Aaron Stefano F. Bonaobra **9/03/2025**

BSIT – 3A

Big Data Analysis

Car Price Visualization Project – Documentation

Introduction

This project focuses on visualizing car price data using Matplotlib and Seaborn. The dataset provides key attributes such as price, engine size, mileage, fuel type, and brand. Through various plots, we explore trends and relationships within the data to uncover insights into car pricing.

Data Source

The dataset is stored in car_price_dataset.csv. It contains multiple features, including:

- Price ($): The selling price of the car.
- Engine Size (L): The engine capacity in liters.
- Brand: The manufacturer of the car.
- Fuel Type: The type of fuel the car uses (e.g., petrol, diesel, electric).
- Mileage (km): The total distance the car has traveled.

Visualizations

We used five different visualizations to analyze the dataset:

1. Histogram - Distribution of Car Prices

- Goal: Understand the spread and frequency of different price ranges.

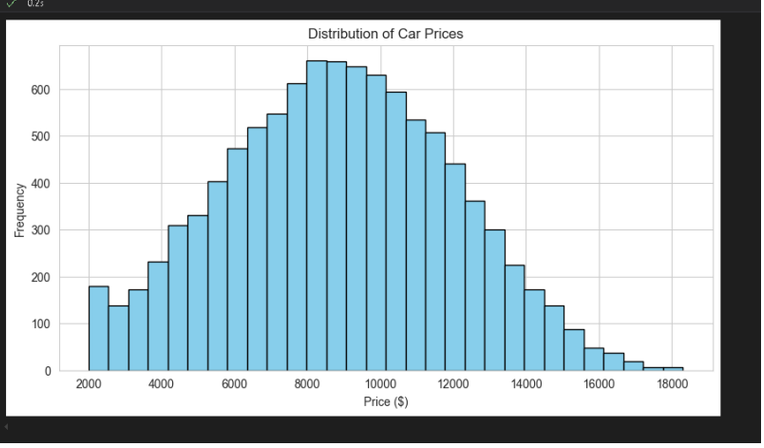
Insights:

Prices are likely skewed towards lower values.

Helps identify common price ranges for used cars.

```
# 1 histogram to show the distribution of car prices
plt.figure(figsize=(10, 5))
plt.hist(df["Price"], bins=30, color='skyblue', edgecolor='black')
plt.xlabel("Price ($)")
plt.ylabel("Frequency")
plt.title("Distribution of Car Prices")
plt.show()
```

[42] ✓ 0.2s



2. Scatter Plot - Engine Size vs. Price

- Goal: Identify the relationship between engine size and price.

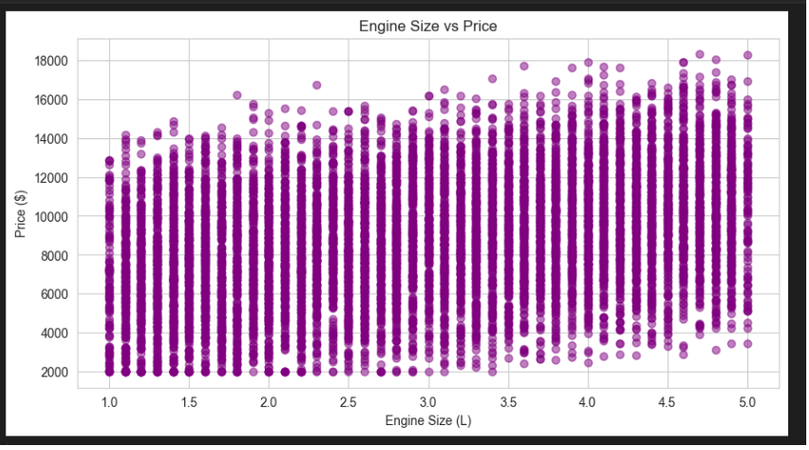
Insights:

Larger engine sizes generally correlate with higher prices.

Helps understand if engine size is a significant price factor.

```
# 2 scatter plot to visualize the relationship between engine size and price
plt.figure(figsize=(10, 5))
plt.scatter(df["Engine_Size"], df["Price"], alpha=0.5, color='purple')
plt.xlabel("Engine Size (L)")
plt.ylabel("Price ($)")
plt.title("Engine Size vs Price")
plt.show()
```

[43] ✓ 0.3s



3. Bar Plot - Average Price of Top 10 Car Brands

- Goal: Compare the average price of the most common car brands.

Insights:

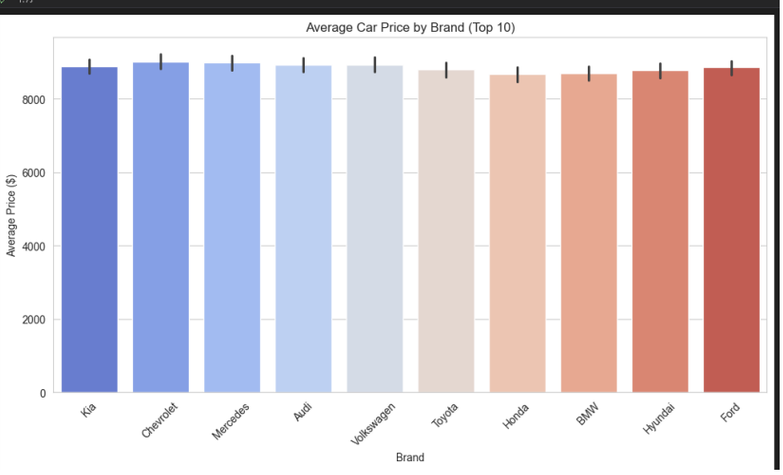
Some brands hold higher market value than others.

