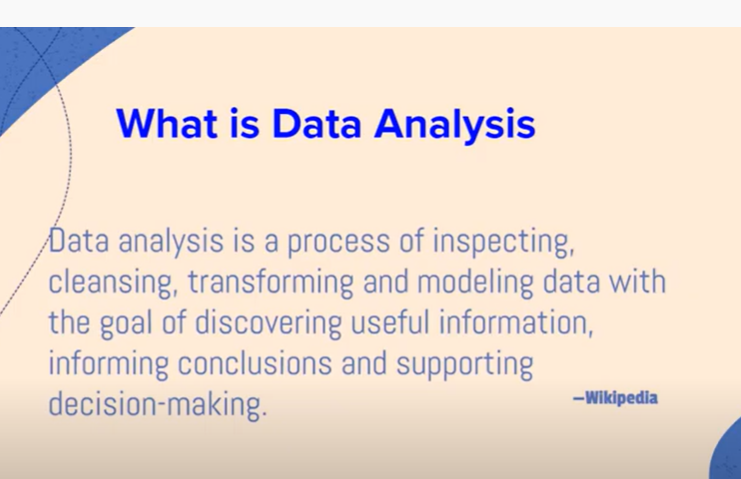
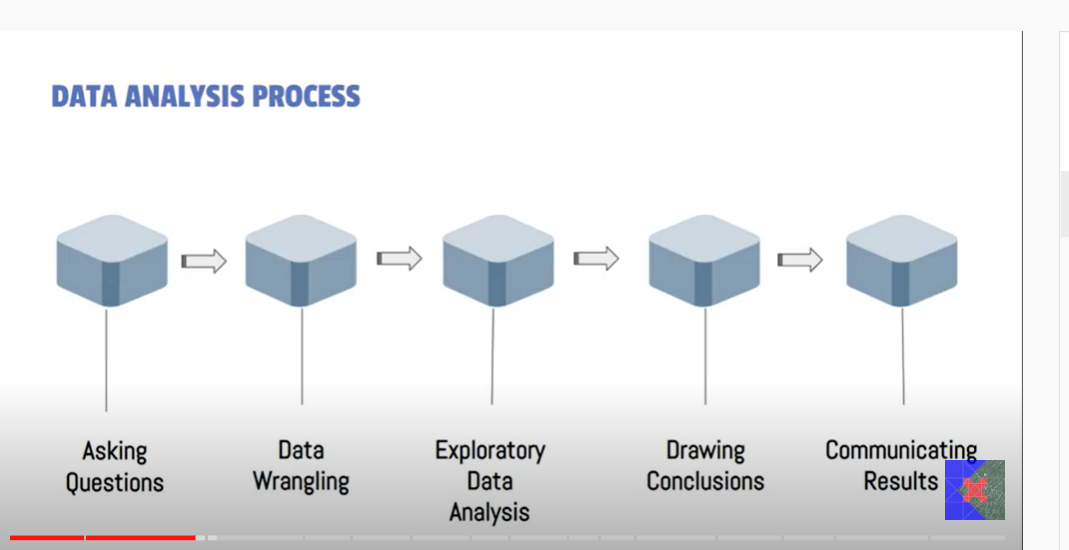
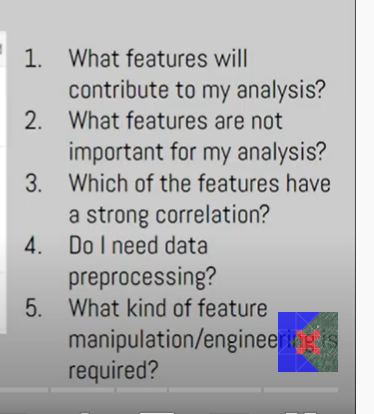
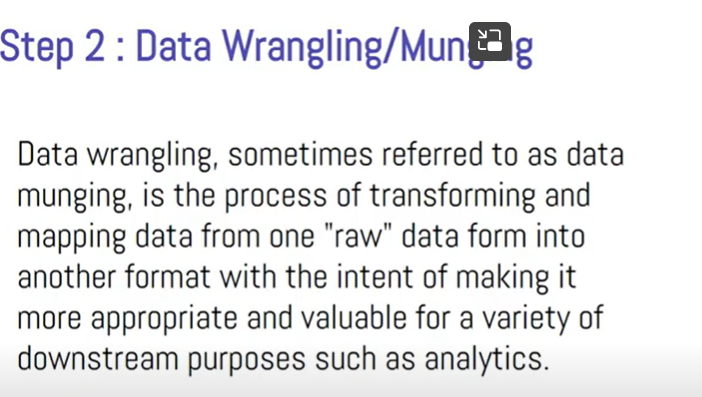
Data Analysis:









