Data Visualization and Exploration Assignment 1

Ru'aan Maharaj - 2446659 James Stanton - 2541773

September 7, 2025

Question 1: Data Cleaning

1.1: Dates

All columns in the raw dataset were initially read in as strings, even the numeric fields. Many columns also contained missing values, and some were in the wrong format (e.g., `date_fueled`). In certain cases, columns even contained a mix of text and numbers.

We first examined the `date_fueled` column. We found that 11.66% of entries could not be parsed as proper dates, these were converted to `NaT`. The `date_captured` column was also converted to datetime format. Where `date_fueled` was missing but `date_captured` contained a valid date, we imputed `date_fueled` with the corresponding `date_captured` value using `df.loc`.

After conversion, we removed invalid entries such as dates earlier than 2005 or dates in the future.

The distribution of fueling dates shows that most records are concentrated in recent years (2020 onwards). Earlier dates have fewer records, likely because fewer users logged their data at that time. There are no entries beyond the current date, confirming that invalid future values were successfully removed. The largest concentration of fueling records comes from 2021–2023, reflecting more consistent data collection in recent years.

Figure: Distribution of fueling dates.

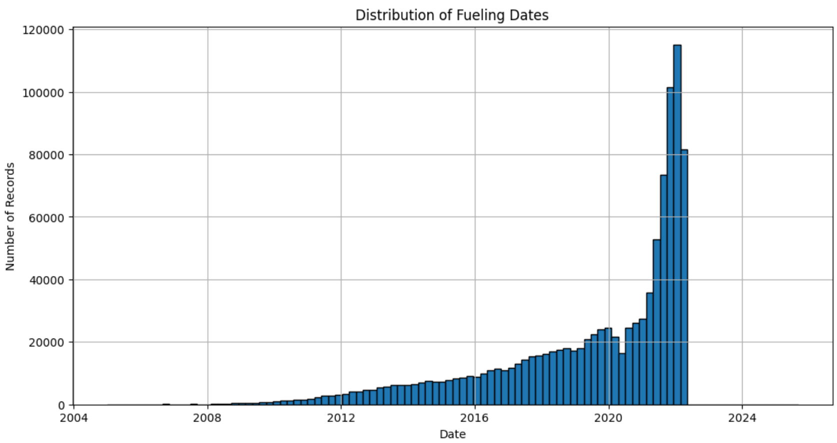


Figure 1: Distribution of fueling dates after cleaning.

1.2 Numeric Fields

1.2.1: Missing Values

The numeric columns `gallons`, `miles`, `odometer`, and `mpg` were first examined for missing values. The `gallons` column has only 3.32% missing entries, indicating that most fueling events include fuel volume data and are largely reliable. In contrast, the `miles` column is missing 88.58% of its values, severely restricting direct mileage analysis and making imputation from gallons and mpg essential. The `odometer` column has 11.61% missing entries, a noticeable but manageable gap that may impact long-term tracking. Finally, the `mpg` column has 3.32% missing values, a relatively small proportion that can often be reconstructed using the other two fields.

1.2.2: Interdependence of Miles, Gallons, and MPG

Initially, we attempted to fill missing values in the interdependent fields directly after coercing them to numeric types. However, this approach led to data loss, since values such as "73,370" were converted to NaN before commas were removed. To avoid this, we first performed data cleaning (Section 1.2.3) and then applied the interdependence formulas. This ensured that reconstruction was based on valid numeric values rather than corrupted NaNs.

Missing entries were reconstructed using the following relationships:

$$\text{mpg} = \frac{\text{miles}}{\text{gallons}}, \quad \text{miles} = \text{mpg} \times \text{gallons}, \quad \text{gallons} = \frac{\text{miles}}{\text{mpg}}$$

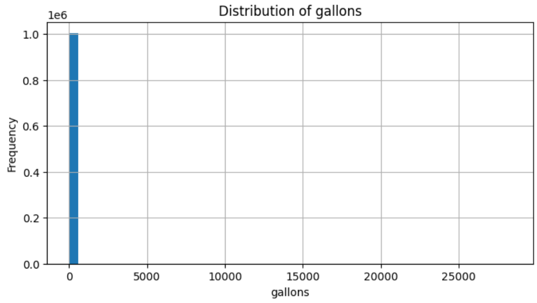
with safeguards against division by zero. This reduced information loss while maintaining internal consistency between the fields.

1.2.3: Cleaning and Conversion

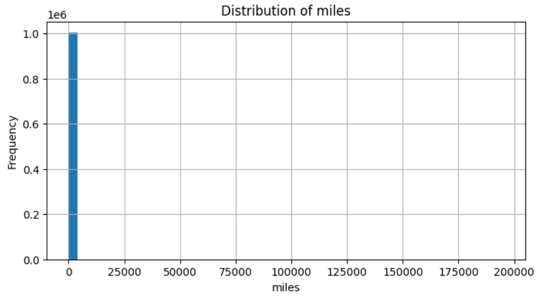
All four numeric columns were stored as strings with commas and other non-numeric characters. These were first cleaned using string replacement (e.g., "73,370" → "73370"), then converted to float using `pd.to_numeric`. This conversion step, performed before imputation, ensured that valid numbers were preserved for reconstruction.

1.2.4: Distributions

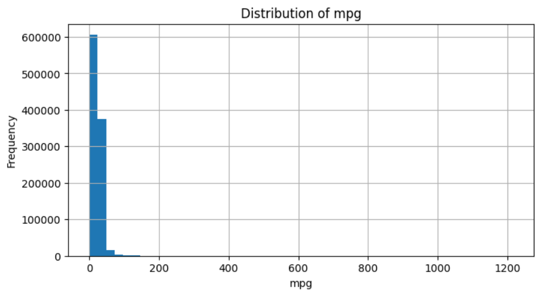
The distribution plots of the numeric fields reveal significant quality issues typical of user-generated datasets. Odometer readings show a uniform distribution with a concerning spike at very low values, suggesting data entry errors or unit confusion. Both `miles` and `gallons` exhibit extreme right skew, with massive concentrations near zero and unrealistic outliers extending to 20,000+ miles and 1000+ gallons. These patterns likely reflect a mixture of personal and commercial vehicles, alongside unit inconsistencies. The `mpg` distribution appears more realistic, with most values between 10–60 mpg and a peak around 20–30 mpg, though some extreme outliers (approaching 800 mpg) clearly result from calculation errors. Overall, these plots highlight the critical need for robust outlier handling before further analysis.



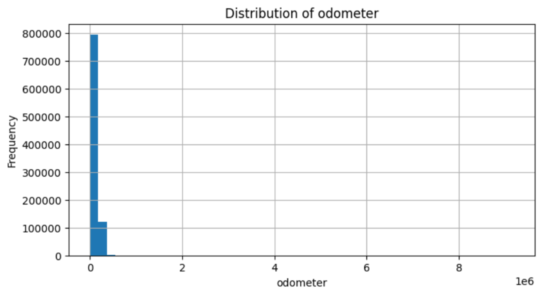
(a) gallons



(b) miles



(c) mpg



(d) odometer

Figure 2: The distribution plots of the numeric fields

1.2.5: Descriptive Statistics

Descriptive statistics for the numeric fields are shown in Table 1.

Column	Count	Mean	Std	Min	25%	50%	75%	Max	Most Freq
Gallons	1,002,644	12.44	6.62	0.0	9.08	12.01	14.97	984.71	10.57
Miles	1,002,641	265.94	190.49	0.0	184.70	268.49	343.30	23,238.40	0.00
Odometer	5,210	496.45	296.89	0.0	259.25	500.00	750.00	999.00	1.00
MPG	1,002,699	22.21	15.59	0.0	15.70	21.80	28.50	786.50	0.00

Table 1: Descriptive statistics for numeric fields.

