ANALYSING AND VISUALISING FACTORS IMPACTING FUEL PRICE FLUCTUATIONS IN THE PERTH METROPOLITAN REGION

Syed Muhammad Ahmed Zaidi

Student ID: 20972008

School of Electrical Engineering, Computing and Mathematical Sciences

Curtin University

This report is submitted for COMP6016

November 2024

DECLARATION

This document is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. It has not been previously submitted, in part or whole, to any university of institution for any degree, diploma, or other qualification.

Syed Muhammad Ahmed Zaidi

Masters of Predictive Analytics

Curtin University

Signed: 

Date: ____

ABSTRACT

This study investigates factors influencing fuel price fluctuations in Western Australia, particularly focusing on the Perth metropolitan suburbs by examining the impact of numerous factors such as public holidays, weather conditions, population density, interest rates, inflation, taxes, etc., to gain a comprehensive understanding of how fuel prices are determined. It utilises multiple data sources, with primary data sourced from the government’s FuelWatch platform, whilst other sources are engineered together to carry out statistical and spatial analysis with the objective of finding correlations and patterns over the period of 2020 to 2024.

The results suggest that certain local factors, such as public holidays, marginally raise fuel prices to increase demand. However, the impact is not too significant and is short-lived. Similarly, weather conditions, particularly the rainfall, show no significant correlation until it goes beyond a certain bracket of showered millimetres. Other analyses, like prices according to weekdays, show interesting cycles that are being followed only by Perth.

In contrast, macroeconomic factors show a more significant impact that lasts longer on how fuel prices are determined. The lagged response is evident with the investigation of interest rates as it proves an increased correlation over time. Factors such as excise taxes demonstrate a minor impact even during periods of massive tax reductions of 50%, indicating the power of other global factors playing an even more vital role in determining equilibrium.

Furthermore, the study contributes by giving recommendations that could allow this research to become even more robust so that deeper insights could be derived that could be beneficial for multiple stakeholders, including consumers, businesses and policymakers.

Acknowledgements

I would like to extend my heartfelt gratitude to Curtin University for giving me the resources and environment to carry out such an extensive project. I am also particularly grateful to Curtin Institute of Data Science (CIDS) for providing me with this opportunity and backing me to complete my research till the end.

I would also like to deeply thank my Supervisors, Dr Nancy Tippaya and Calvin Pang, who guided me through this journey, showering their experience and knowledge to help me learn and excel in my field. I would also like to express sincere appreciation to Dr Valerie whose support and insights were vital to conduct the project.

Finally, I am immensely thankful to my family, especially my parents, wife and brother for their unwavering support, encouragement and understanding. They have been the source of my strength and motivation throughout.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

List of Abbreviations and Acronyms

API – Application Programming Interface

EDA - Exploratory Data Analysis

IDE - Integrated Development Environment

CSV – Comma Separated Values

RBA – Reserve Bank of Australia

DF – Data Frame

IQR - Inter Quartile Range

DBSCAN - Density-Based Spatial Clustering of Applications

CPI - Consumer Price Index

GPS - Global Positioning System

ULP91 - Unleaded Petrol 91 Octane

ABS - Australian Bureau of Statistics

VS Code - Visual Studio Code

LIST OF APPENDICES

1 INTRODUCTION

This section sheds light on the key research problem, providing an overview of how fuel prices fluctuate in metropolitan areas of Perth. It also summarises the scope of the study by discussing the assumptions and limitations. In brief, the objective is to discover the hidden relationships between numerous factors and how fuel prices are determined. The approach is to investigate underlying patterns in publicly available datasets using various visualisation techniques to identify correlations with fuel prices in selected suburbs of Perth. This chapter also outlines the terminologies and working definitions used during the research that were found within the reviewed literature. The contributions of this work include the insights found into how fuel prices are determined when considering multiple factors including public holidays, weather conditions and other macro factors.

1.1 RATIONALE

As inflation and cost of living rise, fuel prices start to play a vital role for individuals and businesses. It comprises of a significant proportion of disposable income directly affecting how people make their daily living decisions. Specifically for Western Australia, where many people are heavily dependent on road transport due to its large geographical area, fuel pricing plays an important to make intelligent purchasing strategies and benefit from it. Identifying key drivers in why such fluctuations occur can offer individuals and businesses valuable information that influences their consumer behaviour and future planning.

There has been a considerable amount of research conducted on this topic. However, a vast majority of it was aimed at finding a relationship between fuel pricing trends and macro factors, including crude oil, global financial changes and world events. Only a few studies have given attention to micro factors that include local influences like public holidays, weather conditions, population density etc. This project fills this gap by focussing more on local influences by searching for cyclical trends and patterns observed over certain suburbs of Perth.

The motivation for this research evolves considering the financial hardships currently being faced by consumers and businesses due to the increase in living costs. Fuel consumption being an inevitable expense, can be something people can strategise to have long-term savings. Having knowledge of these unexplored local factors can allow residents of Perth to make smarter financial decisions, easing the burden on every household. The opportunity to contribute within this academic gap is another noteworthy cause to make efforts to investigate into this matter. All findings will contribute to a practical solution to the overall problem of increasing inflation and subsistence costs.

The beneficiaries of this research include the entire population of Western Australia as the sample of certain metropolitan suburbs of Perth will contribute enough information towards this gap in the study that could help in a clearer understanding of fuel pricing patterns, allowing every individual to make an informed decision to save some dollars in the long run. Another stakeholder would be businesses, especially those directly related to delivery and logistics, as the insights of fuel price fluctuations will allow them to predict future expenditures, enabling better planning for cost reduction techniques. On an even broader level, policymakers will also benefit from the knowledge produced by this project as it can be a vital source of information that can be considered before refining regulation strategies and promoting competition among numerous brands.

1.2 Statement of Problem

There can never be a single reason for any fluctuation in fuel prices. There are always multiple factors pushing the prices in different directions. Powerful ones include the macro factors which are related to global oil price determinants (Hamilton 2009). Whereas factors that have a weaker correlation tend to stay hidden since they remain underexplored. There has been some research regarding the comparison of petrol prices across multiple Australian cities (Ghazanfari 2022), however none that would specifically target Perth and its metropolitan areas.

Despite the vital role fuel prices play in the everyday lives of consumers, businesses and policymakers in Perth, there is still a lack of proper understanding of how this is determined or the local factors that contribute towards its volatility. As mentioned by (Borenstein, Cameron, and Gilbert 1997), fuel price volatility leads to economic inefficiencies. If not eradicated, it can always be reduced by applying a strategic approach to everyday living with the use of the knowledge gained through this research.

Perth, being a regional city, has untapped resources of knowledge that can be explored with more research. This project answers the question of "what local factors contribute to fuel price fluctuations in metropolitan suburbs in Perth and how the knowledge of this can be put into actionable insights." It questions and investigates a range of factors that could potentially contribute towards any kind of pull or push in demand or supply of fuel, causing prices to change.

1.3 Delimitation and Limitations

The study particularly focuses on analysing fuel prices for only a specific variant which is the ULP91 (Unleaded). There are many other categories of fuel currently being offered by the studied area, which include PULP 95, 98 RON, LPG, Diesel, etc.; However, to keep the research concentrated, only the category that is most used is being considered. This limits the generalizability of the findings to other fuel types, as their characteristics capture different patterns of demand and supply.

During the project, the scope of the area was reduced from the whole of Western Australia to the Metropolitan suburbs of Perth. This includes a total of 203 Suburbs, which consists of 738 fuel stations being owned by 29 Brands. The duration was kept for a total of 5 years, starting in 2020 and ending in late 2024. While the scope was reduced considering the time frame of the project and to do a controlled analysis of the region without adding further complexity of other areas being vastly different in many geographical features. By only focusing on certain areas and a certain fuel type, the study aims to provide insights that can be directly relevant to a larger number of consumers, businesses and policymakers.

Additionally, due to budget constraints, the project was heavily dependent on publicly available datasets, which included scraping data from websites like FuelWatch and Open-Mateo. There is an absence of paid datasets which could have helped further into finding more insights into how fuel pricing works. These include data for Population Census, Household Incomes and Sporting Events. Moreover, there was no manual data collection implemented when searching for data due to the nature of the information required. Major proportion of data was already readily available through APIs and others were downloaded from the web.

Delimitations for the project were established early to maintain a manageable scope. As much as the impact of international oil prices is significant in determining prices, such factors are already established and thoroughly discussed by many studies and hence, hidden factors like local influences, public holidays, weather conditions, etc., were researched to explore insights contributing towards a practical solution.

The target audience includes the general public who are the consumers of fuel. Apart from them, the businesses who are direct stakeholders of price fluctuations in fuels are according to the industry they provide service. Lastly, the policymakers who aim to aid the majority use their skills to devise policies that are financially and practically beneficial for the people.

1.4 Definition of Terms

The report consists of numerous pieces of jargon that are used to explain different scenarios in the world of fuel and terminologies that are used when analysing its data. For ease and understanding, we will be using different abbreviations to fuel the types. This includes ULP91 for Unleaded Petrol 91. It refers to Unleaded Petrol with an octane rating of 91, which is considered the standard fuel type for many vehicles in Australia due to its wide availability (Kalligeros et al. 2003).

Another important use of the term Geospatial analysis was done when carrying out regional and brand analysis. This refers to the analysis of spatial data, which often uses visualised maps to identify relationships between variables (Goodchild et al. 2005).

Since most graphs were created with time being the x-axis, the use of Time-Series Analysis frequently occurred. This refers to analysing data points collected or recorded at a specific time interval; that allows tracking and forecasting fuel price trends over time (Box et al. 2015).

The project was highly code intensive with most of the analysis being carried out on Visual Studio Code (VSCode). Some terminologies related to this field that highly occur in the report include Data Frame. It is a data structure used for storing tabular data in a format that is easy to manipulate, query and analyse. It is primarily used to handle rows and columns, making it ideal for filtering and grouping large datasets (McKinney 2010).

Another technique used to fetch data from the internet was the use of API (Application Programming Interface). It refers to the process of code to program a request that would retrieve data from external services or databases (RT 2000). For example, within this project the Google Maps API was just to fetch longitude and latitude data for all the 738 Stations considered.

Lastly, to visualise maps, Shapefiles were used to import geospatial vector data for mapped analysis. It is a way to interactively represent geographic locations through visualisations (ESRI 1998). In the project, such files were used to locate fuel stations in the map of Perth and to carry out cluster analysis for brands.

1.5 ASSUMPTIONS

For this research, it is assumed that the data collected from various websites and repositories, including FuelWatch, Open-Mateo, Government statistics, etc., are accurate and up to date. The analysis and findings rely on the correctness of the data offered by these services. The use of pricing, location coordinates, weather conditions, etc., are all considered to be valid and reliable. Any discrepancies in the data may have an impact on the findings and may affect the selected outcomes of the hypothesis.

The study also assumes that the selected 203 metropolitan suburbs of Perth are sufficient for to be considered as a diverse sample that may represent the overall consumer behaviour of Perth. It is assumed that the primary dataset of fuel pricing, along with the engineered datasets of other factors, cover a range of areas and their characteristics that could infer insights captured to the overall population.

Another assumption kept constant is the quality of Fuel (ULP91) over all the fuel stations being studied. Differences in the fuel quality by different brands or regions are ignored to limit complexity, and it is assumed that the demand and supply are the major sources for price determination and not the quality of the product.

1.6 RESEARCH APPROACH

The approach of this research follows a systematic method of collecting, processing, analysing and interpreting the results (Khan et al. 2003). The overall objective of finding hidden factors impacting fuel prices in metropolitan areas of Perth was the driving point of this research. It employs a mixture of qualitative and quantitative approaches using theoretical and statistical analysis to examine the relationship between fuel prices and other macro and micro factors. The study was initiated with a review of available literature with the aim of gaining knowledge in the field being investigated. It was then followed by a challenging compilation of datasets that were collected through numerous sources. These datasets were engineered together into different notebooks where each factor was processed individually through integrated data frames. Upon cleaning and merging, analysis was carried out using various statistical and geospatial techniques. Hypothesis were created and tested using evidence found to reach conclusions about insights being examined.

1.7 Report Structure

The structure of this report starts with the abstract and introduction section. This specific chapter highlights the research problem and primarily focuses on the main objective of why this research is carried out. It discusses the necessity of the study and the motivations behind the rationale of doing it.

Followed by this is the literature review that thoroughly discusses the past research that has been carried out and how it is related to this investigation. It attempts to extract the hidden factors that can be studied further to see if a hypothesis can be created and examined. This section also evaluates the different methods employed by researchers before for time series and geospatial analysis that can be integrated with the current research.

The next session of this report discusses the Methodology, which explains the processes of data handling in detail. Starting with data sources, it moves forward in giving in-depth details of the methods followed to collect, process and analyse data to reach conclusions. It sheds light on the processes of data extraction, scraping and the use of APIs to fill relevant fields in a dataset. The choice of software, including IDEs, language and libraries, is thoroughly discussed.

Chapter 4 of this report sheds light on the techniques used to complete the exploratory data analysis and results achieved over the course of the complete investigation. It discusses the rationale behind the use of a certain type of visualisation and interprets it to clarify the deduction made from it. The results highlight the patterns found that were hidden under the raw data and better understood using static and interactive plots. Furthermore, this section moves over each factor at a time and looks to see if the evidence is enough to satisfy the hypotheses made at the start of the research.

Conclusion section summarises the key findings of the study and paves way towards further investigation in the future. The section uncovers the left-out areas where the next focus should be and the parts where this research lacked. It gives recommendations for stakeholders like consumers, businesses and policymakers on how they can make use of the findings for the beneficiary of the overall economy.

1.8 THEORETICAL FRAMEWORK

This project is grounded on several economic and statistical theories that explain the reasons and aids in understanding on how price mechanism and external influences impact the determination of pricing in the Fuel Market.

- Price theory being a fundamental economic theory, explains how the prices of goods and services are determined in a competitive market. In context of fuel, the theory helps in examining what factors of supply and demand are causing the price to hit the equilibrium. Considering multiple limitations, competition and demand fluctuations, and this theory is essential to deeply understand how the market of fuel works.
- There are behavioural and psychological theories that are needed to understand the strategic responses from fuel stations in anticipation of demand. For example, understanding their reaction to upcoming public holidays can allow a deeper dive into insights related to cognitive decisions causing fluctuations in the market.
- Lastly, spatial and clustering theories like K nearest Neighbour are required to understand how clustering between brands and fuel stations are carried it. This is also required to understand spatial graphs.

1.9 CONTRIBUTIONS

The project makes considerable contributions to the analysis of fuel pricing, particularly for Western Australia dynamics. The study examines fuel prices on multiple criteria which included daily, weekly and other types of seasonal analysis that provides a comprehensive understanding of its peak and depth over time. It helps in extending the knowledge by experimenting the clustering phenomenon for brands and fuel stations, highlighting areas with comparative price increase and reflecting the causes behind it. It also investigates macro-economic factors like taxation, inflation and interest rates to test the contrary assumptions found in other literature.

All the research allows considerable accumulation of knowledge that contributes to decision making for multiple stakeholders. This may help strategise purchasing for consumers and retailers to lower their costs by effectively making the right choices. It may also act as a tool to policymakers to create regulations based on researched insights that may benefit both the users and the businesses.

2 REVIEW OF LITERATURE

2.1 BACKGROUND AND CRITIQUE

The retail prices of any type of fuel are influenced by a combination of both global and local factors. Considering the global scale, prices of crude oil play the most important role as any shift in the changes of its market price affects running of economies. The changes in crude oil prices are largely due to its demand and supply, geopolitical events and the oil production capabilities and policies. The price shocks and oil prices that happened in 2007-08 had a massive impact on the world economies. Such impactful changes have long-term impacts and can be devastating in changing consumer behaviour and overall purchase patterns. As Hamilton, 2009 mentions in his article how the crude oil price volatility caused consequential hikes in fuel prices around the world, which impacted global transportation costs and led towards an increase in inflation. Hence, global oil prices serve as a foundation for how fuel prices are priced; however, it is equally important to understand the local reasons and their correlation to improved strategic planning. This is especially true for an untapped region of Western Australia where there are numerous regional factors in play.

Past research has consistently proven the impact of local factors on fuel prices in different regions. These factors range from population density, transportation choices, competition levels and more. Similar research was done by (Ghazanfari 2022) where he analysed and compared the fuel prices across multiple Australian cities. He discovered that the prices differ greatly, especially in the rural areas, experiencing a higher rate compared to urban. Few of the reasons being higher distribution costs and little competition from other stations to keep the prices low. Majority of urban areas have multiple brands competing against each other on prices causing the overall price averages to be low. This competitive pricing strategy can cause regional differences with stations densely packed together having a lower average cost for diverse types of fuels. Comparing rural with metropolitan areas does also reflect how their contrasting regional characteristics can also contribute towards different transportation costs and operational efficiencies both impacting how fuel is priced.

Western Australian fuel market is unique compared to other states of Australia. It has its own regulatory interventions and local characteristics that keeps it apart from others. The government has launched multiple platforms related to fuel pricing with the intention of making it easy for consumers and retailers to strategise their daily expenses. One of these is called the FuelWatch, which requires all stations in the area to set their prices in advance and quote it to the website for the entire public to see. These policies were put into practice to make a difference; however, many of them do little when having an impact on the demand and supply of fuel prices (Dewenter, Heimeshoff, and Lüth 2017b). Despite efforts to increase transparency with the hope of creating more stability, local prices continue to fluctuate based on other key factors that are driven by geographic isolation, regional competition and distribution logistics.

Apart from the global factors, local events and conditions are also relevant in Western Australia when considering how fuel prices are being determined. For example, there are a significant number of public holidays which the public enjoys. (Ghazanfari 2022) explains how such minute reasons could cause little spikes, hikes and falls, causing an overall impact on the prices. Such patterns are also hidden in how brands can create monopolistic prices by having more dense ownership within a certain area (Bergeaud and Raimbault 2020).

