import pandas as pd

# Load the car dataset (replace 'car\_data.csv' with your actual file path)

file\_path = 'car\_data.csv'

df = pd.read\_csv(file\_path)

# Display basic information about the dataset

print("Dataset Information:")

print(df.info())

# Display the first few rows of the dataset

print("\nFirst Few Rows of the Dataset:")

print(df.head())

# Summary statistics of numerical columns

print("\nSummary Statistics:")

print(df.describe())

# Analyze categorical variables (e.g., Car Make)

car\_make\_counts = df['Make'].value\_counts()

print("\nCar Make Distribution:")

print(car\_make\_counts)

# Example: Average Price per Make

average\_price\_per\_make = df.groupby('Make')['MSRP'].mean()

print("\nAverage Price per Car Make:")

print(average\_price\_per\_make)

# Example: Correlation between numerical variables

correlation\_matrix = df.corr()

print("\nCorrelation Matrix:")

print(correlation\_matrix)

# Example: Filtering the dataset based on conditions

# For instance, display cars with a price greater than 50000

expensive\_cars = df[df['MSRP'] > 50000]

print("\nExpensive Cars:")

print(expensive\_cars[['Make', 'Model', 'MSRP']])

# Visualization (requires additional libraries like matplotlib or seaborn)

# For example, plotting a histogram of car prices

import matplotlib.pyplot as plt

plt.hist(df['MSRP'], bins=20, color='blue', edgecolor='black')

plt.title('Distribution of Car Prices')

plt.xlabel('Car Price (MSRP)')

plt.ylabel('Number of Cars')

plt.show()

pip install pandas matplotlib