The gallons column averages 12.4 gallons per fueling, consistent with typical fuel tank sizes, though the maximum of 984.71 is unrealistic and reflects data entry errors. Miles average 266 miles between refueling, with a typical range of 185–343 miles, but the maximum of 23,238 is clearly an outlier. Odometer values average 496, which is unusually low for real-world readings, suggesting partial values or truncation. MPG averages 22.2, a reasonable figure, but the maximum of 786.5 mpg is implausible. The most frequent values of 0 in miles and mpg reflect either missing or corrupted entries, while the frequent gallon value around 10.6 aligns with typical real world refueling behavior.

Question 2: Feature Engineering

2.1.1: Currency Extraction

We extracted the currency from the cost_per_gallon column by identifying the prefix symbol or abbreviation before the numeric values. A special case was handled for the Swiss Franc, which can appear as "Ch.f." instead of "CHF". The resulting currency column allows for consistent comparison of fuel prices across countries. The majority of entries were in US dollars ($), with smaller proportions in pounds sterling, euros, rand, and other currencies.

2.1.2: Numeric Conversion of Costs

We created numeric versions of both cost_per_gallon and total_spent by removing currency symbols and other non-numeric characters, then converting the results to floats. This transformation ensures that these fields can be used in statistical analysis and numerical computations. For example, values such as "$67.86" were successfully converted to 67.86.

2.1.3: Vehicle Information from URL

Vehicle details were extracted from the user_url field. The URL structure consistently encodes the make, model, year, and user ID. These were parsed into four new columns (make, model, year, and user_id). This enrichment makes it possible to compare fuel efficiency across vehicle types and manufacturing years. Minor issues were observed in some entries (e.g., missing years), but most were parsed cleanly.

2.2.1 and 2.2.2: Litres and Kilometres

We computed the litres of fuel per entry from the gallons column, taking into account whether the entry represented US gallons (1 gallon = 3.785 L) or UK gallons (1 gallon = 4.546 L, identified by the currency "£"). Distances driven in miles were converted to kilometres using the standard conversion (1 mile = 1.609 km). These new features (litres, km) allow for more standardised analyses of vehicle usage across international contexts.

2.2.3: Fuel Economy

We calculated fuel economy in litres per 100 kilometres using the formula:

$$L/100km = \frac{\text{litres}}{\text{km}} \times 100$$

This provides the international standard metric for fuel efficiency, complementing the original mpg column. Most values fell between 6–11 L/100km, consistent with real-world passenger vehicles. Extreme outliers (e.g., values approaching zero or greater than 50) were treated as errors and removed.

Question 3

3.1: Unique users per country

The analysis of unique users per country reveals a heavily US dominated user base with 77039 users, representing approximately 77% of the top 10 countries, followed by a significant drop to the United Kingdom with 10018 users, indicating the fuel logging appl has achieved strong penetration in English speaking markets. The remaining countries show a steep decline in adoption with Europe at 6437 users, Canada at 4810, and South Africa at 3833, suggesting regional preferences or marketing focus, while smaller markets like New Zealand, Malaysia, Brazil, and Poland demonstrate limited but present adoption. Japan's extremely low user count of only 72 users, despite being a major automotive market, suggests cultural preferences for different apps, language barriers, or limited marketing presence, highlighting how digital adoption patterns don't necessarily correlate with economic development or car ownership rates but rather reflect factors like app store presence, language localization, and regional marketing strategies.

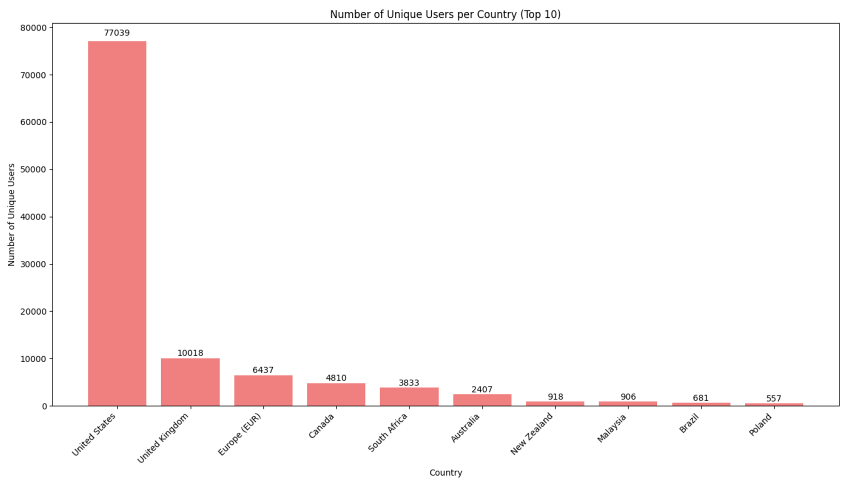


Figure 3: Unique users per country.

3.2: Popularity of the app

The app popularity analysis reveals a classic technology adoption curve with extremely slow initial growth from 2008 to around 2016 (averaging under 50 daily users), followed by steady moderate growth through 2020 (reaching approximately 200-400 daily users), and then explosive growth from 2021-2023 with a dramatic spike reaching the peak of 2375 daily users. The overall average of 183.08 daily unique users reflects the extended period of low adoption in the early years, while the sharp exponential growth pattern from 2020 onwards suggests the app benefited significantly from increased digitisation during the COVID-19 pandemic, heightened fuel cost awareness due to global price volatility, or successful marketing campaigns and app store optimisation.

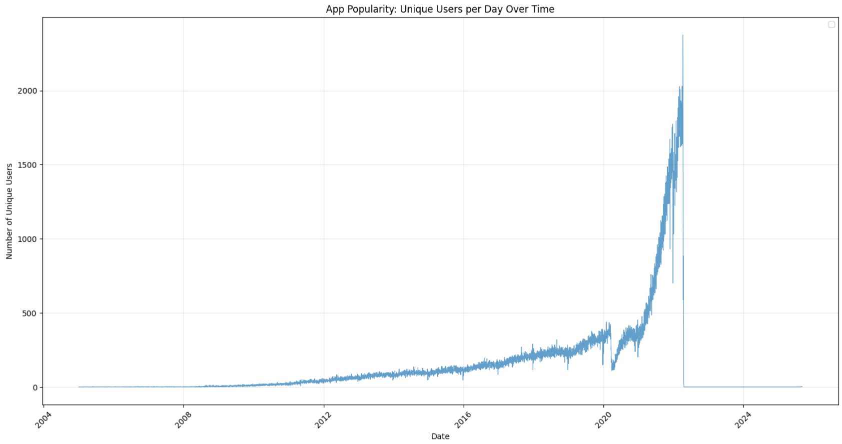


Figure 4: Unique users per country.

3.3: The distribution of age of the vehicles per country

The vehicle age distribution analysis reveals distinct regional patterns with New Zealand showing the oldest vehicle fleet at an average of 14.5 years, followed by the United States and Europe, while Malaysia maintains the youngest fleet at 8 years, with South Africa and the United Kingdom also showing relatively newer vehicles. The histograms demonstrate that most countries follow a right-skewed distribution with peaks around 5-10 years, suggesting active replacement cycles, though New Zealand's distribution is more evenly spread across older age ranges, likely reflecting economic factors such as higher vehicle import costs or cultural preferences for maintaining older vehicles rather than frequent replacements. The relatively young vehicle ages in Malaysia and South Africa may indicate rapidly growing economies with increasing vehicle ownership, newer market penetration of the fuel logging app among younger vehicle owners, or economic conditions that favor more recent vehicle purchases, while the older fleets in developed markets like the US and Europe suggest mature automotive markets where vehicles are kept longer due to improved reliability and maintenance practices.