2.2 Relevant Topics

A unique phenomenon about Western Australia is its fuel pricing cycles. It is a well-documented event that occurs every week of the year. This involves the fuel prices following a cyclical pattern in which it starts to fall from the start of the week while touching their minimum on Tuesday and then suddenly touching their maximum as fuel suppliers wish to cut back on losses incurred on the prior day. These patterns are driven by competitive dynamics among both retailers and market forces. Studies highlight the existence of “Edgeworth price Cycles,” where prices rise and fall in a predictable manner over a certain period (Wills-Johnson 2008). This cyclical behaviour is linked to competition that is found between different retail brands and independent stations that are setting price benchmarks that needs to be followed by the market to avoid any share of customers (Ghazanfari 2020).

Perth, having characteristics of being geographically isolated contribute to the price-setting patterns of retailers. Due to its distance from major cities like Melbourne and Sydney, well-established brands like Caltex-Ampol tend to be the price leaders that determine the price early on which then needs to be followed by smaller independent stations to remain competitive (Wills-Johnson 2008). This research can be better implemented into a filtered area of Perth

Metropolitan suburbs to show how time-series and visualisation techniques can be utilised to see how differently each area responds in respect to this cyclical phenomenon.

Research has also shown that fuel price is also a subject of variability due socioeconomic and geographic characteristics of an area. There could be factors like population density, income levels and even local demand that could have an impact on what the price of the fuel is on that date. A study on spatial variability of fuel prices in the United States discovered that areas that were densely populated and areas where the income levels were comparatively higher tend to have a higher average of prices (Bergeaud and Raimbault 2020). Such findings can be tested on the metropolitan suburbs of Perth since it lies on a huge geographic area having different socioeconomic zones. The possibility of different consumer behaviours, transportation needs, and station density could lead to a differentiated price determination standard.

It is also noted in other studies that due to an area being geographically isolated, there is a higher probability of fuel stations taking advantage of this to charge a higher price. This does not only mean in a certain area but also in a city. Perth having such characteristics allows larger brands to exploit the situation (Ghazanfari 2020). On the other hand, any geographically isolated area also tends to have higher transportation and distribution costs, leading to higher prices being passed down to consumers.

To keep such fluctuations and exploitation under control, government regulations are intervened to promote a healthy price determination. It has also been implemented in historically in Western Australia to keep the system standardised. However, studies prove that the results of these policies have not always come out as expected. For instance, a regulation introduced to reduce the frequency of price changes had little or no impact on the over-price levels (Dewenter, Heimeshoff, and Lüth 2017a). Hence, it shows that policies like taxation, or restrictions may have short-term effects on fuel prices, but it generally tends to find its way back to normal eventually.

Like every other city in the world, Perth is also not insulated from global price shocks. It’s evident from the 2007-08 oil shock with which there was a global rise in the demand with stable production causing a significant increase in retail prices. Studies show that such situations lead to a reduced consumption and even economic slowdowns for cities who are heavily reliant on transportation and logistics (Hamilton 2009). Hence, any discoveries with local factors can never be analysed alone. Global factors will also have a say in how prices are being determined.

Another external factor that has influence on the determination of fuel prices is the environment itself. This can be subdivided into how weather plays a key role in settling the demand for fuel. Particularly, temperature and precipitation play a significant role in the consumption patterns of fuels ultimately impacting how it is priced (Doğanlar et al. 2024). As temperature rises, the demand for fuel does as well since in transportation sector, vehicle consume more fuel to power the air conditioner which leads to a spike during hotter months. Increased rainfall is another factor since this lower down the amount of transportation in an area leading to a lower use of fuel. Ultimately, these “Weather Shocks” both affect the demand for fuel and lead to shifts in daily price averages (Phan 2023).

3 METHODOLOGY

3.1 APPROACH

The approach used to carry out this research is entirely based on secondary resources. The research adopts a quantitative approach with certain scenarios explained through qualitative understanding. The justification behind this choice is due to the nature of data being mostly numerical. Fields like fuel prices, temperatures, population etc. can all be statistically analysed to uncover hidden patterns and relationships. Since the research is based on sample data inferring about the entire population, this type of research is more applicable since it can be generalised over a larger population.

A quantitative research approach allows analysis to be completed over a large dataset. These can also be easily integrated with other datasets to find any correlations between other factors and the fuel prices. Particularly for this project, the primary data source of fuel prices had over 9 million data points which were easily handled through a data frame built in Python using Pandas. Numerical data also allows objective framework for measuring the effects of other variables on the target variable. This connection can help in finding hidden relationships between different variables that cannot be seen through raw data

The core of this analysis lies in understanding the effect of other variables on the target variable being the price of ULP91. This is achieved through multiple analytical techniques which rely on statistical measures. Time is another important consideration that helped in creating numerous time series charts to find insights into the data by studying the change in patterns over time.

Another approach taken for further depth in the analysis was the incorporation of the Geospatial Analysis. This refers to the use of Shapefiles, longitude, latitude and other mapping techniques to interactively visualise scenarios like station density, brand clustering and population characteristics. This allows patterns to be found which are not easily recognisable through data in tabular form but can be seen using coordinates on a map.

This study uses Python Programming language as the primary tool for data analysis and visu. There were numerous imports of libraries in different notebooks to acquire the required functions for analysis purposes. For example, Pandas for manipulation, GeoPandas for geospatial analysis, ipywidgets for interactive dropdowns, folium for visualised mapping, Matplotlib for graphical visualisation, etc. Python was selected because of its robust capabilities to handle huge datasets and the ease through which it can engineer different datasets together. The IDE used throughout was VS Studio Notebooks were created for each section of the code to run Python scripts and simultaneously debug any error that occurred.

3.2 DATA COLLECTION

The data used within this study was collected through publicly available datasets on the web. It was gathered with a combination of manual searching and automated APIs or scraping. The aim was to find that is accurate and fulfils the temporal requirement of 2020 to 2024. For this reason, this process was more inclined to use governmental data compared to third-party sources due to its reliability and robustness.

The primary data for fuel was collected through the government-run platform for Western Australia, called FuelWatch. It provides real-time data about fuel prices for different variants around the whole of Western Australia. This website has evolved to promote transparency allowing people to take strategic decisions by comparing fuel prices between brands and locations before purchases. This website was able to provide CSV files for each month of every year. Each file had information about its Published date, the type of fuel, brand, type of fuel, its region, suburb and its postal codes. This data source is reliable as each station must submit their prices to FuelWatch by 2 pm each day for the next day. These prices are then locked up for the next 24 hours starting 6 am the following day. This is a requirement by law to benefit consumers to plan their fuel purchases before time while being confident that these will not be fluctuating. There were 12 files for each year with 5 years in total. All these 56 files were engineered together using Python Scripts.

Once the data collection for fuel prices, date, and location was complete, there were many missing fields that needed to be filled up to find further insights into the research. One of them was to see visual maps to find patterns of clusters and density. For this an approach related to Application Programming Interface (API) was used (Gordon and Rudin 2022). This allows the retrieval of data by sending a request to a website server and in return getting the data in a structured format. For locations, longitude and latitude coordinates were required. Therefore, the Google Maps API was used utilised to retrieve the grid references of all the fuel stations that are being studied. This paved way to do further geospatial analysis as with such information it was possible to make interactive maps and heatmaps. Google Maps API was also integrated with GeoPandas and Folium to make these chart more interactive and user-friendly.

Once the primary file was fully collected and engineered, the focus was then shifted to other factors that need to be evaluated based on the hypothesis discussed in the next chapter. First one is the Public Holidays of which the data was collected through another publicly available governmental dataset. This was fetched from the website where only the Western Australian Holidays were filtered for a csv file that was later cleaned further. The data included fields related to the exact date the holidays incurred from 2020 to 2024, their categorical name and a brief description about why it is celebrated.

Another important factor explored during the analysis was the relationship between weather conditions with Fuel prices. The data for this was difficult to collect as it was needed to be precisely conveyed for every day from 2020 to 2024. Also, as the geographical area is too vast to generalise, the data was required to present an averaged weather condition of the entire suburb. This was made possible through an API called Open-Meteo, which is a free platform to retrieve information about weather conditions based on the latitude and longitude coordinates of an area. Upon creation of an account and fetching API key, this significant information was extracted from a high-resolution national data service provider that allowed integration of currently available data about fuel pricing dates with the weather conditions of that suburb. There were temperature and rainfall being assessed and for this, its maximum, minimum and average were retrieved for further analysis.

With many local and regional factors analysed, the focus was then shifted towards the macro factors. One of them was the analysis of Interest rates with the aim to find if there is any hidden relationship between this of and the target variable. The data was collected from the official website of Reserve Bank of Australia (RBA) where it was publicly downloadable. This online resource offered historical and current interest rate data in a CSV file that was conveniently engineered together with the primary data frame to move forward with the analysis.

Another macroeconomic factor that had great potential for correlation with fuel prices was inflation itself. This data was collected from the official platform for accessing public datasets in Western Australia called DATA WA. It provides a huge repository of a wide range of datasets that are uploaded by different departments of the government. The categories range from finance, health, transportation, infrastructure and many more. The aim of this platform is to provide transparent data to the public so that data making decision making can be carried out with the views of public being considered. The file for inflation was in CSV format. However, it was thoroughly filtered to find only the relevant fields of consumer price index and year-end inflation rate can be kept.

Taxation rates were perpetually changed during the analysed years to govern the fuel prices into a more affordable commodity. This factor was analysed through the dataset obtained from the Australian Bureau of Statistics (ABS) which is Australia's official national statistical agency offering data and insights on numerous topics including economy, government, environment and population. The data provided is highly reliable as this is also made use by policymakers. The tax rate file obtained from here had countless fields showing excise rates for all the commodities being taxed in Australia. Through Python script, only tax rates for oil that is used for cars and other motor vehicles were extracted for further analysis.

To find if there is any relationship between the densities of the population in a certain area and the fuel prices being charged in that area, the data for every suburb was required. This was made possible through another API called GeoNames which is a repository of geographic databases that provide detailed information about locations around the world. It offers data particularly related to population and its statistics. It requires the field of coordinates to be filled before it fetches the most current data about the population of that area. Google Map API was used to retrieve the longitude and latitude of each suburb which was then further used by GeoNames to find the population for them for further analysis discussed in Chapter 4.

3.3 Data Processing

Once the data was collected, each file was processed individually before they were integrated together with data engineering techniques. Within the processing operation, the first way to clean the data was by finding any missing values within any of the datasets. Every field and row were checked prior to any analysis so that no errors or misleading analysis would come out of it. Any missing entries were treated different with different techniques for example, there were some coordinate data missing which were filled with imputation data. If a row had many missing fields, they were dropped to improve the accuracy of the results.

After this, a thorough check for duplicate values was done to eradicate any chances of multiple entries. Duplicate entries particularly for fuel stations and dates were identified and removed to make sure that there is no data redundancy found. Next the data was transformed into a uniform standard so that the next step of engineering can be more convenient. Majority of datasets downloaded did match the dates that were required but the formats were usually different. For this reason, a consistent date format was copied from the PUBLISH_DATE column from the primary file of fuel prices and pasted on all other factors.

Finally, each dataset file was individually scaled down to reduce memory usage. This was done by filtering only the relevant details required to conduct the analysis and letting go of the columns that were of no use. An example of this is from the primary file of Fuel prices in which there were an extra 4 types of fuel that were not the scope of the investigation. These were removed, helping in reducing a huge amount of data points, which would have used a lot of computational power if analysed. An example of this could be from the tax file, which offered excise rates for all commodities being offered within Australia. Since the study was only about fuels, the filtering reduced the huge dataset to only a few columns. This was then further diminished by keeping only the excise tax that matched the most to our target type of ULP91. There were also instances when the datasets had to be filtered down in terms of the temporal adequacy being offered. For example, the inflation dataset offered historical data touching before 1940s. This has no impact to the research due to which it was filtered down to only the year starting from 2020 and ending in 2024 (August).

3.4 DATA ENGINEERING

One important aspect of this project was the connection between all these datasets so that analyses could be carried out simultaneously with each other. This required the core task of engineering where processed datasets were merged with a common field “Date”.

This process ensured that each day’s fuel price was accurately matching with the information provided by other datasets. For example, the fuel prices of a particular date should give information about what weather conditions were there, and at that point, what was the inflation rate, excise tax and the interest rate that was being charged so that a combined analysis can be done.

Engineered datasets into a common pool allowed robust analysis across both time and space which made it possible to uncover insights that would have been impossible to find if dealt with in a different manner. Data Engineering allowed the whole analysis to connect with each other defending discoveries that were made throughout the research period.

3.5 DATA HANDLING

All data was processed and managed through Visual Studio as the primary IDE. This was done in multiple notebooks being created to reflect each step in the analysis process. The kernel used within these notebooks was Anaconda for Python. The data is organised in different folders categorized by datasets. For example, the name of the dataset of Fuel Prices reflecting the output from the notebook number. All cleaned datasets were stored as CSV files which were then imported directly into the notebooks when needed for a particular analysis to be carried out.

As all data is publicly available, such as fuel prices from FuelWatch, weather data from Open-Meteo and population from Geonames, there was no significant concern about data privacy or security breaches. The datasets were open for use and analysis, which is why there were no restrictions on compliance requirements, making concerns for security at a minimal level.

4 RESULTS

4.1 OVERALL ANALYSIS

In this section, the overall analysis of the entire primary dataset of fuel prices is discussed. It was initially analysed using Python scripts of descriptive statistics. The data reached around 972000 data points that reflected information of fuel prices in Metropolitan suburbs of Perth from 2020 to 2024. It shows an overall average of 161.8 cents per litre with a significant variability of 31.1 cents of standard deviation. The minimum price in all years was 76.7 cents, while the highest turned out to be 239.9 cents. The median price for all years combined was close to the average, which is 167.9 cents per litre. The below graph shows a histogram of the fuel price distribution from the period being analysed.

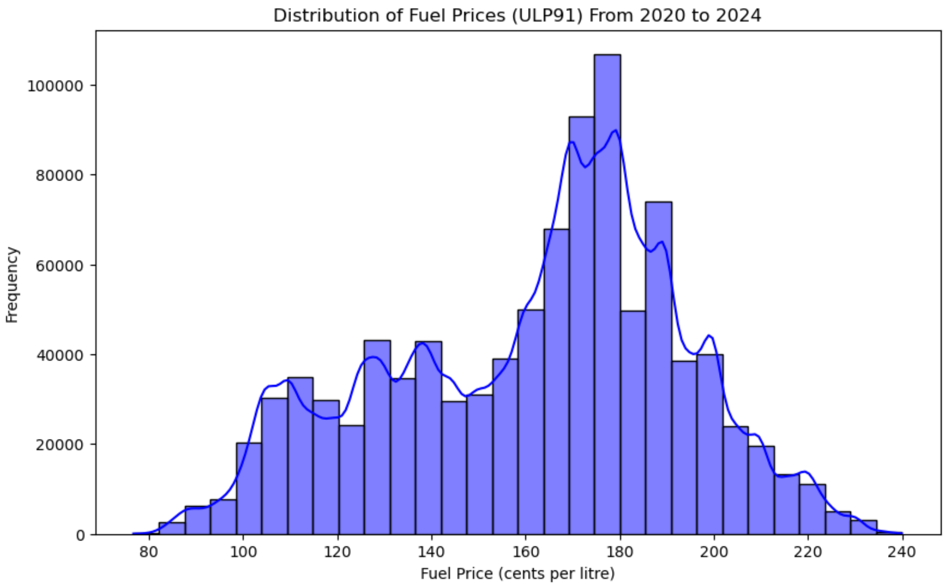


Figure 1: Distribution of Fuel Prices (ULP91) from 2020 to 2024

The above shown distribution shows a rightly skewed graph with most of the prices between the range 160 cents and 180 cents. There are also notable peaks at the point of 180 cents which shows that this price range was most common over the years. Prices below 120 cents or above

200 cents have a lower percentage of occurrence, indicating that over these years, there were only a few periods when there were extreme values like these.

Since, these measures show a combined result having little meaning, a further detailed trend analysis was done to understand the data better. A weekly average fuel price graph was created to see the overall pattern in a smoothed-out version compared to the daily graph.

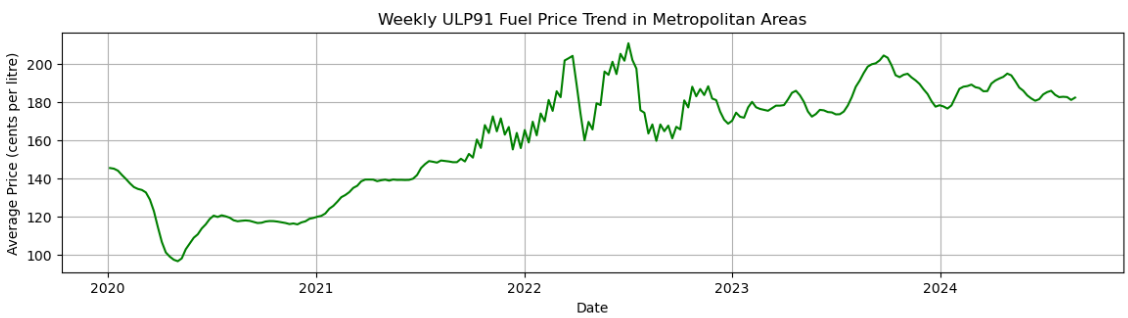
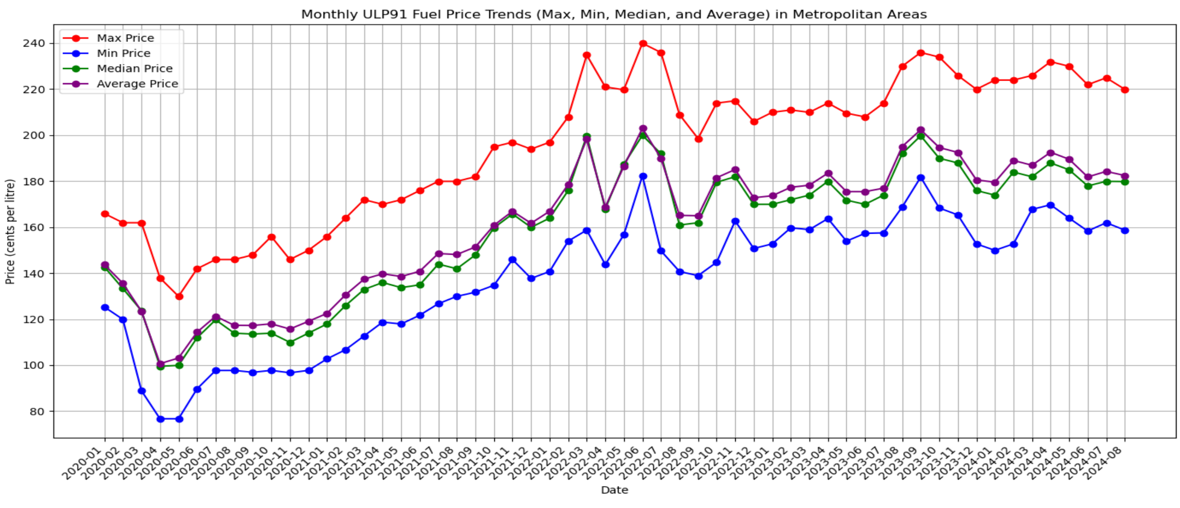


Figure 2: Weekly ULP91 Fuel Price Trend in Metropolitan Areas

Figure 2 shows that fuel prices have generally increased over the years, whereas 2020 shows a relatively stable graph, which eventually starts rising whilst continuing into 2021, which can reflect the period of COVID-19 lockdowns ending and life going back to normal routines. There are noticeable spikes in the year 2022, which are generally due to the global events taking place, for example, the Ukraine war. From mid-2023 onwards, the trend stabilises, and prices show occasional peaks and troughs that are a product of general changes in global oil prices and fluctuations.

