Final Report – Car Price Analysis

Objective

The goal of this project is to identify the key features that impact car prices using data analytics. The project is part of the internship task assigned by Novanectar Pvt Ltd.

Data Cleaning and Preparation

- Extracted `carCompany` from `CarName` column and fixed brand spelling issues.

- Dropped unnecessary columns such as `car_ID`, `CarName`, and highly correlated features like `carlength`, `curbweight`, and `wheelbase`.

- One-hot encoded categorical variables to convert them into numeric form.

- Created a derived feature: `car_stability = wheelbase / carlength` (later dropped due to correlation).

Visual Analysis & Insights

Heatmap:

- Showed strong correlations among several features.

- Dropped highly correlated features to reduce multicollinearity.

Histogram:

- Revealed a **right-skewed price distribution**.

- Most cars fall into **mid-range**, with a few **luxury outliers** (BMW, Porsche, etc.).

Bar Chart:

- Brands like **Jaguar, BMW, Porsche** had significantly **higher average prices**.

- Entry-level brands like **Honda, Mazda** were more affordable.

Conclusion

- **Engine size** is a major driver of car price.

- **Fuel efficiency** (MPG) is negatively associated with cost.

- **Brand identity** plays a large role in pricing — luxury brands command a premium.

- Dataset was clean, making it easier to prepare for analysis.

Status

Project completed successfully and submitted via GitHub.

Prepared and submitted as part of the internship at Novanectar Pvt Ltd.

BY - Ashish sahu

Date-23/05/2025