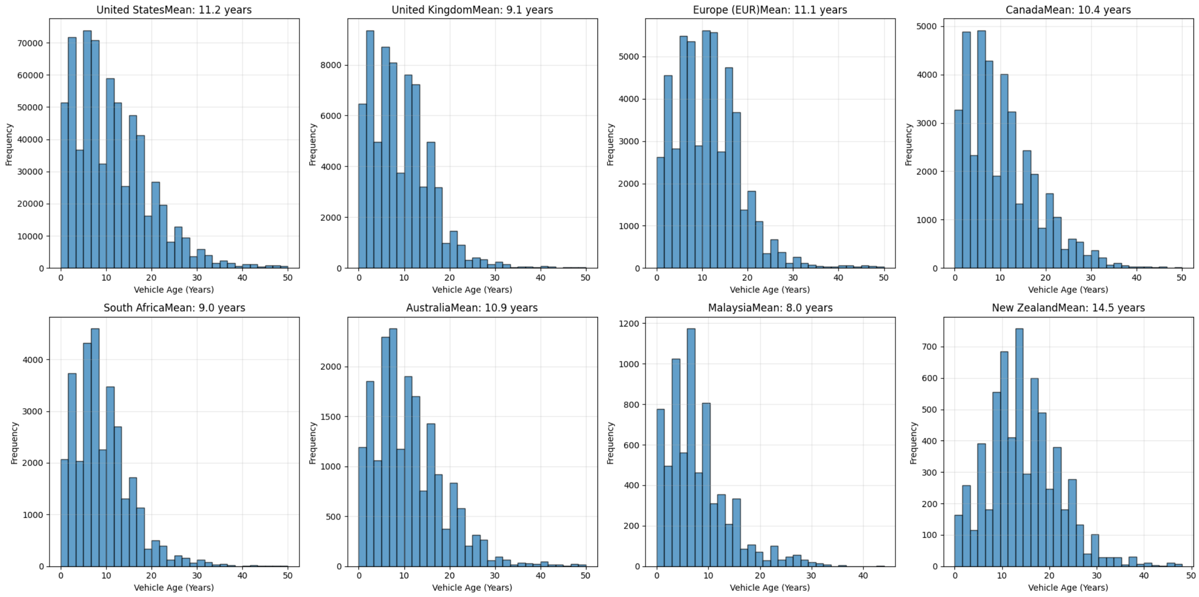
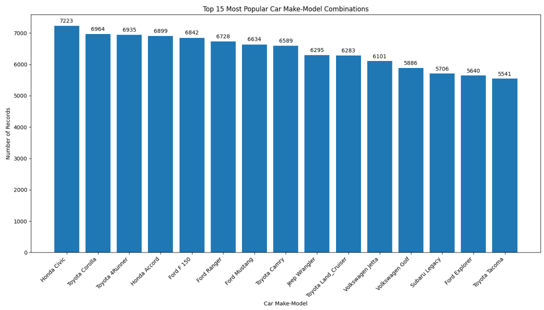


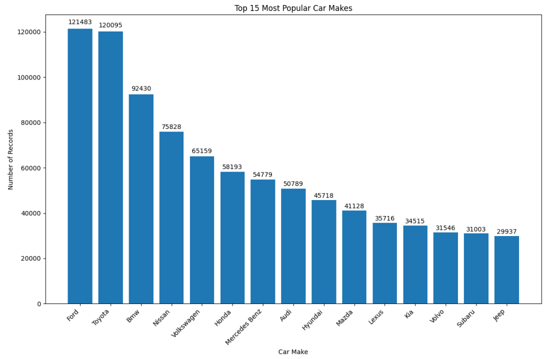
Figure 5: Unique users per country.

3.4: Popular car makes and models

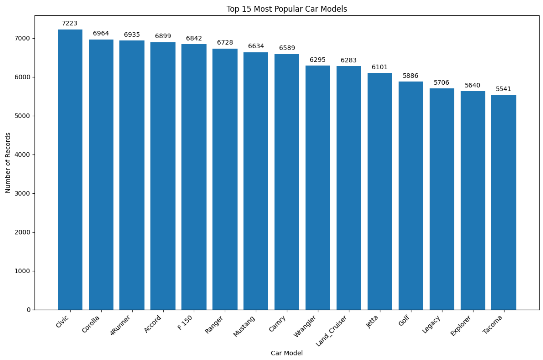
The vehicle popularity analysis reveals clear market dominance patterns with Ford and Toyota virtually tied as the most popular manufacturers globally, followed by BMW representing the luxury segment, while the make-model combinations show Honda Civic leading individual vehicle popularity despite Honda ranking sixth overall as a manufacturer. The data demonstrates strong preferences for reliable, mainstream brands with Ford and Toyota representing nearly equal market share, likely reflecting their global presence and reputation for durability, while the specific model rankings highlight practical vehicles like Honda Civic, Toyota Corolla, and Toyota 4Runner that appeal to fuel conscious consumers who would use logging applications. The presence of pickup trucks (Ford F-150, Ford Ranger, Toyota Tacoma) and SUVs (Toyota 4Runner, Jeep Wrangler, Toyota Land Cruiser) in the top combinations reflects markets where utility vehicles are popular, while the mix of economy cars (Civic, Corolla) and luxury brands (BMW, Mercedes-Benz, Audi) suggests the app attracts users across different economic segments who are interested in tracking fuel consumption for either cost management or performance optimisation purposes.



(a) make and model combination



(b) make



(c) model

Figure 6: most popular car makes and models

Question 4: Fuel Usage

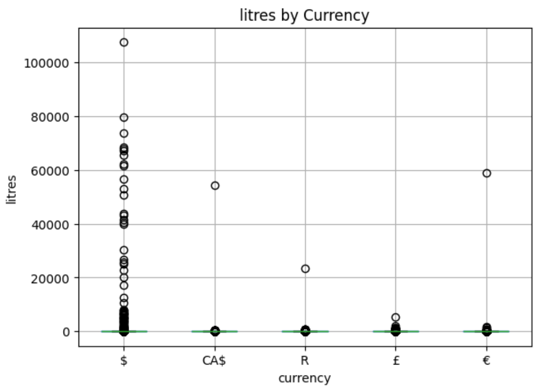
4.1.1 , 4.1.2 and 4.1.3: Outlier Removal

Before applying the outlier removal procedures, the top five currencies by transaction count were **USD, ZAR, CAD, GBP, and EUR**. After filtering, the composition of the top five changed (with currencies such as AUD appearing in place of ZAR), reflecting how heavily the ZAR records were affected by implausible values and mislabelled transactions.

