória académica deve-se a todos eles.

Abstract

This project was based on the study of the bikesharing service, that is, the service of sharing bicycles. In order to make this type of service efficient, it's necessary to optimize the profits generated from the understanding and restructuring of the balance of bicycles between their collection points. For an initial framing, a research was made on the history and evolution of what led to the bicycle sharing that we have today and what are the factors that made it fail previously.

Throughout the project, a detailed analysis was made regarding the main factors that influence a certain influx on the search for bicycles on a specific collection point. These patterns are mainly based on aspects such as the period of the day, the days of the week, the location of stations, available bicycles, and on the other hand, available spots for relocation. After the identification of the main patterns, an extremely important study of the modeling was made, in order to better understand the searches and realization of the simulation. In the modeling, an analysis was made on the distribution of the intervals between searches and travel time, and naturally tests have been made as well, in order to understand the significance of the conclusions taken.

In the simulation we tried to understand to what extent the total number of bicycles has an influence on the satisfaction of demand, as well as to identify behaviors that can induce a redistribution that optimizes the satisfaction of demand and in turn, profits.

Finally, some suggestions were made in the terms of the information that would be interesting to have available for a future study, as well as in the perspective of service and service management.

Keywords: Bikesharing, Bicycle Relocation, Modeling, Simulation

Resumo

Este projeto centrou-se no estudo do serviço de *bikesharing*, isto é, no serviço de compartilhamento de bicicletas. Com o objetivo de tornar eficiente este tipo de serviço, é necessário otimizar os lucros gerados partindo da compreensão e eventual reestruturação do balanceamento de bicicletas entre os seus pontos de recolha. Para um enquadramento inicial, foi realizada uma pesquisa sobre a história e evolução do que levou ao compartilhamento de bicicletas que temos na atualidade e quais os fatores que falharam em compartilhamentos de bicicletas anteriores.

Ao longo de todo o projeto foi feita uma análise detalhada relativamente aos principais fatores que influenciam uma determinada afluência à procura das bicicletas num ponto de recolha específico. Estes padrões centram-se principalmente em aspetos como o período do dia, os dias da semana, a localização das estações, bicicletas disponíveis e por outro lado, vagas disponíveis para realocização.

Após a identificação dos principais padrões foi realizado um estudo da modelação extremamente importante para melhor compreensão das procuras e realização da simulação. Na modelação, foi feita uma análise de qual a distribuição que os intervalos entre procuras e os tempos de viagem seguem, e naturalmente realizados testes para compreender a significância das conclusões tomadas.

Na simulação tentou compreender-se até que ponto o número total de bicicletas tem influência na satisfação da procura, assim como identificar comportamentos que permitam induzir a uma redistribuição que otimize a satisfação da procura e por sua vez, os lucros.

Para finalizar foram feitas algumas sugestões na ótica da informação que seria interessante ter disponível para um futuro estudo, assim como na ótica de serviço e gestão de serviço.

Palavras-chave: Compartilhamento de bicicletas, estratégias de realocização, Modelação, Simulação

Índice

1.	Introdução.....	1
2.	Definição de um sistema de <i>bikesharing</i>	3
3.	Caso de estudo da Gira em Lisboa	7
3.1.	O que é a Gira.....	7
3.2.	Dados fornecidos.....	8
3.3.	Pré-processamento dos dados.....	10
4.	Análise de padrões das estações	13
4.1.	Análise de padrões para cada estação.....	13
4.1.1.	Representação em série temporal	13
4.1.2.	Análise de padrões temporais dos dados	15
4.1.3.	Localização das estações e fatores externos.....	17
4.2.	Análise de dados nas viagens	17
4.2.1.	Duração	17
4.2.2.	Distância das viagens	22
4.2.3.	Número de viagens.....	24
5.	Análise da procura.....	26
5.1.	Modelação dos intervalos entre procuras	26
5.2.	Modelação dos tempos de viagem.....	38
6.	Simulação e resultados computacionais	40
6.1.	Simulação	40
6.1.1.	Descrição geral do modelo de simulação	40
6.1.2.	Funcionamento do modelo de simulação	40
6.1.3.	Resultados do modelo de simulação.....	46
6.2.	Resultados Computacionais	46
7.	Conclusão	54
8.	Bibliografia.....	56
9.	Anexo	57

Índice de Equações

Equação 5.1 Cálculo dos intervalos entre procuras	18
Equação 5.2 Cálculo dos intervalos entre procuras	19
Equação 5.3 Cálculo da duração das viagens entre cada par de estações	21
Equação 6.1 Função densidade de uma variável aleatória X com distribuição Exponencial.....	26
Equação 6.2 Função densidade de uma variável aleatória X com distribuição Exponencial.....	26
Equação 6.3 Equação do estimador de máxima verosimilhança do parâmetro da distribuição exponencial.....	27
Equação 6.4 Função densidade da distribuição de Weibull	27
Equação 6.5 Função distribuição acumulada da distribuição de Weibull	27
Equação 6.6 Hipóteses do teste do Qui-quadrado	27
Equação 6.7 Regra de Mann e Wald: Classes com igual probabilidade	28
Equação 6.8 Regra de Mann e Wald: Valor esperado comum às classes maior ou igual a 5.....	28
Equação 6.9 Regra de Mann e Wald: Expressão e cálculo da primeira classe.....	28
Equação 6.10 Regra de Mann e Wald: Expressão e cálculo da segunda classe	28
Equação 6.11 Regra de Mann e Wald: Expressão e cálculo da (k-1)-ésima classe.....	28
Equação 6.12 Regra de Mann e Wald: Expressão e cálculo da k-ésima classe.....	28
Equação 6.13 Regra de Mann e Wald: Cálculos intermédios para determinar as classes pela regra de Mann e Wald	28
Equação 6.14 Regra de Mann e Wald: Primeira classe	28
Equação 6.15 Regra de Mann e Wald: Segunda classe.....	28
Equação 6.16 Regra de Mann e Wald: (k-1)-ésima classe	28
Equação 6.17 Regra de Mann e Wald: k-ésima classe	28
Equação 6.18 Estatística de teste com aplicação da regra de Mann e Wald.....	28
Equação 6.19 Distribuição da estatística de teste do teste do qui-quadrado de Pearson.....	29
Equação 6.20 Hipótese de teste para o teste de Kolmogorov-Smirnov.....	29
Equação 6.21 Função distribuição empírica associada a uma amostra aleatória de X no teste de Kolmogorov-Smirnov.....	29
Equação 6.22 Estatística de teste do teste de Kolmogorov Smirnov.....	30
Equação 6.23 Estatística de teste do teste de Kolmogorov Smirnov.....	30
Equação 6.24 Hipóteses do teste de Kolmogorov-Smirnov no caso particular da distribuição exponencial.....	30
Equação 8.1 Expressão da taxa de serviço	47
Equação 8.2 Expressão do Lucro.....	53

Índice de Figuras

Figura 2.1 Exemplificação de um sistema de bikesharing	3
Figura 2.2 Exemplificação da capacidade de cada estação num sistema de bikesharing	3
Figura 2.3 Exemplificação da ocupação das bicicletas nas estações num sistema de bikesharing	4
Figura 2.4 Exemplificação da escolha de um utilizador da estação de partida num sistema de bikesharing	4
Figura 2.5 Exemplificação do percurso de um utilizador após o início de uma viagem num sistema de bikesharing	5
Figura 2.6 Exemplificação de uma viagem realizada num sistema de bikesharing	5
Figura 2.7 Exemplo de um gráfico de uma série temporal.....	6
Figura 3.1 Mapa das estações da empresa Gira.....	7
Figura 3.2 Exemplo da informação disponível após a seleção de uma estação	8
Figura 3.3 Série temporal da estação 105 adotando o primeiro procedimento	10
Figura 3.4 Série temporal da estação 105 adotando o segundo procedimento.....	11
Figura 4.1 Série Temporal das observações registadas do número de bicicletas no decorrer do ano 2018 da estação 105-CC Vasco da Gama.....	13
Figura 4.2 Série temporal da estação 105 no decorrer do dia 25 de março de 2018	14
Figura 4.3 Série Temporal das observações registadas do número de bicicletas no decorrer do ano 2018 da estação 109	14
Figura 4.4 Número médio das bicicletas referentes à estação 105-CC Vasco da Gama, no decorrer do ano de 2018, dispostas em blocos de horas	15
Figura 4.5 Número médio de bicicletas referentes ao ano 2018 na estação 105-CC Vasco da Gama, dispostos em dias da semana	16
Figura 4.6 Número médio de bicicletas referentes ao ano 2018 na estação 105-CC Vasco da Gama, dispostos em meses	17
Figura 4.7 Representação do problema existente no cálculo dos intervalos entre procuras.....	18
Figura 4.8 Série Temporal do número de bicicletas na estação 105-CC Vasco da Gama após as devidas alterações mencionadas no Capítulo 2	18
Figura 5.1 Histograma dos intervalos entre procuras	26
Figura 5.2 Construção da estatística de teste do teste de Kolmogorov-Smirnov	30
Figura 5.8 Histograma, função densidade da exponencial e função densidade de weibull do intervalo entre procuras no período das 10h-13h	34
Figura 6.1 Esquema do modelo de simulação	40
Figura 6.2 Exemplificação dos estados	46
Figura 6.3 Cálculo do número médio de bicicletas no instante $[t, t+1[$	51
Figura 6.4 Esquemática dos dados usados nas estratégias.....	51

Índice de Tabelas

Tabela 3.1 Exemplo de um registo do ficheiro dados1	9
Tabela 3.2 Exemplo de um registo do ficheiro dados2	10
Tabela 3.3 Primeiros registos da estação 105 com lapsos	10
Tabela 3.4 Exemplo do ficheiro dados2 no período em que ocorrem procuras na estação 105 entre as 22h09 e as 22h14 no dia 22/02/2018.....	11
Tabela 3.5 Exemplo do ficheiro dados2 no período em que ocorrem entregas na estação 105 entre as 22h09 e as 22h14 no dia 22/02/2018.....	12
Tabela 4.1 Estatísticas descritivas relativamente às observações registadas no decorrer do ano 2018 da estação 105-CC Vasco da Gama	15
Tabela 4.2 Tabela Descritiva da duração dos intervalos entre procuras de bicicletas na estação 105-CC Vasco da Gama.....	19
Tabela 4.3 Tabela Descritiva da duração dos intervalos entre entregas de bicicletas na estação 105-CC Vasco da Gama.....	19
Tabela 4.4 Tabela Descritiva da duração dos intervalos entre procuras na estação 105- CC Vasco da Gama	20
Tabela 4.5 Tabela Descritiva da duração dos intervalos entre entregas na estação 105- CC Vasco da Gama	20
Tabela 4.6 Tabela Descritiva da duração dos Intervalos entre procuras na estação 101-Alameda dos Oceanos-Rua dos Argonautas	20
Tabela 4.7 Tabela Descritiva da duração dos intervalos entre entregas na estação 101-Alameda dos Oceanos-Rua dos Argonautas	20
Tabela 4.8 Parte 1 da tabela descritiva da duração média das viagens tendo como ponto de partida a estação 105-CC Vasco da Gama	21
Tabela 4.9 Parte 2 da tabela descritiva da duração média das viagens tendo como ponto de partida a estação 105-CC Vasco da Gama	21
Tabela 4.10 Tabela Descritiva da duração entre viagens com a estação 105-CC Vasco da Gama como origem.....	22
Tabela 4.11 Parte 1 da tabela descritiva da duração média das viagens tendo como ponto de partida a estação 105-CC Vasco da Gama	22
Tabela 4.12 Parte 1 da tabela descritiva da distância das viagens tendo como ponto de partida a estação 105-CC Vasco da Gama.....	23
Tabela 4.13 Parte 2 da tabela descritiva da duração média das viagens tendo como ponto de partida a estação 105-CC Vasco da Gama	23
Tabela 4.14 Parte 2 da tabela descritiva da distância das viagens tendo como ponto de partida a estação 105-CC Vasco da Gama.....	23
Tabela 4.15 Parte 1 da Tabela do número de viagens para cada estação, tendo como ponto de partida a estação 105-CC Vasco da Gama	24
Tabela 4.16 Parte 2 da Tabela do número de viagens para cada estação, tendo como ponto de partida a estação 105-CC Vasco da Gama	25
Tabela 5.1 Histograma dos intervalos entre procuras com 14 classes para todos os dias da semana. 31	
Tabela 5.2 Histograma dos intervalos entre procuras com 7 classes para todos os dias da semana... 31	
Tabela 5.3 Tabela do histograma de 7 classes da estação 105-CC Vasco da Gama respetivo ao intervalo entre as 10h e 13h, para todos os dias da semana	32
Tabela 5.4 Estatísticas descritivas da estação 105-CC Vasco da Gama respetivo ao intervalo entre as 10h e as 13h, para todos os dias da semana.....	32
Tabela 5.5 Teste do Qui-quadrado em vários períodos para verificar se os intervalos entre procuras seguem uma distribuição exponencial nos respetivos períodos	33
Tabela 5.6 Teste de Kolmogorov-Smirnov ao intervalo entre procuras no período entre as 10 e as 13 horas	33

Tabela 5.7 Teste do Qui-quadrado em vários períodos para verificar se os intervalos entre procuras seguem uma distribuição de weibull nos respetivos períodos	35
Tabela 5.8 Parte 1 da Tabela com os valores estimados e o p-value para todos os intervalos entre procuras possíveis ao longo do dia na estação 105 seguindo a distribuição de Weibull.....	36
Tabela 5.9 Valores estimados e o p-value para todos os intervalos entre procuras possíveis ao longo do dia na estação 105 seguindo um distribuição exponencial	36
Tabela 5.10 Intervalos a considerar na simulação que seguem a distribuição de Weibull.....	37
Tabela 5.11 Tabela com a distribuição categórica da estação 105	37
Tabela 5.12 Probabilidade de destinos onde a estação 101 é a estação de partida.....	38
Tabela 5.13 Duração mínima e máxima das viagens realizadas partindo da estação 101 até a estação 453	38
Tabela 6.1 Ficheiro autoMulti110 com intervalos do tipo 1	41
Tabela 6.2 Ficheiro autoMulti102 com intervalos do tipo 2	42
Tabela 6.3 Ficheiro autoMulti487 com intervalos do tipo 2	43
Tabela 6.4 Parte da agenda simulada	44
Tabela 6.5 Parte das viagens realizadas com base na agenda da Tabela 7.4	45
Tabela 6.6 Parte da agenda atualizada após algumas viagens.....	45
Tabela 6.7 Taxa de serviço de 3 simulações realizadas com base num sistema com 384 bicicletas.....	47
Tabela 6.8 Taxa de serviço em sistemas com uma dimensão de bicicletas diferente	47
Tabela 6.9 Percentagem de estados únicos de 3 simulações realizadas com base num sistema com 384 bicicletas.....	48
Tabela 6.10 Percentagem de estados únicos em sistemas com uma dimensão de bicicletas diferente .	48
Tabela 6.11 Percentagem de estados únicos das 3 simulações realizadas com base num sistema com 384 bicicletas juntas	48
Tabela 6.12 Taxa de serviço acima dos 90% e respetivas estações	49
Tabela 6.13 Taxa de serviço abaixo dos 25% e respetivas estações	49
Tabela 6.14 Percentagens de tempo em que as estações se encontram sem bicicletas inferior a 12 % e respetivas estações.....	50
Tabela 6.15 Percentagens de tempo em que as estações se encontram sem bicicletas superior a 50% e respetivas estações.....	50
Tabela 6.16 Taxa de Serviço da estratégia 1, estratégia 2 e estratégia 3	52

1. Introdução

Neste capítulo é feito um enquadramento sobre o serviço de *bikesharing* onde é explicado no que consiste o serviço de *bikesharing*, o seu contexto histórico, quais as suas vantagens. São explicados os objetivos deste trabalho e é feito um pequeno resumo sobre as tarefas realizadas neste projeto.

Bikesharing é um serviço de compartilhamento de bicicletas no qual um utilizador do serviço aluga uma bicicleta num determinado período.

Este serviço seguiu quatro fases principais, também designadas por quatro gerações distintas. A primeira geração designa-se como “*White Bikes*”, (também conhecida como “*Free Bikes*”), onde o sistema de compartilhamento de bicicletas consiste num sistema onde as bicicletas se encontram disponíveis gratuitamente pela cidade, sem qualquer custo atribuído e cuja utilização é de tempo ilimitado. A primeira geração teve início na década de 60, nos Países Baixos e Escandinávia (o primeiro sistema deste tipo foi lançado em Amesterdão, em 1965). O insucesso desta geração rapidamente ocorreu devido a vandalismo, roubos e falta de manutenção que acabaram por tornar o serviço ineficiente.

A segunda geração do serviço de *bikesharing* teve início em Copenhaga, em 1995, e o seu sistema consistia no estabelecimento de estações fixas, nas quais ao contrário da primeira geração só seria possível retirar e devolver as bicicletas através do depósito de moedas. Este pequeno avanço tecnológico inicialmente surtiu alguns efeitos, pois ganhou algum interesse comercial através de patrocínios privados. Contudo, este sistema de compartilhamento de bicicletas continuou a ser vulnerável uma vez que o valor pouco significativo dos depósitos e o anonimato dos utilizadores não era suficientemente relevante para garantir a devolução das bicicletas.

A terceira geração, conhecida também como um sistema baseado em Tecnologia da informação foi lançada em Rennes, França em 1998. Nesta geração, o utilizador efetua o seu registo no sistema e faz o respetivo depósito com recurso a uma determinada aplicação, sendo ainda possível definir a sua rota, ter conhecimento da disponibilidade de bicicletas e vagas nas estações da sua rede. Por outro lado, o sistema desta geração é igualmente benéfico para os operadores uma vez que têm acesso às procuras realizadas podendo assim fazer um estudo de redistribuição logística que otimize a resposta à procura, através da identificação de certos padrões do serviço.

Na quarta e atual geração as bicicletas são elétricas, tem um sistema de travamento automático e as estações contem painéis solares para recarregar as bicicletas.

A notoriedade do serviço de *bikesharing* começou em 2007, quando dois relevantes sistemas de compartilhamento de bicicletas foram lançados: o *Vélib* com 10000 bicicletas, em Paris, e o *Bicing* com 1500 bicicletas, em Barcelona.

Em 2014 havia já mais de 700 cidades que possuíam sistemas de compartilhamento de bicicletas, sendo o *Vélib* o maior na Europa, com mais de 20 mil bicicletas e cobrindo a totalidade da cidade de Paris. Ainda em 2014 o maior sistema do mundo localiza-se em Wuhan, na China, usufruindo de mais de 90 mil bicicletas.

É possível encontrar informação mais detalhada em [3] e [8].

O crescimento da adesão ao *bikesharing* tem sido bastante significativo no decorrer dos últimos anos. Tal deve-se às vantagens que este tipo de serviço acarreta para o utilizador.

Numa civilização moderna e ocupada, que vive numa correria diária e constante, o tempo é um bem essencial e por isso valioso. Um dos fatores mais importantes na produtividade de cada indivíduo é uma utilização eficiente do seu tempo. Logicamente, o meio com que cada pessoa adota para se deslocar entre os seus afazeres reflete uma utilização eficaz ou redundante do seu tempo. Felizmente, atualmente existem várias e diferentes opções que permitem uma deslocação adequada a cada indivíduo como os transportes públicos, veículos próprios, o serviço de compartilhamento de bicicletas, entre outros.

A questão mais pertinente a se colocar é qual o benefício neste tipo de serviço face aos demais. Na verdade, existem várias vantagens no serviço de *bikesharing* face aos restantes serviços de transportes

mencionados anteriormente. Por exemplo, a aquisição de um veículo próprio acarreta vários custos. Os custos na aquisição de um veículo são diversos, como perda de tempo significativo em encontrar estacionamento, uma contribuição para o aumento da poluição, um uso irracional do veículo (os carros em média encontram-se parados 90% do tempo), entre outros. Por outro lado, os transportes públicos têm uma cobertura de serviço muitas vezes insuficiente, os horários são inflexíveis e são fortemente condicionados por eventuais greves.

O serviço de *bikesharing* consegue contornar muitos aspetos negativos dos exemplos mencionados anteriormente. O serviço de *bikesharing* engloba uma deslocação racional, conservando, tal como os transportes públicos o meio ambiente com o impacto que reflete, por exemplo na redução de utilização de veículos, contribuindo assim para a redução de dióxido de carbono. Contempla ainda um modo fácil de proceder a uma eventual reserva e devolução (sem que seja necessária a intervenção de terceiros), sempre consciente de que o único custo a acarretar é o já inicialmente proposto.

O principal objeto de estudo na elaboração de modelos de otimização e simulação de serviços de *bikesharing* é conseguir contornar a curto e longo prazo o balanceamento entre bicicletas nas respetivas áreas de operação. É necessário garantir que a procura é maioritariamente satisfeita, assim como oferecer a possibilidade ao cliente de ter a liberdade de se dirigir para qualquer área de operação, assegurando que em princípio terá uma vaga disponível para devolução da bicicleta. Todos os atos referidos anteriormente devem se realizar recorrendo a uma menor quantidade de gastos possíveis de modo a maximizar o lucro das empresas que prestam este tipo de serviços.

Dito isto, uma das tarefas iniciais será, portanto, compreender os modelos que já anteriormente foram realizados no estudo deste serviço, conseguindo conservar os benefícios dos mesmos e se possível, contornar as suas falhas, sendo esta uma fase fulcral desta tese. Pretende-se então garantir o equilíbrio de bicicletas entre estações de modo a otimizar a qualidade de serviço prestado aos clientes, tendo como vantagem um maior lucro para a empresa.

Os dados utilizados neste estudo são usados para realizar uma estimativa realista das oscilações de bicicletas em tempo real. Estas estimativas são extremamente importantes para o desenvolvimento de um bom modelo. Os dados utilizados são referentes à empresa Gira, que liberou dados referentes ao ano 2018 para o desafio de otimização das rotas de rebalanceamento das bicicletas partilhadas, proposto pela Câmara Municipal de Lisboa.

Com este trabalho pretende-se compreender de forma clara e global o serviço de *bikesharing* e o seu impacto na mobilidade urbana e determinar um sistema de apoio à decisão para a relocalização de bicicletas, associado a métricas de desempenho de negócio.

Assim, será realizado primeiramente um levantamento e contextualização de *bikesharing*, seguindo-se a identificação e modelação de padrões de procura. Os dados disponibilizados pela empresa Gira denotam um papel essencial na fase de identificação de padrões. Posteriormente, é realizada a construção de um modelo de desenvolvimento e visualização. Por fim, serão realizados testes computacionais para avaliação do desempenho e da robustez do modelo.

Toda a análise deste trabalho teve como base o recurso a uma ferramenta construída através de uma aplicação em R, que permite ter uma visualização mais detalhada sobre cada estação. É importante ter em conta que embora ao longo deste projeto seja apresentado um número restrito de estações como exemplo, esta ferramenta permite uma análise automática para todas as estações. Além da ferramenta estar adaptada à análise de todas as estações, apresenta ainda a possibilidade de manipular diversos fatores nos dados em análise (ver anexo).

2. Definição de um sistema de *bikesharing*

Para um melhor desenvolvimento do projeto é importante começar-se por compreender detalhadamente o funcionamento de um sistema de *bikesharing*. Neste capítulo será apresentado o funcionamento de tal sistema explicando quais as principais informações a reter durante o processo e como é feito o registo dessa informação.

Num sistema de *bikesharing* existe um conjunto de estações, onde se encontram bicicletas disponíveis para utilização. Considere-se um sistema com n estações, sendo n estações designadas por S_i , onde $i=1, \dots, n$ estão espalhadas fixamente numa determinada região.