Another attempt to find further smoothed-out details of the fuel prices was carried out by analysing the data based on monthly basis.



;

Figure 3: Monthly ULP91 Fuel Price Trends (Max, Min, Median, and Average)

In this analysis, monthly trends of ULP91 were displayed using all four statistical measures: maximum, minimum, median, and average prices. This was done to find insights without relying solely on averages as it can mask some of the important characteristics of the data like ignoring max and minimum values that can be considered outliers. By capturing the maximum and minimum monthly values, the complete range of variability can be captured which helps in highlighting price hikes and drops due to several reasons.

The median price is an important statistical tool which is the middle value of the data set and hence it is much less sensitive to extreme values when compared through a time series graph. In the case of this dataset, the median and average prices did show a similar and overlapping pattern which suggests that the data distribution is relatively symmetrical with little or no extreme skewness in the results. It reflects the fact that due to not having too many extreme outliers, the average itself shows consistent results which can be a good measure for analysis in the future.

Using all four measures on the same graph shows a deeper understanding of the variability in prices and what consumers faced during this period considering if there was a dramatic peak or drop. The pattern is similar to the above-discussed reasons for weekly analysis which will be further discussed through individual factors in the next section.

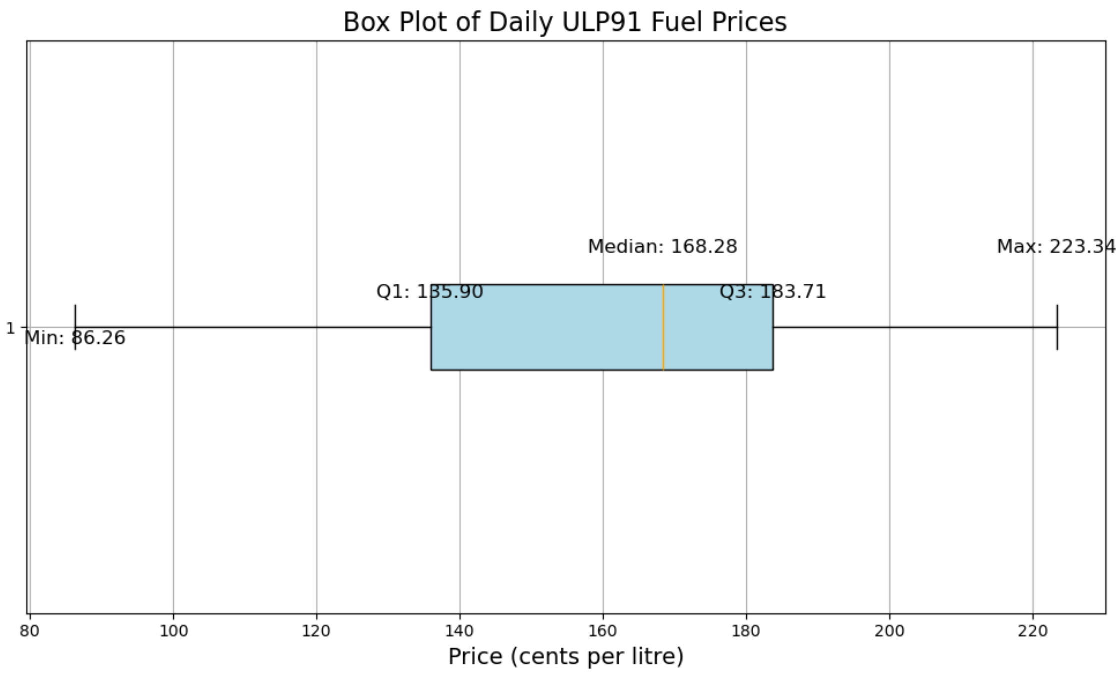


Figure 4: Box Plot of Daily ULP91 Fuel Prices

To analyse the distribution of prices on daily based, another type of statistical visualisation was carried out. It's a highly effective technique for identifying the spread, central tendency and the variability within the given data by highlighting the most important aspects like max, min, median and interquartile ranges.

The above graph reflects the same results discussed earlier for max, min and median values. However, this goes a step ahead to discuss the IQR (Inter Quartile Range). This spans from 135.90 cents in Q1 to 183.71 in Q3, with 50% reflecting the median values. It once again suggests a lack of extreme values in the dataset, implying most prices remain in the expected ranges. It is still to be noted that there is a wide gap between maximum and minimum values, which does reflect a higher variability but nothing that can be seen as something too important as the standard deviation is moderate.

4.2 ANALYSIS OF INDIVIDUAL FACTORS

4.2.1 WEEKDAY PRICES

As discussed in the previous chapter, Western Australia is known to have a cyclical pattern for following fuel prices. However, for this research, the hypothesis is based on the idea that fuel prices vary depending on what day of the week it is. This experiment is known to be done specifically for metropolitan areas of Perth. The assumption is made from the common observations where petrol stations tend to reduce price in the start of the week (typically Monday and Tuesday) and increasing towards the end of the week (Thursday and Friday). Hence the hypothesis for this section is **"Different days of the week will exhibit significantly different fuel prices."**

To test this hypothesis, there were analysis done to find insights into the data. There were multiple visualisations created to find this hidden trend. One of them is shown below

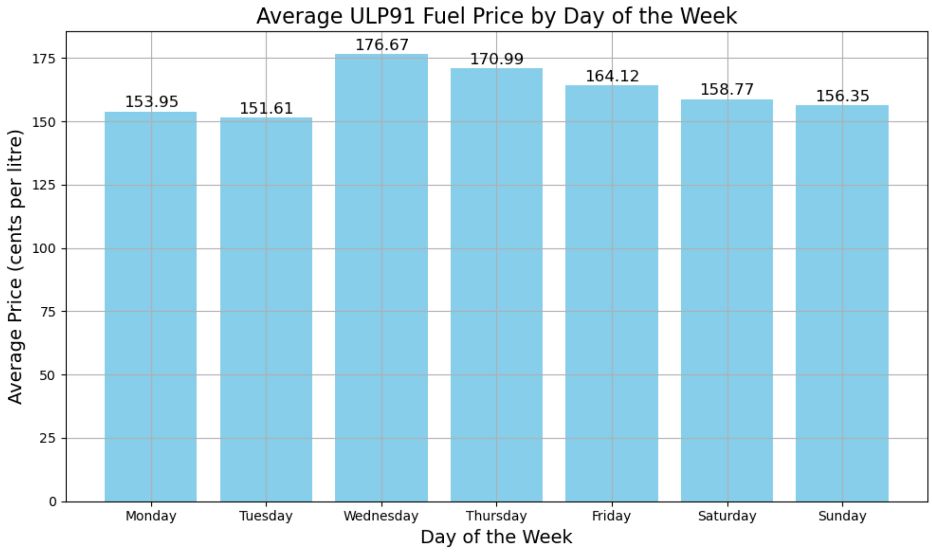


Figure 5: Average ULP91 Fuel Price by Day of the Week

In this illustration, the average price of ULP91 fuel type is shown on a weekly basis for all years. It reveals a clear midweek price hike and a lower price early in the week. The bars show that Monday and Tuesday often have the lowest prices at 153.95 and 151.61 cents per litre. Midweek prices peak when it reaches Wednesday, as it touches almost 176.67 cents per litre. This is the highest average day of the week which then decreased on Thursday. Weekends show a moderate average of around 157 cents with the overall pattern suggesting that stations do indeed increase prices when consumer demand is higher in the midweek and then lower it again to attract drivers on weekends.

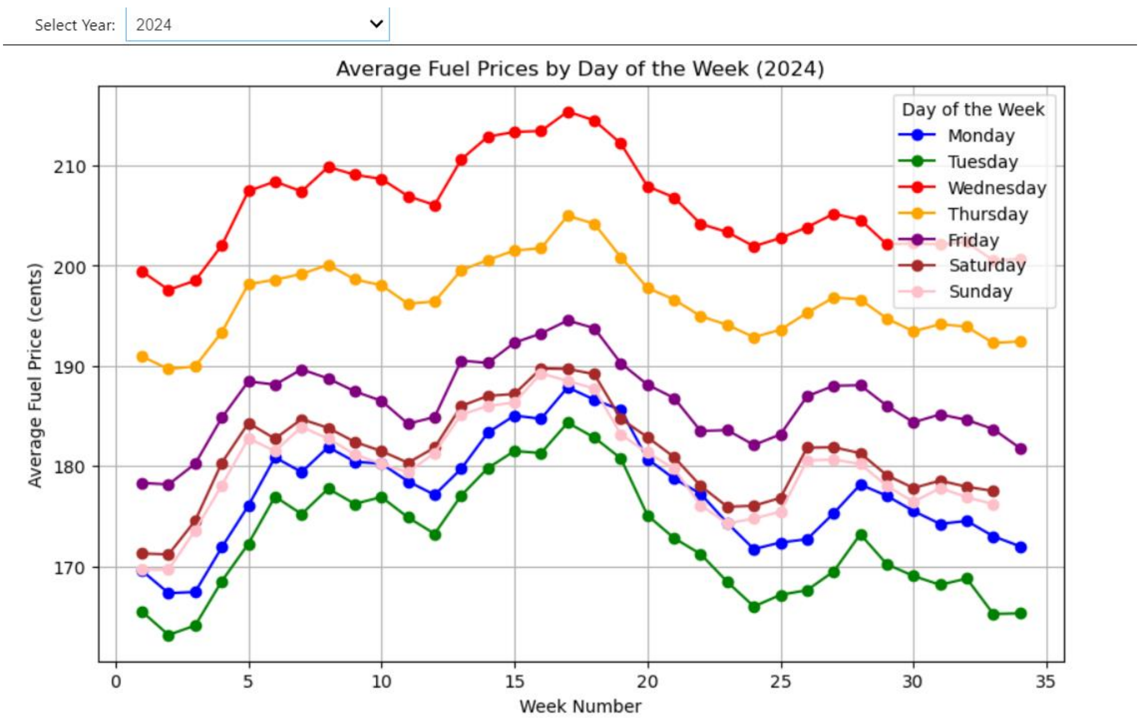


Figure 6: Interactive Plot for Fuel Prices by Day of the Week (2024)

For further analysis, an interactive plot was created to find deeper insights into each year. The graph particularly follows the analysis of the bar histogram above and represents the average fuel price as per the day of the week for 2024. It can evidently prove the point that Wednesdays consistently show a higher price with the average being more than 200 cents per litre for the entire year. This is closely followed by the Yellow line representing Thursdays. On the other hand, Monday and Tuesday maintain the lowest prices, with Tuesday hovering at the lowest, around 190 cents a litre. Additionally, the graph highlights periods where there was stability and when there were high fluctuations, offering insights into how fuel stations strategise around the week. Similar analysis is also done for median prices of different years. An example chart is given in Appendix 2 of the report, which ones again reflect a similar pattern to average prices.

To dive deeper into each day, a monthly box plot was created for each day of the week to analyse the outliers and further trends over the years. This can be seen in Appendix 1 of the report. From the analysis, we can confirm the same fact about prices being different on different days. It does also further give insight into outlier analysis with months like June, July and August where prices spike for certain on days like Saturday and Sunday. Hence, compared to other days some fuel stations tend to charge a higher price on weekends in these months. This could be a way for retailers to cut down on losses by getting the most out of consumers for an increased demand on Weekends.

In conclusion to this analysis, there is enough evidence that supports the initial hypothesis that fuel prices do vary based on day of the week. Even after using different charts, time and analysis, it can be concluded that Tuesday is the day when the prices touch the minimum. This is followed by Mondays, then weekends and finally the latter half of the week being Thursdays and Fridays. Hence, in conclusion, it is advised to get fuel usually at the start of the week, preferably Tuesday, and always avoid the latter half of the week, which is typically Thursdays.

4.2.2 Regional Differences

The primary data for this analysis is divided into three different regions for Western Australia. These include the North of the River, South of the River and the East Side Hills. Therefore, to analyse it further, the hypothesis of this section is if the fuel prices vary significantly across different regions. The assumption is that based on differences in topography and other characteristics of these regions, the prices may also be different. The analysis aims to identify whether each regions exhibits distinct enough factors that allows a change in pricing patterns.

Hence, the hypothesis being tested in this section will be **“Different regions will exhibit significantly different fuel prices.”**

For this purpose, the investigation started with a general fuel average across the time frame being analysed for each region. The chart is shown in Figure 9.

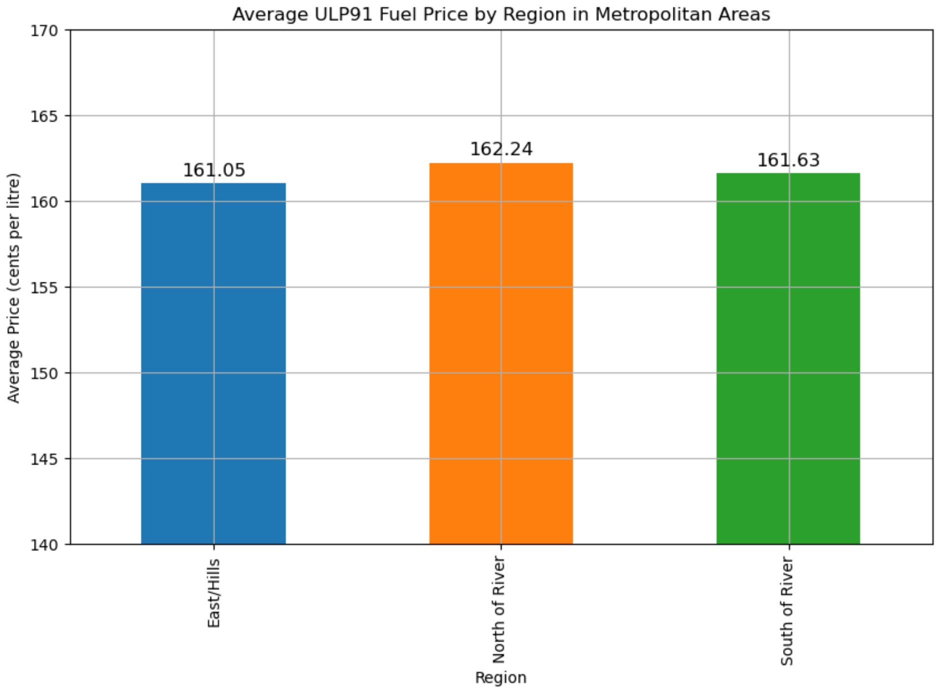


Figure 9: Average ULP91 Fuel Price by Region in Metropolitan Areas

The graph compares the average prices for all three regions discussed above. The data shows that North of the River, has the highest average fuel price of 162.24 Cents per litre which is followed by South of the River with 161.63 and then East/hills of 161.05 Cents. Even though the differences are not too large, but it does support the hypothesis that regional factors to play a role in affecting how prices are being determined in different regions. It reflects that there could be socioeconomic factors along with different demands structures that may be leading to such a difference. However, since an overall average can be a bit misleading, below is a box plot for the latest year.

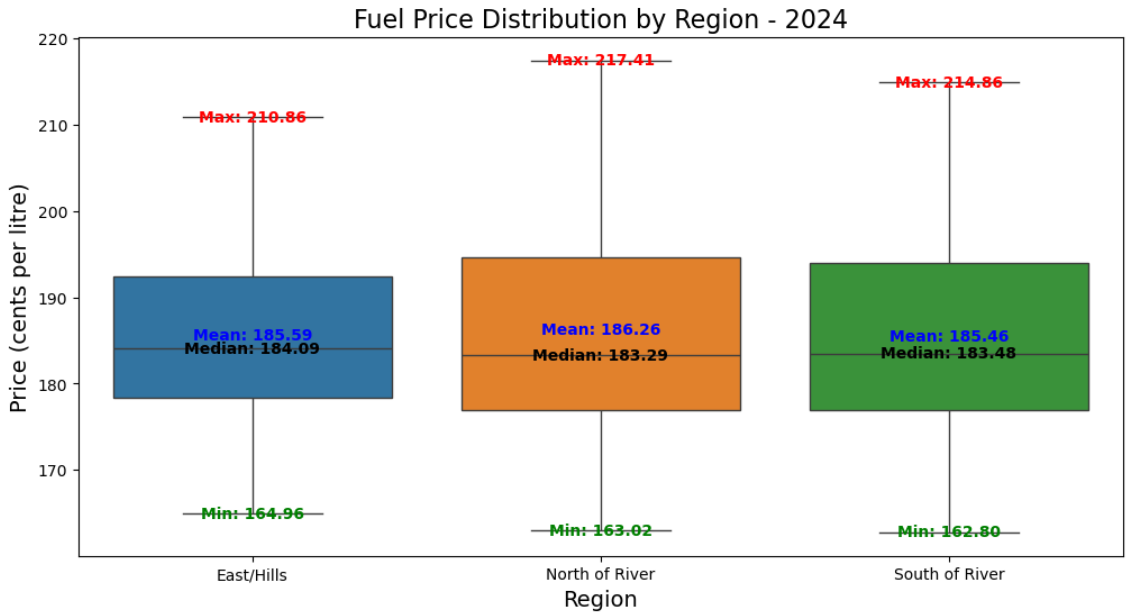


Figure 10 : Box Plot for Different Regions (2024)

From the box plot analysis, the research was done on a deeper level with each year being considered at a time. As an example, the report shows the latest year of 2024 in which, the trends continue to be the same with the highest mean for suburbs North of the River. This also shows that the maximum in the entire year compared to the other two regions is again North of the River, touching a high of 217.41 cents per litre. the East/Hills continues to show a relatively lower average as it also got the minimum price among all three regions of about 164.96 Cents a litre. The median prices across the region are close to each other, which proves that based on variability, Suburbs that are North of the River, experience more highs and lows with a wider range of price. On the other hand, Suburbs in the south and Hill side experience a more stable pricing strategy. For this purpose, another analysis is done to depict this on a Time series graph. This is shown below.

