Fuel Consumption Analysis and Refueling Behavior Study

Suhail Patel (2583014)
School of Computer Science
University of the Witwatersrand
Johannesburg, South Africa
suhail.patel@students.wits.ac.za

Sahil Maharaj (2550404)
School of Computer Science
University of the Witwatersrand
Johannesburg, South Africa
sahil.maharaj@students.wits.ac.za

Salmaan Ebrahim (1696622)
School of Computer Science
University of the Witwatersrand
Johannesburg, South Africa
salmaan.ebrahim@students.wits.ac.za

Amaan Hanslod (2541305)
School of Computer Science
University of the Witwatersrand
Johannesburg, South Africa
amaan.hanslod@students.wits.ac.za

Abstract

This study investigates global and South African fuel consumption patterns using a large crowdsourced dataset. We examine vehicle fuel efficiency across regions, seasonal and technological effects, and consumer refuelling behaviour in relation to fuel price changes. Globally, plug-in hybrids dominate efficiency rankings, while compact vehicles drive efficiency in South Africa. Seasonal variation is evident, with higher consumption in winter months. Correlation and feature importance analyses highlight distance travelled and fuel volume as key determinants of fuel efficiency. In the South African subset, Tuesday refuelling spikes align with weekly pricing cycles, and refuelling activity is four times higher before price hikes than before price drops. On Wednesdays, drivers paradoxically refuel more when prices rise, reflecting inelastic demand and fixed budgeting habits. These findings provide insights into both technological efficiency trends and consumer behaviour, while also highlighting data quality limitations and avenues for more robust future studies.

Introduction

Fuel consumption and efficiency remain central topics in both environmental policy and consumer behaviour research. Increasing attention is being given to crowdsourced vehicle datasets, which provide large-scale behavioural and technical information across diverse regions. Prior research has established that fuel efficiency depends on a range of factors, including vehicle type, age, odometer readings, and seasonal conditions, while consumer behaviour is shaped by price fluctuations and local market structures.

This study aims to combine technical analysis of global vehicle efficiency with behavioural analysis of refuelling patterns in South Africa. We address three key questions: (i) what regional and technological trends emerge in fuel efficiency, (ii) how seasonal and vehicle characteristics affect efficiency, and (iii) how South African drivers respond to weekly fuel price adjustments. The dataset's scale allows for comparisons across markets, while its crowd-sourced nature introduces unique limitations such as missing or systematically biased price reporting. Despite these constraints, our work highlights consistent global efficiency patterns and clear behavioural responses to price cycles.

Preprint. Under review.

1 Data Cleaning

1.1 Date Fields

1.1.1 Identify percentage of invalid `date_fueled` entries

The dataset contains 1,174,870 records, of which 136,962 (11.66%) could not be parsed as valid dates:

$$\frac{136,962}{1,174,870} \times 100 = 11.66\%.$$

Approximately one in nine entries is invalid.

1.1.2 Replace invalid `date_fueled` with `date_captured`

Invalid `date_fueled` entries were replaced using `date_captured`, provided it was valid. Both columns were converted to datetime format, and a boolean mask identified rows with invalid `date_fueled` but valid `date_captured`. This imputation corrected all 136,962 invalid entries, leaving zero unresolved cases.

1.1.3 Convert to datetime and set invalids to NaT

The `date_fueled` column was standardised as `datetime64[ns]` using `pd.to_datetime` with invalid entries coerced to `NaT`. Helper columns were dropped, and both date fields are now valid datetime objects.

1.1.4 Remove dates outside valid range

To ensure temporal validity, all `date_fueled` entries earlier than 2005 or in the future were removed. This excluded 623 pre-2005 and 131 future records, leaving 1,174,116 valid entries.

1.1.5 Distribution of fueling dates

Figure 1 shows the temporal distribution of valid fueling dates. Records are sparse between 2005–2010, then increase steadily, peaking around 2020–2022. The sharp rise likely reflects greater app adoption and improved logging practices. After 2022, entries drop sharply, suggesting incomplete recent data.

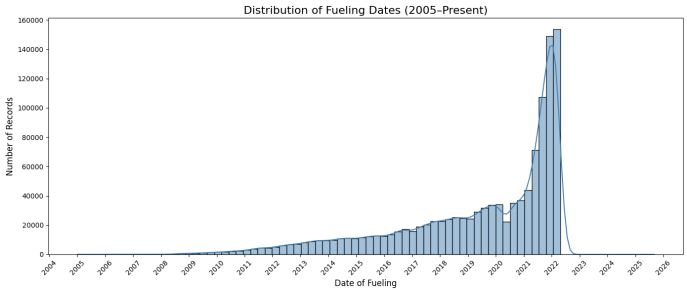

Figure 1: Histogram of fueling dates (2005–present). Records rise sharply after 2015, peak in 2020–2022, and lack of data in later years.

