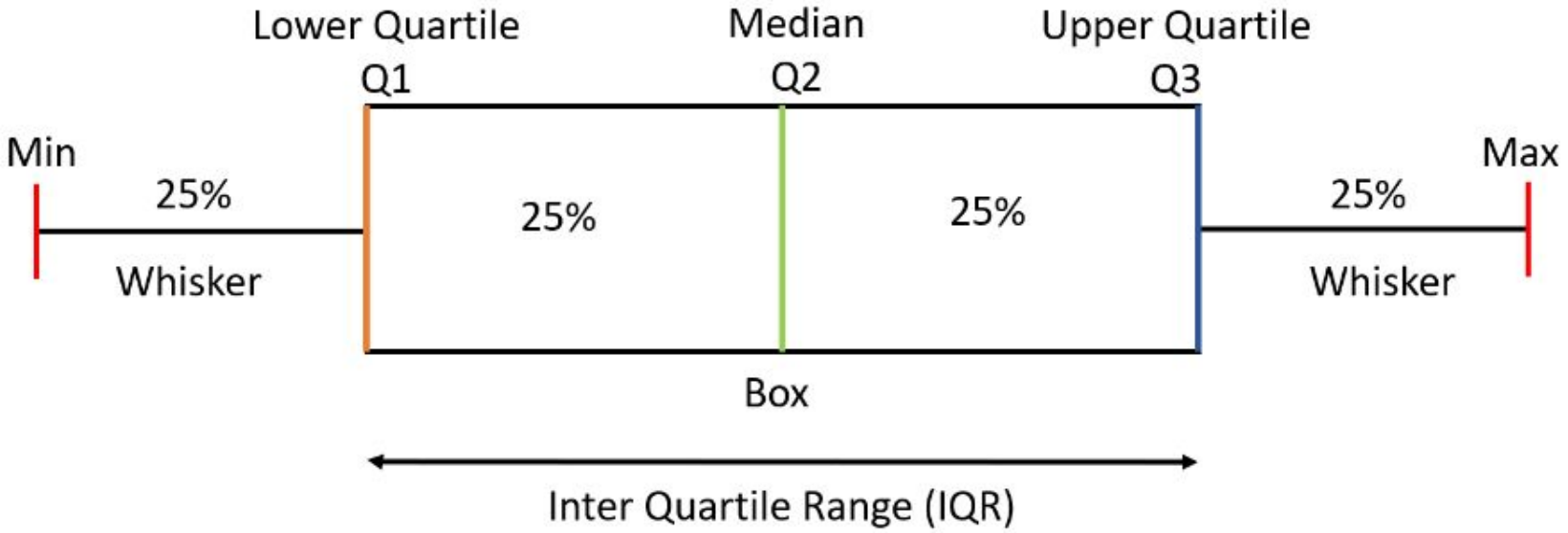
Module-5

EDA Graphs Details Explained

Boxplots


Lower Quartile
Q1
Median
Q2
Upper Quartile
Q3
Min
Max
25%
25%
25%
25%
Whisker
Whisker
Box
Inter Quartile Range (IQR)