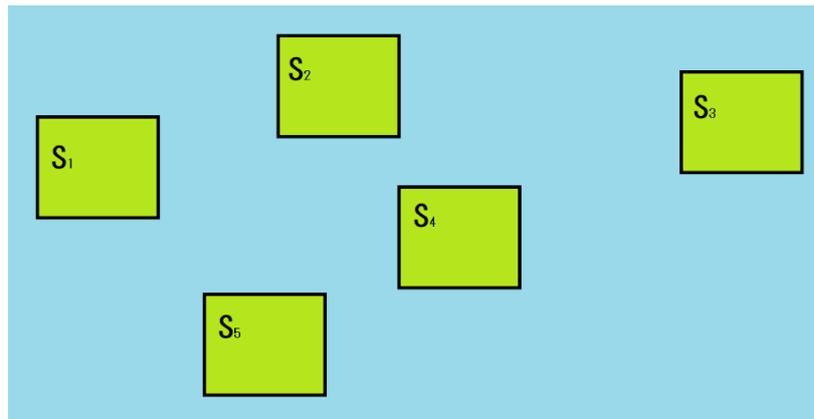


Figura 2.1 Exemplificação de um sistema de *bikesharing*

Na Figura 2.1, a zona colorida a azul representa a região onde as estações de um determinado sistema de *bikesharing* se encontram, já os quadrados verdes representam as estações. No exemplo ilustrado na Figura 2.1 está a ser considerado um sistema com apenas 5 estações, isto é, entende-se que n é igual a 5.

Cada estação, por sua vez, tem uma respetiva capacidade para recolha de bicicletas, isto é, existe um limite de quantas bicicletas se podem encontrar por estação. Designe-se esta capacidade como c_i , onde $i = 1, \dots, n$.

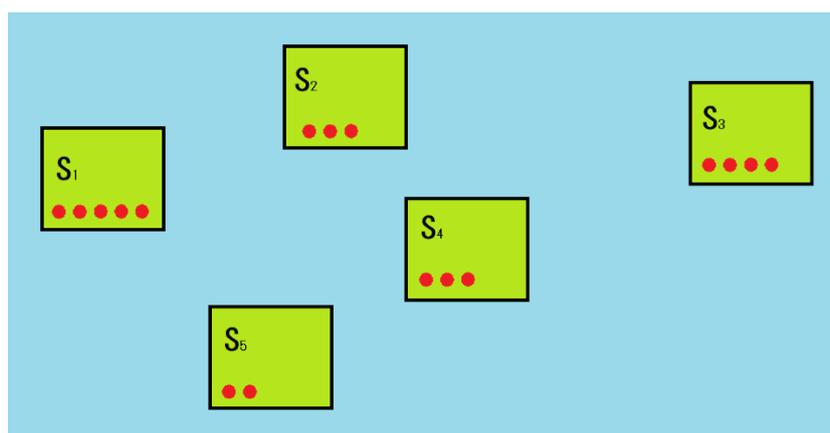


Figura 2.2 Exemplificação da capacidade de cada estação num sistema de *bikesharing*

Na Figura 2.2, os pontos ilustrados a vermelho representam a capacidade de cada estação. Pela exemplificação da Figura 2.2 entende-se que a estação S_1 tem uma capacidade designada por c_1 igual a 5. Por outro lado, observa-se que as estações S_2 e S_4 têm, respetivamente, as capacidades c_2 e c_4 iguais a 3. A estação S_3 tem uma capacidade c_3 igual a 4 e a estação S_5 tem uma capacidade c_5 igual a 2.

Num determinado instante no tempo existe um certo número de bicicletas em cada estação. Designe-se o número de bicicletas no instante t como $y_{i,t}$ onde $i = 1, \dots, n$ e $t = 1, \dots, m$, sendo m o número total de instantes observados.

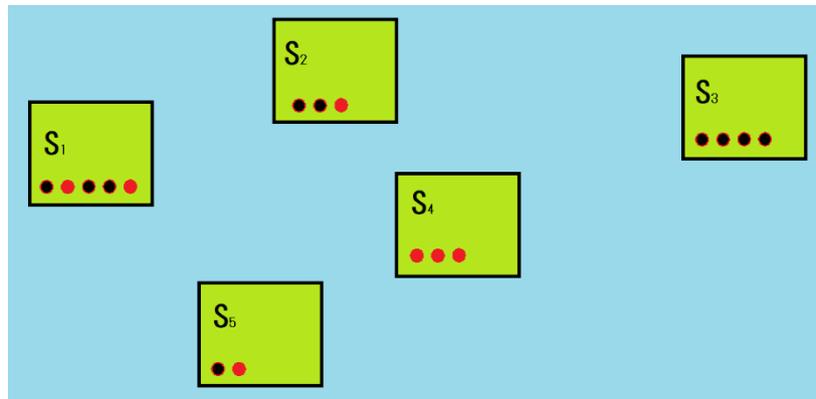


Figura 2.3 Exemplificação da ocupação das bicicletas nas estações num sistema de bikesharing

Na Figura 2.3 os pontos a preto representam as bicicletas presentes nas respetivas estações, enquanto que os pontos vermelhos representam as vagas disponíveis para a alocação de mais bicicletas. Pelo exemplo da Figura 2.3 sabe-se que, por exemplo, na estação S_1 estão disponíveis 3 bicicletas para utilização e que existem 2 vagas para a entrega de mais bicicletas. Por outro lado, na estação S_3 existem 4 bicicletas disponíveis, mas não é possível entregar nesta estação mais bicicletas no instante em causa. Na estação S_4 não existem, no instante em causa, bicicletas disponíveis para utilização.

Assim, um utilizador para usufruir deste serviço deve recorrer a uma aplicação num dispositivo móvel. Através da aplicação consegue ter acesso sobre quais as estações que se encontram mais próximas de si, bem como se existem bicicletas disponíveis nessas estações, nesse instante. Na j -ésima viagem o utilizador realiza então uma escolha sobre qual a estação de partida da sua viagem, que será designada por e_j .

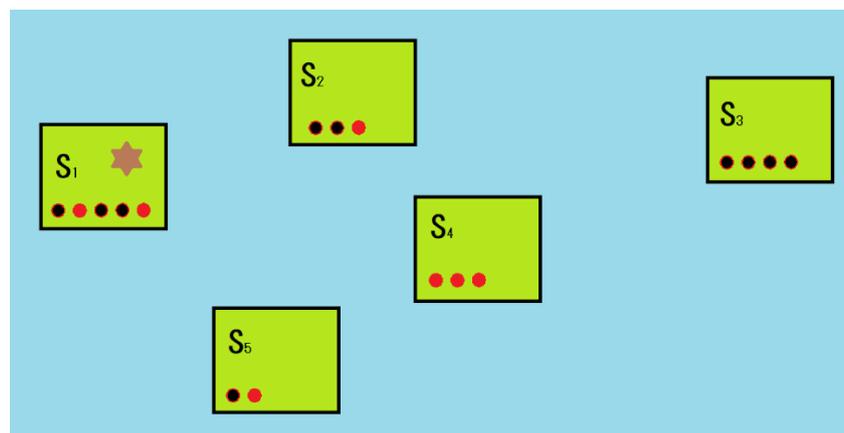


Figura 2.4 Exemplificação da escolha de um utilizador da estação de partida num sistema de bikesharing

Na Figura 2.4 a estrela colorida a castanho representa o utilizador. Esta ilustração representa a viagem k , onde o utilizador escolhe como estação de partida a estação S_1 , ou seja, entende-se que e_k é igual a S_1 . O instante em que o utilizador recolhe a bicicleta na estação de partida designa-se como a_k e determina o tempo em que se inicia a viagem.

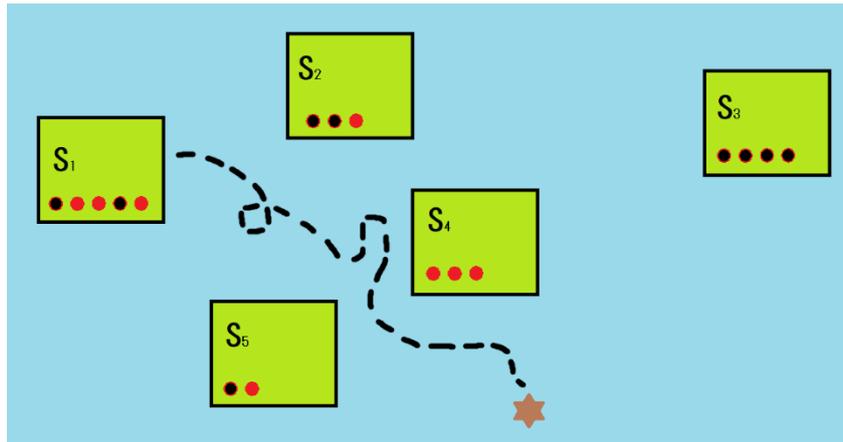


Figura 2.5 Exemplificação do percurso de um utilizador após o início de uma viagem num sistema de bikesharing

O tracejado ilustrado na Figura 2.5 representa o percurso percorrido pelo utilizador num dado momento, ainda durante a viagem. Note-se que a estação de partida e_j que no exemplo que ilustra a viagem k , designa-se por e_k é a estação S_1 , tem agora apenas 2 bicicletas disponíveis para utilização. Cada viagem realizada acaba por ter uma determinada distância percorrida. Designe-se a distância percorrida na j -ésima viagem como d_j , no caso do exemplo da viagem k como d_k .

Para dar como encerrada a viagem, o utilizador deve devolver a bicicleta numa das n estações presentes na região que se está a considerar. Note-se que a estação onde a bicicleta será entregue poderá coincidir ou não com a estação de partida. A estação onde a bicicleta é entregue na j -ésima viagem designa-se como estação de chegada f_j .

Existem duas situações que podem acontecer ao utilizador relativamente á estação mais próxima do local onde pretende devolver a bicicleta:

- A primeira situação ocorre quando a estação mais próxima não tem qualquer vaga disponível para a realocação da bicicleta. Por exemplo, considerando o exemplo da Figura 2.5, se a estação onde o utilizador pretende devolver a bicicleta for a estação S_3 , tal não será possível pois não existem vagas disponíveis.
- A segunda situação ocorre quando a estação mais próxima tem vagas disponíveis para realocar a bicicleta. Na Figura 2.5, à exceção da estação S_3 , todas as restantes estações estão naquele instante disponíveis para o utilizador entregar a bicicleta.

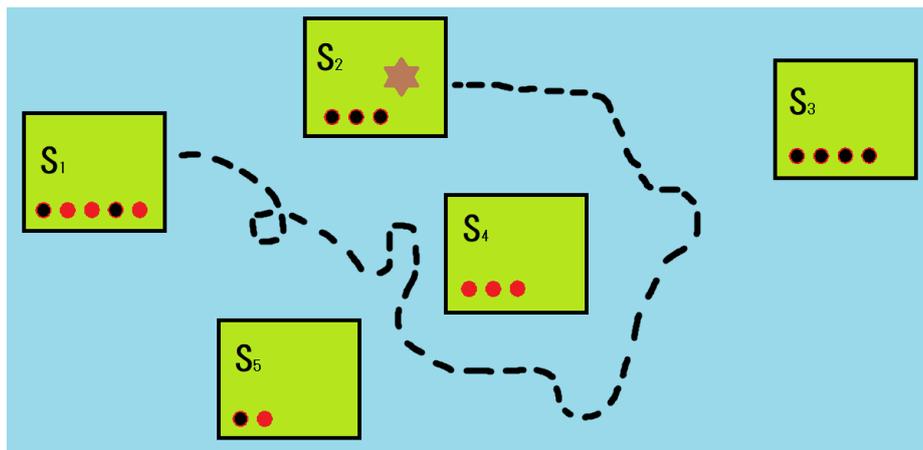


Figura 2.6 Exemplificação de uma viagem realizada num sistema de bikesharing

Na Figura 2.6 está ilustrado um exemplo cuja estação de origem e_k é a estação S_1 e a estação de chegada f_k é a estação S_2 . Note-se que após a chegada da bicicleta á estação S_2 , deixam de existir vagas para a realocação de futuras bicicletas, até uma bicicleta dessa estação ser recolhida por outro utilizador. Designe-se por b_j o instante em que uma bicicleta é entregue na j -ésima viagem. Na viagem k , do mesmo modo que a bicicleta é requerida no instante a_k é entregue no instante de chegada b_k . Mais uma vez a

devolução da bicicleta por parte do utilizador é realizada através da aplicação referida inicialmente. Em suma, um sistema de *bikesharing* consiste em várias observações simultâneas ao longo do tempo. A requisição e devolução de bicicletas traduzem a imprevisibilidade sobre quantas bicicletas estão em cada estação num determinado instante, pois a distância e duração da viagem são inicialmente desconhecidos.

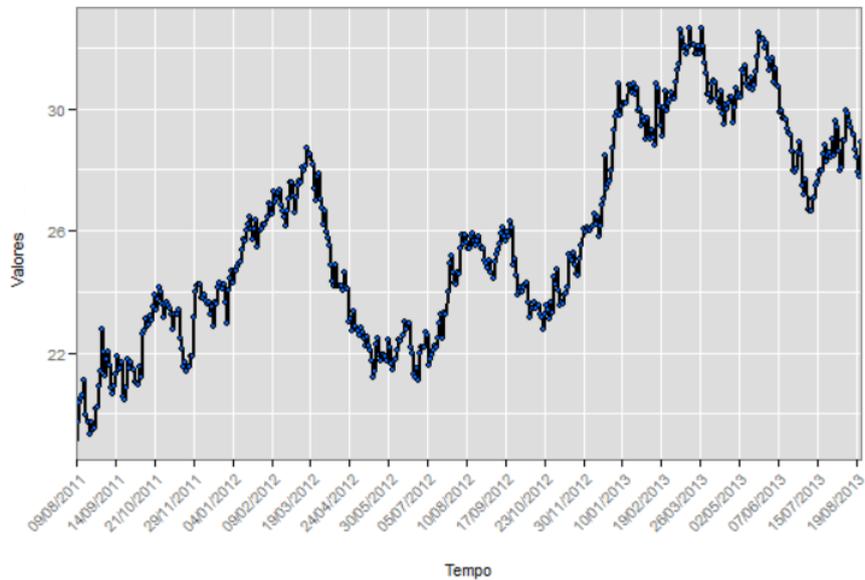


Figura 2.7 Exemplo de um gráfico de uma série temporal

A Figura 2.7 ilustra o exemplo de um gráfico de uma série temporal, onde os dados são dispostos ao longo do tempo. O eixo horizontal representa o tempo e o eixo vertical as observações associadas ao tempo. De modo a compreender melhor o comportamento do número de bicicletas numa estação ao longo do tempo, mais tarde serão utilizadas séries temporais. Uma série temporal é um conjunto de observações feitas ao longo do tempo, sendo a ordem das observações fundamental. Uma característica muito importante deste tipo de dados é que as observações vizinhas são dependentes. Neste caso as observações correspondem ao número de bicicletas e o interessante desta representação é analisar e modelar essa dependência de modo a conseguir deduzir qual o número de bicicletas no instante seguinte, sabendo o número de bicicletas num determinado instante presente.

3. Caso de estudo da Gira em Lisboa

Neste capítulo realiza-se uma breve abordagem sobre a empresa que disponibilizou os dados utilizados ao longo deste projeto. É explicada a informação contida nos ficheiros usados na construção deste projeto. São também apresentados os procedimentos tomados relativamente ao tratamento que os dados sofreram, algo que foi necessário para dar continuidade ao projeto.

3.1. O que é a Gira

A Gira é um serviço de bicicletas partilhadas de Lisboa. Os principais propósitos da Gira centram-se em tornar Lisboa numa cidade mais acessível, menos poluída e com menos ruído. Têm 60 quilómetros de ciclovias e brevemente este valor subirá para 150 quilómetros, tendo como objetivo alargar a área interveniente. Têm à disposição dos utilizadores bicicletas clássicas e elétricas, as quais se pode recolher utilizando a aplicação do Gira. Através da app o utilizador poderá escolher o seu passe, desbloquear bicicletas, encontrar estações e muito mais.

Assim, para recorrer ao serviço da Gira deve-se seguir os seguintes passos:

- a) Descarregar a aplicação;
- b) Escolher o passe, este pode ser anual, mensal ou diário;
- c) Encontrar a estação, tendo conhecimento de qual a doca mais próxima do utilizador através da aplicação, sabendo as bicicletas que estão disponíveis;
- d) Escolher a bicicleta, tendo a opção de escolher elétrica ou clássica. No caso de a escolha ser a elétrica é disponibilizada informação quanto à bateria da bicicleta;
- e) Selecionar a bicicleta, devendo assinalar na aplicação a bicicleta que pretende utilizar e de seguida retirar a mesma da doca, podendo dar início à viagem. O utilizador tem 15 segundos para realizar a recolha da bicicleta;
- f) Realizar a viagem;
- g) Devolver a bicicleta, arrumando a bicicleta na doca mais próxima confirmando que a viagem terminou. Neste passo deve-se ter em atenção os lugares vagos em cada estação, informação também esta disponível na app;
- h) Avaliar a viagem, sendo pedido um feedback ao utilizador relativamente à viagem realizada;

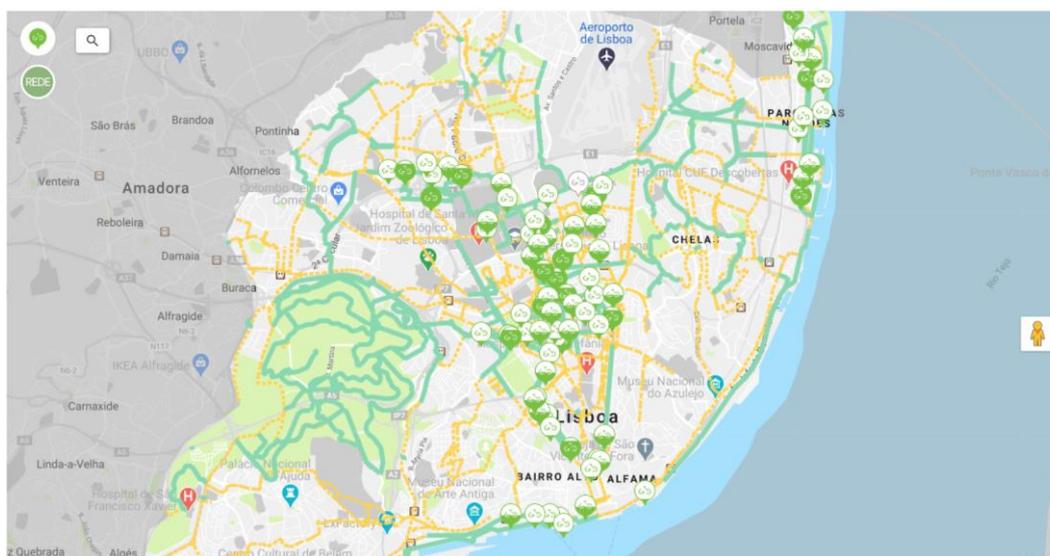


Figura 3.1 Mapa das estações da empresa Gira

A Figura 3.1 ilustra a distribuição das estações da Gira, onde cada sinalização verde representa uma estação. Por outro lado, as linhas verdes representam as ciclovias. Ao clicar numa determinada estação é disponibilizada ao utilizador mais informação sobre a estação selecionada.

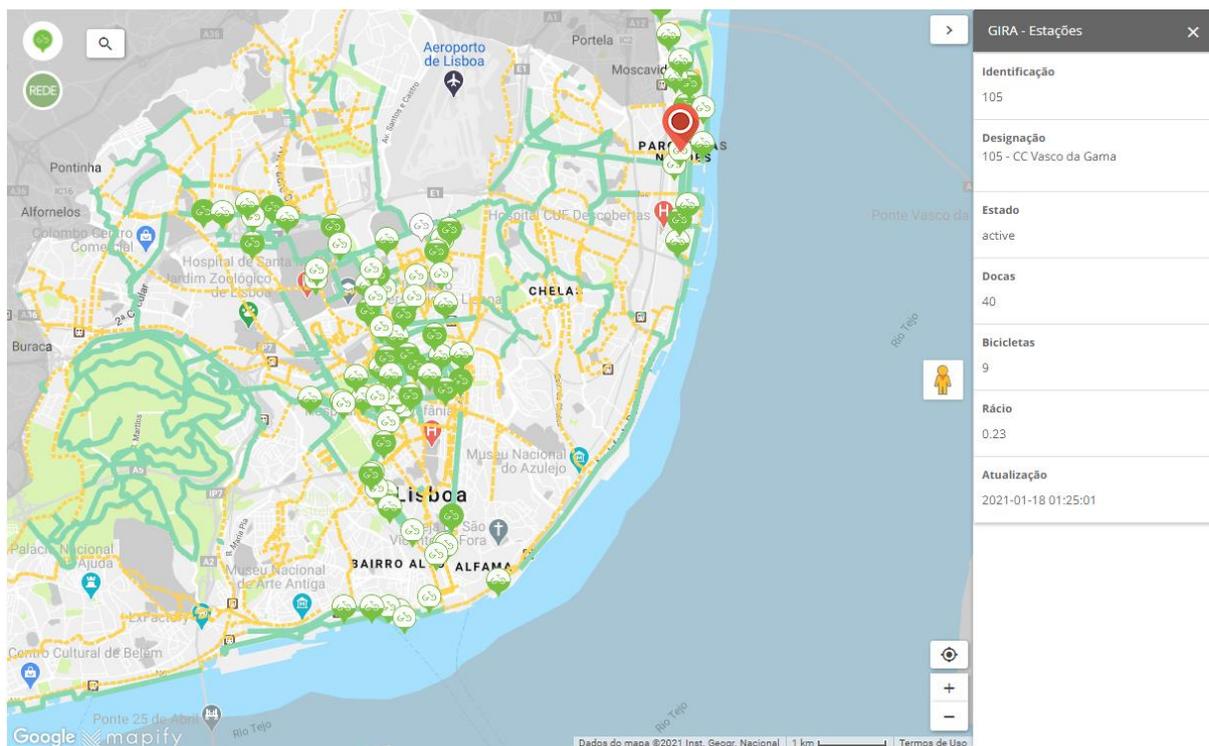


Figura 3.2 Exemplo da informação disponível após a seleção de uma estação

A Figura 3.2 ilustra um exemplo da informação que é disponibilizada ao selecionar uma estação. Por exemplo, ao clicar na estação 105 na atualização registada às 01h25m01 no dia 18 de janeiro sabe-se que a estação 105 – CC Vasco da Gama estava ativa e tinha 9 bicicletas disponíveis. É ainda dada a informação sobre a capacidade desta estação, neste caso sabe-se que a capacidade da estação 105 é igual a 40. Uma vez que existem 9 bicicletas presentes na estação 105 e a sua capacidade é de 40 bicicletas existem $40-9=31$ vagas para a realocação de mais bicicletas no instante da atualização. O rácio corresponde à proporção de vagas ocupadas na estação, isto é, $9/40=0,23$.

3.2. Dados fornecidos

Para uma boa gestão do serviço de *bikesharing* é necessário realizar vários registos sobre todo o processo. Assim, é importante também registar quantas bicicletas se encontram disponíveis em cada estação, assim como quantas vagas para entrega de bicicletas encontram-se livres.

Existem dois ficheiros utilizados na elaboração deste projeto. Designemos os ficheiros por **dados1** e **dados2** respetivamente. Ambos os ficheiros podem ser interpretados como uma tabela, onde cada coluna representa um tipo de informação específica, e cada linha representa um registo.

As observações do ficheiro **dados1** ocorrem entre dia 23 de fevereiro de 2018 às 22h09 e dia 10 de Junho de 2018 às 9h45. No ficheiro **dados1** entende-se como registo a anotação das informações respetivas a cada coluna relativamente a uma determinada estação, num determinado instante. O ficheiro **dados1** tem 11 colunas.

