1-ExData

November 24, 2022

```python
import pandas as pd
import numpy as np
import seaborn as sns #visualisation
import matplotlib.pyplot as plt #visualisation
%matplotlib inline
sns.set(color_codes=True)
```

```python
df = pd.read_csv("data\cardata.csv")
```

```python
df.dtypes
```

```python
df.head(2)
```

```python
df = df.drop(["Engine Fuel Type", "Market Category", "Vehicle Style",
    "Popularity", "Number of Doors", "Vehicle Size"], axis=1)
df.head(2)
```

```python
df = df.rename(columns={"Engine HP": "HP", "Engine Cylinders": "Cylinders",
    "Transmission Type": "Transmission", "Driven_Wheels": "Drive Mode","highway
    MPG": "MPG-H", "city mpg": "MPG-C", "MSRP": "Price" })
```

```python
df.head(2)
```

```python
df.shape
```

```python
duplicate_rows_df = df[df.duplicated()]
print("number of duplicate rows: ", duplicate_rows_df.shape)
```

```python
df.count()
```

```python
df = df.drop_duplicates()
df.head(5)
```

```python
df.count()
```

```python
print(df.isnull().sum())
```

```
df = df.dropna()
df.count()
```

```
sns.boxplot(x=df["Price"])
```

```
sns.boxplot(x=df["HP"])
```

```
sns.boxplot(x=df['Cylinders'])
```

```
Q1 = df.quantile(0.25)
Q3 = df.quantile(0.75)
IQR = Q3-Q1
print(IQR)
```

```
df = df[~((df < (Q1-1.5 * IQR)) |(df > (Q3 + 1.5 * IQR))).any(axis=1)]
```

```
df.shape
```

```
df.Make.value_counts().nlargest(40).plot(kind="bar", figsize=(10,5))
plt.title("Number of cars by make")
plt.ylabel("Number of cars")
plt.xlabel("Make");
```

```
plt.figure(figsize=(20,10))
c= df.corr()
sns.heatmap(c,cmap="BrBG",annot=True)
c
```

```
fig, ax = plt.subplots(figsize=(10,6))
ax.scatter(df["HP"], df["Price"])
ax.set_xlabel("HP")
ax.set_ylabel("Price")
plt.show()
```