

Analysis of popular cars in Indian Market

Group 20: Aviral, Pulak, Arunansh, Rutvik

1. Introduction

1.1 Introduction to Cars

The automotive industry is undergoing significant changes with the rise of electric vehicles (EVs), autonomous driving technologies, and increasing connectivity within vehicles. In India, the car market is diverse, offering a wide range of vehicles with varying specifications, prices, and features. This presents an opportunity to analyze car data to better understand consumer preferences, pricing trends, and the characteristics of different car models. With advancements in technology and data science, there is an increasing focus on data-driven projects that provide insights into this dynamic industry.

1.2 Motivation

As a group, our motivation for this project is to leverage data science techniques to make automotive information more accessible and valuable to consumers. By developing a Shiny app, we aim to provide an interactive platform where users can explore and compare car specifications, pricing, and other important factors. Our goal is to create a tool that simplifies the process of making informed decisions when purchasing a car and helps users understand the trends and dynamics of the Indian automotive market. This project demonstrates the power of data science in transforming how car data is analyzed, presented, and utilized by consumers and industry professionals.

The Car Data Explorer is a Shiny web application that provides an interactive experience for exploring dataset of most sold cars in India with their corresponding diverse attributes.

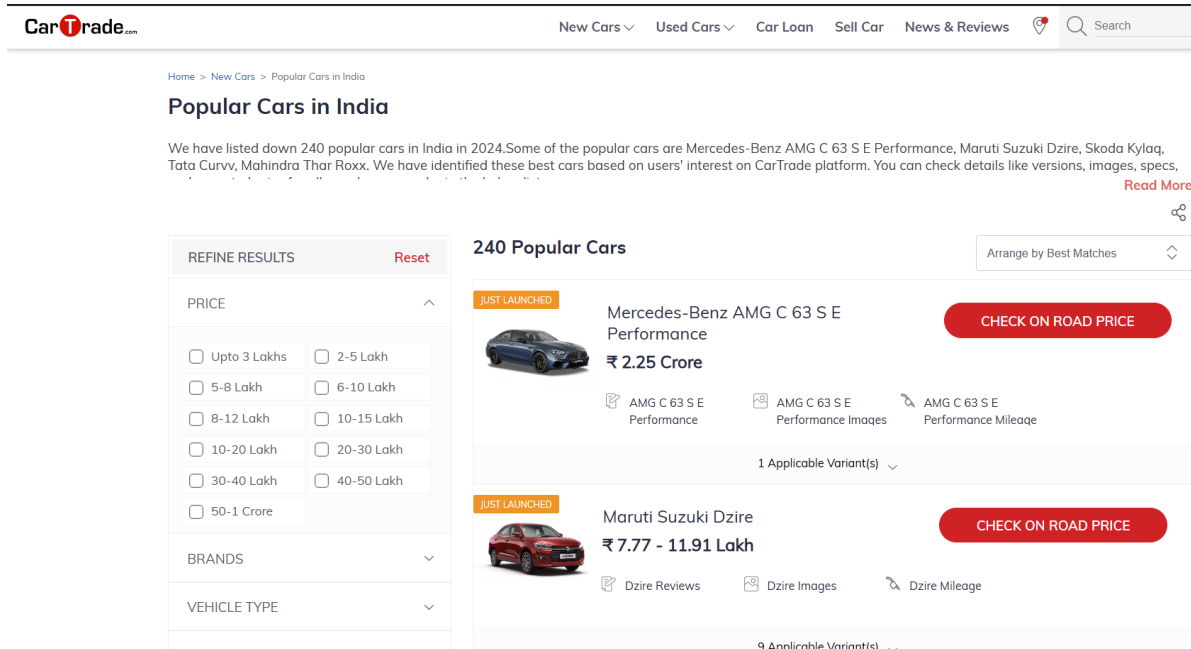


Figure 1: An image of CarTrade website used to collect data

2. Data Used

The data collection process involved web scraping and organizing various car attributes from an external website, [CarTrade](#). The code leveraged libraries such as `rvest`, `tidyverse`, and `httr` to scrape data on car models across multiple pages and subsequently store the data in a structured format.