- 1) A coluna **id** onde os valores são algarismos distintos em cada linha e representam a identificação de cada registo;
- 2) A coluna **geom** onde os valores são representados por letras e algarismos. Não foi obtida qualquer explicação sobre este campo;

- 3) A coluna **id_expl** é representada por algarismos e representa a identificação numérica de cada estação;
- 4) A coluna **id_planeamento** é idêntica à coluna **id_expl**;
- 5) A coluna **desig_comercial** é representada pelo **id_expl** e o nome da estação respetiva ao **id_expl**;
- 6) A coluna **tipo_serviço_niveis** é representada por uma letra que pode assumir os valores de A ou B;
- 7) A coluna **num_bicicletas** é representada por algarismo e representa o número de bicicletas na estação identificada pela coluna **id_expl** em determinado instante;
- 8) A coluna **num_docas** é representada por algarismos e representa o número total de vagas em funcionamento para colocar as bicicletas em cada estação;
- 9) A coluna **rácio** é representada por algarismos e identifica o rácio de bicicletas na estação identifica pela coluna **id_expl** num determinado instante;
- 10) A coluna estado pode assumir os valores “*active*” no caso em que a estação está aberta sem quaisquer inconvenientes, “*repair*” no cado da estação estar sobre algum tipo de manutenção e “*stock*”;
- 11) A coluna **update_date** é representada por uma data e hora. Esta coluna representa o horário em que é feito cada registo. O registo do número de bicicletas por cada estação é feito de 5 em 5 minutos;

id	geom	id_expl	id_planeamento	desig_comercial	tipo_serviço_niveis	num_bicicletas	num_docas	rácio	estado	update_date
157	I101000000D55C6	104	104	104-Gare do Oriente	A	23	40	0.58	active	2018-02-23 22:09:16

Tabela 3.1 Exemplo de um registo do ficheiro dados1

Na Tabela 3.1 está ilustrado um exemplo de uma das várias linhas que constituem o ficheiro **dados1**. Através da linha exemplificada na Tabela 3.1 sabe-se que o registo com id igual a 157 corresponde a um registo da estação 104 de seu nome Gare do Oriente no dia 23 de fevereiro de 2018 às 22h09m6, altura em que se encontram 23 bicicletas. Na data do registo a estação tem em funcionamento de 40 vagas para colocação de bicicletas, e uma vez que se encontram 23 bicicletas na estação o rácio de bicicletas na estação é igual a $23/40=0,58$.

As observações do ficheiro **dados2** ocorrem entre dia 25 de janeiro de 2018 às 15h38 e dia 31 de julho de 2018 às 23h39. No ficheiro **dados2** entende-se como registo a anotação das informações respetivas a cada coluna relativamente a uma determinada viagem. O ficheiro **dados2** tem 10 colunas.

- 1) A coluna **id** onde os valores são algarismos distintos em cada linha e representam a identificação de cada registo;
- 2) A coluna **date_start** é representada por uma data e hora. Esta coluna representa o horário em que uma bicicleta partiu de uma determinada estação;
- 3) A coluna **date_end** é representada por uma data e hora. Esta coluna representa o horário em que uma bicicleta chegou a uma determinada estação;
- 4) A coluna **distance** é representada por algarismos e representa a distância percorrida durante cada viagem;
- 5) A coluna **station_start** é representada pelo id de cada estação onde se iniciam as respetivas viagens;
- 6) A coluna **station_end** é representada pelo id de cada estação onde se finalizam as respetivas viagens;
- 7) A coluna **bike_rfid** onde os valores são representados por letras e algarismos. Representa a identificação da bicicleta utilizada em cada viagem;
- 8) A coluna **geom** onde os valores são representados por letras e algarismos;
- 9) A coluna **num_vertices**. Não foi obtida qualquer explicação sobre este campo;
- 10) A coluna **Tipo_Bicicleta** é representada por uma letra que pode assumir os valores de C ou E, onde C representa as bicicletas clássicas e E as bicicletas elétricas;

id	date_start	date_end	distance	station_start	station_end	bike_rfid	geom	num_vertices	Tipo_Bicicleta
424661	25/01/2018 15:40	25/01/2018 15:47	1716.67237595586	446	481	9ddb19d7a	1000734C22	20	E

Tabela 3.2 Exemplo de um registo do ficheiro dados2

Na Tabela 3.2 é ilustrado um exemplo de uma das várias linhas que constituem o ficheiro dados2. Através da linha exemplificada na Tabela 3.2 sabe-se que o registo com id igual a 424661 corresponde ao um registo de uma viagem realizada entre estação de id 446 até à estação de id 481. A distância percorrida no decorrer da viagem cuja informação se encontra na Tabela 3.2 foi igual a aproximadamente 1716 metros. A viagem teve início às 15h40 e terminou às 15h47 do dia 25 de janeiro de 2018. Esta viagem foi realizada com recurso a uma bicicleta elétrica.

3.3. Pré-processamento dos dados

A primeira fase deste projeto consistiu no tratamento de dados. Em estudos que envolvem um tratamento de dados de grande volume gerados a partir de mecanismos informáticos em tempo real, muitas vezes ocorrem pequenos lapsos, provocando assim erros nos registos. O problema dos lapsos nos registos ocorre nos dados que foram fornecidos para este projeto. Muitos dos registos efetuados mostravam em instâncias idênticas na escala dos segundos, um diferente número de bicicletas. Essa situação não corresponde de todo a algo que fosse expectável. Para cada estação, cada instância deve ocorrer 5 minutos depois da instância anterior e 5 minutos antes da instância seguinte.

id	geom	id_expl	id_planeamento	desig_comercial	tipo_servico_niveis	num_bicicletas	num_docas	racio	estado	update_date
158	1000000B9	105	105	105-CC Vasco da Gama	A	26	40	0.65	active	2018-02-23 22:09:16.
211	1000000B9	105	105	105-CC Vasco da Gama	A	24	40	0.6	active	2018-02-23 22:09:16.
255	1000000B9	105	105	105-CC Vasco da Gama	A	24	40	0.6	active	2018-02-23 22:09:16.
299	1000000B9	105	105	105-CC Vasco da Gama	A	24	40	0.6	active	2018-02-23 22:09:16.
343	1000000B9	105	105	105-CC Vasco da Gama	A	24	40	0.6	active	2018-02-23 22:09:16.
387	1000000B9	105	105	105-CC Vasco da Gama	A	24	40	0.6	active	2018-02-23 22:09:16.
431	1000000B9	105	105	105-CC Vasco da Gama	A	24	40	0.6	active	2018-02-23 22:09:16.
475	1000000B9	105	105	105-CC Vasco da Gama	A	24	40	0.6	active	2018-02-23 22:09:16.
519	1000000B9	105	105	105-CC Vasco da Gama	A	22	40	0.55	active	2018-02-23 22:09:16.
563	1000000B9	105	105	105-CC Vasco da Gama	A	21	40	0.53	active	2018-02-23 22:09:16.

Tabela 3.3 Primeiros registos da estação 105 com lapsos

Na Tabela 3.3 está ilustrado um exemplo do lapso acima referido. Como se pode observar, na estação 105 existem várias observações registadas dia 23 de fevereiro de 2018 às 22h09m16 com diferentes valores no que diz respeito ao número de bicicletas.

Para solucionar este tipo problema é possível seguir três procedimentos.

- 1) O primeiro procedimento consiste em desconsiderar os registos repetidos, isto é, eliminar todas as datas duplicadas. Nesse caso, os dados ilustrados na Tabela 3.3 seriam removidos. Esta hipótese representa uma grande perda de informação.

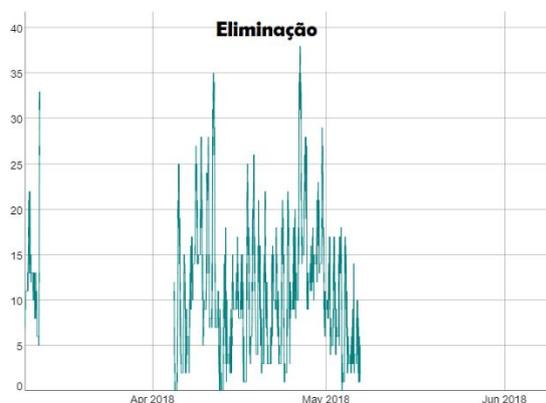


Figura 3.3 Série temporal da estação 105 adotando o primeiro procedimento

A Figura 3.3 ilustra a série temporal do número de bicicletas na estação 105 adotando o primeiro procedimento. Como se pode ver pela Figura 3.3 a informação fica consideravelmente reduzida, algo que interfere com as conclusões retiradas a partir dos dados.

- 2) O segundo procedimento consiste em corrigir as datas. Sempre que são detetados lapsos nos registos, isto é, datas duplicadas, é feita uma alteração na data destes registos, assumindo que a informação corresponde a atualizações de 5 em 5 minutos. Esta hipótese começa por considerar que o primeiro registo ilustrado na Tabela 3.3 se mantém igual. Já o segundo registo sofrerá uma alteração na data e corresponderá a 23 de fevereiro de 2018 às 22h14m16, o terceiro registo corresponderá a 23 de fevereiro de 2018 às 22h19m16, e assim sucessivamente. Esta abordagem permite conservar a informação dada sobre o número de bicicletas. Com isto, as datas foram alteradas, modificando cada segmento de instantes duplicados. Para cada secção de dados duplicados deve-se alterar as datas dos registos de modo a que tenham 5 minutos de diferença entre si, até que não existam mais dados duplicados. Note-se que estas alterações em nada comprometeram os registos que se encontravam corretos, isto é, mesmo alterando as datas repetidas para 5 em 5 minutos, não houve sobreposições sobre datas correctas.

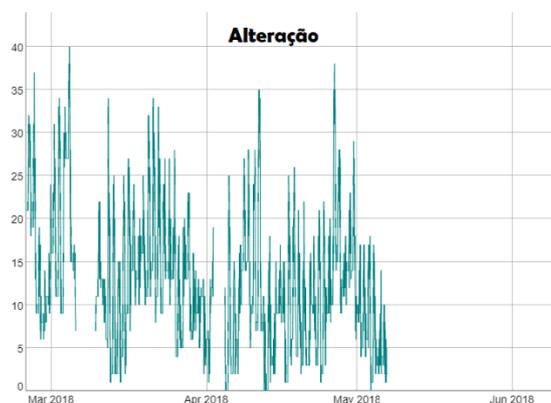


Figura 3.4 Série temporal da estação 105 adotando o segundo procedimento

A Figura 3.4 ilustra série temporal e a disposição dos dados ao se alteram as datas dos instantes duplicados da estação 105.

- 3) O terceiro procedimento consiste na consolidação dos dois ficheiros utilizados. Observando novamente a Tabela 3.3 sabe-se que a data da primeira sequência de lapsos nos registos corresponde à data 22 de fevereiro de 2018 às 22h09, onde o número de bicicletas corresponde a 26 bicicletas. Através dos dados disponibilizados no segundo ficheiro sabe-se quais os instantes em que as bicicletas saem de uma estação e os instantes em que chegam a uma estação.

id	date_start	date_end	distance	station_start	station_end	bike_rfid	geom	num_vertices	Tipo_Bicicleta
447990	22/02/2018 21:17	22/02/2018 21:22	982.55651	105		109 e9db90d8;01050000;		16	E
447999	22/02/2018 21:55	22/02/2018 22:10	1073.9075	105		109 435997b8;01050000;		41	C
447984	22/02/2018 22:07	22/02/2018 22:11	923.57171	105		109 fedb814ee;01050000;		11	E
448025	22/02/2018 22:37	22/02/2018 22:41	1154.8760	105		110 1edba48b;01050000;		10	E

Tabela 3.4 Exemplo do ficheiro dados2 no período em que ocorrem procuras na estação 105 entre as 22h09 e as 22h14 no dia 22/02/2018

Na Tabela 3.4 está contida a informação sobre parte das viagens que tiveram como estação de partida a estação 105 e tiveram início antes e após 23 de fevereiro de 2018 às 22h09m16, instante em que ocorreu o primeiro registo de dados duplicados. Uma vez que a próxima atualização deveria ser realizada 5 minutos após o registo anterior, deveria ocorrer às 22h14. Pela Tabela 3.3 percebe-se que entre as 22h09 e as 22:14 nenhuma bicicleta saiu da estação.

id	date_start	date_end	distance	station_start	station_end	bike_rfid	geom	num_vertices	Tipo_Bicicleta
447816	22/02/2018 22:03	22/02/2018 22:55	8099.6931	430	105	0adba269:01050000;		150	E
447757	22/02/2018 22:12	22/02/2018 22:17	989.77361	102	105	1edba48b:01050000;		14	E
447766	22/02/2018 23:29	22/02/2018 23:38	881.48973	103	105	e759769d:01050000;		26	C

Tabela 3.5 Exemplo do ficheiro dados2 no período em que ocorrem entregas na estação 105 entre as 22h09 e as 22h14 no dia 22/02/2018

Pela Tabela 3.5 entende-se que entre as 22h09 e as 22h14 chegou uma bicicleta à estação 105. Assim, na segunda atualização o número de bicicletas deveria ser $26 + 1 = 27$ bicicletas, contudo como pode observar-se na Figura 3.3 a segunda observação dos dados publicados registou 24 bicicletas. Uma vez que os valores não coincidem, não existem bases suficientemente fortes que permitam dar continuidade a este procedimento.

Como se pode observar pela Figura 3.4, o aspeto gráfico do comportamento do número de bicicletas numa escala temporal parece ser mais realista adotando o segundo procedimento, onde se alteram os dados, uma vez que se registam menos espaços em branco (que representam perda de informação), sendo esta a hipótese escolhida. Em suma, alteraram-se as datas face a cada ocorrência de registos temporais repetidos.

Porém, note-se que, embora a perda de informação seja substancialmente inferior optando pelo segundo procedimento face ao primeiro, continuam a existir períodos sem qualquer tipo de informação no que diz respeito ao número de bicicletas. Estes dados omissos são totalmente desconsiderados em todos os cálculos realizados ao longo do projeto.

Em suma, ao longo deste projeto toda a informação extraída é proveniente de dois ficheiros distintos. Num dos ficheiros encontra-se toda a informação sobre o número de bicicletas e limite de vagas por estação em cada instância. Esta informação é extremamente útil uma vez que permite tomar conhecimento sobre qual o comportamento do número de bicicletas numa escala temporal.

Noutro ficheiro é dada a informação sobre cada viagem entre pares de estações, dando conhecimento da duração e distância percorrida entre cada par de estações. Este ficheiro é igualmente importante pois permite deduzir qual a probabilidade de partindo da estação e ir para a estação f .

Assim sendo, facilmente se percebe que é imprescindível a consolidação da informação proveniente entre os ficheiros mencionados anteriormente. Todavia, as estações tratadas num ficheiro não correspondem na sua totalidade às estações tratadas no segundo ficheiro. É importante que as estações em análise existam em ambos os ficheiros uma vez que no passo seguinte, na Modelação, é necessário ter conhecimento sobre a capacidade de cada estação (informação que se encontra no ficheiro **dados1**), bem como sobre a informação das viagens realizada (informação que se encontra no ficheiro **dados2**). Assim, é necessário desconsiderar as estações que não têm informação em ambos os ficheiros, ficando em análise somente as estações cuja informação se encontra completa.

4. Análise de padrões das estações

O objetivo principal deste projeto é colmatar as ineficiências existentes no atual sistema de bicicletas da empresa Gira, de modo a otimizar os lucros finais. Para tal, uma das soluções é aumentar a taxa de satisfação de serviço face à procura dos utilizadores. Para aumentar a taxa de serviço deve-se aumentar o número de bicicletas no sistema e/ou melhorar a distribuição das bicicletas. De modo a compreender de que forma o serviço poderá otimizar a sua resposta, deve-se primeiramente estudar o comportamento das procuras e do número de bicicletas por estação, para que assim seja possível proceder a uma eventual modelação. Nesta primeira abordagem segue-se o estudo de padrões relativamente ao número de bicicletas numa escala temporal de modo a compreender quais os aspetos que se mostram mais relevantes. Posteriormente, será feita uma análise relativamente aos pares de estações e e f . As principais informações a serem estudadas consistem naturalmente em tomar conhecimento sobre qual a probabilidade de partindo de uma determinada estação e , chegar até outra estação f , bem como a duração e a distância percorrida entre as mesmas. Relativamente ao fator tempo, é também importante analisar o seu comportamento em termos de procura e entrega de bicicletas, uma vez que é uma condicionante extremamente significativa na distribuição temporal do número de bicicletas em cada estação, visto que tem uma influência direta na taxa de serviço, e por outro lado, na disponibilidade de entrega atendendo ao número de vagas disponíveis.

4.1. Análise de padrões para cada estação

É feita uma análise no que diz respeito ao número de bicicletas numa escala de tempo contínuo, numa escala composta em blocos de horas ao longo do dia, numa escala diária e numa escala mensal. Por fim, são abordados alguns aspetos externos que poderão ter influencia sobre a procura em cada estação.

4.1.1. Representação em série temporal

Neste capítulo realiza-se uma análise relativamente ao comportamento do número de bicicletas ao longo do tempo. Ao analisar o comportamento das bicicletas numa escala de tempo continua é possível identificar mais facilmente padrões temporais que permitam outro tipo de agrupamento dos dados. Entende-se como padrões temporais os blocos de horas de um dia, dias da semana e meses. Esta análise é extremamente importante uma vez que permite fazer uma previsão do que se sucede nas secções seguintes.

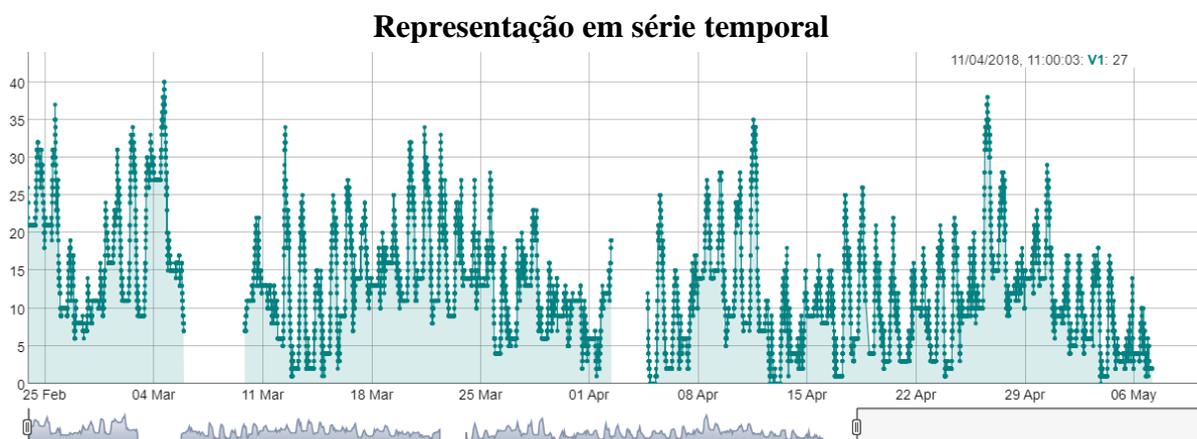


Figura 4.1 Série Temporal das observações registadas do número de bicicletas no decorrer do ano 2018 da estação 105-CC Vasco da Gama

No gráfico ilustrado na Figura 4.1 o eixo vertical representa o número de bicicletas na estação 105-CC Vasco da Gama e o eixo horizontal representa cada instante de tempo dos registos realizados. Assim, cada ponto do gráfico corresponde ao número de bicicletas num determinado instante na estação 105. Os números das bicicletas na estação 105 sofre várias oscilações, tendo alternadamente um comportamento crescente e decrescente ao longo do tempo. Note-se que o número máximo de bicicletas na estação 105 ao longo do tempo corresponde a 40 bicicletas, que por sua vez coincide com a capacidade da estação 105.

É perceptível na Figura 4.1 que de facto existem pequenas perdas de informação que se mostraram impossíveis de contornar, todavia pode-se observar que o comportamento da série temporal não denota um comportamento constante. A barra a cinzento localizada na parte inferior da Figura 4.1 limita as observações do número de bicicletas para um certo intervalo. Quando se altera a escala temporal de todas as observações registadas para um intervalo menor, o aspeto do gráfico da Figura 4.1 altera-se em conformidade, podendo a série ficar com um aspeto mais monótono. Por exemplo, reduzindo o intervalo de tempo em estudo percebe-se que dias diferentes assumem um comportamento consideravelmente semelhante.

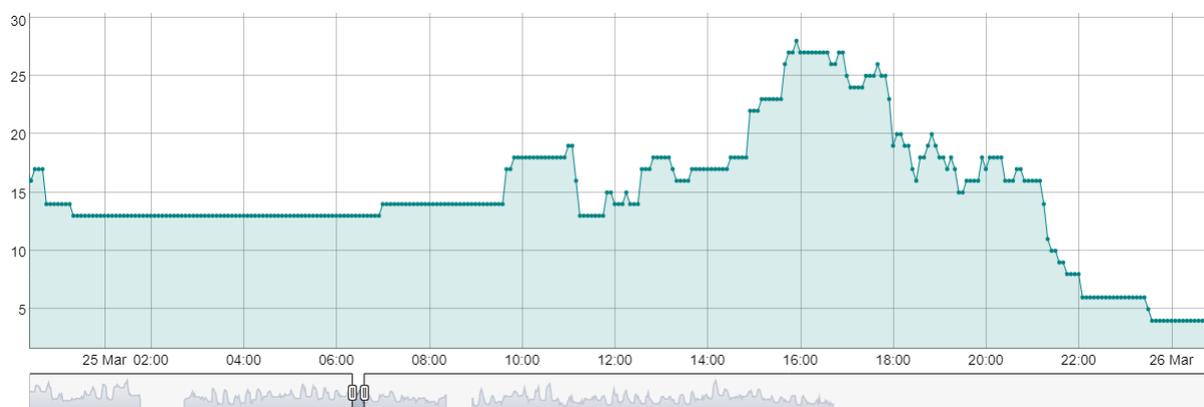


Figura 4.2 Série temporal da estação 105 no decorrer do dia 25 de março de 2018

Observando a Figura 4.2 facilmente se percebe o que foi dito anteriormente. Nessa Figura 4.2 a barra cinzenta reduziu o intervalo das observações, mudando o aspeto gráfico no que diz respeito ao número de bicicletas numa escala temporal. Ao observar o número de bicicletas no decorrer do dia 25 de março de 2018 percebe-se que até aproximadamente às 10 horas, o número de bicicletas mantém-se consideravelmente constante. O número de bicicletas começa a aumentar com ligeiras oscilações entre as 10 horas e as 17 horas, sofrendo posteriormente uma diminuição também com ligeiras oscilações entre as 17 horas e as 24 horas.

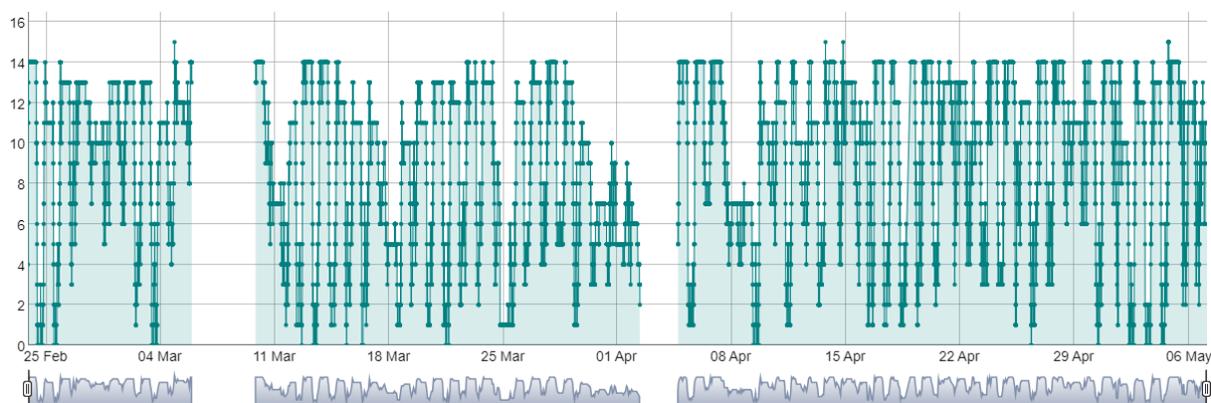


Figura 4.3 Série Temporal das observações registadas do número de bicicletas no decorrer do ano 2018 da estação 109

A Figura 4.3 ilustra a série temporal do número de bicicletas na estação 109. Comparando a Figura 4.1 com a Figura 4.3 facilmente se percebe que o comportamento do número das bicicletas ao longo do tempo varia consoante a estação em estudo. Ao contrário da estação 105, o número de bicicletas na estação 109 parece ter um comportamento mais homogéneo. Comparando as séries temporais da estação 105 e 109 respetivamente, entende-se que conhecendo o presente é mais fácil prever o futuro do número de bicicletas na estação 109 do que na estação 105.

