Outlier-Detection-and-removal using InterQuartile Range(IQR)

Various algorithms like Linear,logistic regressions are Robust to outliers so we need to remove them.Removing outliers is a hectic process , let us remove outliers directly from a given Dataset using InterQuartile Range(IQR).

Implementation :

Import dataset with outliers and observe the upper and lower outlier in it,

```
import pandas as pd
df=pd.read_csv("outlier.csv")
print(df.head())
print(df.tail())
```

Output:

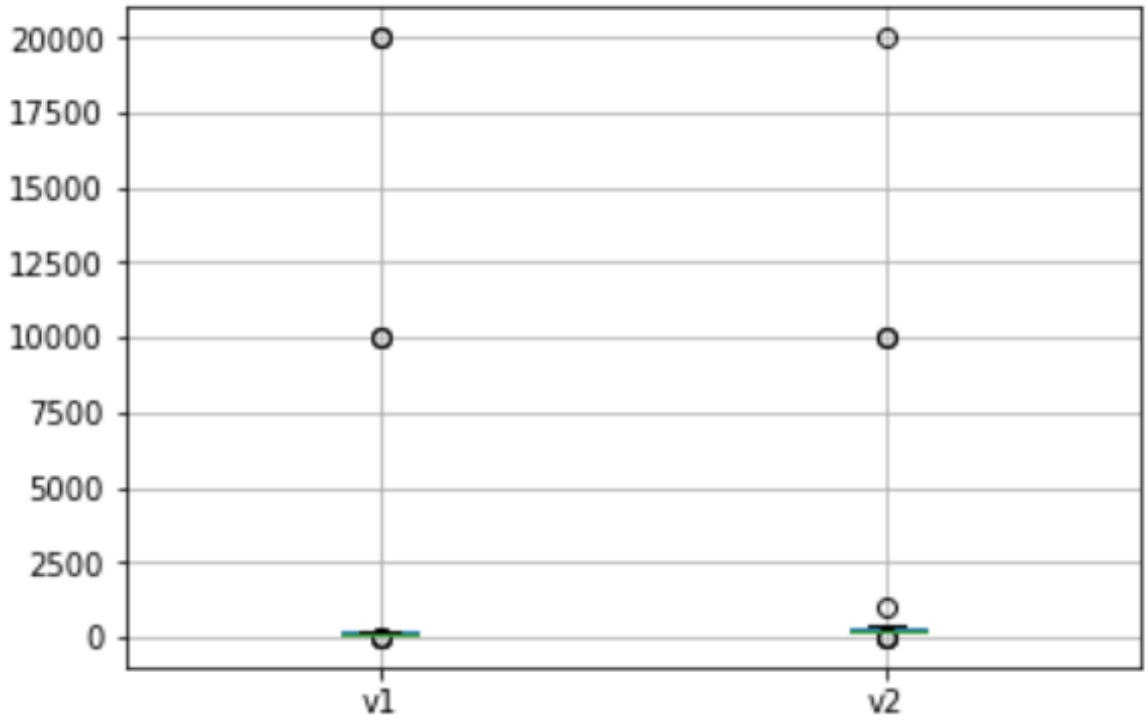
```
     v1   v2
0     1    2
1     2    4
2     3    1
3   120  236
4   120  354
        v1     v2
46     140    235
47   10000  10000
48   20000   1000
49   10000  10000
50   20000  20000
```

Plotting Dataset:

Boxplot shows significant outliers in dataset and can be implemented as follows,

```
#plot dataset
df.boxplot()
```

Output:



This shows significant outliers in the dataset.

Apply IQR:

In this section we remove the outliers from dataset using percentage quantile as follows,

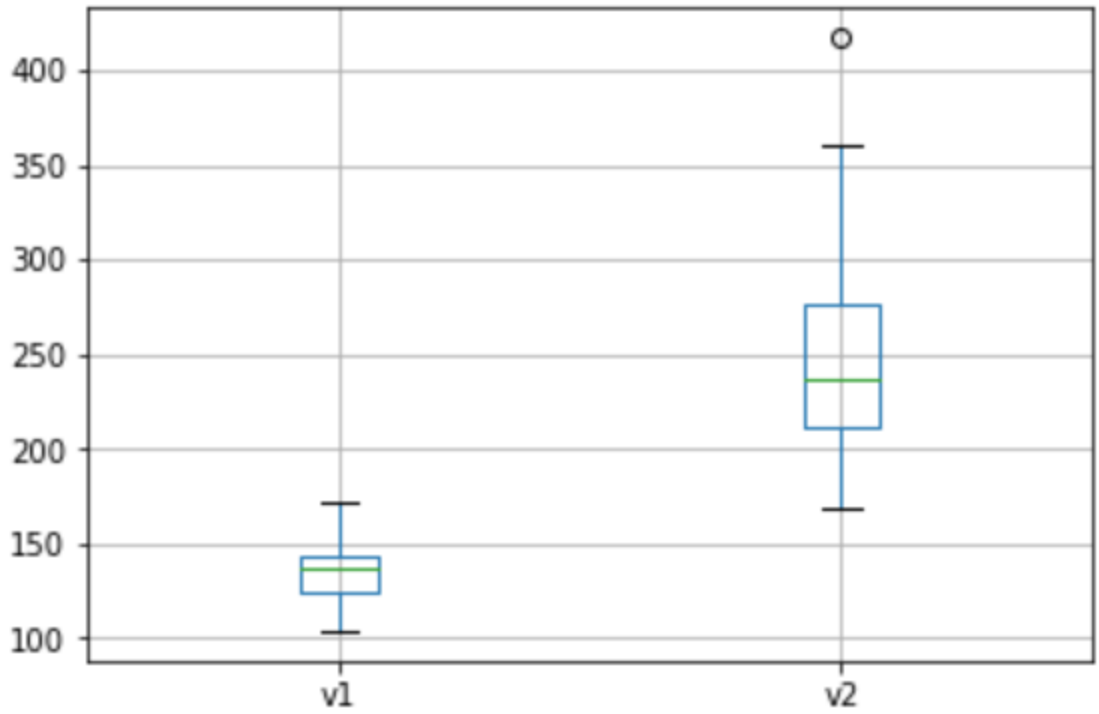
```
Q1=df.quantile(0.25)
Q3=df.quantile(0.75)
IQR=Q3-Q1

#calculate
lower_outlier=Q1-(1.5*IQR)
higher_outlier=Q3+(1.5*IQR)

df_without_significant_outlier = df[~((df < lower_outlier)|(df > higher_outlier))]
```

Plot the result:

In this section we plot dataset after removing dataset as below

```
df_without_significant_outlier.boxplot()
```

Output:



Hence outliers are removed successfully.