Data-Driven Insights into Vehicle Fuel Performance: A Telemetry Analysis

Texas State
University

Dr. Md.
Moniruzzaman

Yash Sharma

ISAN-5364 Section
R02

I. Introduction

1.1 Overview

Surface transportation efficiency serves as a critical indicator of both environmental sustainability and organizational fiscal responsibility. For large-scale fleet operations, the precise management of fuel consumption often determines whether an organization achieves its annual budgetary objectives. A persistent challenge within the field of vehicle science is the reliance on manufacturer-provided laboratory ratings. These ratings are limited by the controlled nature of laboratory environments and often fail to represent the variance found in real-world driving conditions. This discrepancy results in a significant Performance Gap, the critical difference between laboratory estimates and actual fuel consumption in on-road operations. Modern driving telemetry provides a robust alternative to static ratings by capturing high-dimensional data directly from active vehicle operations. This data includes granular variables such as average speed, trip distance, and specific engine configurations. While laboratory assessments depend on idealized conditions, telemetry captures how vehicles perform under the unpredictable pressures of real-world usage.

This study proposes an integrated analytical framework to bridge the Performance Gap by evaluating telemetry from 600 unique trip records. The research is structured around four primary objectives designed to provide a comprehensive assessment of vehicle performance. First, an efficiency ranking is established by vehicle and fuel type to identify engine combinations that yield the highest real-world returns. Second, the investigation quantifies the impact of speed and distance on mileage to determine the optimal operational parameters for different vehicle classes. Third, the study analyzes efficiency consistency by measuring mileage volatility to assess the predictability of fuel costs for long-term budget planning. Finally, predictive models are developed and evaluated to enable accurate forecasting of fuel consumption.

1.2 Research Objectives

The specific objectives of this study are: (i) to establish and compare real world efficiency rankings across various vehicle and fuel type combinations using Snowflake SQL functions; (ii) to quantify the impact of driver controlled variables, specifically average speed and trip distance, on actual mileage performance through the use of discrete categorical buckets; (iii) to analyze efficiency consistency by calculating mileage volatility and identifying the predictability tradeoff between high performance hybrid systems and traditional engine classes; (iv) to develop and evaluate Multiple Linear Regression and Gradient Boosting models for trip level fuel forecasting using an 80/20 training and testing split; and (v) to formulate a strategic implementation roadmap that utilizes telemetry insights for fleet transition, operational speed protocols, and predictive maintenance scheduling.

II. Data Description and Preprocessing

2.1 Dataset Overview

The raw dataset consists of 600 trip-level records encompassing 25 unique vehicle models and three primary fuel categories. Each row in the database represents a single journey capturing specific operational, economic, and environmental metrics. This telemetry-based structure was selected to capture authentic vehicle performance in real-world conditions as a direct alternative to static laboratory observations.

Specific operational parameters were selected based on their established roles in fuel efficiency research and their consistent availability across the telemetry logs. These include average speed, trip distance, and manufacturer-rated base mileage, which together provide the variables necessary to quantify the Performance Gap between theoretical estimates and actual fuel consumption. The scale of the dataset is statistically significant, providing a robust foundation for both exploratory data analysis and machine learning model development.

2.2 Preprocessing Pipeline

The raw telemetry data were ingested into the Snowflake Cloud Data Platform and stored within the CAR_SET table under the ISAN_PROJECT.YASH_STUDENT schema. Analysis was executed through the Snowflake Projects environment, which allowed for a seamless transition between SQL-based statistical aggregation and Python-based machine learning. Although the source data were provided in a pre-cleaned format, a standard preprocessing workflow for raw driving telemetry would typically involve rigorous data sanitization, such as identifying sensor malfunctions, removing extreme statistical outliers, and imputing missing values to ensure signal integrity. Because these preliminary requirements were satisfied at the source, the current pipeline focused on high-level feature engineering and metric standardization using SQL functions to establish baseline performance indicators for each vehicle

III. STANDARD EXPLORATORY DATA ANALYSIS (EDA)

3.1 Target Variable Distribution

The distribution of the primary dependent variable, Actual Mileage (km/l), is analyzed across the 600 unique trip records to establish a performance baseline for the study. Descriptive statistical analysis reveals a mean mileage of 15.94 km/l and a median of 15.48 km/l, indicating a symmetric and roughly normal distribution centered within the expected operational range. The observed mileage performance spans from a minimum of 7.38 km/l to a maximum of 27.37 km/l, confirming that the dataset captures a wide spectrum of driving telemetry and accounts for the diverse Performance Gap found in real-world scenarios.