Another graph in appendix 3 of the report shows the monthly average price differences between regions over a time from 2020 to 2024. It is included in the analysis to find the peaks and drops for different regions at the same point in time. We can derive that many peaks and troughs are overlapping however, for North of the River the prices are most consistently higher. The variability in price difference does suggest that this could be due to market conditions or seasonal factors. The divergence at different points from similar patterns does reflect that local economic reasons do play a role in how prices are being determined.

For this reason, price change was required to be investigated, for which a scatter plot was created that is shown in Figure 12.

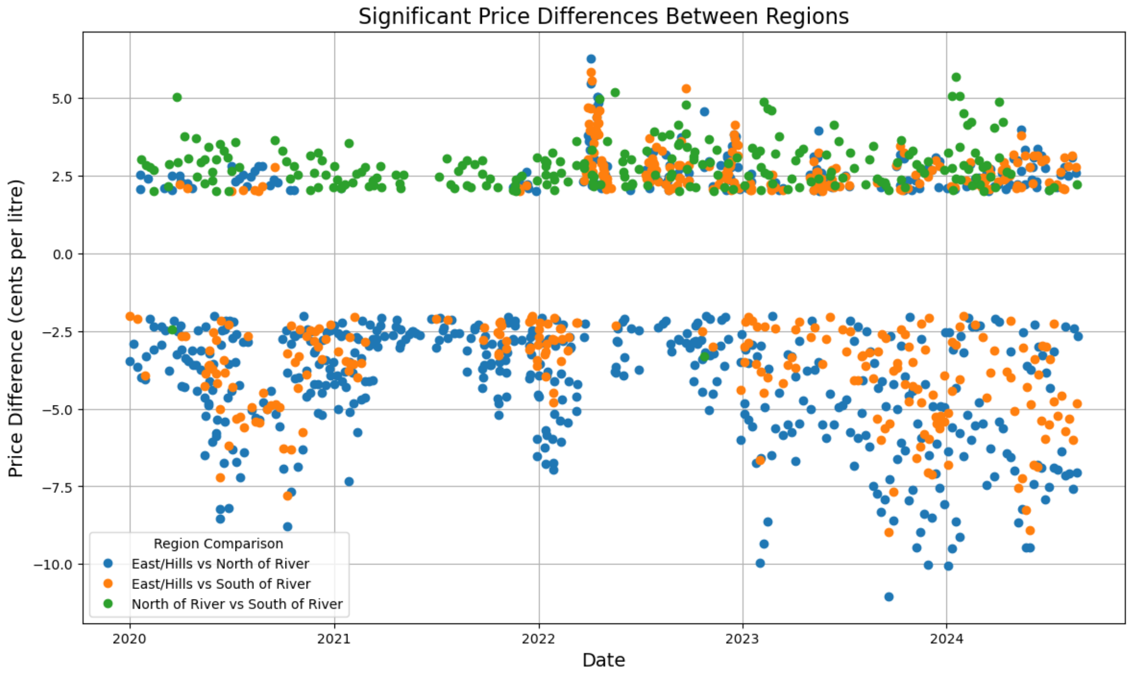


Figure 12: Scatter Plot for Significant Price Differences in Regions

This scatter plot shows that the dots above the 0 line represented all those instances where the first region in comparison is more expensive than the other. For example, through the green dots we can see that there is majority of positive price changes for North of the River compared to the South of the River. On the other hand, the blue dots are going in negative showing the opposite since East Hills is being compared to North of the River. Therefore, we can see that times there were huge differences in the prices specifically in 2024.

These price change difference can be a subject of many factors that are happening regionally. Few of them that will be the topic of discussion in the next few sections like brand clustering, population density and station density.

In conclusion, the analysis of regional differences does provide enough evidence to support the hypothesis that there are differences in how prices are determined in different regions. All above charts and discussion shows that North of River is consistently charging higher prices while the exact opposite is happening with East Hills. Additionally, the variability in price differences between regions does suggest that there could be other factors into play that are specifically related to the characteristics of the region. For example, competition, demand supply logistics etc. Further in this chapter, we will be discussing few of these in detail to find what regional factors can contribute to this found insight.

4.2.3 FUEL STATION DENSITY

Continuing the discussion from the previous analysis, there could be regional factors that play an important role in how prices are determined. Such characteristics can impact how the forces of demand and supply work in that area hence determining the fuel prices. Therefore, the analysis moves further in finding whether the having more fuel station in a particular region, area or suburb has an impact on how prices are determined. The hypothesis for this section is **"Areas with a higher density of fuel stations will have higher fuel prices."**

Since we are only considering certain metropolitan suburbs of different regions, an attempt was made to find the exact area of the chosen suburbs. This was made possible through the GeoPandas API which was able to extract the surface area of the selected suburbs using the Longitude and Latitude of the suburbs generated from the Google Maps API. The following table shows the area of each region.

Table 1: Regions and Their Respective Areas

Region	Exact Area (km²)
North of River	1206.27
South of River	1422.59
East/Hills	1050.81

Considering these areas the quantity of Fuel Stations available in that particular year, a correlation analysis was conducted to find the results of the relationship between them. It came out to be close to 1 (0.86371) which suggests that there is a strong relationship between higher station density and higher prices. From research point of view, the more stations there are in an area the higher the competition and hence the lower the prices would be. However, in this case the data suggests quite the opposite. This could be due to the following reasons

- Price Control: In regions with higher fuel station density, it could be a possibility of clustering taking place. Retailers and brands trying to target areas with consumers who are less price sensitive to increase overall profits.
- High Traffic Areas: There is also chance that regions that are more urban or have higher traffic levels will have consistently higher demands of fuel. Hence, stations may feel a

lesser pressure to lower the prices since the purchasing is also high, irrespective of the prices being set.

- Operational Cost: It is possible that due to having open fuel stations in a densely populated area, the operational costs would be higher too. The government may also be charging higher taxes or the overall cost of service there could be higher due to labor and rent cost. These costs are transferred to consumers in the form of higher prices which may reflect the reason why more stations have higher prices.

To prove the point further, another statistical measure of the median was also used to see if there might be any changes compared to averages. However, the results once again reflect the same scenario. The high-density areas were separated from those that had lower density. From the analysis the higher-density areas had average price of 162.24 Cents per litre whereas the lower-density areas had an average of 161.34 Cents a litre. To analyse further standard deviations were taken out from the average price for each region for each year. It is shown in the table 2.

Table 2: Regional Fuel Price Analysis from 2020 to 2024

Year	AREA_DESCRIPTION	Price_Std_Dev	Avg_Price	Num_Stations
2020	East/Hills	16.30611	118.4327	54
2020	North of River	16.94611	119.487	166
2020	South of River	16.8021	118.9659	190
2021	East/Hills	17.83223	145.1292	59
2021	North of River	18.47532	146.0372	198
2021	South of River	18.32695	145.466	221
2022	East/Hills	17.46823	180.3756	58
2022	North of River	17.83869	180.4863	187
2022	South of River	17.88752	179.573	223
2023	East/Hills	14.73339	183.4028	68
2023	North of River	15.80607	184.3133	189
2023	South of River	15.84874	183.5366	221
2024	East/Hills	14.12584	185.5897	60
2024	North of River	15.37538	186.2571	191
2024	South of River	15.36452	185.4665	220

From the data shown above, it can be clearly deduced that the trend for average prices for all the regions is upward. However, with increasing stations over the year, South of the River did not lead to a decrease in prices. The standard deviation indicates that price variability was higher in the earlier years, particularly in 2021, but slowly decreased over time, showing increased price stability across all regions as the market matured.

For diving in deeper, an interactive visualisation was made using Folium library and Shapefiles extracted from the internet. This shows graphs of areas divided into the three regions discussed above. Also, the charts can be chosen for different years depending on the analysis. It shows clusters with prices higher than the average of the region with different colors. An example graph is shown in Figure 13.

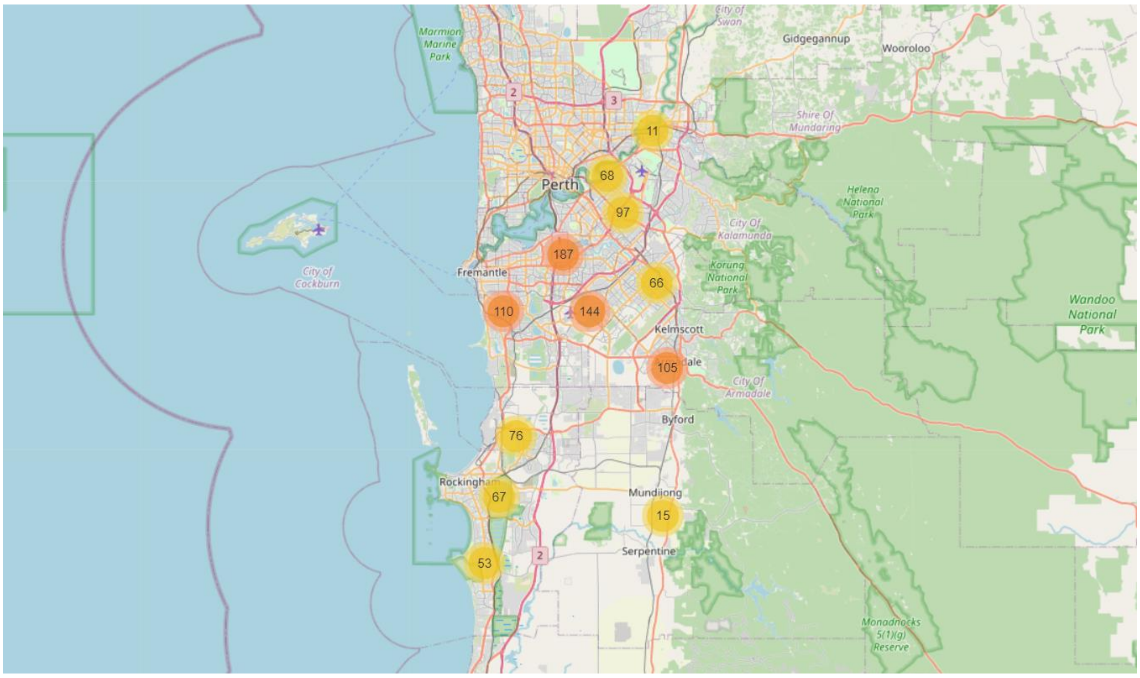


Figure 13: Interactive Graph of Cluster Averages (South, All Years)

As seen from the chart, there are different clusters formed in the South of the River. The ones in orange color show charging a higher price compared to the average of the region. More numbers in those clusters show that, indeed, prices are higher for more areas where there are more fuel stations. This analysis was also done for other regions. Another example of North of the River is given in Appendix 4 of the report.

Similar to the above graph, this graph shows clusters of fuel stations in the region of North of the river within 2024. The graph once again presents the same facts about more fuel stations leading to higher price charging. Zooming it in on the IDE environment shows exactly how even smaller clusters are contributing towards a higher price average.

Overall, the analysis of fuel station clustering clearly demonstrates that regions, areas or suburbs with a density of fuel stations lead to a higher average price. Despite assumptions of competition bringing price down, there are other factors like operational costs or demand/supply that play an even vital role in determining prices. Therefore, there is enough evidence to suggest that the hypothesis is true. It reinforces the idea that competition can affect the pricing of fuel; however, there can be many other factors that could act as being more important than the competition itself.

4.2.4 POPULATION

Since the last analysis discussed the potential of how higher density of fuel stations can lead to higher prices. This section of the report discusses another social factor related to how population density can have an impact on the fuel prices being determined. The assumption stems from the expectation that with higher density of people living in one area may mean there is a higher demand of fuel. This could lead to retailers charging a higher price causing the overall average of that area to rise. Therefore, the detailed hypothesis of the report will be "**Higher population areas will have higher fuel prices due to increased demand for fuel.**"

To conduct this analysis, the primary file was engineered with API-generated data to add population of each suburb. This was extracted through a Geonames which is an extensive geographical database that provides population data for more than 11 million locations worldwide. By feeding it with latitude and longitude of each suburb, extract from Google Maps API, Geonames was able to provide the latest data for each if the 203 suburbs being considered for this study.

Once the required data was engineered into a data frame, a correlation analysis was carried out between the population size and the fuel prices specifically for the year of 2024. This relation showed an unexpected result of correlation coefficient coming out to be very close to 0 (-0.328). This indicates that there is no meaningful relationship between fuel prices and the population in that region. With increasing population density in an area, it does not mean that the demand will lead to an increase in fuel prices. There are other factors that play a more vital role than population in determining the prices. In other words, from this analysis, it seems that that even though population size may intuitively look to affect the prices, but the data does not support this assumption.

To do a further in-depth analysis, 20 of the top suburbs were selected which present the highest average cost for the year of 2024 and their respective population data is plotted on a separate axis. This visualisation can be seen in Figure 15.

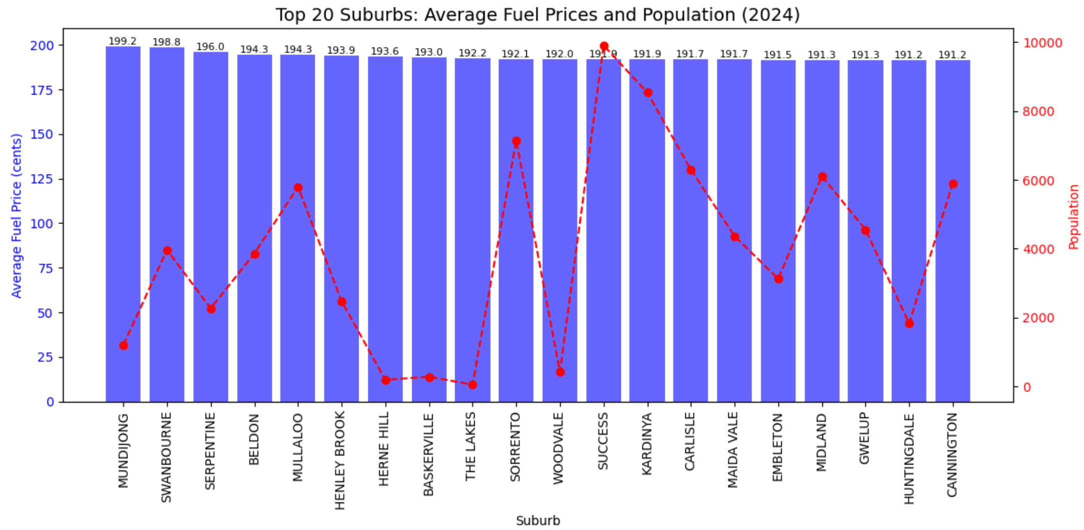


Figure 15: Top 20 Suburbs: Average Fuel Prices and Population (2024)

The above-shown graph provides a comparison between the factors being discussed (Population) and the target variable (Fuel Prices). The blue bars represent the average fuel prices of that suburb in the year 2024, whereas the red line displays the respective population figures. As seen from the top till the bottom, there is a variability of around 10 cents between the highest and the lowest. Mundjong shows the high average, touching almost 200, while Cannington is in the 20th position with a 191.2 Cents per litre average. However, if we compare the population of that area, we can see that even the Mundjong has more average the population is quite less. Similarly, in Cannington the results show the opposite of what was assumed through the hypothesis. Even though the population is quite high, the average fuel price in the area is still low. Despite the significant differences in the population, the fuel prices do not show any pattern of corresponding increase or decrease. This supports the notion of correlation analysis that even if there is a relationship, it’s not a significant one.

From the above discussion about population against fuel prices, there was no evidence found to support the hypothesis. It can be deduced from this analysis that population does seem to be an important factor, but in terms of fuel price, it is not a significant one that may impact the prices. There are other factors, especially the ones that add to the characteristics of the different areas, which can be regarded as more important in determining fuel prices.

4.2.5 BRAND CLUSTERING

Another important factor to discuss for differences in area could be if the fuel stations tend to cluster in a certain area which may cause fuel prices to rise. This could be because one brand is in majority quantity, which may allow them to charge a price lower or higher depending on the conditions. Hence, the analysis for this section would be to examine if brands creating a cluster can lead to an effect on prices. The hypothesis, therefore, will be **"Fuel stations clustered geographically and belonging to the same brand are likely to charge higher prices than their overall brand average, potentially exploiting their dominant market position within specific clusters."**

The following chart represents a high-level view of how many brands are being considered and the average pattern of how they price the fuel.