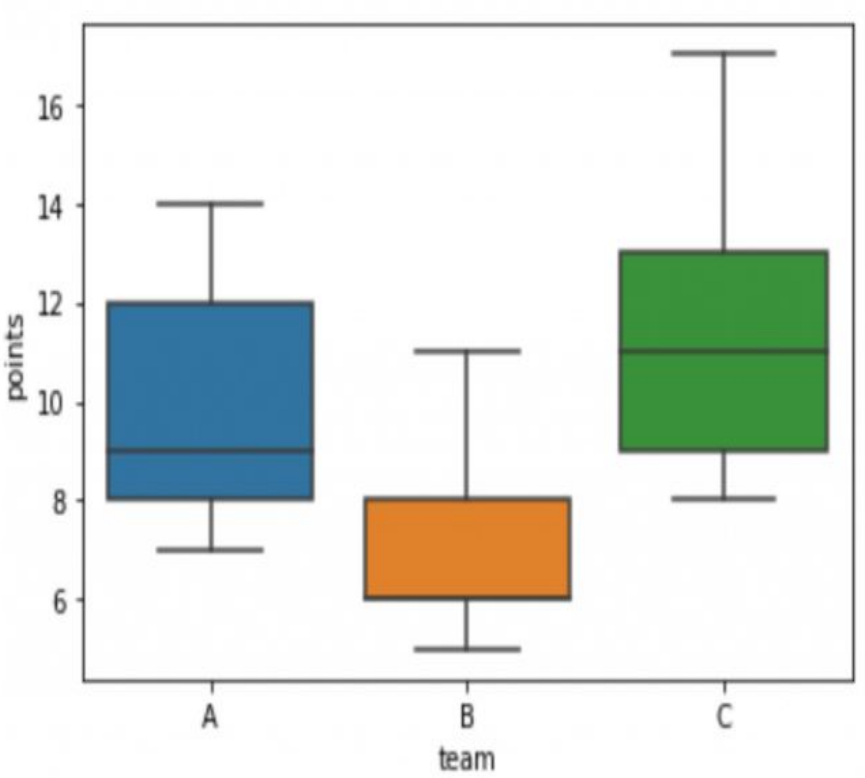
Boxplots also known as Whisker plot

The area inside the box (50% of the data) is known as the **Inter Quartile Range.** IQR = Q3-Q1

Outliers are the data points **below and above** the **lower and upper limit**.

Lower Limit = Q1 - 1.5*IQR

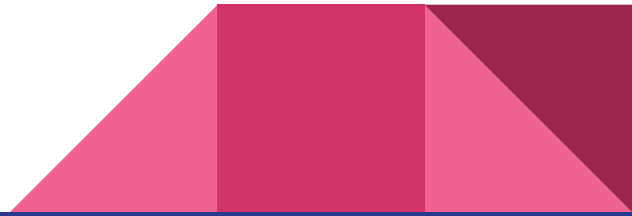
Upper Limit = Q3 + 1.5*IQR



Uses of a Box Plot

Box plots provide a visual summary of the data with which we can

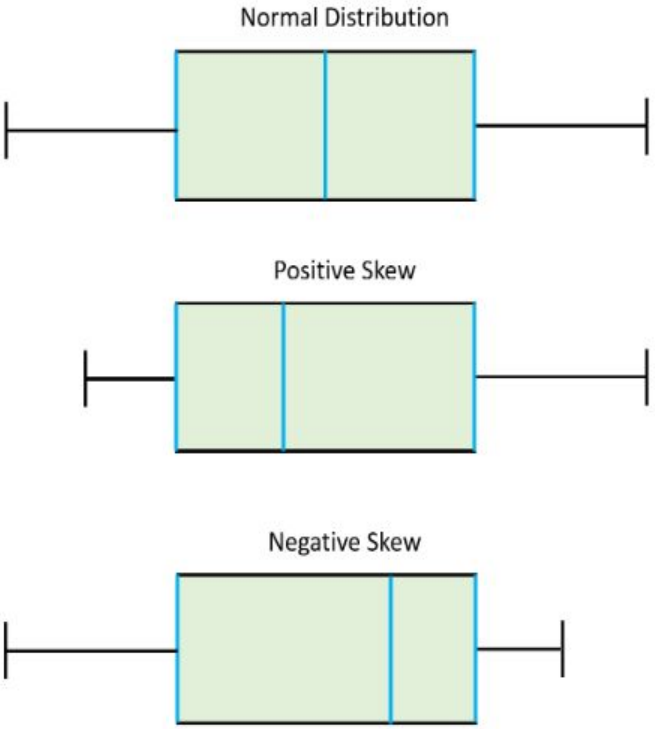
- quickly identify the average value of the data (Median),
- how dispersed the data is,
- whether the data is skewed or not (skewness).



