Bahria University,

Karachi Campus

A logo with text on it

Description automatically generated

LAB EXPERIMENT NO.

10

LIST OF TASKS

|  |  |
| --- | --- |
| TASK NO | OBJECTIVE |
| **01** | An automotive company wants to identify and treat outliers in their "mtcars" dataset. which contains information about various car models. They suspect that certain car models might have outlier values in certain variables and want to analyze and handle them using Python. |
| **02** | An automotive company has identified outliers in their "mtcars" dataset and wants to treat them to ensure accurate analysis and modeling. They want to detect and handle the outliers using Python. Performed the techniques that provided in the lab to treat the outliers. |
|  |  |

Submitted On:

08-05-2024

(Date: DD/MM/YYYY)

**Task No. 01:** An automotive company wants to identify and treat outliers in their "mtcars" dataset. which contains information about various car models. They suspect that certain car models might have outlier values in certain variables and want to analyze and handle them using Python.

Perform the following outlier detection techniques.

* Box Plot
* Z-score
* IQR range
* Scatter

**Solution:**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from scipy import stats

df = pd.read\_csv('mtcars.csv')

# Display the first few rows

A table with numbers and letters

Description automatically generateddf.head()

**Box Plot**

plt.figure(figsize=(12, 8))

df.boxplot()

plt.title('Box Plot for all numeric columns in mtcars dataset')

plt.show()

A graph with lines and dots

Description automatically generated

# Box plot for a specific column (e.g., 'mpg')

plt.figure(figsize=(8, 6))

sns.boxplot(x=df['mpg'])

plt.title('Box Plot for mpg')

plt.show()

A blue rectangular object with black lines

Description automatically generated

**Z-score**

# Calculate Z-scores

z\_scores = np.abs(stats.zscore(df.select\_dtypes(include=[np.number])))

# Define a threshold for identifying outliers

threshold = 3

# Identify outliers

outliers = np.where(z\_scores > threshold)

outlier\_rows = np.unique(outliers[0])

print(f"Rows with outliers based on Z-score:\n{df.iloc[outlier\_rows]}")

A close up of numbers

Description automatically generated

# Remove outliers based on Z-score

df\_no\_outliers = df[(z\_scores < threshold).all(axis=1)]

print(f"Data without outliers based on Z-score:\n{df\_no\_outliers}")

# Verify removal

plt.figure(figsize=(12, 8))

df\_no\_outliers.boxplot()

plt.title('Box Plot for all numeric columns in mtcars dataset (Outliers removed)')

A screenshot of a computer screen

Description automatically generatedplt.show()

A graph with lines and dots

Description automatically generated

**IQR range**

def find\_outliers\_iqr(data):

Q1 = data.quantile(0.25)

Q3 = data.quantile(0.75)

IQR = Q3 - Q1

outliers = (data < (Q1 - 1.5 \* IQR)) | (data > (Q3 + 1.5 \* IQR))

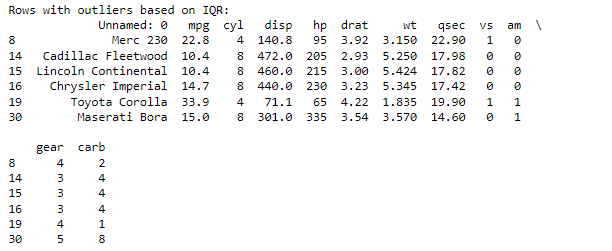
return outliers

numeric\_cols = df.select\_dtypes(include=[np.number])

outliers\_iqr = numeric\_cols.apply(find\_outliers\_iqr)

outlier\_rows\_iqr = df[outliers\_iqr.any(axis=1)]

print(f"Rows with outliers based on IQR:\n{outlier\_rows\_iqr}")



**Scatter**

# Scatter plot for mpg vs. hp (Horsepower)

plt.figure(figsize=(10, 6))

sns.scatterplot(x=df['hp'], y=df['mpg'])

plt.title('Scatter Plot of hp vs mpg')

plt.xlabel('Horsepower (hp)')

plt.ylabel('Miles per Gallon (mpg)')

plt.show()

# Scatter plot for mpg vs. wt (Weight)

plt.figure(figsize=(10, 6))

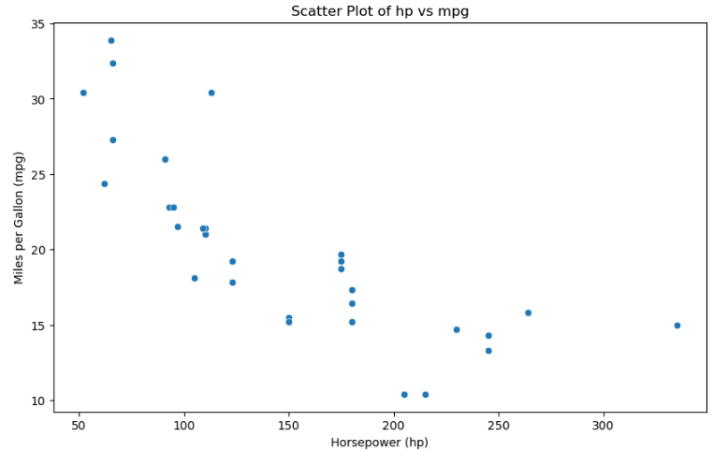
sns.scatterplot(x=df['wt'], y=df['mpg'])

plt.title('Scatter Plot of wt vs mpg')

plt.xlabel('Weight (wt)')

plt.ylabel('Miles per Gallon (mpg)')

plt.show()



A diagram of a scatter plot

Description automatically generated

**Task No. 02:** An automotive company has identified outliers in their "mtcars" dataset and wants to treat them to ensure accurate analysis and modeling. They want to detect and handle the outliers using Python. Performed the techniques that provided in the lab to treat the outliers.

**Solution:**