The Analysis of Car Parking Spaces in Dublin City Centre.

National College of Ireland

Ailis Curran

BSc in Computing

Data Analytics

2018/2019

x15325146

Declaration Cover Sheet for Project Submission

SECTION 1 *Student to complete*

Name: Ailis Curran
Student ID: X15325146
Supervisor: Ade Fajemisin

SECTION 2 Confirmation of Authorship

The acceptance of your work is subject to your signature on the following declaration:

I confirm that I have read the College statement on plagiarism (summarised overleaf and printed in full in the Student Handbook) and that the work I have submitted for assessment is entirely my own work.

Signature: A/S Curran Date: 05-5-19

NB. If it is suspected that your assignment contains the work of others falsely represented as your own, it will be referred to the College's Disciplinary Committee. Should the Committee be satisfied that plagiarism has occurred this is likely to lead to your failing the module and possibly to your being suspended or expelled from college.

Complete the sections above and attach it to the front of one of the copies of your assignment,

3
Table of Contents

5

Executive Summary

This project is compiled for the BSc Honours in Computing with a specializing in Data Analytics, at the National College of Ireland.

The overall idea for this project is to analyse different trends in the car parking spaces in Dublin City Centre. Data was taken from a website and formatted in different ways that were interesting and easy for others to understand in a clear and concise manner. The reasoning behind choosing car parking spaces was that it was a topic never really explored before and this overall information would be beneficial to a multitude of people.

Once data was stored through a python file in Amazon Web Services and then the analysing of the data was completed on R Studio and Tableau. Which gives us clear representations of the data.

This will all be discussed in the following document.

6

Introduction

Background

Cars in general have been a prominent topic for the past couple of years. With global warming and fumes being the centre of this topic. Now by no means in Ireland, Dublin in this study the worst of car fumes but there has been a huge insensitive for people to take public transport. The public in Dublin City Centre has hugely improved over the past few years, with the extension of the green line on the Luas being the most recent. Companies now have TaxSaver Leap cards which come out of people's wages at a reduced cost. On the other hand Dublin is the place of multi-story car parks. These car parks are not cheap and if you are looking at spending the whole day in one of the car parks it could set you back €36.00 for 24 hours (Ilac Carpark- Parnell Street).

Moving out from the City Centre there are Pay and Display machines on roads and smaller carparks. These work on biases that you put the money in for as long as you need otherwise the car will be clamped and with that comes another fee of maximum €150. Cars can be clamped for various other reasons, for example if a car is parked on double yellow lines or parked outside a no parking sign.

A new investment opportunity that people are now looking into is renting out their car parking spaces. People who live in town or just outside are taking full advantage of the chance. These people who live in town don't have cars as everything is just on their doorstep and they probably work very local to. With websites like Park pnp, it makes it so much easier to rent out your space but to also find a space that suits where you want to go. Prices are anything from €2.50 per hour just outside the City Centre or as high as €65 per hour in Grand Canal. For people who want to rent per monthly, Rent.ie looks at spaces per month, from as little as €100 per month in the IFSC area of town or as high as €1,200 per month in St Anne's Street. There are a lot of people who use these kind of services, but unfortunately, these figures won't be included in my overall report as it is impossible to tell who is renting out each car parking space that is owned privately.

In 2015 a study was conducted to see what form of transport do people use the most percentage wise. The bus came up top with 42%, the car was the second highest at 17%, passengers in vehicles 3rd with 15%, cycling at 4% and walking at 3%. According to the study there are 13,000 car parking spaces provided to the public which are free of charge. The car parks that were analysed at for my project are all priced per hour or per day. Having so many car parking spaces for free is never a good incentive for people to use public transport. In August 2015 the National Transport Agency and Dublin City Council submitted the joint transport proposals for the future, now this brought up the debate of weather the 13,000 spaces in Dublin City Centre should be charged and this money is put into public transport.

All of these factors incredibly interesting as when the idea struck for the project from the fact that when working in at the internship in ESB Networks in Clanwilliam House, the only car parking there was for higher members of staff but was completely unsure where to park myself. It wasn't until a colleague of mine told me she parked a 20 minute walk from the office as there was no bus that she could take to get to work. This is where my interest began.

Aims

The overall aim for this project is to develop a prediction algorithm to try and obtain the parking spaces in Dublin City over time. As we have seen in the above sub-heading that with 31% of people driving into town, this will only increase with time. How many spaces will these car parks have left? To compare other car parks against each other and do analysis on them through R Studio and Tableau would also be hugely interesting.

Aim 1: Which car park has the most spaces taken up during the time investigated, and include factors (would shopping centres like Brown Thomas be busier on a Saturday then say Parnell).

Aim 2: Which area has the most spaces full, ie. North/South side.

Aim 3: Compare Monday-Friday 9am- 6pm with Saturday & Sunday 9am- 6pm (This will show how many people use the spaces on a typical 9am-5pm working day).

Aim 4: To overall have a clean and concise view the most and least busy car parking spaces and understanding the reasons as to why?

Aim 5: For this data to be easily understandable and can be used in the future of building and analysing car parks.

Technology

R: Software environment perfect of statistical computing and displaying a multitude of graphics. Widely used in developing data analysis.

R Studio: Open source (IDE) Integrated Development Environment for R. Also like R used for statistical computing and displaying graphics.

Microsoft Excel: Spreadsheet program which is included in the overall Microsoft Office applications. The spreadsheets present tables of values which are arranged in rows and columns that can be manipulated by using arithmetic operations and functions.

Microsoft Word: Graphical word processing program which users can type with.

Amazon Web Service: Secure cloud service platform which offers database storage and content delivery. This project was carried out on the Ubuntu platform.

Sublime Text: Proprietary cross-platform source code editor with a Python application programming interface. Supports multiple mark-up languages and functions which can be added with plugins.

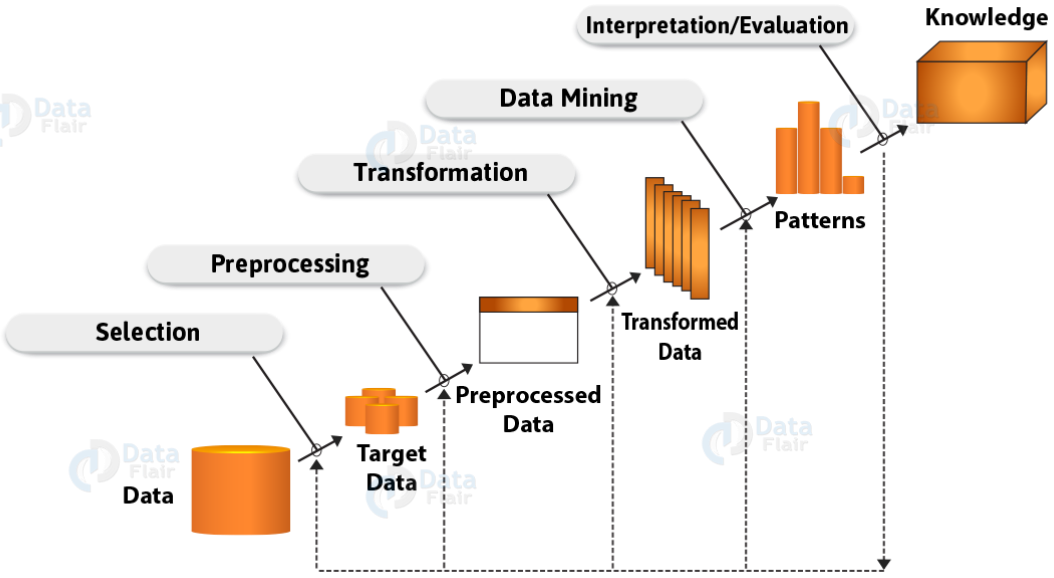
Python3: High-Level, interpreted, interactive and object oriented scripting language. It is extremely readable.

Google Drive: Personal cloud storage letting users to be able to sync digital content from a number of devices.

Filezilla: Powerful software which transfers files over the internet. Using the popular SFTP client to be able to work with AWS and PuTTY.

PuTTy: Open-source terminal emulator, serial console and network file transfer application. It supports several network protocols including the one used in this project SSH.

Methodology



For the project it was necessary to complete a KDD Methodology, as it was the best process to extract knowledge from the data that was web scraped. KDD refers to the process of fining useful knowledge from the data. It completes this by involving the evaluation of the patterns, this helps the overall decision of what actually qualifies as knowledge.

Selection: Once the data was web scraped from the Dublin City Council website. It was then time to decide what exactly did questions did want to find from this

and who was the target audience? Once my aims for the project was set it was onto the next part of the KDD process.

Pre-processing: This was the stage where cleaning the data and preparing it for the next stage of the KDD process, this stage is extremely important as without having clean data there would be incorrect data and numerous outliers.

Transformation: Then used R and R Studio and loaded the data through a .csv file, for this section of the KDD process, this was when the overall analysis of the data could been seen in a more visually satisfactory.

Data Mining: This is where both R Studio and Tableau were used to help show the data in a graphical and table presentation. This is where it was able to see patterns and start putting my aims of the project into place.

Interpretation/ Evaluation: This is the final step of the KDD methodology and this is where everything started coming together and was able to visually prove the questions and aims with graphical evidence. Then this step is complete, having have made sure that it has met all the requirements of the project.

Research

Before stating my project, research went into the different modes of transport that people used going to work in Dublin City Centre. This was the first start and mentioned before in the background section of the project that the reason behind the project was that when completing the third year internship. Overall unsure of the car parking spaces around the local area and other colleagues had to park quite far away. Originally found the statistics of means to travel to work from the Central Statistics Office. This gave the statistics from 1986 to 2016 of how people got to work. What was noted was that from 1986 to 2011 the amount of people driving to work got steadily larger and larger. From being a high number as 44.9% in 1986 to an unbelievable 63% in 2011. In 2016 the number of people driving to work was down to 61.8%. We could look at this as possibly better transport links or new bus routes added.

Also need to remember that some people also carpool to work, these statistics are also included on the website. With the number gradually going down and down these are interesting numbers. In 1986 were looking at around 10% of

people carpooling and in 2016 the number has gone down to 4.1% which is surprising.

Can't forget that cars aren't the only ones who use car parking spaces, with trucks, vans and articulated trunks also being on our road and with an increasing number of people using them as a mode of transport to work with 7.3% of people in 2016 using them to get to and from work.

This information is what really attracted me to looking into car parking spaces in Dublin and showed me new and interesting ways that people are now using car parking spaces, for example renting them out to websites so people can park closer to work at possibly a lower cost than it would be to park at a multi-storey parking space in the centre of town. With actual websites like Park pnp and Parking Motel renting out parking spaces on a one time deal, to people renting out or selling their parking space completely on daft.ie or rent.ie or even people putting their spaces on gumtree or done deal. It has amazed me how important it is for people to park nearer work.

Then went onto researching how to scrape using python, this was the first year that we were exposed to Python and R Studio, which was a shame as these languages would have really benefited me in my degree in the past. As the web scraping part of the project was being completed in the first semester of fourth year this meant that there was a lack of confident working with Python but after much learning and trial and error got the web scraping working and also got the data to save to a .csv file.

When second semester came around, the Python script needed to be automated which meant it need to be on a cloud service platform. As this was the first time to even look at a cloud service platform, proceeded to dive into Amazon Web Service this took a lot of reaching and a few hiccups through the way.

Had a basic knowledge of R Studio before working on that side of the project but it also took a good bit of researching to get the best out of the software for the project.

12
Acronyms and Abbreviations

Acronyms

NCI: National College of Ireland

GUI: Graphical User Interface

DCC: Dublin City Council

CSO: Central Statistics Office

AWS: Amazon Web Service

CMD: Command Prompt

DA: Data Admin

Definitions

Linear Regression Model

Linear Regressions is used to predict the value of one variable based on a second or more input predictor variable.

R-Squared Value

R-Squared is a statistical measure of how close the data is to the fitter regression line. Also known as coefficient of determination.

Knowledge Discovery in Databases

Knowledge Discovery in Databases also known as KDD is a process of using techniques including data mining to discover a comprehension within data.

Car Park

Is an area or a building where cars and other vehicles can be left for a short period of time.

Multi-Story Car Park

Is a building designed for parking cars and other vehicles which is over a number of floors and levels.

Project Restrictions

Data

The data that was web scraped is from the Dublin City Council website which is reliable enough, during the time completing this project it came to attention the website never had any data for Abby Street even though it was still on the website. So I made a decision to completely get rid of the column as an attempt to start cleaning the data.

Amazon Web Service

With Amazon Web Service all was going smoothly the instance was running and there was no problems all was green and saying running and 2/2 checks passed. Originally set up the AWS it was the 16th of March the idea was to collect the data on the 30th of April. Essentially having just over a months' worth of data, which would be ideal. To my shock and horror when it came to looking through my Filezilla it said Error: No Connection, automatically putting me into panic mode. Once being able to fix the issue, it came to my attention that there was no data from when analysing the Filezilla it last took data from 22nd of March. Now the AWS is running smoothly but sadly will not have the amount of data that would have been ideal. Being unable to contact Amazon as it was saying it was running correctly, as currently on a free tier of the package. This situation has been discussed with my supervisor.

Time

The project has been running from September 2018 till May 2019. Due to exams, other projects and learning Python, R Studio and web scraping from scratch this year that have distracted me from the final year project.

Software

As mentioned before not learning Python and R Studio in the past has made this project harder then it needed to be.

Cost

AWS, although on a "Free" tier of the program still took 89 cent from my bank account every week. This was a small price to pay so really didn't mind.

System

Requirements

The data requirements that was needed to make sure the project was completed to a perfect degree was as followed:

The dataset that was needed for the project was open source and was found on the Dublin City Council website. The data contained was 14 car parking spaces in Dublin City, these included: Parnell, Ilac, Jervis, Arnotts, Marloboro, Abby Street, Thomas Street, Christ Church, Setanta, Dawson, Trinity, Greenrcs, Drury and Brown Thomas. As mentioned above Abby Street stopped displaying information so this meant the decision was made to remove this car park from the list.

Parnell	Ilac	Jervis	Arnotts	Marlboro	Thomas	ChristC
185	694	403	350	156	173	74
229	769	417	359	157	186	73

Setanta	Dawson	Trinity	Greenrcs	Drury	BrownT	Date_Time
19	83	145	270	147	378	2019-03-19 14:07:35
146	118	165	1200	182	267	2019-03-19 15:45:02

Table 1: Example of car parking spaces dataset after cleaning in R.