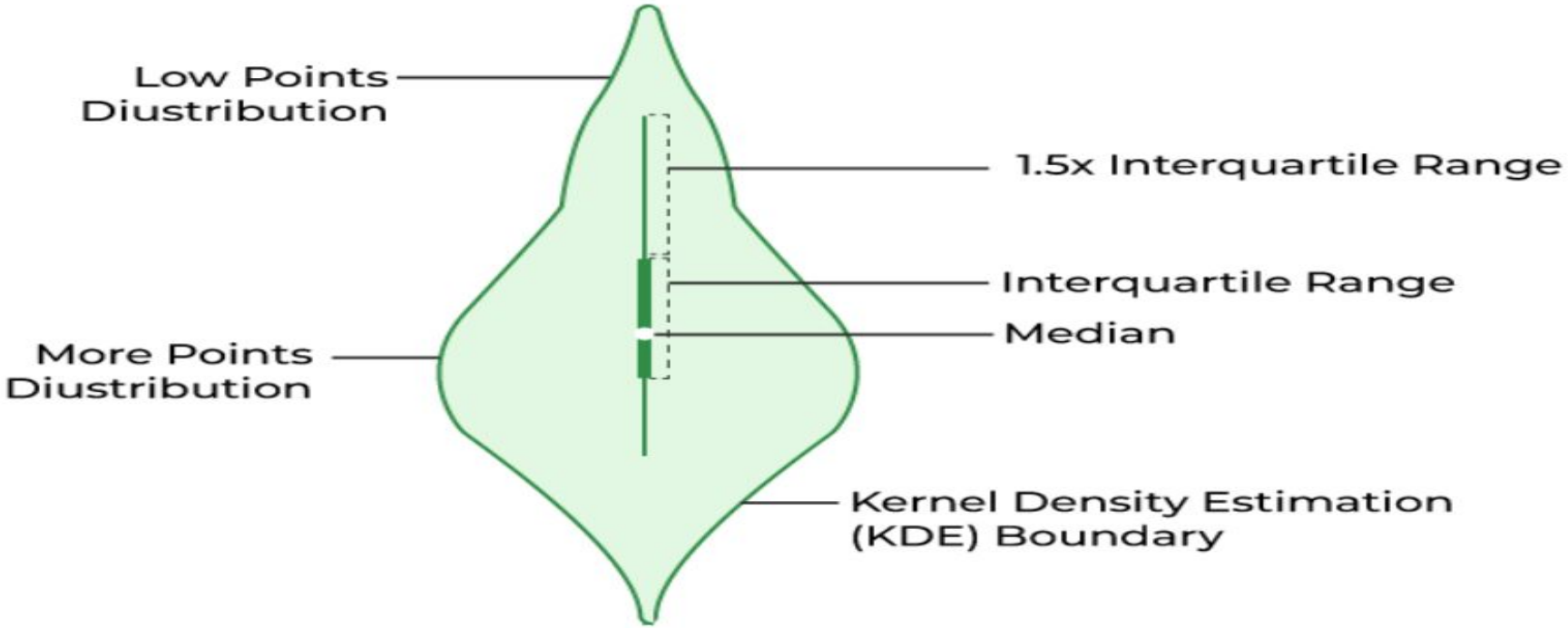
a) If the Median is at the **center** of the Box and the **whiskers** are almost the **same on both the ends** then the data is **Normally Distributed**.