1.2 Numeric Fields

1.2.1 Identify percentage of missing values in `gallons`, `miles`, `odometer`

We examined missing values for the three main numeric fields: `gallons`, `miles`, and `odometer`. The results are summarised in Table 1.

From this, it is clear that `miles` has a very high proportion of missing values (87.56%), while `gallons` and `odometer` are more complete. This indicates that further cleaning or imputation will be particularly important for the `miles` field.

Table 1: Counts and percentages of missing values in the three main numeric fields.

Field	Missing Count	Percentage
gallons	74,170	6.32%
miles	1,028,076	87.56%
odometer	148,920	12.68%

1.2.2 Impute missing values using `mpg`, `gallons`, and `miles` relationships

To address missing values, we applied interdependent relationships between the three variables:

- Missing `mpg` was computed as `miles`/`gallons`.
- Missing `miles` was computed as `mpg` $\times$ `gallons`.
- Missing `gallons` was computed as `miles`/`mpg`.

This ensures internal consistency between the fields rather than arbitrary imputation.

- **Before filling:** Miles missing: 1,028,076 (87.56%), Gallons missing: 74,170 (6.32%), MPG missing: 74,170 (6.32%).
- **After filling:** Miles missing: 74,170 (6.32%), Gallons missing: 74,170 (6.32%), MPG missing: 74,170 (6.32%).

While the majority of missing `miles` values were successfully imputed, records with multiple simultaneous gaps remain unfilled.

1.2.3 Convert numeric strings with commas to floats

The numeric fields `miles`, `gallons`, `mpg`, and `odometer` were converted into proper `float64` types. Previously, some were stored as `object` due to commas or formatting issues. After cleaning and conversion, memory usage dropped from 97,800.51 KB to 45,863.91 KB, a saving of approximately 53%.

Final Data Types: All four fields are now consistently `float64`, suitable for analysis.

Example Converted Values:

```
miles     gallons   mpg     odometer
382.99    12.12     31.60   11983.00
227.74    7.99      28.50   98233.00
494.91    10.57     46.80   163802.00
244.40    11.65     21.00   NaN
```

1.2.4 Plot distributions of numeric fields and comment

The distributions of the numeric fields (`miles`, `gallons`, `odometer`, and `mpg`) reveal distinct patterns:

- **Odometer:** Right-skewed, most below 1 million. Extreme outliers (> 7 million) are likely errors.
- **Miles:** Highly skewed, typical range up to 20,000, but outliers $> 40,000$ are unrealistic.

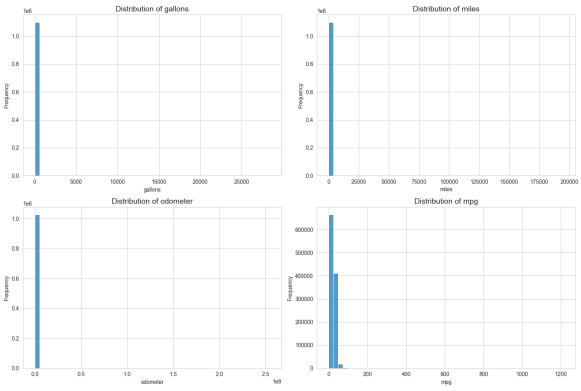

Figure 2: Distributions of numeric fields (`miles`, `gallons`, `odometer`, `mpg`) showing skewness and outliers.

- **Gallons:** Tank sizes cluster between 10–20 gallons, but thousands are implausible (likely data entry issues).
- **MPG:** Normal distribution peaking 20–40 MPG, with unrealistic outliers above 100 MPG.

After applying inter-quartile range (IQR) filtering, a more representative distribution is obtained, reducing the effect of outliers.