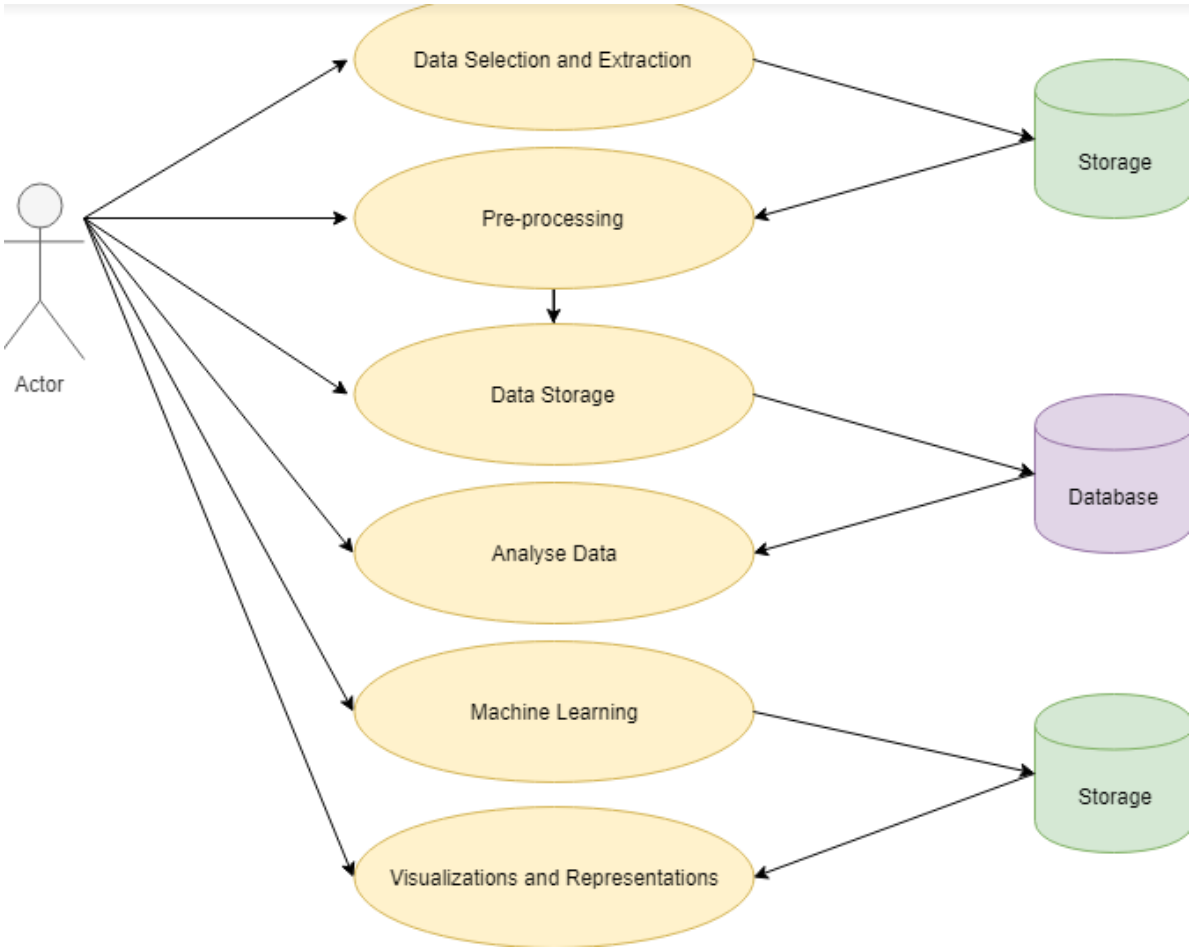
*Data on R is displayed across 14 columns (columns are divided in half for the report).

Dublin City Council website is used for a multitude of different reasons, for planners, people living in the Dublin City area and also the Dublin City Council workers would be on the website. The website does more than just car parking spaces it also looks at location elections, roads and traffic, culture and amenities and water, waste and environment. Having one dataset with a lot of information wasn't hard to manage.

Functional Requirements

The functional requirements that were used in my project was the KDD methodology. This is the methodology that follows the concept of Data Selection, Pre-processing, Data Storage, Data Mining and then finally Interpretation/Evaluations. This process was mentioned before in the report in the Methodology section

Use Case Diagrams



16

Use Case Priority Table

This table is a priority table, it sets a ranking and shows what each ranking determines and the meaning of each of the use cases. This is important to have as we can have an index to the importance to each of the use cases.

Priority 1	Considered a must and is vital to the overall scope to the project.
Priority 2	Considered a must as well as priority 1, but there are multiple solutions to the priority.
Priority 3	Not considered a must for the project but it does help the overall analysis.

Requirements Spec 1- Data Selection & Extraction

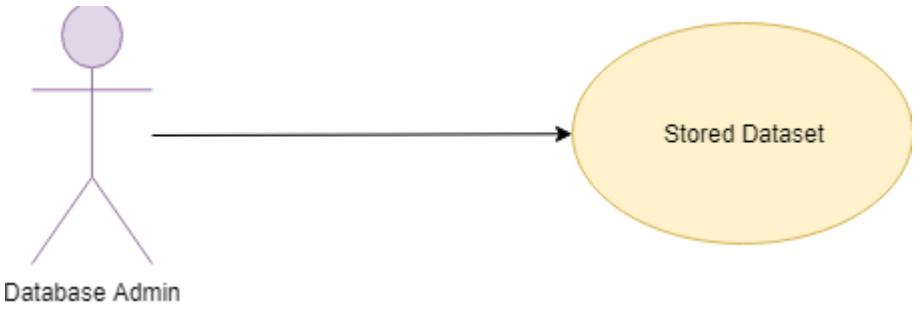
Data Selection is the first of the requirements and one of the most important, it is important to choose a data set that it interesting and unique. You need to have an interest or a knowledge of the area to get an idea of how to manage it and know what the aim of the project is. As it is so important to the function of the project this requirement would be classified as a priority 1.

Use Case

Scope

The data that we will be working with has been web scraped through a Python script and then this script has been running through AWS to run constantly, which gives us the data we need. Once this is completed for enough days the AWS script will have had a complete .csv file for me to work with.

Use Case Diagram



Flow Description

Pre-Condition

The user must be able to access the public website and then be able to code a Python script for the data to be updated and saves as a .csv file for the analysis of the data. User would need to have an idea of how to use a cloud service platform, for example Amazon Web Service. This would be needed to have the data running, if you had the Python script running through the CMD, it would be impossible to keep on top of the data as if the computer died it would mean that everything would be lost.

Activation

The use case happens with the Data Admin or the User retrieves the data that is required for the project.

Main Flow

1. The Data Admin/ User finds the website and analyses when the data is updated (every hour).
2. The DA then writes a script that is able to web scrape the data and save the updated car parking spaces every hour.
3. The DA then puts the Python script onto a cloud service platform, and makes sure it runs correctly.
4. Data is then stored within the AWS.

Exceptional Flow

1. The Dublin City Council website isn't working/isn't displaying car parking spaces.
2. Software isn't working or accessible.
3. AWS is down.
4. Dataset becomes corrupted.
5. Script isn't working correctly.
6. Software isn't up-to date.

Termination

When the process of gathering and storing the data is complete and correct. Therefore once final dataset is collected the AWS instance can be stopped.

18

Requirements Spec 2 -Pre-Processing

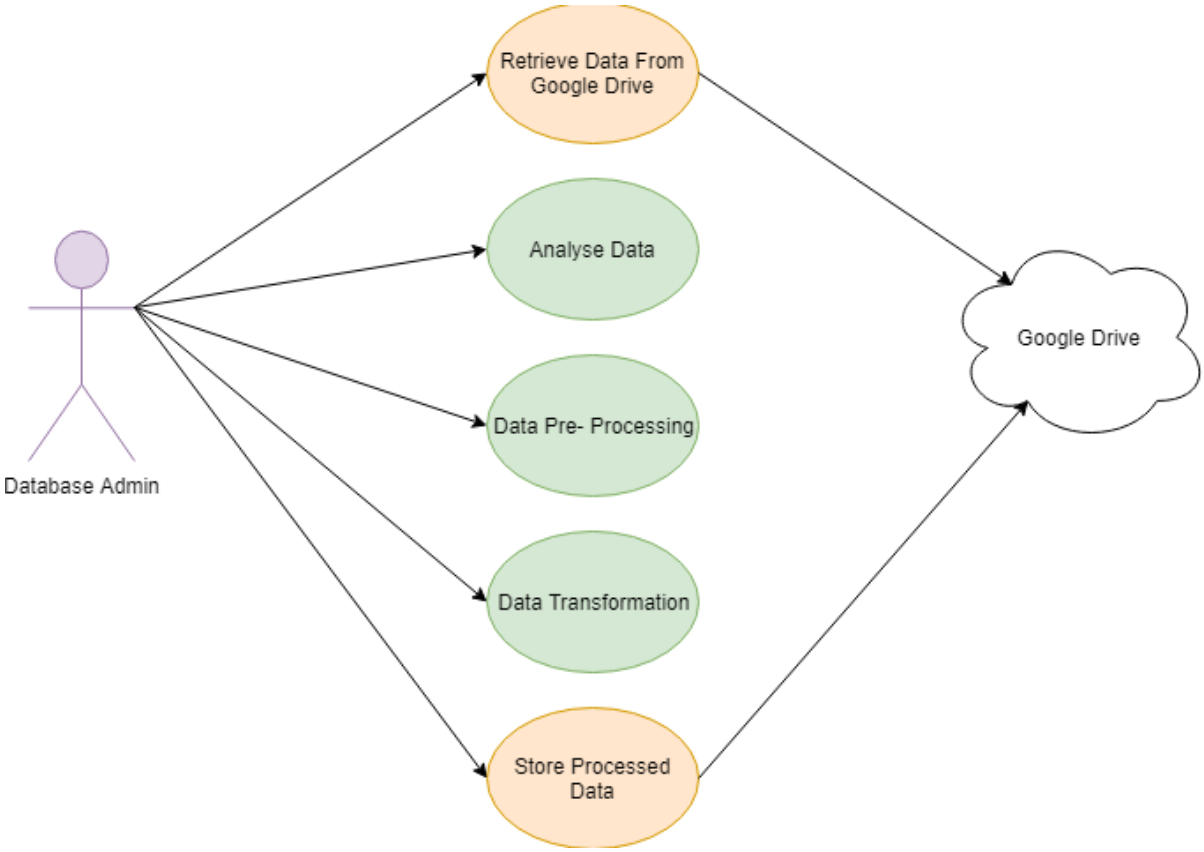
The overall pre-processing requirements means that we can now start augmenting the dataset, which allows us to complete out our analysis in the next few stages of the project. This section would be rated a priority 1 as without such an important part of the KDD process we would have data extremely off and multiple outliers. This would mean that the graphs and tables of the data that we will be looking at in the next part of the project will not be correct.

Use Case

Scope

The pre-processed data allows us to have the correct data to be able to conduct a correct analysis of the data.

Use Case Diagram



Flow Description

Pre-Condition

The data must be accessible to the users in able for the per-processing to take place.

Activation

The use case can begin with the Data Admin retrieves the data from AWS which is required to start the analysis.

Main Flow

1. The Data Admin gathers a local copy of the data which has been web scraped.
2. Data is then examined to make sure it is all correct and notes taken on what needs to be cleaned.
3. Cleaning can commence on the data.
4. Pre-processing then starts on the data.
5. Data starts to now become transformed.
6. Processed data is now updated and stored safely.

Exceptional Flow

If the data wouldn't be retrievable.

Termination

Our process of pre-processing the data is complete, now we have an organised dataset which is ready to be analysed for the project. This means termination of the process can begin.

Requirements Spec 3 –Data Storage

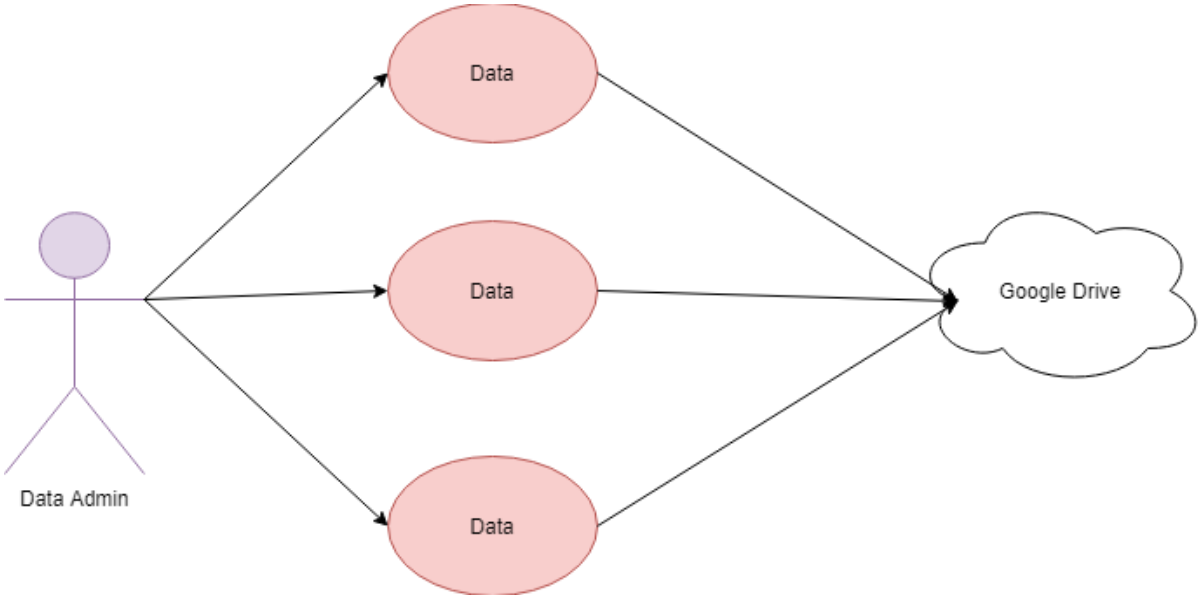
Data Storage is extremely important as with the data stored in a safe and easy to locate place it means it is easier to access. This is also the case if something goes wrong and if you lose the data on your local machine, it is very important to have backups. As it is important to have these backups it is rated priority 3 on our list as it is not completely necessary.

Use Case

Scope

The Data Storage develops a dataset with specific tables including rows and columns for where the data can occur.

Use Case Diagram



Flow Description

Pre-Condition

The data would have to pre- processed as well as being transformed to the correct format to be stored correctly.

Activation

The activation can commence when the Data Admin has made the correct data available.

Main Flow

1. The Data Admin has received the data.
2. The DA has then created a copy of the data set and has correctly saved it in the right format.
3. The data is then saved onto a cloud server.

Exceptional Flow

If the data cannot be retrieved at all.