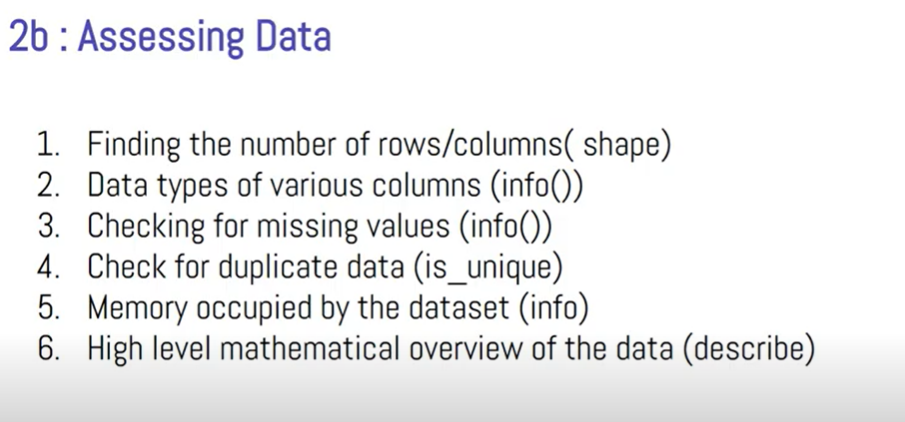
Raw form to another form

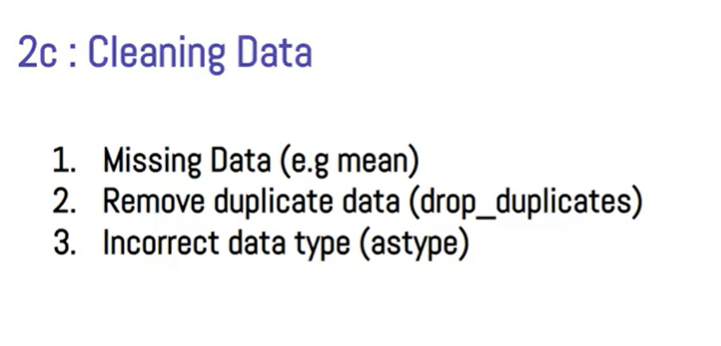
Gathering data

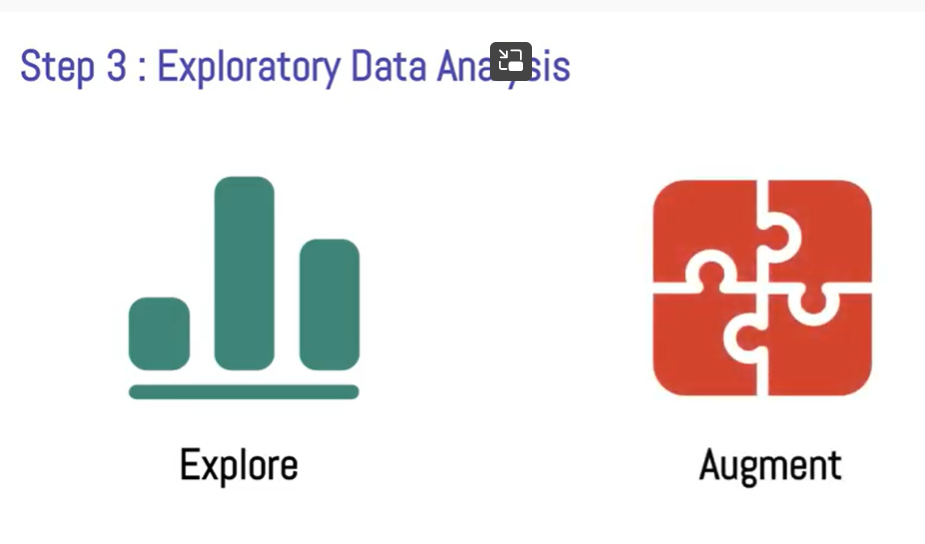
Assessing data

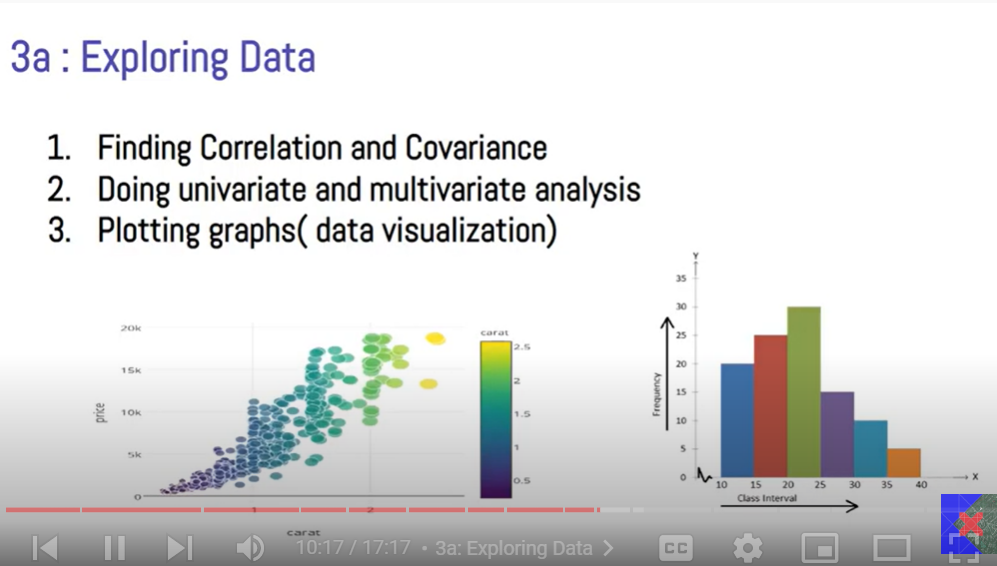