Media	DesvioPadrao	Maximo	Minimo
12.41	7.39	40	0

Tabela 4.1 Estatísticas descritivas relativamente às observações registadas no decorrer do ano 2018 da estação 105-CC Vasco da Gama

A Tabela 4.1 fornece algum tipo de informação relevante. Por exemplo, visualizando a Tabela 4.1 sabe-se que no decorrer do ano de 2018, na estação 105-CC Vasco da Gama, o número máximo de bicicletas presentes na respetiva estação foi igual a 40 bicicletas. Por outro lado, a estação chegou a estar completamente vazia, sem qualquer bicicleta disponível em caso de procura. É ainda dada a informação de que o número médio de bicicletas disponíveis na doca no decorrer do tempo mencionado foi igual a aproximadamente 12,41 bicicletas. Como o desvio-padrão registado é consideravelmente elevado face aos valores em causa, a média de bicicletas varia em torno de 5,02 a 19,8 bicicletas disponíveis na respetiva doca.

4.1.2. Análise de padrões temporais dos dados

Neste capítulo realiza-se uma análise relativamente ao comportamento do número de bicicletas em blocos de uma hora ao longo do dia, em cada dia da semana, e em cada mês, no que diz respeito a uma determinada estação.

Seja y_h o número médio de bicicletas no intervalo entre as $[h, h+1[$ horas. A estimativa para y_h é calculada somando o número de bicicletas de todas as observações que se realizaram entre as $[h, h+1[$ horas ao longo de todos os dias, dividindo esta soma pelo número total de observações nesta estação.

Existem ao todo 24 blocos de 1h, sendo o primeiro bloco o intervalo entre as 0h e as 0h e 59 minutos, o segundo bloco o intervalo entre a 1h e a 1h e 59 minutos, assim adiante, até ao último bloco que corresponde ao intervalo entre as 23h e as 23h e 59 minutos.

Por exemplo, para calcular o número médio de bicicletas no intervalo entre as 13 horas e as 13:59, isto é, y_{13} têm de se ter conhecimento do total de observações e da soma do número de bicicletas registadas neste intervalo de tempo. Uma vez que se registaram 779 observações na estação 105 entre as 13 horas e as 13h59, e a soma do número de bicicletas neste intervalo é 13653 então $y_{13}=13653/779=17,53$.

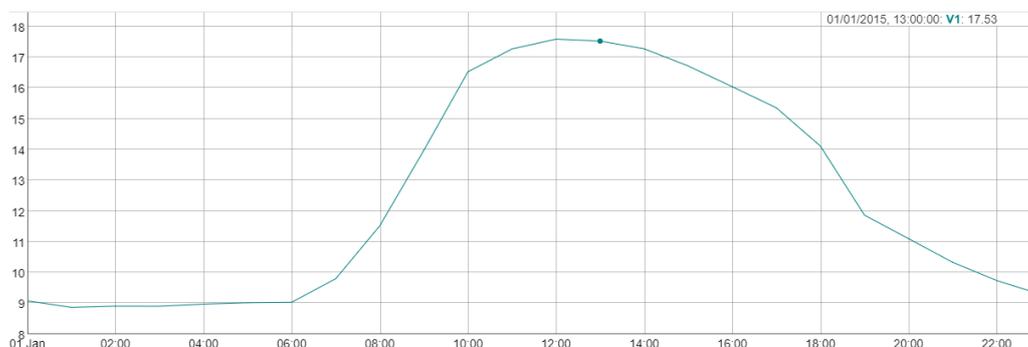


Figura 4.4 Número médio das bicicletas referentes à estação 105-CC Vasco da Gama, no decorrer do ano de 2018, dispostas em blocos de horas

A Figura 4.4 ilustra a representação gráfica do número médio de bicicletas por cada bloco de horas. O eixo horizontal é representado pelas horas enquanto que o eixo vertical representa o número médio de bicicletas em cada bloco de horas. Por exemplo, sabe-se entre as 13h e as 13h e 59 minutos o número médio de bicicletas é aproximadamente igual a 17 bicicletas na estação 105.

Observando a Figura 4.4, tal como foi mencionado anteriormente que seria esperado, já é perceptível a identificação de um comportamento mais linearizado. Percebe-se que no caso particular da estação 105 no período entre a meia noite e as 6h da manhã não existe grande afluência no número médio de bicicletas, mantendo-se constante. Já entre as 6h da manhã até perto do meio dia, o número médio de bicicletas tende a crescer exponencialmente, acabando por decrescer de forma menos acentuada entre o meio dia e a meia noite.

As conclusões que se retiram deste gráfico são bastante importantes para a etapa que será abordada mais à frente, a etapa da modelação.

Seja y_d o número médio de bicicletas no dia da semana d . A estimativa para y_d é calculada somando o número de bicicletas de todas as observações da respetiva estação que se realizaram no dia da semana d , dividindo esta soma pelo número total de observações registadas na respetiva estação em todos os dias da semana.

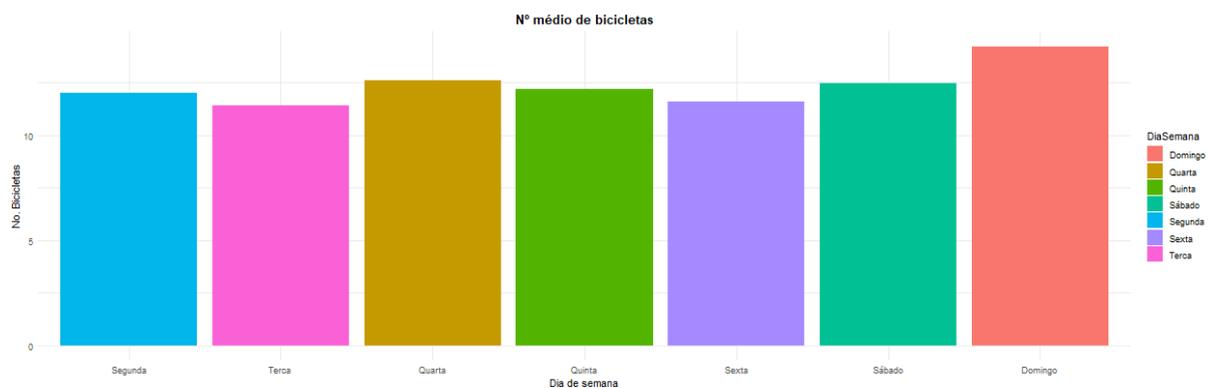


Figura 4.5 Número médio de bicicletas referentes ao ano 2018 na estação 105-CC Vasco da Gama, dispostos em dias da semana

O eixo vertical do gráfico de barras ilustrado na Figura 4.5 representa o número de bicicletas enquanto que o eixo horizontal representa os dias da semana. Quando os dados são tratados numa escala temporal retratada em dias da semana, o número médio de bicicletas não varia consideravelmente, pelo menos no que diz respeito à estação em questão. Ainda assim, verifica-se que nesta estação existe um registo de número de bicicletas maior ao domingo e por outro lado, menor à terça-feira, sendo razoavelmente constante nos restantes dias da semana. Saliente-se que o maior número médio de bicicletas ao domingo pode justificar-se por ser um dia de fim-de-semana. Uma vez que muitas bicicletas são usadas não só para lazer, mas também como recurso para eventuais deslocações dos usuários até ao seu local de trabalho, em dias da semana fora do contexto laboral o uso deste serviço poderá diminuir e consequentemente ficarem mais bicicletas alojadas nas estações. Sendo domingo um dia da semana onde a maioria dos portugueses se encontra fora do contexto laboral, esta diferença acaba por ser expectável.

Seja y_m o número médio de bicicletas no mês m . A estimativa para y_m é calculada somando o número de bicicletas de todas as observações da respetiva estação que se realizaram no mês m , dividindo esta soma pelo número total de observações desta estação realizadas em todos os meses.

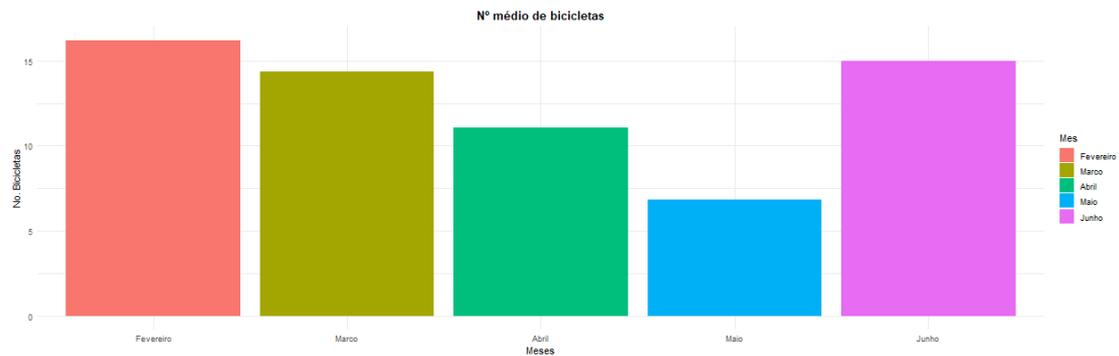


Figura 4.6 Número médio de bicicletas referentes ao ano 2018 na estação 105-CC Vasco da Gama, dispostos em meses

O eixo vertical do gráfico de barras ilustrado na Figura 4.6 representa o número de bicicletas enquanto que o eixo horizontal representa os meses em estudo. Na Figura 4.6 observa-se que no decorrer do mês de fevereiro até ao mês de maio, a variação entre o número médio de bicicletas decresce substancialmente, sofrendo um acentuado aumento em junho. Mais uma vez, o valor superior do número médio bicicletas no mês de junho pode justificar-se pelo recurso das bicicletas como um meio de deslocação até à área de trabalho dos usuários deste serviço. Uma vez que junho é um mês onde muitas pessoas iniciam o seu período de férias em Portugal, existe um menor uso deste serviço, nesta estação em particular, convergindo assim para um número médio de bicicletas superior aos meses de época laboral. Por outro lado, o decréscimo do número médio de bicicletas entre o mês de fevereiro e maio pode refletir um maior recurso das bicicletas disponíveis nesta estação, algo que pode justificar-se por eventuais fatores externos, como por exemplo, o término do inverno, uma vez que o recurso de bicicletas como meio de transporte torna-se mais apelativo quando as condições atmosféricas se tornam mais sugestivas. Por outro lado, tratando-se a estação 105-CC Vasco da Gama de uma estação em área turística, pode-se considerar que um maior número de usuários tem como destino esta estação, acabando tal fenómeno por se refletir num maior número de bicicletas num mês onde existe mais disponibilidade para o lazer.

4.1.3. Localização das estações e fatores externos

Embora não seja algo que esteja ao alcance deste estudo, é importante de referir e salientar que a localização das estações é um fator importante no estudo das procuras, uma vez que a localização das mesmas irá influenciar uma maior ou menor aderência ao recurso do serviço de *bikesharing*.

Por outro lado, dada a natureza do transporte em questão, as condições atmosféricas são um fator determinístico das procuras, uma vez que é expectável que em dias chuvosos ou com vento mais forte, a aderência a este serviço reduza.

Registam-se ainda outras externalidades, como por exemplo a presença de obras que afetem determinados trajetos entre estações, como eventuais manutenções das mesmas, entre outros.

4.2. Análise de dados nas viagens

Neste capítulo realiza-se uma análise sobre a duração entre procuras, entre entregas e são retiradas conclusões comparando as estatísticas estimadas respetivas a estes dois fatores. É ainda feita uma análise à duração das viagens entre cada par de estações e à distância percorrida no decorrer das mesmas.

4.2.1. Duração

Como já foi várias vezes referido, o estudo dos intervalos de tempo é crucial para a modelação, uma vez que é um aspeto totalmente relacionado com a afluência das bicicletas entre estações. Assim, será feita

uma análise que se decompõe entre a duração da procura numa determinada estação, a duração da entrega de bicicletas numa determinada estação, bem como a duração entre viagens, partindo de uma determinada estação.

4.2.1.1. Duração entre procuras

Seja a_j o instante em que é feita a recolha de uma bicicleta na j -ésima viagem de uma determinada estação e , onde $j = 1, \dots, w$, sendo w o número total de viagens realizadas. O cálculo da duração entre procuras da estação e dá origem a um vetor. Cada elemento do vetor dos intervalos entre procuras p_j , representa o j -ésimo intervalo entre procuras, onde $j=1, \dots, (w-1)$. A expressão da duração do k -ésimo intervalo entre procuras p_k é a seguinte:

$$p_k = a_{k+1} - a_k, k = 1, \dots, (w - 1)$$

Equação 4.1 Cálculo dos intervalos entre procuras

Para realizar o estudo da duração dos intervalos entre procuras numa determinada estação mais uma vez teve de ser realizada uma importante manipulação dos dados. Note-se que as alterações realizadas nos dados referidas no capítulo de Pré-processamento dos dados fez com que os dados contornassem consideravelmente o problema anteriormente existente quanto aos pontos temporais sem informação relativamente ao número de bicicletas, mas não o resolveram totalmente. Esta condicionante traz consigo um problema no estudo da duração dos intervalos entre procuras.

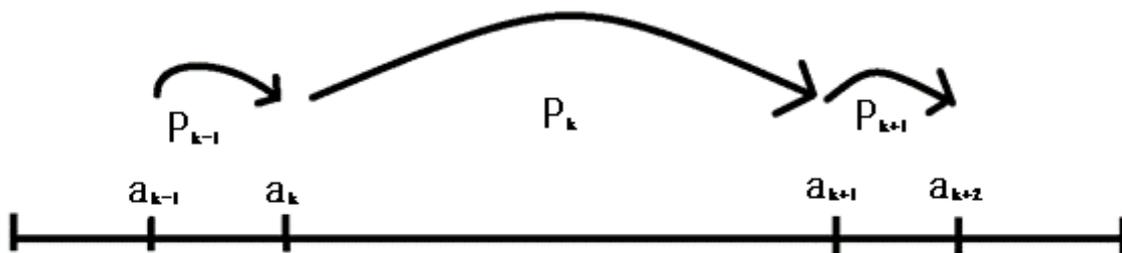


Figura 4.7 Representação do problema existente no cálculo dos intervalos entre procuras

Na Figura 4.7 o eixo horizontal representa os instantes k em que houve uma procura numa determinada estação. Todavia, existem espaços no tempo sem qualquer tipo de informação registada. Considere-se que a duração entre o instante a_k e a_{k+1} , isto é, o intervalo entre procuras p_k , é superior a 24h, então considera-se que este intervalo de tempo corresponde a uma ineficiência de informação e com isto elimina-se o intervalo entre procuras p_k .

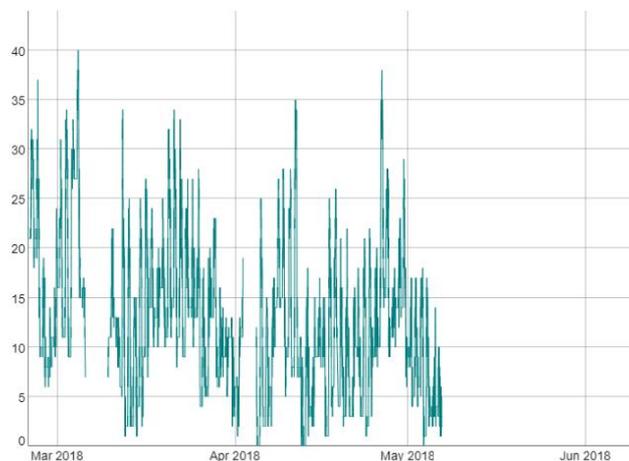


Figura 4.8 Série Temporal do número de bicicletas na estação 105-CC Vasco da Gama após as devidas alterações mencionadas no Capítulo 2

O eixo vertical do gráfico ilustrado na Figura 4.8 representa o número de bicicletas enquanto que o eixo horizontal representa o tempo decorrido. Como se pode perceber na Figura 4.8 mesmo com as alterações realizadas, continuam a existir espaços temporais sem qualquer tipo de informação. Por exemplo, entre as 20h49m16 horas de 5 de março de 2018 e as 20h45m02 horas de 9 de março de 2018 não existe qualquer tipo de informação, com isto, o cálculo da duração da procura entre estes dois instantes deve ser ignorado, uma vez que se traduz num erro do sistema, pois iria considerar-se que a procura entre bicicletas durou aproximadamente 4 dias, quando tal ocorrência dificilmente se traduz na realidade. Posto isto, e tendo em consideração a eliminação da informação obtida pela diferença entre espaços temporais sem informação, obteve-se a Tabela 4.2 respetiva à estação 105-CC Vasco da Gama.

minimo	maximo	mediana	media
0.00	1061.11	7.98	22.49

Tabela 4.2 Tabela Descritiva da duração dos intervalos entre procuras de bicicletas na estação 105-CC Vasco da Gama

Pela Tabela 4.2 sabe-se que na estação 105-CC Vasco da Gama o tempo mínimo dos intervalos entre procuras é de 0 minutos, o tempo máximo dos intervalos entre procuras é de 1061,11 minutos que se traduz em aproximadamente 18h. Cerca de 50% das observações registadas de intervalos entre procuras tiveram uma duração de aproximadamente 8 minutos, e a média da duração dos intervalos entre procuras é aproximadamente igual a 22 minutos e 30 segundos.

4.2.1.2. Duração entre entregas

Seja b_j o instante em que é feita a entrega de uma bicicleta na j -ésima viagem de uma determinada estação f , onde $j = 1, \dots, w$, sendo m o número total de viagens realizadas. O cálculo da duração entre entregas da estação h dá origem a um vetor. Cada elemento do vetor dos intervalos entre entregas g_j , representa o j -ésimo intervalo entre entregas, onde $j=1, \dots, (w-1)$. A expressão do k -ésimo intervalo entre entregas g_k é a seguinte:

$$g_k = b_{k+1} - b_k$$

Equação 4.2 Cálculo dos intervalos entre procuras

Na duração dos intervalos entre procuras foi referido como o lapso de informação em certos instantes de tempo pode ser um problema para o estudo da duração entre procuras. Do mesmo modo, tal ocorrência também se pode traduzir num problema no estudo da duração dos intervalos entre entregas, contudo, tal equívoco é solucionado de maneira semelhante ao caso anterior, isto é, não são considerados os intervalos que resultam da diferença entre instantes de tempo sem qualquer tipo de informação no que diz respeito ao número de bicicletas presentes em determinada questão.

minimo	maximo	mediana	media
0.00	1061.63	7.98	22.49

Tabela 4.3 Tabela Descritiva da duração dos intervalos entre entregas de bicicletas na estação 105-CC Vasco da Gama

Pela Tabela 4.3 sabe-se respetivamente à estação 105-CC Vasco da Gama que o tempo mínimo de intervalos entre entregas é de 0 minutos, o tempo máximo de intervalos entre entregas é de 1061,63 minutos que se traduz em aproximadamente 18h. Cerca de 50% das observações registadas de intervalos entre entregas tiveram uma duração de aproximadamente 8 minutos, e a média da duração dos intervalos entre entregas é aproximadamente igual a 22 minutos e 30 segundos.

4.2.1.3. Comparação entre a duração do tempo de procura e o tempo de entrega

É interessante comparar os resultados ilustrados na Tabela 4.4 e na Tabela 4.5 dos intervalos entre procura e entrega uma vez que tal permite antecipar conclusões interessantes.

minimo	maximo	mediana	media
0.00	1061.11	7.98	22.49

Tabela 4.4 Tabela Descritiva da duração dos intervalos entre procuras na estação 105- CC Vasco da Gama

minimo	maximo	mediana	media
0.00	1061.11	7.98	22.49

Tabela 4.5 Tabela Descritiva da duração dos intervalos entre entregas na estação 105- CC Vasco da Gama

Por exemplo, analisando as tabelas ilustradas na Tabela 4.4 e Tabela 4.5 conclui-se imediatamente que ambas são bastante semelhantes, ou seja, a duração média dos intervalos com que se procura uma bicicleta é aproximadamente igual à duração média dos intervalos com que se entrega uma bicicleta, o que leva antecipadamente a supor que o número médio de bicicletas na estação 105-CC Vasco da Gama é razoavelmente constante.

minimo	maximo	mediana	media
0.00	1352.58	19.87	55.85

Tabela 4.6 Tabela Descritiva da duração dos Intervalos entre procuras na estação 101-Alameda dos Oceanos-Rua dos Argonautas

minimo	maximo	mediana	media
0.00	4505.31	7.98	23.55

Tabela 4.7 Tabela Descritiva da duração dos intervalos entre entregas na estação 101-Alameda dos Oceanos-Rua dos Argonautas

Por outro lado, pelo exemplo das tabelas ilustradas na Tabela 4.6 e Tabela 4.7 entende-se que ao contrário do que acontece na Tabela 4.4 e na Tabela 4.5 referentes à estação 105, os valores obtidos no que diz respeito às durações dos intervalos da procura e entrega de bicicletas na estação 101, quando comparados são substancialmente diferentes. A duração média dos intervalos de procura na estação 101-Alameda dos Oceanos-Rua dos Argonautas é superior à duração média dos intervalos de entrega na mesma estação. Neste caso é difícil retirar conclusões, pois se por um lado pode ser feita a suposição de que a satisfação da procura será substancialmente superior a estações onde a duração média dos intervalos entre procuras são inferiores à duração média dos intervalos entre entregas, e ainda supor que a frequência com que a estação 101 fica lotada e conseqüentemente sem vagas para devolução de bicicletas é superior em relação a estações onde a duração média dos intervalos entre procuras é inferior à duração média dos intervalos entre entregas. Por outro lado, pode considerar-se que a duração dos intervalos entre procuras poderá ser superior uma vez que pode não ter nenhuma bicicleta disponível em vários instantes de possível procura, algo que diminui efetivamente a taxa de satisfação da procura.

4.2.1.4. Duração das viagens entre pares de estações

Seja a_j o instante em que é feita a recolha de uma bicicleta na j -ésima viagem de uma determinada estação e e b_j o instante em que é feita a entrega de uma bicicleta na j -ésima viagem de uma determinada estação f , onde $j = 1, \dots, m$, sendo m o número total de viagens realizadas entre a estação e e a estação f . O cálculo da duração das viagens realizadas entre a estação e e a estação f dá origem a um vetor. Cada

elemento do vetor de durações de viagens entre e e f , designa-se por v_j , e representa a duração de uma viagem entre estas duas estações, onde $j = 1, \dots, m$. A expressão da duração da viagem k , v_k é a seguinte:

$$v_k = b_k - a_k$$

Equação 4.3 Cálculo da duração das viagens entre cada par de estações

No estudo da duração das viagens, atempadamente se percebe que existe uma relação proporcional direta entre a duração, e a distância uma vez que se a distância a percorrer entre duas estações for elevada comparativamente com as restantes, a sua duração dificilmente será pequena. Porém, existem mais uma vez fatores externos que poderão ter influência neste padrão, tais como o propósito da viagem, uma vez que uma viagem por lazer provavelmente terá uma duração superior a uma viagem com um objetivo concreto. Outro fator retrata-se no percurso consistir essencialmente numa subida, onde provavelmente a duração será maior, do que numa descida e vice-versa.