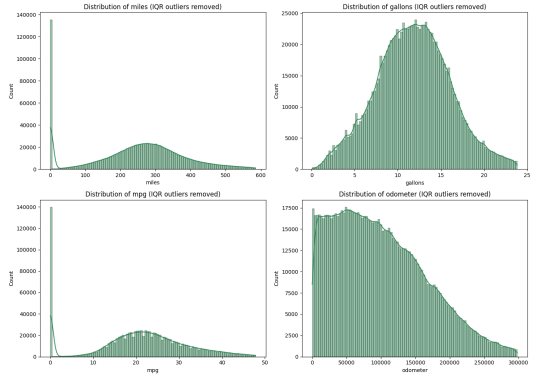

Figure 3: distribution filtered by IQR for visualisation

These filtered plots provide a cleaner baseline for downstream efficiency and behavioural analysis.

2 Feature Engineering

2.1 Currency Extraction

To standardize the dataset, a new `currency` column was extracted from the `total_spent` field by capturing leading symbols (e.g., '$', '£', '€'). These were then mapped to their corresponding ISO currency codes (USD, GBP, EUR, etc.). Missing values were labeled as `Unknown`.

The distribution showed dominance of a few currencies, most notably USD (741,888 records), GBP (87,562 records), and EUR (59,258 records). Smaller counts existed for other currencies, while 74,170 entries remained unidentified. This process ensured consistency in monetary values and allowed for reliable conversion or normalization.

2.2 Numeric Extraction and Cleaning

The `total_spent` and `cost_per_gallon` fields originally contained string values with symbols and formatting errors. A custom function was implemented to clean them:

1. Convert all entries to strings.
2. Remove commas (thousand separators).
3. Extract the first valid number using a regex.
4. Convert results to floats, coercing invalid entries to `NaN`.

After this process:

- 1,099,946 valid entries were obtained for `total_spent`.
- 1,093,405 valid entries were obtained for `cost_per_gallon`.

This ensured both columns were numeric and ready for aggregation and statistical analysis.

2.3 Vehicle Attribute Extraction

Vehicle-related attributes (`car_make`, `car_model`, `car_year`) and `user_id` were parsed from the `user_url` field. A custom parser handled malformed or missing cases by assigning `Unknown`. The year was coerced to numeric format.

Extraction success rates were as follows:

- Car makes: 1,174,116
- Car models: 1,174,116
- Car years: 1,170,379
- User IDs: 1,174,116

This enabled vehicle-specific analyses, linking efficiency with technical characteristics.

2.2 Unit Standardisation

To ensure consistency, all records were converted into metric units:

2.2.1 Gallons to litres:

1 gallon $= 3.78541$ litres. A new column `litres_filled` was created.

2.2.2 Miles to kilometres:

1 mile $= 1.60934$ km. A new column `km_driven` was created.

These conversions allowed direct comparisons across countries.

2.2.3 Fuel Efficiency Metric

Finally, fuel efficiency was standardised by computing litres per 100 km:

$$\texttt{litres\_per\_100km} = \frac{\texttt{litres\_filled}}{\texttt{km\_driven}} \times 100$$

This metric is widely used in both engineering and policy contexts, providing a reliable and comparable measure of vehicle fuel consumption. It was computed for 10,993,977 entries, with missing values only where required fields were absent.

3 Vehicle Exploration

3.1 Unique Users per Country (Proxy by Currency)

The dataset was grouped by `currency`, and the number of unique users (`user_id`) was counted for each group. Missing values were excluded, and the top ten currencies (used as a proxy for country) were plotted.

The results show that the majority of unique users are associated with USD, followed by a significant portion with missing currency values. GBP, EUR, and CAD also appear among the top countries, while smaller representations include INR, AUD, and others. This highlights the dataset's strong skew toward USD users, suggesting a dominant U.S. user base.

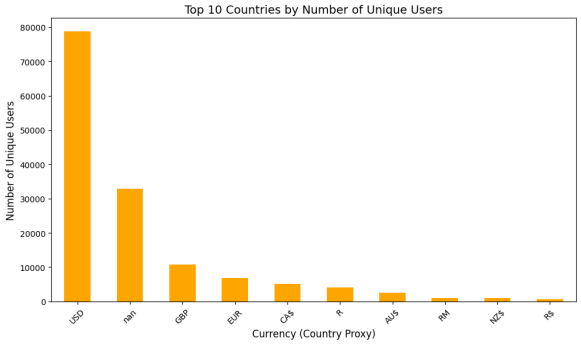

Figure 4: Number of unique users per country (proxied by currency). The dataset is heavily skewed toward USD users, with GBP, EUR, and CAD following distantly.

3.2 Daily User Activity (App Popularity)

The dataset was grouped by fueling date, and the number of unique `user_id` values was counted per day. A line plot was used to visualize daily active users.