Cleaning data

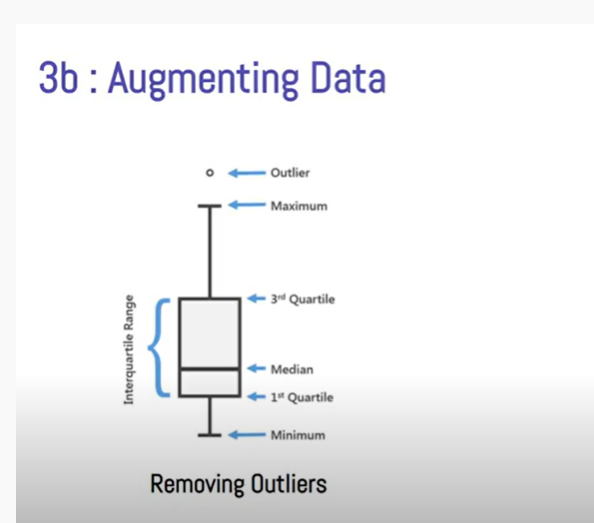




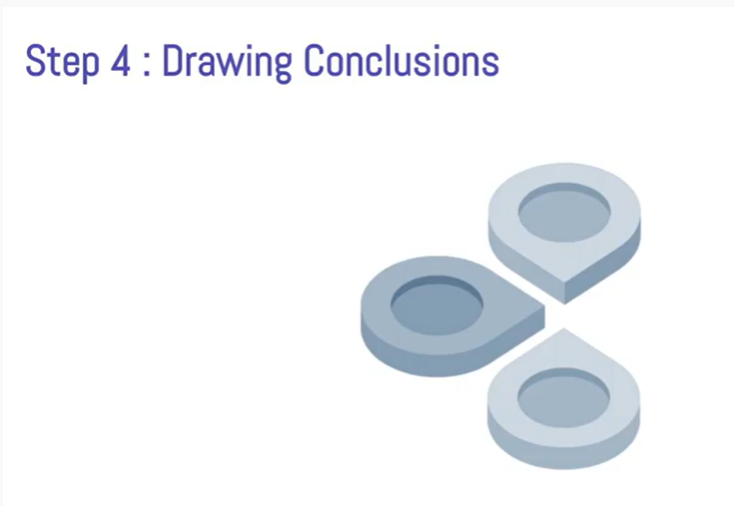


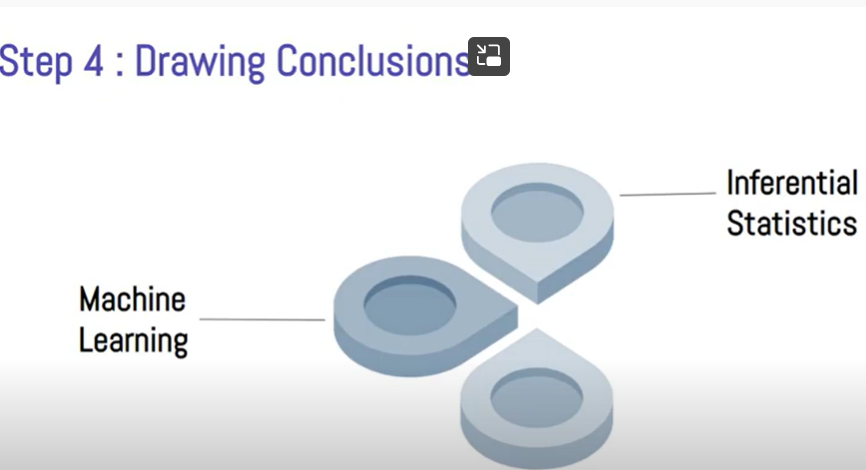


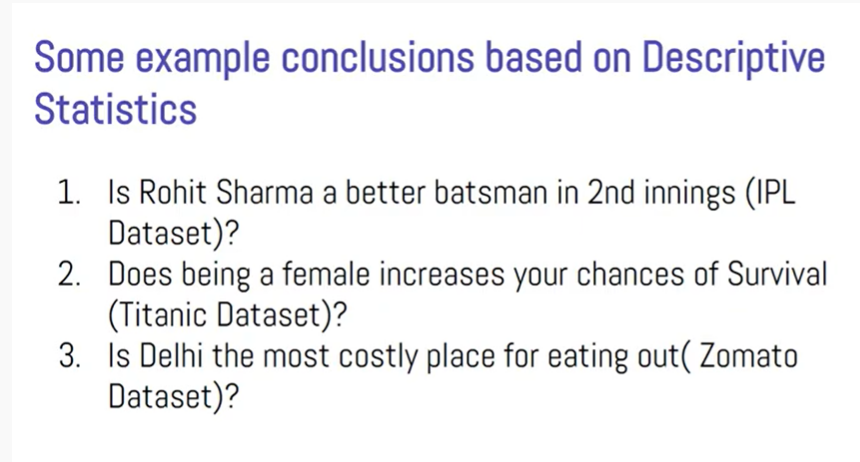