Estacao	minimo	maximo	mediana	media	chart
224	31.00	43.00	39.50	38.17	
225	41.00	47.00	43.00	44.00	
304	34.00	142.00	46.50	64.44	
305	31.00	98.00	42.50	53.29	
306	33.00	38.00	38.00	36.33	
307	35.00	149.00	39.00	67.71	
308	35.00	42.00	36.00	37.29	
309	34.00	42.00	41.00	39.00	
403	26.00	74.00	35.00	36.84	
406	28.00	108.00	33.00	49.20	
407	24.00	44.00	29.00	31.54	
408	23.00	95.00	32.00	38.20	

Tabela 4.8 Parte 1 da tabela descritiva da duração média das viagens tendo como ponto de partida a estação 105-CC Vasco da Gama

A Tabela 4.8 ilustra informação sobre a duração mínima e máxima relativamente às viagens realizadas entre a estação 105-CC Vasco da Gama e algumas outras estações ilustradas na coluna **Estacao** numa escala de minutos. Assim, sabe-se por exemplo que, a duração mínima de uma viagem entre a estação 105 e a estação 224 foi igual a 31 minutos, por outro lado a duração máxima de uma viagem entre a estação 105 e 224 foi igual a 43 minutos. É ainda dada a informação de que a mediana da duração entre as duas estações referidas anteriormente é 38,17, o que significa que pelo menos 50% das viagens tiveram uma duração igual ou inferior a 38 minutos e aproximadamente 10 segundos, ou por outro lado, 50% das viagens teve duração igual ou superior a 38 minutos e aproximadamente 10 segundos.

Estacao	minimo	maximo	mediana	media	chart
101	2.00	164.00	6.00	8.62	
102	2.00	140.00	5.00	8.64	
103	2.00	130.00	4.00	8.70	
104	1.00	160.00	21.00	22.82	
105	0.00	379.00	21.00	26.15	
106	0.00	704.00	4.00	14.42	
107	1.00	215.00	3.00	6.42	
108	2.00	149.00	4.00	10.30	
109	2.00	308.00	5.00	7.09	
110	3.00	643.00	5.00	7.97	
114	6.00	231.00	12.50	19.56	
115	8.00	117.00	14.00	19.16	

Tabela 4.9 Parte 2 da tabela descritiva da duração média das viagens tendo como ponto de partida a estação 105-CC Vasco da Gama

Na Tabela 4.9 é fornecida a mesma informação que na Tabela 4.8, todavia referente a um grupo de estações diferentes ainda que a estação de origem se mantenha igual. É interessante visualizar estes resultados, pois atendendo aos valores mínimos das durações atempadamente se compreende que todas as estações da coluna **Estacao** da Tabela 4.9 se encontram próximas da estação 105, pois de outro modo seria irrealista uma duração tão curta entre estes pares de viagem.

minimo	maximo	mediana	media
0.02	704.98	5.70	12.08

Tabela 4.10 Tabela Descritiva da duração entre viagens com a estação 105-CC Vasco da Gama como origem

A Tabela 4.10 contém informação respetiva às viagens realizadas a partir da estação 105-CC Vasco da Gama. Através da Tabela 4.10 pode concluir-se que a duração entre viagens partindo da estação 105 teve como duração mínima aproximadamente 0 minutos, algo que facilmente se percebe que pode e deve derivar de um colapso de um usuário. Por outro lado, a duração máxima entre viagens partindo da estação 105 é de aproximadamente 12 horas. Sabe-se ainda que cerca de 50% das viagens realizadas a partir da estação 105 tiveram uma duração de aproximadamente 6 minutos e que a média da duração é de aproximadamente 12 minutos.

4.2.2. Distância das viagens

Tal como foi referido no capítulo da Duração, existe uma relação proporcional direta entre a distância percorrida e a duração das viagens relativamente a cada par de estações. Assim sendo, quando a distância é maior, naturalmente a duração da viagem será maior, e analogamente, quando a distância for menor, a duração da viagem será menor.

Estacao	minimo	maximo	mediana	media	chart
106	0.00	704.00	4.00	14.42	
107	1.00	215.00	3.00	6.42	
108	2.00	149.00	4.00	10.30	
109	2.00	308.00	5.00	7.09	
110	3.00	643.00	5.00	7.97	
114	6.00	231.00	12.50	19.56	
115	8.00	117.00	14.00	19.16	
212	21.00	127.00	34.00	37.66	
219	27.00	86.00	42.50	44.59	
224	31.00	43.00	39.50	38.17	
225	41.00	47.00	43.00	44.00	
304	34.00	142.00	46.50	64.44	

Tabela 4.11 Parte I da tabela descritiva da duração média das viagens tendo como ponto de partida a estação 105-CC Vasco da Gama

Estacao	minimo	maximo	mediana	media	chart
106	0.00	6750.00	3.81	531.27	
107	0.00	7758.70	29.95	538.46	
108	0.09	10753.22	27.42	707.76	
109	0.00	19143.32	28.21	585.68	
110	0.00	15499.55	42.73	594.74	
114	0.26	4005.75	3.39	382.41	
115	0.52	8758.79	11.27	746.46	
212	1.30	9598.65	10.68	1160.43	
219	5.77	6893.30	1252.58	2018.90	
224	3.61	6961.51	238.02	1519.10	
225	2.11	5243.53	138.59	1399.17	
304	285.69	54479.97	5398.25	10159.64	

Tabela 4.12 Parte 1 da tabela descritiva da distância das viagens tendo como ponto de partida a estação 105-CC Vasco da Gama

Estacao	minimo	maximo	mediana	media	chart
446	25.00	49.00	29.50	34.21	
449	26.00	37.00	36.00	33.43	
450	30.00	34.00	32.00	32.00	
452	17.00	90.00	23.00	32.00	
453	21.00	70.00	30.00	32.26	
457	21.00	43.00	31.00	31.42	
459	18.00	44.00	24.50	26.83	
460	19.00	104.00	21.00	29.50	
463	23.00	32.00	27.00	27.25	
464	22.00	41.00	30.50	30.75	
468	19.00	63.00	26.50	30.59	
471	18.00	57.00	36.00	38.14	

Tabela 4.13 Parte 2 da tabela descritiva da duração média das viagens tendo como ponto de partida a estação 105-CC Vasco da Gama

Estacao	minimo	maximo	mediana	media	chart
446	155.19	13587.33	2064.44	3160.22	
449	6.17	2008.26	186.22	550.95	
450	36.09	7306.46	3671.28	3671.28	
452	1.29	6713.71	567.68	1655.15	
453	2.55	7109.10	1255.60	2313.10	
457	8.90	6152.74	1539.12	1536.18	
459	1.30	8080.29	80.99	1739.97	
460	2.77	4254.75	1751.17	1636.22	
463	0.00	3330.22	71.22	868.16	
464	78.81	6309.82	2264.80	2737.15	
468	2.92	4994.94	580.62	1562.55	
471	2.72	3604.18	141.79	896.35	

Tabela 4.14 Parte 2 da tabela descritiva da distância das viagens tendo como ponto de partida a estação 105-CC Vasco da Gama

Note-se que os dados apresentados na Tabela 4.14 dizem respeito a viagens que tomaram início na estação 105.

Ao comparar-se a Tabela 4.11 e Tabela 4.13 respetivamente com a Tabela 4.12 e a Tabela 4.14 percebe-se facilmente a relação proporcional referida anteriormente. Através da Tabela 4.11, Tabela 4.12, Tabela 4.13 e Tabela 4.14 é possível ainda perceber quais as estações que se encontram mais próximas da estação 105, como por exemplo, a estação 106. Por outro lado, pode também tomar-se conhecimento sobre qual as estações mais distantes da estação 105, como por exemplo a estação 450. É ainda interessante perceber que ordenando pela média a tabela descritiva das durações das viagens e a tabela descritiva das distâncias, a ordenação das estações fica muito semelhante quando se comparam as duas tabelas, algo que vem reforçar a relação direta entre a duração e distância.

4.2.3. Número de viagens

Para a elaboração da modelação e seguinte simulação, é importante ter conhecimento sobre qual a probabilidade de ter uma determinada estação como destino, sabendo a estação de origem. Para tal, é essencial saber o número de viagens realizadas entre pares de estações.

estacao	nviagens	freqviagens
419	2	0.02
430	2	0.02
450	2	0.02
414	3	0.04
428	4	0.05
463	4	0.05
426	6	0.07
305	7	0.09
412	7	0.09
423	7	0.09
442	7	0.09

Tabela 4.15 Parte 1 da Tabela do número de viagens para cada estação, tendo como ponto de partida a estação 105-CC Vasco da Gama

A Tabela 4.15 está ordenada de forma crescente pela coluna **nviagens**, querendo isto dizer que as estações ilustradas são as estações com o menor número de viagens realizadas partindo da estação 105. Assim, pela partição da Tabela 4.15 sabe-se que a frequência de viagens partindo da estação 105 para as estações ilustradas na coluna **estacao** da tabela são inferiores a 0,1%. Assim sendo, deduz-se que, por exemplo, a probabilidade de deslocação partindo da estação 105 para a estação 419 é muito diminuta.

estacao	nviagens	freqviagens
481	97	1.18
108	138	1.68
105	184	2.24
106	230	2.80
104	486	5.91
107	588	7.15
110	747	9.09
103	1062	12.92
109	1092	13.28
102	1269	15.44
101	1318	16.03

Tabela 4.16 Parte 2 da Tabela do número de viagens para cada estação, tendo como ponto de partida a estação 105-CC Vasco da Gama

Por outro lado, pela Tabela 4.16 têm-se conhecimento que partindo da estação 105, as 11 estações com um maior número de viagens são as ilustradas na coluna **estacao**. Por exemplo, partindo da estação 105-CC Vasco da Gama, a estação com maior probabilidade de seguir como destino é a estação 101, onde a frequência de viagens é igual a 16,03%. Posteriormente, na simulação a escolha das estações de chegada será gerada aleatoriamente segundo uma distribuição categórica recorrendo à informação ilustrada na Tabela 4.16.

5. Análise da procura

Após a análise dos padrões e interpretação dos diversos resultados, estão reunidas as condições para realizar uma modelação do sistema, particularmente necessária para uma eventual futura simulação.

Para simular o sistema de *bikesharing* é necessário simular as durações entre procuras, bem como as durações entre viagens. É imprescindível ter conhecimento sobre qual a distribuição destes fatores. Posto isto, está-se em condições de introduzir a modelação.

5.1. Modelação dos intervalos entre procuras

Comece-se por compreender a distribuição dos intervalos entre procuras de bicicletas por estação. Para tal, será realizado um estudo e eventuais testes para uma avaliação consistente das distribuições propostas.

A distribuição exponencial caracteriza-se por ter uma função de taxa de falha constante, sendo a única com esta propriedade. Usualmente, é usada como um modelo para o tempo de vida de certos produtos e materiais.

É de conhecimento estatístico que a distribuição exponencial é bastante aconselhável no que diz respeito à modelação de intervalos de tempo, em particular.

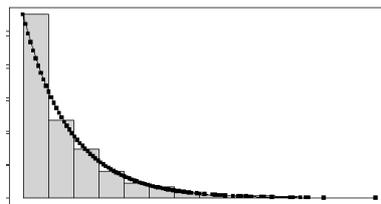


Figura 5.1 Histograma dos intervalos entre procuras

Observando o histograma da Figura 5.1 a ideia de que os intervalos entre procuras poderão seguir uma distribuição exponencial é fortalecida uma vez que o seu comportamento das observações é bastante semelhante ao da exponencial.

Sendo X uma variável aleatória com distribuição exponencial e parâmetro λ , $\lambda > 0$, então X tem uma função densidade $f(x)$ tal que:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Equação 5.1 Função densidade de uma variável aleatória X com distribuição Exponencial

onde λ é o parâmetro da taxa de distribuição e deve satisfazer a condição $\lambda > 0$. No caso particular das bicicletas, pode depreender-se que λ é o tempo médio de intervalos de procura e x é o tempo de falha.

A função de distribuição acumulada $F(x)$ é dada por:

$$F(x) = \int_0^x f(s) ds = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Equação 5.2 Função densidade de uma variável aleatória X com distribuição Exponencial

Note-se que λ é um parâmetro desconhecido, tendo de ser estimado. O estimador de máxima verosimilhança de λ é dado pela seguinte expressão:

$$\hat{\lambda} = \frac{1}{\bar{x}}, \bar{x} = \frac{\sum_{i=1}^n x_i}{n}, n = \text{dimensão da amostra}$$

Equação 5.3 Equação do estimador de máxima verosimilhança do parâmetro da distribuição exponencial

A exponencial tem a particularidade do valor de λ ser igual ao valor esperado de x , isto é, $\lambda = E[x]$. Este é um bom indicador para compreender se os dados seguem ou não uma distribuição exponencial.

Observe-se agora a distribuição de *Weibull*. A distribuição de *Weibull* foi proposta em estudos relacionados ao tempo de falha devido a fadiga de metais. É uma distribuição deveras interessante devido à sua adaptação a diversas formas.

Seja uma variável aleatória X com distribuição de *Weibull*. A função densidade de x é dada por:

$$f(x) = \begin{cases} \frac{\alpha}{\beta} x^{\alpha-1} \exp\left[-\left(\frac{x}{\beta}\right)^\alpha\right], & \text{se } x \geq 0 \\ 0, & \text{se } x < 0 \end{cases}$$

Equação 5.4 Função densidade da distribuição de Weibull

E a função distribuição acumulada é dada por:

$$F(x) = \begin{cases} 1 - \exp\left[-\left(\frac{x}{\beta}\right)^\alpha\right], & \text{se } x \geq 0 \\ 0, & \text{se } x < 0 \end{cases}$$

Equação 5.5 Função distribuição acumulada da distribuição de Weibull

Note-se exponencial é um caso particular da distribuição de *Weibull*, onde $\alpha = 1$ e $\beta = \frac{1}{\lambda}$.

É possível encontrar informação mais detalhada em [2].

Para avaliar se os intervalos entre procuras seguem uma distribuição exponencial e/ou de *Weibull* seguiu-se a realização de testes. Foram realizados dois testes, o teste do ajustamento do qui-quadrado de *Pearson* e o teste de *Kolmogorov Smirnov*, este último trata-se de um teste mais robusto.

O teste do qui-quadrado de *Pearson* pode ser aplicado quando os dados são discretos ou contínuos. Uma das vantagens do recurso a este teste é a facilidade com que se pode modificar os graus de liberdade da estatística de teste de modo a permitir a estimação de parâmetros, todavia também apresenta algumas desvantagens pois além de só poder ser aplicado em amostras de grande dimensão, no caso dos dados contínuos obriga ainda a um agrupamento de dados, algo que tem como consequência a perda de informação fornecida, acabando por ser um teste não tão potente como o de *Kolmogorov-Smirnov* que será abordado em seguida.

Note-se que o teste do qui-quadrado é dado pelas seguintes hipóteses:

$$H_0: X \cap F \text{ vs } H_1: X \cap G, G \neq F$$

Equação 5.6 Hipóteses do teste do Qui-quadrado

Como os dados em causa tratam-se de dados contínuos para realizar o teste do qui-quadrado é necessário classificar os dados em classes, e para este efeito será aplicada a regra de *Mann e Wald*. Considere-se k o número de classes. Para determinar k segundo a regra de *Mann e Wald* seguem-se os seguintes passos:

1º Passo: Classes com igual probabilidade

$$p_j = 1/k \quad e_j = np_j \Rightarrow e_j = n/k = e, j = 1, \dots, k$$

Equação 5.7 Regra de Mann e Wald: Classes com igual probabilidade

2º Passo: Valor esperado comum às classes ≥ 5

$$n/k \geq 5 \Leftrightarrow k \leq n/5 \Rightarrow k = \lceil n/5 \rceil$$

Equação 5.8 Regra de Mann e Wald: Valor esperado comum às classes maior ou igual a 5

3º Passo: Construção das classes:

$$C_1 = (-\infty, a_1), a_1: P(X \in C_1) = 1/k \Leftrightarrow P(X < a_1) = 1/k \Leftrightarrow a_1 = F^{-1}(1/k)$$

Equação 5.9 Regra de Mann e Wald: Expressão e cálculo da primeira classe

$$C_2 = [a_1, a_2), a_2: P(X \in C_2) = 1/k \Leftrightarrow P(a_1 \leq X \leq a_2) = 1/k \Leftrightarrow F(a_2) - F(a_1) = 1/k \Leftrightarrow F(a_2) = 2/k \Leftrightarrow F^{-1}(2/k)$$

Equação 5.10 Regra de Mann e Wald: Expressão e cálculo da segunda classe

$$C_{k-1} = [a_{k-2}, a_{k-1}), a_{k-1}: P(X \in C_{k-1}) = 1/k \Leftrightarrow P(a_{k-2} \leq X \leq a_{k-1}) = 1/k \Leftrightarrow F(a_{k-1}) - F(a_{k-2}) = 1/k \Leftrightarrow F(a_{k-1}) = (k-1)/k \Leftrightarrow F^{-1}((k-1)/k)$$

Equação 5.11 Regra de Mann e Wald: Expressão e cálculo da (k-1)-ésima classe

$$C_k = [a_{k-1}, +\infty)$$

Equação 5.12 Regra de Mann e Wald: Expressão e cálculo da k-ésima classe

Em conclusão, determinam-se:

$$F^{-1}(1/k), F^{-1}(2/k), \dots, F^{-1}((k-1)/k)$$

Equação 5.13 Regra de Mann e Wald: Cálculos intermédios para determinar as classes pela regra de Mann e Wald

E as classes da partição do suporte de X são:

$$C_1 = \left(-\infty, F^{-1}(1/k)\right)$$

Equação 5.14 Regra de Mann e Wald: Primeira classe

$$C_2 = [F^{-1}(1/k), F^{-1}(2/k))$$

Equação 5.15 Regra de Mann e Wald: Segunda classe

...

$$C_{k-1} = [F^{-1}((k-2)/k), F^{-1}((k-1)/k))$$

Equação 5.16 Regra de Mann e Wald: (k-1)-ésima classe

$$C_k = [F^{-1}((k-1)/k), +\infty)$$

Equação 5.17 Regra de Mann e Wald: k-ésima classe

Assim, simplificando, a expressão da estatística de teste será:

$$e_j = n/k = e \Rightarrow X^2 = \frac{k}{n} \sum_{j=1}^k O_j^2 - n$$

Equação 5.18 Estatística de teste com aplicação da regra de Mann e Wald

Quando a distribuição não está completamente especificada, a menos de r parâmetros, a estatística de teste do Qui-Quadrado continua a ter a mesma distribuição assintótica, mas a estimação de cada parâmetro faz “perder” 1 grau de liberdade, pelo que

$$X^2 \sim X_{(k-r-1)}^2$$

Equação 5.19 Distribuição da estatística de teste do teste do qui-quadrado de Pearson

Observe-se agora o caso do teste de *Kolmogorov-Smirnov*. A realização do teste de *Kolmogorov-Smirnov* aos intervalos entre procuras acaba por ser relevante pois trata cada observação de modo individual, aproveitando assim toda a informação fornecida pelas mesmas, acabando por ser um teste mais potente relativamente ao teste do qui-quadrado de *Pearson*, tal como já foi referido anteriormente. O teste de *Kolmogorov-Smirnov* só pode ser aplicado em dados contínuos e requer que a distribuição em H_0 esteja completamente especificada, ou seja, não permite a estimação de parâmetros, contudo existe uma exceção para o caso da distribuição normal e exponencial, casos em que são usados os quantis de *Lilliefors*. Outra vantagem neste teste centra-se no facto de poder ser aplicado em amostras de qualquer dimensão.

Seja X uma variável aleatória contínua. As hipóteses de teste são:

$$H_0: X \cap F \text{ vs } H_1: X \cap G, G \neq F, (G(x) \neq F(x) \text{ para algum } x \Rightarrow \text{Teste bilateral})$$

Equação 5.20 Hipótese de teste para o teste de Kolmogorov-Smirnov

Seja (X_1, X_2, \dots, X_n) uma amostra aleatória de X , a função distribuição empírica associada a este amostra será:

$$F_n(x) = \begin{cases} 0, & x < X_{(1)} \\ 1/n, & X_{(1)} < x < X_{(2)} \\ \dots & \dots \\ i/n, & X_{(i)} < x < X_{(i+1)} \\ \dots & \dots \\ (n-1)/n, & X_{(n-1)} < x < X_{(n)} \\ 1, & x \geq X_{(n)} \end{cases}$$

Equação 5.21 Função distribuição empírica associada a uma amostra aleatória de X no teste de Kolmogorov-Smirnov

É possível encontrar informação mais detalhada em [1].

A estatística de teste de *Kolmogorov-Smirnov* “compara” a função distribuição F com a função distribuição empírica F_n , construída com base na amostra aleatória de X , tal como se pode observar na Figura 5.2.

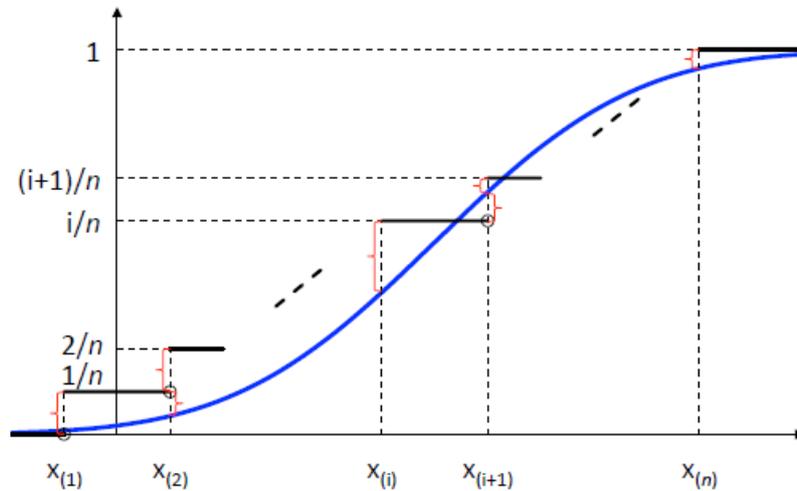


Figura 5.2 Construção da estatística de teste do teste de Kolmogorov-Smirnov

Assim, a estatística de teste é dada por:

$$D_n = \max(D_n^+, D_n^-), \text{ com } D_n^+ = \max_{1 \leq i \leq n} \{i/n - F(X_{(i)}), 0\} \text{ e } D_n^- = \max_{1 \leq i \leq n} \{F(X_{(i)}) - (i-1)/n, 0\}$$

Equação 5.22 Estatística de teste do teste de Kolmogorov Smirnov

Rejeita-se a hipótese nula para valores de D_n grandes, estando os quantis da estatística de *Kolmogorov-Smirnov* tabelados.

No caso particular da distribuição exponencial, considera-se a amostra

$$(X_1^*, X_2^*, \dots, X_n^*), \text{ onde } X_i^* = X_i / \bar{X}, \text{ com } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Equação 5.23 Estatística de teste do teste de Kolmogorov Smirnov

E testa-se as seguintes hipóteses:

$$H_0: X^* \cap \text{Exp}(1) \text{ vs } H_1: X^* \cap G, G \neq \text{Exp}(1)$$

Equação 5.24 Hipóteses do teste de Kolmogorov-Smirnov no caso particular da distribuição exponencial.

Utilizando os quantis da estatística de *Lilliefors* para a Exponencial.

Como se pode concluir do capítulo onde é realizado um estudo de padrões relativamente às estações, percebeu-se que os blocos de tempo em estudo têm realmente interferência na análise do número de bicicletas, assim como os dias de semana em questão. Dado isto, a estimação do valor de λ será influenciado pelo intervalo de horas que se está a considerar, bem como se refere aos 7 dias da semana, ao fim de semana, ou aos dias úteis da semana.

Mais uma vez, será analisado o caso particular da estação 105. Para todos os dias da semana, num intervalo entre as 0h e as 24 horas o valor estimado para λ é aproximadamente igual a 0,044 minutos, enquanto que para um intervalo, entre as 6h e 8h é aproximadamente igual a 0,070 minutos. Por outro lado, entre as 10h e as 13h, esse valor é aproximadamente igual a 0,055 minutos. Observa-se que o período em estudo realmente tem influência, ainda que na escala das centésimas, na estimação do λ .

Uma vez estimado o valor de λ estão reunidas todas as condições para realizar testes de ajustamento para cada intervalo de tempo, de modo a perceber se os intervalos entre procuras seguem, ou não, uma distribuição exponencial.