- **Car_data.Rdata:** Contains *239 obs. of 10 variables* which include details like
 - * **Rank:** The rank of the car based on popularity or other criteria (numeric).
 - * **Name:** The name of the car model.
 - * **Company:** The car manufacturer or company that produces the car.
 - * **Seating_capacity:** The number of seats available in the car.
 - * **Fuel_type:** The type of fuel the car uses, such as Electric, Petrol, Diesel, etc.
 - * **Price:** The price range of the car in terms of currency (e.g., 13.50–15.50 Lakh).
 - * **No_of_Colours:** The number of different colors the car is available in.
 - * **No_of_Variants:** The number of different variants or configurations available for the car.

- * **Cleaned_Price:** The cleaned price of the car (likely numerical, without the range or special characters).
- * **Price_Category:** The price category into which the car falls, such as "10-20 Lakh."
- **Image_Path.Rdata:** Contains the paths to saved car images
- **Colour_Data.Rdata:** Contains color options for each car
- **Variants_Data.Rdata:** Contains variants available for each car
- **Images Folder:** Contains Images of each car whose path has been saved *Image_Path.Rdata*

3. Obtaining the Data

We did not use any *csv* files. All the data used was scraped from [CarTrade](#) website. Following details the process:

1. Car Links Collection:

- A loop was used to access a list of car models across multiple pages, generating links to each car's individual webpage.
- Using the *read_html* function, the code collected URLs for each car's details page, stored them, and appended the base URL to create complete links.

2. Scraping Data from Each Car Page:

- For each car page, the following information was extracted:
 - * Car Name
 - * Seating Capacity
 - * Fuel Type
 - * Price
 - * Color Options
 - * Variant Options
- Prices scraped were cleaned to standardize the format, accounting for prices in units like Crores or Lakhs

3. Image Extraction:

- A URL pointing to each car's image was extracted and saved with a unique filename based on the car name and company.
- Images for each car were scraped and saved locally.

4. Data Cleaning:

- The price data was standardized to convert units into numerical values. Prices were then categorized into ranges: "0-10 Lakh," "10-20 Lakh," "20-50 Lakh," "50 Lakh - 1 Crore," and "Above 1 Crore".
- Any non-standard fuel type (not petrol, diesel, or electric) was categorized as "Hybrid" for consistency

Finally the scraped data is stored *.RData* blobs as mentioned before.