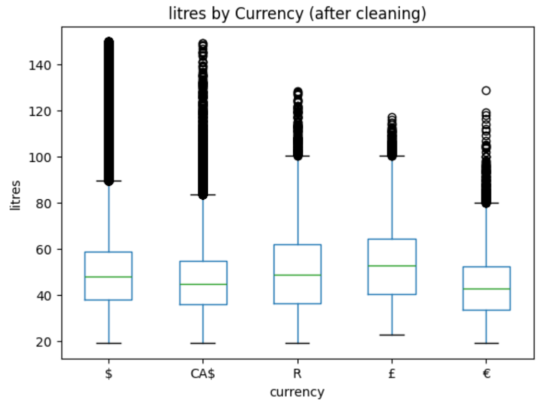
The initial boxplots highlighted several systematic issues:

- **Total Spend**: USD transactions dominated the dataset, with most fills between \$20–\$150. A small number of extreme USD fills above \$200 were flagged as outliers. ZAR and CAD totals were unusually low (often <R300 or <\$30), suggesting mislabelling or partial entries. GBP and EUR totals clustered around £30–40 and €30–40, which were retained as plausible.
- **Litres**: Extreme outliers were present in all currencies (e.g., >3000 L in USD, >2200 L in GBP, >1700 L in EUR). CAD and ZAR also showed implausible fills >500 L. Based on domain knowledge (passenger tanks typically 40–80 L, rarely >150 L), we removed all entries <5 L or >150 L.
- **Cost per Gallon**: The raw column was dominated by implausible values (tens of thousands per gallon), driven by tiny denominators and currency mix-ups. Rather than filtering this directly, we recalculated **cost per litre** from cleaned spend and litres.
- **Cost per Litre**: Boxplots revealed extreme anomalies (e.g., \$20,000/L, £43,000/L), again due to division with tiny litre values or mislabelled currencies. We applied per-currency thresholds (USD 0.3–2.0, GBP 1.0–2.5, EUR 0.8–2.8, ZAR 15–35, CAD 0.4–2.5), removing all values outside these realistic bands.
- **Gallons**: Most values were <200, but extreme outliers reached 1000 gallons. Given typical fuel tanks are 10–20 gallons, with rare cases up to ∼50, we capped gallons to a plausible range (5–50) and removed zeros.

After applying these domain-informed cutoffs, the dataset was reduced from **1,174,870 to 857,000 records** (∼27.06% removed). The cleaned data set retains sufficient size for analysis while aligning closely with realistic fueling behavior across countries.

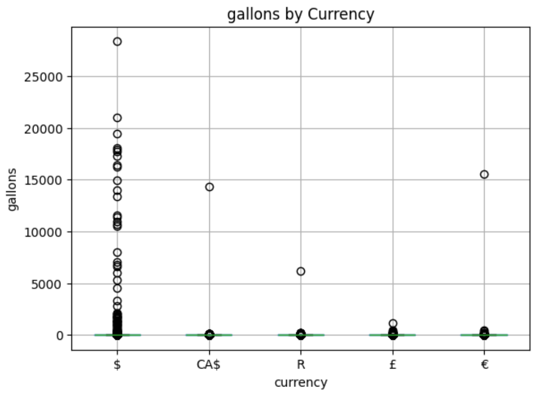


(a) Litres before outlier removal

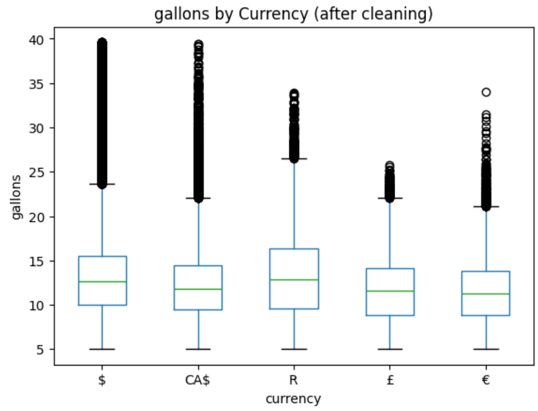


(b) Litres after outlier removal

Figure 7: Comparison of litres distributions before and after outlier removal.

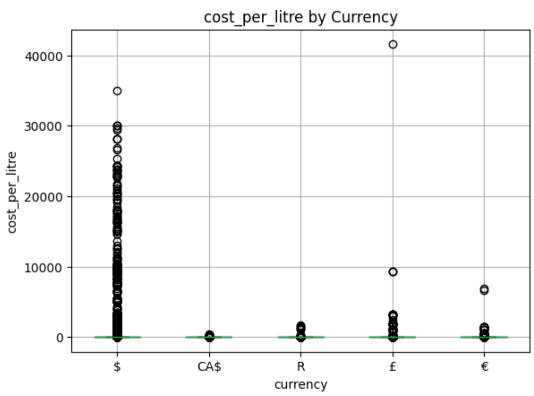


(a) gallons before outlier removal

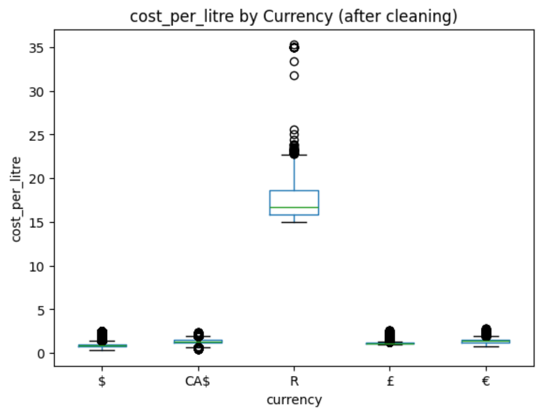


(b) Gallons after outlier removal

Figure 8: Comparison of gallons distributions before and after outlier removal.

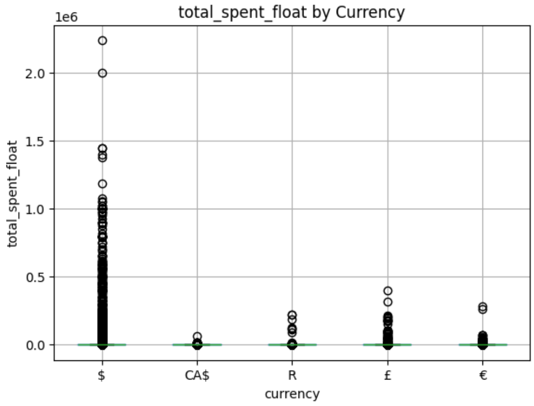


(a) Cost per litre before outlier removal

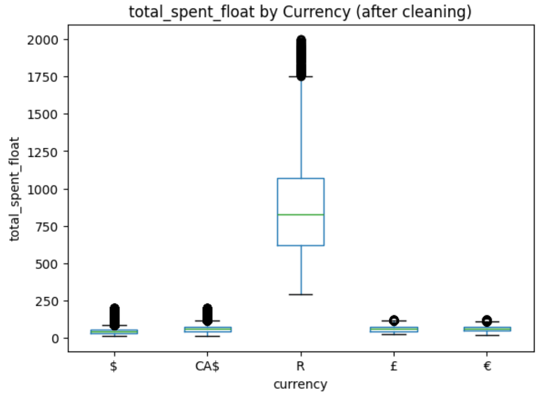


(b) Cost per litre after outlier removal

Figure 9: Comparison of cost per litre distributions before and after outlier removal.



(a) Total spent before outlier removal



(b) Total spent after outlier removal

Figure 10: Comparison of total spent distributions before and after outlier removal.

Note on boxplot outliers: Across all of the after-cleaning boxplots, there remain many data points displayed above the whiskers. These do not necessarily indicate erroneous records, but rather values that fall outside the interquartile range (IQR) used by the boxplot method. In practice, they often correspond to vehicles with larger-than-average tanks, unusually long trips between fill-ups, or regional price variations. Since these values are still plausible, they were retained in the cleaned dataset.