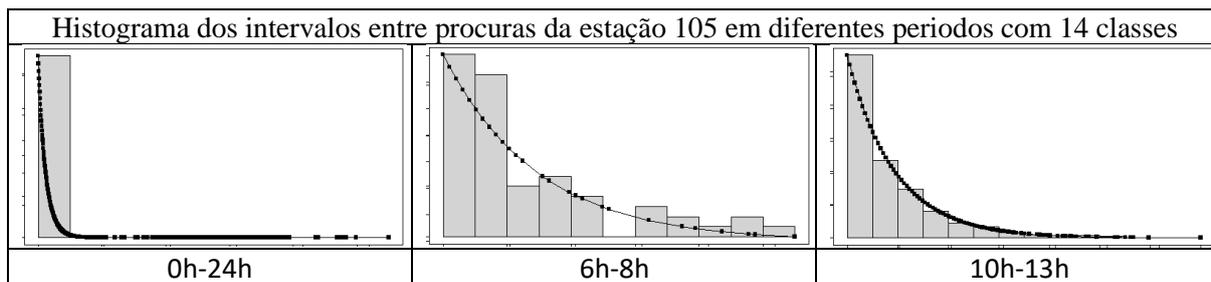


Tabela 5.1 Histograma dos intervalos entre procuras com 14 classes para todos os dias da semana

Na Tabela 5.1 pode ver-se os histogramas dos intervalos de tempo entre procuras num periodo de tempo especifico. Nos histogramas estão sobrepostas as linhas das funções densidades da distribuição exponencial dos respetivos intervalos. O parâmetro utilizado corresponde aos valores de λ anteriormente estimados. Deste modo, pode-se visualmente antecipar o resultado dos testes realizados. Por exemplo, no intervalo entre as 0h e as 24h é previsível que os testes venham a rejeitar o ajustamento à distribuição exponencial. Pelo contrário, para os intervalos entre procuras entre as 6h e as 8h, e entre as 10h e as 13h é possível que os testes não rejeitem esse ajustamento.

Note-se que os histogramas têm 14 classes uma vez que este foi o valor dado como input, contudo pode alterar-se este valor. Porém, alterar o número de classes pode naturalmente modificar a perceção a tirar dos dados, podendo levar a diferentes conclusões ou, pelo contrário, reforçar as mesmas.

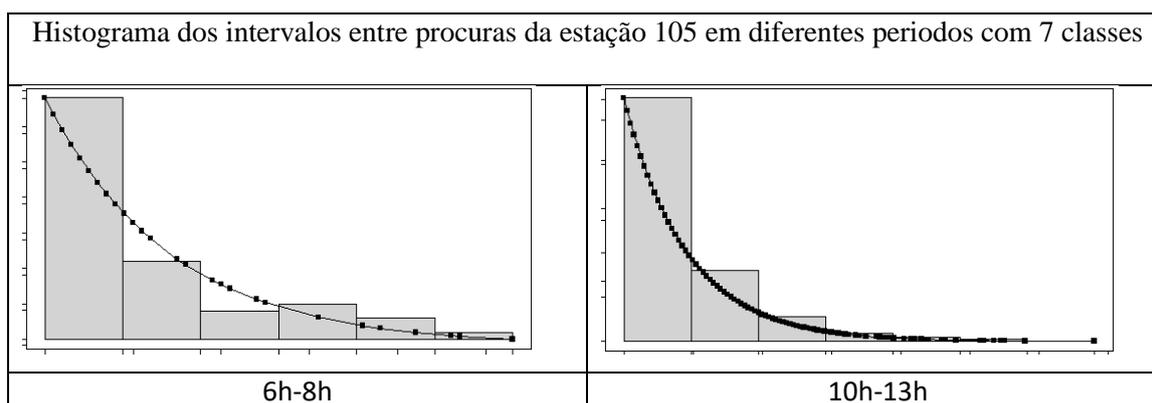


Tabela 5.2 Histograma dos intervalos entre procuras com 7 classes para todos os dias da semana

Neste caso, comparando da Tabela 5.1 com os da Tabela 5.2 nos respetivos periodos as conclusões retiradas nos histogramas com 14 classes são análogas às conclusões que se podem retirar dos histogramas com 7 classes.

Por outro lado, o valor estimado para λ quando são considerados apenas os dados relativamente aos dias úteis da semana num intervalo entre as 0h e as 24h é igual a 0,049 minutos. Já nos dados relativos apenas ao fim de semana, o valor estimado para λ é igual a 0,033 minutos. A diferença entre o valor estimado para λ nos dias úteis e o valor estimado para λ no fim de semana tem uma diferença significativa na casa das centésimas.

Classei	classef	observados
0.00	20.00	806
20.00	40.00	234
40.00	60.00	82
60.00	80.00	27
80.00	100.00	15
100.00	120.00	6
120.00	140.00	1

Tabela 5.3 Tabela do histograma de 7 classes da estação 105-CC Vasco da Gama respetivo ao intervalo entre as 10h e 13h, para todos os dias da semana

A Tabela 5.3 ilustra os valores observados para cada classe, considerando-se 7 classes. A coluna **Classei** representa o limite inferior das classes, e **classef** representa o limite superior. É interessante observar a Tabela 5.3, pois permite compreender que de facto, tal como seria de esperar numa distribuição exponencial, a maioria dos valores estão centrados nos menores valores.

Media	Variância	DesvioPadrao	POATmax	Dimensao	CoefAssim
18.22	381.33	19.53	0.00	1171	1.95

Tabela 5.4 Estatísticas descritivas da estação 105-CC Vasco da Gama respetivo ao intervalo entre as 10h e as 13h, para todos os dias da semana

Na Tabela 5.4 são fornecidas várias informações estatísticas acerca dos intervalos de tempo entre as 10h e as 13h, em todos os dias da semana, referentes à estação 105-CC Vasco da Gama. A média dos intervalos entre procuras é igual a aproximadamente 18,22 minutos, e o desvio-padrão igual a 19,53 minutos. Note-se que na distribuição exponencial o valor médio e o desvio-padrão devem ser iguais. Embora os valores mencionados não sejam exatamente iguais a diferença entre os mesmos não aparenta ser substancialmente elevada, sendo este um indicador de que a distribuição dos dados se ajusta bem a uma distribuição exponencial. O indicador **POATmax** corresponde à percentagem de observações acima da duração máxima estipulada para cada intervalo entre procuras. Portanto, **POATmax** é a percentagem de informação que se rejeita ao eliminar os intervalos que ultrapassem o limite estipulado. Neste caso particular são excluídos os intervalos entre procuras superiores a 24 horas, contudo não existem valores nestes dados concretos cuja duração dos intervalos entre procuras sejam superiores a 24h. A restrição de um limite para o tempo dos intervalos entre procuras deve-se ao facto de que o ajustamento não era muito bom, uma vez que quando se analisa existem casos em que a procura seguinte estava fora do intervalo selecionado. Sabe-se ainda que a dimensão da amostra em estudo é igual a 1171 intervalos de tempos entre procuras, mais uma vez, deve realçar-se a ideia que estas estatísticas são referentes a várias condicionantes como o bloco de horas em estudo, os dias da semana e ainda o número máximo de horas que os intervalos de tempo entre procuras podem ter. Sabe-se ainda o coeficiente de assimetria que neste exemplo é igual a 1,95 que é um valor positivo, indicando que de facto a cauda do lado direito é consideravelmente mais larga do que a cauda do lado esquerdo, tal como no caso da função densidade da exponencial, sendo esta mais uma indicação de que de facto a distribuição dos intervalos poderá ter uma distribuição exponencial.

Uma vez realizados todos os cálculos necessários com recurso a várias ferramentas do R, foram realizados testes do qui-quadrado de *Pearson* para verificar se os intervalos de tempo entre procuras têm uma distribuição exponencial aos níveis de significância usuais.

Teste do Qui-Quadrado de <i>Pearson</i>

significancia	pvalue	Conclusao	significancia	pvalue	Conclusao	significancia	pvalue	Conclusao
0.01	1.00	Nao rejeitamos a Hipotese nula, ou seja, tem uma distribuicao exponencial	0.01	0.00	Rejeitamos a Hipotese nula, ou seja, Nao tem uma distribuicao exponencial	0.01	0.00	Rejeitamos a Hipotese nula, ou seja, Nao tem uma distribuicao exponencial
0.05	1.00	Nao rejeitamos a Hipotese nula, ou seja, tem uma distribuicao exponencial	0.05	0.00	Rejeitamos a Hipotese nula, ou seja, Nao tem uma distribuicao exponencial	0.05	0.00	Rejeitamos a Hipotese nula, ou seja, Nao tem uma distribuicao exponencial
0.10	1.00	Nao rejeitamos a Hipotese nula, ou seja, tem uma distribuicao exponencial	0.10	0.00	Rejeitamos a Hipotese nula, ou seja, Nao tem uma distribuicao exponencial	0.10	0.00	Rejeitamos a Hipotese nula, ou seja, Nao tem uma distribuicao exponencial
10h-13h			0h-24			2h-22h		

Tabela 5.5 Teste do Qui-quadrado em vários períodos para verificar se os intervalos entre procuras seguem uma distribuição exponencial nos respectivos períodos

Pela Tabela 5.5 sabe-se que não se rejeita a hipótese nula, ou seja, os intervalos entre procuras no período entre as 10 e 13 horas tem uma distribuição Exponencial aos níveis de significância usuais.

Por outro lado, rejeita-se a hipótese nula para o intervalo entre as 0 e 24 horas e entre as 2 e 22 horas respectivamente, quer isto dizer que para os níveis de significância usuais estes intervalos não seguem uma distribuição exponencial.

No exemplo do intervalo entre as 0 e 24 horas verifica-se que as conclusões retidas pela ilustração gráfica dos intervalos entre procuras acabaram por coincidir com as conclusões dos respectivos testes.

Uma vez realizados todos os cálculos necessários com recurso a várias ferramentas do R, foram posteriormente realizados testes de *Kolmogorov-Smirnov* para verificar se os intervalos de tempo entre procuras têm uma distribuição exponencial aos níveis de significância usuais.

significancia	pvalue	Conclusao
0.01	0.00	Rejeitamos a Hipotese nula, ou seja, Nao tem uma distribuicao exponencial
0.05	0.00	Rejeitamos a Hipotese nula, ou seja, Nao tem uma distribuicao exponencial
0.10	0.00	Rejeitamos a Hipotese nula, ou seja, Nao tem uma distribuicao exponencial

Tabela 5.6 Teste de Kolmogorov-Smirnov ao intervalo entre procuras no período entre as 10 e as 13 horas

Observando a Tabela 5.6 percebe-se facilmente que no período entre as 10h e as 13h rejeita-se a hipótese nula para os níveis de significância usuais, ou seja, os intervalos entre procuras neste período, segundo o teste de *Kolmogorov-Smirnov*, não seguem uma distribuição exponencial.

Este resultado é contrário ao resultado obtido no teste do qui-quadrado de *Pearson* onde inclusive o *p-value* obtido era significativamente elevado, contudo é algo que se justifica com o facto do teste do *Kolmogorov-Smirnov* ser mais potente do que o teste do qui-quadrado de *Pearson*.

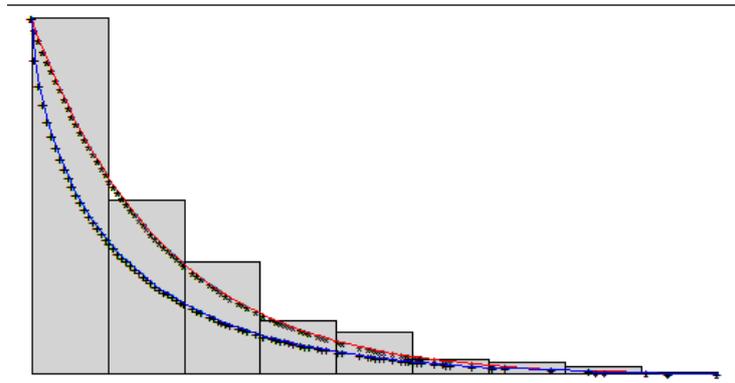


Figura 5.3 Histograma, função densidade da exponencial e função densidade de weibull do intervalo entre procuras no período das 10h-13h

Os gráficos ilustrados na Figura 5.3 dizem respeito ao período entre as 10 e 13 horas, em todos os dias da semana. A linha a azul representa a função densidade de *Weibull*, enquanto que a linha a vermelho representa a função densidade da exponencial. Esta distribuição tem dois parâmetros, um parâmetro de forma, α e outro de escala, β . A distribuição exponencial é um caso particular da distribuição de *Weibull* onde o parâmetro forma é igual a 1.

No período em estudo, os estimadores de máxima verosimilhança para os parâmetros desconhecidos são dados por $\hat{\alpha}=0,887$ e $\hat{\beta}=29,811$.

Após a estimação dos parâmetros e observando a Figura 5.3 entende-se que o gráfico da função de *Weibull* tem uma curva mais acentuada do que a curva do gráfico da exponencial, dando assim a sugestão de que a distribuição dos intervalos entre procuras segue uma distribuição de *Weibull* que não se enquadra no caso particular da exponencial.

Para verificar se os dados seguem de facto uma distribuição de *Weibull* é necessário a realização de testes, contudo ao contrário do caso particular da distribuição exponencial em que foi possível realizar o teste do qui-quadrado de *Pearson* e o teste de *Kolmogorov-Smirnov*, nesta distribuição só é possível realizar o teste do ajustamento do qui-quadrado de *Pearson*, pois o teste de *Kolmogorov-Smirnov* como já foi dito não aceita a estimação de parâmetros com a exceção da exponencial e normal.

O procedimento em termos de cálculo para a realização do teste do qui-quadrado de *Pearson* é análogo ao que foi explicado inicialmente no que diz respeito à distribuição Exponencial, com a única exceção de que a distribuição agora em estudo é a distribuição de *Weibull*.

Posto isto, após a realização dos devidos cálculos com recurso a várias ferramentas do R, obteve-se o seguinte resultado do teste:

Teste de Kolmogorov.Smirnov					
significancia	pvalue	Conclucao	significancia	pvalue	Conclucao
0.01	1.00	Nao rejeitamos a Hipotese nula, ou seja, tem uma distribuicao weibull	0.01	1.00	Nao rejeitamos a Hipotese nula, ou seja, tem uma distribuicao weibull
0.05	1.00	Nao rejeitamos a Hipotese nula, ou seja, tem uma distribuicao weibull	0.05	1.00	Nao rejeitamos a Hipotese nula, ou seja, tem uma distribuicao weibull
0.10	1.00	Nao rejeitamos a Hipotese nula, ou seja, tem uma distribuicao weibull	0.10	1.00	Nao rejeitamos a Hipotese nula, ou seja, tem uma distribuicao weibull
10h-13h			2h-22		

Tabela 5.7 Teste do Qui-quadrado em vários períodos para verificar se os intervalos entre procuras seguem uma distribuição de weibull nos respectivos períodos

Pela Tabela 5.7 sabe-se que não se rejeita a hipótese nula para os níveis de significância usuais, ou seja, no período entre as 10h e as 13h e no período entre as 2 e 22 horas respectivamente, os intervalos entre procuras seguem uma distribuição de *Weibull*.

Note-se que no período entre as 10 e 13 horas também não foi rejeitada a hipótese de os intervalos entre procuras seguirem uma distribuição exponencial, contudo para o período entre as 2 e 22 horas a hipótese dos intervalos entre procuras seguirem uma distribuição exponencial foi rejeitada, mas tal não aconteceu no mesmo período para a hipótese de seguirem a distribuição de *Weibull*.

Em suma, após uma automatização foi perceptível que a distribuição de *Weibull* era aceite numa quantidade superior de períodos do que a exponencial, e visto que a exponencial se trata de um caso particular da distribuição de *Weibull*, a modelação dos intervalos entre procuras para uma futura modelação centrou-se na distribuição de *Weibull*.

Foram realizados para vários intervalos entre procuras, com diferentes amplitudes, o teste do ajustamento do qui-quadrado. Sempre que o *p-value* obtido era inferior a 0,05, rejeitava-se a hipótese de o intervalo entre procuras seguir a distribuição de *Weibull*. No total existem 300 intervalos a considerar ao longo de um dia, cuja amplitude varia entre 1 a 24 horas.

LimInf	LimSup	PVWeibull	Shape	Scale	Amplitude
0	1	1,000	0,710	6,618	1
8	9	1,000	0,938	7,953	1
8	10	1,000	0,947	10,323	2
8	11	1,000	0,902	13,554	3
8	12	1,000	0,894	15,485	4
8	13	1,000	0,873	15,838	5
8	14	1,000	0,867	14,857	6
8	15	1,000	0,867	14,079	7
8	16	1,000	0,863	14,350	8
8	17	1,000	0,857	14,320	9
8	18	1,000	0,861	13,521	10
8	19	1,000	0,853	12,420	11
8	20	1,000	0,847	11,727	12
8	21	1,000	0,846	11,743	13
8	22	1,000	0,838	12,062	14
8	23	1,000	0,830	12,251	15
8	24	1,000	0,828	12,483	16
9	10	1,000	0,962	10,396	1

Tabela 5.8 Parte 1 da Tabela com os valores estimados e o *p-value* para todos os intervalos entre procuras possíveis ao longo do dia na estação 105 seguindo a distribuição de Weibull

A Tabela 5.8 ilustra o limite inferior de cada intervalo na coluna **limInf**, o limite superior na coluna **limSup**, e a amplitude do respetivo intervalo na coluna **Amplitude**. É dada também informação relativamente aos valores estimados dos parâmetros da distribuição de Weibull α e β nas colunas **Shape** e **Scale** respetivamente, e o valor do *p-value* obtido para o intervalo em consideração na coluna **PVWeibull**. Note-se que os intervalos representados dizem respeito apenas aos intervalos que não tem um *p-value* superior a 0,05 ou não tem blocos de horas com uma dimensão de registos inferior a 30, algo que é necessário uma vez que o teste do ajustamento do qui-quadrado tal como já foi referido, não é apropriado para amostras de pequena dimensão.

Quanto maior for o *p-value*, mais significativa será a conclusão obtida a partir do teste realizado. Deste modo os *p-values* foram ordenados de forma decrescente para se perceber quais os intervalos que mais significativos. No entanto, todos os testes realizados segundo o teste de ajustamento do qui-quadrado que não foram rejeitados, obtiveram um *p-value* igual a 1, pelo que não é possível definir quais os intervalos mais significativos pelo valor dos *p-values*. Posto isto, obteve-se a mesma tabela para a distribuição exponencial, caso em que tal como já foi referido é possível aplicar o teste de Kolmogorov-Smirnov.

LimInf	LimSup	PVExponencial	Lambda	Amplitude
10	11	0,955	0,072	1
12	13	0,768	0,098	1
13	14	0,728	0,114	1
14	15	0,702	0,117	1
17	18	0,316	0,133	1
10	12	0,249	0,057	2
12	14	0,200	0,090	2
11	12	0,193	0,081	1

Tabela 5.9 Valores estimados e o *p-value* para todos os intervalos entre procuras possíveis ao longo do dia na estação 105 seguindo um distribuição exponencial

A Tabela 5.9 ilustra tal como na Tabela 5.8 o *p-value* e a amplitude dos intervalos, com a diferença que os testes realizados dizem respeito à distribuição exponencial, estando também ilustrados os valores

estimados para o parâmetro λ na coluna **Lambda**. Os *p-values* da Tabela 5.9 tiveram como base a estatística de teste do teste de *Kolmogorov-Smirnov* ao contrário dos *p-values* obtidos na Tabela 5.8 que tiveram como base a estatística de teste do teste do ajustamento do qui-quadrado de *Pearson*. Note-se que os intervalos que seguem a distribuição de exponencial são consideravelmente menos do que os que seguem uma distribuição de *Weibull*. Os valores da Tabela 5.9 estão ordenados de forma decrescente segundo os valores estimados dos *p-values*. Note-se que existe uma certa predominância para intervalos de amplitude 1. Uma vez que a exponencial é um caso particular da distribuição de *Weibull*, e segundo os resultados obtidos nos testes de ambas as distribuições foi então definido que os períodos a considerar serão de hora a hora.

LimInf	LimSup	PVWeibull	Shape	Scale	Amplitude
0	1	1,000	0,676	6,382	1
8	9	1,000	0,952	8,007	1
9	10	1,000	0,960	10,383	1
10	11	1,000	1,039	14,036	1
11	12	1,000	0,997	12,325	1
12	13	1,000	1,026	10,379	1
13	14	1,000	0,976	8,665	1
14	15	1,000	0,998	8,544	1
15	16	1,000	0,939	10,034	1
16	17	1,000	0,923	9,616	1
17	18	1,000	0,985	7,480	1
18	19	1,000	1,060	6,166	1
19	20	1,000	0,981	6,581	1
20	21	1,000	0,973	8,568	1
21	22	1,000	0,879	9,422	1
22	23	1,000	0,854	9,590	1
23	24	1,000	0,860	10,594	1

Tabela 5.10 Intervalos a considerar na simulação que seguem a distribuição de *Weibull*

A Tabela 5.10 ilustra os intervalos a considerar na fase da simulação. Contudo, nem todos os intervalos seguem a distribuição de *Weibull*, como se pode observar pela inexistência de informação para o período entre a 1h e as 8h. De facto, observando o comportamento dos intervalos em diversas estações nota-se que a partir de parte do período da noite e do período da madrugada existem poucos registos observados, algo que é natural atendendo ao tipo de serviço em estudo. Para estes intervalos seguiu-se uma distribuição categórica. Por exemplo, para no caso da estação 105 foi calculada a probabilidade de durante este período ocorrer apenas uma procura, duas, três e por aí adiante até ao número máximo de procuras alguma vez registado no decorrer deste período. Obteve-se assim a Tabela 5.11 que contém informação sobre a probabilidade de cada ocorrência no intervalo entre a 1h e as 8h.

Obs	Prob	LimInf	LimSup
0	0,804	1	8
1	0,114	1	8
2	0,054	1	8
3	0,005	1	8
4	0,016	1	8
5	0,000	1	8
6	0,005	1	8

Tabela 5.11 Tabela com a distribuição categórica da estação 105

Pela Tabela 5.11 percebe-se que o número máximo de ocorrências que se registou nos dados fornecidos entre o intervalo da 1h e das 8h na estação 105 foi de 6 observações.

Em suma os períodos da noite e madrugada não se enquadram na distribuição exponencial e/ou na distribuição de *Weibull*. Tal fenómeno foi registado predominantemente no que se pode considerar período da noite e madrugada, algo que é compreensível uma vez que a utilização deste serviço no decorrer desse período é naturalmente diminuta. Assim, para modelar os intervalos entre procuras neste período utilizou-se uma distribuição categórica onde foi calculada a probabilidade de ocorrer um certo número de procuras nesse período.

5.2. Modelação dos tempos de viagem

A distribuição uniforme é usada para se ter o número de chances possíveis de um determinado evento ocorrer, dentro de um limite de duas variáveis lógicas. Assim, a distribuição uniforme teoricamente adequa-se ao comportamento da duração das viagens. Realizou-se uma análise gráfica no que diz respeito às durações das viagens realizadas entre cada par de estações. O aspeto gráfico das durações das viagens entre cada par de estações fortaleceu a ideia de que as mesmas seguem uma distribuição uniforme. Isto é, conhecida a estação de partida da viagem e a estação de chegada, sabe-se qual o tempo mínimo e máximo registado entre estes dois pontos, gerando assim o tempo de viagem segundo uma distribuição uniforme.

Considere-se que a estação de partida é a estação 101.

Origem	Destino	Viagens
101	101	0,096
101	102	0,056
101	103	0,055
101	104	0,248
101	105	0,184
101	106	0,033
101	107	0,081
101	108	0,013
101	109	0,081
101	110	0,046
101	114	0,003
101	115	0,006
101	212	0,015
101	219	0,007

○○○

101	453	0,003
101	457	0,001
101	459	0,000
101	460	0,001
101	463	0,000
101	464	0,001
101	468	0,000
101	471	0,000
101	473	0,004
101	474	0,001
101	481	0,004
101	484	0,002
101	485	0,001
101	486	0,001
101	487	0,001
101	488	0,003

Tabela 5.12 Probabilidade de destinos onde a estação 101 é a estação de partida

Na Tabela 5.12 sabe-se qual a probabilidade de cada estação ser a estação de chegada partindo da estação 101. Pela geração aleatória de uma multinomial com as probabilidades apresentadas na Tabela 5.12, imagine-se que a estação de chegada resultou, por exemplo, na estação 453.