4. Features of the App

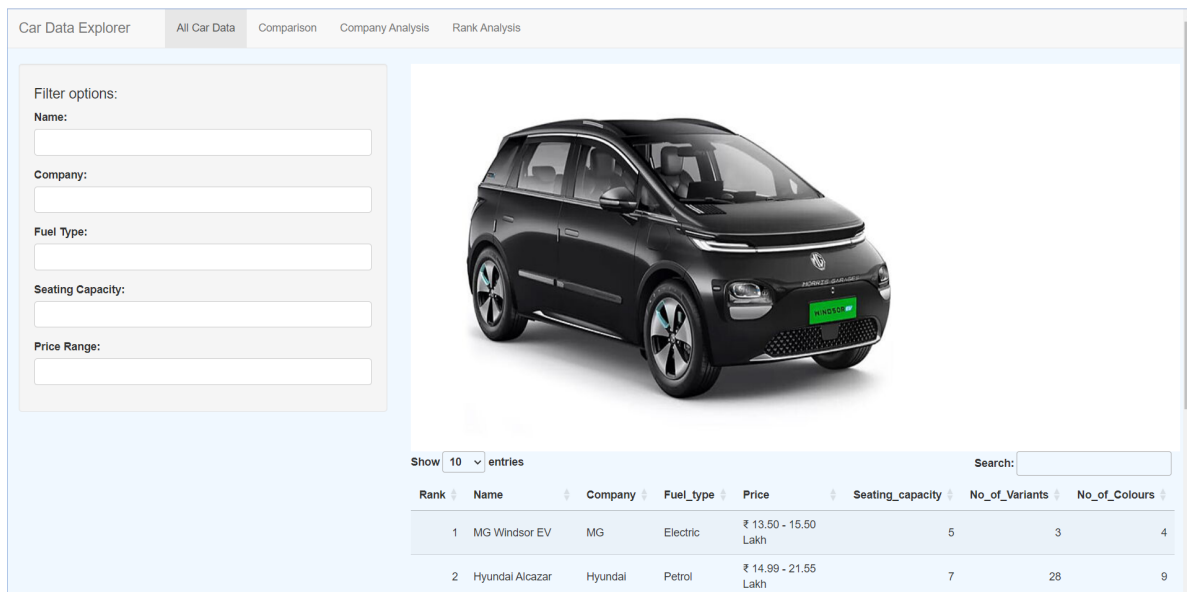


Figure 2: An Image of our ShinyApp

1. All Cars Data Allows users to view the entire car dataset with filter options:

- Name: Filter by car name.
- Company: Filter by car company.
- Fuel Type: Filter by fuel type (Petrol, Diesel, Electric, Hybrid).

- Seating Capacity: Filter by seating capacity.

Displays a car image plot and a data table of filtered results.

2. **Compare Two Cars** Provides an interface for side-by-side comparison of two selected cars. Displays:

- Images of both cars.
- Comparison table showing attributes such as name, company, seating capacity, fuel type, price, number of colors available, and number of variants.

3. **Comparison of Companies** Visualizes car rankings by company and fuel type, with options to filter by:

- Company: Select one or multiple companies to compare.
- Fuel Type: Select a fuel type to refine the comparison.
- Plots:
 - * Rank vs. Company: Shows car rankings for selected companies.
 - * Number of Cars by Company: Histogram showing the count of cars per company.

4. **Comparison of Rank of Cars**

- Visualizes car rank distributions based on fuel type and seating capacity.
- Includes:
 - * Rank Range Slider: Select a range of ranks to display.
 - * Fuel Type Filter: Choose a fuel type for additional filtering.
- Plots:
 - * Rank vs. Fuel Type: Shows rank distributions across fuel types.
 - * Rank vs. Seating Capacity: Shows rank distributions across seating capacities.

5. Possible biases in Data

Potential biases in the dataset may arise from several sources:

- **Brand Overrepresentation:** Certain car companies may have more models represented than others, leading to an imbalance in comparisons. This might lead to incorrect generalizations about results drawn about certain brands viz a viz others.
- **Fuel Type Distribution:** If fuel type distribution is not uniform, it could lead to biased insights regarding the popularity or performance of fuel types.

Such biases could affect the overall analysis and might need to be accounted for when interpreting results.

6. Potential Insights and Questions

Building on the recent COP29 discussions, we can explore the adoption trends of electric and hybrid cars—reflecting India’s progress toward clean energy—using the collected data.

To evaluate India’s readiness for the clean energy transition, we can analyze commitments from both Indian and foreign brands by examining their efforts to offer electric or hybrid models.

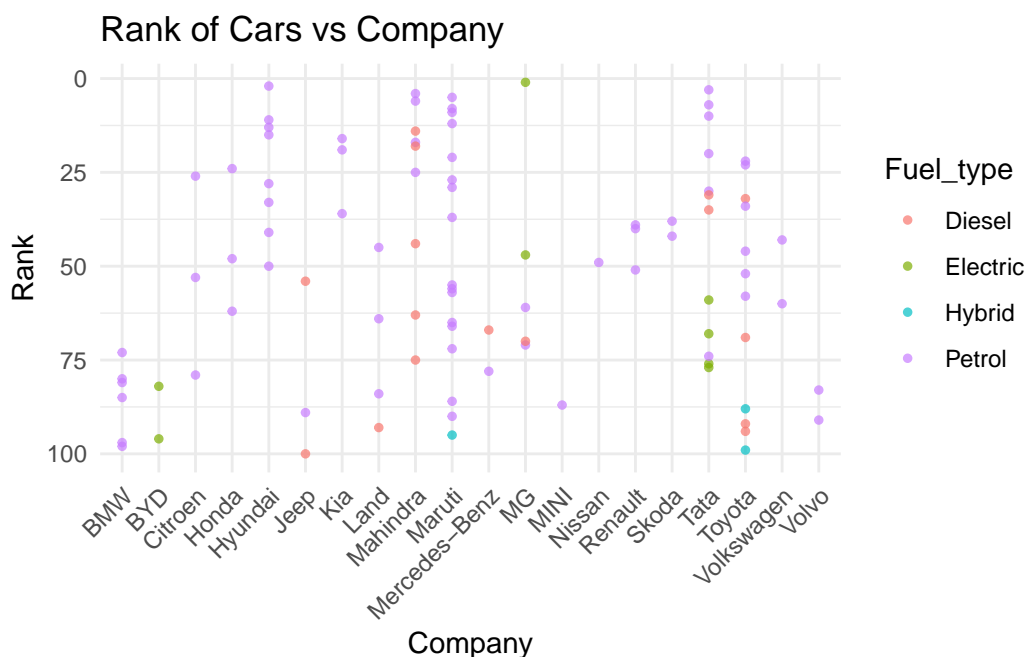
Additionally, we can assess the influence of Indian players in the domestic market through app-based graphs and data, providing insights into their growing role in advancing clean energy.