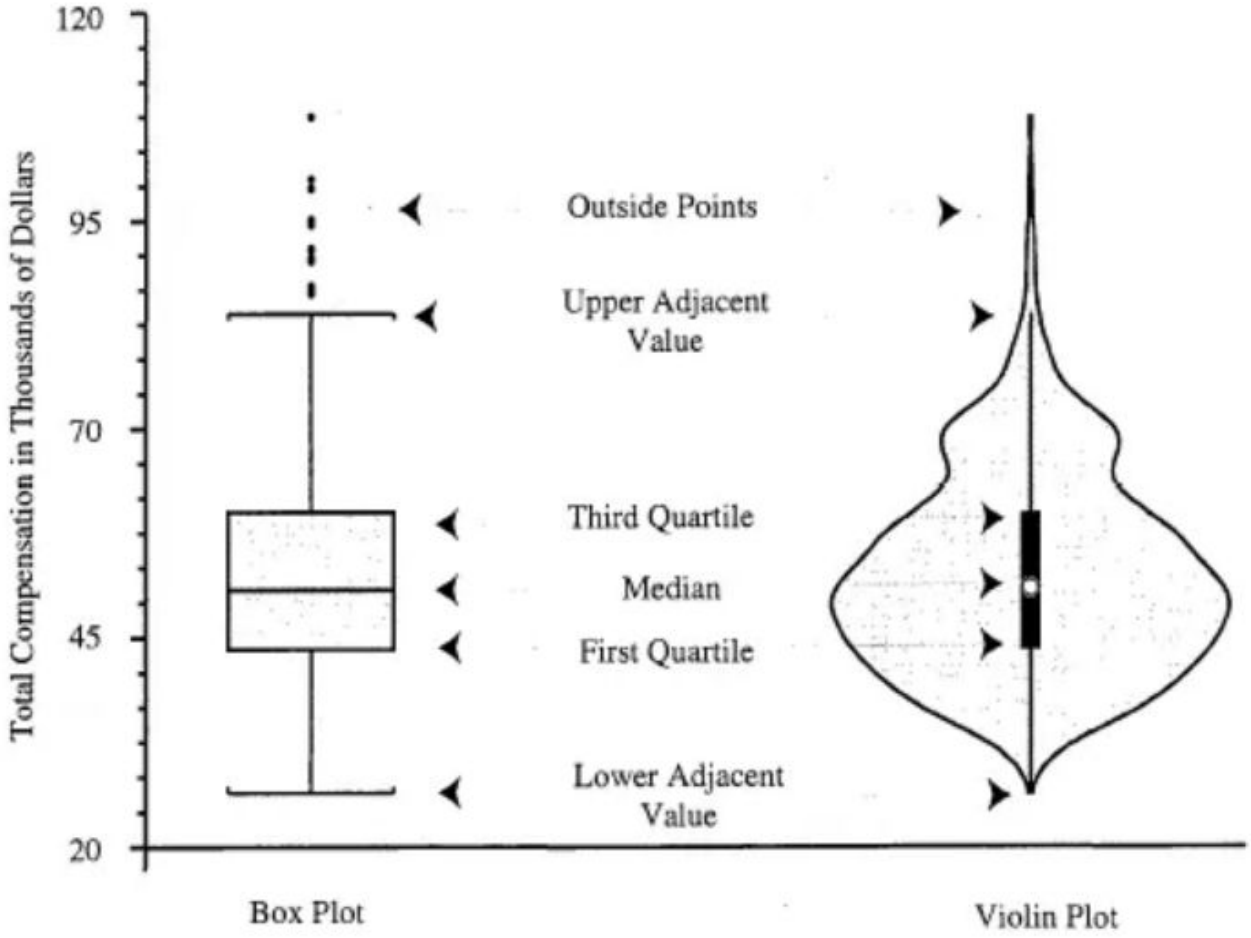
b) If the Median lies **closer to the First Quartile** and if the **whisker at the lowerend is shorter** (as in the above example) then it has a **Positive Skew (Right Skew)**.