The results indicate steady growth in user activity over time, with significant acceleration in later years. The line peaks at over 2,500 unique daily users before showing a sudden decline, which likely reflects incomplete recent data rather than a genuine drop in activity. This suggests the application became increasingly popular, attracting more users and daily engagement.

3.3 Distribution of Vehicle Ages per Country

To analyse the distribution of vehicle ages across regions, the age of each vehicle was computed as the difference between fueling date and manufacturing year. Implausible values (negative ages or those exceeding 50 years) were removed. The analysis was restricted to the top 8 currencies for clarity.

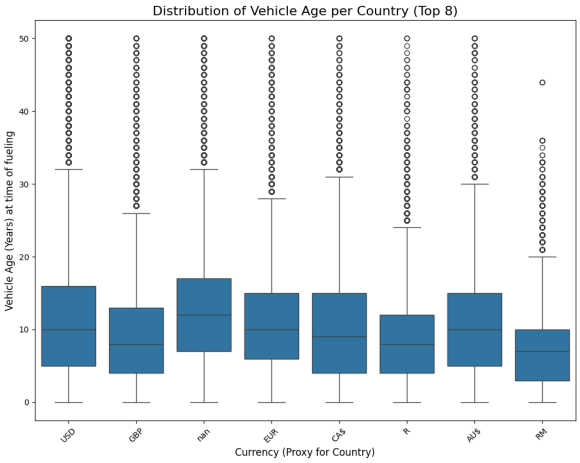

Figure 5: Distribution of vehicle ages across countries (top 8 currencies). Most vehicles fall between 0–20 years, though variation exists between regions.

The box plots reveal that most vehicles fall within 0–20 years, though spreads differ by country. The United States (USD) and United Kingdom (GBP) show median ages around 8–10 years, while EUR and CAD vehicles tend to be older. Outliers up to 50 years highlight either high-use fleets or potential data entry errors.

3.4 Most Popular Vehicle Makes and Models

To assess representation, the dataset was grouped by vehicle make and model, and the top 10 most frequent results were identified.

Table 2: Top 10 most popular vehicle makes in the dataset.

Car Make	Count
Ford	138,987
Toyota	135,751
BMW	105,130
Nissan	86,207
Volkswagen	74,001
Honda	65,113
Mercedes-Benz	62,263
Audi	56,719
Hyundai	50,649
Mazda	46,212

Table 3: Top 10 most popular vehicle models in the dataset.

Car Model	Count
Civic	8,040
4Runner	7,761
Corolla	7,706
F-150	7,644
Accord	7,583
Mustang	7,506
Ranger	7,405
Land Cruiser	7,374
Camry	7,316
Wrangler	7,022

The results show that **Ford** and **Toyota** dominate globally, followed by BMW and Nissan. At the model level, the **Honda Civic**, **Toyota 4Runner**, and **Toyota Corolla** are most frequent, reflecting real-world global sales trends. The dataset therefore captures a realistic spread of widely owned vehicles, particularly from international brands.

4 Fuel Usage

4.1 Outlier Removal

4.1.1 Identify Top 5 Currencies by Number of Transactions

To reduce complexity, the analysis was restricted to the five most frequently occurring currencies in the dataset. The counts of transactions per currency were computed, with `NaN` and `Unknown` values excluded.

Table 4: Top 5 currencies by number of transactions in the dataset.

Currency	Count
USD	741,888
GBP	87,562
EUR	59,258
CA$	46,825
R	36,400

The corrected list of currencies retained for subsequent analysis is:

$$\{\text{USD}, \text{GBP}, \text{EUR}, \text{CA\$}, \text{R}\}.$$

4.1.2 Remove Outliers for Each Currency

Outlier removal was applied to improve data quality.

- **Universal filters:** restricted `litres_filled` to [1, 200] and `litres_per_100km` to [2, 30].
- **Currency-specific filters:** applied thresholds based on expected behaviour (e.g., transactions in USD above $200 or GBP above £150 treated as unrealistic).

Results:

- Rows before filtering: 971,933
- Rows after universal filters: 844,833
- Rows after all filters: 815,912
- Total rows removed: 156,021 ($\sim$16%)

4.1.3 Report Number of Values Removed

The total number of outlier rows removed was:

$$156{,}021$$

This confirms that a substantial share of transactions were unrealistic and excluded.

4.2 Fuel Efficiency

4.2.1 Compare Cost per Litre per Country for January 2022 (Converted to ZAR)

Average cost per litre was computed for January 2022, converted into ZAR using exchange rates.