Termination

This can occur once the project is completed and unnecessary to the course any more. This would be after the results.

21

Requirements Spec 4 –Analyse Data

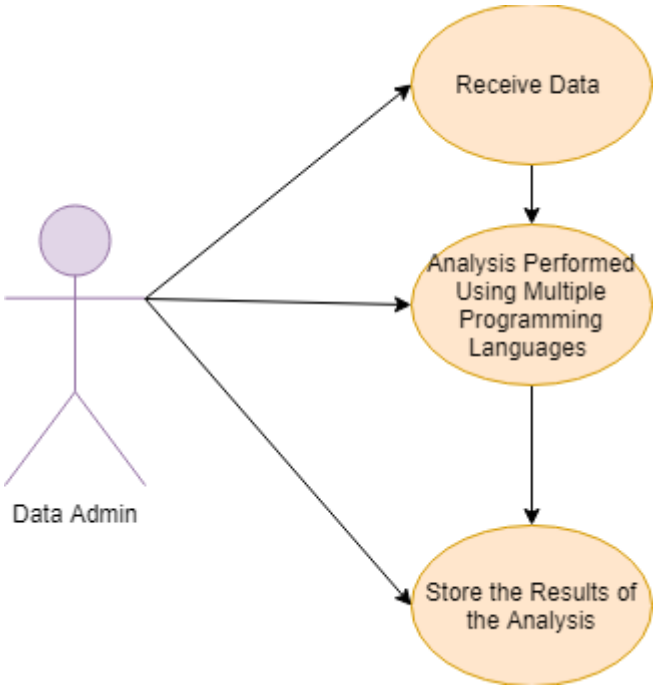
Analysing the data is the fundamental requirement for this project and is most certainly ranked at a priority 1. Analysing the data is extremely important for this project, without the analysis we wouldn't be able to answer the questions or tackle the aims of the project. Completing the analysis of the dataset needs a lot of attention and a massive amount of detail to help with achieving the project goals.

Use Case

Scope

Data is retrieved from the data storage and put into the chosen software to perform a clear analysis of the data. The evaluation and interpretation of the results will occur and once this is complete the results are then saved onto the local device and also onto the cloud in case of anything going wrong.

Use Case Diagram



Flow Description

Pre-Condition

The data must be saved onto the local device to it is accessible to the data analyst.

Activation

The use case starts when the data analyst calls the data from the local device.

Main Flow

1. The Data Analyst receives the data from the cloud server.
2. The data is then downloaded onto the local device.
3. Data is then analysed using certain programming languages.
4. The results start to come through once the code has been wrote on the certain programming languages.
5. The results are then saved to a save location.

Exceptional Flow

Data wasn't able to be downloaded from the cloud server.

Termination

This use case is finished when the analysis is completely finished.

Post Condition

The data is waiting for more analysis.

Requirements Spec 5 –Machine learning

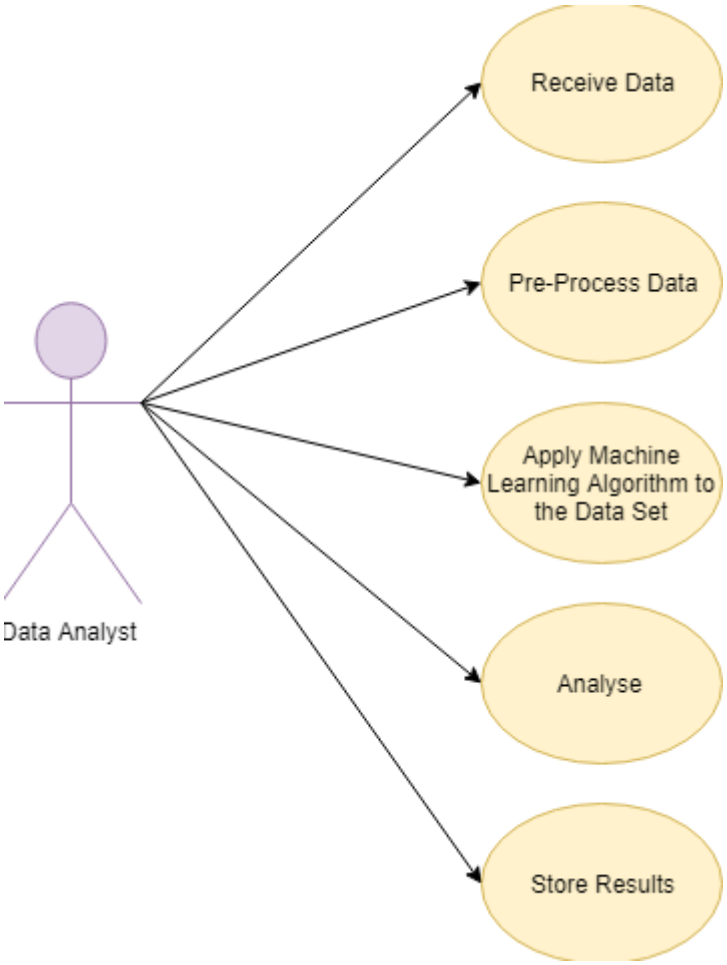
This requirement is an exploratory requirement and I would rank it as a priority 2. Rated priority 2 is because it is not as necessary as the overall analysis on the project. The dataset will be running through machine learning algorithms using R Studio and Tableau to discover patterns that are interesting and unexpected.

Use Case

Scope

Datasets are retrieved and pre-processed. Once this is all complete the data then runs through the machine learning algorithms and then when satisfied with the results then I will start an analysis and then results will be stored safely.

Use Case Diagram



Flow Description

Pre-Condition

Data is able to be retrieved from a source.

Activation

The use case starts when the Data Analyst has received the data.

Main Flow

1. The Data Analyst receives the data.
2. The Data Analyst pre-processes the data.
3. The data runs through the machine learning algorithms code.
4. Results are then analysed.
5. Results are then stored in a safe environment.

Exceptional Flow

The code is incorrect, which you bring up errors for the Data Analyst.

Termination

This use case is completed once the results have been stored.

Requirements Spec 6 –Visualizations and Representations

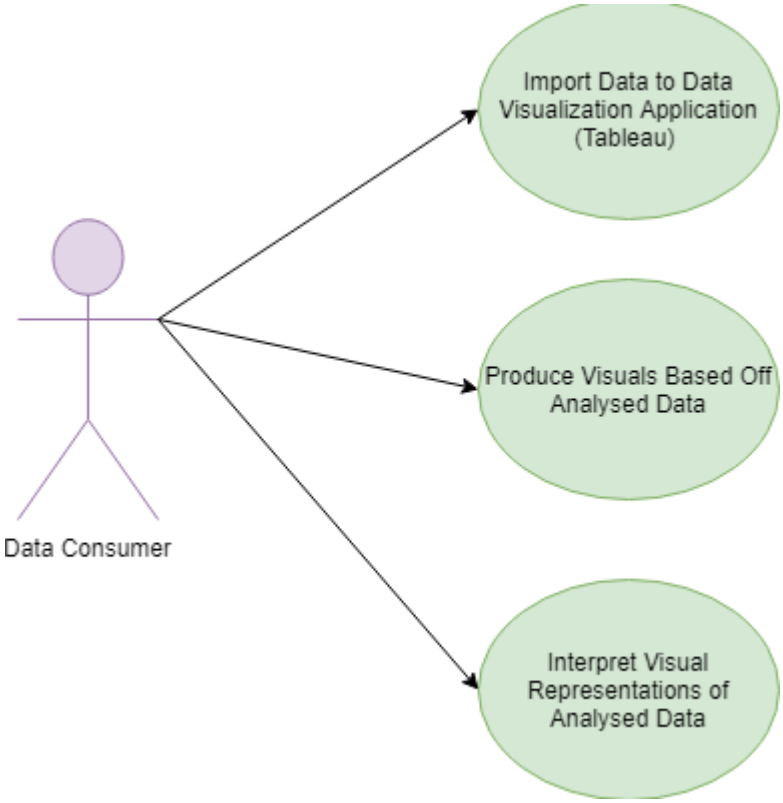
This is the final requirement for this part of the project. Visualizations and Representations are ranked at a priority 2 as it is important to show the data in a clear and concise manner.

Use Case

Scope

Import the analysed data into a data visualization application. Once the Visualizations representations have been produced then help the data consumer to interpret the results of the analysis.

Use Case Diagram



Flow Description

Pre-Condition

All analysis of the dataset must be complete at this stage of the requirements as it is necessary to create a final visualization and representation of the overall data that has been looked at throughout the project.

Activation

This is then activated once the data admin has the final load of data.

Main Flow

1. The analysed data is imported to a data visualization application.
2. This application then creates a final visualization and representation of the results.
3. Visualizations and representations are available to look at for the target user.

Exceptional Flow

Importing the data to the application isn't possible, or the application needs an update or needs to be installed.

Termination

The use case is completely finished when all the visualizations and representations are completed.

Non-Functional Requirements

Security Requirement

Having this data, although it is open source it is important to be careful with it. The likes of GDPR wouldn't come into it as it is only car parking spaces and not important data with people personal details on it. I believe that all data should be dealt with a similar level of security. My AWS was encrypted with keys and the Filezilla had usernames and passwords so they were protected to a high standard.

Availability Requirement

The data is still available on the Dublin City Council, although it is impossible to receive data from the past on the website. The data is updated every hour and the past space history is deleted.

Integrity Requirement

The dataset contains all the necessary attributes and values in order for the correct analysis to be performed and analysed. This will result in assuming accurate outcomes and will answer questions about the data.

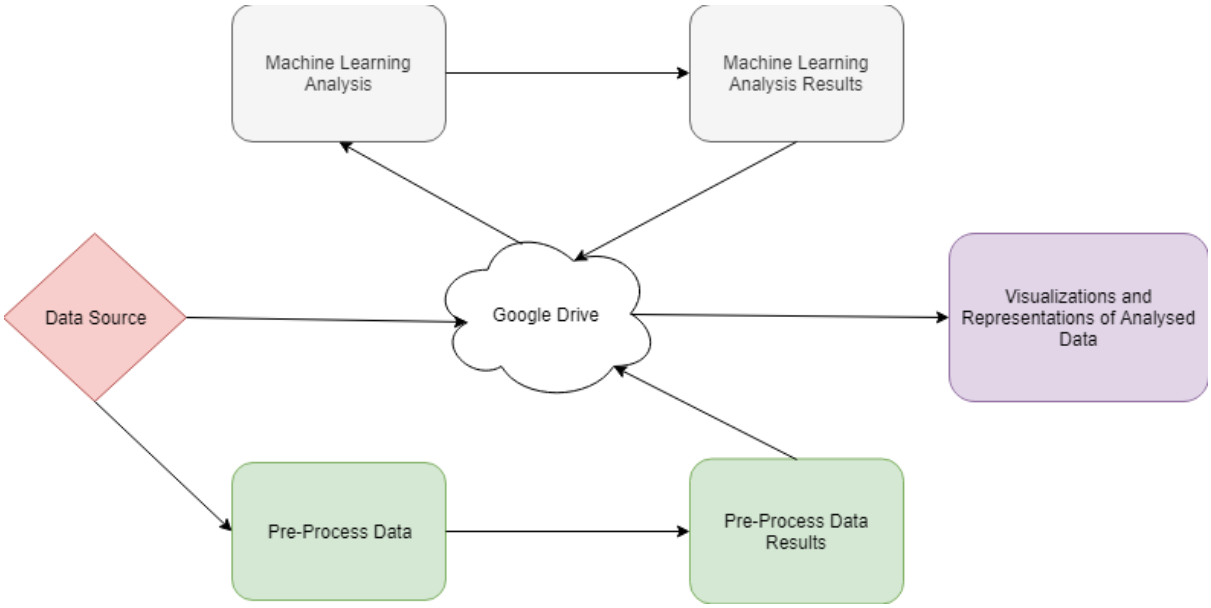
User Requirement

The user requirement is that in the future when city planners are looking at Dublin City they will be able to quickly analyse the data and receive answers to these questions. These questions could be in terms of where to put new car parking spaces in the area, if some car parking spaces on the South side are still available from 9am till 6pm compared to the North side which is all full up would it be more practical to build a car park on the North side. This is only an example of the many questions this data could answer. The overall requirement is that cars in general are used more and more yet there is a constant debate about global warming and the need for more public transport. Seeing how many car parking spaces that are taken up on the daily is a very interesting topic and can lead to future discussion.

Design and Architecture

The diagram below shows the system architecture that has been completed during this project. The storage has been the heart of the system as without it the data and results wouldn't be housed properly. As you can see from the diagram that when the data is retrieved and stored, an overall analysis can begin on the data. This analysis is through R, R Studio and Tableau. Using these three programmes I feel like I achieved the correct about of graphs and important detail needed to make this project the best that it could be. Learning and investigating from the data has made me aware of the parking in Dublin a lot more. From seeing different areas in town being busier than others or that at 9am the spaces are mostly taken up till around 6pm, when rush hour is

beginning to end, there is loads of spaces left. Interesting but not surprising. Using the applications to visualise the graphs and tables is very important for the end user as they might not have a computer background, this is why the graphs need to be displayed in a certain way so people can understand clearly.



Implementation

Low level system Architecture

This is where the outline of the significant analysis conducted throughout the project lifespan. All of the R code used for the analysis which was implemented in R Studio is attached with the submission. Next would be to look at the horizontal low level view of the system architecture.

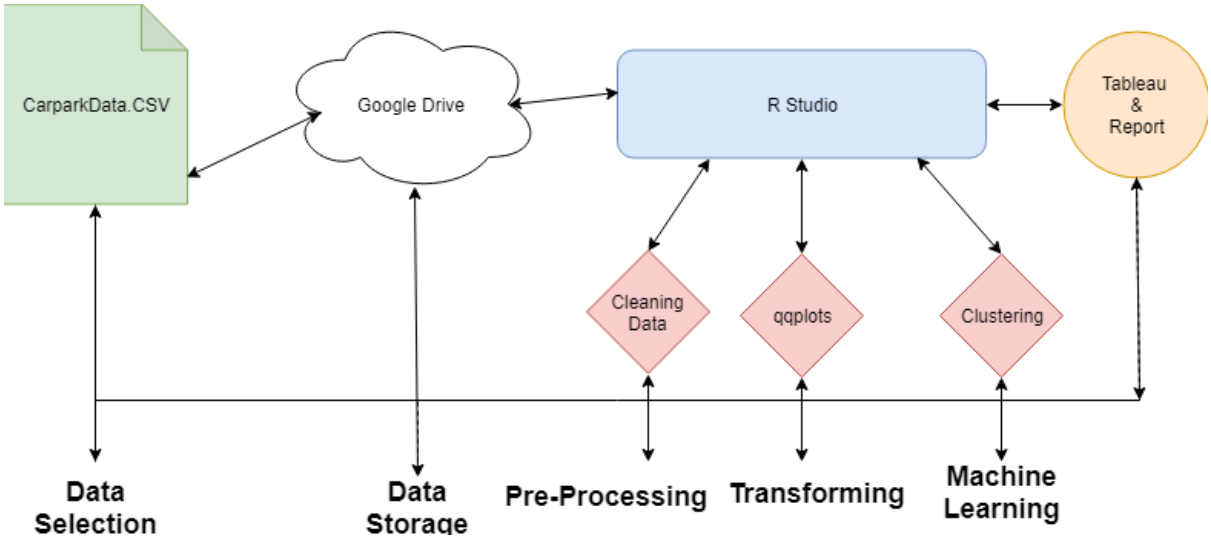


Figure1: The Display of the Low – Level View of System Architecture.

