## Portfolio Project Report: Car Industry Insights through Data Analysis

**1. Project Overview**

This project aimed to extract valuable insights from car data to inform data-driven strategies for the automotive industry. The data was obtained through web scraping a comprehensive car information website (<https://www.cars-data.com/>). The Python libraries BeautifulSoup and Requests facilitated data extraction, while Pandas was used to create a structured DataFrame. After cleaning in Python, Microsoft Excel was used for further data refinement and data exploration. Finally, the prepared data was loaded into a PostgreSQL database for in-depth analysis using SQL queries. Power BI served as the platform for data visualization, presenting key findings in an interactive format.

**2.1 Data Acquisition**

Web scraping techniques were employed to gather car information from a designated website. Python libraries, BeautifulSoup and Requests, were used in retrieving the website's HTML code across various pages. By parsing the code, specific details such as car names, years of production, number of doors, and body styles were extracted.

**2.2 Data Cleaning and Preprocessing**

The acquired data was cleaned in Python using Pandas. This involved and selecting relevant texts to ensure a well-structured format. Thereafter, Microsoft Excel was used for further data cleaning and exploration, preparing the data for advanced analysis.

**2.3 Data Storage and Analysis**

PostgreSQL served as the chosen database management system for storing the prepared car data. SQL queries were then executed to perform in-depth analysis:

1: Identifying Popular Body Styles

SELECT bodystyles, COUNT(\*) AS BodyStylesCount

FROM cars

GROUP BY bodystyles

ORDER BY BodyStylesCount DESC;

This query identifies the most common body styles in the dataset by counting the occurrences of each unique body style and ordering the results in descending order based on their frequency.

2: Average Doors per Body Style

SELECT bodystyles, AVG(number\_of\_doors) AS avg\_doors

FROM cars

GROUP BY bodystyles;

This query calculates the average number of doors for each body style. It groups the data by body style and then calculates the average number of doors within each group.

3: Car Production Volume by Year

SELECT years\_of\_production, COUNT(\*) AS count

FROM cars

GROUP BY years\_of\_production

ORDER BY years\_of\_production;

This query analyzes car production trends by identifying the number of cars produced within each year. It groups the data by year of production and then counts the number of cars within each year.

4: Average Doors Over Time (Considering Body Style)

SELECT

years\_of\_production,

bodystyles,

AVG(number\_of\_doors) AS avg\_doors\_per\_style

FROM cars

GROUP BY years\_of\_production, bodystyles

ORDER BY years\_of\_production, bodystyles;

This query analyzes how the average number of doors for cars has changed over time, based on different body styles. It groups the data by both year of production and body style, then calculates the average number of doors within each unique combination.

**2.4 Data Visualization**

Power BI, was used to create interactive visualizations based on the analyzed data. These visualizations provided clear insights into:

* Car production trends over time.
* The popularity of various body styles.
* Consumer preferences regarding the number of doors.
* The relationship between body style and door configuration.

**3. Findings and Insights**

The data analysis provided valuable insights into consumer preferences and car production trends. Key findings include:

* **Understanding Body Style Popularity:** Analyzing body style distribution revealed the most prevalent styles within the dataset. This information can guide production focus towards high-demand styles like Hatchbacks or Sedan cars.
* **Door Configuration Preferences:** Identifying popular door configurations within each body style empowers manufacturers to optimize production and potentially streamline offerings for less popular configurations.
* **Car Production Trends:** Analyzing production trends over time helps pinpoint emerging market segments. This allows for research and development efforts to be directed towards these areas, such as 5 door hatchbacks.
* **Consumer Preferences and Marketing:** Insights on body style and door preferences can be leveraged to tailor marketing campaigns. For instance, SUVs can be highlighted for their spaciousness and family-friendly features.

**4. Recommendations for the Automotive Industry**

Based on the data-driven insights, the following recommendations are proposed for the automotive industry:

* **Optimize Door Configurations:** Identify the most popular door configurations within each body style. Consider offering a wider range of options within those configurations while potentially streamlining production for less popular configurations. For example, Based on the findings, the most popular SUVs have 5 doors, focus on offering a variety of trim levels and features within those configurations
* **Production Focus:** Focus production efforts on body styles showing consistent demand or increasing popularity. Simultaneously, optimize door configurations within each body style to cater to consumer preferences. This could involve allocating more resources to manufacturing Hatchback vehicles (Based On findings).

**5. Conclusion**

This project successfully utilized web scraping, data cleaning, SQL analysis, and Power BI visualization to extract valuable insights from car data. The findings provide a strong foundation for data-driven strategies in the automotive industry, from production optimization and targeted marketing to product development that aligns with evolving consumer preferences. By implementing these recommendations, automotive companies can gain a competitive edge and enhance their market success.