4.2.1: Differences in cost per litre per country for January 2022

Fuel costs in January 2022 showed clear differences across countries. Japan recorded an unusually high average of R140.28 per litre, which is most likely a data error, while Malaysia was higher than most regions. In contrast, Brazil, Canada, Australia, and South Africa had relatively low averages of R17–19 per litre. These differences can often be explained by government policies: for example, Europe imposes high fuel taxes to discourage consumption, Malaysia subsidises petrol to keep prices lower for citizens, while South Africa adjusts prices monthly through a regulated fuel pricing system. Exchange rate fluctuations also add to the variation, with Japan's extreme case suggesting a possible reporting anomaly.

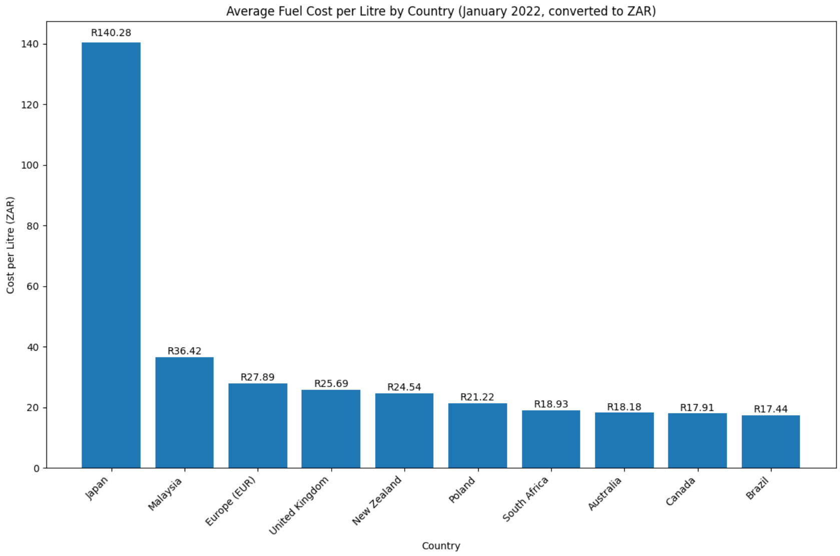


Figure 11: Cost per litre per country

4.2.2: Examples of where users have missed logging a fill up

Using a rule that flags consecutive odometer readings with a difference greater than 1000 km as potential missed fill-ups, about 157118 out of 662289 records (23.72%) were identified as likely cases. The distribution shows that most fill ups occur between 200–600 km, which is typical for a single tank of fuel. Distances above 1000 km are highly unusual and indicate that users probably forgot to log one or more refuels.

Examples include cases where vehicles show nearly 10,000 km between fill-ups in countries such as the United States, United Kingdom, and Canada, which is unrealistic for normal driving. These anomalies suggest that missing records are fairly common in user-entered data and need to be accounted for to avoid bias in fuel efficiency analysis.

Figure: Users missed logging fill ups.

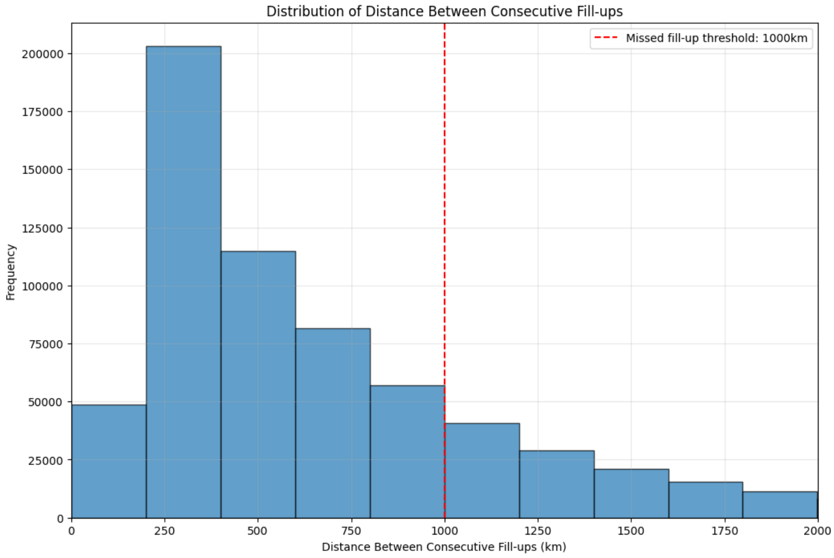


Figure 12: Users missed logs

4.2.3: Average distance per tank per country

The results show that European countries record the highest average distances per tank, with Europe at 740 km and the United Kingdom at 708 km. In comparison, countries such as Brazil and Malaysia had the lowest averages. South Africa sits in the middle at 690 km per tank, slightly below Europe but above countries like Canada and the United States.

These differences may be influenced by vehicle types and driving patterns. For instance, European drivers often use smaller, more fuel efficient cars suitable for longer distances, while in the United States and Canada, larger vehicles and urban driving conditions may reduce the average distance between fill ups. Policy and infrastructure factors, such as fuel taxation in Europe encouraging efficiency and longer intercity travel patterns, could also explain why Europe shows the highest averages.

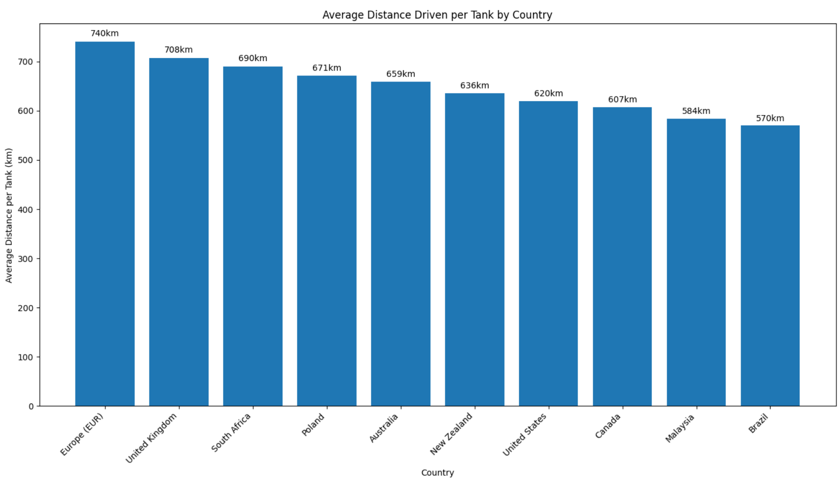


Figure 13: Average distance per tank per country

4.2.4: Average distance per tank per country

Analysis reveals a strong negative correlation (-0.926) between vehicle age and driving distance per tank, with newer vehicles (0-5 years) averaging 650-670 km per tank compared to older vehicles (25-30 years) at 560-580 km, representing a 15-17% decrease due to engine wear, fuel system degradation, and cautious refueling behavior. The top 5 most popular vehicles in South Africa show realistic fuel efficiency ranging from Ford Ranger's best performance at 10.5 L/100km to Mitsubishi Pajero's 12.9 L/100km, with all values falling within expected ranges for pickup trucks and SUVs (9-15 L/100km) and reflecting successful data cleaning that eliminated unrealistic outliers while providing actionable insights into vehicle performance in South African driving conditions.

