Name: Perez, Junmar A.
Git Repo: https://github.com/arnthegreat/Bonaobra_Elec2.git

Lab No. 5

Date: 03/20/2025

Objective

This lab visualizes car price data using Matplotlib and Seaborn. It aims to find trends and relationships between price, engine size, mileage, fuel type, and brand.

Introduction

Data visualization helps in understanding trends. This lab uses Matplotlib and Seaborn to analyze car prices based on key attributes like engine size, brand, fuel type, and mileage.

Methodology

1. Loaded the dataset into a Pandas DataFrame.

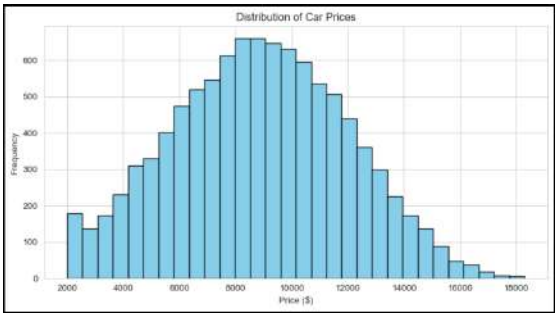
```
# load the dataset
file_path = "car_price_dataset.csv"
df = pd.read_csv(file_path)

# set seaborn style for better visuals
sns.set_style("whitegrid")
```

[3] ✓ 0.0s

2. Created a histogram to show car price distribution.

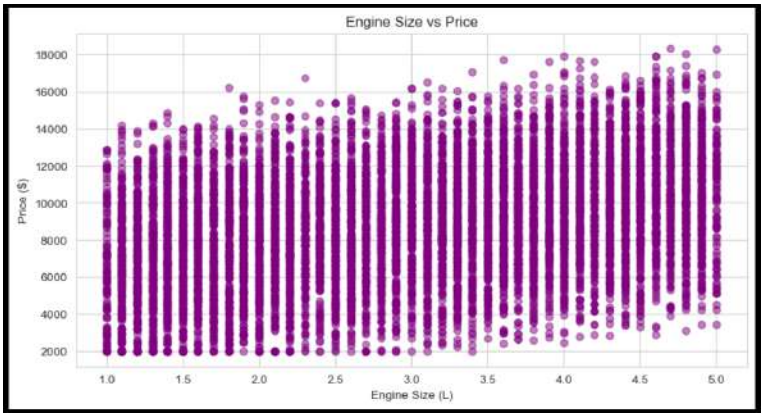
```
# set seaborn style for better visuals
sns.set_style("whitegrid")

# 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()
```

✓ 0.1s



3. Made a scatter plot to analyze engine size vs. price.

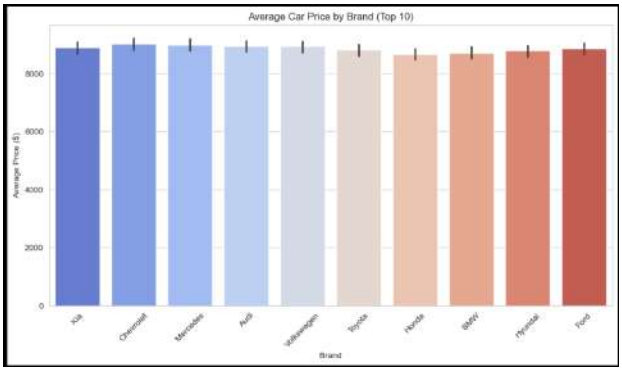
```
# 2 scatter plot to visualize the relationship between engine size and price
plt.figure(figsize=(10, 5))
plt.scatter(df["Engine_Size"], df["Price"], alpha=0.5, color='purple')
plt.xlabel("Engine Size (L)")
plt.ylabel("Price ($)")
plt.title("Engine Size vs Price")
plt.show()
```



4. Used a bar plot to compare the average price of the top 10 brands.

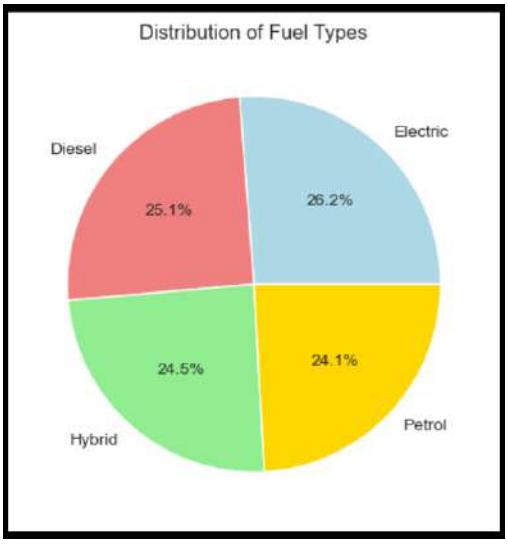
```
# 3 bar plot showing the average price for the top 10 most common car brands
top_brands = df["Brand"].value_counts().head(10).index
df_top_brands = df[df["Brand"].isin(top_brands)]

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=False, palette="coolwarm", legend=
plt.xticks(rotation=45)
plt.xlabel("Brand")
plt.ylabel("Average Price ($)")
plt.title("Average Car Price by Brand (Top 10)")
plt.show()
```

Python



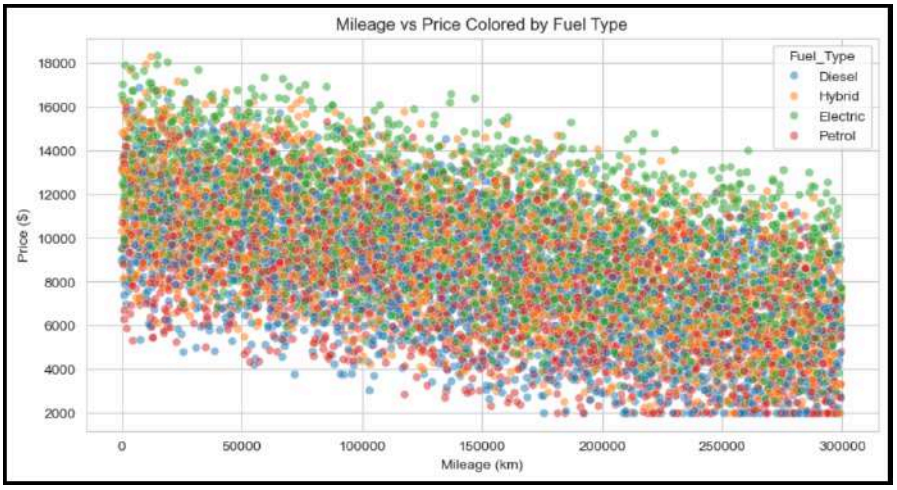
5. Created a pie chart to show the proportion of fuel types used in the dataset.

```
# 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()
```



6. Created a scatter plot to analyze mileage vs. price.

```
# 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()
```



Results and Analysis

The histogram showed most cars are in the lower price range. The scatter plot indicated that larger engines generally have higher prices. The bar plot revealed that some brands are priced significantly higher than others.

Challenges and Solutions

Handling missing values was a challenge, solved by filling or removing them. Adjusting visualization settings for better readability was fixed by modifying figure sizes and labels.

Conclusion

This lab demonstrated how data visualization helps analyze car prices. The charts provided insights into pricing patterns, and Matplotlib and Seaborn proved useful for extracting trends from data.