team members’

Names ID Department

Ahmed Hatem Abdelaziz Abdelftah 2018170510 SC

Ahmed Hamdy Abdelhalim 2018170511 SC

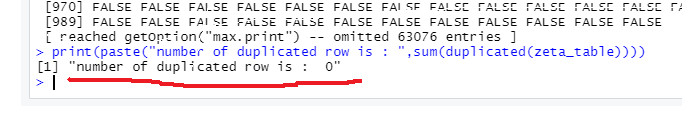
Omar Mohamed farhat 2018170264 SC

How many rows of data are there in the zeta table?

64076 rows.

Are there any duplicate rows of data in the zeta table? If so, how can you tell?

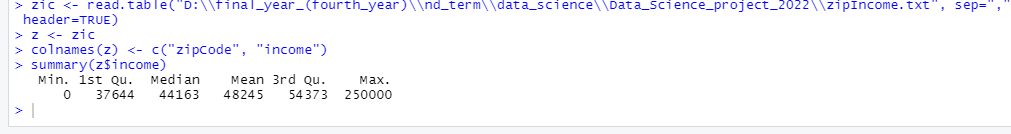
No duplicate in rows.



Analyze the summary of your data. What are the mean and median average incomes?

Median = 44163.

Mean = 48245.



do you see any outlier values? If so, what are they?

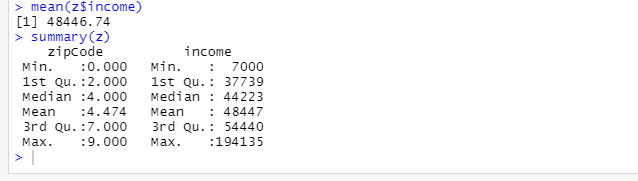
Yes, data have outlier.



What’s your new mean?

Median: 44223.

Mean: 48447.



How has the clustering changed after log10 scale? Why?

Number of clustering changed from 2 to 5 because numbers are close to each other and getting smaller.

Re-evaluate the choice of K

1. Before log 10

Re-evaluate the choice of K = 2 because there is large change around this valueGraphical user interface, histogram

Description automatically generated with medium confidence

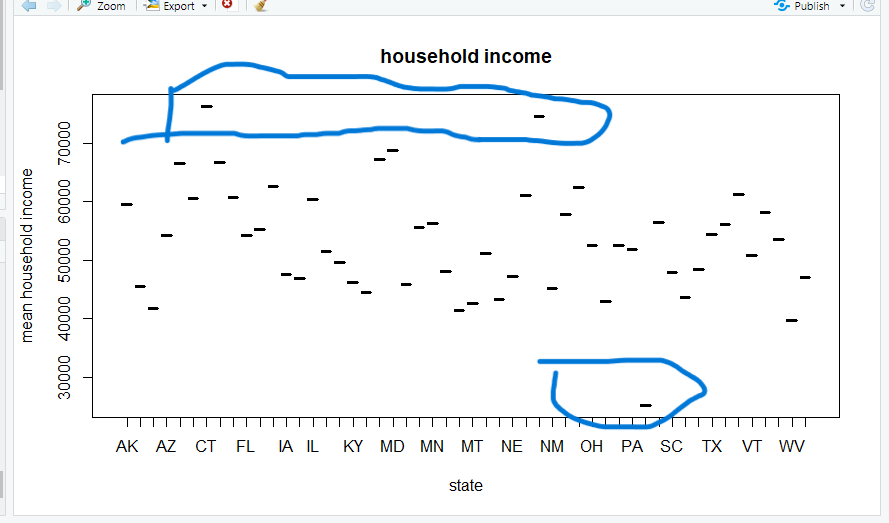
1. After log 10

Re-evaluate the choice of K = 5 because there is large change around this value

Diagram

Description automatically generated with low confidence

Have you observed an outlier in the data?

Yes, data have outlier 

number of cluster become 2