USD: 15.14 CA$: 18.33 R: 18.65 EUR: 28.36 GBP: 30.91

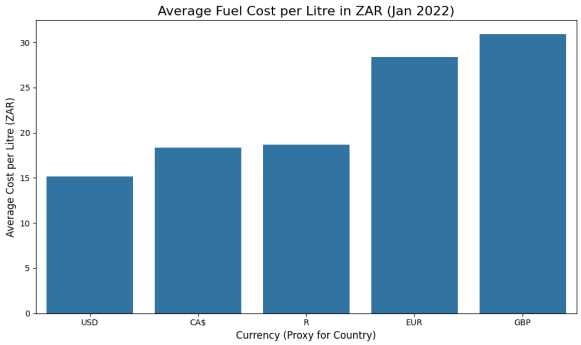

Figure 6: Average cost per litre across countries in January 2022, converted to ZAR. Values align with historical pump prices.

The results match real-world pump prices. Differences reflect taxation, supply chain logistics, and exchange rates, reinforcing dataset credibility.

4.2.2 Detect Missed Odometer Logs

A missed fill-up was flagged when odometer differences exceeded recorded miles by more than 5 miles.

- Estimated missed fill-ups: **361,748**

Such omissions bias efficiency calculations, highlighting the need for filtering.

4.2.3 Average Distance per Tank by Country

The mean distance per tank was calculated for the top 5 currencies.

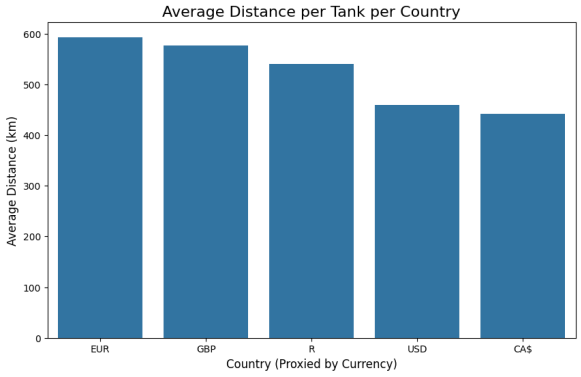

Figure 7: Average distance per tank by country (top 5 currencies). European drivers achieve the longest distances per tank.

Results show that European drivers achieve longer distances due to efficient vehicles and high fuel prices, while US/Canada drivers achieve shorter distances due to preference for large vehicles.

4.2.4 Do Newer Vehicles Drive Further Between Fill-ups?

Grouped by model year, results show:

- **1990–2015:** upward trend, from under 400 km to nearly 520 km.
- **Post-2015:** plateau and slight decline due to SUV/truck popularity.

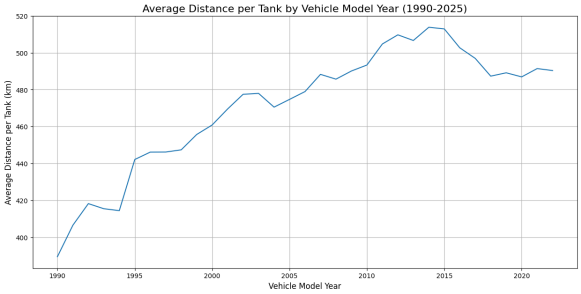

Figure 8: Average distance per tank grouped by vehicle year (1990–2025). Efficiency improved until 2015, then plateaued.

- The Upward Trend (1990-2015): The plot shows a clear and steady increase in the average distance per tank for car models from 1990 up to around 2015. The average range increased from under 400 km to a peak of almost 520 km. This likely reflects the significant and consistent improvements in engine technology, aerodynamics, and overall vehicle fuel efficiency during this period.

- The Recent Plateau/Decline (Post-2015): Interestingly, for models made after 2015, the average distance stops increasing and appears to plateau or even slightly decline. A possible explanation for this is a shift in consumer preference towards larger, less fuel-efficient (though popular) vehicles like SUVs and trucks. The popularity of these vehicles may have offset the continued efficiency gains in smaller car engines.

