



IIB40303 ASSESSMENT COVERSHEET

Attach this coversheet as the cover of your submission. All sections must be completed.

Section A: Submission Details

Programme	:	BACHELOR'S IN INFORMATION TECHNOLOGY (HONS) IN INTERNET OF THINGS (IoT)
Course Code & Name	:	IIB40303 – PROGRAMMING FOR DATA ANALYTIC
Assessment	:	Group Project Report
Dateline	:	26 DECEMBER 2025
Penalties	:	<ul style="list-style-type: none">• 10% will be deducted for late submission.• Plagiarised work is an Academic Offence in University Rules & Regulations and will be penalized accordingly.• AI content will affect your mark

Section B: Academic Integrity

Tick (✓) each box below if you agree:

<input checked="" type="checkbox"/>	I have read and understood the UniKL's policy on Plagiarism in University Rules & Regulations.
<input checked="" type="checkbox"/>	This submission is my own, unless indicated with proper referencing.
<input checked="" type="checkbox"/>	This submission has not been previously submitted or published.
<input checked="" type="checkbox"/>	This submission follows the requirements stated in the course.

Section C: Submission Detail

Office Receipt of Submission

No	Student Name(s)	Student ID(s)
1	RAJA MOHAMAD HAFIZH HAIKAL BIN RAJA ZAZALI	52224223348
2	NUR ALIA MAISARAH BINTI MANSOR	52224223354
3	MUHAMMAD RAZMEE BIN MOHD AZME	52224223306

Student Receipt of Submission

The submission receipt of the student task will be based on the **VLE submission date and time**.

Data Analysis Report: Cars Dataset 2025

1. Introduction

Data analysis plays a crucial role in understanding trends, patterns, and relationships within real-world datasets. This report presents an end-to-end data analysis of a Cars Dataset (2025) using Python and popular data science libraries such as Pandas, NumPy, Matplotlib, and Seaborn.

The objective of this analysis is to:

- Understand the structure and quality of the dataset
- Perform exploratory data analysis (EDA)
- Identify meaningful trends and relationships between variables
- Extract insights that can support decision-making

2. Dataset Overview

The dataset used in this analysis (cars_2025.csv) contains information related to various car models and their attributes. The dataset was loaded using the Pandas library. It contains 614,068 records with seven main attributes:

- Date of registration
- Vehicle type
- Car maker
- Car model
- Colour
- Fuel type
- State

2.1 Data Loading

```
df = pd.read_csv("cars_2025.csv")
```

2.2 Initial Inspection

- The first few rows were examined using *df.head()*
- Dataset dimensions were checked using *df.shape*
- Data types and missing values were inspected using *df.info()*

This initial inspection helps in understanding:

- Number of rows and columns
- Types of variables (numerical or categorical)
- Presence of missing or null values

3. Data Cleaning and Preparation

Before conducting the analysis, several data preparation steps were performed to ensure data quality and consistency.

First, the dataset structure and data types were examined using *.info()*, and missing values were checked using *.isnull().sum()*. The results showed that there were no missing values in any of the columns, indicating that the dataset is clean and complete. Nevertheless, a *dropna()* operation was applied as a precautionary step to ensure data consistency.

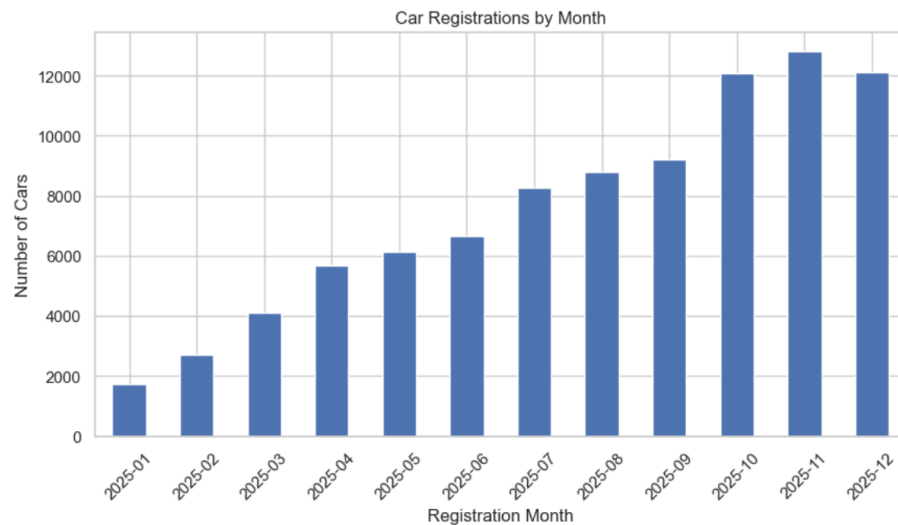
Next, the *date_reg* column was converted into a datetime format to enable time-based analysis. From this column, two new features were extracted:

- *year_reg*: used to analyze yearly trends
- *reg_month*: used to analyze monthly registration patterns

These steps prepared the dataset for accurate visualization and analysis.