This low level view of the system architecture as shown above. The system architecture displays the workflow during the project. The KDD methodology was a very important base on where to start with the project. It helped with planning and moving forward with the project when times had got hard. The graph above shows a clear indicator of all the different stages of the project which included obtaining the data, storing the data, analysing and preparing the data and then finally storing the data.

Important R Functions Used

colnames()	qqline()
duplicated()	Bgtest
mean()	box.test()
summary()	boxplot()
ggplot()	adf.test()
geom_jitter()	adf()
geom_smooth	pacf()
geom_point()	cor()
hist()	predict()
qqnorm()	cro()

Important R Packages Used

Tidyverse	Forecast
Car	Magrittr
Lmtest	qwraps2
Tseries	Expss

Initial Analysis

The dataset that shows all the car parking spaces in Dublin City Centre was web scraped from Dublin City Council website which is open source, which made the overall experience of web scraping from the website was a lot easier. Once I wrote the Python script to web scrape the information I was able to collect the data into a .CSV file which displayed the data across 15 columns, one which included the date and time.

Test2.py | Test4.py | Test3.py | Script2.py

```
import csv
import requests
from bs4 import BeautifulSoup
import time
from datetime import datetime

finalTime = 4665600
currentTime = 0
sleepTime = 3600

while currentTime < finalTime:
    # Collect data
    url = 'http://www.dublincity.ie/dublintraffic/cpdata.xml?1543254514266'

    res = requests.get(url)
    soup = BeautifulSoup(res.content,"xml")
    data = []
    for item in soup.select("carpark"):
        ditem = {}
        ditem['Name'] = item.get("name")
        ditem['Spaces'] = item.get("spaces")
        data.append(ditem)

    collectionTime= datetime.now().strftime('%Y-%m-%d %H:%M:%S')
    mylist= [d['Spaces'] for d in data]
    mylist.append(collectionTime)
    print(mylist)

    with open("CarparkData1.csv","a",newline="") as f:
        writer = csv.writer(f)
        writer.writerow(mylist)


    # wait 1 hour
    time.sleep(sleepTime)

    # update current time
    currentTime = currentTime + sleepTime
```

Figure 2 –Python Script for the web scraping of the website.

Next was to delete one of the columns (Abby Street) as the website still had the car park up on the website but no data what so ever. This meant the data set now had 14 columns.

Filter

	Parnell	Ilac	Jervis	Arnotts	Malboro	Thomas	ChristC	Setanta	Dawson	Trinity	G
1	188	665	421	340	160	173	77	21	70	142	
2	183	683	408	344	160	173	76	19	87	143	

Figure 3 – A sample of the data set on R.

The AWS to look at was running correctly, or so was to be believed, when it came to retrieving the final dataset, no data had been collected to my complete dismay. This meant that unfortunately was left with less data then originally planned. It meant the data set consisted of 5 days of data from May.

Figure 4 –What AWS displayed, no issues to the eye (this project is the second instance).

When running the data through Filezilla, in the status box it would come out as an Error. Once rebooting the AWS it ran smoothly as you can see below that the right hand side has what is exactly on the AWS instance. Filezilla was used to show and easily access the AWS instance, otherwise it would be impossible to access the EC2 instance.

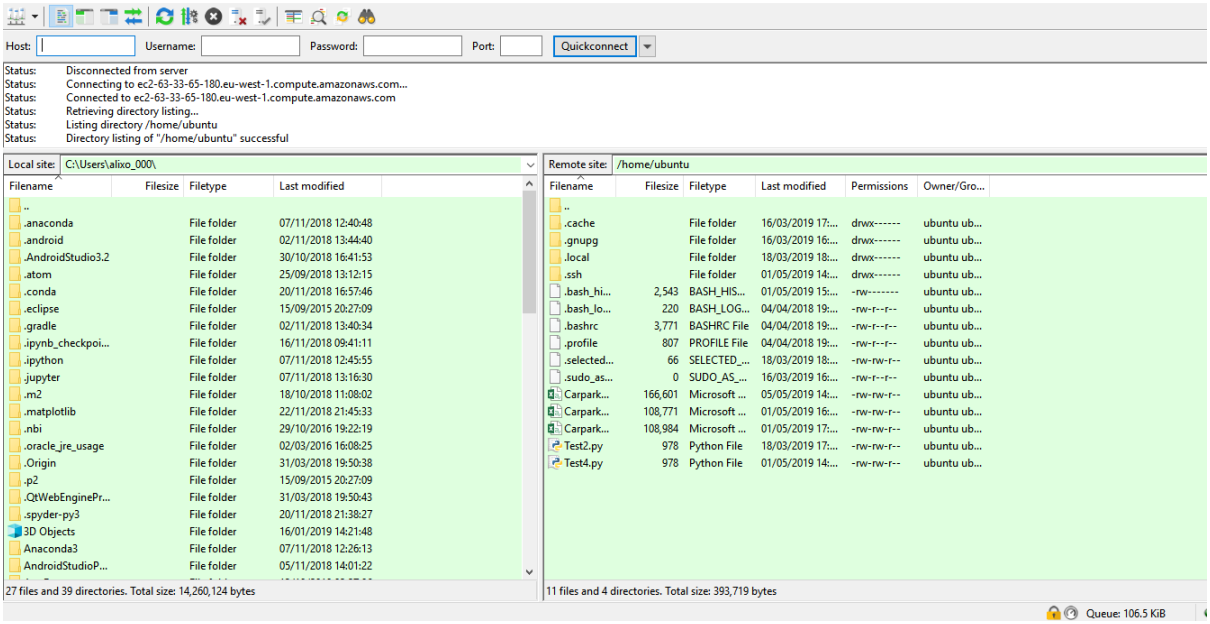


Figure 5- What Filezilla looks like when it runs successfully, left hand side displays what is running on the AWS.

Carpark with most spaces taken up (Shopping Centres Vs Non Shopping Centres).

This is the first of the questions that was to be looked at. Analyse the car parks to see which shopping car parking facility has the most spaces taken up on any given day as opposed to just regular car parks. The overall assumptions was that the shopping centres for example Brown Thomas or Jervis would be more busy then say the likes of Trinity car park, especially on a weekend.

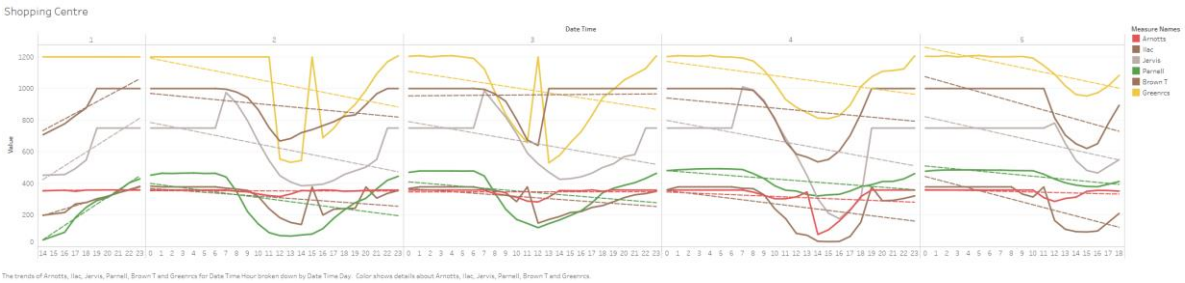


Figure 6- Tableau of car park data of the shopping car parking centres.

As mentioned before the Greenrcs car park spaces is extremely high, although web scraped from the Dublin City Centre website it doesn't look 100% correct. Second highest with the free spaces is the Ilac centre. Stevens Green (Greenrcs) car park and Ilac have the highest amount of spaces. So it is not surprising that they would have the freest spaces available. Arnotts, Brown Thomas and Parnell are competing to be the busiest car parks and most interestingly that Saturday the 4th of May has the least amount of spaces free, this also leads into our other question in regards spaces free on the weekend compared to the week-day.

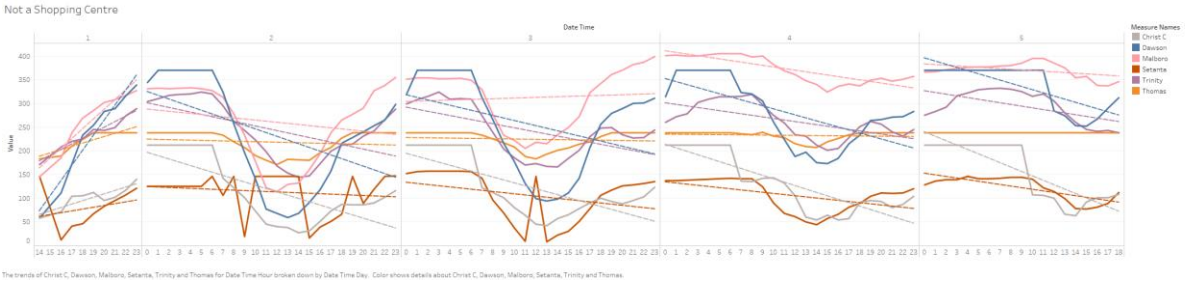


Figure 7- Tableau of car park data of the car parking centres that are not attached to shops.

From analysing the data, first thing to note is that parking spaces that are not in shopping centres are considerably lower than in shopping centres. With the highest being Marlboro at around 400 spaces. The two car parks that are full the most are Setanta and Christchurch. Setanta is always quite full, with being in the centre of the South-side could be a huge indicator.

Car parks in shopping centres are much bigger and more convenient to customers who possibly need to do a quick shop, as in most of the shops that these car parks are connected to have nearly everything you could imagine. Another thing to note is that the car parks that were not connected to shops are fuller during the week-day. Which is interesting but not too surprising.

32

Area with the most spaces full

This is the second question that was to answer, I wanted to see which side of the city was the busiest, and the reasoning behind it. There are a lot of companies particularly on the South-side, and going down toward Grand Canal, so personally it wouldn't be a huge shock to myself if there was less spaces free on the South-side, particularly during the average working week and hours.

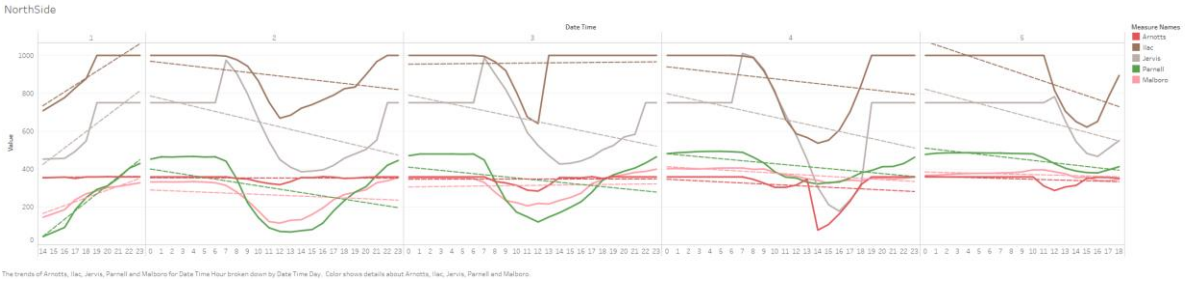


Figure 8- Tableau of car park data of parking spaces on the North-side of Dublin City Centre.

Although hard to read the North-side of Dublin City Centre looks like it has more movement in the graph. Ilac car park clearly has the most spaces available at any given time. Parnell having the least amount of spaces free (on average). Parnell is a small car park with only 500 spaces, Ilac on the other hand has 1000. This would also be another factor as to why Ilac is so high on the graph. If going into town and parking on the north side was needed, Ilac would be the best to park in as there is more chance of a parking space. With Ilac charging 30 cent less an hour then Parnell, this is also another interesting reason to park in Ilac.

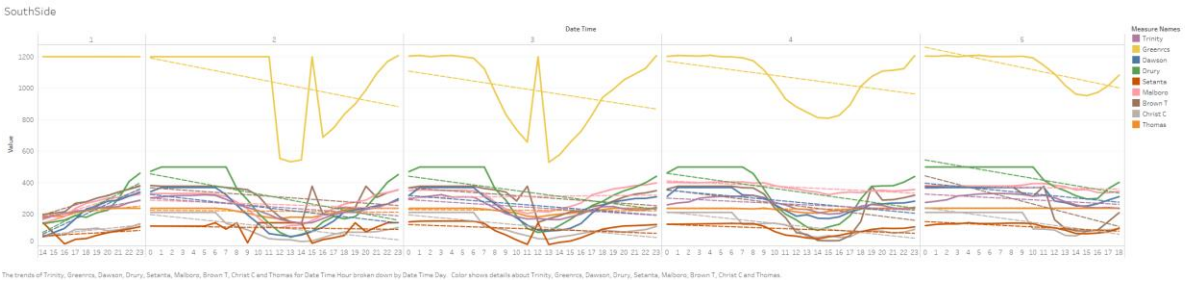


Figure 9- Tableau of car park data of parking spaces on the South-side of Dublin City.

After analysis of the South-side car parking spaces, it was surprising to see so many spaces available in Greenrcs. Dublin City Council website had the same figures as the data set. Stephens Green shopping centre does have 1200 spaces available, this is a massive number compared to the other car parking spaces on the South-side and 200 more than the highest on the north side (Ilac).

From analysis of both of the plots its does show that, yes overall there is more parking on the south side, there is more activity in and out of the spaces. Looking into the North-side Arnotts hardly moves much but interestingly enough both Parnell and Marlboro have a constant wave per day of people. This could indicate that people park there on a regular basis. The South-side however has less spaces available over all and Setanta and Christchurch have the least amount of spaces available. Thomas Street stays at a constant.

Overall it depends where you are going in the city centre that influences available parking. Parking on the North-side when you have a meeting on the South-side makes little to no sense. Ilac is the car park with the most spaces and best value, coming in at €3.20 per hour on the North-side and Stephens Green is the best for spaces and is a reasonable price at €3.80 per hour.

