MINI PROJECT -1 : HISTOGRAMS, BAR CHARTS & SCATTERPLOTS

Prepared by: Jagadeesh Reddy Vanga
SBU ID: 115396272
GitHub Link: github.com

Dataset Source: Principal Component Analysis | Kaggle

Kaggle Dataset Attributes:

```
symboling,make,fuel_type,aspiration,num_of_doors,body_style,drive_wheels,engine_location,wheel_base,length,width,height,curb_weight,engine_type,num_of_cylinders,engine_size,fuel_system,bore,stroke,compression_ratio,horsepower,peak_rpm,city_mpg,highway_mpg,price
```

We have discarded `'make', 'fuel_type', 'aspiration', 'num_of_doors', 'body_style', 'drive_wheels', 'engine_location', 'engine_type', 'num_of_cylinders', 'fuel_system'` as these are categorical data

Modifications done to dataset:

- Generated new data points using

```
df = pd.read_csv('pca/autos.csv')

# Generate new data
new_data = pd.DataFrame()

for column in df.columns:
    if df[column].dtype == np.number:
        mean = df[column].mean()
        std = df[column].std()
        new_data[column] = np.around(np.random.normal(mean, std, size=len(df)), 2)
    else:
        new_data[column] = np.random.choice(df[column], size=len(df))

# Save the new data to a CSV file
```

```
new_data.to_csv('autos_generated.csv', index=False)
```