

Data Science – Report Task 1

Task - Descriptive Statistics and Data Profiling

Analyse a small dataset (e.g., student scores, product reviews) and provide summary statistics (mean, median, mode, etc.) and insights.

Summary - The project involves analysing a dataset titled "CAR DETAILS FROM CAR DEKHO.csv" to perform descriptive statistics and data profiling. Using Python libraries such as Pandas, NumPy, and Matplotlib, the data is explored to compute summary statistics like mean, median, and mode for key attributes, including the year of manufacture, selling price, kilometres driven, and fuel type. The notebook organizes this information into a dictionary for easy reference, highlighting trends in both numerical and categorical data. This approach provides insights into the dataset's structure and characteristics, forming the basis for deeper analysis or visualization.

```
Importing Necessary Libraries and Load Dataset

[19] # Importing necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# Loaded the dataset
file_path = '/content/CAR DETAILS FROM CAR DEKHO.csv'
car_data = pd.read_csv(file_path)

Calculating Mean, Median & Mode

[21] stats = {
    'name': {
        'mode': car_data['name'].mode()[0]
    },
    'year': {
        'mean': car_data['year'].mean(),
        'median': car_data['year'].median(),
        'mode': car_data['year'].mode()[0]
    },
    'selling_price': {
        'mean': car_data['selling_price'].mean(),
        'median': car_data['selling_price'].median(),
        'mode': car_data['selling_price'].mode()[0]
    },
    'km_driven': {
        'mean': car_data['km_driven'].mean(),
        'median': car_data['km_driven'].median(),
        'mode': car_data['km_driven'].mode()[0]
    },
    'fuel': {
        'mean': None,
        'median': None,
        'mode': car_data['fuel'].mode()[0]
    }
}
```

```
Statistical Analysis

Visualize Year Statistics

[22] plt.figure(figsize=(8, 6))
plt.bar(['Mean', 'Median', 'Mode'], [stats['year']['mean'], stats['year']['median'], stats['year']['mode']], color='skyblue')
plt.title('Year Statistics')
plt.ylabel('Year')
plt.text(1, 1250, f"Mean: {stats['year']['mean']:.2f} | Median: {stats['year']['median']} | Mode: {stats['year']['mode']}",
        fontsize=10, ha='center', bboxdict(boxstyle='round', facecolor='lightgrey', alpha=0.5))
plt.savefig('year_stats.jpeg', format='jpeg')
plt.show()

print(f"Year Statistics:")
print(f"Mean: {stats['year']['mean']:.2f}")
print(f"Median: {stats['year']['median']}")
print(f"Mode: {stats['year']['mode']}")
print(f"{'.' * 50}")
```

Visualize Selling Price Statistics

```
plt.figure(figsize=(9, 6))
plt.bar(['Mean', 'Median', 'Mode'], [stats['selling_price']['mean'], stats['selling_price']['median'], stats['selling_price']['mode']], color='green')
plt.title('Selling Price Statistics')
plt.ylabel('Price ($)')
plt.text(1, 200000, f"Mean: {stats['selling_price']['mean']:.2f}\nMedian: {stats['selling_price']['median']:.2f}\nMode: {stats['selling_price']['mode']}",
        fontsize=10, ha='center', bbox=dict(boxstyle="round", facecolor="lightgrey", alpha=0.5))
plt.savefig('selling_price_stats.jpeg', format='jpeg')
plt.show()

print(f"Selling Price Statistics:")
print(f"Mean: {stats['selling_price']['mean']:.2f}")
print(f"Median: {stats['selling_price']['median']:.2f}")
print(f"Mode: {stats['selling_price']['mode']}")
print('-' * 50)
```

Visualize KiloMeter Driven Statistics

```
plt.figure(figsize=(9, 6))
plt.bar(['Mean', 'Median', 'Mode'], [stats['km_driven']['mean'], stats['km_driven']['median'], stats['km_driven']['mode']], color='orange')
plt.title('KM Driven Statistics')
plt.ylabel('KM')
plt.text(1, 40000, f"Mean: {stats['km_driven']['mean']:.2f}\nMedian: {stats['km_driven']['median']:.2f}\nMode: {stats['km_driven']['mode']}",
        fontsize=10, ha='center', bbox=dict(boxstyle="round", facecolor="lightgrey", alpha=0.5))
plt.savefig('km_driven_stats.jpeg', format='jpeg')
plt.show()

print(f"KM Driven Statistics:")
print(f"Mean: {stats['km_driven']['mean']:.2f}")
print(f"Median: {stats['km_driven']['median']:.2f}")
print(f"Mode: {stats['km_driven']['mode']}")
print('-' * 50)
```