Useful for brand-based price comparison.

```
# 3 bar plot showing the average price for the top 10 most common car brands
top_brands = df["Brand"].value_counts().head(10).index
df_top_brands = df[df["Brand"].isin(top_brands)]

plt.figure(figsize=(12, 6))
sns.barplot(data=df_top_brands, x="Brand", y="Price", estimator=lambda x: sum(x)/len(x), hue="Brand", dodge=Fal
plt.xticks(rotation=45)
plt.xlabel("Brand")
plt.ylabel("Average Price ($)")
plt.title("Average Car Price by Brand (Top 10)")
plt.show()
```

[44] ✓ 1.7s



4. Pie Chart - Distribution of Fuel Types

- Goal: Show the proportion of different fuel types in the dataset.

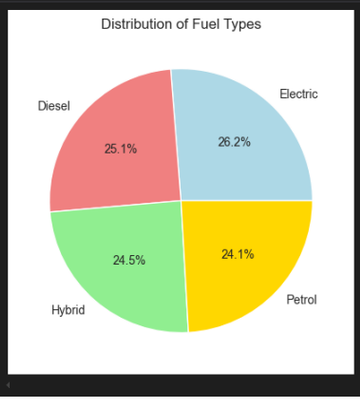
Insights:

Identifies the most common fuel type in the dataset.

Can reveal trends in consumer preferences (e.g., shift to electric vehicles).

```
# 4 pie chart to show the distribution of fuel types
plt.figure(figsize=(8, 5))
df["Fuel_Type"].value_counts().plot(kind="pie", autopct='%1.1f%%', colors=['lightblue', 'lightcoral', 'lightgreen', 'gold'])
plt.title("Distribution of Fuel Types")
plt.ylabel("")  # hide y-label for cleaner look
plt.show()
```

[45] ✓ 0.0s



5. Scatter Plot - Mileage vs. Price (Colored by Fuel Type)

- Goal: Understand how mileage impacts car pricing.

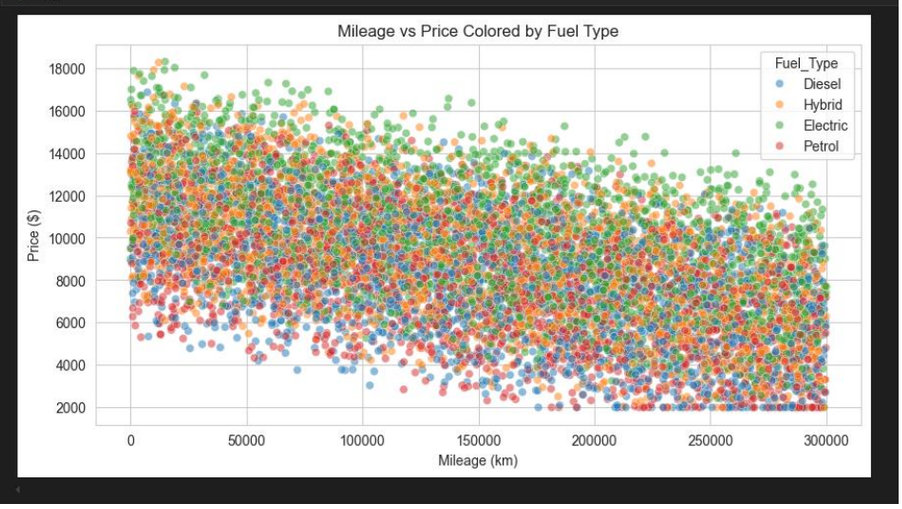
Insights:

High-mileage cars generally have lower prices.

Different fuel types may have different depreciation rates.

```
# 5 scatter plot to compare mileage and price, colored by fuel type
plt.figure(figsize=(10, 5))
sns.scatterplot(data=df, x="Mileage", y="Price", hue="Fuel_Type", alpha=0.5)
plt.xlabel("Mileage (km)")
plt.ylabel("Price ($)")
plt.title("Mileage vs Price Colored by Fuel Type")
plt.show()
```

[46] ✓ 0.6s



Technical Details

- Matplotlib is used for basic plotting (plt.hist, plt.scatter, etc.).
- Seaborn enhances the visual aesthetics (sns.barplot, sns.scatterplot).
- Color Schemes: Applied for better readability (e.g., coolwarm, skyblue, purple).
- Formatting: Titles, axis labels, and figure sizes are customized for clarity.

Conclusion

This visualization project provides insights into car pricing based on various attributes. The different charts allow for a comprehensive understanding of price distribution, brand value, fuel preferences, and key factors affecting price variation. These findings can be useful for car buyers, sellers, and analysts in the automotive industry.

Thank you so much sir! God bless po palagi! ^_^