Estacao1	Estacao2	Min	Max
101	453	19,35	26,15

Tabela 5.13 Duração mínima e máxima das viagens realizadas partindo da estação 101 até a estação 453

Para simular a duração da viagem realizada entre o par de estações 101 e 453 é gerada aleatoriamente pela distribuição uniforme cujos parâmetros correspondem respectivamente aos valores **Min** e **Max** ilustrados na Tabela 5.13. Neste caso, a viagem terá uma duração entre 19,35 minutos a 26,15 minutos.

6. Simulação e resultados computacionais

Neste capítulo é apresentado o modelo de simulação que foi desenvolvido de modo a permitir retirar importantes conclusões sobre o funcionamento do sistema e, muito especialmente, sobre a resposta do sistema à variação de alguns dos seus parâmetros. Posteriormente, é feita uma análise aos resultados obtidos após várias simulações e são delineadas estratégias com o objetivo de obter um maior lucro para a empresa.

6.1. Simulação

Primeiramente será feita uma descrição do modelo de simulação, designadamente referindo quais os inputs e outputs de simulação que envolve. Em segundo lugar, apresenta-se uma descrição do funcionamento do modelo, com o objetivo de se compreender como é gerido todo o processo. Por último, é feita uma descrição de indicadores importantes que podem ser obtidos a partir do modelo de simulação.

6.1.1. Descrição geral do modelo de simulação

O modelo de simulação tem por objetivo mimetizar a situação real, dadas certas condições existentes. Isso significa que o modelo deve partir da descrição de um certo momento do sistema e, a partir desse ponto, simula todo o seu funcionamento. No final, observa-se o comportamento do sistema e são retiradas conclusões.

Para que se cumpra o objetivo de simular o comportamento do sistema, é necessário gerar dados aleatórios, nomeadamente para os intervalos entre procuras, para a escolha de destinos das viagens e para a duração dessas mesmas viagens. Os dados deverão seguir as distribuições de probabilidade de acordo com os modelos devidamente ajustados para cada situação.

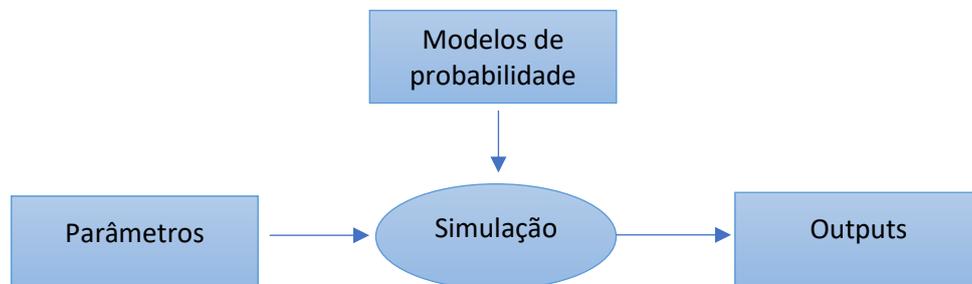


Figura 6.1 Esquema do modelo de simulação

A Figura 6.1 mostra o esquema geral do modelo de simulação. Em cada execução do modelo, são fornecidos os parâmetros e os modelos de probabilidade escolhidos. A partir desses dados são obtidos outputs para posterior análise. A vantagem deste tipo de modelos assenta no facto de ser possível reproduzir de forma rápida um período extenso de funcionamento do sistema e ainda de ser possível observar o comportamento dos outputs em função da alteração de alguns parâmetros.

Entre os parâmetros do sistema incluem-se o número de estações, a capacidade de cada estação, o número de bicicletas no sistema e ainda o período em que decorre a simulação. Nos modelos de probabilidade estão as distribuições escolhidas para as variáveis acima indicadas, incluindo os parâmetros que essas distribuições possam envolver de acordo com comportamentos temporais.

6.1.2. Funcionamento do modelo de simulação

No caso de estudo considerado, o número de estações foi considerado ser igual a 64. As capacidades das estações, sendo um dos parâmetros de grande relevância do modelo, são fornecidas como input. Considerou-se a execução do modelo entre janeiro e julho, para que seja possível capturar suficientes elementos de variação temporal do sistema e, simultaneamente, corresponde ao horizonte temporal do caso de estudo da Gira.

Por outro lado, considera-se que o sistema tem, no seu momento inicial, um número igual de bicicletas em cada estação. Seria possível, naturalmente, considerar uma situação diversa. Todavia, espera-se que o modelo de simulação acabe por convergir para uma situação final em que esses números vão sendo trocados em função do funcionamento do serviço, e que, desse modo, se acabe sempre por retratar a realidade.

Na situação particular do caso de estudo, começou por se considerar 6 bicicletas por estação, o que significa um total de 384 bicicletas no sistema. Posteriormente, para se observar o impacto no sistema, foi o modelo de simulação utilizado para observar o impacto de ir aumentando o número de bicicletas por estação até um máximo de 10, representado assim a disponibilização de 640 bicicletas em todo o sistema.

Na realização da simulação foram utilizados vários ficheiros, os quais foram obtidos após a automatização da modelação:

- 1) Ficheiro *autopares* que fornece informação sobre qual a probabilidade de partindo de uma estação *s* terminar a viagem numa estação *f*;
- 2) Ficheiro *autotp* que fornece informação sobre qual o tempo mínimo e o tempo máximo das viagens realizadas entre cada par de estações;
- 3) Ficheiro *autodm* que fornece informação sobre a distância média entre cada par de estações;
- 4) Ficheiro *autoXXX*, onde XXX corresponde ao número de cada estação. Este ficheiro contém informação sobre quais os estimadores de máxima verosimilhança para cada intervalo com uma distribuição de *Weibull*.;
- 5) Ficheiro *autoMultiXXX*, onde XXX corresponde ao número de cada estação. Este ficheiro contém informação sobre a estação XXX identificando os intervalos da estação que não seguem a distribuição de *Weibull*. Apresenta a probabilidade de ocorrerem *x* procuras em cada intervalo que não segue a distribuição de *Weibull*. Contem ainda informação sobre que tipo de intervalo representa. A necessidade de criar tipos de intervalos por uma questão computacional deve-se ao facto de nem todos os intervalos ao longo de um dia seguirem uma distribuição de *Weibull*. Existem 3 tipos de intervalos:

- a) O intervalo do tipo 1 corresponde a um intervalo onde as observações foram bastante reduzidas no período da madrugada. Esse período da manhã não segue uma distribuição de *Weibull*.

Obs	Prob	LimInf	LimSup	Tipo
0	0,766	0	7	1
1	0,109	0	7	1
2	0,049	0	7	1
3	0,054	0	7	1
4	0,000	0	7	1
5	0,005	0	7	1
6	0,005	0	7	1
7	0,005	0	7	1
8	0,005	0	7	1

Tabela 6.1 Ficheiro *autoMulti110* com intervalos do tipo 1

A Tabela 6.1 ilustra o ficheiro *autoMulti110*. Observando a Tabela 6.1 percebe-se que esta estação tem um intervalo do tipo 1, onde o período da manhã entre as 0h e 7h não segue uma distribuição de *Weibull*. O número máximo de observações que se registaram ao longo deste período foram 8 observações. A probabilidade de ocorrer uma observação é igual 0,109;

- b) O intervalo do tipo 2 corresponde a um intervalo onde as observações foram bastante reduzidas em parte do período da noite e ao período da madrugada. Esse período não segue uma distribuição de *Weibull*.

Obs	Prob	LimInf	LimSup	Tipo
0	0,000	23	8	2
1	0,011	23	8	2
2	0,005	23	8	2
3	0,027	23	8	2
4	0,011	23	8	2
5	0,011	23	8	2
6	0,044	23	8	2
7	0,011	23	8	2
8	0,027	23	8	2
9	0,022	23	8	2
10	0,044	23	8	2



30	0,027	23	8	2
31	0,005	23	8	2
32	0,011	23	8	2
33	0,022	23	8	2
34	0,000	23	8	2
35	0,005	23	8	2
36	0,011	23	8	2
37	0,011	23	8	2
38	0,005	23	8	2
39	0,000	23	8	2
40	0,005	23	8	2
41	0,005	23	8	2

Tabela 6.2 Ficheiro *autoMulti102* com intervalos do tipo 2

A Tabela 6.2 ilustra o ficheiro *autoMulti102*. Observando a Tabela 6.2 percebe-se que esta estação tem um intervalo do tipo 2, que corresponde a parte do período da noite e ao período da madrugada. O intervalo entre as 23h e 8h não segue uma distribuição de *Weibull*. O número máximo de observações que se registaram ao longo deste período foram 41 observações. A probabilidade de ocorrer uma observação é igual 0,011;

- c) O intervalo do tipo 3 corresponde a um dia inteiro, isto é, é um intervalo entre as 0h e 24h. Este tipo de intervalo é o menos comum.

Obs	Prob	LimInf	LimSup	Tipo
0	0,000	0	24	3
1	0,000	0	24	3
2	0,000	0	24	3
3	0,008	0	24	3
4	0,008	0	24	3
5	0,000	0	24	3
6	0,017	0	24	3
7	0,008	0	24	3
8	0,017	0	24	3
9	0,017	0	24	3
10	0,000	0	24	3



30	0,008	0	24	3
31	0,000	0	24	3
32	0,000	0	24	3
33	0,017	0	24	3
34	0,008	0	24	3
35	0,000	0	24	3
36	0,008	0	24	3
37	0,000	0	24	3
38	0,008	0	24	3
39	0,000	0	24	3
40	0,008	0	24	3

Tabela 6.3 Ficheiro *autoMulti487* com intervalos do tipo 2

A Tabela 6.3 ilustra o ficheiro *autoMulti487*. Observando a Tabela 6.3 percebe-se que esta estação tem um intervalo do tipo 3. O intervalo entre as 0h e 24h não segue uma distribuição de *Weibull*. O número máximo de observações que se registaram ao longo deste período foram 40 observações. A probabilidade de ocorrer três observações é igual 0,008;

O processo de simulação tem início na geração aleatória de intervalos entre procuras em cada estação. Esses intervalos são gerados com base nos respetivos modelos de probabilidade. As procuras geradas tomaram como data de partida 25/01/2018 às 00h00, a qual corresponde à data mais antiga dos ficheiros fornecidos pelo caso de estudo.

Na simulação de procuras nas estações com intervalos do tipo 1 se a última data registada pertencer a um intervalo que não segue a distribuição de *Weibull*, são geradas aleatoriamente as procuras irão ocorrer nesse período, com recurso à distribuição categórica (uma generalização da distribuição de Bernoulli). Os tempos entre essas observações são gerados de forma aleatória segundo a distribuição uniforme. Após esta ocorrência, até que se volte a este período, todas as procuras foram geradas pela distribuição de *Weibull*. O tempo gerado pela distribuição de *Weibull* é somado à observação anterior, sendo repetido este processo.

A diferença entre a simulação das procuras em estações com intervalos do tipo 2 comparativamente a intervalos do tipo 1 é que o recurso da distribuição categórica em vez de ocorrer apenas no período da madrugada, ocorre também em parte do período da noite. Todo o restante procedimento é idêntico em ambos os intervalos.

Na simulação de procuras em estações com intervalos do tipo 3, para cada dia são simuladas quantas observações irão ocorrer pela distribuição categórica, e por sua vez o tempo entre estas observações é estimado por uma distribuição uniforme. Neste tipo de estações, para a simulação das procuras nunca se recorre à distribuição de *Weibull*.

Agenda0	s1	Tipo	Nb
25/01/2018 00:00	427	P	6
25/01/2018 00:00	427	P	6
25/01/2018 00:01	428	P	6
25/01/2018 00:01	102	P	6
25/01/2018 00:01	485	P	6
25/01/2018 00:01	423	P	6
25/01/2018 00:02	473	P	6
25/01/2018 00:02	419	P	6
25/01/2018 00:03	449	P	6
25/01/2018 00:03	485	P	6
25/01/2018 00:04	474	P	6
25/01/2018 00:04	426	P	6
25/01/2018 00:04	219	P	6
25/01/2018 00:04	419	P	6
25/01/2018 00:05	488	P	6

Tabela 6.4 Parte da agenda simulada

A Tabela 6.4 ilustra parte de uma agenda gerada na simulação. Esta agenda corresponde à agenda inicial, pois só contém os instantes das procuras. As procuras são representadas pela letra *P* na coluna **Tipo**, as estações onde ocorrem as procuras (ou entregas no caso em que o **Tipo** tem a letra *C*) estão representadas na coluna **s1**, e os instantes em que as procuras (ou entregas) ocorrem estão ilustradas na coluna **agenda0**. A coluna **Nb** representa o número de bicicletas que ficam na estação **s1** no instante respectivo da **agenda0**.

Uma vez que as procuras da Tabela 6.4 dizem respeito às procuras iniciais da simulação e que se considerou que no início da simulação o sistema tem 6 bicicletas em cada estação, as procuras iniciais são garantidamente satisfeitas.

É necessário juntar e ordenar as datas onde ocorrem as procuras em cada estação, uma vez que o número de bicicletas disponíveis para satisfazer cada procura dependerá sempre das procuras anteriores, assim como das futuras entregas nas estações. Ao conjunto das instâncias das procuras de todas as estações dá-se o nome de agenda.

O passo seguinte consiste em simular as viagens. Para o efeito parte-se da estação da primeira ocorrência da agenda, designe-se por estação *e*. Sabendo qual a estação de partida *e*, é simulada a estação de chegada *f* com base na distribuição categórica associada que se encontra indicada no ficheiro *autopares*. Assim, no instante em que ocorre a procura, retira-se uma bicicleta da estação *e*. É simulado o tempo da viagem entre a estação *e* e a estação *f*, com base na distribuição uniforme ajustada para o par de estações em causa, a qual se encontra no ficheiro *autotp*. O instante em que a bicicleta é entregue na estação *f* é então introduzido na agenda, sendo identificado como um instante de chegada e não de procura. Este processo é repetido consecutivamente para todas as datas da agenda. Quando um instante da agenda está identificado como uma chegada, verifica-se se a estação para onde se destinava o fim da viagem tem vagas para lá colocar a bicicleta. Caso a estação tenha possibilidade de receber a bicicleta, acrescenta-se uma bicicleta a essa estação. Por outro lado, se a estação já não tem qualquer vaga disponível, observa-se qual a estação *g* mais próxima da estação *f* que tenha vagas disponíveis para entregar a bicicleta. O tempo de viagem entre a estação *f* e a estação *g* é calculado novamente de acordo com a distribuição uniforme ajustada a esse par de estações, cuja informação se encontra no ficheiro *autotp*. O instante estimado para a entrega da bicicleta na estação *g* é introduzido na agenda, eliminando o instante em que estava pressuposto que a bicicleta seria entregue na estação *f*. Quando ocorre uma procura numa estação sem bicicletas, nada acontece, acabando esta procura por não ser satisfeita. Todo o processo referido é repetido até que não haja mais procuras para processar.

s1	agenda1	s2	agenda2
427	25/01/2018 00:00	408	25/01/2018 00:36
427	25/01/2018 00:00	403	25/01/2018 00:04
428	25/01/2018 00:01	481	25/01/2018 02:32
102	25/01/2018 00:01	107	25/01/2018 09:42
485	25/01/2018 00:01	488	25/01/2018 00:12
423	25/01/2018 00:01	460	25/01/2018 00:27
473	25/01/2018 00:02	412	25/01/2018 00:24
419	25/01/2018 00:02	472	25/01/2018 00:19
449	25/01/2018 00:03	416	25/01/2018 00:19
485	25/01/2018 00:03	307	25/01/2018 00:57
474	25/01/2018 00:04	481	25/01/2018 01:27
426	25/01/2018 00:04	423	25/01/2018 01:16
219	25/01/2018 00:04	408	25/01/2018 00:46

Tabela 6.5 Parte das viagens realizadas com base na agenda da Tabela 6.4

Na Tabela 6.5 é possível observar as estações de partida (estações da coluna **s1**) e as estações de entrega (estações da coluna **s2**). Note-se que as estações **s1** correspondem às estações do **Tipo P**, e as estações **s2** correspondem às estações do tipo **C** da Tabela 6.4. Na Tabela 6.5 a coluna **agenda1** corresponde às instâncias em que é iniciada a viagem, e a coluna **agenda2** às instâncias em que a viagem termina. Cada linha só é introduzida na tabela se na estação em **s1** existirem bicicletas, sendo que se acrescenta uma bicicleta no instante da **agenda1**. Os valores de **s2** e da **agenda2** poderão mudar caso a estação **s2** tenha a sua capacidade lotada no instante em que se encontra na **agenda2**. Quando a estação **s2** tem vagas para entregar a bicicleta no instante registado da **agenda2**, então esse valor é introduzido na **agenda0** ilustrada na Tabela 6.4, e acrescenta-se uma bicicleta à estação.

Note-se que a Tabela 6.4 só ilustra uma parte da agenda, sendo que a procura mais antiga ocorreu dia 25/01/2018 às 00h00 e a procura mais recente ocorreu dia 25/01/2018 às 00h05. Observando a Tabela 6.5 sabe-se que ocorre uma entrega na estação 403 no dia 25/01/2018 às 00h04, instante que está contido entre o período da Tabela 6.4.

agenda0	s1	Tipo	Nb
25/01/2018 00:00	427	P	5
25/01/2018 00:00	427	P	4
25/01/2018 00:01	428	P	5
25/01/2018 00:01	102	P	5
25/01/2018 00:01	485	P	5
25/01/2018 00:01	423	P	5
25/01/2018 00:02	473	P	5
25/01/2018 00:02	419	P	5
25/01/2018 00:03	449	P	5
25/01/2018 00:03	485	P	4
25/01/2018 00:04	474	P	5
25/01/2018 00:04	426	P	5
25/01/2018 00:04	403	C	7
25/01/2018 00:04	219	P	5

Tabela 6.6 Parte da agenda atualizada após algumas viagens

A Tabela 6.6 corresponde à Tabela 6.4 mas atualizada após as viagens registradas na Tabela 6.5. Note-se como os valores da coluna **Nb** se alteraram. Tal como seria suposto, foi introduzida a chegada da estação 403 na agenda.

6.1.3. Resultados do modelo de simulação

Os dados agregados na agenda final, contida após a execução do modelo de simulação, permitem retirar um conjunto diverso de informações sobre o comportamento do sistema. Em particular, é possível observar viagens que foram realizadas e quais as procuras que acabaram por não ser satisfeitas. É também possível observar o vetor constituído pelo número de bicicletas em cada estação.

Em primeiro lugar, um resultado importante que é deduzido a partir do modelo de simulação envolve o conjunto de procuras não satisfeitas. Recorde-se que a agenda do modelo vai guardando essa informação com base nos instantes em que ocorre uma procura numa estação e quando, nesse instante, a estação não tem bicicletas. Isto permite quantificar a taxa de serviço do sistema e de cada estação.

Em segundo lugar, o estado final do sistema pode dar pistas sobre uma eventual convergência, ou não, para um certo estado. É ainda possível observar quais são os estados principais por que passa o sistema, de modo a tirar conclusões sobre o seu funcionamento. Como referido, entende-se por estados a disposição do número de bicicletas num certo instante.

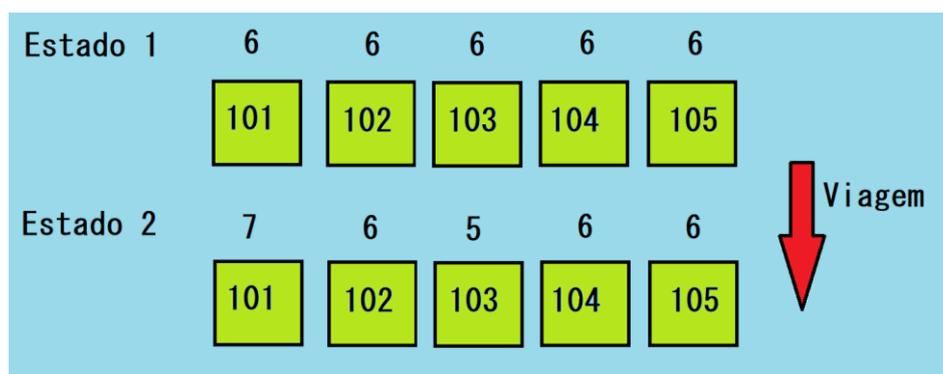


Figura 6.2 Exemplificação dos estados

Na Figura 6.2 estão representados dois estados. O estado 1 corresponde a uma situação em que existem 6 bicicletas em cada estação. Contudo ao realizar-se uma viagem onde a estação de partida corresponde à estação 103 e a estação de chegada corresponde à estação 101 passam a existir 7 bicicletas na estação 101, 5 na estação 103 e 6 bicicletas nas restantes estações. É interessante a análise destes resultados após sucessivas viagens, pois permite entender se o sistema entra num grupo de estados específicos. Note-se que, na realidade, existe um período de tempo entre os dois estados durante o qual já só há 5 bicicletas na estação 103 mas ainda não há 7 bicicletas na estação 101. O fator tempo tem alguma influência na apreciação dos estados. Ainda assim, é interessante analisar os estados para perceber-se se existe ou não uma tendência para o balanceamento de bicicletas entre estações.

6.2. Resultados Computacionais

Neste capítulo é apresentada uma análise aos resultados computacionais obtidos a partir da simulação. Estes resultados são extremamente importantes, pois a partir deles foi possível desenvolver diversas estratégias com o objetivo de otimizar a taxa de serviço. A partir dos resultados obtidos em cada estratégia foi feita uma análise sobre as vantagens e desvantagens de cada estratégia em termos de lucro. A taxa de serviço do sistema é calculada com base na proporção de procuras que foram satisfeitas. As procuras satisfeitas correspondem ao número de viagens realizadas. Assim, o cálculo da taxa de serviço é obtido a partir da seguinte expressão:

$$\text{Taxa de Serviço} = \frac{\text{Dimensão das Viagens}}{\text{Dimensão das Procuras}}$$

Equação 6.1 Expressão da taxa de serviço

Numa primeira abordagem foram realizadas 3 simulações para um sistema com 384 bicicletas. Este passo teve como objetivo compreender se à partida a simulação estava a realizar-se corretamente.

	1ª Simulação	2ª Simulação	3ª Simulação
Número de Procuras	2010848	2025758	1975116
Número de Viagens	448145	449235	449570
Taxa de Serviço	22,3%	22,2%	22,8%

Tabela 6.7 Taxa de serviço de 3 simulações realizadas com base num sistema com 384 bicicletas

Consultando a Tabela 6.7 é possível conhecer a dimensão dos valores obtidos das procuras e viagens em cada uma das 3 simulações. É ainda disponibilizada a informação quanto à taxa de serviço em cada uma das simulações. Comparando a taxa de serviço percebe-se que não varia consideravelmente, sendo que a maior diferença não chega aos 0,6 pontos percentuais. Uma variação substancialmente baixa entre 3 simulações com os mesmos inputs é um bom indicador de que a simulação está a ser bem feita. Existem duas tomadas de decisão para otimizar a taxa de serviço:

- i) Aumento da dimensão da frota;
- ii) Medidas de realocação;

Comece-se por observar o impacto que o aumento da dimensão da frota reflete. Foi feita uma simulação da procura no horizonte temporal fixado entre 25 de janeiro e 31 de julho. No início da simulação, cada estação tem o mesmo número de bicicletas que corresponde ao número médio de bicicletas no sistema. A partir de uma procura simulada foram efetuadas 5 simulações das viagens ocorridas com um sistema com 384, 448, 512, 576 e 640 bicicletas, o que dá respetivamente uma média de 6,7,8,9 e 10 bicicletas.

	Número de Bicicletas				
	384	448	512	576	640
Número Total de Procuras	2010848				
Número Total Viagens Realizadas	448145	504144	550239	587745	630449
Taxa de serviço	22,3%	25,1%	27,4%	29,2%	31,4%
Aumento da Taxa de Serviço		2,8p.p.	2,3p.p.	1,9p.p.	2,1p.p.