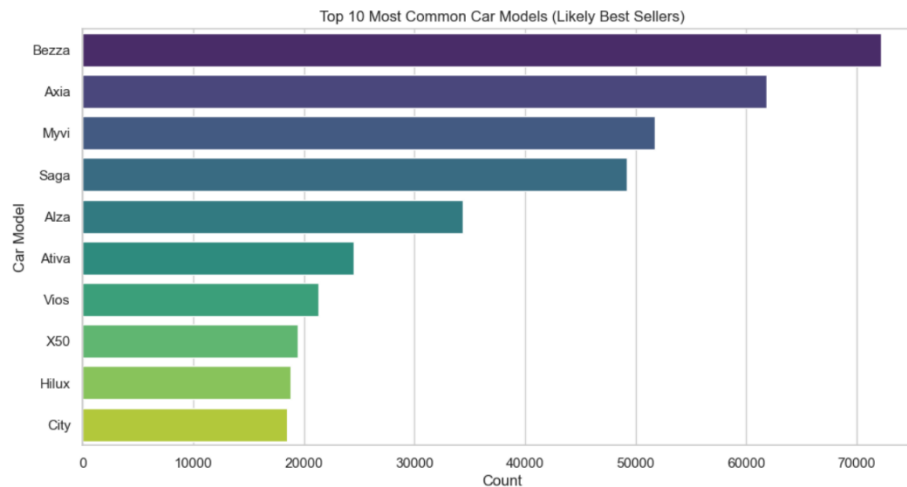
4. Exploratory Data Analysis and Visualization (EDA)

EDA is used to uncover hidden patterns, trends, and relationships in the dataset.



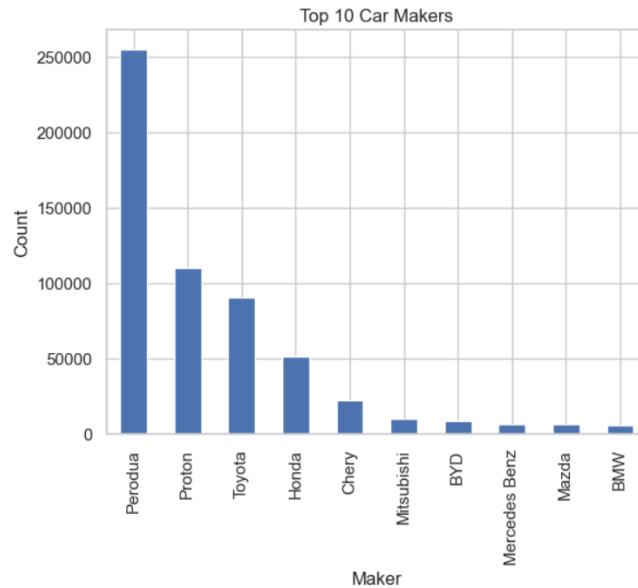
4.1 Car Registrations by Month

A bar chart was created to visualize the number of car registrations for each month. The analysis shows a steady increase in registrations throughout the year, with the highest numbers occurring toward the end of the year, particularly from October to December. This trend may be influenced by year-end promotions, festive seasons, and increased consumer purchasing before the year ends.