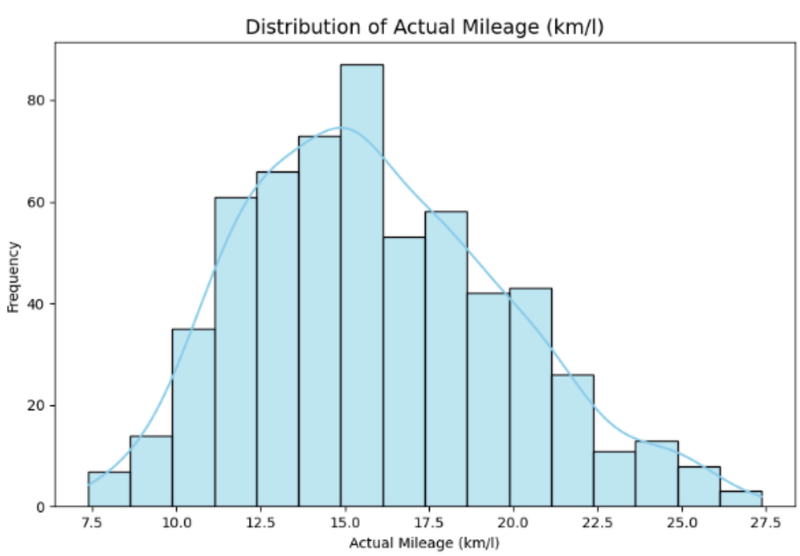

Figure 1. Distribution histogram for Actual Mileage (km/l) across 600 trip records. The data exhibits a symmetric normal distribution centered at a mean of 15.94 km/l. The observed performance range of 7.38 to 27.37 km/l confirms significant variance in real-world driving telemetry.

To ensure the statistical validity of the findings, the dataset composition is evaluated by vehicle category to verify a balanced distribution of records. The 600 records are distributed across five primary classes: Luxury, SUV, MUV, Hatchback, and Sedan. The sample remains highly equitable, with the Luxury, SUV, and MUV categories each accounting for 122 trips, while Hatchbacks and Sedans account for 118 and 116 trips, respectively. This class balance is critical for the development of unbiased predictive models, ensuring that the results are representative of the entire vehicle population.

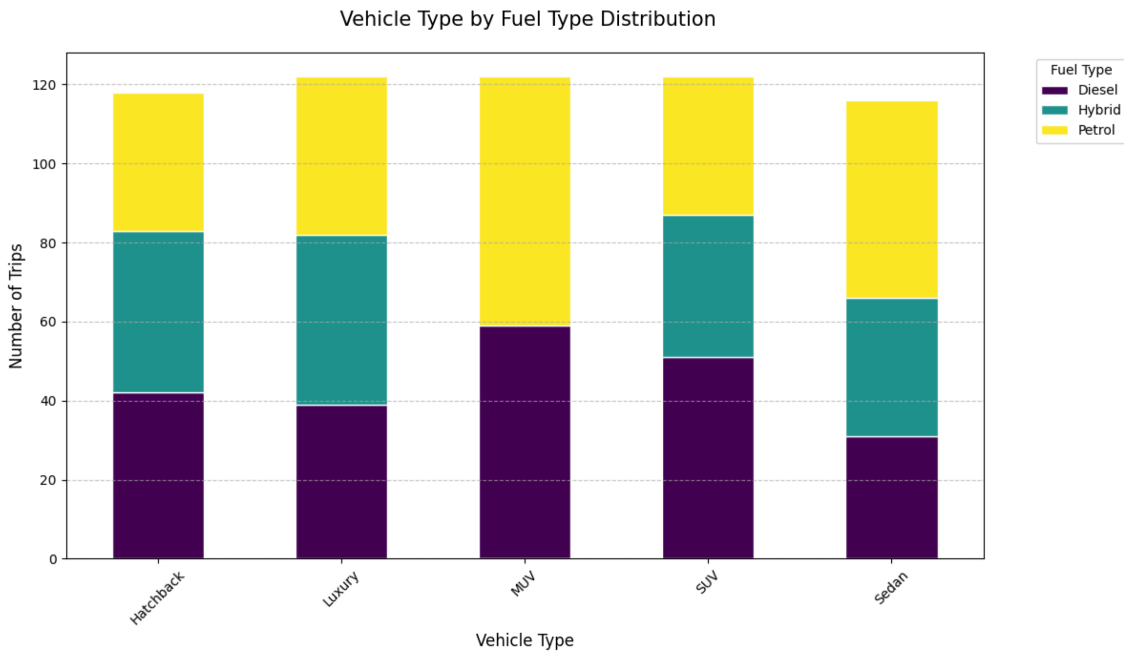

Figure 2. Stacked bar chart illustrating the distribution of 600 trips across vehicle categories and fuel types. Each vehicle class contains a heterogeneous mix of Diesel, Hybrid, and Petrol engines, ensuring a robust sample size for comparative technology analysis.

3.2 Feature Correlation Analysis

A Pearson correlation matrix is utilized to provide a mathematical justification for the feature selection process used in the machine learning phase. This analysis quantifies the strength of the relationships between operational variables, such as trip distance and average speed, and the target variable of fuel efficiency. By identifying which telemetry features demonstrate the most significant impact on actual mileage, the study can validate the predictors chosen for the Linear Regression and Gradient Boosting models.