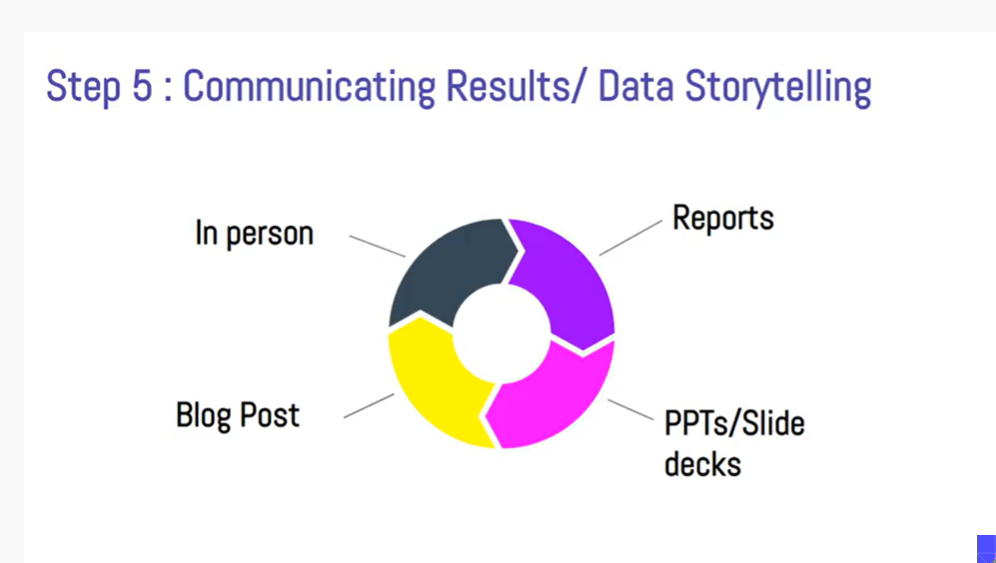




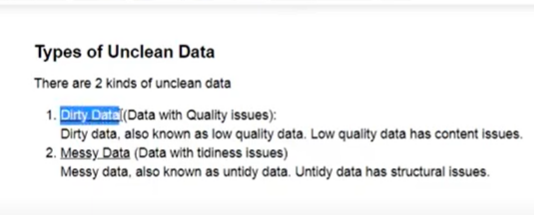


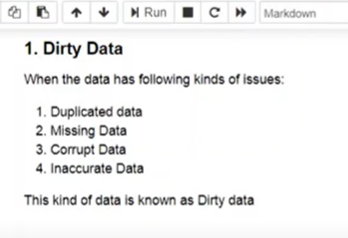


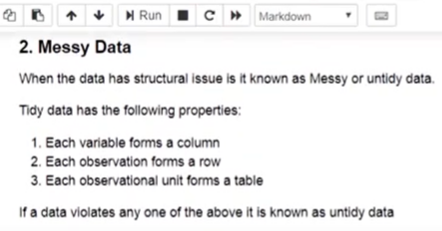


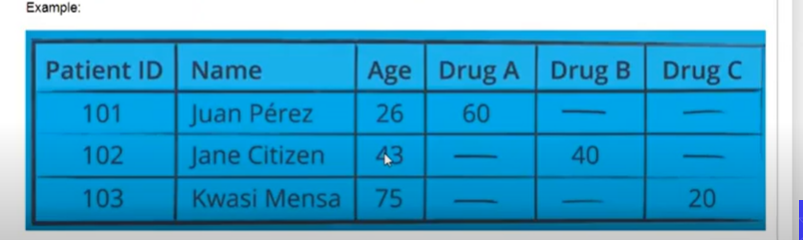


Assessing data

’

