**Code book**

test: the data of test

train: the data of train

colname: the names of columns

Q1\_df: dataset for question 1

Q2\_df:dataset for question 2

Q3\_df:dataset for question 3

Q4\_df: dataset for question 4

Q5\_df\_mean: dataset for question 5