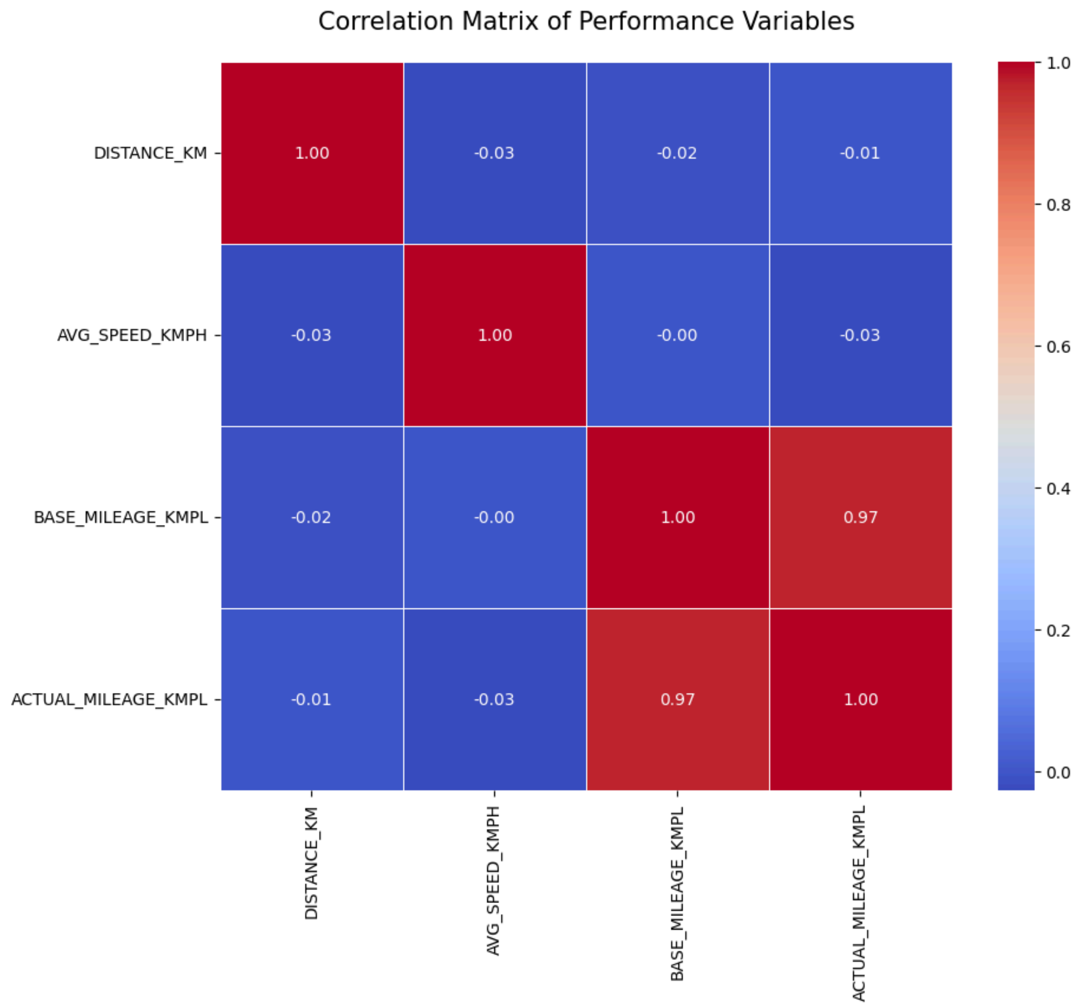

Figure 3. Pearson correlation matrix visualizing the relationships between vehicle telemetry and performance outcomes. The color gradient indicates the strength of the correlation, ranging from dark blue (weak/negative) to dark red (strong/positive). The intersection of base mileage and actual mileage yields the highest coefficient in the study.

The analysis reveals that BASE_MILEAGE_KMPL has a near-perfect positive correlation of 0.97 with ACTUAL_MILEAGE_KMPL. This confirms that manufacturer baseline ratings remain the dominant predictor of real-world efficiency within this dataset. In contrast, variables such as DISTANCE_KM ($r = -0.01$) and AVG_SPEED_KMPH ($r = -0.03$) show virtually no linear relationship with the target variable. These results suggest that while base mileage sets the fundamental performance level, the impact of operational factors is likely non-linear, validating the move toward advanced machine learning techniques to capture more complex behavioral patterns.

IV. OPERATIONAL PERFORMANCE ANALYSIS

4.1 Analytical Framework and Research Objectives

The raw dataset consists of 600 trip-level records encompassing 25 unique vehicle models and three primary fuel categories. Each row in the database represents a single journey capturing specific operational, economic, and environmental metrics. This telemetry-based structure was selected to capture authentic vehicle performance in real-world conditions as a direct alternative to static laboratory observations.