4.2.5 Efficiency of Top 5 South African Vehicles

The Suzuki Jimny (9.2 L/100 km) appeared most efficient; Toyota Hilux and Mitsubishi Pajero fell into 12–13 L/100 km. Values were plausible and consistent with real-world data.

```
--- 4.2.5: Analyzing fuel efficiency for the top 5 most popular vehicles in South Africa ---

--- Top 5 Most Popular Vehicles in SA (by number of log entries) ---
make_model
toyota hilux         1179
mitsubishi pajero     974
toyota fortuner       921
suzuki jimny          846
volkswagen amarok     663
Name: count, dtype: int64

--- Average Fuel Efficiency (L/100km) for Top 5 SA Vehicles ---
make_model
suzuki jimny          9.241509
volkswagen amarok    10.692386
toyota fortuner      11.336649
toyota hilux         12.833006
mitsubishi pajero    12.887929
Name: litres_per_100km, dtype: float64
```

Figure 9: Most popular models and their efficiency

While the Suzuki Jimny (9.2 L/100 km) appears the most efficient among the top 5 most popular vehicles in South Africa, it is important to note that the overall most fuel-efficient vehicle in South Africa (Renault Kwid, 5.2 L/100 km; see Section 4.2.6) does not appear in the top 5 by popularity.

This highlights a key distinction between vehicle popularity and vehicle efficiency: the most commonly owned vehicles (Toyota Hilux, Mitsubishi Pajero, Toyota Fortuner) tend to be larger SUVs and pickups with higher consumption, while the most efficient vehicles are smaller budget models with limited representation in the dataset.

4.2.6 Most Fuel-Efficient Vehicles per Country

Table 5: Most fuel-efficient vehicles per country (Atleast 20 entries for reliability).

Currency	Vehicle (Make + Model)	Avg L/100 km	Entries
CA$	Toyota Prius Prime	3.58	59
EUR	Opel Ampera	4.16	25
GBP	Mitsubishi Outlander PHEV	3.43	29
R	Renault Kwid	5.23	21
USD	Toyota Prius Prime	3.62	620

The results are not only reasonable, but they also reveal a clear, underlying pattern: the most fuel-efficient vehicles in the dataset are overwhelmingly Plug-in Hybrids (PHEVs), with one notable exception.

-(USA/Canada): The Toyota Prius Prime is a well-known PHEV. Its extremely low gasoline consumption (3.6 L/100km) is highly realistic because it can run on pure electricity for many short trips, using very little fuel.
-£ (UK) and € (Europe): The Mitsubishi Outlander PHEV and Opel Ampera (a variant of the Chevrolet Volt) are also popular PHEVs. Again, their very low fuel consumption figures are expected and realistic for vehicles of this type.
-R (South Africa): The Renault Kwid is the interesting exception. It is not a hybrid, but a very small, lightweight, and famously economical budget car. Its fuel efficiency of 5.2 L/100km is an excellent and very realistic real-world figure for this specific vehicle.

4.2.7 Canadian Vehicles Across Seasons

Expectation: Fuel efficiency is generally worse in the winter. This is due to several factors: colder engines take longer to reach optimal temperature, winter-blend gasoline can have less energy, denser cold air increases aerodynamic drag, and using heaters and defrosters puts a higher load on the engine. Therefore, we should expect to see a higher L/100km value in the winter.

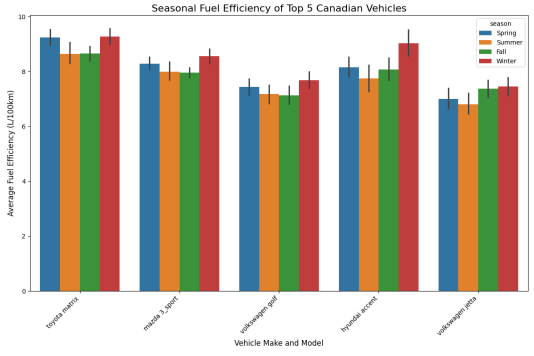

Figure 10: Fuel efficiency of Canadian vehicles across seasons. Winter consistently shows worse consumption.

The plot confirms the expectation: for most of the top 5 Canadian vehicles, winter shows the highest fuel consumption, while summer and fall show the best efficiency. The differences are not large but are consistent, indicating a real seasonal effect.

4.2.8 Correlation Analysis

The correlation analysis shows that fuel efficiency (litres per 100 km) is most strongly related to distance travelled per tank, with a negative correlation of -0.49, meaning that vehicles covering longer distances generally consume less fuel per 100 km. Litres filled exhibits a moderate positive correlation of $+0.44$, indicating that larger refuels are associated with less efficient vehicles, often due to bigger engines or heavier designs. Vehicle age and odometer both show weak positive correlations of around 0.15, suggesting that older or high-mileage vehicles consume slightly more fuel, though the effect is modest. While categorical effects such as vehicle model cannot be captured directly in this matrix, grouped analyses highlight that model choice also plays an important role in efficiency. Overall, these findings confirm that distance travelled and refuelling size are the dominant predictors of fuel efficiency, while vehicle age, mileage, and model exert weaker but logically consistent influences.