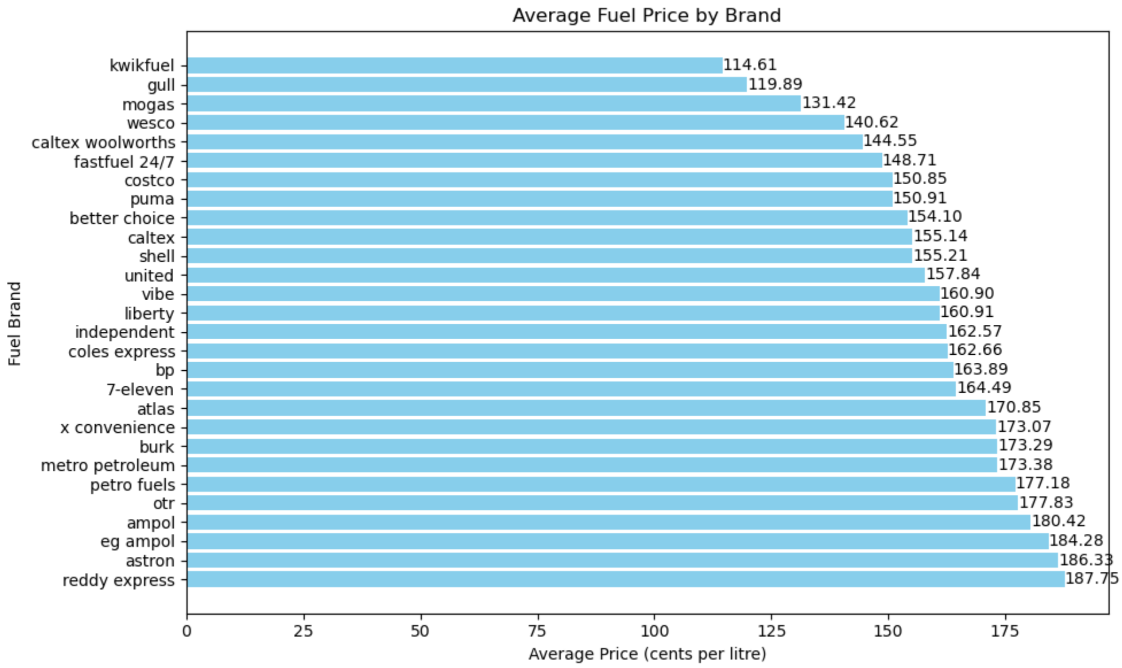


Figure 16: Average fuel Prices by Brand Names

This bar chart above in Figure 16 shows the first towards visualisation the patterns of being charged by fuel brands throughout the metropolitan areas of Perth. It presents a noticeable variation across the brands. At the lower end, we find Kwikfuel and gull that exhibit the lowest average fuel prices with around 114.61 per cent. On the other hand, the highest prices are charged by Reddy Express and Astron which touch around 190 Cents a litre. The big giants like Shell, Caltex, BP and 7 Eleven, sit in the middle, charging almost the same average. This shows a balance between competition and brand recognition. Their market strategies reflect consumer trust and an existence of competition which shows why they maintain a similar pricing pattern to keep their share of the market.

For in depth analysis there was another interactive map created with the use of Folium and OpenStreetMap that shows how in the brands are clustered in different regions of Perth.

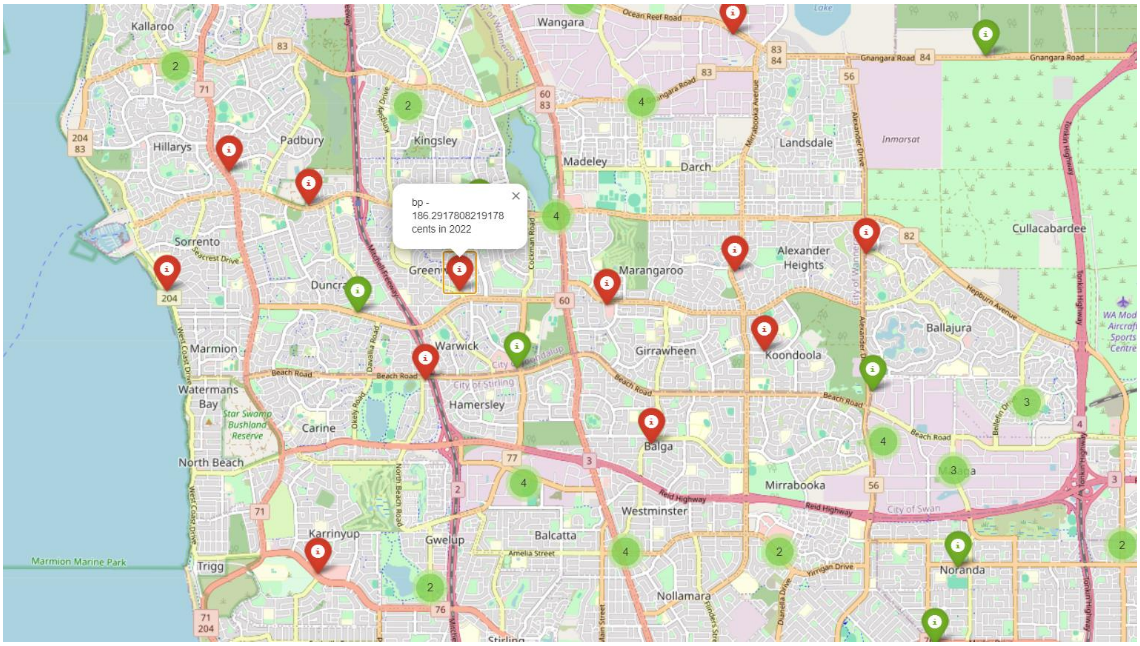


Figure 17: Interactive Map to find insights into Brand Clustering

This was done to find a visual pattern to see if there are any brands who are situated near to each other and are charging a higher price because of dominance in the area. The graph above shows the area North of the River in 2022. However, even after a deeper look, there was not much evidence collected to show that brands that are together can try to exploit the consumers by charging a higher price because of area dominance.

To take this analysis even further, K-means clustering was carried out using the DBscan library. This was done to group fuel stations based on their geographic coordinates (latitude and longitude). This technique was used to detect the areas in different regions where fuel stations are densely located so that it can identify clusters. The objective was to see if DBscan can isolate clusters and compare their pricing patterns with those in the regions to check for insights into whether distinct brands are taking an advantage from local dominance by charging a higher price.

From the analysis, there were a total of 51 Clusters created. The Results showed that Caltex and BP charged significantly higher prices than their overall average. For example, Caltex itself was dominant in clusters 1,6,13, and 38. In these clusters, it charged a much higher price (25

Cents) more than its average in other clusters where it was not in a dominant position. This suggests that Caltex may have been exploiting its position in these clusters where competition was not as high. BP on the other also did the same in multiple scenarios. For example, in Cluster 6 specifically, it charges 20 Cents more per litre compared to its normal price average.

In conclusion, the brand clustering analysis initially did not show any promising results, however with this more technical analysis of using DBscan and K means clustering method, the hypothesis can be supported by enough evidence to suggest that geographic concentration of brands can lead to a dominating position that could lead to them charging a higher price. This can be a valuable insight for consumers as well as the policy makers when considering their pricing strategies to see how spatial factors play a role in how the prices are being determined in the market.

4.2.6 Public Holidays

The following section of the report discusses in detail the relationship of Public Holidays with Fuel Prices. The dataset combines the fuel price data with public holidays that are enjoyed by Western Australia. The analysis is divided into two hypotheses where the first one will explore **"Fuel Price are higher on Public Holidays than Normal Days."**

To start the experiment, the first general visualisation produced is an average of fuel prices on Public Holidays compared to those not on public holidays. The graph is shown in Figure 18.

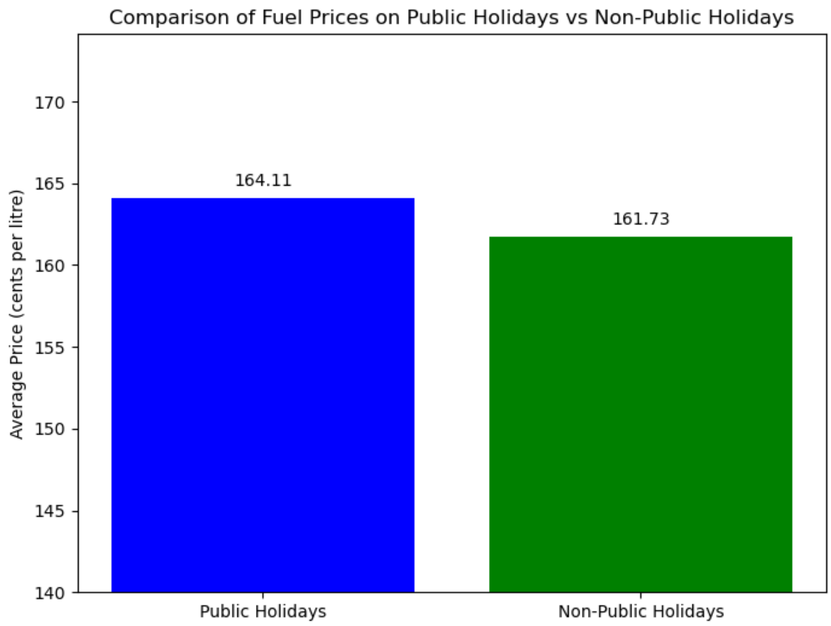


Figure 18: Comparison of Fuel Prices on Public Holidays vs Non-Public Holidays

With this high-level analysis, the bar chart shows that the Public Holidays tend to charge a higher fuel price for ULP91 compared to normal days. This gives an initial suggestion that is in support of the hypothesis claiming a higher average price for a Public Holiday. However, the difference is not so drastic with Public Holidays exceeding the normal days by only 4 cents a litre. Therefore, an attempt for a further and deeper dive took place to analyse each holiday in detail. The following graph shows a Box Plot for each holiday.

Another visualisation, Appendix 5, shows box plots for each of the Public Holidays enjoyed by Western Australia. The whiskers show the range of the prices within 1.5 times the IQR whereas the ones outside the box are outliers. It can be observed from the chart that holidays like New Year's Day, Australia Day and Anzac Day tend to have a higher median price compared to other holidays like Good Friday and Boxing Day. Additionally, there is a significant variation in price ranges for different holidays. For example, holidays like Australia Day and Easter Monday, indicate that the prices can fluctuate widely. Whereas holidays like Boxing Day and Anzac Day have more price stability with other holidays like Good Friday indicate numerous outliers either too high or too low.

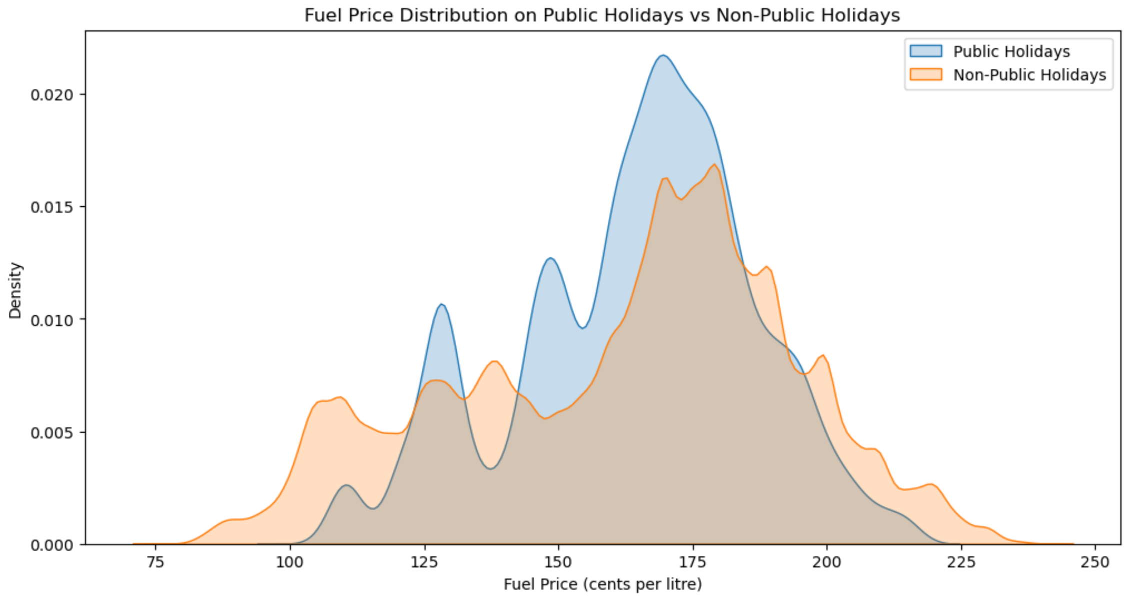


Figure 20: Fuel Price Distribution (Public Holiday Vs Non-Public Holiday)

The graph above is a density plot that compares the two scenarios being analysed. From the plot it can be observed that public Holidays generally exhibit a higher price trend compared to Normal days. The peak for public holidays around 165-175 cents per litre whereas peak for Non-public holidays is slightly lesser which is around 150-160 cents per litre. Additionally, the public holidays have their right tail extended towards higher peak prices, touching almost 250

Cents per litre. However, this is not the case with normal days as it is more centred, indicating that prices are generally more stable. This also proves that there is higher instability in prices due to a change in demand. These fluctuations provide support to hypothesis that the prices are indeed charged higher on these dates.

In the next part of this same section the data was analysed to test the hypothesis that "**Fuel prices either begin to rise before or after public holidays.**" The assumption is that fuel prices start increasing when the public holiday is coming near since people are generally getting ready for long trips which is why there will be an increase in demand. It could also be seen from a delayed price adjustment immediately after the holiday showing another part where the prices would be generally higher. Understanding the pattern would help consumers in a way that they can strategise before the time when to get their tanks filled up, anticipating the part that the prices would go higher around the day.

To conduct this analysis, there were two plots created with the purpose of finding the percentage changes in fuel prices before and after the public holidays occur. The graph is shown in Figure 21 (a) and (b).

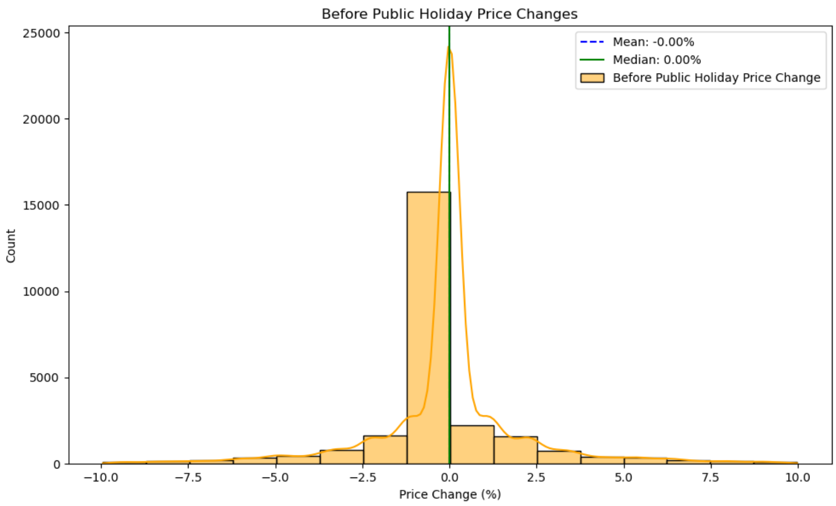


Figure 21(a): Percentage change before Public Holidays

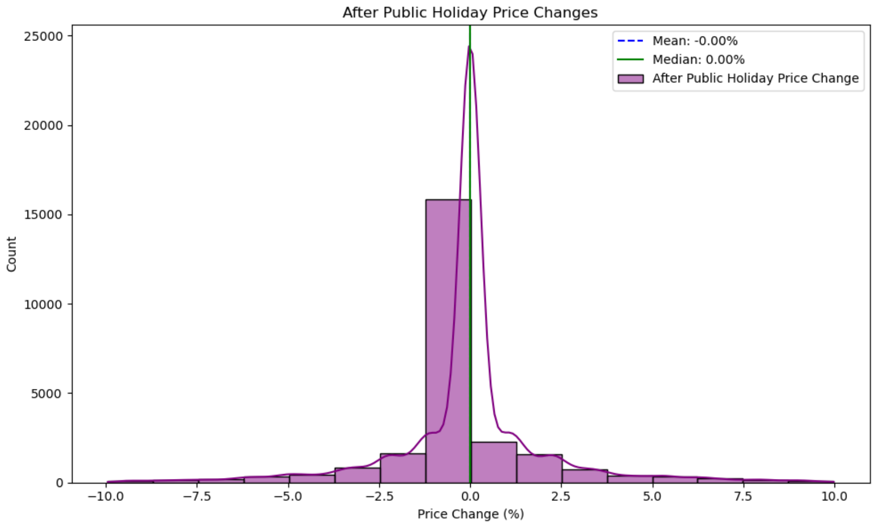


Figure 21(b): Percentage change after Public Holidays

The following visualisation uses the mean and median of fuel prices to find what happens around the public holiday dates. It shows that the mean and median price for both the graphs are near 0% which indicates that prices do not significantly change before and after the public holiday has incurred. Both distributions are heavily centred on the 0-mark suggesting almost 0 change in prices when days are approach towards and after any public holiday. This was further analysed through time series graphs as well, in which common holidays were plotted against the price averages. However, there was no major fluctuation that can be seen to deduce that prices have enough correlation with days leading up to public holidays. A few example plots are shown below, showing a random pattern with prices acting normally, indicating there are other more important reasons that are determining the prices for fuel around the days of Public Holidays. These are generated for random holidays for random days showing no trend or pattern.