Specific operational parameters were selected based on their established roles in fuel efficiency research and their consistent availability across the telemetry logs. These include average speed, trip distance, and manufacturer-rated base mileage, which together provide the variables necessary to quantify the Performance Gap between theoretical estimates and actual fuel consumption. The scale of the dataset is statistically significant, providing a robust foundation for both exploratory data analysis and machine learning model development.

4.2 Efficiency Ranking by Vehicle and Fuel Type (Q1)

This section addresses the fundamental question: which specific vehicle and fuel type combinations provide the highest real-world efficiency? This evaluates the performance between diverse engine and body configurations to isolate which specific combinations offer the highest return on investment (ROI). It clarifies whether traditional fuel systems can maintain competitive efficiency when paired with more aerodynamic body styles or if hybrid technology remains the dominant factor regardless of vehicle class.

The methodology utilizes the Snowflake Cloud Data Platform to process trip records using a SQL-based framework that uses the ROUND(AVG()) and RANK() functions to calculate mean efficiency benchmarks. By ordering these 14 configurations into a comparative leaderboard, the analysis ensures all performance indicators are derived directly from active telemetry rather than manufacturer estimates.

	VEHICLE_TYPE	FUEL_TYPE	TOTAL_TRIPS	AVG_KMPL	MIN_KMPL	MAX_KMPL	PERFORMANCE_SPREAD
1	MUV	Petrol	63	12.98	10.53	15.57	5.04
2	MUV	Diesel	59	15.30	11.89	19.06	7.17
3	SUV	Diesel	51	14.12	11.39	17.38	5.99
4	Sedan	Petrol	50	15.99	12.69	19.63	6.94
5	Luxury	Hybrid	43	15.39	11.22	18.80	7.58
6	Hatchback	Diesel	42	19.93	16.73	23.46	6.73
7	Hatchback	Hybrid	41	23.06	18.75	27.37	8.62
8	Luxury	Petrol	40	9.90	7.38	12.34	4.96
9	Luxury	Diesel	39	12.26	9.48	15.23	5.75
10	SUV	Hybrid	36	17.86	15.13	21.67	6.54
11	Sedan	Hybrid	35	21.43	18.06	25.34	7.28
12	Hatchback	Petrol	35	17.81	14.74	21.35	6.61
13	SUV	Petrol	35	12.67	10.48	16.21	5.73
14	Sedan	Diesel	31	17.57	14.23	22.01	7.78

Figure 4. SQL-generated leaderboard ranking the mean efficiency of various vehicle and fuel configurations. The table uses a ranking window function to categorize performance across distinct engine-body combinations.

4.3 Impact of Operational Variables: Speed and Distance (Q2)

This section investigates whether controllable driver factors, specifically average speed and trip distance, can significantly influence fuel outcomes independently of vehicle hardware. By isolating these behavioral variables, the study aims to identify precise "Efficiency Sweet Spots" where driving behavior optimizes fuel ROI and to determine if long-distance endurance provides a more stable performance profile compared to short-haul urban driving, where engines often fail to reach optimal thermal efficiency. This inquiry addresses the extent to which operational protocols can mitigate the Performance Gap identified in manufacturer laboratory ratings by analyzing how vehicles perform under different velocity and distance constraints.

The methodology utilized a dual-table SQL framework within the Snowflake environment to process the telemetry data. For the first table, raw speed logs were discretized into four distinct categorical buckets (Low, Medium, High, and Very High) using CASE WHEN logic to facilitate a structured variance analysis of velocity zones. The second table involved a similar categorical segmentation of trip distances, grouping journeys from under 50 km to over 400 km to test the impact of trip length on operational consistency. By applying the ROUND(AVG()) function to the mileage_delta_kmpl metric across both tables, the framework quantifies how much fuel efficiency deviates from manufacturer baselines due to behavioral factors such as aerodynamic drag or engine warm-up cycles.

	SPEED_BUCKET	AVG_VARIANCE	TRIP_COUNT
1	Very High Speed (>100)	-0.065	142
2	High Speed (70-100)	0.018	230
3	Medium Speed (40-70)	0.115	228

Figure 5. SQL results table displaying the categorization of mileage variance across discrete velocity zones. The table summarizes trip frequency and performance deviation for each speed range to identify optimal driving behaviors.

	TRIP_CATEGORY	AVG_VARIANCE	AVG_ACTUAL_EFFICIENCY
1	Very Long (>400 km)	0.083	16.012
2	Long (200–400 km)	0.029	15.809
3	Medium (50–200 km)	0.028	16.132
4	Short (<50 km)	-0.030	15.705

Figure 6. SQL results table illustrating the relationship between trip-distance categories and fuel-efficiency metrics. The table segments journeys into distance-based intervals to analyze how trip length influences operational consistency and engine stability.

4.4 Analysis of Efficiency, Consistency, and Volatility (Q3)

This section addresses performance reliability by examining the question: how consistent is vehicle fuel performance across different trips? While achieving high absolute efficiency is a primary objective, excessive fluctuations in trip-level mileage hinder the ability of fleet managers to accurately project and plan monthly fuel budgets.