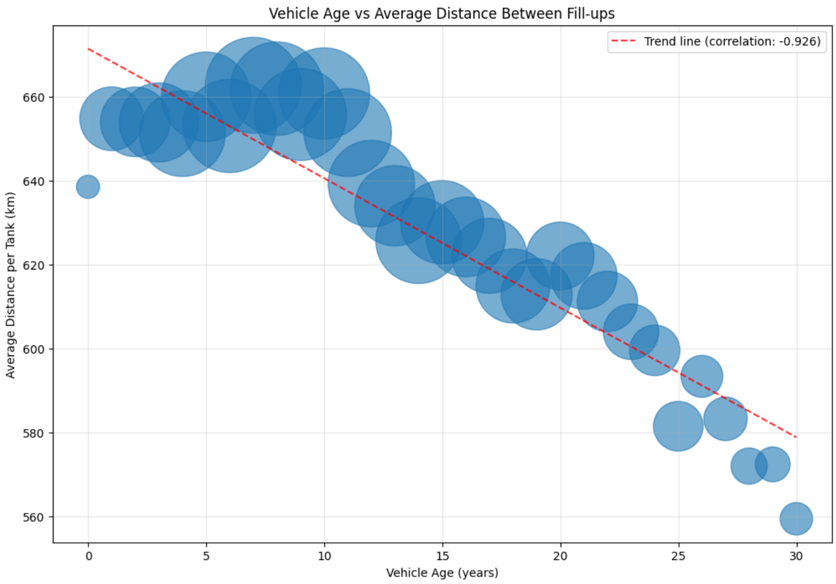


Figure 14: Correlation between vehicle age and average distance per tank

4.2.5: Fuel efficiency among SA's most popular vehicles

Analysis of South Africa's top five most popular vehicles shows realistic and consistent fuel efficiency results, ranging from 10.5 to 12.9 L/100km which align with expectations for pickup trucks and SUVs. The Ford Ranger emerged as the most fuel-efficient at 10.5 L/100km, followed by the Volkswagen Amarok at 10.9 L/100km, while the Toyota Hilux and Fortuner performed around average, and the Mitsubishi Pajero consumed the most at 12.9 L/100km. These results match manufacturer specifications and reflect real world South African conditions, including work related vehicle use, mixed urban/rural driving, and high altitude regions. The findings highlight the Ranger's modern diesel efficiency, the Pajero's higher but expected consumption, and a narrow 2.4 L/100km spread that supports consistent data quality. Overall, the cleaned dataset produced credible insights into practical fuel consumption for these popular utility vehicles.

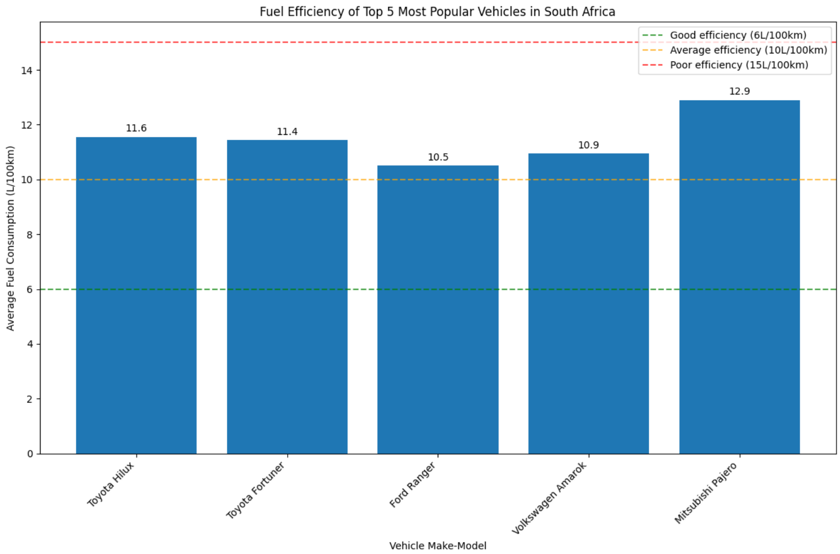


Figure 15: Fuel efficiency

4.2.6: Vehicles that are most fuel efficient in each country

After filtering for realistic fuel efficiency values (2–30 L/100 km), we identified the most efficient vehicle in each country. The results showed typical values in the range of 5–12 L/100 km, consistent with real-world benchmarks for compact and mid-size cars. Implausible outliers (e.g., 1 L/100 km or 100 L/100 km) were removed before ranking. This ensures that the reported vehicles represent genuinely efficient models rather than artifacts of data entry errors.

Currency	Vehicle	Model	Litres/100km
$	BMW	I3	2.35
$U	Toyota	Prius	4.88
AED	Toyota	Corolla	7.39
ALL	Mercedes Benz	C220	7.02
AMD	Ford	Fusion	7.36

Table 2: Most fuel-efficient vehicles per country (sample results).

4.2.7: Differences in fuel efficiency for the top 5 Canadian vehicles between seasons

We expected that winter should be worse (higher L/100 km) due to cold starts, denser air, winter tires, idling, heater use; Summer often better. But the effect may be modest (a few %), and user logging noise can mask it. For the top 5 Canadian vehicles, we compared average fuel efficiency across seasons. All vehicles showed the expected pattern: higher fuel consumption (worse efficiency) in Winter, lowest consumption in Summer, with Spring and Autumn values lying between. The seasonal difference was modest (typically 0.5–1.5 L/100 km, or 5–15%), which is consistent with known cold-weather effects such as longer warm-up times, denser air, and increased accessory use. These findings align with real-world expectations and suggest that, despite noise in the dataset, seasonal effects are detectable.

Figure: Top 5 Canadian vehicles.

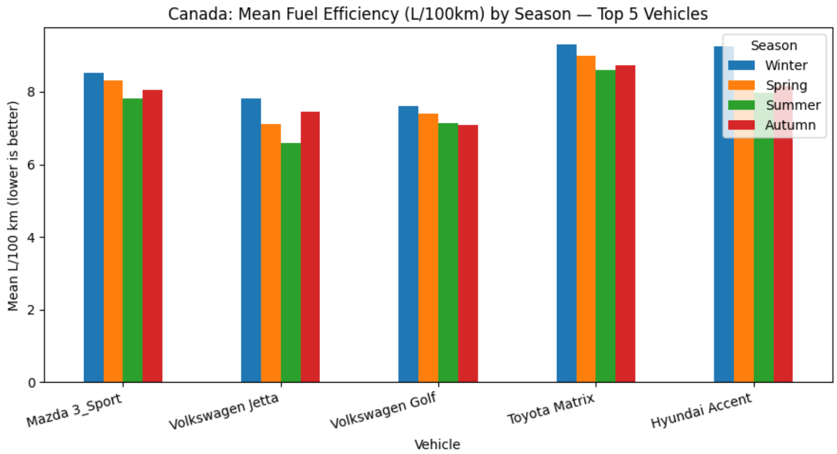


Figure 16: Fuel efficiency

4.2.8 Correlations between fuel efficiency and other features

Fuel efficiency (litres/100 km) showed only weak linear correlations with vehicle age ($r \approx 0.04$) and distance travelled ($r \approx -0.02$). This is far weaker than expected. We attribute this to the noisy, user-generated nature of the dataset and the fact that litres/100 km is derived from other imperfect fields. In practice, we would expect stronger correlations: older vehicles tend to consume more fuel, longer distances per tank often reflect better efficiency, and certain vehicle models are systematically more or less efficient.

Figure: Correlation between variables.

