

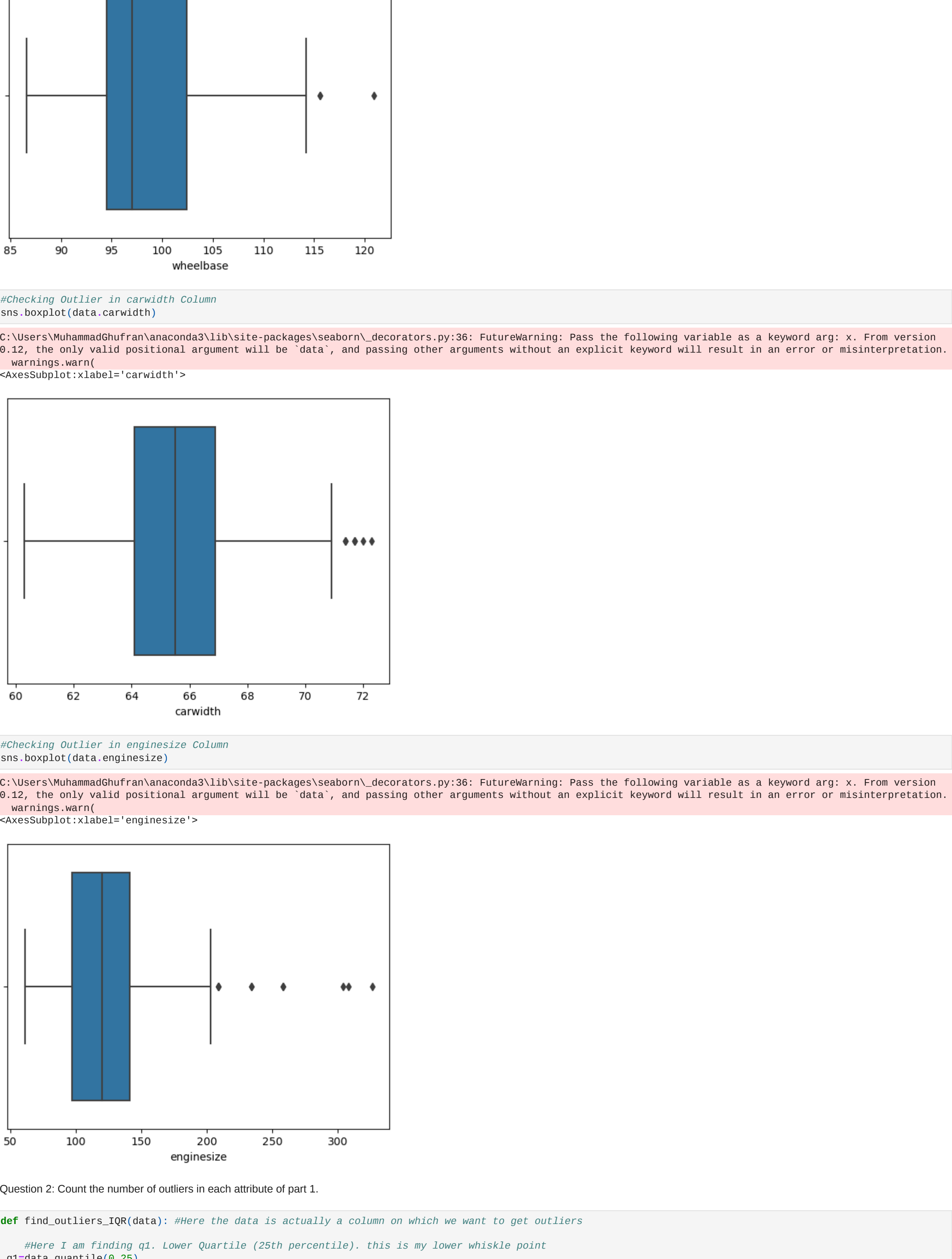
NAME : MUHAMMAD GHUFRAN ASHFAQ

SID:9897

Problem-1

An outlier is a data point which is different from other points in dataset or we can say that data is not normally distributed. For example we have a dataset having attribute age inn which data is distributed between age 10 to 70 but there is few data which is between 80-100 and 1-5 . These range of data will called as outliers .