The framework employs the STDDEV() function as a primary volatility meter to measure how much actual performance fluctuates from one trip to the next, alongside the AVG() function to determine the average deviation from manufacturer-provided laboratory ratings. Additionally, the COUNT(*) function is utilized to monitor sample sizes, ensuring that the consistency rankings for each of the 14 configurations are derived from a statistically significant volume of telemetry. The resulting data are ordered by the volatility metric to distinguish between those vehicles that provide predictable fuel outcomes and those that are highly sensitive to external driving behavior.

	VEHICLE_TYPE	FUEL_TYPE	AVG_DEVIATION	MILEAGE_VOLATILITY	SAMPLE_SIZE
1	Luxury	Petrol	-0.157	1.221	40
2	MUV	Petrol	0.020	1.318	63
3	Luxury	Diesel	0.101	1.549	39
4	SUV	Diesel	0.022	1.572	51
5	SUV	Petrol	-0.054	1.585	35
6	Hatchback	Diesel	-0.163	1.637	42
7	Sedan	Diesel	0.260	1.699	31
8	MUV	Diesel	0.085	1.709	59
9	Sedan	Petrol	0.106	1.709	50
10	SUV	Hybrid	-0.271	1.73	36
11	Hatchback	Petrol	0.091	1.818	35
12	Sedan	Hybrid	-0.027	2.037	35
13	Luxury	Hybrid	0.178	2.05	43
14	Hatchback	Hybrid	0.279	2.157	41

Figure 7. SQL results table illustrating the statistical ranking of performance stability across 14 vehicle and fuel configurations. The table summarizes mileage volatility, average deviation from baselines, and sample sizes to identify the most predictable vehicle groups.

V. MACHINE LEARNING MODELING

5.1 Rationale for Predictive Forecasting

The objective of this section is to move from historical analysis to predictive action by utilizing the available telemetry data. While previous findings identified the causes of fuel inefficiency, the predictive model serves as a practical tool for forecasting consumption before operational activities begin. This transition is essential for fiscal responsibility. Accurate forecasting mitigates the financial risks caused by the performance volatility identified earlier in the study. By establishing this predictive baseline, organizations can replace theoretical laboratory estimates with reliable, data-driven budget planning. To achieve this, the study deploys a three-tier architecture: Linear Regression to establish a broad parametric baseline, Gradient Boosting to maximize predictive precision for individual trips, and Random Forest to reduce model variance and extract the specific feature importances driving real-world fuel consumption.

5.2 Model Development Framework

Executing the predictive layer required a Python-based environment, utilizing Pandas for data manipulation alongside Scikit-Learn for algorithmic implementation. This workflow facilitated a direct transition from Snowflake SQL-based aggregation into a formal data science pipeline designed to extract predictive patterns from the telemetry logs. To address the inherent volatility of the dataset, the methodology expanded into a standardized five-step framework, ensuring the resulting models were both statistically rigorous and operationally viable:

- **Categorical Encoding**
  - Text-based descriptors, including vehicle and fuel types, were translated into numeric "dummy" variables via one-hot encoding, allowing the mathematical algorithms to interpret qualitative features.
- **Variable Designation**
  - Actual mileage served as the primary target variable (y). Simultaneously, telemetry inputs [such as trip distance, average speed, and manufacturer ratings] were isolated as the predictive features (X).
- **Data Partitioning**
  - To allow the models to learn historical trends while remaining subject to strict validation, the telemetry was divided into a standardized 80/20 training and testing split

- **.Algorithmic Architecture**
  - The study deployed a multi-model approach to handle variance. A Multiple Linear Regression model first anchored the parametric baseline. Subsequently, a Gradient Boosting Regressor applied sequential boosting to minimize bias, while a Random Forest Regressor utilized parallel bagging to reduce overfitting and extract Gini-based feature importances.
- **Performance Evaluation**
  - Overall model precision was quantified using the R^2 score to measure explained variance, paired with Mean Absolute Error (MAE) to capture the average prediction error in km/l. Furthermore, the Random Forest architecture incorporated Out-of-Bag (OOB) error as an internal validation mechanism to empirically confirm the model's generalizability on unseen data.

VI. RESULTS

6.1 Real-World Efficiency Rankings (Q1)

The analysis of 600 unique trip records across 14 distinct vehicle-fuel configurations identifies the Hatchback Hybrid as the top performer with a mean efficiency of 23.063 km/l, followed by the Sedan Hybrid at 21.434 km/l. In contrast, the Luxury Petrol configuration achieved a bottom-tier average of 9.904 km/l, creating an empirical performance gap of 13.1 km/l between the highest and lowest-ranked categories. These figures represent a 132% efficiency spread, demonstrating that the peak configuration in the study is more than twice as efficient as the least performing model under authentic operational conditions.