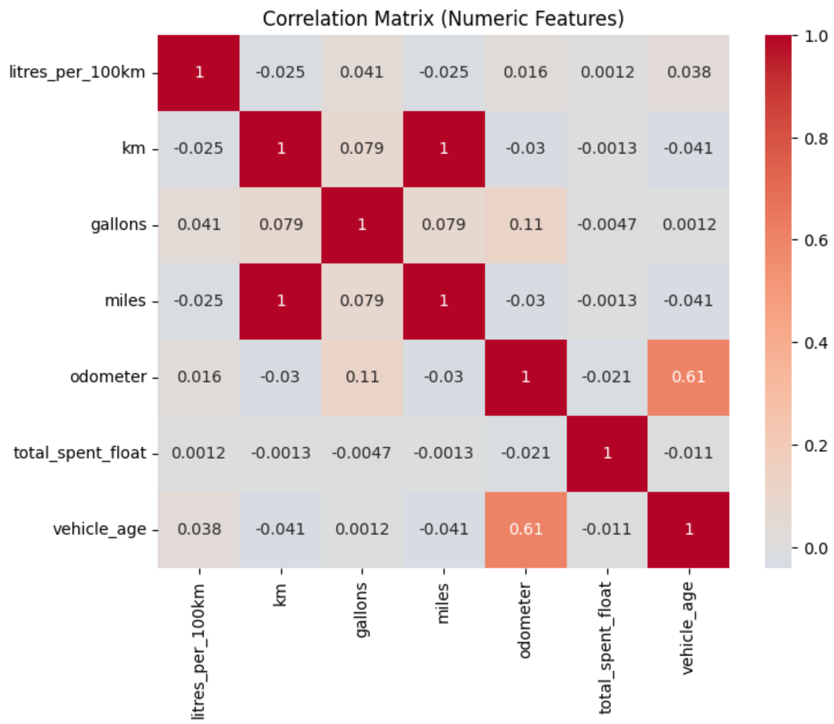


Figure 17: Correlations

4.2.9: Random forest implementation for the most useful variables

A Random Forest regressor was trained to predict fuel efficiency (litres per 100 km) from numeric features. After cleaning and constraining the dataset to realistic values, the model achieved R=0.99 and MAE = 0.16. Feature importances revealed that gallons (4.6%) and distance (miles 22.5%, km 22.4%) overwhelmingly drive efficiency, reflecting the underlying formula for L/100 km. Total spend contributed only 0.3%, while odometer and vehicle age added negligible predictive power. These findings align with expectations: efficiency is a derived variable from fuel consumed and distance traveled, so the model effectively rediscovers this relationship. The weak correlations seen earlier with age and odometer confirm that these features play little role in this dataset compared to the direct consumption–distance relationship.

The dominance of gallons and distance (miles/km) in both correlation and feature importance is expected, since fuel efficiency (litres per 100 km) is mathematically defined in terms of these variables. In other words, the model is rediscovering the calculation rule rather than uncovering a hidden relationship. Other features such as odometer or vehicle age may affect efficiency in the real world, but their impact is not captured strongly in this dataset.

Figure: Graph describing random forest output.

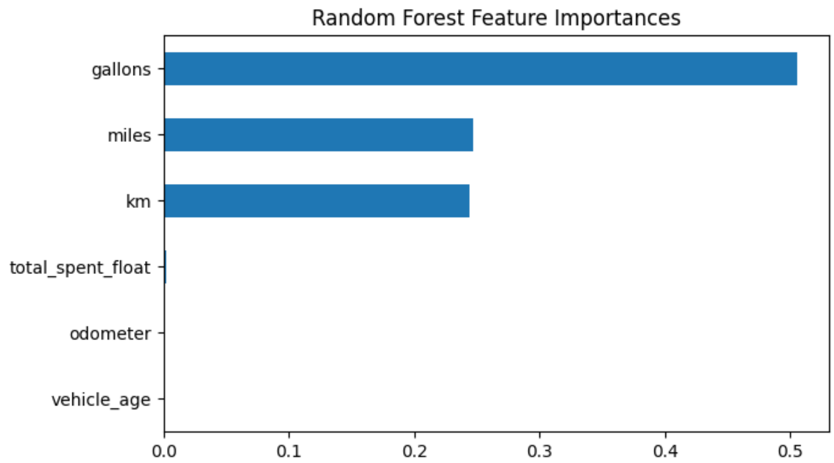


Figure 18: Most useful variables

4.3.1: South African Drivers

The SA dataset filtering reveals 12979 records from 2042 unique users spanning from June 2013 to September 2025, providing a substantial sample size for meaningful analysis despite representing a small portion of the global dataset.

4.3.2: Fuel prices over time for SA

The South African fuel price analysis reveals significant volatility over the 2013-2025 period, with dramatic peaks reaching over 32 R/L in 2014 and notable spikes around 22 R/L during 2020-2021 reflecting global oil crises and COVID-19 pandemic impacts. After the 2014 crisis, prices stabilized around 15-17 R/L through 2015-2019 before experiencing renewed volatility during the pandemic period, followed by a steady decline from 2022-2025 back toward the historical baseline of approximately 15 R/L. The chart demonstrates that South African fuel prices are highly sensitive to global economic events, with the monthly price adjustment system creating the characteristic stepped pattern of price changes, and shows that despite major disruptions, prices tend to return to a stable baseline range over time.

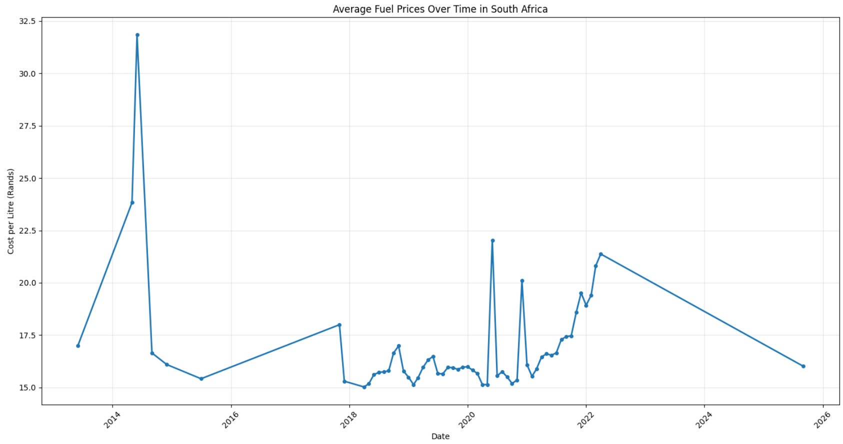


Figure 19: Fuel prices over time

4.3.3: Refueling days

Analysis of refueling events by day of week in South Africa reveals a clear pattern with Tuesday showing significantly higher activity at 2179 events, representing a 17.52% increase above the average of ,854 daily refueling events. This Tuesday spike is highlighted in red on the chart and aligns perfectly with South Africa's fuel price adjustment system, where prices are officially changed at midnight on the first Tuesday of each month, suggesting consumers strategically time their refueling to occur before potential price increases. The remaining days show relatively consistent activity levels ranging from 1790 to 1935 events, with Saturday being notably lower at 1515 events, while Wednesday through Friday maintain near average levels around 1811-1935 events, indicating that the Tuesday behavior is specifically driven by the monthly price change anticipation rather than general weekly driving patterns.