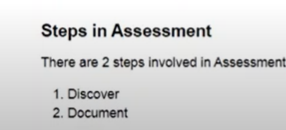


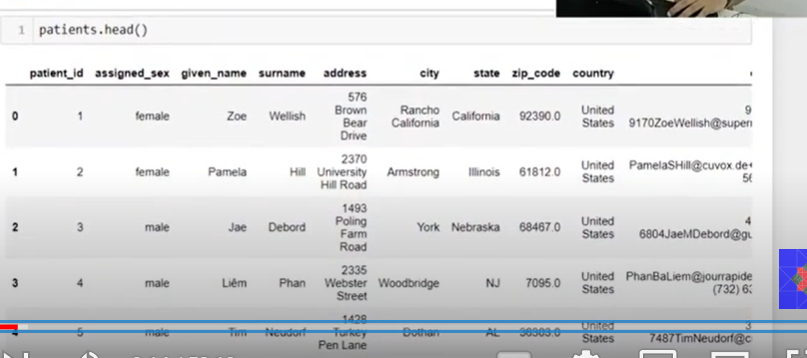




Structural issue because 3 drug columns makes data sparse



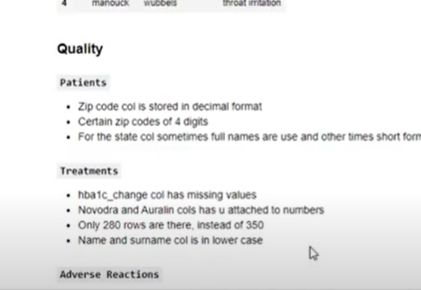








Manual assessments



Programmatic assessments

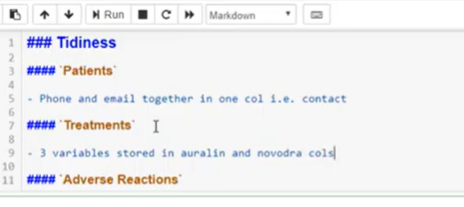
Patient.info()-

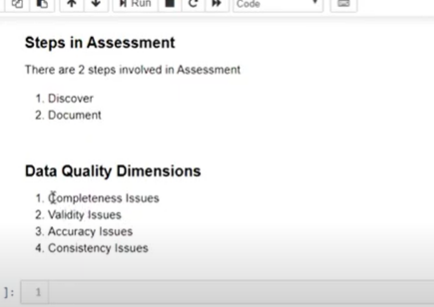
Not null, data type, memory usage

Describe

Df.sample(5)--- random 5 sample

Df.duplicated





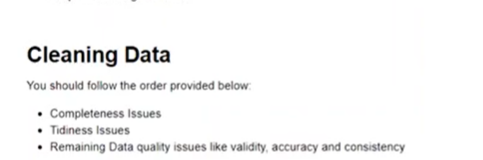
Completeness – null

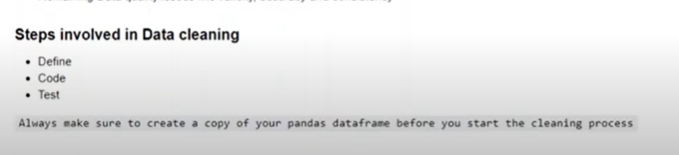
Validity- height negative

Accurate-height non negative but 1 cm

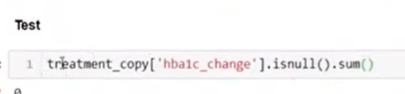
Constancy : one place newyork and one place uk