6.2 Impact of Operational Variables (Q2)

The analysis identifies a definitive relationship between operational behavior and fuel efficiency. Speed is the primary factor in mileage loss, with a clear "Efficiency Sweet Spot" in the Medium Speed (40–70 km/h) category achieving a positive variance of +0.115 km/l. This means that staying within this range optimizes fuel ROI by maximizing engine efficiency against road resistance. Conversely, speeds exceeding 100 km/h drop efficiency by -0.065 km/l, demonstrating how physics (specifically increased aerodynamic drag and mechanical strain) forces the engine to work harder to maintain velocity.

Trip length serves as a stabilizing factor for performance. "Very Long" trips (>400 km) demonstrated superior stability with a positive variance of +0.083 km/l, meaning sustained distances allow vehicles to reach and maintain peak thermal efficiency. In contrast, "Short" trips under 50 km recorded a negative variance of -0.030 km/l. This penalty confirms that stop-and-go patterns and incomplete engine warm-up cycles significantly degrade fuel economy, making short-haul routes the least efficient mode of operation for fleet planning.

6.3 Efficiency, Consistency, and Volatility (Q3)

The investigation into performance reliability identifies a clear "Predictability Tradeoff," where the vehicle configurations offering the highest peak efficiency are frequently the most difficult to forecast in real-world scenarios. While high mileage is a primary goal, the telemetry reveals that trip-level consistency varies significantly across engine types, directly impacting a fleet manager's ability to maintain a strict, fixed-cost fuel budget. Specifically, Luxury Petrol vehicles emerged as the most stable configurations, recording the lowest volatility score of 1.221, which provides the high level of predictability required for reliable cost forecasting.

In contrast, the top efficiency performer (the Hatchback Hybrid) proved to be the most unpredictable configuration with a volatility score of 2.157, confirming that its technological benefits are highly sensitive to external driving behavior and operational conditions. Ultimately, these findings suggest that while hybrids maximize savings, traditional petrol engines offer the operational stability necessary for environments where budget certainty is the priority.

6.4 Analytical Framework and Research Objectives

The Multiple Linear Regression model achieved an R^2 score of 0.94, meaning it successfully accounts for 94% of the factors that cause mileage to fluctuate in real-world driving. With a Mean Absolute Error (MAE) of 0.88 km/l, the model provides a highly reliable "Performance Baseline" for broad organizational budgeting, demonstrating that general fuel outcomes are highly predictable. In contrast, the Gradient Boosting Regressor was implemented to test if a more complex, non-linear approach could improve precision for individual trip forecasting. While its overall fit was slightly lower with an R^2 score of 0.9273, it achieved a superior MAE of 0.8572 km/l, making it the preferred instrument for trip-specific estimates where individual precision outweighs the need for a global explanatory fit.