Hybrid vehicles are touted as an middle ground alternative, we can look into the affordability of this alternative for a common man.

Finally, we can examine the impact of government interventions and subsidies on EVs to determine whether EVs are truly more cost-effective for consumers compared to petrol or diesel vehicles.

7. Key Visualizations for Data Insights

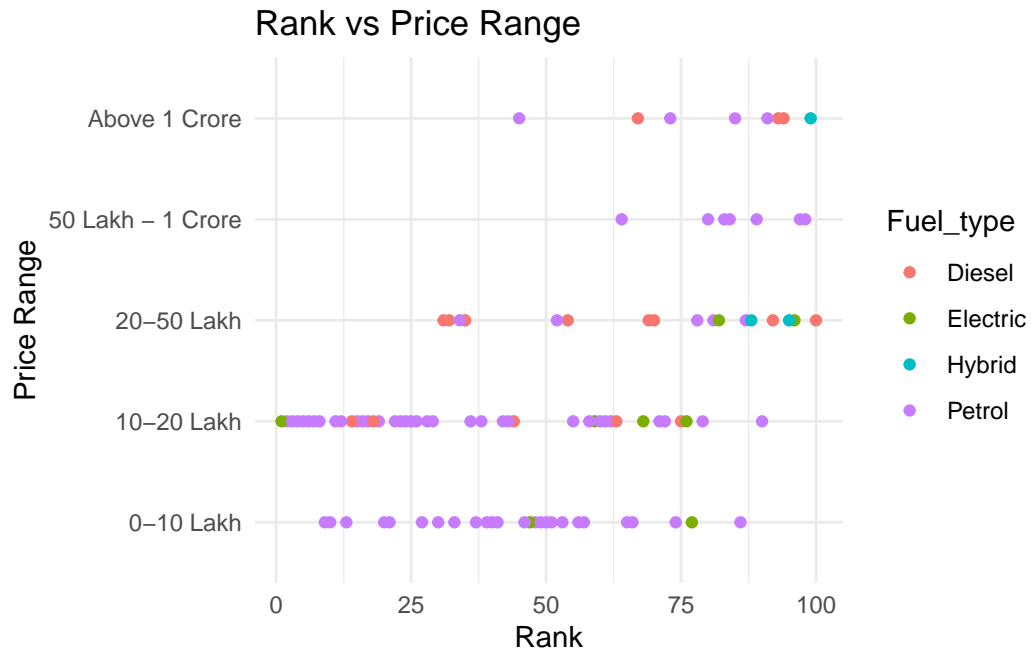
Below drawn plot shows the popularity rank of cars by company, filtered to only show top 100 most popular cars and corresponding brands. The graph clearly indicates that **petrol-powered cars lead the popular Indian market**, with diesel-powered cars following. Moreover as seen from the graph, **the top 100 best-selling cars are predominantly from Indian brands**, including Maruti, Tata, and Mahindra.