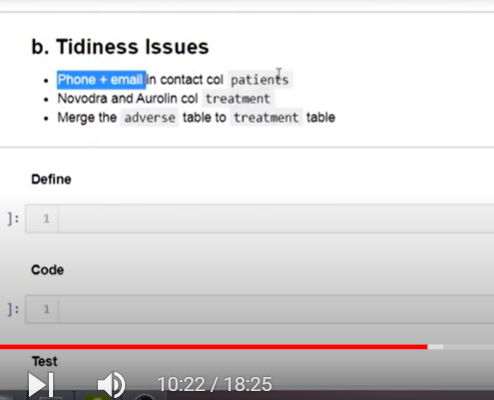
**Data Cleaning | Data Wrangling Step**

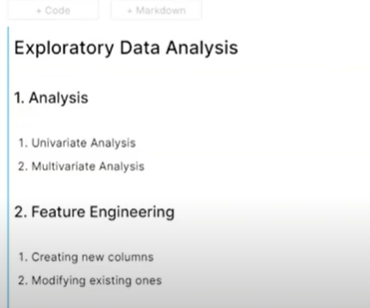


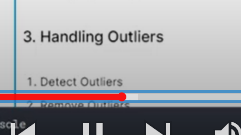


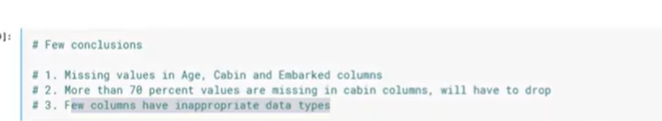


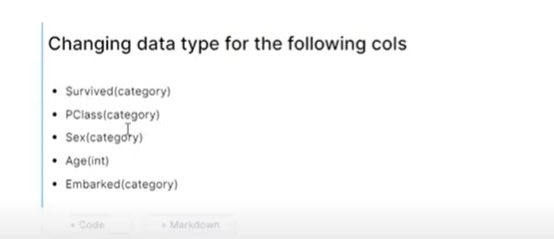


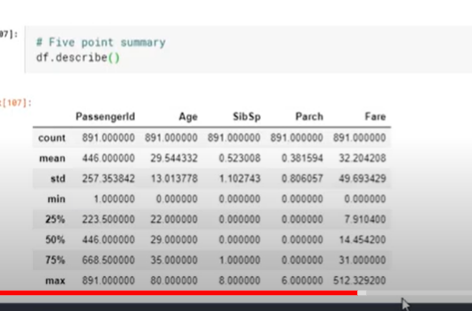




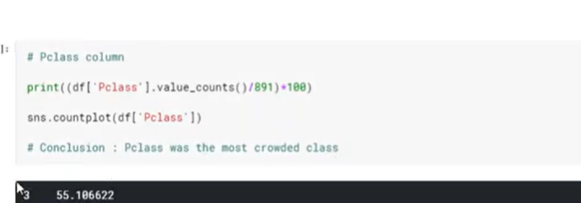


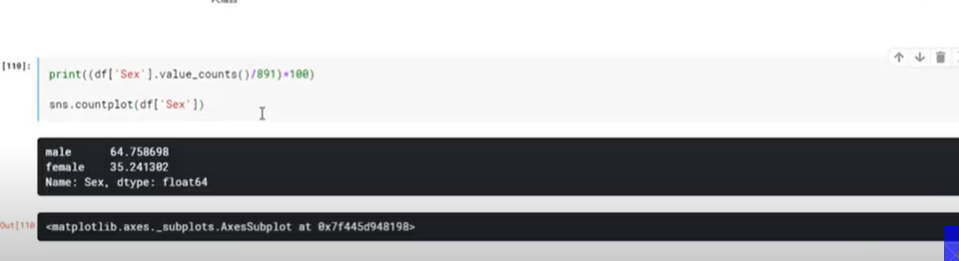


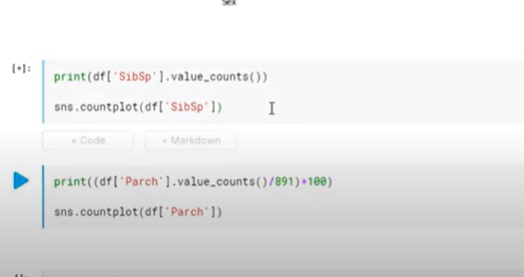