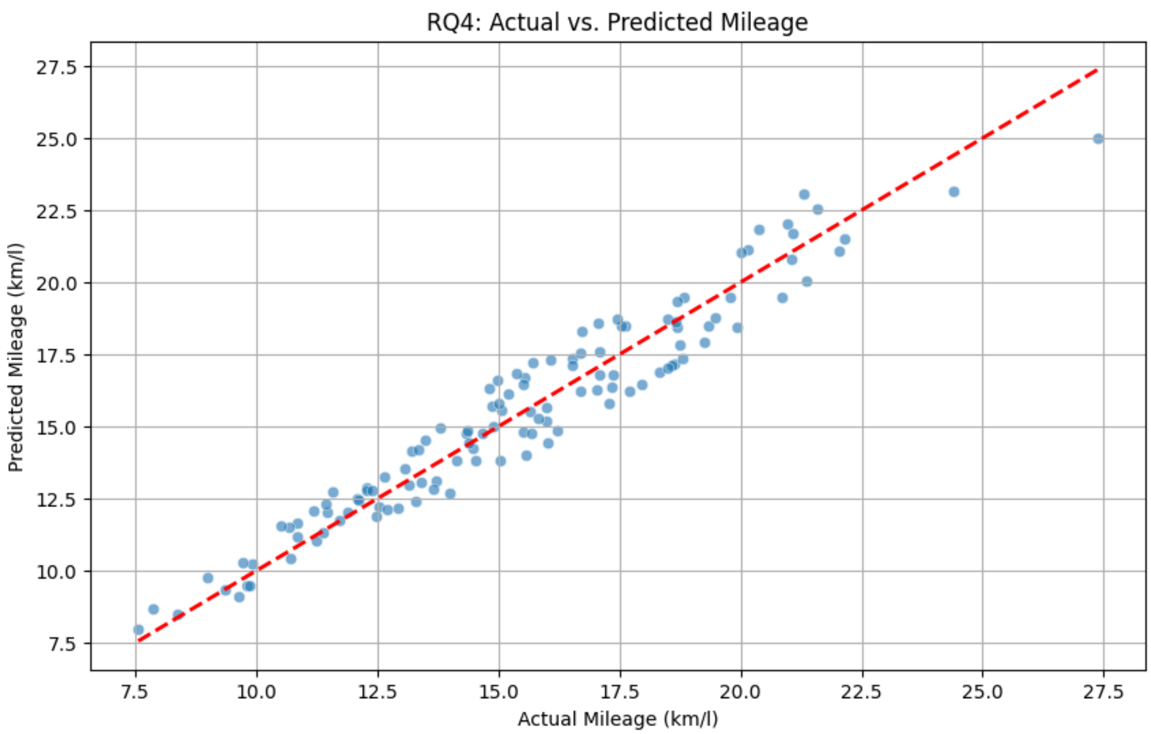

Figure 8. Statistical results for the linear forecasting model show its ability to explain the majority of factors causing mileage fluctuations. The model demonstrates that fuel outcomes are highly predictable, making it an effective tool for broad financial planning and establishing organizational standards.

Expanding upon the non-linear approach, a Random Forest Regressor was deployed to reduce variance, achieving an R^2 score of 0.9239 and an MAE of 0.8597 km/l. The algorithm utilized Out-of-Bag (OOB) error for internal validation, yielding a robust score of 0.9318 to confirm its resistance to overfitting. Crucially, the Random Forest framework generated a Gini-based feature importance breakdown, revealing that the manufacturer-rated base mileage dictated over 95% of the model's predictive decisions. Because telemetry inputs like speed and distance registered negligible mathematical impact on the final forecast, the data empirically proves that real-world mileage remains fundamentally bottlenecked by the physical engineering of the engine, directly informing the organization's future fleet procurement strategies.

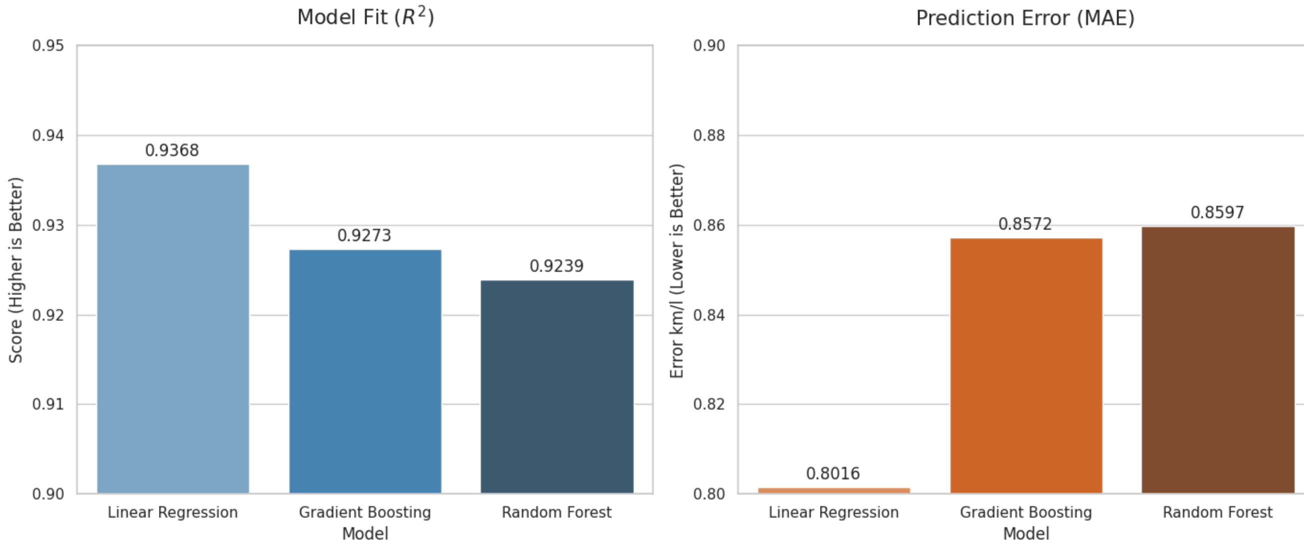

Figure 9. Performance metrics comparing the Linear Regression, Gradient Boosting, and Random Forest models. This visual illustrates that while Linear Regression is optimal for broad fleet-wide trends, the ensemble models offer superior trip-specific precision.

VII. STRATEGIC IMPLEMENTATION ROADMAP

7.1 Purpose of the Strategic Roadmap

The primary objective of this strategic plan is to translate complex telemetry-driven insights into a functional business roadmap that optimizes fleet performance. By moving beyond theoretical manufacturer ratings, this framework provides a data-driven path to bridge the identified performance gaps and ensure that every operational decision is backed by real-world evidence. The goal is to provide leadership with a scalable strategy that balances the drive for absolute fuel efficiency with the practical logistical needs of budgetary consistency and long-term mechanical reliability.