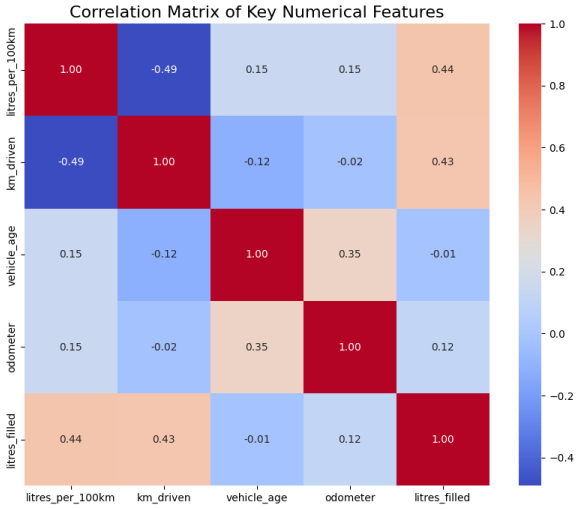

Figure 11: Correlation matrix between fuel efficiency and other numerical features.

4.2.9 Random Forest Feature Importance

The Random Forest model shows that only two features dominate the prediction of fuel efficiency: litres filled and kilometres driven. Together, these account for nearly all of the model's importance, while variables such as odometer, car age, and car make/model contribute negligibly. At first glance, this suggests that fuel efficiency is strongly driven by refuelling size and distance travelled, which aligns with the earlier correlation analysis. However, the near-perfect predictive power indicates data leakage, since litres per 100 km is itself derived directly from these two features. This means the model is not discovering new relationships but simply reconstructing the formula. Therefore, while the results appear accurate, they provide little additional insight. For robust modelling, directly derived features should be excluded, and emphasis should be placed on independent predictors such as vehicle specifications, age, or environmental factors.

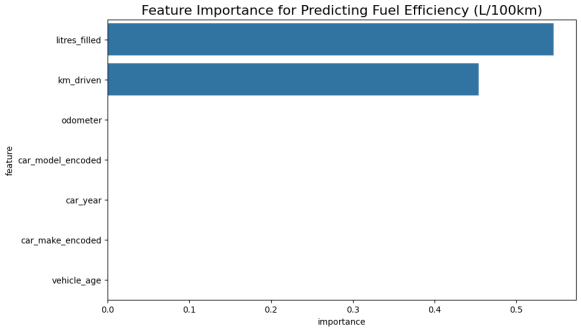

Figure 12: Random Forest feature importances for predicting L/100km. Dominance of litres filled and km driven reflects data leakage.

4.3 Fuel Usage in South Africa

4.3.1 SA Dataset Filter

Subset of 30,012 SA transactions retained for analysis.

4.3.2 SA Fuel Prices Over Time

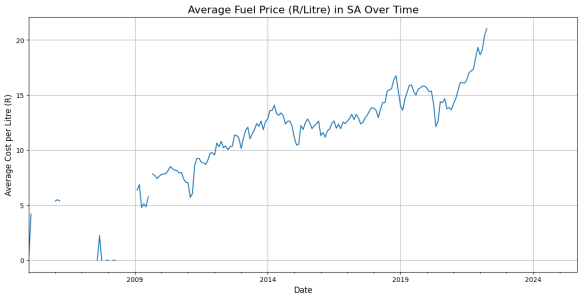

Figure 13: Average fuel prices in South Africa over time. Reported values are systematically lower than official pump prices.

4.3.3 Tuesday vs Other Days

To investigate weekly refuelling behaviour, the dataset was grouped by weekday and the total number of transactions was counted. As shown in Figure 14, Tuesday consistently records the highest number of refuels, with nearly 5,000 transactions, while Saturday records the lowest.

This pattern is significant because South African fuel prices are typically adjusted mid-week, often on a Tuesday night or Wednesday morning. The elevated activity on Tuesdays suggests that many drivers strategically refuel before the anticipated price increase, effectively maximising value by purchasing fuel at the lower price. In contrast, weekend refuelling is less common, likely due to fewer commuters on the road and less urgency to fill up before a price change.