Comparing working days with the weekend, same time.

Looking at these two sets of data was interesting and personally though it would give an idea of who uses the spaces for work, would more people be inclined to bring the car into work rather than to bring the car into town for shopping. This could also indicate people who need to bring their car into town as they live further then the transport belt.

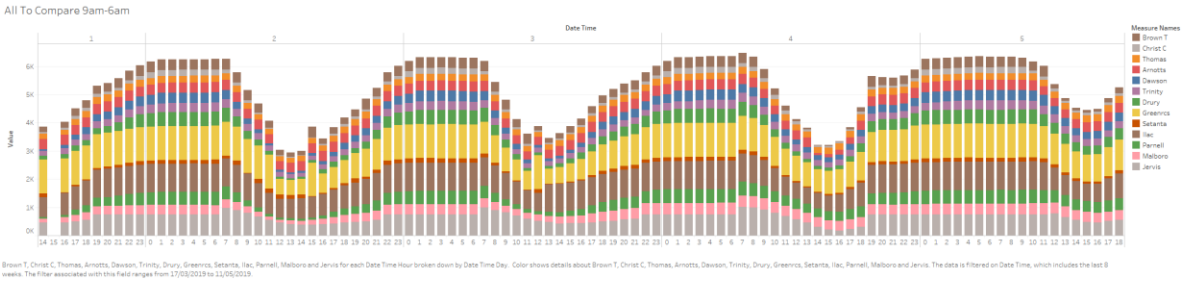


Figure 10- Tableau of all the car parks and the various times and days.

From even just glancing at the graph it is very clear that each wave is bang on 8am till 6pm, like clockwork it's interesting to see. This is to be expected with 61.4% of people going to work drive. What was interesting to see that at 10am (a little later then the week-day) on Saturday that the dip occurs, something to note that this weekend is a bank holiday and also stores in town had a lot of sales happening, so it was worth people coming in. Sunday wasn't too busy with the car parks which was interesting but then again it is a bank holiday, it would have been interesting to see if the bank holiday Monday would look like compared to another average working Monday.

34

Analysing Boxplots and Histograms

Conducting a histogram on the density of the car parking spaces was an effective graphical method which showed off the distribution of data over the period of time that the data was for. As we can see from the histogram below, it is clear that Parnell Car Park has the mostly 500 car parking spaces free. This isn't too surprising as there is 500 parking spaces in Parnell and it would make sense as most people don't park their cars overnight. But we can also see that the car park isn't normally completely full. Zero would indicate that there is no car parking spaces available.

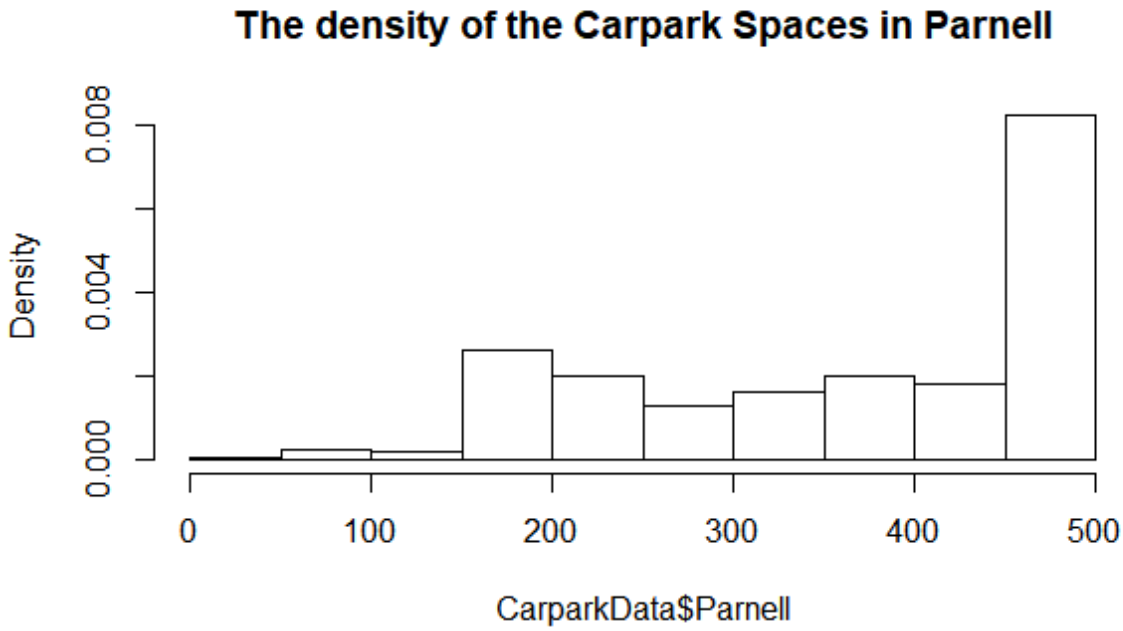


Figure 11- Histogram of the density of Parnell car park overall.

After analysing all of the car parks, Setanta came up with the facts and figures that Setanta is the car park that is Full the most often, compared to the other car parks. Setanta only has 150 spaces, so this could be a huge factor as to why it's mostly full.

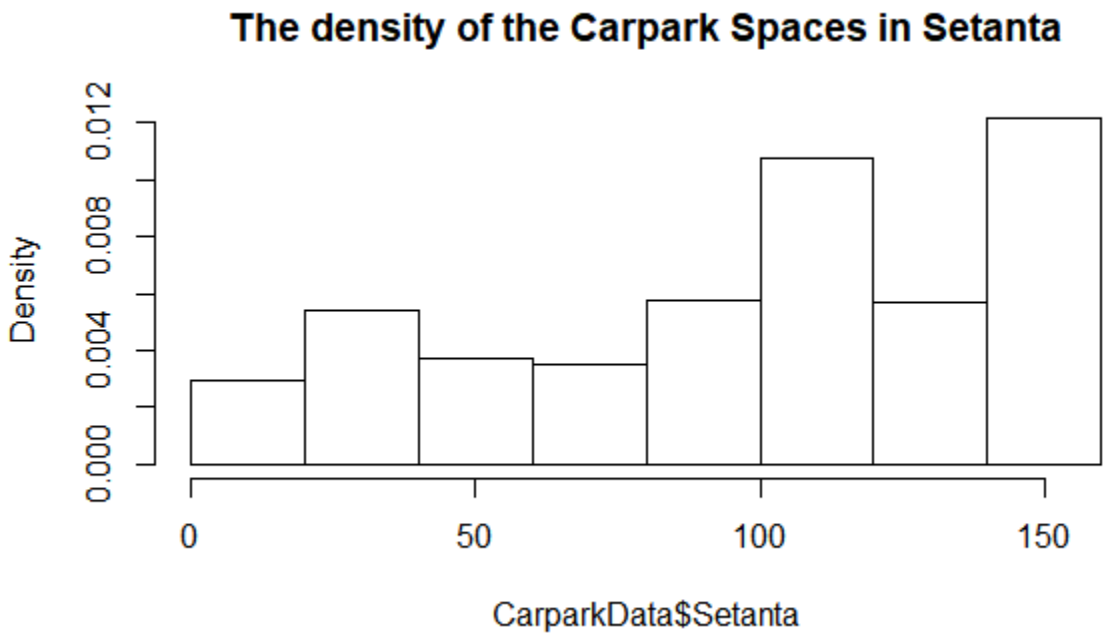


Figure 12- Histogram of the density of the Setanta car park overall.

For the likes of showing where most of the data lie, boxplots would be an effective graph to use. Viewing the upper and lower quartiles and the interquartile range and using another line of code, being able to view the median with a dotted horizontal line. These two different type of graphs were conducted on R. Box Plots are also ideal to visually show any outliers that can occur in the data. Surprisingly analysing the data I decided to analyse Setanta as discovered above we can see that it is the car park that is full the most. Expecting to see a lot of outliers towards the bottom of the data surprising there was none. As the car park is full a lot of the time, it would make the overall box shape towards the middle, which is what happened as pictured below.

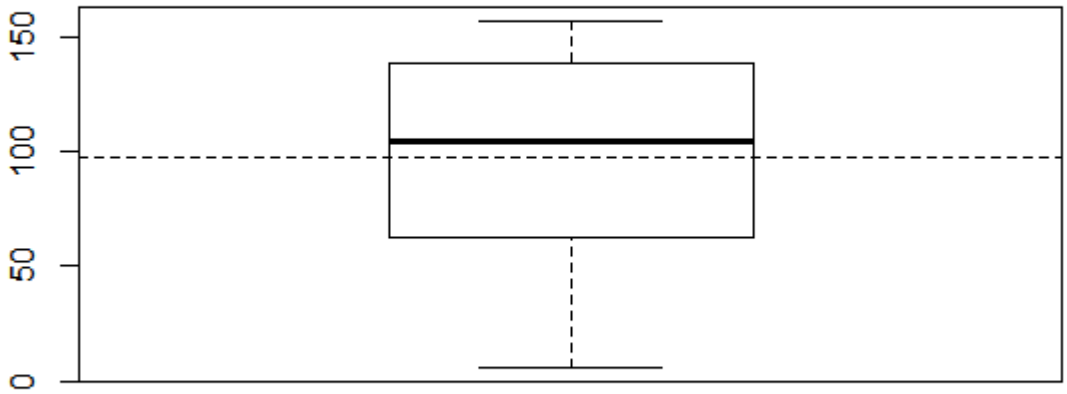


Figure 13- Boxplot of the Setanta car park including mean line.

Moving on through the analysis it came to my attention of Arnotts car park. Arnotts is hardly ever full. So automatically the box plot was going to be towards the top of the graph. The mean line is also much higher than Setanta. However it was a shock to see exactly how many outliers there was exactly.

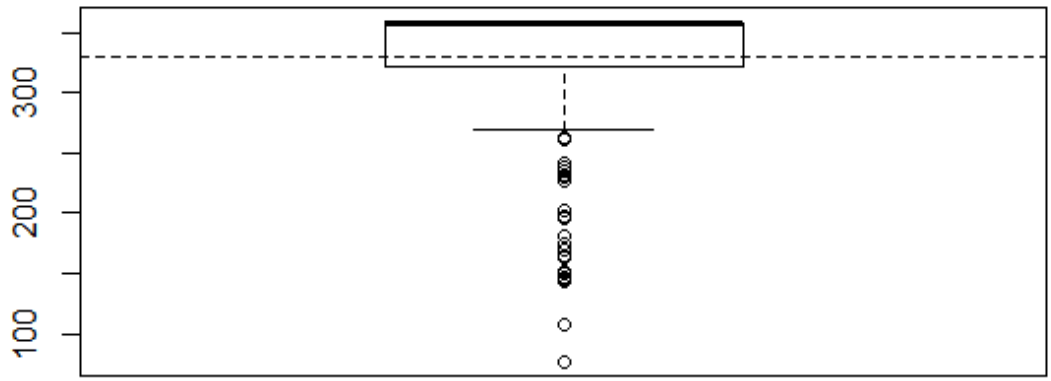


Figure 14- Boxplot of the Arnotts car park, including mean line and outliers.

37

Analysing Q-Q plots

Conducting an analysis using Q-Q plots (Quantile- Quantile plots) show us two quantiles against one another. Essentially a quantile is a fraction where a certain value can fall below. There is a line on the Q-Q plot and the closer the Q-Q plot is to the line the more normal the data is. This analysis was conducted on R Studio.

Trinity car park was the data that was closest to the line, as shown below. Trinity was so close to the line which really shows us that is a normal plot. Comparing the Sample Quantiles and the Theoretical Quantiles gives the best result.

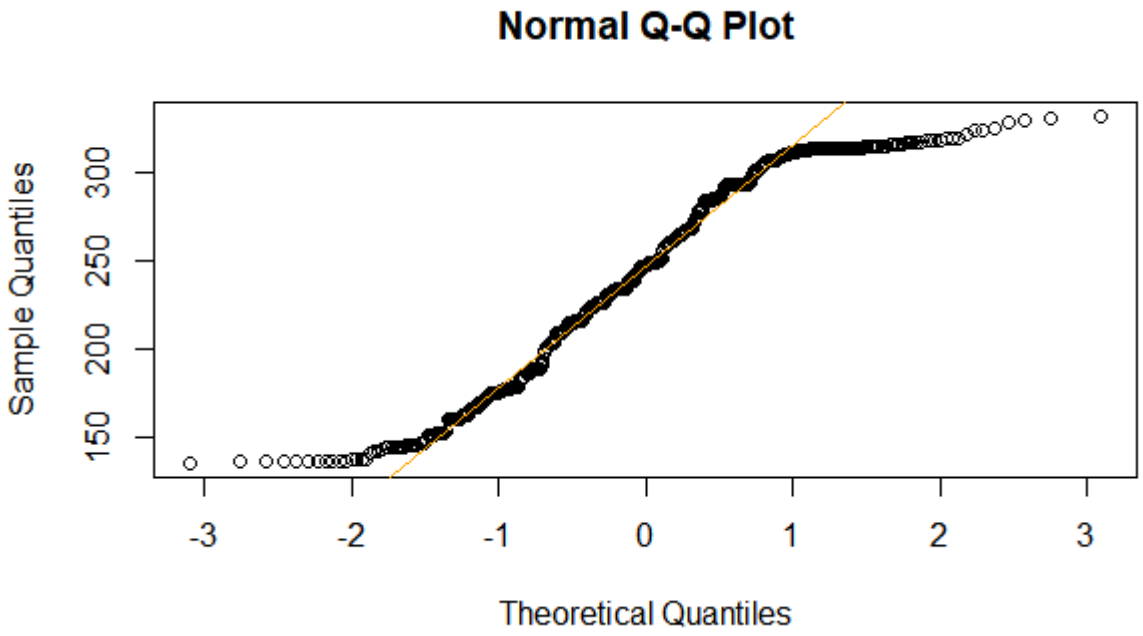


Figure 15- Q-Q plot of Trinity car park, note how close to the line the graph is.

Analysing Arnotts car park using the Q-Q plot, gives a very clear indication that it is not as normal, as it only hits the line twice and most of the points are nowhere near the line. Clear indication that the above box plot is correct.