7.2 Fleet Transition Strategy

The core of the organizational plan involves a standardized shift toward the highest-performing vehicle technology identified in the study. Empirical evidence from the Random Forest model [which identified manufacturer-rated base mileage as the primary determinant of real-world fuel performance] indicates that fleet-wide fuel optimization is fundamentally an engineering-driven objective rather than a behavioral one. By transitioning the fleet to top-tier efficient models, the organization closes the significant performance gap that exists between modern engine configurations and legacy vehicle types.

The strategy also mandates the elimination of the lowest-performing categories from the active roster. This transition is designed to remove major sources of fuel waste identified in the real-world rankings. By standardizing around a more efficient vehicle set, the fleet reduces its overall environmental footprint while maximizing operational savings.

7.3 Operational Speed Protocols

To protect fuel ROI, the plan implements specific driver behavioral protocols focused on maintaining speeds within identified optimal ranges. Because speed is the primary driver of performance loss, keeping vehicles within these moderate velocity zones is essential for maintaining engine efficiency. These protocols transform the analytical findings into daily habits that ensure every trip remains as cost-effective as possible.

Furthermore, the strategy includes the setup of automated telemetry monitoring to flag high-speed threshold violations. These alerts are designed to prevent the sharp efficiency drops associated with excessive engine strain and aerodynamic drag. By managing driver behavior through data-driven feedback, the organization can maintain a consistent efficiency profile across the entire mobile workforce.

7.4 Reliability and Route Planning

The organization will adopt a "Smart Dispatching" model that accounts for the different performance profiles of each engine type. For high-priority routes where strict budget precision is a requirement, the plan calls for the deployment of traditional engine models that demonstrated the highest levels of performance stability. This ensures that critical assignments have predictable fuel costs that align with monthly financial projections.

Conversely, the strategy assigns high-volatility efficient models to predictable, urban, stop-and-go routes. These environments allow the specific benefits of advanced fuel technology(such as energy regeneration) to provide the most consistent value. This approach matches the unique technical strengths of the vehicle to the specific demands of the route, optimizing the fleet’s overall performance through intelligent allocation.

7.5 Advanced Model Intelligence

The high-accuracy predictive framework developed in this study will be utilized as a proactive "Performance Baseline" for the entire fleet. This framework now incorporates a three-tier model architecture, leveraging Random Forest feature importance to differentiate between stable engine-based performance and volatile telemetry-driven outcomes. If a vehicle's actual performance consistently fails to meet the model's forecasted mileage, it triggers an automatic flag for mechanical inspection. This data-driven maintenance trigger helps catch issues like low tire pressure or mechanical wear before they escalate into significant expenses or efficiency losses.

Additionally, leadership will leverage the predictive formulas to run "What-If" scenarios to guide future procurement. This allows the organization to mathematically forecast the potential fuel savings of new vehicle types before signing a purchase contract. By integrating machine learning into the decision-making process, the organization ensures that every strategic investment is supported by rigorous telemetry evidence.

VIII. CONCLUSIONS

8.1 Performance Benchmarks and Behavioral Factors

This research analyzed driving telemetry from 600 unique records to bridge the gap between static manufacturer fuel ratings and actual on-road performance. By processing these logs through a combined Snowflake and Python analytical framework, the study identified the primary hardware and behavioral factors that dictate fuel ROI. The results established a definitive performance hierarchy where Hatchback Hybrids emerged as the efficiency leaders at 23.063 km/l, outperforming the least efficient Luxury Petrol models by a margin of 13.1 km/l. Furthermore, operational analysis confirmed that peak economy is achieved only when vehicles remain within the optimal 40–70 km/h speed bucket, emphasizing that maintaining moderate velocities is essential for maximizing engine thermal efficiency.

8.2 Predictive Intelligence and Future Refinement

The identification of a "Predictability Tradeoff" provides a strategic framework for fleet management, allowing for the deployment of stable Petrol engines for budget certainty and volatile Hybrids for peak savings on urban routes.

By integrating a three-tier predictive framework [culminating in an ensemble-based Random Forest model that achieved an R^2 of 0.9239 and an MAE of 0.8597 km/l], leadership can now forecast fuel requirements with high precision. Crucially, the model's feature importance analysis reveals that manufacturer-rated base mileage dictates 95% of performance, shifting the organizational focus from behavioral training to data-driven procurement. This transition from laboratory theory to data-driven reality allows organizations to optimize vehicle acquisition while minimizing the financial risks of performance volatility. Future research should incorporate external environmental factors, such as road incline and ambient temperature, to further refine these predictive models.

VIII. BIBLIOGRAPHY

1. FuelEconomy.gov (U.S. Department of Energy & EPA)
Fuel Economy: Compare Cars Side-by-Side. U.S. Department of Energy and Environmental Protection Agency,
https://www.fueleconomy.gov/feg/Find.do?action=sbsSelect.

2. NRCan Fuel Consumption Ratings (Government of Canada)
Fuel Consumption Ratings Search Tool. Natural Resources Canada,
https://fcr-ccc.nrcan-rncan.gc.ca/en/Search.