Visualize Fuel Type Distribution

```
fuel_counts = car_data['fuel'].value_counts()
plt.figure(figsize=(8, 6))
plt.bar(fuel_counts.index, fuel_counts.values, color='purple')
plt.title('Fuel Type Distribution')
plt.ylabel('Count')
plt.savefig('fuel_type_distribution.jpeg', format='jpeg')
plt.show()

print(f"Fuel Type Mode: {stats['fuel']['mode']}")
print('-' * 50)
```

Visualize Top 10 Car Models

```
top_cars = car_data['name'].value_counts().head(10)
plt.figure(figsize=(8, 8))
plt.pie(top_cars.values, labels=top_cars.index, autopct='%1.1f%%', colors=plt.cm.tab10.colors)
plt.title('Top 10 Cars Available')
plt.savefig('top_10_cars_pie_chart.jpeg', format='jpeg')
plt.show()

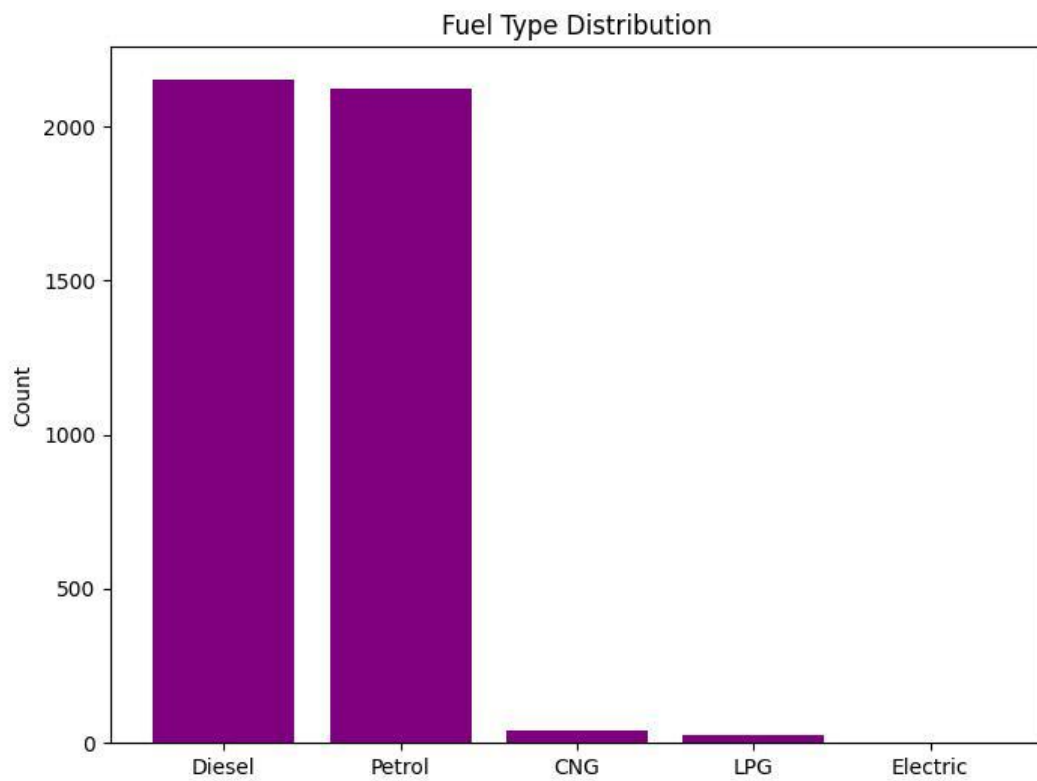
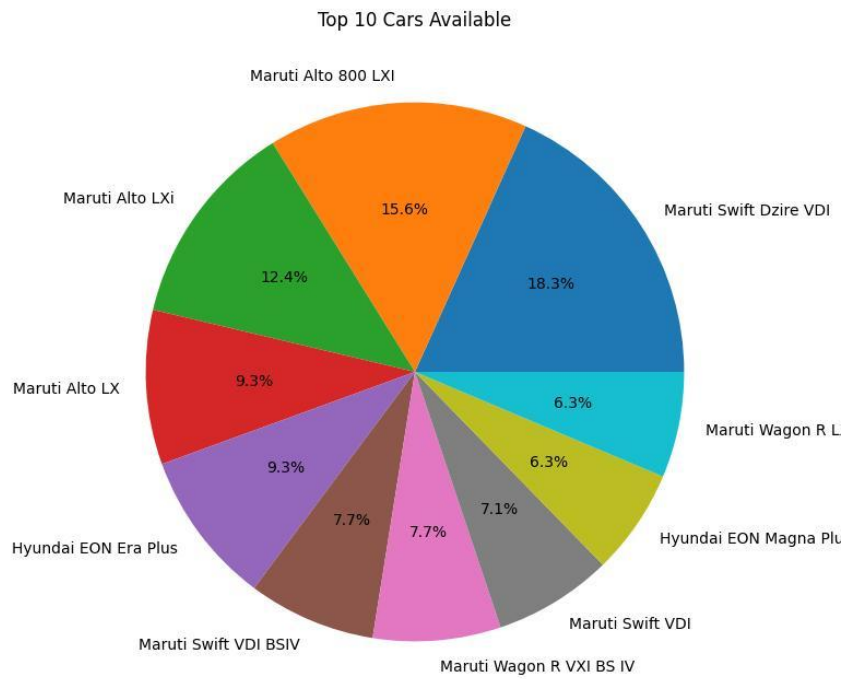
print(f"Top 10 Car Models (Most frequent):")
for car, count in top_cars.items():
    print(f"({car}): {count} cars")
print('-' * 50)
```