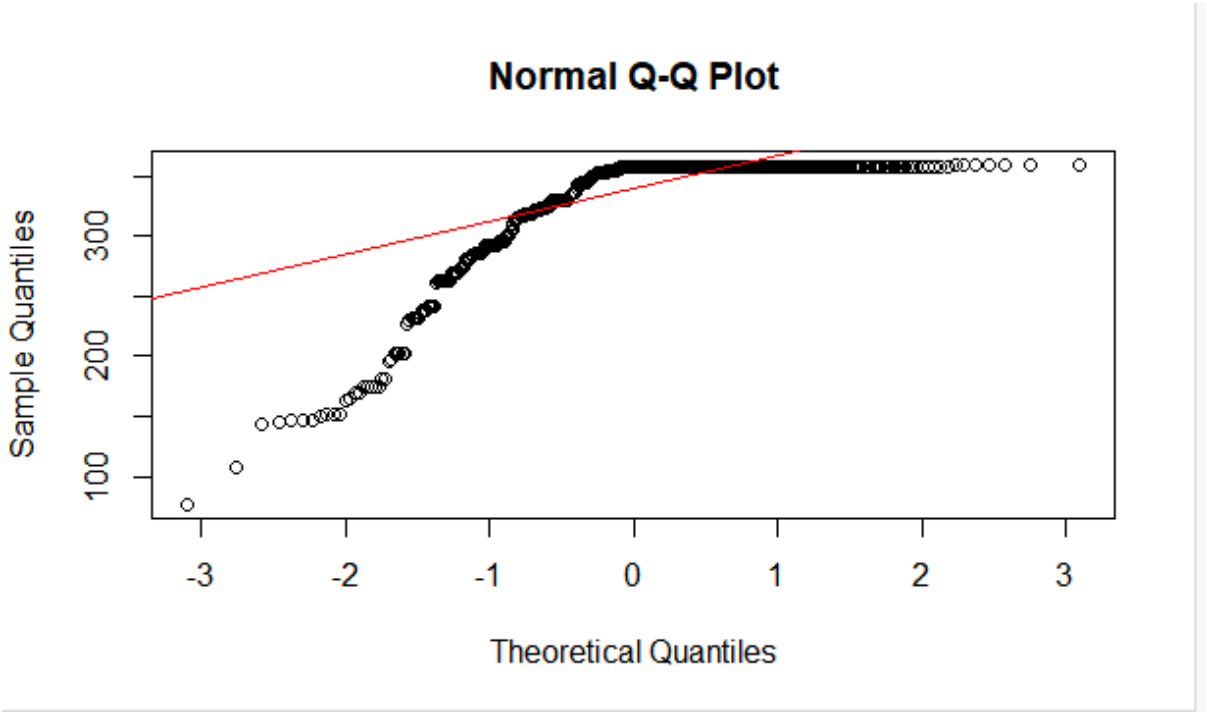


Figure 16- Q-Q Plot of Arnotts car park, note how far it is from the line.

As the above Q-Q plot isn't normal, a log.malic test would be ideal. This just double checks if the data is correct.

Multiple Regression

Multiple Regression analysis helps to understand how a dependent variable could change when an independent variable is changed. Multiple Regression can predict trends and future values and can be used to get estimates. This was carried out on R Studio. The multiple regression tests that was used on the data set was, Breusch-Godfrey test, Box-Ljung test and the Durbin-Watson test. The following test and results are shown below.

```
> bgtest(model)

        Breusch-Godfrey test for serial correlation of order up to 1

data:  model
LM test = 408.51, df = 1, p-value < 2.2e-16
```

Figure 17- Image of Breusch-Godfrey Test carried out.

```
> Box.test(model$residuals, type="Ljung-Box", fitdf=1)

        Box-Ljung test

data:  model$residuals
X-squared = 391.8, df = 0, p-value < 2.2e-16
```

Figure 18- Image of Box-Ljung Test carried out.

```
> dwtest(model)

        Durbin-Watson test

data:  model
DW = 0.19934, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0
```

Figure 19- Image of Durbin-Watson Test carried out.

Testing for stationarity

Stationarity testing is used mostly for investigating time series, it is used before building an ARIMA model. Essentially it is used to check if the series is stationarity, if it is then the data would be suited to perform a seasonality or trend. There are two tests at needed to be carried out for stationarity, Augmented Dickey-Fuller Test and KPSS Test for Level Stationarity. Both of these tests were carried out on R Studio and the results are as followed. Tests were both carried out on Parnell and Thomas Street.

```
> adf.test(CarparkData$Parnell)

        Augmented Dickey-Fuller Test

data:  CarparkData$Parnell
Dickey-Fuller = -3.7114, Lag order = 7, p-value = 0.02343
alternative hypothesis: stationary
```

Figure 20- Image of Augmented Dickey-Fuller Test carried out on Parnell car park.

```
> kpss.test(CarparkData$Parnell)

        KPSS Test for Level Stationarity

data:  CarparkData$Parnell
KPSS Level = 0.16798, Truncation lag parameter = 6, p-value = 0.1
```

Figure 21- Image of KPSS Test for Level Stationarity carried out on Parnell car park.

```
> adf.test(CarparkData$Thomas)

        Augmented Dickey-Fuller Test

data:  CarparkData$Thomas
Dickey-Fuller = -3.1912, Lag order = 7, p-value = 0.08945
alternative hypothesis: stationary
```

Figure 22- Image of Augmented Dickey-Fuller Test carried out on Thomas Street car park.

```
> kpss.test(CarparkData$Thomas)

        KPSS Test for Level Stationarity

data:  CarparkData$Thomas
KPSS Level = 0.19793, Truncation lag parameter = 6, p-value = 0.1
```

Figure 23- Image of KPSS Test for Level Stationarity carried out on Thomas car park.

From the analysis of the two test the next is to complete auto corrections and now to choose the correct model for the data. Using the Autocorrelation Function (ACF), which refers to how correlated a data set is and Partial Autocorrelation Function (PACF) function it retrieves two different graphs. The graphs below is all the data in the dataset, these functions were also analysed on each car park.

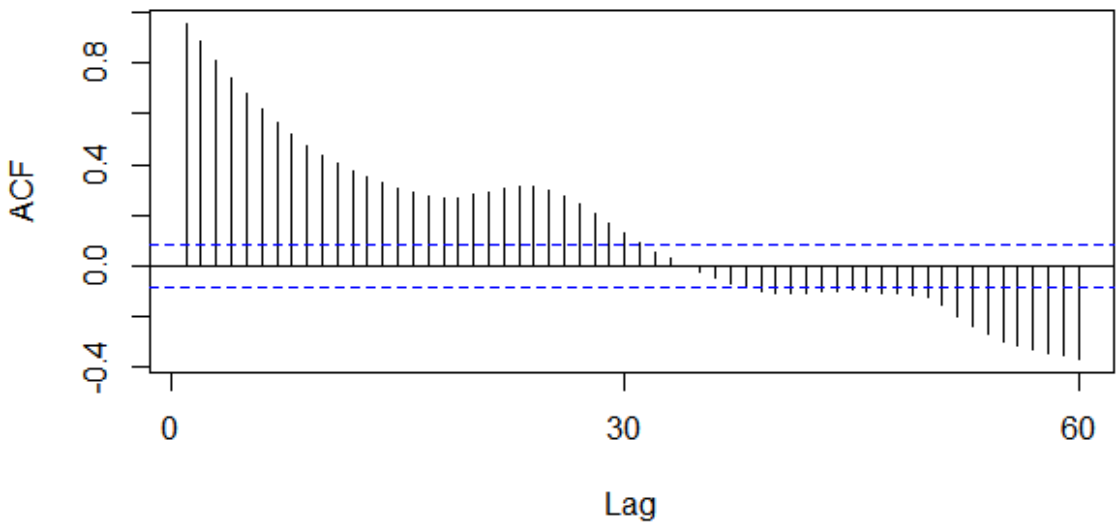


Figure 24- Autocorrelation Function for the whole dataset.

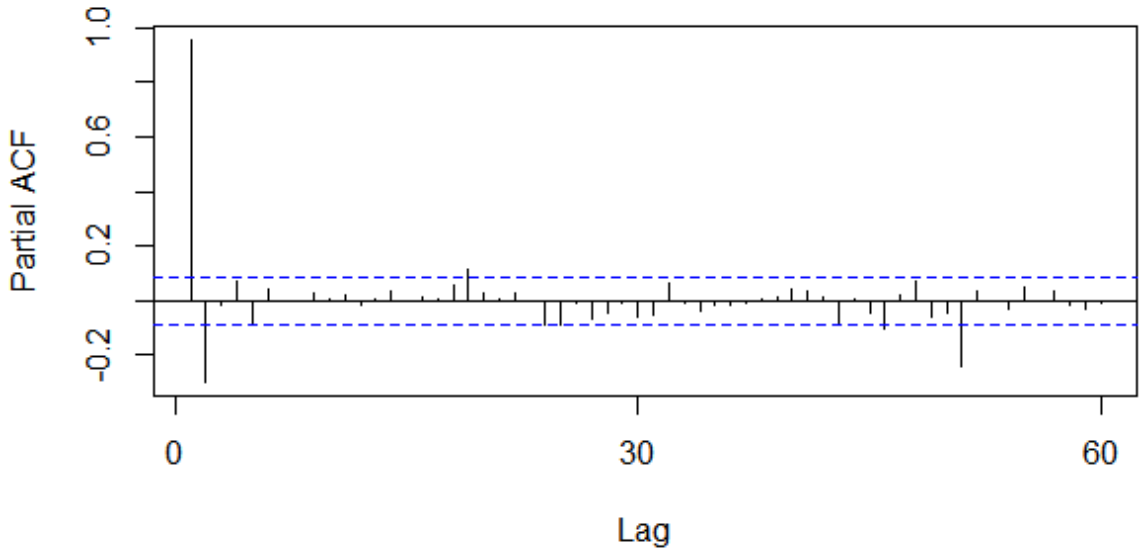


Figure 25- Partial Autocorrelation Function for the whole dataset.

Next the best thing to do would be to analyse the decomposing data in each of the car parks, below is an example of Jervis car park and the decomposing data.

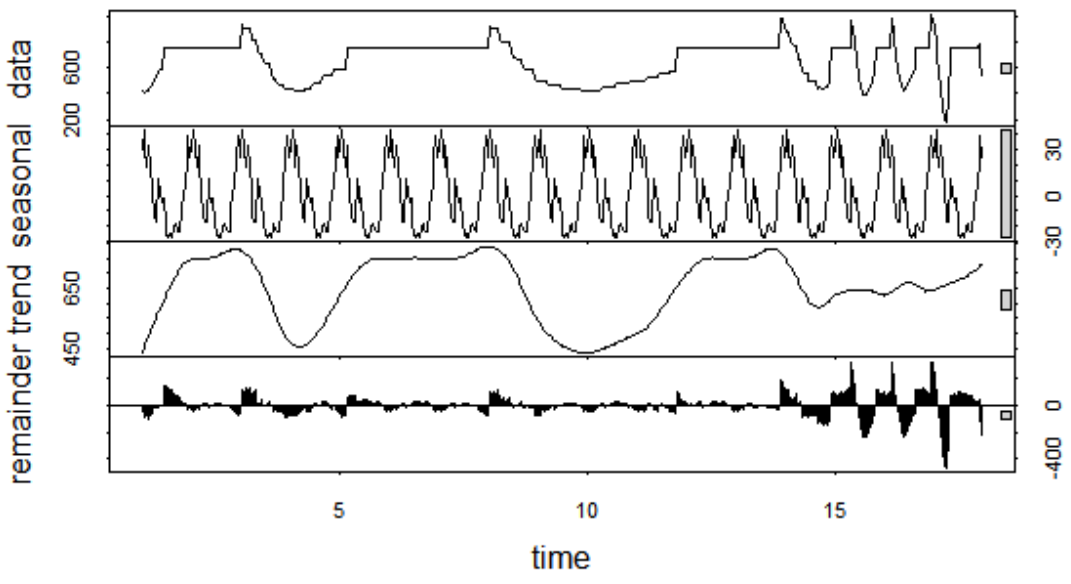


Figure 26- Decomposing data for Jervis car park.

42

Data Quality Report

Conducting a data quality report is important to have a good understanding of the data quality, if the there is an issue with the data quality, it would need to be sorted or re-acquired. Preforming a data quality report as well as an EDA (Exploratory Data Analysis) can give us a clear indication if the data is what we expect it to me. Using the tidyverse package in R Studio is perfect for what we need to complete this section of the report. Originally the plan was that the data would filter into a table. Below shows the code and the outcome and the table will be manually inputted.

```
summary <-function(x){
  funs <- c(mean,median,sd,mad,IQR)
  lapply(funs,function(f) f(x, na.rm=TRUE))
}
sapply(CarparkData, function(x) {if(is.numeric(x)) summary(x)})
```

Figure 27-Code for the Data Quality Report.

```
$BrownT
$BrownT[[1]]
[1] 316.1863

$BrownT[[2]]
[1] 353

$BrownT[[3]]
[1] 73.35964

$BrownT[[4]]
[1] 37.065
```

Figure 28- Image of the result of the Data Quality Report.

	Mean	Median	Standard Deviation	Median Absolute Deviation	Inner Quartile Range
Parnell	365.8549	404	119.7911	117.1254	232.75
Ilac	914.0686	990	112.6694	14.826	167
Jervis	652.5176	750	149.2941	63.7518	248
Arnotts	329.3863	358	50.6544	0	37
Marlboro	255.0157	289.5	86.67309	51.1497	136.75

Thomas Street	222.3353	238	21.466	0	31
Christchurch	93.65217	90	29.63113	34.0998	47
Setanta	97.52745	105	43.06912	56.3388	75.25
Dawson	252.0412	275.5	112.3589	140.1057	217
Trinity	243.8647	248	56.30714	68.1996	92.75
Greenrcs	1054.563	1200	273.3288	0	266
Drury	320.6882	348	158.9228	225.3552	352.5
Brown Thomas	316.1863	353	73.35964	37.065	125.75

Figure 29- Table of the Data Quality Report results.

Clustering

Still continuing with the tidyverse package looking at plotting against two different car parks and analysing the clusters. From this diagram can see that we have two car parks (Ilac & Parnell, taken as an example), each cluster represents a parking space. The closer the clusters are to the line the more normal the data is. Analysing a clustering algorithm shows the strength of the relationship between the data objects. This ggplot is similar to the Q-Q plots, the big difference is that it is comparing the two different car parking spaces. Each plot on the graph is a different space in the two car parks.