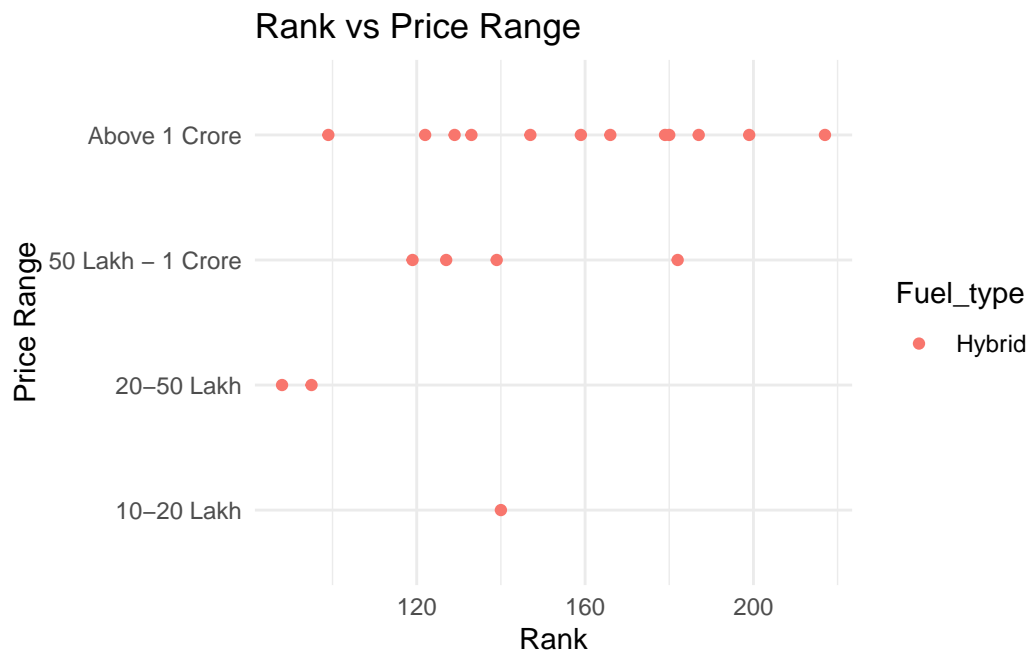
The graph shows the relationship between car rank and price range, categorized by fuel type. From the plot, it is evident that petrol-run cars dominate the lower price segments, particularly in the range from 0 to 20 Lakh. Even with the government incentives aimed at promoting greener technologies, **petrol-powered cars still represent the majority of the more affordable options.**

This trend is particularly notable in the **"0-10 Lakh" price range, where petrol cars outnumber other fuel types by a significant margin.** Electric and hybrid vehicles, although gaining popularity, remain less prevalent in these lower price ranges, which may be attributed to higher initial costs despite the incentives.

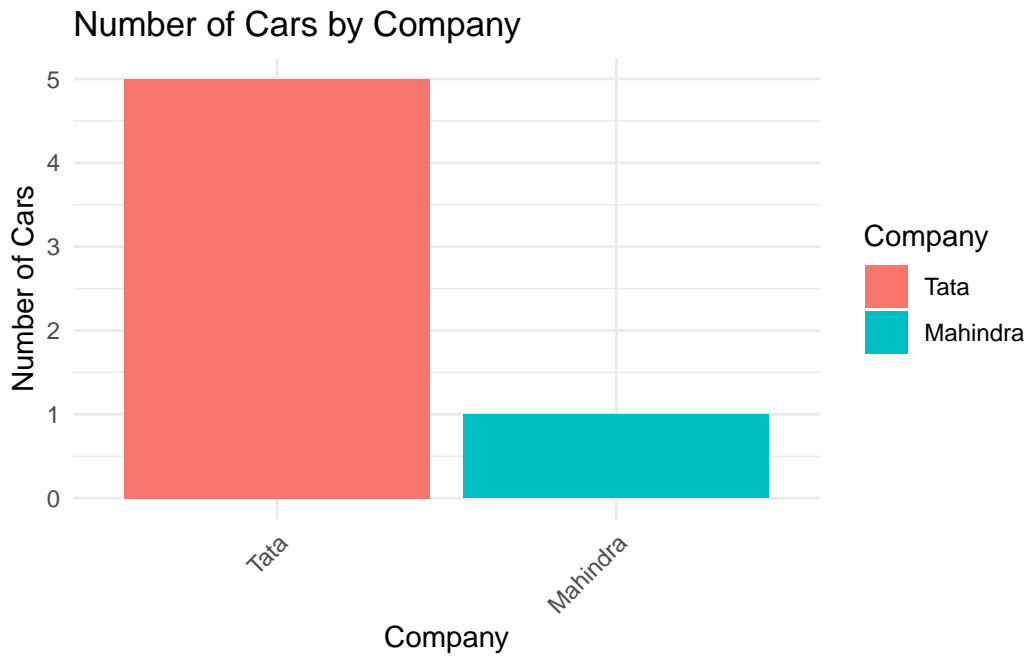
The presence of diesel cars in the mid-price ranges (20-50 Lakh) suggests that consumers are still choosing fuel-efficient options within this category, particularly where price sensitivity is less of an issue. The electric cars and hybrid cars, represented by green and blue points, show a scattering of models across various price ranges, with a concentration in the higher price categories, reflecting a growing interest in eco-friendly options but at a higher price point.