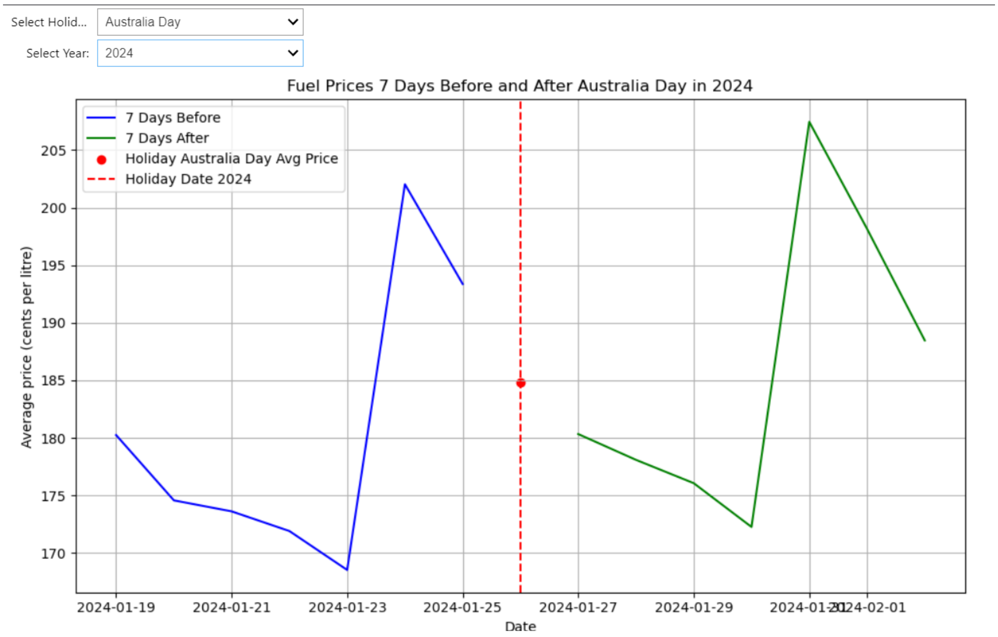


Figure 22 : 7 Day Before and After Average for Australia Day 2024

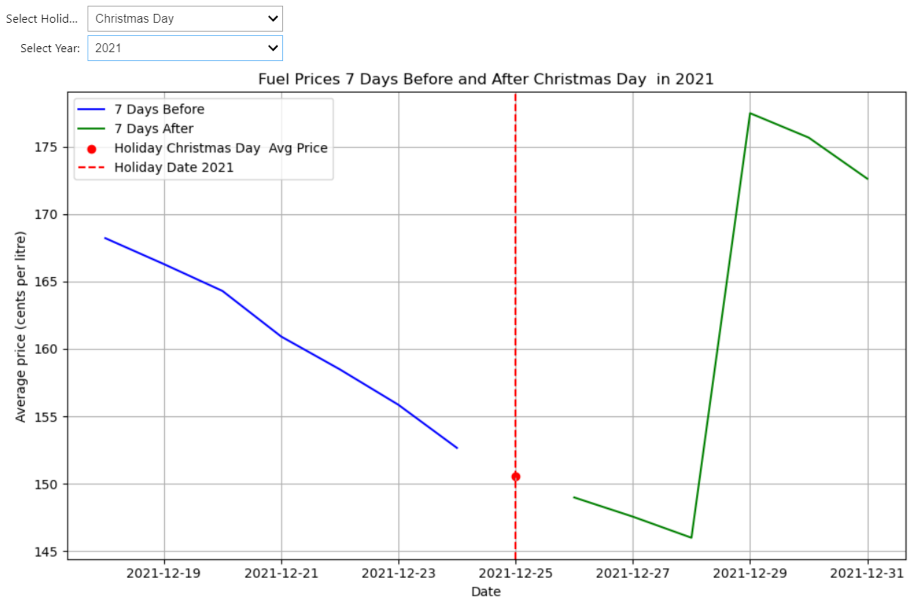


Figure 23 : 7 Day Before and After Average for Christmas Day 2021

Based on the above analysis in Figure 23 and 24, it can be concluded that even though enough evidence was found that prices tend to be higher on the dates of public holidays, there is not substantial evidence showing that prices around these days tend to increase as well. Hence, the data does not support the common assumption the prices tend to rise around the public holiday dates as people anticipate a higher demand. Therefore, for this section, one hypothesis is accepted, but the other is rejected.

4.2.7 Weather Conditions

For this section, one very important variable is being tested against the target variable of this study, Fuel Prices. The factor being analysed is the weather conditions, and the approach will be divided into two parts and hence two hypotheses. The first being tested is **"Higher Temperature causes higher Fuel prices."** With the help of Open-Mateo API, the data was engineered to add fields for average, minimum and maximum temperature based on the longitude and latitude of the selected suburbs. The investigation will quantify the relationship between prices and temperatures to find if there is any correlation between the two.

To carry out this research, a correlation analysis was carried out to find any underlying relationship between any of the created new fields against the prices. The results are shown in Table 3.

Table 3: Correlation Analysis between Temperature and Fuel Prices

	max_temp	**min_temp**	**avg_temp**	**PRODUCT_PRICE**
max_temp	1	0.831508	0.960499	0.027935
min_temp	0.831508	1	0.943538	-0.01207
avg_temp	0.960499	0.943538	1	0.005624
PRODUCT_PRICE	0.027935	-0.01207	0.005624	1

However, with this analysis there was not much to be found since the correlation with all field came out to be very close to 0. Temperature variables (max, min, and avg) show no close relationship between these factors. This implies that temperature alone doesn't seem to have a direct or strong linear effect on the fuel prices in the dataset.

To further the study, time series analysis was also conducted to keep temperature and fuel prices and separate axis to find pattern that may support the hypothesis. An example chart for 2020 is shown below.

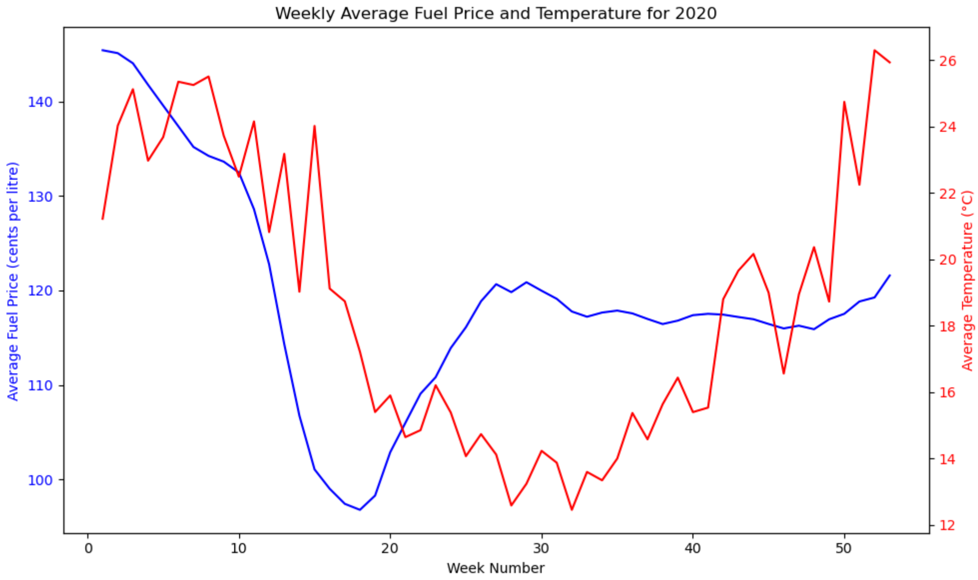


Figure 24: Weekly Average of Fuel Prices and Temperature (2020)

An example graph is shown above, which makes it evident that there is no clear or consistent relationship between fuel prices and temperature. Whenever there is a steep decline or a quick rise, the temperatures do not support it. For example, around week 20 the temperature starts to rise which does not show the same pattern in the fuel average prices. Hence, this is showing the same that fuel prices generally have no meaningful relationship with the temperatures at that point. To confirm it one more time, a seasonal average was also taken out to see if summers or winters face a different trend. However, the results show hardly a difference with prices remaining almost similar for all 4 years. This can be seen in Figure 25.

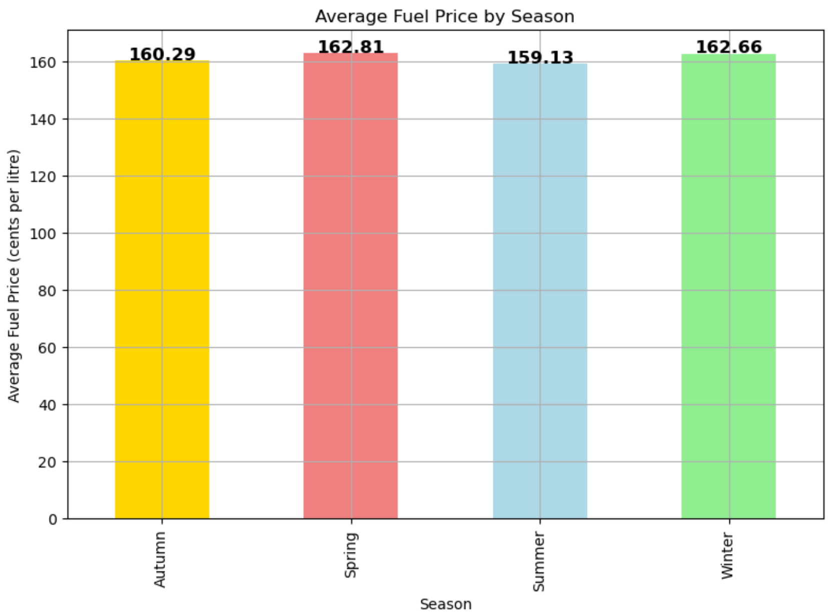


Figure 25: Average Fuel prices by Season

For the first part of this section, it can be concluded that there is enough evidence to reject the hypothesis that prices tend to be higher in higher temperatures. The underlying assumption that people use more Air Conditioners which uses up fuel more quickly and increases demand is not true. It could be due to cars being more fuel efficient or other factors playing a more important role in how prices are determined in any suburb. Therefore, this hypothesis can be rejected with confidence.

The other part of this hypothesis is related to another aspect of weather conditions, which is rainfall. The hypothesis for this section is that **"Higher amount of rainfall will lead to an increase in Fuel Prices."** This assumption is derived from the view that due to higher rainfall, there are more distributional challenges faced, causing the operational costs to rise and, ultimately, the prices. It also assumes that due to the rainfall, cars drive slowly, leading to longer traffic congestion and causing more fuel consumption. This leads to a higher demand causing prices to rise.

A similar approach was taken to find insights into the data, and hence, time series charts were created with a dual axis showing the weekly average against the Rainfall in mm. Figure 26 shows an example graph of 2020.

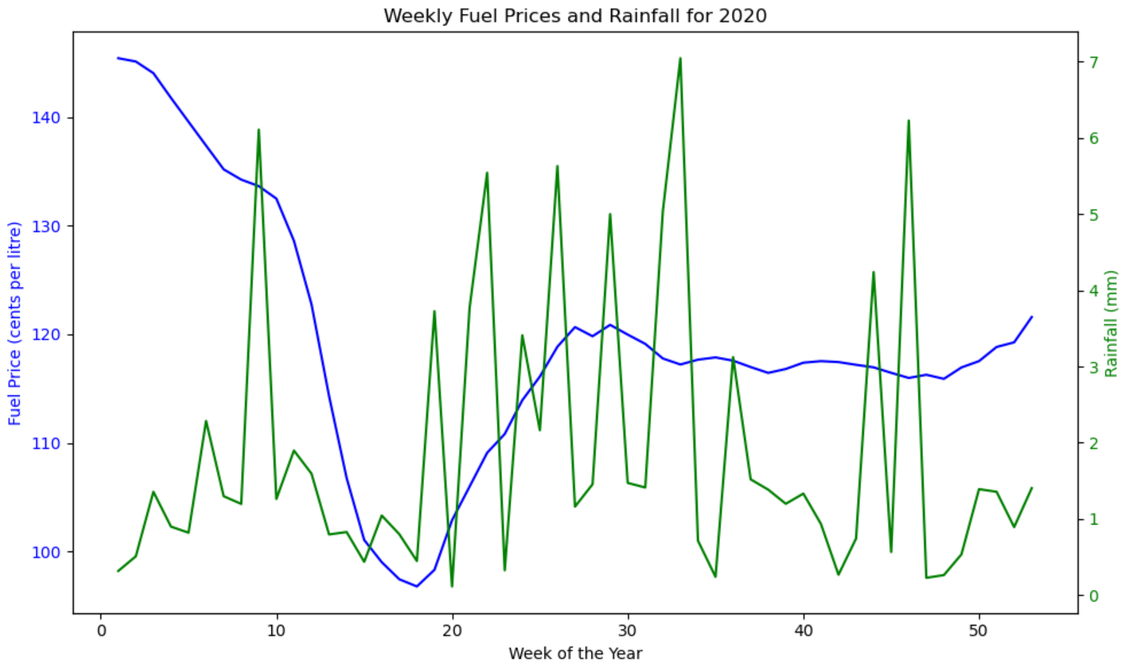


Figure 26: Weekly Fuel prices vs Rainfall (mm) in 2020

From this high-level initial analysis, we can observe that there is no clear correlation between rainfall and fuel prices. Even where there are spikes of high rainfalls during the whole week with 20 and 40 mm, there is still no corresponding increases or decreases. In fact, the fuel price independently makes their own determination without any viewable impact from rainfall.

To check for extreme rainfall, there were bins created to see if there is any impact when rainfall is of extreme mm. The graph in Figure 27 shows the results.

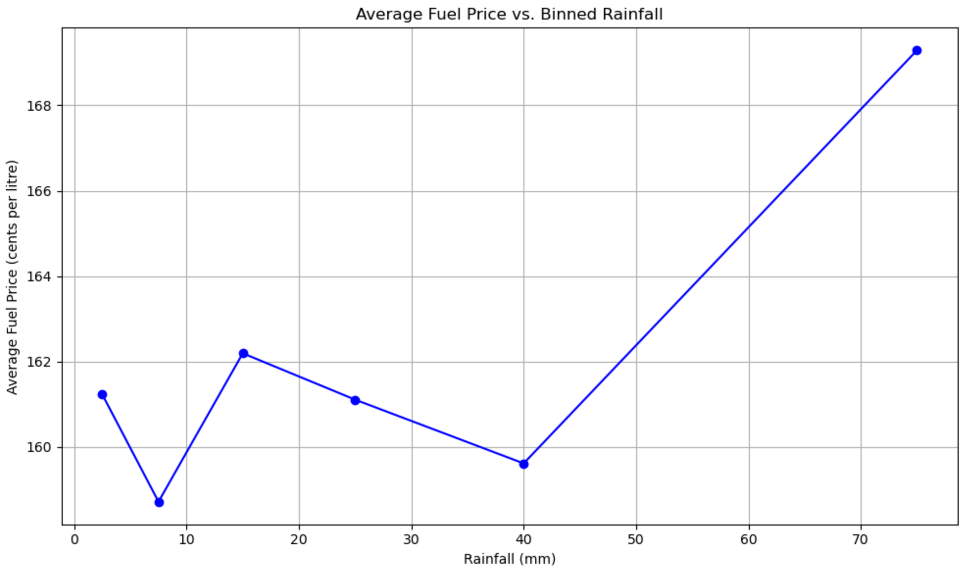


Figure 27: Average Fuel Prices against Binned Rainfall

The graph depicts relationship of both factors with binned rainfall. This shows an interesting insight that if the rainfall is 40mm or below, there is hardly any impact on the fuel prices since it is a normal weather phenomenon. However, once it goes past 50mm, a clear upward trend can be seen. This showing a higher correlation between higher rainfall and fuel prices.

From the discussion above, the hypothesis for rainfall is partially accepted since the rainfall does not have an impact on fuel prices. However, in rare cases, once it goes past a certain level of mm, the prices do tend to get impacted, showing an upward trend.

4.2.8 Interest Rates

In the later part of the analysis, some macro factors were considered to find if there is any relationship between these. The factor being discussed in this section is the interest rate being charged over the years in Western Australia. The assumption behind this basis is that due to increased borrowing, there is chance that the fuel prices rise with the increase in interest rates. Hence, the established hypothesis **is "Increased interest rates lead to an increase in Fuel Prices."**

Once the data was engineered to add a field of interest rates in WA with the primary file of fuel prices, time series analysis was carried out to find patterns in how the fuel prices are charging regarding interest rates. An example graph of 2023 is shown below to find insights into this scenario.

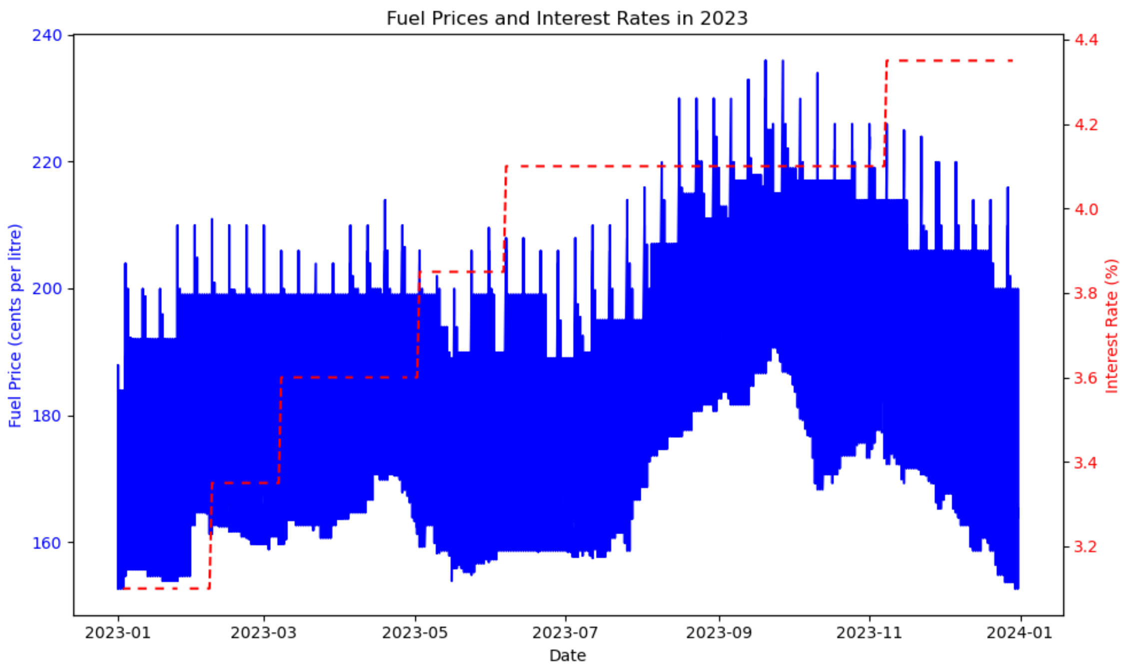


Figure 28: Time Series Analysis Fuel Prices with Interest Rates 2023

The graph above highlights the changes in interest rates in a step-like manner reflecting the interest rate policies given by the government. An overall, it can be seen that fuel prices are generally on the rise; however, they are not rising alongside the increase in interest rates. The patterns look different as it can be seen in March 2023 where fuel prices follow an independent pattern of volatility. This suggests that there are other factors that are impacting fuel prices more compared to just the interest rates. To prove it further, a correlation analysis was done to find the underlying relationship between both the factors.