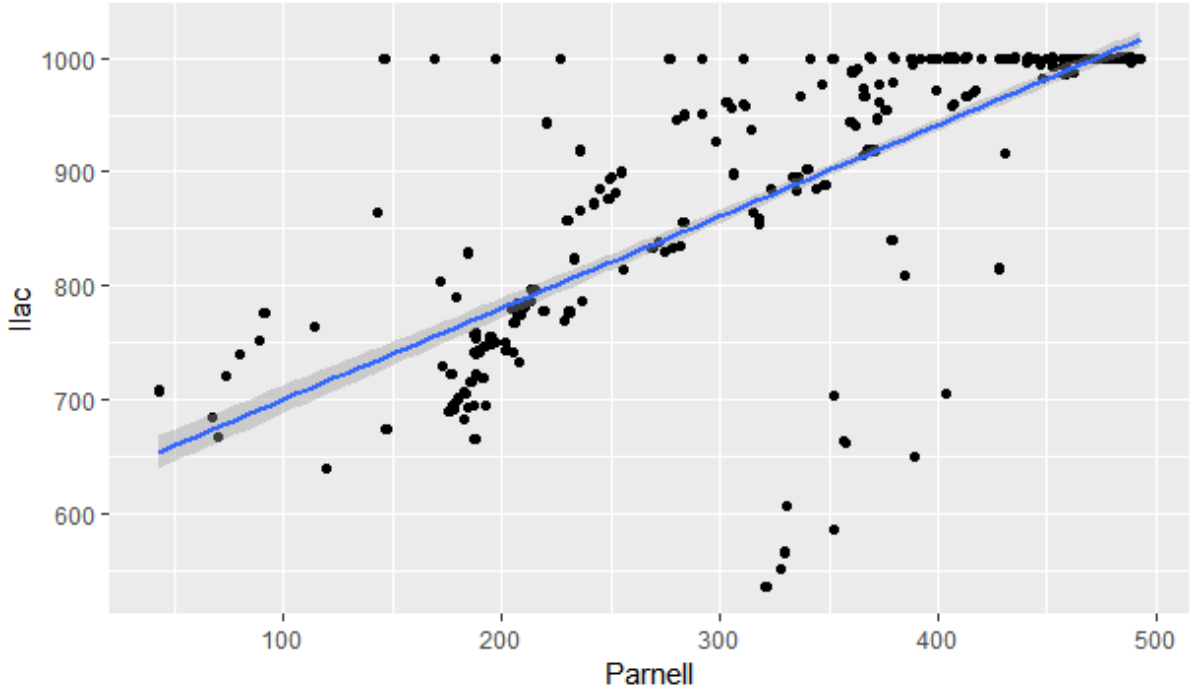


Figure 30- Clustering of car parking spaces between Ilac and Parnell car parks.

ARIMA Model

ARIMA (auto-retrogressive integrated moving average) is mostly used as a technique to a time series and forecasting. Our ARIMA is in a table format and was conducted on R Studio. As mentioned before completing a test of stationarity is the first step before completing the ARIMA Model. The Augmented Dickey-Fuller Test is the test in particular that will help with the ARIMA Model.

```
> auto.arima(deceasonal_cnt, seasonal=FALSE)
Series: deceasonal_cnt
ARIMA(2,0,1) with non-zero mean

Coefficients:
         ar1      ar2      ma1      mean
      1.5604  -0.6002  -0.3784  644.9330
s.e.  0.0974   0.0930   0.1105   28.4179

sigma^2 estimated as 1748:  log likelihood=-2626.85
AIC=5263.69   AICc=5263.81   BIC=5284.86
```

Figure 31- Output of the ARIMA table and the coefficients.

Once the ARIMA Model is fitted, as shown above, evaluation and iterate is the next stage. This will also show us the forecasting of the future.

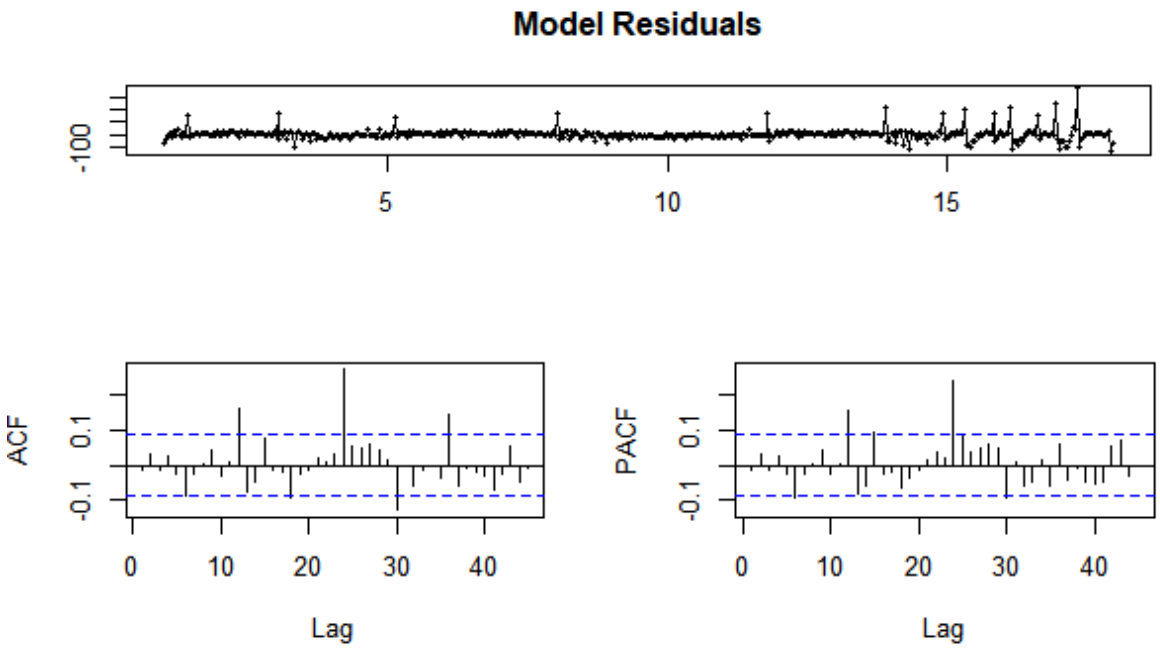


Figure 32- Graphs of the evaluation of the Model Residuals of the whole data set.

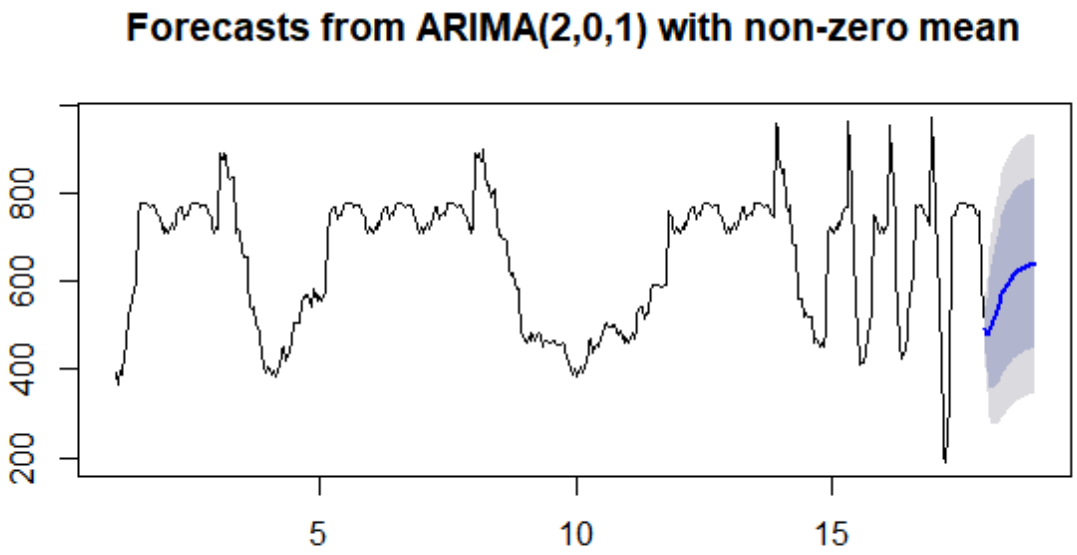


Figure 33-Forcast from the ARIMA of the whole data set.

Testing

Testing is hugely important to a project like this, you need to test to make sure that the answers you have been receiving are correct and the data is not obscure.

For the testing the data was divided into 25/75, the 25% was tested on, this is mentioned in Test Case 3.

Testing wasn't just limited to the code. Testing was completed on the Python Script that web scraped the data and also testing was carried out on the AWS script as it was essential that the instance was running smoothly, to have the satisfied amount of data. These will be discussed in more detail below.

Test Cases

Test Case Field	Description
Test Case ID	Test 1
Test Priority	High
Test Designed by	Ailis Curran
Date of Test Execution	3rd of May 2019
Test Title	Checking the Python Script is Running
Description of Test	To test if the Python Script is successfully scrapping the data and then saving the data into a .csv file.
Pre-Condition	Dublin City Website is automatically displaying the available car parking spaces.
Test Steps	1. 1.Run the following commands through the cmd: • cd desktop • cd software • Python Test4.py 2. The available car parking spaces will automatically load, including the exact date and time that the data was collected. 3. Exit cmd. 4. Open Software folder.

	5. Check the CarparkData.csv file, which is saved. 6. The file should have updated automatically with the new free car parking spaces.
Requirements	• Test4.py • Carpark.csv
Expected Results	The cmd show show the current available spaces for that time. The .csv file should have the updated version.
Post Condition	The script shows the updated values with no issue.
Status (Fail/Pass)	Pass

Test Case Field	Description
Test Case ID	Test 2
Test Priority	High
Test Designed by	Ailis Curran
Date of Test Execution	3rd of May 2019
Test Title	Checking the AWS is Running
Description of Test	To test if the Amazon Web Service is successfully running the data with no issue and then automatically saving the data into the .csv file.
Pre-Condition	Amazon Web Service is running smoothly. Filezilla is running smoothly.
Test Steps	1. Open up AWS.

	2. Find the instance you want to view. 3. Analysis the properties of the AWS instance, looking for any major errors. 4. Open up Filezilla. 5. Open site manager. 6. Connect Carparks.
Requirements	• Carparks.csv • Amazon Web Services account (Ubuntu). • Filezilla. • PuTTY.
Expected Results	Left hand side of the page should show the AWS instance with no errors.
Post Condition	Filezilla can't connect and the AWS instance needs to be rebooted.
Status (Fail/Pass)	Pass

Test Case Field	Description
Test Case ID	Test 3
Test Priority	Medium
Test Designed by	Ailis Curran
Date of Test Execution	3rd of May 2019
Test Title	Checking for normalities in the R Studio code.
Description of Test	To test that there are normalities in the R Studio code and nothing completely off.

Pre-Condition	R Studio is up to date and running smoothly.
Test Steps	1. Open R Studio. 2. Once Carpark.csv file has been read in and all the necessary analysis has been carried out, the testing can start. 3. First divide the data set into 25/75. 4. Next start by checking the structure of the test data. 5. Move on and test the dimension of the test data. 6. Next check out the predictor variable. 7. Then finally check the number of unique values in the testing data.
Requirements	• Carpark.csv • R Studio • Code to conduct the tests.
Expected Results	R Studio will show us that there is nothing to unexpected.
Post Condition	R Studio will show us the results and then these results can be saved.
Status (Fail/Pass)	Pass

Testing Output

First test that was analysed was just to check the overall structure of the testing data.

```
> str(testing)
'data.frame':   128 obs. of  14 variables:
 $ Parnell  : int  183 237 323 334 372 392 435 470 482 483 ...
 $ Ilac     : num  683 786 886 895 947 1000 1000 1000 1000 1000 ...
 $ Jervis   : num  408 423 509 521 586 750 750 750 750 750 ...
 $ Arnotts  : num  344 359 351 355 352 353 358 358 358 358 ...
 $ Malboro  : int  160 173 243 259 292 293 303 304 298 296 ...
 $ Thomas   : num  173 189 212 214 233 238 238 238 238 238 ...
 $ ChristC  : num  76 83 108 111 110 104 108 123 148 NA ...
 $ Setanta  : num  19 146 74 78 86 146 114 129 131 131 ...
 $ Dawson   : num  87 122 190 209 254 278 304 330 331 370 ...
 $ Trinity  : num  143 167 233 233 248 249 252 279 294 304 ...
 $ Greenrcs : num  267 250 264 1200 305 319 327 1200 378 378 ...
 $ Drury    : num  143 193 195 172 153 180 250 493 497 499 ...
```

Figure 34- Image of the output of the structure.

Moving on, checking the dimension of the test data, this helps to measure the quality of the data and helps to measure and communicate the quality of data.

```
> dim(testing)
[1] 128  14
```

Figure 35- Image of the dimension of the data.

Next was checking out the predictor variable, which predict the value of a variable based on the value of another independent variable in the dataset.

```
> apply(testing,2,class)
    Parnell        Ilac      Jervis     Arnotts     Malboro      Thomas     ChristC
"character" "character" "character" "character" "character" "character" "character"
    Setanta      Dawson     Trinity    Greenrcs       Drury      BrownT   Date_Time
"character" "character" "character" "character" "character" "character" "character"
```

Figure 36- Image of the predictor variable.

Finally checking the number of the unique values in the data set.

```
> apply(testing,2,function(t) round(length(unique(t))/nrow(testing),3)*100)
  Parnell      Ilac    Jervis   Arnotts   Malboro    Thomas   ChristC   Setanta
     63.3      41.4      43.0      35.9      64.1      28.1      46.9      48.4
   Dawson   Trinity  Greenrcs     Drury    BrownT Date_Time
     51.6      59.4      32.8      59.4      51.6      99.2
```

Figure 37- Image of the number of unique values in each variable.

Evaluation

The results of all the analysis that has been undergone in the project has been presented and discusses in the above titles. The KDD methodology was used and adapted for the whole of the project. Having the KDD methodology gave me a bases on follow my project along and did help in times of trouble. It also insured that the data selected to do the analysis on would be satisfactory, the pre-processing of the data would be prepared correctly for the in depth analysis that

was underwent for this project. Once all the analysis was concluded from the data it meant that the results could be saved onto the Google Drive.

Key Analysis

The key analysis was the implementation of the data from the Python script, cleaned and organising the questions around the data. Questions that would be interesting and would be something the user would possibly need in the future of planning or organising of future car parks and multi stories. Tableau was very important to the overall graphs for these questions, it gave the tools needed to have very colourful, yet clearly concise graphs. These graphs are easy on the eye and also easy to understand. The graphs and tables on R Studio are a little more complex than the ones on Tableau. These graphs aren't as colourful or as clearly mapped out. Although not as visually satisfying till very important to the findings of the project as they are all mentioned above.

Recommendations

Now after the processing a rich dataset of car parking spaces in Dublin, and also the analysis to go with the dataset. After answering questions on the analysis we can now see that the shopping centres have the highest amount of car parking spaces, and it would be interesting to view a car park with 1000 plus car parking spaces that wasn't attached to a shopping centre, to see how busy it would get.