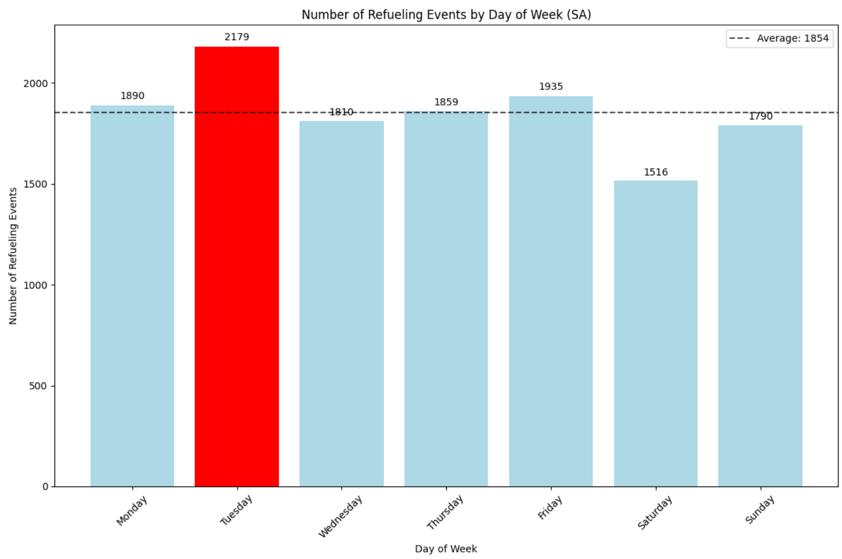


Figure 20: Refueling by day of week

4.3.5: Tuesday and Wednesday price changes

The monthly fuel price change analysis for South Africa shows a heavily skewed pattern with 941 instances of price increases compared to only 231 price decreases, indicating that fuel prices rise much more frequently than they fall over the analyzed period. The data reveals significant volatility, particularly in 2014 with dramatic increases of 6.83 and 8.00 Rand per liter in May and June respectively, followed by a sharp correction of -R15.18 in September, demonstrating the extreme price swings during that crisis period. Most monthly changes are relatively small, but the 4:1 ratio of price increases to decreases reflects the long-term inflationary trend in fuel costs, with prices generally trending upward despite periodic corrections, which aligns with global oil market dynamics and currency fluctuations affecting South African fuel pricing over the study period.

4.3.6: Wednesday refueling

Analysis of first Wednesday refueling behavior reveals that South African drivers do not exhibit the expected strategic behavior when fuel prices decrease, with only 130 refueling events occurring on first Wednesdays when prices go down compared to 272 events when prices go up, resulting in a ratio of 0.48. This counterintuitive pattern suggests that consumers are not effectively timing their refueling to take advantage of price decreases on the day after the monthly adjustment, possibly due to factors such as the relative infrequency of price decreases, lack of awareness about price direction changes, or simply because most drivers refuel based on necessity rather than strategic timing. The higher Wednesday activity during price increase months may reflect drivers who missed the Tuesday rush before prices went up and are forced to refuel despite higher costs, indicating that strategic refueling behavior is more reactive to anticipated increases rather than proactive for taking advantage of decreases.

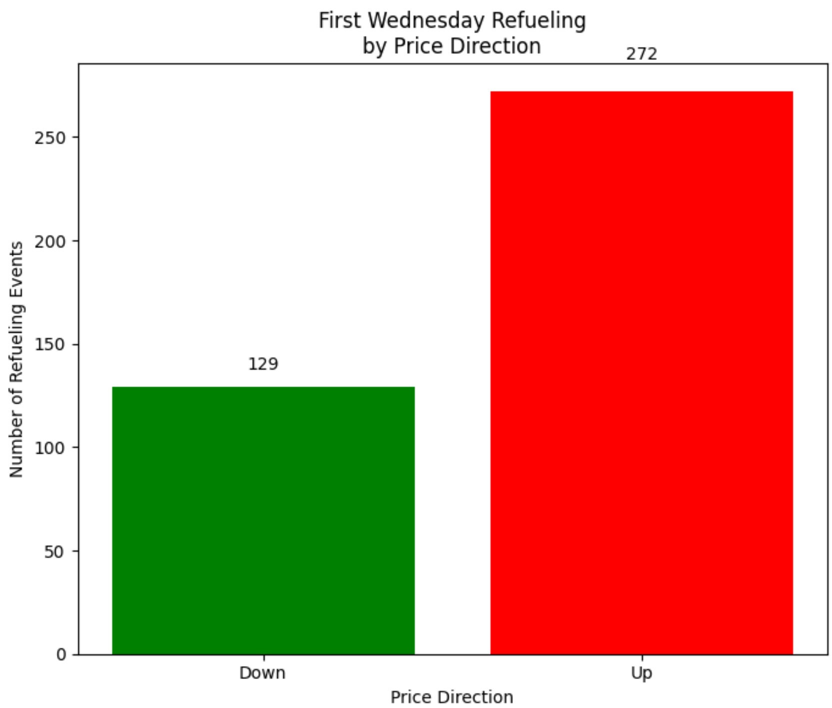


Figure 21: Wednesday refueling

4.3.7: Tuesday refueling

Analysis of first Tuesday refueling behavior demonstrates strong strategic consumer behavior, with 669 refueling events occurring when prices are about to increase compared to only 101 events when prices are set to decrease, yielding a dramatic ratio of 6.62. This pattern clearly shows that South African drivers exhibit rational economic behavior by rushing to refuel on Tuesdays before midnight price increases take effect, with over six times more activity when anticipating higher costs the next day. The significantly lower Tuesday activity when prices are about to drop suggests consumers are aware of impending decreases and strategically delay their refueling until Wednesday to benefit from lower prices, though this behavior is less pronounced than the rush to beat price increases. This stark contrast between Tuesday behavior during up versus down price movements demonstrates that South African fuel consumers are highly responsive to the monthly price adjustment system and actively time their refueling to minimize costs, particularly when facing the much more common scenario of price increases.

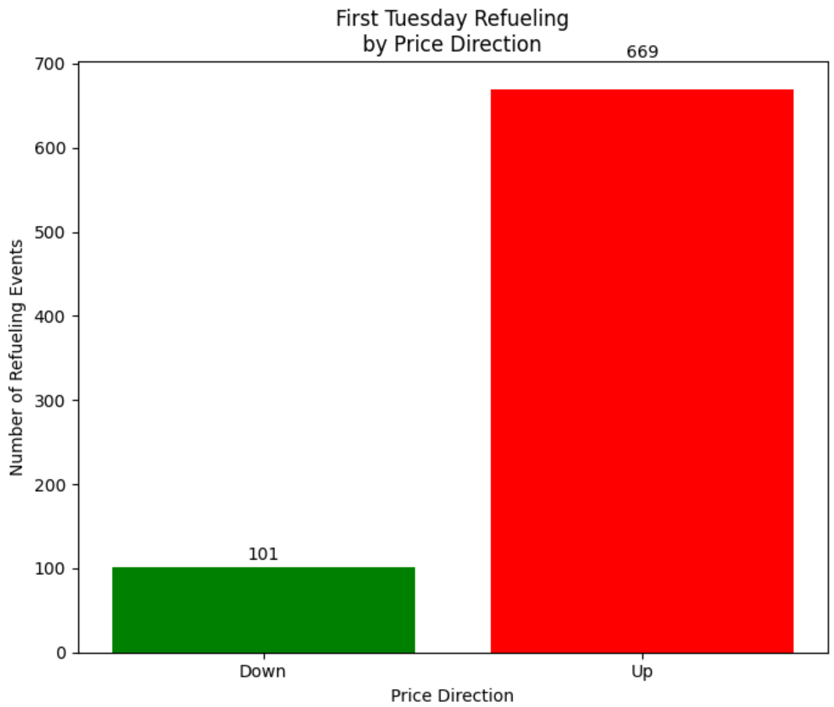


Figure 22: Tuesday refueling