The results show a moderate relationship with a figure of 0.626, indicating a noticeable but not very significant connection between the two. It can be assumed that interest rates are causing an effect on another macro factor like consumer spending or inflation which may be causing this lagged effect on fuel prices. For this reason, another analysis was carried out to find insights into the lagged effect of interest rates on fuel prices. This can be seen below.

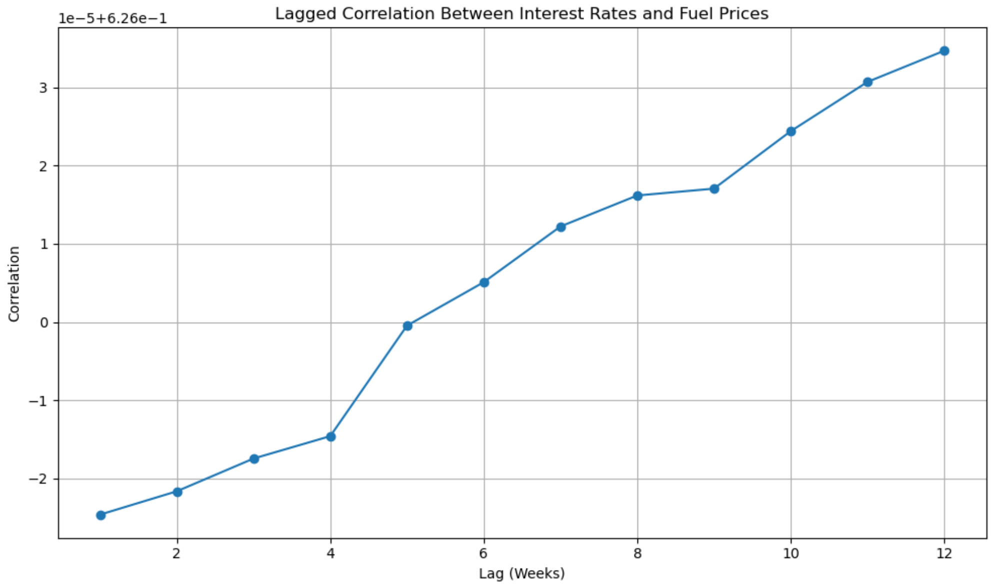


Figure 29: Lagged Correlation between Interest Rates and Fuel prices

Figure 29 assesses how changes in the interest rates affect the fuel prices over time rather than immediately. From this case we can see that as the time periods increase after a change the lagged effect turns into more positive. This shows that due to the delays in the market adjustment or other external factors, the price tends to react a bit slower. In other words, in the short term, while considering around 4 weeks, the trend indicates the prices may either not be affected or may be negatively affected. This may be because the market forces are underway towards a movement and hence, there is no such immediate response. When considering beyond 6 weeks, the long-term trends show a more positive response showing that interest rates impact other principal factors like inflation that then has a lagged effect on the fuel prices.

As a conclusion to this macro factor, it’s hard to accept or reject the hypothesis as on a larger scale there are many factors playing a role in determining prices. However, if we solely consider interest rates, we can partially accept the hypothesis since the correlation and the delayed response analysis does indicate a rising trend due to interest, but it is through a lagged interval of about 6 Weeks after any change to the policy.

4.2.9 Inflation

With the last analysis on interest rates indicating a lagged response, this part of the report analyses what factor could be causing that delay. Inflation being something that directly impacted by interest rates, is engineered with the primary dataset to finder if this has any connection with the fuel prices. The consumer price index and year-ended inflation rate was extracted to do an even deeper analysis. Hence, the hypothesis for this section will be **"Higher Inflation Rates lead to a similar increase in Fuel Prices."**

A time series analysis was carried out on both fields to find deeper insights into the factor of inflation being tested. Since CPI tracks a basket of goods and year-end inflation rate is the overall rate of inflation, both act as good measures to detect any patterns. By analysing them together allows to discuss the potential of correlation with the overall factor better. The graph s are shown in Figure 30.

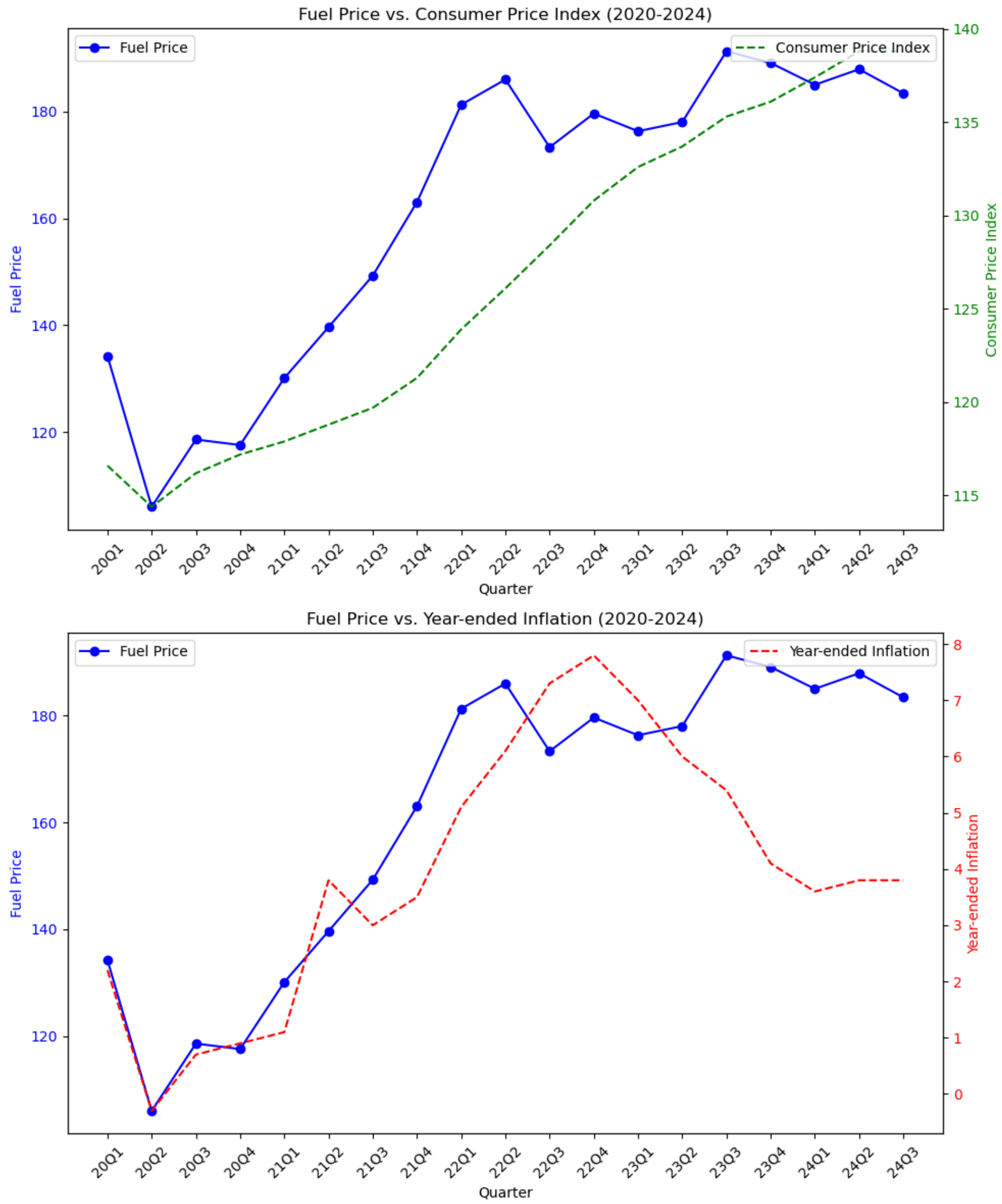


Figure 30: Time Series Analysis of CPI and Inflation Rate with Fuel Prices

From the first graph, the fuel prices are consistently on the rise in all considered years and at the same time so is the CPI. However, CPI not rising at the same pace which means that fuel pricing being a component of CPI does have an impact on the inflation to rise however it a appears to be a moderate relationship since other components are also having an impact. Like this, if we see the second graph for yearend rates, we can observe a similar upward trend; however, there are points on the graph that show that even though at times the fuel prices were declining, the inflation rate at that point was still on the rise. Hence, both factors may have a connection between them, but because these are macro reasons, it cannot be easily deduced that one has a complete effect on other since there are so many other factors impacting the results at the same time.

Analysis was further done with box plots and for lagged analysis. It can be found in Appendix 1. However, all results were reflecting the same findings. It can be accepted that these factors are interrelated. However, the hypothesis of them following a similar pattern is rejected since, at one point in time, there are areas where the dependent and the independent factors are going in opposite directions. Also, the rates at which they both are increasing also reflect that when considering a macro factor, there are other significant factors that playing a role in pulling or pushing the fuel prices.

4.2.10 TAXES

The last factor being discussed is also another that falls under the category of macro impactors. This explores the relationship between the excise taxes imposed on fuels against fuel prices overtime. The assumption being that taxes are imposed to control the prices being determined for the fuel to bring stability in the economy. Hence, the hypothesis for this section will be **"Fuel Prices directly respond to the changes in Excise Taxes."**

For the analysis, the first part was to conduct a correlation analysis between the two studied variable. From the results the correlation came out to be very weak with almost the figure of 0.053. This shows a positive yet very weak correlation, indicating there is no meaningful relationship between the two factors. In other words, from this analysis we can find that from the changes in excise tax there is minimal impact on fuel prices. There are other important variables that play a more significant role in determining the prices of fuel in WA.

To take the research deeper, a particular scenario was considered in the mid of 2022. This was the part of the year when a significant world event was taking place with Russia and Ukraine having a war. As Russia is one of the largest oil producer and exporter, having sanctions and restrictions led to the global supply and demand to get disrupted leading to an increase in fuel prices. For this reason, Australian government responded by cutting the taxes into half in March. This went down from 44.2 Cents to 22.1 Cents/l. This reduction was in place for the next 6 months where small adjustments were made. This can be seen in the graph below

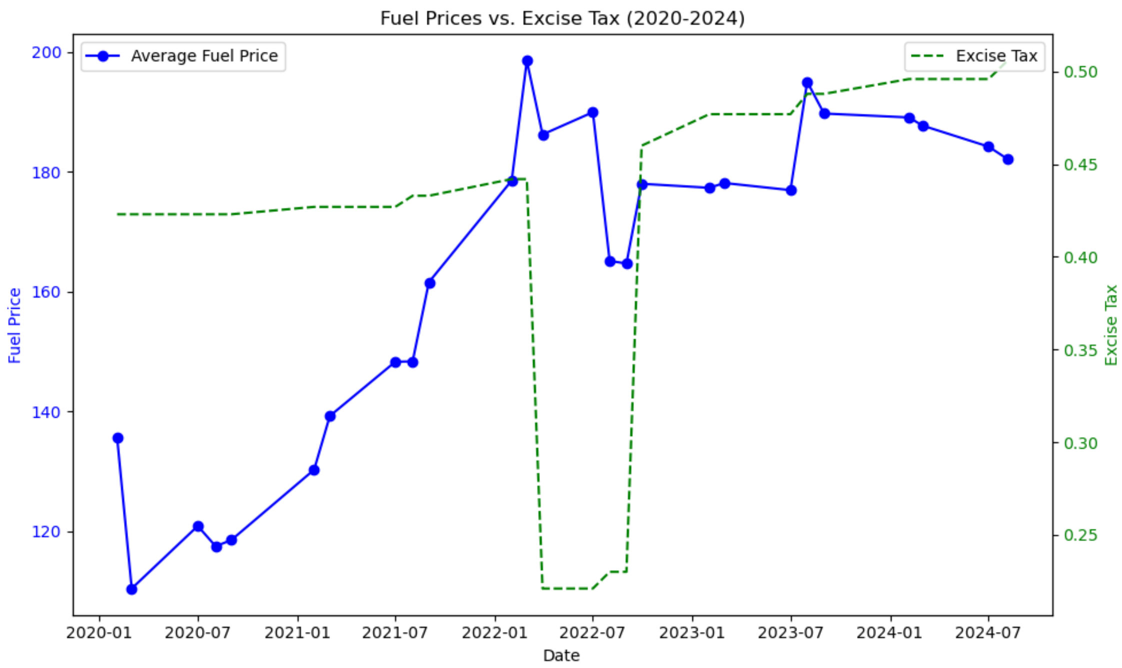


Figure 31: Time Series Analysis of Excise Taxes with Fuel Prices

Figure 31 shows the attempt by the government to reduce the cost of living. However, despite this cut in taxes, the prices continued to increase. This shows that even though taxes might have reduced a bit of the fuel prices at that point, there was not much impact on the overall determining. It was the forces of the market that had more control, and hence, it still went its own way even after adjustment. Therefore, it can be concluded from this analysis that there is even support against the hypothesis that the prices will move towards the direction of the taxes. It is not always that impactful as there are many other major factors that impact the determination of fuel prices.

5 CONCLUSION

5.1 IMPLICATIONS

- This research provides significant insights into the numerous factors influencing fuel price fluctuations. The major portion discusses the local or micro factors that could be contributing towards the price determination of fuel prices in Perth's Metropolitan suburbs. This allows a thorough understanding of the local dynamics that are at play, causing the peaks and dips in the demand and supply of ULP91. The study gave an overview of multiple factors and tested the assumptions to find known and unknown trends that could aid in the understanding of how fuel prices are set in regional areas like Perth.
- The research demonstrates that the majority of local factors, such as public holidays, weather conditions, fuel station density, etc., are those that add to the short-term variations towards prices. It causes short impacts with little or no lagged effects. However, this study does help understand how macro factors like interest rates, taxes and inflation never cause an impact on prices as an individual factor. Their push or pull towards an equilibrium fades, with so many other factors being more significant. Compared to local or micro factors, they do exert a stronger influence, and hence have a bigger impact on what the prices are.
- Stakeholders like consumers, businesses and policymakers will have an advantage since it will allow them to make strategic decisions considering all the 10 factors discussed in the study. Financially, consumers and businesses will be able to reduce their daily costs by strategising how they make fuel purchases. Policymakers can bring more stability to the economy by understanding how macro factors play a role in the determination of prices.

5.2 Recommendations

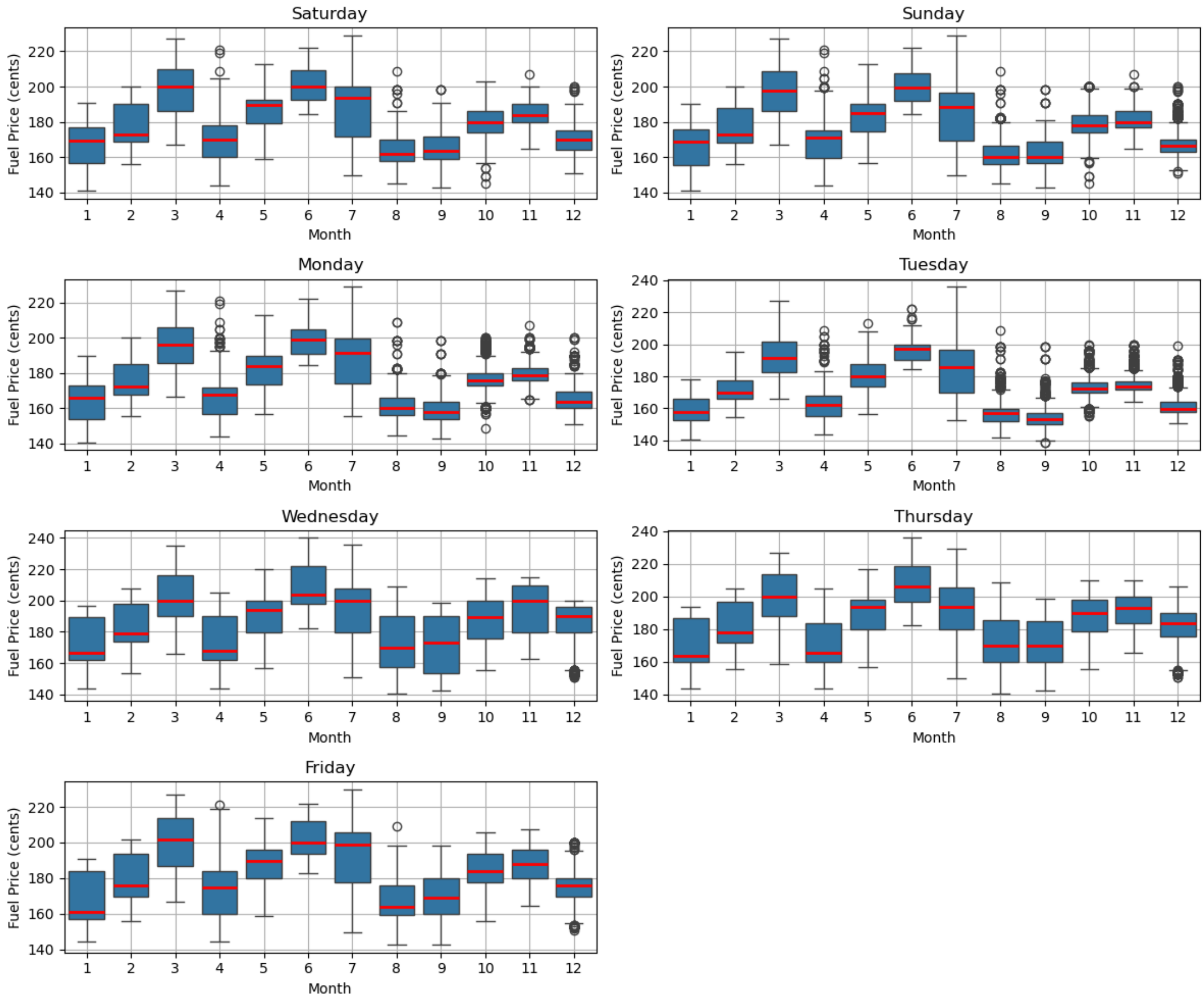
Based on the project, there are many recommendations that could help with the current analysis or could be a way to improve it in the future. Few of them are as follows:

- This research considered selected Metropolitan suburbs from Perth. This was done to have similar characteristics so that complexity can be reduced and there are lesser chances of Outliers. However, this can be expanded in future research to include rural and remote areas for an even more comprehensive understanding of the overall fuel pricing for Perth or Western Australia as a whole.
- The research was not able to reach its full potential due to the constraints of budget and timing. However, in the future, paid datasets like WA Census data can be engineered together with the primary file to find more interesting insights that involve the demographics of people living here.
- Some datasets were outdated or inaccessible like household income or sporting events which could be collected through other approached of research. These can be vital in understanding other local factors that could also have a potentially high correlation with fuel prices.
- As per this research, only one type of fuel was investigated, which is ULP91; however, other types, like Ron 98 or PULP95, have different characteristics that can also provide variations in the analysis that has already been done.
- As of now, in macro factors, only a certain component of the overall variable was discussed. For example, in regulations only excise taxes were considered. This could be further increased to other types of regulations and their impact on fuel prices.
- Once majority of data is collected, the research can also have a project part B where a predictive model can be created that could allow prices to get determined before time. This could also act as beneficial for the same stakeholders as this way they can make strategies considering the future rather than just historical data.