It seems that there is enough car parking spaces right now for the people of Dublin, but it should be something mindful in the future and more effort needs to be put into better and more effective public transport to cut down the car traffic in the city centre.

Key Findings

1. The car parks are busier during the week compared to the weekend.
2. Car parking spaces are larger once attached to shopping centres.
3. The South-side have much more car parks then the North-side.
4. There are obvious trends with possibly regular people parking in certain areas.

Project Changes

After evaluating the entire project life span and overviewing the analysis that now occupies the project. I can now make an overall idea of how I would change the project.

From the beginning of the project was confusing. The showcase last year gave me some ideas, now with that being said the data analysis side of the project that originally was unsure of. From the past students talking about the final year project, the original idea was an app, not internally data analytic related. This overall wasted a lot of time.

Another thing that would have helped the project and the effectiveness of the data as well as the analysis would have been to check the AWS. Now to the eye the AWS was working and looked fine. It wasn't until the final dataset was to be downloaded that the errors occurred. Looking back now the AWS should have been looked at periodically, although it is a cloud service and the data should have been running smoothly already.

Conclusions

The findings in this project have concluded that there should be an incentive for more people to use public transport. Dublin is famous for people complaining about busses and trains running late or in some cases not even turning up. This needs to improve or else more people will be inclined to bring their cars into town rather than use public transport. Driving is a very important skill, driving gets people places that would normally take two or three modes of transport to. Going into town, driving wouldn't be my first choice of transport, yet the public transport can be unreliable at times.

Overall there needs to be more car parks on the North-side of town as there are plenty of shops that people who would park on the South-side would miss. To increase the North-sides business one or two more car parks wouldn't go a miss.

Further Development and Research

A second phase of this project would be to look at the other parts of the country, for example big cities like Galway and Cork. Possibly in the future other countries in Europe. This data is extremely needed for the future of planning and also the future of transport in an out of cities. This analysis would be really useful for the likes of London City, there they are now charging drivers who come into the city as London has one of the highest congestion area in Europe which is really harming the OZ layer.

53

References

1) PRICING GUIDE: HOW MUCH CAN I MAKE FOR SPARE SPACE IN DUBLIN 2 - PARKPNP BLOG.

In-text: (Parkpnp Blog, 2019)

Your Bibliography: Parkpnp Blog. (2019). *Pricing Guide: How Much Can I Make For Spare Space in Dublin 2 - Parkpnp Blog*. [online] Available at: https://parkpnp.com/blog/pricing-guide-how-much-can-i-make-for-spare-space-in-dublin-2/ [Accessed 3 May 2019].

2) GUINEY, R.

Cars: good or bad? It's not that simple in Dublin city centre

In-text: (Guiney, 2015)

Your Bibliography: Guiney, R. (2015). *Cars: good or bad? It's not that simple in Dublin city centre*. [online] The Irish Times. Available at: https://www.irishtimes.com/opinion/cars-good-or-bad-it-s-not-that-simple-in-dublin-city-centre-1.2318854 [Accessed 3 May 2019].

3) PARKING IN DUBLIN - IRISH PARKING ASSOCIATION - IRELAND'S PARKING INDUSTRY REPRESENTATIVE

In-text: (Irish Parking Association - Ireland's Parking Industry Representative, 2019)

Your Bibliography: Irish Parking Association - Ireland's Parking Industry Representative. (2019). *Parking in Dublin - Irish Parking Association - Ireland's Parking Industry Representative*. [online] Available at: http://www.parkingireland.ie/car-parking/parking-in-dublin/ [Accessed 3 May 2019].

4) ILAC SHOPPING CENTRE | PARKING IRELAND | PARK RITE

In-text: (Parkrite.ie, 2019)

Your Bibliography: Parkrite.ie. (2019). *Ilac Shopping Centre | Parking Ireland | Park Rite*. [online] Available at: https://www.parkrite.ie/carparks/ilac-shopping-centre [Accessed 3 May 2019].

5) PARKING FEES, FINES AND VEHICLE CLAMPING

In-text: (Citizensinformation.ie, 2018)
Your Bibliography: Citizensinformation.ie. (2018). *Parking fees, fines and vehicle clamping*. [online] Available at: https://www.citizensinformation.ie/en/travel_and_recreation/traffic_and_parking/parking_fines_and_vehicle_clamping.html [Accessed 3 May 2019].

6) TEAM, D.

Data Mining and Knowledge Discovery Database(Kdd Process) - DataFlair

In-text: (Team, 2018)
Your Bibliography: Team, D. (2018). *Data Mining and Knowledge Discovery Database(Kdd Process) - DataFlair*. [online] DataFlair. Available at: https://data-flair.training/blogs/data-mining-and-knowledge-discovery/ [Accessed 3 May 2019].

7) DBD, U.

KDD Process/Overview

In-text: (DBD, 1996)

Your Bibliography: DBD, U. (1996). *KDD Process/Overview*. [online] Www2.cs.uregina.ca. Available at: http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html [Accessed 3 May 2019].

8) CENSUS OF POPULATION 2016 – PROFILE 6 COMMUTING IN IRELAND

In-text: (Central Statistics Office, 2016)
Your Bibliography: Central Statistics Office. (2016). *Census of Population 2016 – Profile 6 Commuting in Ireland*. [online] Available at: https://www.cso.ie/en/releasesandpublications/ep/p-cp6ci/p6cii/p6mtw/ [Accessed 3 May 2019].

9) DUBLIN CITY COUNCIL | DUBLIN CITY COUNCIL

In-text: (Dublincity.ie, 2019)

Your Bibliography: Dublincity.ie. (2019). *Welcome to Dublin City Council | Dublin City Council*. [online] Available at: http://www.dublincity.ie/ [Accessed 3 May 2019].

10) PREFACE

In-text: (Cran.r-project.org, n.d.)

Your Bibliography: Cran.r-project.org. (n.d.). *Preface*. [online] Available at: https://cran.r-project.org/web/packages/dlookr/vignettes/diagonosis.html [Accessed 3 May 2019].

56

Appendix

Poster

National College of Ireland

ANALYSIS OF CAR PARKING SPACES IN DUBLIN CITY CENTRE

Ailis Curran
Data Analytics
22nd of May 2019

Analysis was carried out on car parking spaces in Dublin City Centre for the future of planning in the city.

Main Technologies Used

Python
Amazon Web Service
R
PuTTy
R Studio
Filezilla
Mircosoft Excel

57

Project Proposal

Objectives

Car parks are part of some people's daily routines, even if they don't even realize it. It's where we assume our cars will remain safe and in place for a certain amount of time or even overnight. Car parks are a huge necessity in this country and with huge increase in cars being the most popular way to travel to work and overall the most popular. This influx in cars coming in and out of Dublin city means that the cars have to have a place for cars to be parked, now most companies have underground parking but for the likes of ESB Networks the car parking spaces were for higher ups only, leaving other people stuck, which is where the project idea originated. Finding out which parking spaces are available the most or not. This is only one of the questions that being able to look at and being in the data analytic stream the data will be put into tableau and input my information and have it displaying in a user friendly format.

In this project proposal, will set out the aim, requirements and ambitions to analyse the car parks in Dublin city center and possibly other cities in Ireland. These ambitions will be met by using Python and R to retrieve relevant data and exporting it is a csv file. From this information I will input it into Tableau and show this data as a user friendly environment.

Motivation

Purpose

The purpose of this document is to set out the requirements for the development of data analysis project to look at the available parking spaces in the Dublin city area. With this information can then inform users about availability and if the spaces are available for people. The project will be important to people who need to park in the city centre and any planners that will need this information in the future.

Project Scope

The scope of the project is to develop a data analysis project that will web scrape the information from an html code which is generated every 20 mins or so. This website that is there already isn't user-friendly and is multiple clicks on the website till you receive the correct page. This will be the website the data

will be received from. The data sets will be a couple of months of information. Programming languages Python and R will be used. The system shall have a tableau connected to the information so the findings will be clearly displayed. There will not be any extra costs for this project, just a lot of time and hard work will be used in this project, as balancing other modules and assignments. I will also be looking at traffic at certain times of the day, mostly rush-hour to gain an estimate of the amount of cars on the road.

Definitions, Acronyms, and Abbreviations

PS Parking Space

DCC Dublin City Council

NCI National College of Ireland

Background

The background of my project starts from when the idea came to mind, which was in my Internship, as a driver and a car owner, asking about the underground car parking spaces as there was two floors of underground facilities and not all the spaces were full. Unfortunately after being refused a space, it was then the project came to mind.

Originally an app was what came to mind that people used and could check which parking space was free and if it was suitable for their needs, or even if they wanted to be parked on a certain part of town ie: North or South side. This idea of the app is great and would like to look into it in more detail in the next semester but as I'm doing data analytics, analytics needed to be looking at the statistics of the car parks and not to be 100% focused on the app. I'll be scraping the real time information on the car parks and inputting it into tableau.

With this project will be using tableau to display the information that has been web scraped using beautiful soup 4 and an xml web parser. This allows the information to be scraped whenever it is needed. As the data updates every five minutes.

In the research there is only one website at all in regards to car parking spaces in Ireland. The billboards once you are going into town there it tells you the parking space and the availability beside it is the information that is being scraped. On the Google Play store there is only car parking games bar one application which allows you to book parking spaces in England. There is another app which is called free parking but people in the comments were complaining about the apps accessibility as they wanted people's contacts and access to camera. Although this app is for the big cities like London, Paris and New York, it is only for free parking. In Dublin free parking is near impossible to find.

From the research there is sufficient evidence in favour for this application to be a commercial success. This application is unique with limited competitors but none that are same as my approach.

In 2016 61.4% of people drove in a car to get to work and the percentage of people using buses, trains and walking have gone down from the 1980's. Seeing this change it is clear that people are taking the car more over the years and this will be set to continue, so an app for the parking spaces will be perfect for the large majority of people who drive to work.

Although this project will be successful, the idea is clear but the technical details just need to be finalised and worked on to make it the most seamless usage for the user.

Technical Approach

For any project to work it is key to plan your time accordingly, time management will be key for managing this project. This will be shown in a Gant Chart in the Project Plan section. Research is key for the project as well, technology moves fast, so I need to keep on the lookout for a similar application and if that happens I'll need a way to better the competition.

- Strategic Planning: Microsoft word, Microsoft Project.
- Diagrams: Microsoft Project, Draw.io, Tableau
- Backup Data: USB Stick, Git Hub

- Documentation: Microsoft Word
- Research: Internet, books
- Data: Found on website
- Language: python, xml

Project Plan

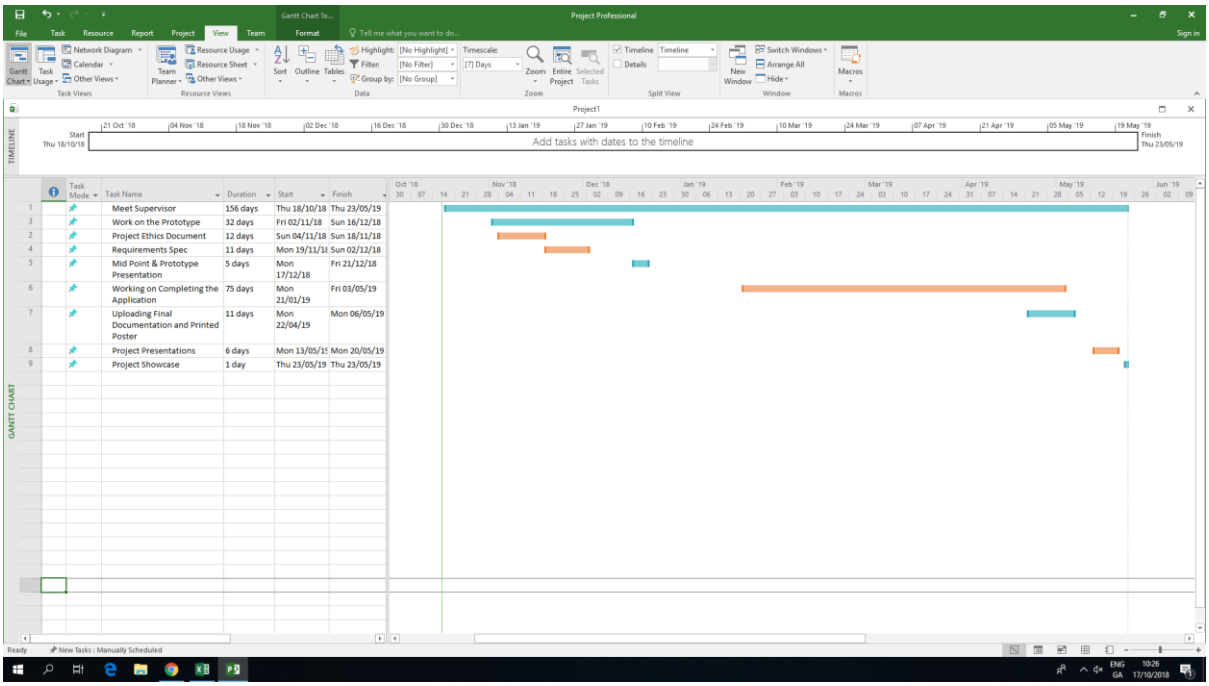
Gantt chart from the Project Proposal shown in appendix below.

- Project Ethics Document: 18th November 2018
- Requirements Spec: 1st December 2018
- Mid-Point and Prototype Presentation: 17th December 2018 – 21st December 2018
- Project Final Documentation:6th May 2019
- Project Presentations: 13th – 20th May 2019
- Project Showcase: 22rd May 2019

Evaluation

On reflection this application will be an enjoyable challenge haven't used python till this year in college, but it will be exciting to use this for the Data Analysis and see what the results come out as. Testing will be important as will time management once it is kept in the back of my mind then the rest is just hard work and determination. Testing will be carried out by myself. The time management has been introduced during the time.

Project Plan


Project Professional
Project1
TIMELINE
Add tasks with dates to the timeline
GANTT CHART
Meet Supervisor 156 days Thu 18/10/18 Thu 23/05/19
Work on the Prototype 32 days Fri 02/11/18 Sun 16/12/18
Project Ethics Document 12 days Sun 04/11/18 Sun 18/11/18
Requirements Spec 11 days Mon 19/11/18 Sun 02/12/18
Mid Point & Prototype Presentation 5 days Mon 17/12/18 Fri 21/12/18
Working on Completing the Application 75 days Mon 21/01/19 Fri 03/05/19
Uploading Final Documentation and Printed Poster 11 days Mon 22/04/19 Mon 06/05/19
Project Presentations 6 days Mon 13/05/19 Mon 20/05/19
Project Showcase 1 day Thu 23/05/19 Thu 23/05/19