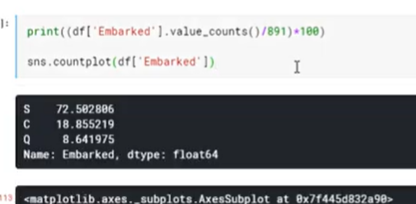




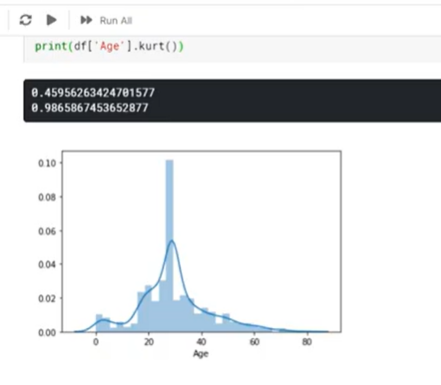








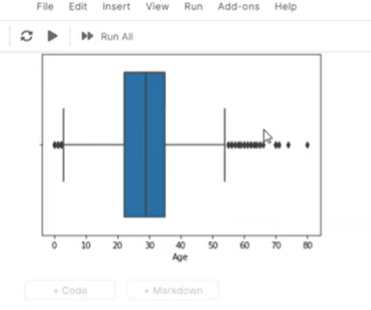




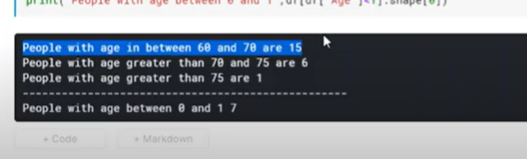
Skew=-5 to +5 – normally distributed

Kurtosis –











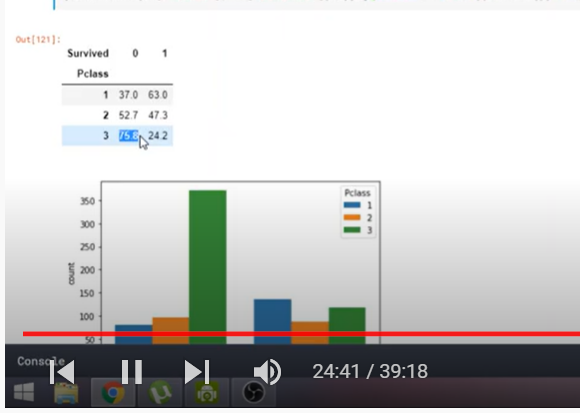




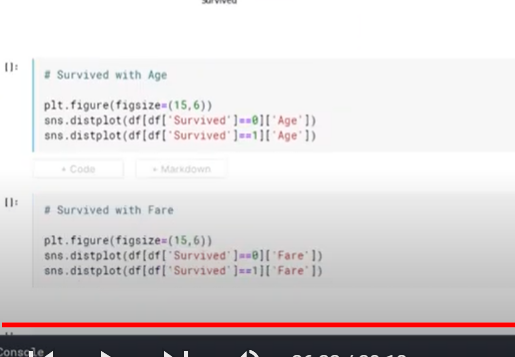




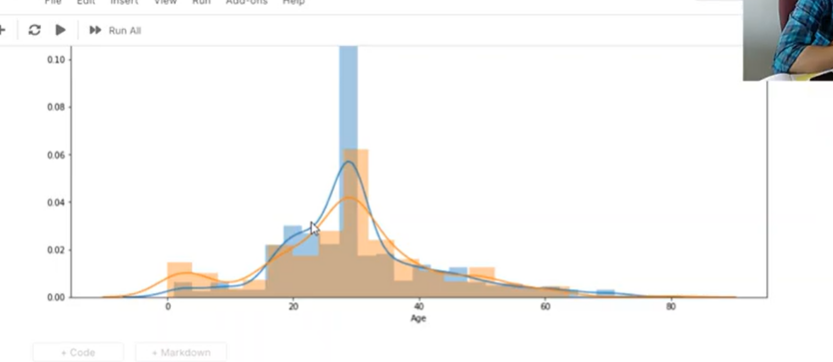








Survival with age