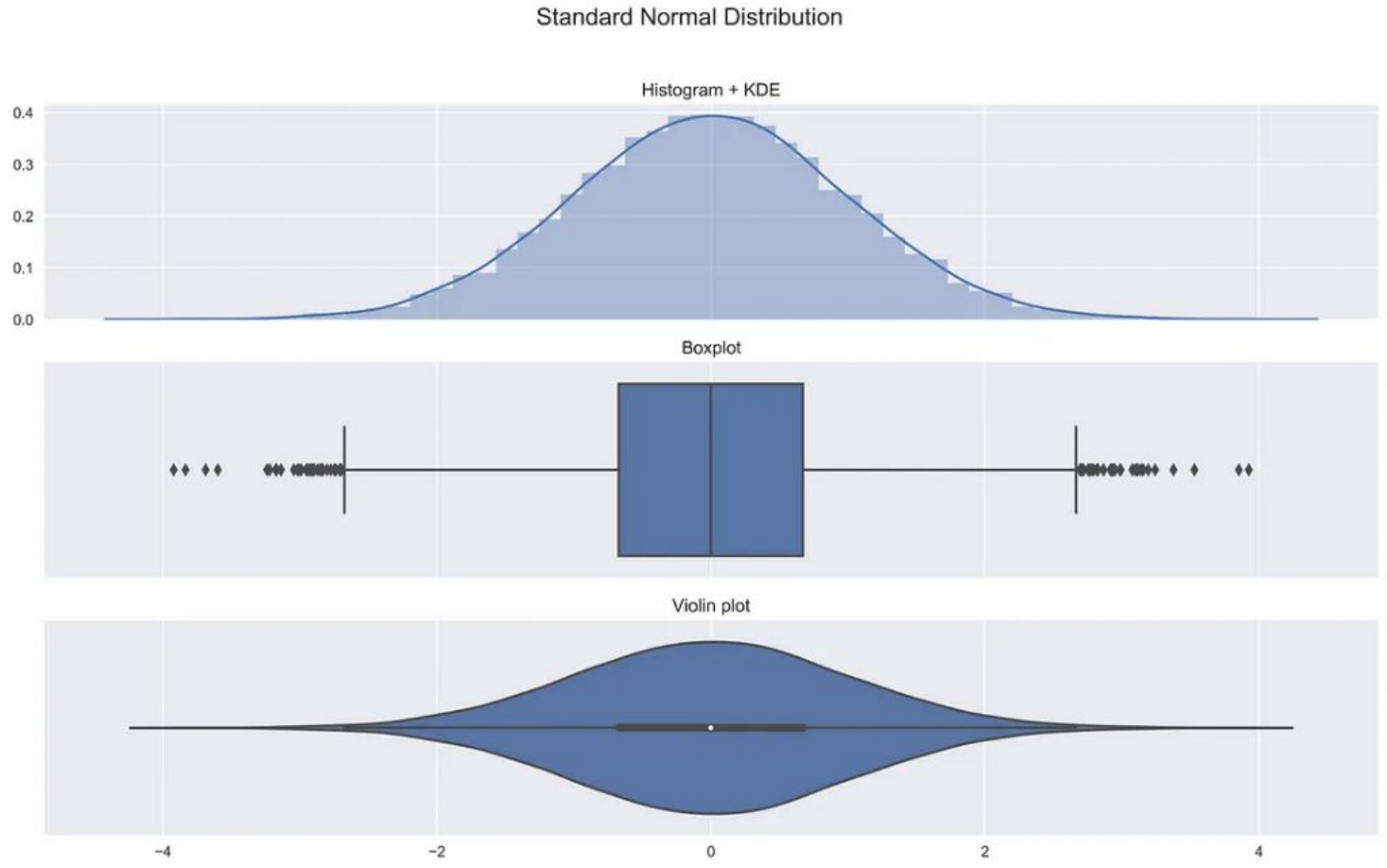
c) If the Median lies **closer to the Third Quartile** and if the **whisker at the upper end is shorter** then it has a **Negative Skew (Left Skew).**



Violin Plots




120
95
70
45
20
Total Compensation in Thousands of Dollars
Outside Points
Upper Adjacent Value
Third Quartile
Median
First Quartile
Lower Adjacent Value
Box Plot
Violin Plot


Standard Normal Distribution
Histogram + KDE
0.4
0.3
0.2
0.1
0.0
Boxplot
Violin plot
−4
−2
0
2
4

THANK YOU