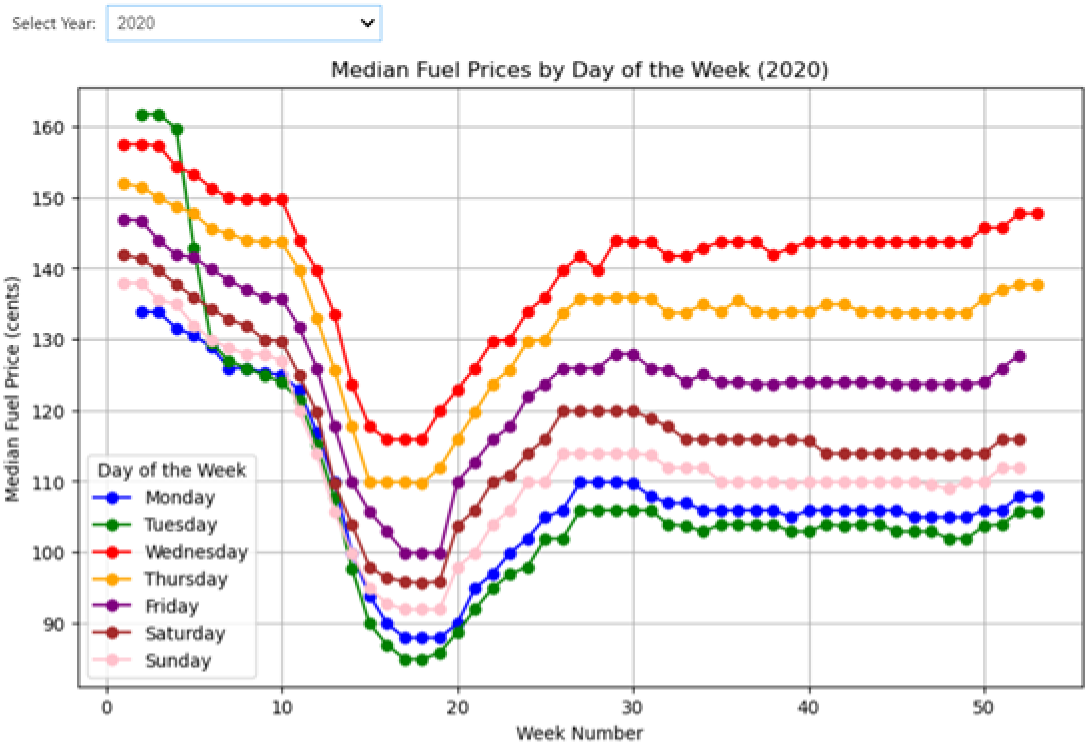
These recommendations offer pathways for further study of the topic and could provide valuable insights into how prices are determined, specifically in a selected region. With an increased understanding of the complex patterns involved with multiple factors at a time, contributions can be made towards informed decision-making, helping different stakeholders, including the people of Perth.

6 APPENDICES

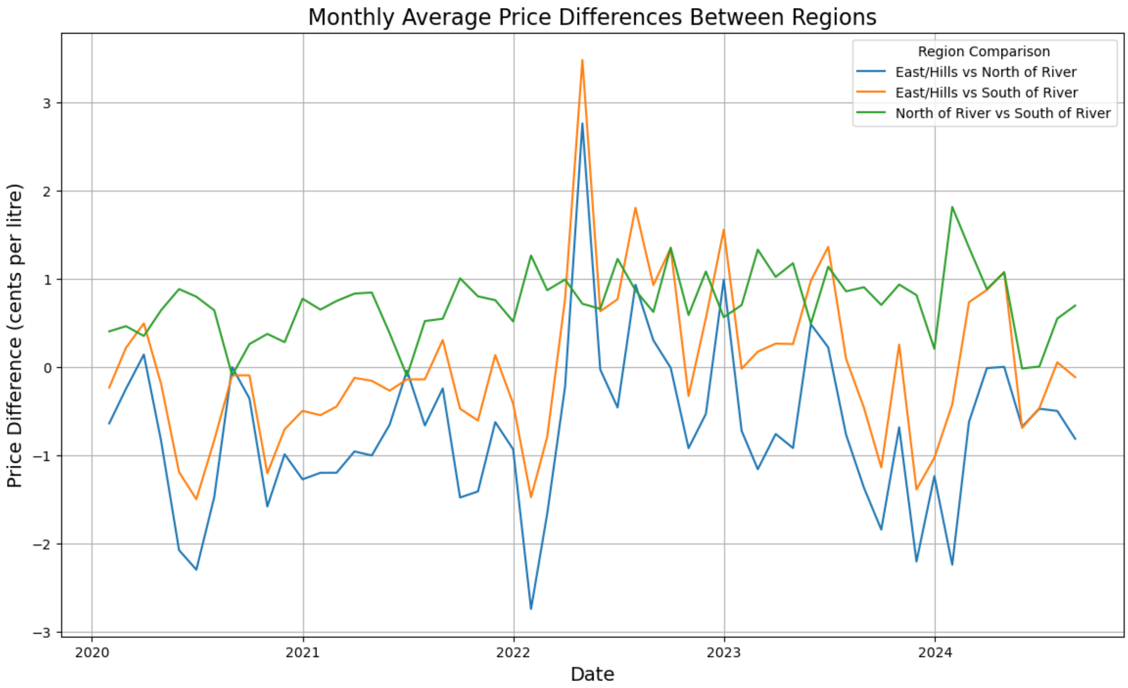
APPENDIX 1: MONTHLY BOXPLOTS FOR EVERY WEEKDAYS



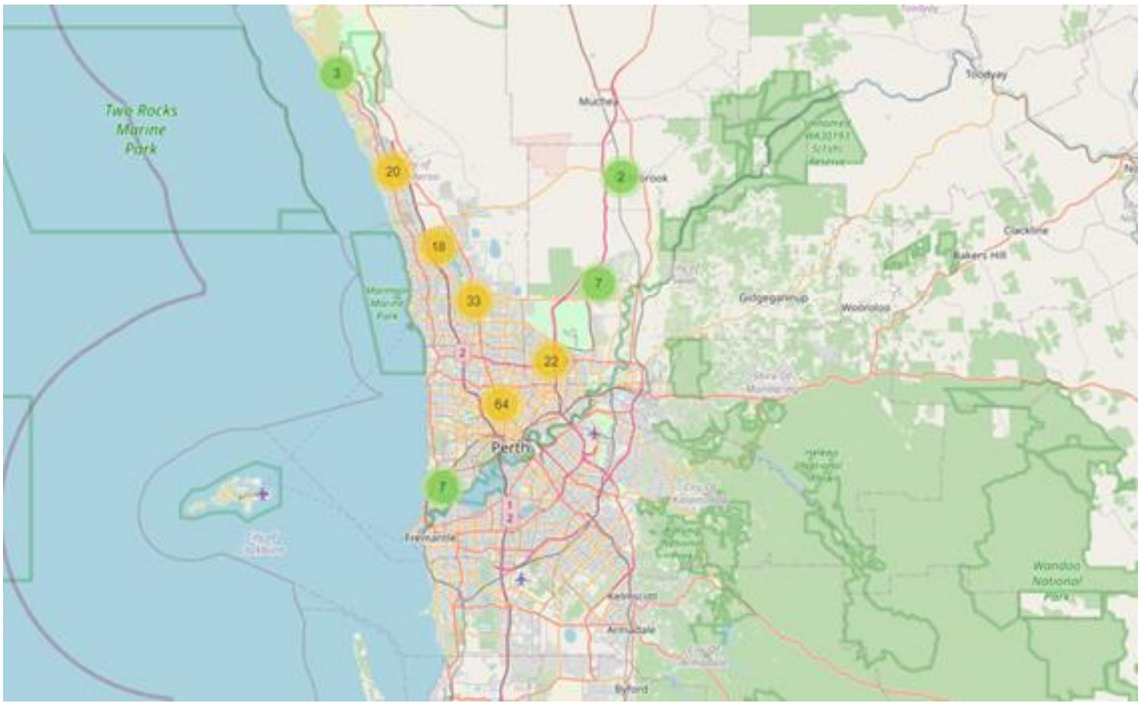
APPENDIX 2: MEDIAN FUEL PRICES BY DAY OF THE WEEK



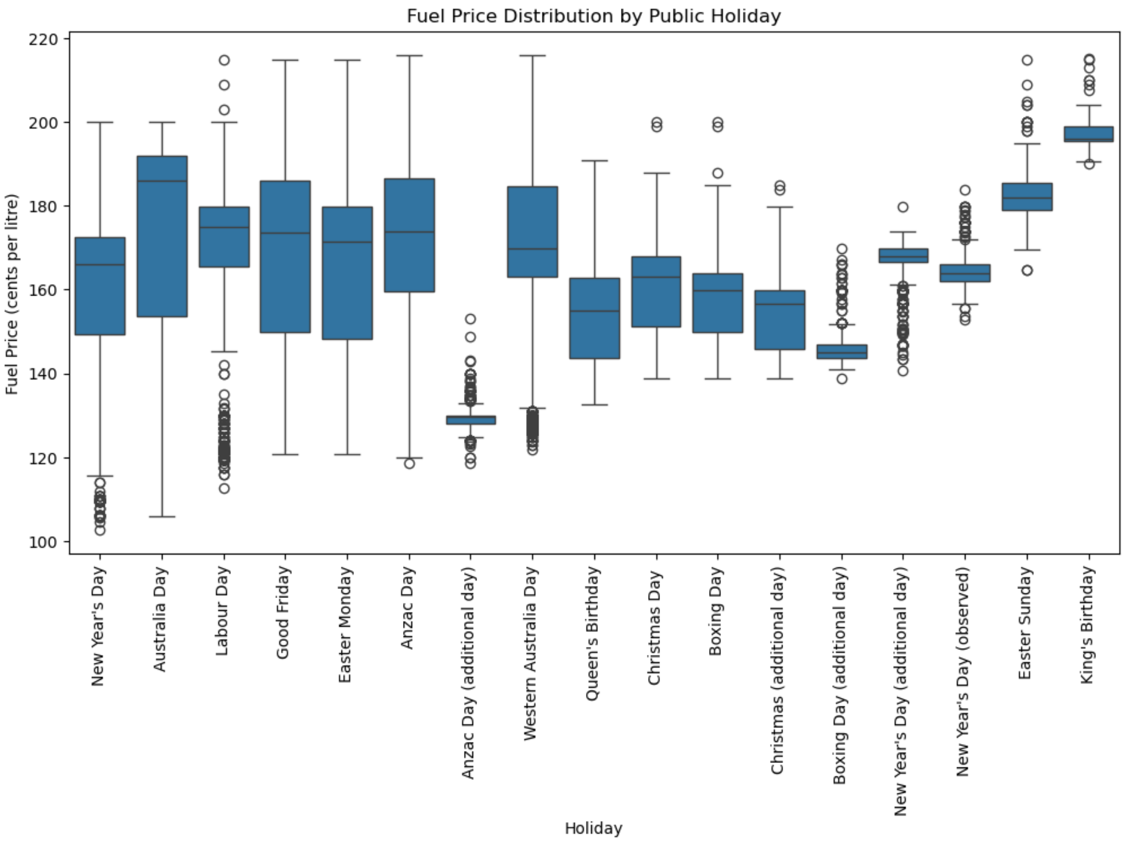
Appendix 3: Monthly Average Price Differences between Regions



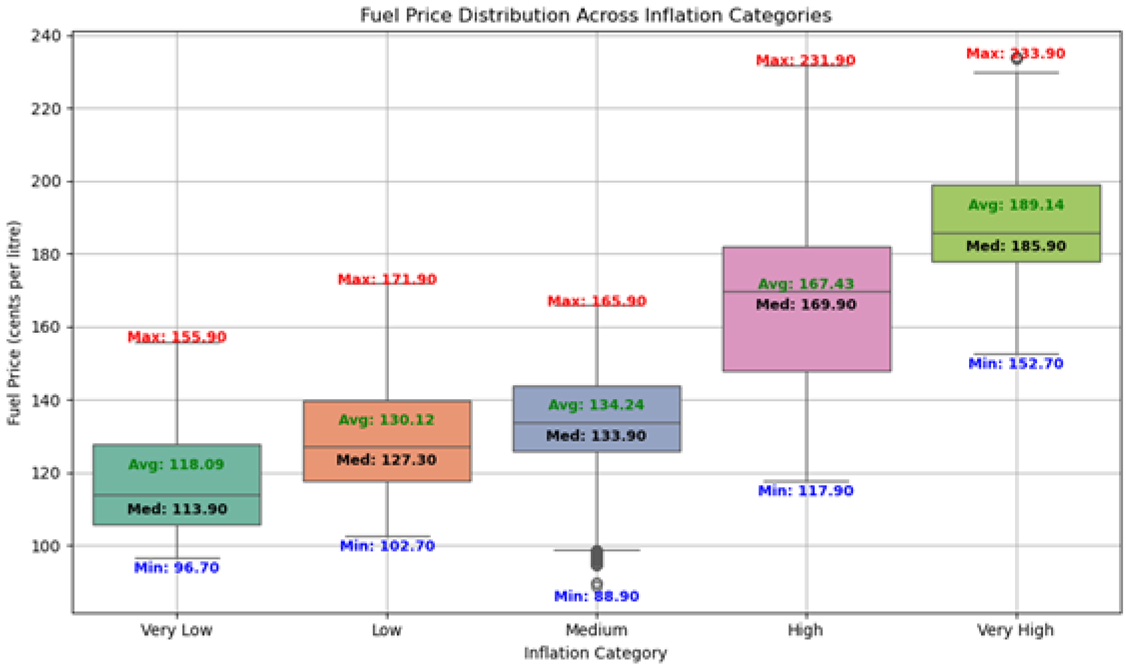
APPENDIX 4: INTERACTIVE GRAPH OF CLUSTER AVERAGES (NORTH OF THE RIVER, 2024)



APPENDIX 5: INTERACTIVE GRAPH OF CLUSTER AVERAGES (NORTH OF THE RIVER, 2024)



APPENDIX 6: BOXPLOT ANALYSIS FOR INFLATION CATEGORIES



7 REFERENCES

Bergeaud, Antonin, and Juste Raimbault. 2020. "An empirical analysis of the spatial variability of fuel prices in the United States." *Transportation Research Part A: Policy and Practice* 132: 131-143.

Borenstein, Severin, A Colin Cameron, and Richard Gilbert. 1997. "Do gasoline prices respond asymmetrically to crude oil price changes?" *The Quarterly journal of economics* 112 (1): 305-339.

Box, George EP, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. 2015. *Time series analysis: forecasting and control*. John Wiley & Sons.

Dewenter, Ralf, Ulrich Heimeshoff, and Hendrik Lüth. 2017a. "The impact of the market transparency unit for fuels on gasoline prices in Germany." *Applied Economics Letters* 24 (5): 302-305.

Ulrich Heimeshoff. 2017b. "Less pain at the pump? The effects of regulatory interventions in retail gasoline markets." *Applied Economics Quarterly* (3): 259-274.

Doğanlar, Murat, Faruk Mike, Oktay Kızılkaya, and Ahmet Kardaşlar. 2024. "Temperature, Precipitation and Economic Growth: The Case of the Most Polluting Countries." *International Journal of Environmental Research* 18 (1): 4.

ESRI, ESRI. 1998. "Shapefile technical description." *An ESRI white paper* 4 (1).

Ghazanfari, Arezoo. 2020. "Analysis of the retail petrol price patterns in Western Australia." RMIT University.

Arezoo Ghazanfari,2022. "What drives petrol price dispersion across Australian cities?" *Energies* 15 (16): 6025.

Goodchild, Michael F, Paul A Longley, David J Maguire, and David W Rhind. 2005. "Geographic information systems and science." *Wiley & Sons, West Sussex, UK* 17: 517.

Gordon, William J, and Robert S Rudin. 2022. "Why APIs? Anticipated value, barriers, and opportunities for standards-based application programming interfaces in healthcare: perspectives of US thought leaders." *JAMIA open* 5 (2): ooac023.

Hamilton, James D. 2009. *Causes and Consequences of the Oil Shock of 2007-08.* National Bureau of Economic Research.

Kalligeros, S, F Zannikos, S Stournas, E Lois, G Anastopoulos, and D Karonis. 2003. "Impact of gasoline quality on engine performance and emissions." Proceedings of the 8th International Conference on Environmental Science and Technology, Lemnos Island, Greece.

Kemerer, Chris F. 1995. "Software complexity and software maintenance: A survey of empirical research." *Annals of Software Engineering* 1: 1-22.

Khan, Khalid S, Regina Kunz, Jos Kleijnen, and Gerd Antes. 2003. "Five steps to conducting a systematic review." *Journal of the royal society of medicine* 96 (3): 118-121.

McKinney, Wes. 2010. "Data structures for statistical computing in Python." SciPy.

Phan, Toan. 2023. "A New Look at the Effects of Weather Shocks Over Time, No. 23-25."

RT, FIELDING. 2000. "Architectural Styles and the Design of Network-based Software Architectures." *Ph. D. Thesis, UC Irvine*.

Shafranovich, Yakov. 2005. Common format and MIME type for comma-separated values (CSV) files.

Wills-Johnson, Nick. 2008. "Price Cycles in Perth Petrol Markets: A Spectral Analysis." *Available at SSRN 1188062*.