Outliers affect the ml model in the way that if we are training our model using linear regression it will train the model according to most distributed dataset and ignore the abnormal data



Date: \_\_\_\_\_

## Problem: 3

Question 1:-

$$x = [24, 35, 19, 122, 41, 16, 136, 46, 132, 400, 28, 56, 329, 19, 294]$$

$$x = [16, 19, 19, 24, 28, 35, 41, 46, 56, 122, 132, 136, 294, 329, 400]$$

$$Q_1 = 0.25 \times 15 = 4^{\text{th}} (24)$$

$$Q_2 = 0.5 \times 15 = 8^{\text{th}} (46)$$

$$Q_3 = 0.75 \times 15 = 11.25 = 11^{\text{th}} (132)$$

$$\text{IQR} = Q_3 - Q_1$$

$$= 132 - 24$$

$$\text{IQR} = 108$$

L.B

$$Q_1 - 1.5(\text{IQR})$$

$$24 - 1.5(108)$$

$$-138$$

U.B

$$Q_3 + 1.5(\text{IQR})$$

$$132 + 1.5(108)$$

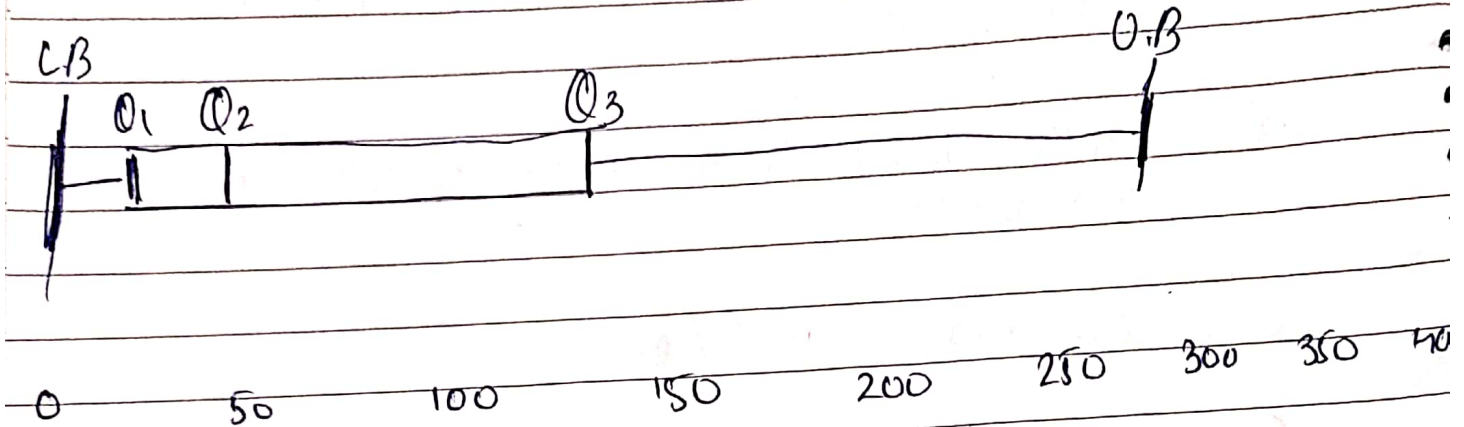
$$294$$



Date: \_\_\_\_\_

$(-138, 294)$  ~~is~~  $\rightarrow$  whisker point

## Part 2



## Q4 Part 4:-

As we have have whisker points are

$(-138, 294)$

So greater than 294 and smaller than -138 are outliers

outliers = 329, 400, these

## Part 5

There are three methods

- 1) The Imputation (Mean, Median, Mode)
- 2) Remove the outliers
- 3) Quantile based flooring and capping.

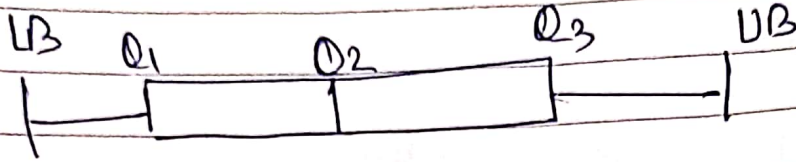
## Part 6

According to the boxplot we can see that Quartile 2 or is on the right ~~most~~ most side of plot so we can say that data is right skewed.

Skewness  $\rightarrow$  right

Date: \_\_\_\_\_

### Problem 2:-



In boxplot we have basically a three quartile.  $Q_1$  is the median of 50% of quartile  $Q_1$  is 25% of dataset and  $Q_3$  is 75% of dataset. In boxplot we find L.B and U.B using these quartile. Any value out of it. If any value is less than L.B and greater than U.B will count as outliers.