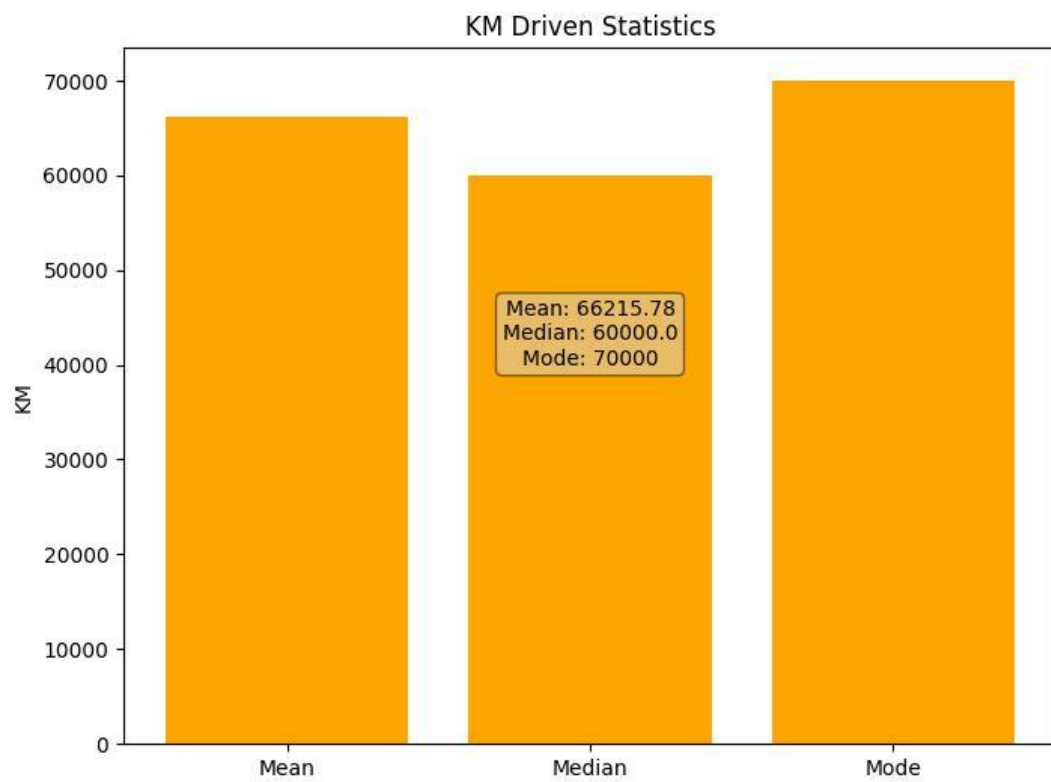
Display of all Stats

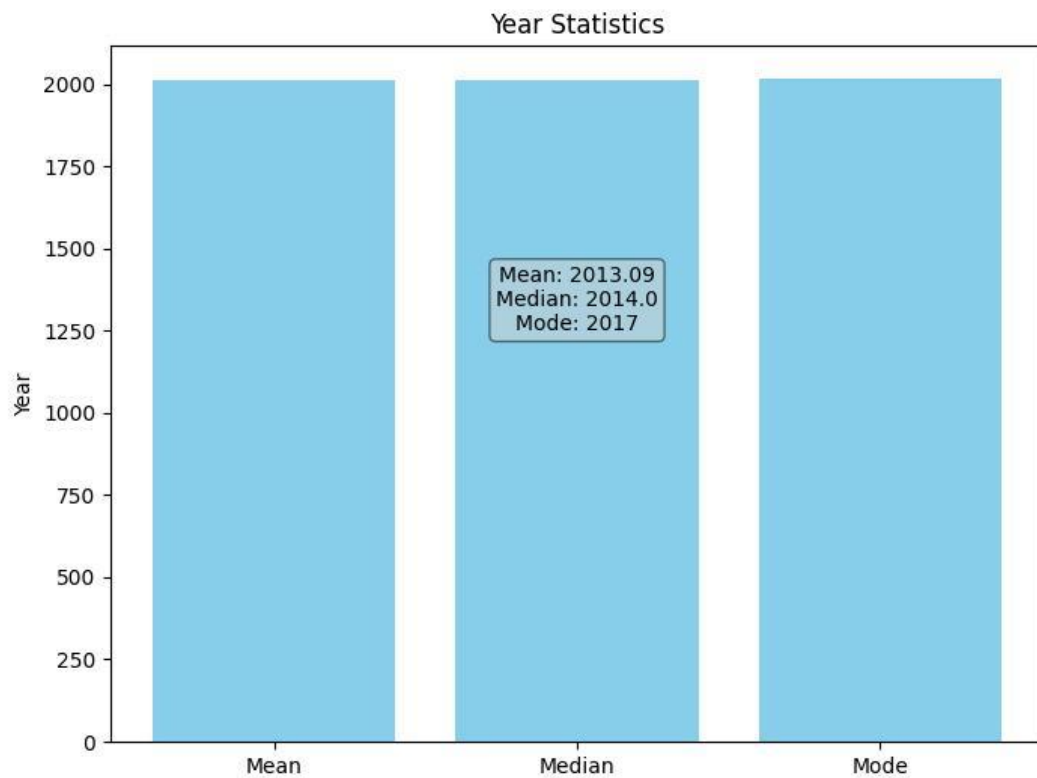
```
# Create a pandas DataFrame to display the statistics in table format
summary_df = pd.DataFrame({
    'Column': ['Year', 'Selling Price', 'KM Driven', 'Fuel'],
    'Mean': [
        stats['year']['mean'],
        stats['selling_price']['mean'],
        stats['km_driven']['mean'],
        None # No mean for categorical 'fuel'
    ],
    'Median': [
        stats['year']['median'],
        stats['selling_price']['median'],
        stats['km_driven']['median'],
        None # No median for categorical 'fuel'
    ],
    'Mode': [
        stats['year']['mode'],
        stats['selling_price']['mode'],
        stats['km_driven']['mode'],
        stats['fuel']['mode']
    ]
})

# Print the summary table
print("\nSummary Statistics:")
print(summary_df)
```

Output —







Skills Learned - Data exploration, statistics.

Tools - Python (Pandas, Numpy, Matplotlib).

Conclusion - In conclusion, the analysis provided key insights into the dataset by calculating summary statistics like mean, median, and mode for important attributes. This profiling highlights trends such as average car prices, popular models, and fuel types, offering a clear understanding of the data. These insights serve as a solid foundation for further analysis or decision-making.