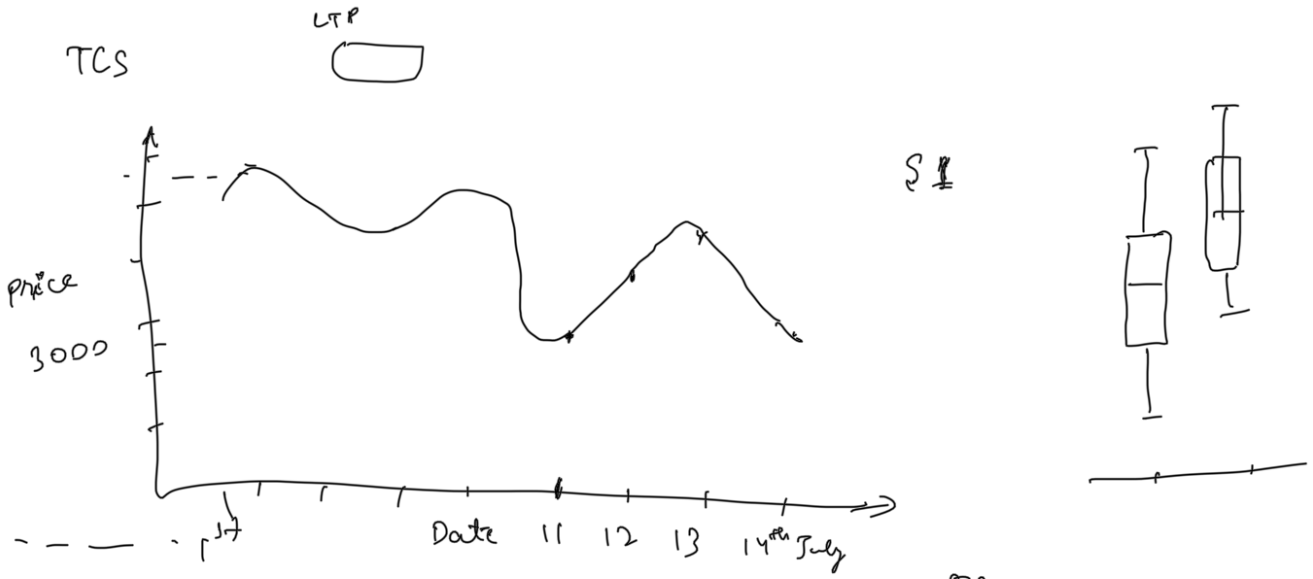
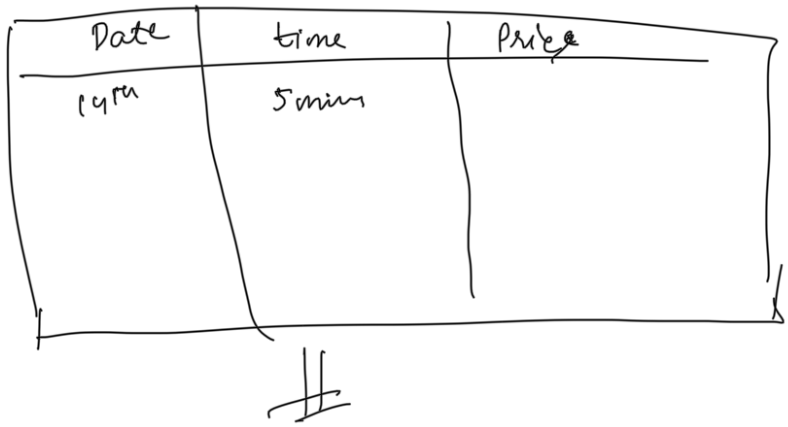
Data Visualisation 1

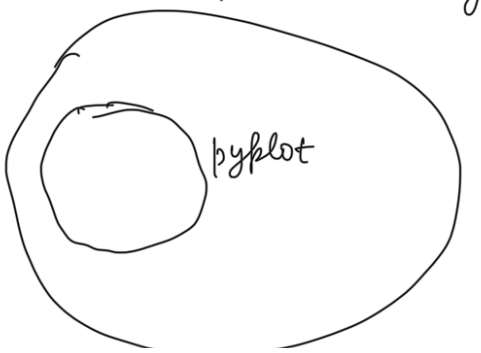
- Categorical: Platform, Genre
- Continuous: (sales, Global-sales ...) Rank
  - Nominal (Platform, Genre) (You can't order diff. possible values)
  - Ordinal (star rating; (low, medium, high, very high)
    1 2 3 4





Matplotlib

Matplotlib = library



matplotlib.pyplot.