Overall, the results provide strong evidence that weekly fuel price adjustments influence consumer behaviour, with Tuesdays becoming the preferred day to refuel. This behaviour aligns with economic theory on anticipatory purchasing, where consumers act in advance of expected cost increases.

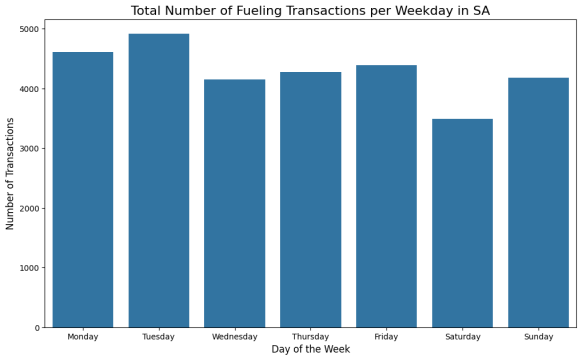

Figure 14: Refuelling counts by weekday in South Africa. Tuesdays show the highest activity, consistent with weekly price adjustments.

4.3.4 Reduce to First Tuesday/Wednesday

Subset reduced to 2,477 rows within first 7 days of each month.

```
--- Verifying the days in the new, filtered dataset ---
date_fueled  weekday
1            Tuesday      301
             Wednesday    112
2            Tuesday      317
             Wednesday    160
3            Tuesday      248
             Wednesday    118
4            Tuesday      148
             Wednesday    105
5            Tuesday      250
             Wednesday    154
6            Tuesday      168
             Wednesday    151
7            Tuesday      129
             Wednesday    116
```

Figure 15: distribution of Wednesday and Tuesday

4.3.5 Price Change Indicator

Indicators showed: 1,774 increases (Up), 700 decreases (Down), 3 Stable prices. Price increases dominate.

year_month	avg_price	price_change
2021-08	17.041551	Up
2021-09	17.199094	Up
2021-10	17.372709	Up
2021-11	18.407713	Up
2021-12	19.330390	Up
2022-01	18.653225	Down
2022-02	19.136512	Up
2022-03	20.414522	Up
2022-04	21.040175	Up
2025-09	15.004945	Down

Figure 16: indicators for a sample of the last 10 months in the data

4.3.6 Do people refuel more often on Wednesday When Prices Drop

Yes, on average slightly more people refuel on the first Wednesday of the month when prices go down (7.16) compared to when prices go up (6.51). The difference is not very large, but it indicates that drivers are marginally more likely to refuel when prices decrease, consistent with the intuition that consumers respond positively to lower fuel costs.

4.3.7 Tuesday Refuels Before Price Increases

Refuels were four times higher before price hikes (1,253 vs 306), showing strong anticipatory behaviour.

5 Discussion and Interpretation

The analyses reveal several clear behavioural patterns. Fuel efficiency results highlight technological and regional differences: plug-in hybrids dominate in developed markets, while budget-oriented compact vehicles drive efficiency in South Africa. Seasonal analysis confirms expected engineering effects, with higher consumption in winter. Correlation and feature importance tests consistently show that distance travelled and fuel volume are the strongest determinants of fuel efficiency, though this also illustrates risks of data leakage when using derived variables.

In the South African subset, strong behavioural patterns emerge around refuelling. Prices are systematically under-reported compared to official pump prices, limiting external validity, but internal consistency allows for meaningful trend analysis. Tuesdays consistently show the highest refuelling activity, aligning with weekly price cycles. Further, drivers refuel more often when anticipating price increases, with a fourfold difference in activity before hikes compared to decreases. On Wednesdays, refuelling rises when prices decrease, reflecting consumers' sensitivity to lower costs, consistent with evidence that fuel demand, though generally inelastic, does respond to price reductions in the short term (1).

6 Conclusion

This study examined global and South African vehicle fuel efficiency and refuelling behaviour using a large crowdsourced dataset. Key contributions include: (i) demonstrating regional contrasts

in efficiency linked to technology adoption, (ii) confirming seasonal and age-related effects on consumption, and (iii) uncovering clear anticipatory and habitual responses to price changes in South Africa.

Limitations include incomplete or systematically under-reported price data, high levels of missing mileage values, and the tautological nature of certain feature-importance models. Future work should combine this dataset with official pump price data and independent vehicle specifications to reduce bias, while behavioural analyses could be extended to test robustness across other regions.

References

[1] Goodwin, P., Dargay, J., & Hanly, M. (2004). Elasticities of road traffic and fuel consumption with respect to price and income: A review. *Transport Reviews*, 24(3), 275–292.