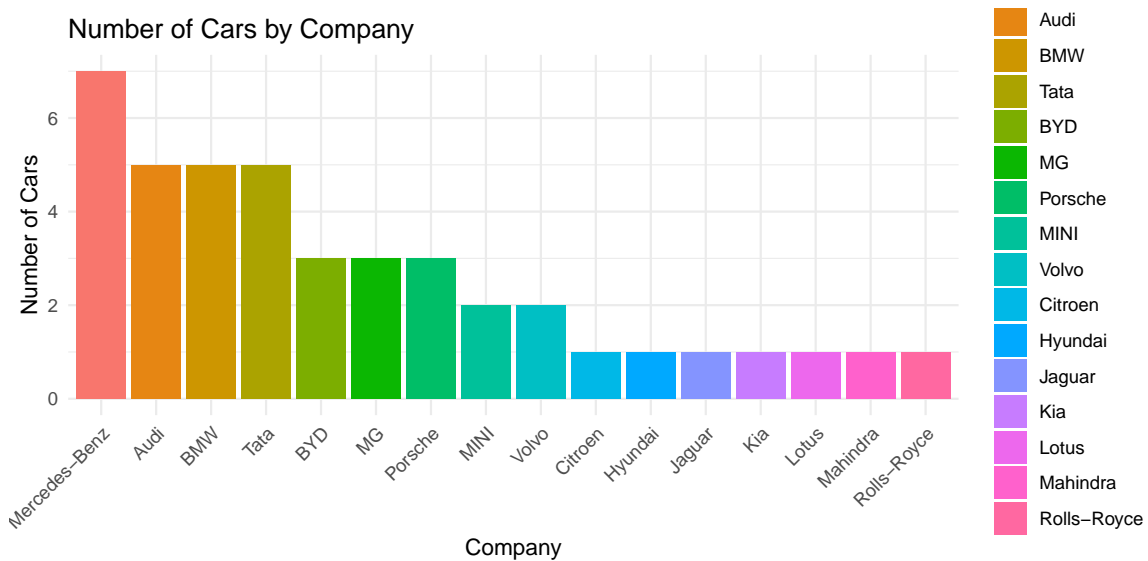
If we chart the prices of hybrid options available in the Indian market, it becomes clear that most fall within a higher price range (with a few exceptions), indicating that this technology is not yet affordable on a large scale.



Furthermore, if we plot histogram of electric vehicles available three most popular (which we could infer from previous plots) Indian car brands i.e. Mahindra, Tata and Maruti. We see that there are much less electric vehicle options available, with Maruti having none at all!



The previous inference of **Indian brands having smaller selection of electric cars** can be contrasted with the abundant availability of electric options by foreign players from the following histogram.



Consolidating all the inferences drawn from the graph, we can say:

1. Petrol-powered cars lead the popular Indian market, and constitute the majority of the cheaper options even after govt. incentives.
2. Indian brands having smaller selection of electric cars

This suggests there is less preparedness for greener transition among Indian players. This can be attributed to ineffective policy measures and lack of commitment by Indian firms.

8. Final Conclusions

We created Car Data Explorer, an interactive Shiny web application that allows users to explore a dataset of cars with various attributes, including name, company, seating capacity, fuel type, price, and rank. The app includes features to filter, compare, and visualize insights into car rankings and company distributions.

Furthermore, using the app's graphs and the dataset collected, we were able to draw inference regarding the clean future of Indian roads.

References

- [CarTrade](#) website
- [ggplot2](#) documentation
- [ShinyApp](#) documentation