How to decide which graph we want:

(1) How many variables

variable = 1 ⇒ Univariate Analysis

variables = 2 Bi variate Analysis

variables = 3 or more Multi-variate Analysis

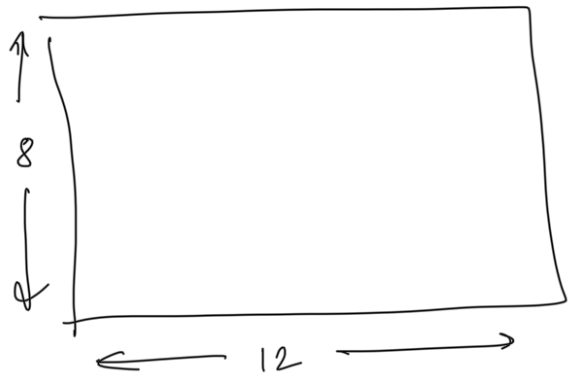
Univariate Analysis

(1) Numerical
(2) Categorical

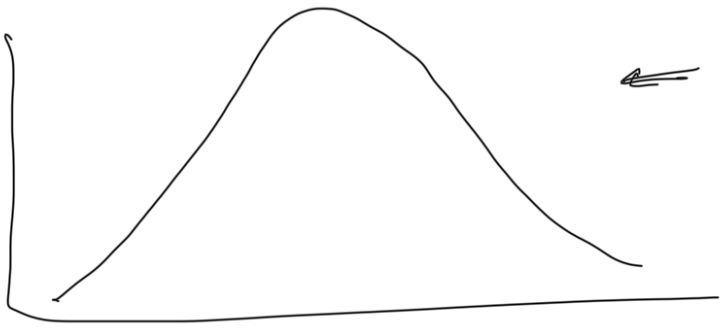
Bivariate Analysis:

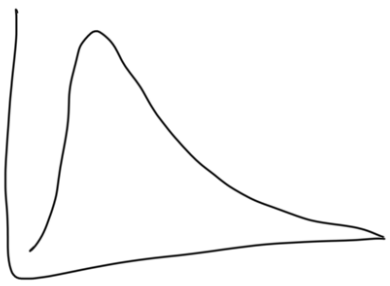
1) N C (v1 v2) (2) C C (v1 v2) (3) N N. (v1 v2)

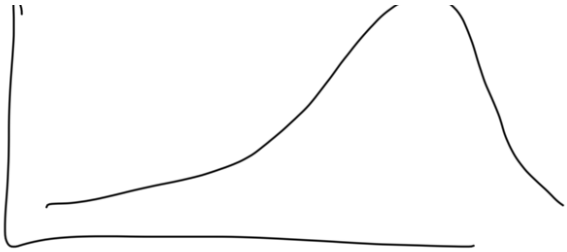
Univariate / Categorical



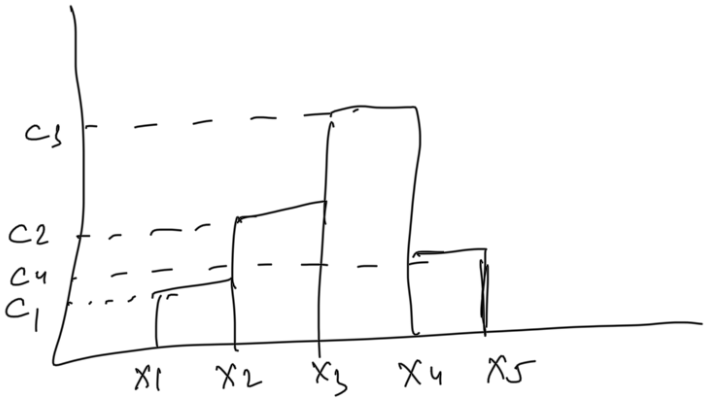
Histogram:







Barplot:



bins = 4

plt.hist(data["year"], bins = 4)

count:

Earray(c1, c2, c3, c4)

bins array(x_1, x_2, x_3, x_4, x_5)

count-

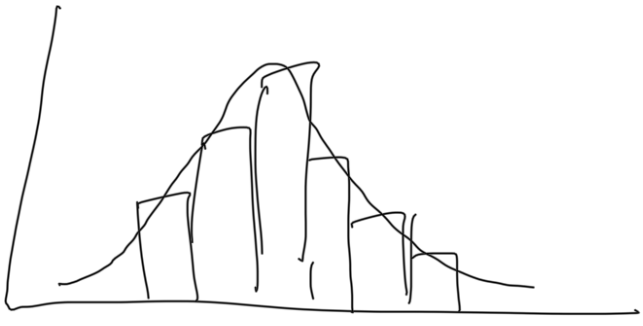
bin 1: $x_1 - x_2 = c_1$

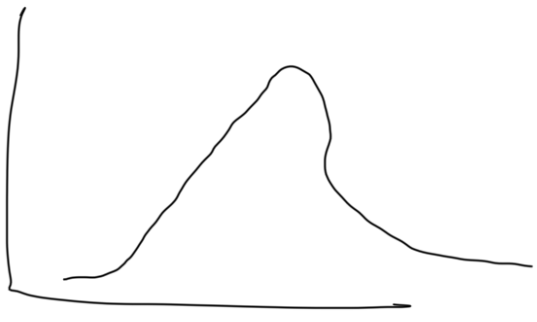
2 : $x_2 - x_1 = c_2$

3 : $x_3 - x_4 = c_3$

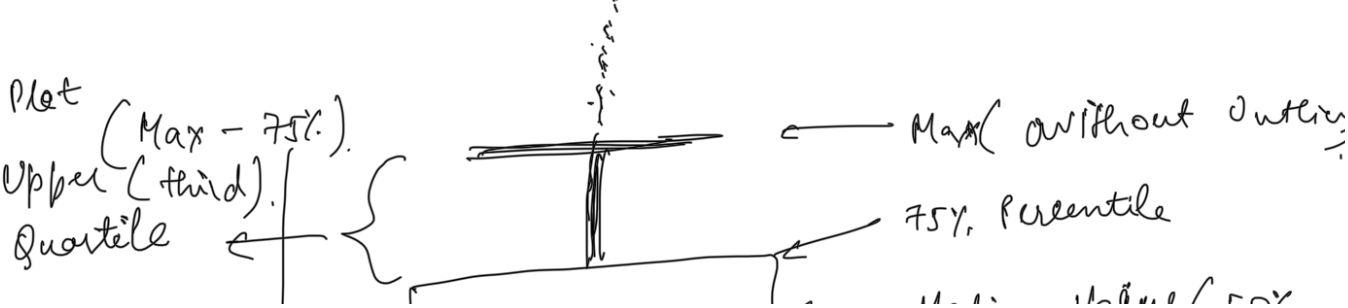
4 : $x_4 - x_5 = c_4$