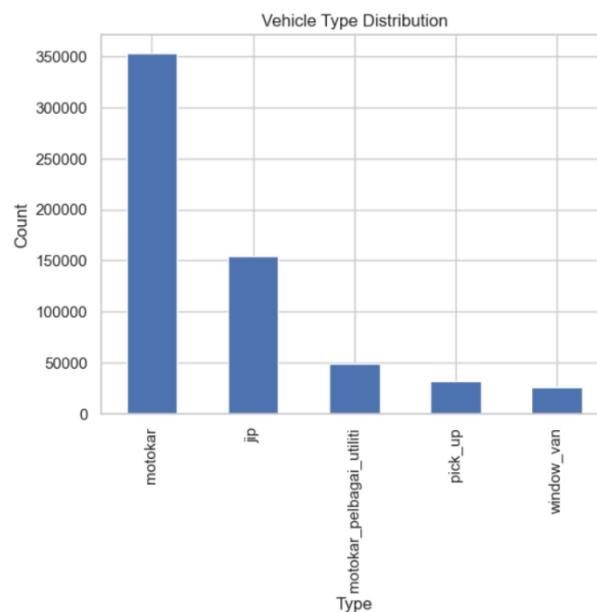
4.2 Top 10 Most Common Car Models

A horizontal bar chart was used to display the top 10 most registered car models. The results show that models such as Perodua Bezza, Axia, Myvi, and Proton Saga dominate the list. These models are generally affordable, fuel-efficient, and suitable for daily use, indicating that consumers prioritize practicality and cost-effectiveness when purchasing vehicles.



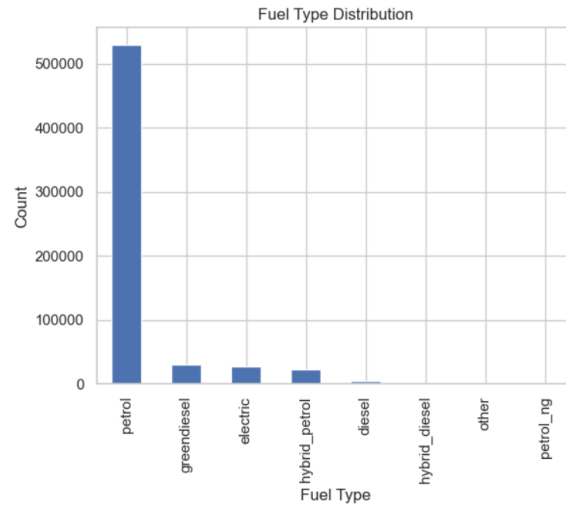
4.3 Top 10 Car Makers

This bar chart presents the top 10 car manufacturers based on total registrations. Perodua leads by a significant margin, followed by Proton and Toyota. The dominance of local brands suggests strong consumer trust in national manufacturers and highlights the importance of affordability in the Malaysian automotive market.



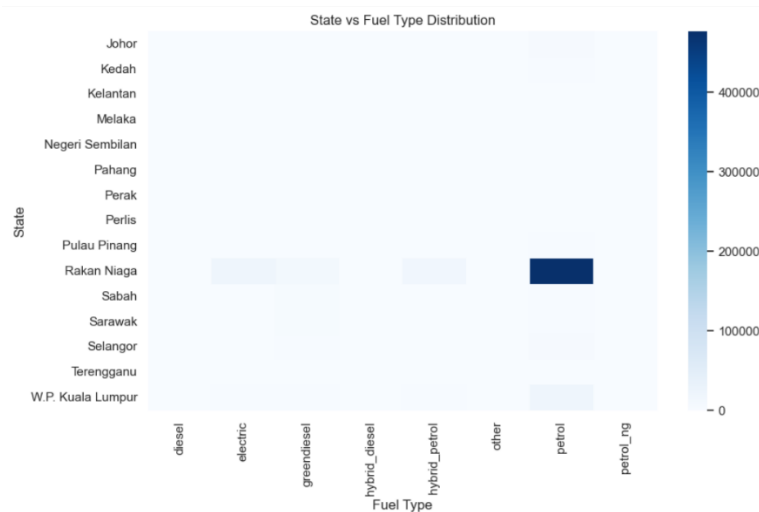
4.4 Vehicle Type Distribution

The vehicle type distribution chart shows that passenger cars (motokar) account for the majority of registrations. Other vehicle types such as jeeps, commercial utility vehicles, pick-up trucks, and window vans have significantly lower counts. This indicates that most registered vehicles are intended for personal transportation rather than commercial use.



4.5 Fuel Type Distribution

The fuel type distribution analysis reveals that petrol-powered vehicles dominate the market. Diesel, electric, and hybrid vehicles account for a much smaller proportion of registrations. This suggests that petrol vehicles remain the preferred choice due to lower cost, widespread refueling infrastructure, and consumer familiarity. Although electric and hybrid vehicles are present, their adoption is still relatively low.



4.6 State vs Fuel Type Distribution

A heatmap was used to analyze the relationship between state and fuel type. The visualization shows that petrol vehicles dominate across almost all states. States such as Selangor and W.P. Kuala Lumpur have higher registration counts compared to others.

An important category in this dataset is Rakan Niaga, which does not represent a physical state. Instead, it refers to business or corporate vehicle registrations, typically by car dealers or companies. This explains the high concentration of petrol vehicles under this category.