Tabela 6.8 Taxa de serviço em sistemas com uma dimensão de bicicletas diferente

Pela Tabela 6.8 verificou-se que a taxa de serviço obtida num sistema com 384, 448, 512, 576 e 640 bicicletas foi respetivamente igual a 22,3%, 25,1%, 27,4%, 29,2% e 31,4%. Comparando a taxa de serviço entre sistemas com uma diferença na dimensão da frota igual 64 bicicleta verificou-se que o aumento da taxa de serviço variou entre 1,9 a 2,8 pontos percentuais.

A estratégia de aumentar a dimensão da frota acarreta custos associados à aquisição de mais bicicletas, manutenção das mesmas, entre outros. Atendendo aos valores do aumento da taxa de satisfação, se a tomada de decisão passar pelo aumento da dimensão de bicicletas no sistema, a dimensão que aparenta ser mais sustentável face às 384 bicicletas atualmente existentes no sistema, é a de aumentar a dimensão para 448 bicicletas, já que é a decisão que revela ter o aumento marginal maior da taxa de serviço. No entanto, uma vez que esse aumento foi igual a 2,8 pontos percentuais, esta estratégia poderá não ser muito aliciante.

No capítulo da simulação discutido sobre o conceito de estados. A dimensão total de estados num sistema corresponde naturalmente à dimensão das viagens. Para compreender se, após várias viagens, os estados do sistema acabam por se fechar num ciclo de estados únicos, ou não, foi calculada a dimensão de estados únicos existentes no sistema.

	1ª Simulação	2ª Simulação	3ª Simulação
Número Total de Estados	448145	449235	449570
Número Total de Estados únicos	420261	418221	418673
Percentagem de Estados únicos	93,8%	93,1%	93,1%

Tabela 6.9 Percentagem de estados únicos de 3 simulações realizadas com base num sistema com 384 bicicletas

A Tabela 6.9 ilustra a dimensão total de estados, a dimensão total de estados únicos e a percentagem de estados únicos para as 3 simulações realizadas com o mesmo input. Verificou-se que a percentagem de estados únicos varia entre 93,8% e 93,1%, valores que são elevados.

	Número de Bicicletas				
	384	448	512	576	640
Número Total de Estados	448145	504144	550239	587745	630449
Número Total de Estados únicos	420261	469787	513222	548316	588725
Percentagem de Estados únicos	93,8%	93,2%	93,3%	93,3%	93,4%

Tabela 6.10 Percentagem de estados únicos em sistemas com uma dimensão de bicicletas diferente

A Tabela 6.10 ilustra a dimensão total de estados, a dimensão total de estados únicos e a percentagem de estados únicos para as simulações realizadas para sistemas com 384, 448, 512, 576 e 640 bicicletas. Verificou-se que a percentagem de estados únicos varia entre 93,8% e 93,2%, valores que são consideravelmente elevados. Uma percentagem de estados únicos elevados permite concluir que os estados das bicicletas não entram em nenhum ciclo restrito de estados, sendo os estados uma característica completamente aleatória do sistema.

De forma a reforçar a conclusão de que os estados são uma característica completamente aleatória, compararam-se os estados obtidos nas 3 simulações com um input totalmente idêntico.

	Total
Soma do # de Estados Únicos das 3 Simulações	1346950
Número Total de Estados únicos das 3 Simulações	1198828
Percentagem de Estados únicos	89,0%

Tabela 6.11 Percentagem de estados únicos das 3 simulações realizadas com base num sistema com 384 bicicletas juntas

Pela Tabela 6.11 percebe-se que a percentagem de estados únicos obtidas é igual a 89%. Este valor é um indicativo de que 11% dos estados entre simulações eram iguais. Contudo, 89% continua a ser uma percentagem de estados únicos substancialmente elevada, o que leva a querer que a conclusão retirada sobre os estados se encontra correta.

Embora a comparação de estados seja uma característica pouco relevante, quando se comparam os estados finais das bicicletas, obtidos no final de cada simulação, percebe-se que independentemente do número de bicicletas no sistema, existe uma certa tendência para algumas estações ficarem lotadas, e por outro lado, para algumas estações ficarem com um défice de bicicletas.

Numa abordagem mais específica foram realizados cálculos para a taxa de serviço em cada estação num sistema com 384 bicicletas. Deste modo foi possível compreender quais as estações cujo as procuras são mais ou menos satisfeitas.

Estação	Taxa de Serviço
471	100,00%
103	99,49%
108	99,32%
106	99,29%
460	99,03%
486	95,71%
450	94,51%
305	94,44%
212	93,68%
487	93,17%
463	93,11%

Tabela 6.12 Taxa de serviço acima dos 90% e respectivas estações

Na Tabela 6.12 estão representadas todas as estações cuja taxa de serviço é superior a 90%. Após cada simulação observou-se a distribuição das bicicletas entre as 64 estações. Verificou-se que as estações apresentadas na coluna Estação em várias simulações terminavam com a frota lotada.

Estação	Taxa de Serviço
308	23,2%
225	22,5%
101	22,0%
426	13,7%
473	10,1%
457	7,3%
306	5,6%
407	4,6%
472	2,5%
430	2,5%
419	0,7%

Tabela 6.13 Taxa de serviço abaixo dos 25% e respectivas estações

A Tabela 6.13 ilustra todas as estações onde a taxa de serviço é inferior a 25%. Observando a distribuição das bicicletas nas estações representadas na coluna **Estação** no final de cada simulação, verificou-se que à exceção das estações 101 e 419 o número de bicicletas em cada estação variou entre as 0 e 3 bicicletas. Ao observar-se o número de bicicletas em cada estação, no final de cada simulação, e a taxa de serviço por estação pode concluir-se que existe uma relação direta entre estes dois fatores. Dada esta relação é interessante observar a percentagem de tempo em que cada estação se encontra sem bicicletas.

Estação	% de cada estação sem bicicletas
471	0,00%
108	0,22%
103	0,24%
106	0,26%
460	0,71%
486	3,35%
463	3,74%
450	3,99%
305	4,14%
212	4,61%
487	6,50%

Tabela 6.14 Percentagens de tempo em que as estações se encontram sem bicicletas inferior a 12 % e respectivas estações

Na Tabela 6.14 é possível observar as estações cuja percentagem de tempo sem bicicletas é inferior a 12%. Comparando a Tabela 6.12 com a Tabela 6.14 é possível perceber que as estações com uma taxa de serviço superior a 90% correspondem às estações com uma percentagem de tempo de cada estação vazia inferior a 12%. Observando a Tabela 6.12 e Tabela 6.14 parece existir uma relação inversa entre a taxa de serviço e a percentagem de tempo em que uma estação fica vazia. No caso da estação 471 onde a taxa de serviço é igual a 100%, a percentagem de tempo da estação 471 sem bicicletas é igual a 0%, o que realça a proposição anterior.

Estação	% de tempo de cada estação sem bicicletas
419	73,26%
457	71,17%
407	70,15%
430	69,56%
473	64,50%
426	61,76%
308	58,62%
485	58,15%
306	56,56%
484	54,93%
225	54,55%
452	54,14%
427	54,12%
428	53,94%
101	53,70%
114	53,18%
423	52,76%
472	51,27%

Tabela 6.15 Percentagens de tempo em que as estações se encontram sem bicicletas superior a 50% e respectivas estações

A Tabela 6.15 ilustra as estações cuja percentagem de tempo sem ocupação de bicicletas é superior a 50%. Comparando a Tabela 6.13 com a Tabela 6.15 verifica-se que as estações com uma taxa de serviço inferior a 25% pertencem ao conjunto de estações com uma percentagem de tempo em que cada estação está vazia superior a 50%. Mais uma vez se verifica uma relação inversa entre a taxa de serviço e a percentagem de tempo em que cada estação se encontra vazia.

As conclusões apontadas a partir da Tabela 6.12 e Tabela 6.15 mostram que uma estratégia acertada de realocação de bicicletas poderá trazer melhorias ao nível da taxa de serviço do sistema. Esse facto é sustentado na constatação de haver estações com uma percentagem elevada de tempo sem bicicletas durante o período de simulado e que, na globalidade desse período apresentam uma taxa de serviço baixa. Conseguir repor bicicletas nessas estações a partir de outras que não apresentem igual condição poderá facilitar a concretização de mais serviço.

Com base nos resultados até aqui referidos, foram realizadas 4 estratégias diferentes de medidas de realocação de bicicletas. Para se compreender melhor as estratégias, é importante ter presente o que significa o conceito de número médio esperado de bicicletas. Para calcular o número médio esperado de bicicletas num certo instante, numa determinada estação, é necessário ter conhecimento sobre o número de bicicletas ao longo do tempo. Deste modo, o número médio de bicicletas no intervalo $[t, t+1]$, onde t representa as horas e é dado por $t=0, \dots, 23$ é dado pela seguinte expressão:

$$\text{Número Médio de Bicicletas}_{[t,t+1]} = \frac{\sum \text{número de bicicletas no período } [t, t+1]}{\sum \text{número de bicicletas}}$$

Figura 6.3 Cálculo do número médio de bicicletas no instante $[t,t+1]$

De seguida é apresentada uma descrição de cada uma das 4 estratégias.

- a) A 1ª estratégia consiste em detetar os momentos em que cada estação fica com um número de bicicletas inferior a 20% do número médio de bicicletas nesse instante. Quando a situação anterior ocorre são recolhidas bicicletas de uma estação com mais de 70% do número médio de bicicletas nesse instante, sendo privilegiadas as estações que reúnem esta condição e que estejam mais próximas. O número de bicicletas a transitar para a estação com défice de bicicletas corresponde ao valor necessário para que a estação possa ter pelo menos 20% do número médio de bicicletas nesse período.

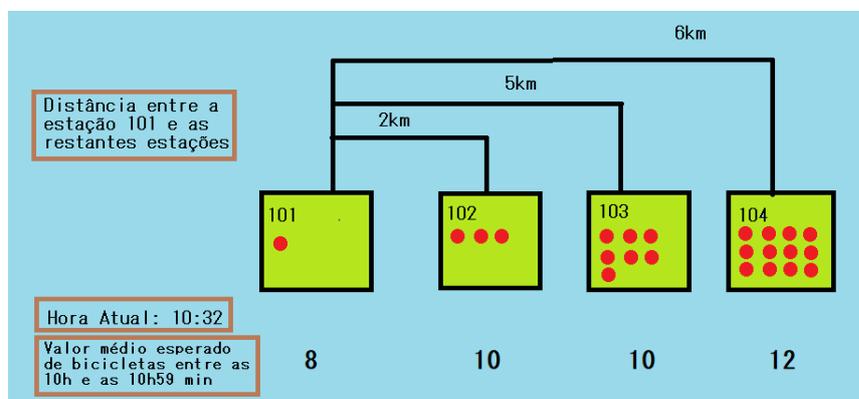


Figura 6.4 Esquematização dos dados usados nas estratégias

Na Figura 6.4 os quadrados a verde representam as estações designadas como 101, 102, 103 e 104. Os pontos a vermelho dentro de cada quadrado representam o número de bicicletas presentes em cada estação. Os valores por baixo de cada estação representam o número médio esperado de bicicletas em cada uma das respetivas estações no intervalo entre as 10h e as 10h e 59 minutos. No exemplo da Figura 6.4 a estação 101 tem às 10h32, apenas 1 bicicleta. Uma vez que esse valor é inferior a 20% ($1 < 0,2 \times 8 = 1,6$) é necessário transferir bicicletas para a estação 101. De modo a atingir os requeridos 20% do número médio de bicicletas é necessário transferir $[1,6 - 1] = [0,6] = 1$ bicicletas para que a estação 101 fique com mais de 20% do número médio de bicicletas neste período. A estação 102, 103 e 104 possuem respetivamente 30%, 70% e 100% do valor médio esperado entre as 10h e as 10h e 59 minutos. Uma vez que se deve ir buscar o número de bicicletas em falta a uma estação que tenha pelo menos 70% do número

médio de bicicletas, automaticamente rejeita-se a estação 102 como estação de recolha, mesmo sendo esta a estação mais próxima da estação 101. Como estações de recolha de bicicletas sobram as estações 103 e 104 uma vez que satisfazem a restrição dos 70%. A estação 103 encontra-se a 5km de distância da estação 101 enquanto que a estação 104 encontra-se a 6km de distância. Uma vez que a estação 103 é a estação mais próxima da estação 101, será nesta estação que é feita a recolha de uma bicicleta de modo a transferir a mesma para a estação 101.

- b) A 2ª estratégia é análoga à 1ª estratégia diferindo apenas no número de bicicletas que são deslocadas entre estações. Assim, é necessário colocar bicicletas numa estação que tenha um número de bicicletas inferior a 20% do valor esperado e as bicicletas são recolhidas da estação mais próxima que tenha pelo menos 70% do valor esperado de bicicletas na estação, nesse instante. Contudo, o número a deslocar de bicicletas corresponde à diferença entre o número que existe estação de recolha e 50% do número médio de bicicletas da estação com défice de bicicletas. Pelo exemplo da Figura 6.4 continuam a ser recolhidas bicicletas da estação 103 para a estação 101. Ao contrário do que acontece com a estratégia 1 onde se transfere apenas uma bicicleta, com a estratégia 2 são deslocadas $[7 - 0,5 * 10] = [7 - 5] = [2] = 2$ bicicletas. Note-se que se 50% do número médio de bicicletas da estação com défice de bicicletas, no intervalo em análise for superior a 70% do número médio de bicicletas de uma potencial estação de recolha, essa estação é rejeitada, uma vez que não possui um número suficiente de bicicletas para satisfazer o critério usado. A necessidade de criar esta 2ª estratégia deve-se ao facto de a 1ª estratégia ter como consequência uma necessidade excessiva de deslocações de bicicletas entre estações, o que levou a pensar numa solução que diminuísse o volume de deslocações.
- c) A 3ª estratégia teve como ponto de partida a informação retida relativamente ao número de bicicletas por cada estação no final de cada simulação. Em várias simulações um determinado grupo de estações terminou sempre com um número de bicicletas coincidente à capacidade das estações. Esta estratégia é análoga à 2ª estratégia diferindo apenas no grupo de estações onde se realiza a recolha. As estações a considerar para ser feita a recolha correspondem apenas às estações lotadas no final de cada simulação. Contudo, permanece a necessidade da recolha se realizar na estação mais próxima com mais de 70% do número médio esperado de bicicletas.
- d) Observando a Tabela 6.16 percebe-se que a estratégia que obteve uma melhor taxa de serviço foi a estratégia 1. Embora as estratégias 2 e 3 tenham como propósito a diminuição do número de deslocações, tal não se verificou, o que reafirma a estratégia 1 como a melhor estratégia entre as 3 estratégias adotadas. Comparando a taxa de serviço da simulação sem qualquer estratégia aplicada, e a taxa de serviço aplicando a estratégia 1 têm-se um aumento de 12,8 pontos percentuais. Na estratégia do aumento de bicicletas no sistema concluiu-se que a decisão mais vantajosa face aos resultados e custos poderia ser a de um sistema com 448 bicicletas. Já nas estratégias de medidas de realocação entende-se que a melhor estratégia em termos de resultados e custos associados é a estratégia 1. Deste modo, foi realizada uma quarta estratégia que, de modo resumido, corresponde à estratégia 1 aplicada num sistema com 448 bicicletas.

	Simulação s/ estratégias	Estratégia 1	Estratégia 2	Estratégia 3	Estratégia 4
Número de Procuras	1975116	1975116	1975116	1975116	1975116
Número de Viagens	449570	701456	574005	556670	792225
Número de Deslocações		192724	283446	251080	222890
Taxa de Serviço	22,8%	35,5%	29,1%	28,2%	40,1%

Tabela 6.16 Taxa de Serviço da estratégia 1, estratégia 2 e estratégia 3

Observando a Tabela 6.16 percebe-se que a estratégia 4 foi a estratégia que obteve uma maior taxa de serviço. Embora a taxa de serviço obtida pela estratégia 4 seja superior à taxa de serviço obtida pela estratégia 1, o número de deslocamentos da estratégia 4 é superior ao número de deslocamentos da estratégia 1. Note-se que por cada deslocamento existe um custo associado. Sendo assim, quanto maior o número de deslocamentos maior será o custo associado. A estratégia 4 além de estar associada ao custo das deslocamentos realizadas, bem como as estratégias 1, 2 e 3, tem também o acréscimo de estar associada ao custo da aquisição e manutenção de novas bicicletas.

A estratégia 4 tem uma maior taxa de serviço do que a estratégia 3, mas também apresenta mais custos associados. Para decidir se deve-se seguir a estratégia 4 ou a estratégia 3 é necessário avaliar os ganhos obtidos pelas taxas de serviço e os custos associados a cada estratégia. A estratégia escolhida deverá ser a que gera um maior lucro.

$$\textit{Lucro} = \textit{Ganhos} - \textit{Custos}$$

Equação 6.2 Expressão do Lucro

7. Conclusão

O sistema de compartilhamento de bicicletas é algo bastante complexo uma vez que se trata de um processo construído a partir de uma coleção de ações tomadas ao longo do tempo. Cada ação tomada é influenciada por fatores internos e externos ao sistema. Como fatores internos encontram-se o número de estações e de bicicletas, assim como a capacidade de cada estação. Por outro lado, como fatores externos têm-se o número de bicicletas em cada estação, a categoria das bicicletas disponíveis, as condições ambientais, entre outros.

A procura apresenta diversos padrões temporais. A afluência da procura nos dias úteis de uma semana é consideravelmente diferente da afluência da procura no decorrer do fim de semana. A procura acaba por ser superior nos dias úteis, algo que leva à suposição de que a maioria dos utilizadores do sistema de *bikesharing* da empresa Gira, recorre a este tipo de serviço para deslocações até os seus locais de trabalho e/ou de ensino. Por outro lado, também é significativa a diferença entre a procura em parte do período da noite e madrugada comparativamente ao resto do dia, altura em que existe uma predominância maior da procura. A maioria dos intervalos entre procuras seguem a distribuição de *Weibull*.

Os resultados computacionais permitem concluir que com o sistema atual a taxa de serviço é substancialmente baixa, rondando uma taxa de satisfação igual a 22%. É oportuno adotar uma nova estratégia para que a taxa de serviço aumente e conseqüentemente a satisfação da necessidade por este importante sistema de mobilidade, levando a um aumento dos resultados da empresa Gira. As estratégias a adotar devem passar pelo aumento do número de bicicletas no sistema e por novas medidas de realocação das bicicletas. Ambas as estratégias têm custos associados, se por um lado o aumento do número de bicicletas traz associado um custo de aquisição e manutenção das mesmas, por outro as medidas de realocação estão associadas aos custos de transporte de bicicletas. Observando os resultados obtidos uma boa estratégia passaria por repor as bicicletas numa estação sempre que a mesma possuir um número de bicicletas inferior a 20% do número médio de bicicletas na hora em que tal se verifica. O número a repor deve ser o necessário para que a estação passe a ter um número de bicicletas igual ou superior a 20% do número médio de bicicletas nesse instante. As bicicletas a deslocar para uma estação com défice de bicicletas devem ser recolhidas da estação mais próxima, com um número de bicicletas superior a 70% do número médio esperado de bicicletas nesse instante. A aquisição de mais bicicletas dependerá de uma avaliação face aos ganhos e custos.

Seria interessante existir uma oferta para cada utilizador que optasse por entregar a bicicleta em uso no final da sua viagem numa estação que tivesse um número de bicicletas inferior a 20% do número médio de bicicletas esperado nesse instante. Esta estratégia seria vantajosa para o utilizador e para a empresa Gira. Para tal, seria necessário encontrar um conjunto de benefícios para o utilizador que o estimulasse a realizar tal escolha com vantagens evidentes para o sistema.

Existem dados que seriam interessantes de analisar caso num estudo futuro fossem disponibilizados. Seria interessante tomar conhecimento sobre qual as estações onde o utilizador pretende efetivamente iniciar e terminar a sua viagem. Atualmente só existe informação sobre as estações de início e fim da viagem que acabou por ocorrer, podendo essas ser fruto das circunstâncias em que se encontravam as estações que o utilizador verdadeiramente gostaria de utilizar. Com a disponibilização deste tipo de informação, seria possível realizar um estudo mais robusto da procura em cada estação, uma vez que podem acontecer casos, como o referido, em que o utilizador se dirige até uma estação que não tem bicicletas. Por outro lado, é interessante também ter conhecimento sobre qual a estação que o utilizador pretende entregar a bicicleta uma vez que pode acontecer o utilizador querer entregar a bicicleta numa estação lotada. Outras informações interessantes seriam o conhecimento sobre a categoria das viagens realizadas e a identificação do perfil do utilizador. As categorias de uma viagem podem dividir-se em viagens para deslocações laborais, ou em viagens de lazer, enquanto que relativamente ao utilizador poderia ser disponibilizada a informação sobre o tipo de passe utilizado. Este tipo de informação

permitira tirar conclusões mais robustas sobre onde seria interessante construir futuras estações, e fazer uma melhor gestão dos passes atuais.

Seria interessante realizar-se também um estudo no que diz respeito à localização das estações. É interessante perceber quais as localizações das estações com uma maior taxa de procura, de modo a compreender se existe algum aspeto relevante que justifique uma maior procura. Esta informação permitiria a realização de uma avaliação sobre quais as áreas de Lisboa onde seria vantajosa a construção de novas estações.

A escolha de uma bicicleta realiza-se 15 segundos antes da recolha da mesma. O utilizador pode consultar na aplicação o número de bicicletas na estação para a qual se pretende dirigir e a informação disponibilizada ser que naquele instante a estação encontra-se com bicicletas disponíveis. Contudo, no instante posterior em que o utilizador chega ao local, corre o risco de já não existirem bicicletas, pois nesse espaço de tempo outros utilizadores podem ter recorrido à utilização das mesmas. Assim, o utilizador corre sempre o risco de chegar a uma estação vazia. Seria possível contornar esta situação se for permitido ao utilizador a reserva de uma bicicleta, na estação desejada, através da aplicação.

8. Bibliografia

1. Severino, E. (2020). Apontamentos da Cadeira de Análise de Dados.
2. PortalAction (s.d). Distribuição de Weibull. URL: <http://www.portalaction.com.br/probabilidades/613-distribuicao-weibull>
3. CaosPlanejado (s.d). Bikesharing: A bicicleta além do lazer. URL: <https://caosplanejado.com/bike-sharing-a-bicicleta-alem-do-lazer/>
4. ShinyRStudio (s.d). Add Control Widgets. URL: <https://shiny.rstudio.com/tutorial/written-tutorial/lesson3/>
5. Gira (s.d). Sobre a Gira. URL: <https://www.gira-bicicletasdelisboa.pt/>
6. SupportMinitab (s.d). O que é uma série temporal. URL: <https://support.minitab.com/pt-br/minitab/18/help-and-how-to/modeling-statistics/time-series/supporting-topics/basics/what-is-a-time-series/>
7. Martins, M. (2014). Dissertação de uma plataforma online de apoio a um serviço de *bikesharing*.
8. Christopher Moon-Milklaucic, Anna Bray-Sharpin, Ivan de la lanza, Azra Khan, Luca Lo Re, Anne Maassen. (2019). The Evolution of Bike Sharing: 10 Questions on the Emergence of New Technologies, Opportunities, and Risks

9. Anexo

Grafico do Nr de Bicicletas por Estacao

Escolha os dados das estações

Browse... dados.csv Upload complete

Selecione a estação

- 104
- 104
- 107
- 105
- 109
- 110
- 452
- 481
- 482

Serie Temporal e Tabela de Freqüencias Por hora do dia Por dia da semana Por Mes Duracao das viagens Distancia das viagens Modelacao

Modelacao Pares de Viagem

dimensao = 475

Periodo de Tempo: 0 3 6 9 12 15 18 21 24

Horas maximas dos intervalos: 21

Selecione os dias para estudo Todos os dias

Number de Classes: 1 5 10 15 20 25 30 35 40 45 50

Shape 0.8112054 Scale 50.59412

Distribucioes:
 Exponencial
 Weibull

Estimador Weibull shape e scale:
Shape= 0.8091418 Scale= 50.5115