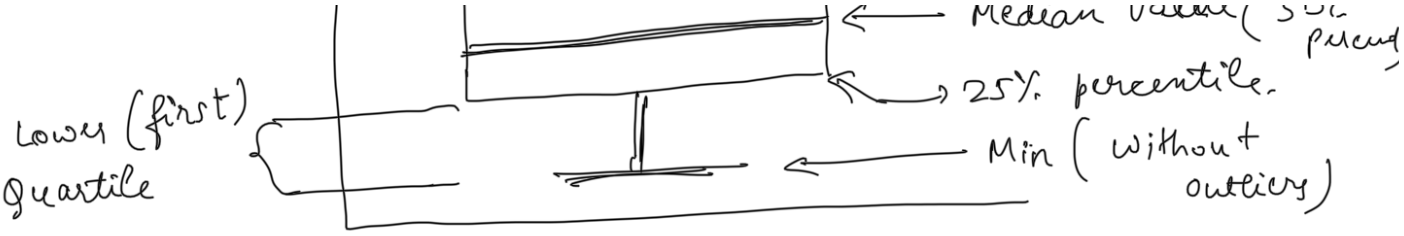
KDE



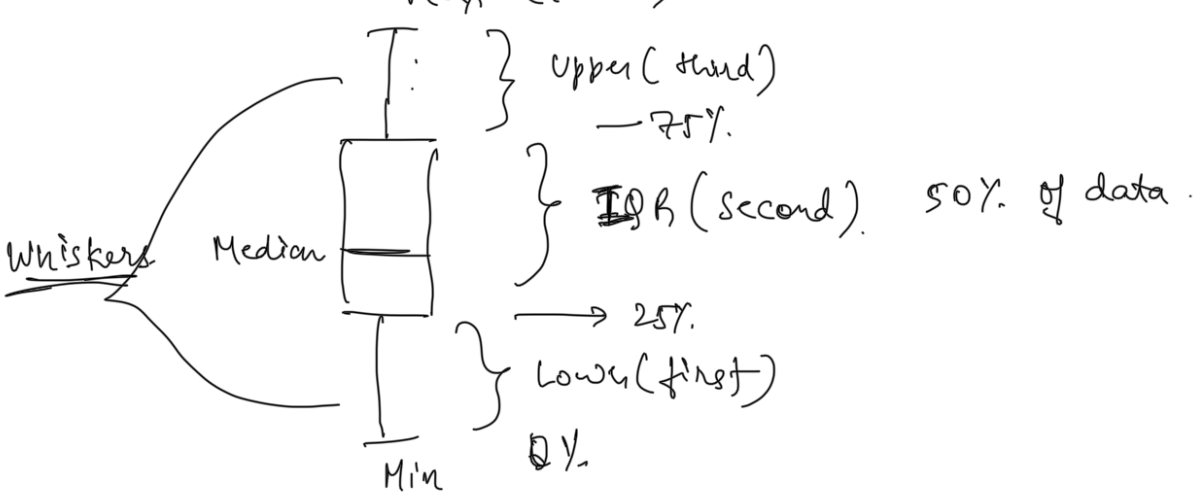


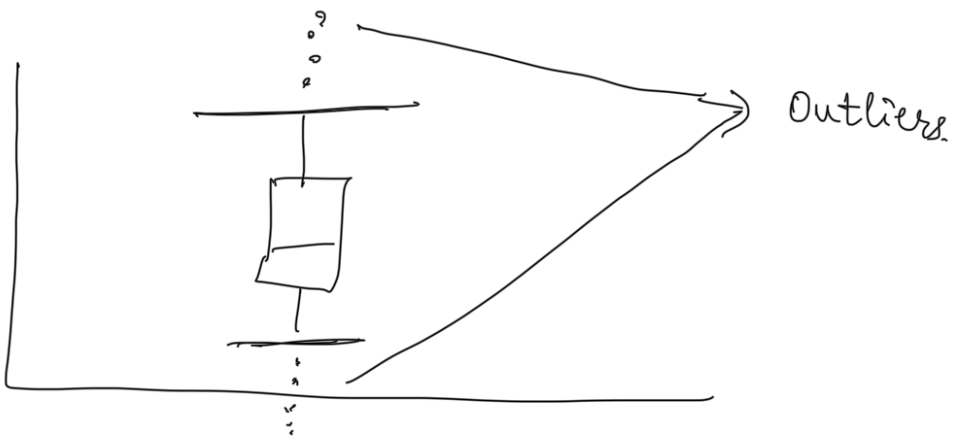
Box Plot




Median value (50% percentile)
25% percentile.
Min (without outliers)
Lower (first) Quartile

Min → 25%.


Max (100%)
Upper (third)
75%.
IQR (second) 50% of data.
Whiskers
Median
25%.
Lower (first)
Min
0%.


Outliers