5. Actionable Insight

Based on the analysis conducted, several meaningful and actionable insights can be identified from the findings. The monthly car registration analysis shows a clear seasonal pattern, with registrations increasing toward the end of the year, especially from October to December. This trend suggests that consumers are more likely to purchase vehicles during year-end periods, possibly due to festive seasons, promotional offers, tax incentives, or clearance sales by car dealers. Automotive companies can take advantage of this pattern by planning marketing campaigns, promotions, and product launches during these peak months, while dealers can better manage inventory to meet higher demand.

In terms of fuel preference, the results indicate that petrol-powered vehicles continue to dominate the market, while diesel, hybrid, and electric vehicles make up only a small portion of total registrations. This suggests that although awareness of environmental sustainability is increasing, many consumers still prioritize factors such as affordability, convenience, and the availability of refueling infrastructure. These findings highlight the need for policymakers and industry players to address challenges such as limited charging infrastructure and higher initial costs if wider adoption of alternative fuel vehicles is to be encouraged.

The analysis of the most registered car models further shows a strong preference for affordable, practical, and fuel-efficient vehicles. Models such as Perodua Bezza, Axia, Myvi, and Proton Saga appear most frequently, indicating that consumers value reliability and cost-effectiveness for everyday use. This insight can help manufacturers focus on producing vehicles that meet these expectations while keeping maintenance and ownership costs low.

In addition, the dominance of local car manufacturers, particularly Perodua and Proton, reflects strong consumer confidence in national brands. This trend highlights the importance of competitive pricing, accessibility, and reliable after-sales support in influencing purchasing decisions. While international brands are still present in the market, they tend to cater to a smaller segment due to higher prices. These findings can assist policymakers in assessing current automotive policies and help manufacturers strengthen the competitiveness of local brands.

Lastly, the vehicle type distribution shows that passenger cars make up most registrations, indicating that personal transportation remains the main driver of vehicle demand. Commercial vehicles such as pick-up trucks and vans form a smaller segment, suggesting that business-related usage is more niche in comparison. Regional analysis also shows higher registration volumes in urban areas such as Selangor and W.P. Kuala Lumpur, which is consistent with higher population density and economic activity. The presence of the Rakan Niaga category further highlights the role of corporate and dealer registrations, particularly for petrol vehicles. Overall, these insights demonstrate how data-driven analysis can support better planning and decision-making for automotive manufacturers, dealers, policymakers, and infrastructure planners.

6. Conclusion

This project showed how a structured data analysis process can be used to explore and understand a real-world dataset effectively. Using the Cars 2025 dataset, the analysis was carried out step by step, starting from data loading and inspection, followed by data cleaning, exploratory data analysis (EDA), and data visualization.

By examining the dataset at an early stage, it was easier to understand the structure of the data, identify different types of variables, and detect potential issues such as missing values or outliers. This step was important because it helped ensure that the analysis was based on reliable and meaningful data.

Exploratory data analysis and visualization were especially useful in this project. Visualizing the data made it easier to observe patterns, compare different categories, and understand the relationships between car prices and performance-related features. Instead of relying only on numerical values, charts and graphs helped present the findings in a clearer and more intuitive way.

Overall, this analysis highlights the importance of:

- Checking and understanding the dataset before starting the analysis
- Using data visualization to better interpret patterns and trends
- Making conclusions based on data rather than assumptions

From a student's perspective, this project provided practical experience in applying data analysis concepts learned in class to a real dataset. It also helped build a stronger understanding of how data-driven analysis is carried out in real-world scenarios and prepared a good foundation for more advanced topics in data analytics.

7. Future Work

Although this analysis provided useful insights, there are several ways the project could be improved or extended in the future. One possible improvement is featuring engineering, where existing features could be modified or combined to better represent important characteristics of the data. This may help improve the quality of future analysis or predictive models.

Another extension of this project would be to build a machine learning model to predict car prices. For example, regression models could be used to estimate prices based on car specifications. This would allow the project to move beyond data exploration and into predictive analysis. Additionally, future work could include a deeper study of consumer preferences and market trends by using additional data such as brand popularity, user reviews, or regional demand. This could provide more realistic insights into the factors that influence car pricing.

Overall, these improvements would help make the project more comprehensive and allow the analysis to better reflect real-world data analytics applications.

Submission Links

GitHub Link:

<https://github.com/RazmeeAzme/Car-Sales-Malaysia-Analysis>

YouTube Link:

<https://youtu.be/bpvf6Nwirs4>