Orange : not died

Blue : not survived

Lower agr: probability of death is lower than survival

Middle age : probability of death is high than death

Higher age:

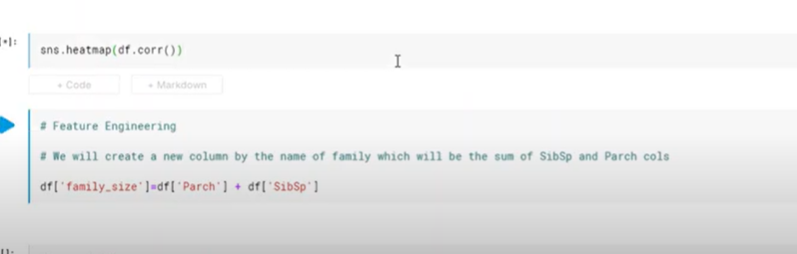
Survived with fare



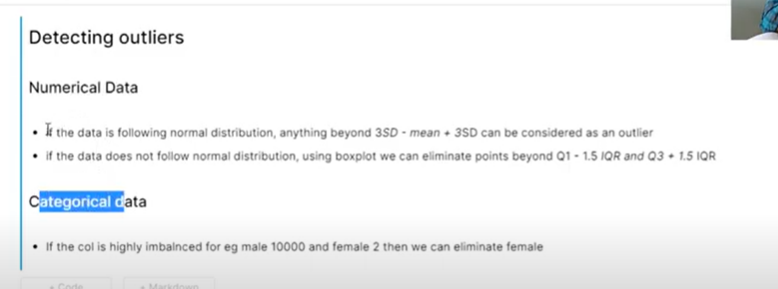
Less price – high probability of death

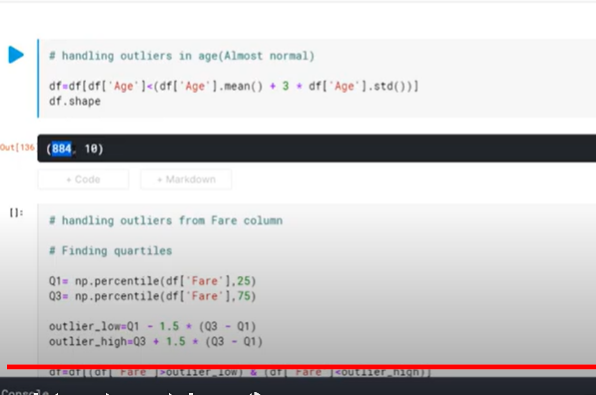
Pair plot and treemap













Corr

Heatmap

Drawing conclusion

