upGrad and
IIITB Machine
Learning & AI
Program - July   **Learn**   **Live**   **Discussions**   **2**
2024   Navigate   Q&A

Summary

You may decide to go through the solutions to the previous segment of practice problems before we summarize the learnings from the session.
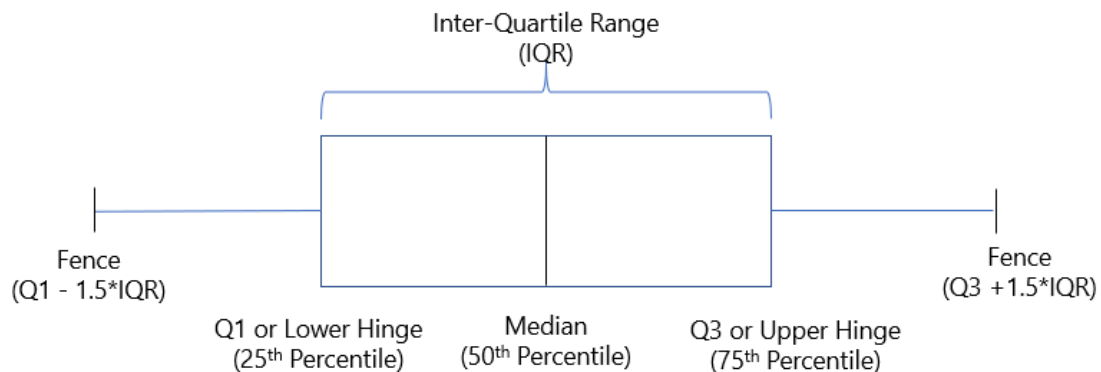
Here's a summary of what you have learnt so far in this very intense session:

- First, you began with understanding the necessity of visualisation using the example of Anscombe's Quartet. There, you saw four different data sets having the same exact summary statistics but completely different visual plots.

upGrad and
IIITB Machine
Learning & AI
Program - July
2024

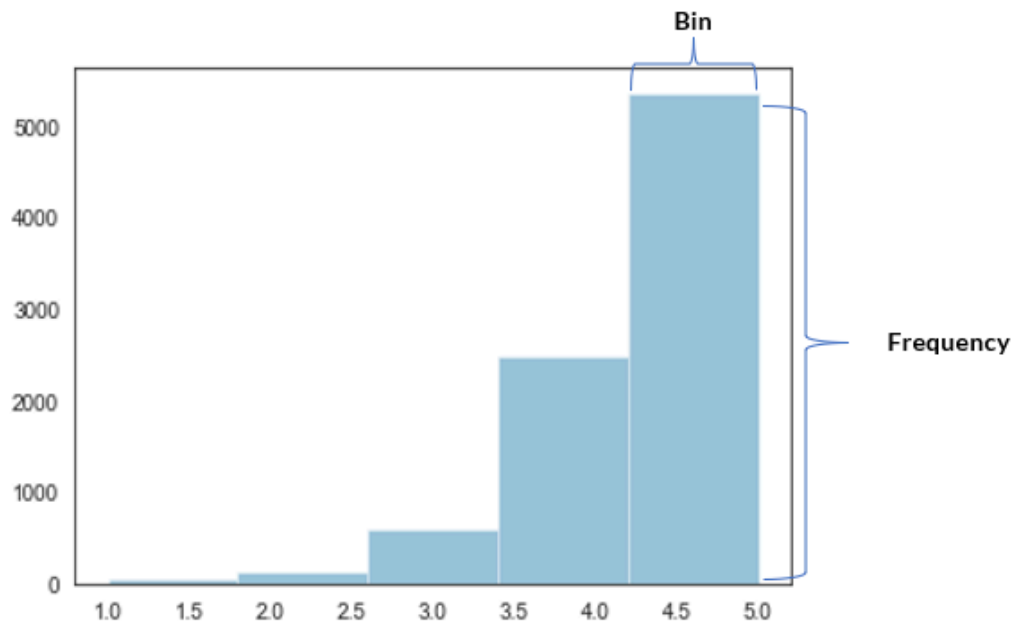🏠 **Learn**   ((◦)) **Live**   ⬭ **Discussions**   2
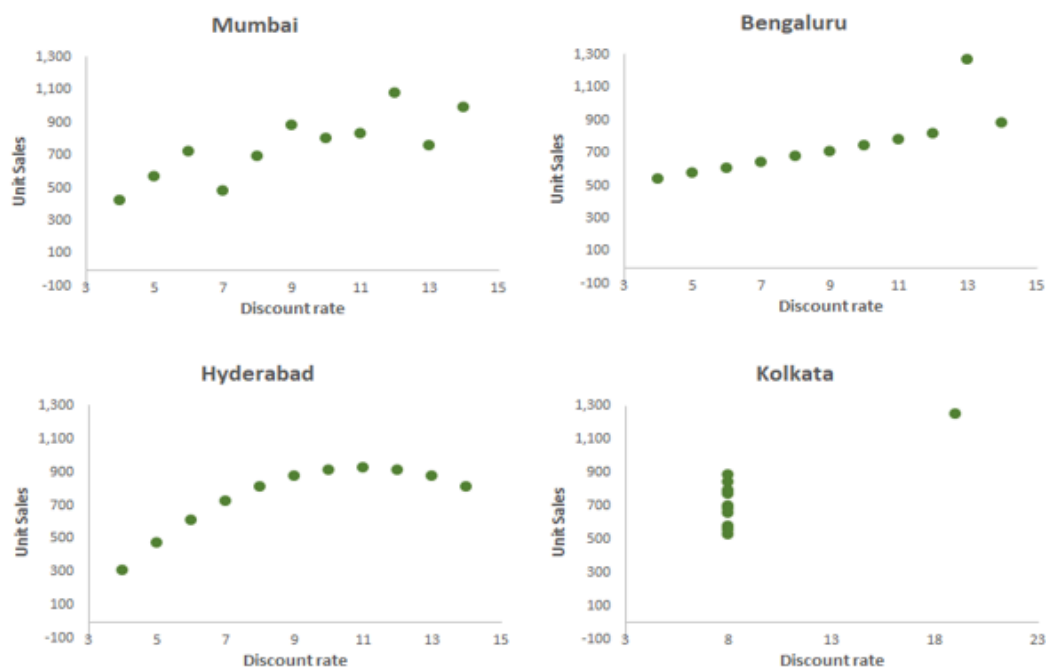
- Next, you began your data visualisation journey with the help of a case study. Here, you had to analyse the ratings of various Google Play Store apps in order to determine the features that decide whether an app is performing well.

- In the beginning, you performed some basic data-handling and cleaning tasks using your knowledge of Python libraries, which you have learnt earlier. You converted data types, cleaning certain columns containing erroneous entries, and even deleted records having a lot of null values.

- After that, you learnt how visualisation tools like box plots and histograms can help you with certain data-cleaning and analysis tasks. Box plots are required for understanding the spread of a numeric variable and identifying extreme values or outliers.

Inter-Quartile Range
(IQR)

Fence
(Q1 - 1.5*IQR)

Fence
(Q3 +1.5*IQR)

Q1 or Lower Hinge
(25th Percentile)

Median
(50th Percentile)

Q3 or Upper Hinge
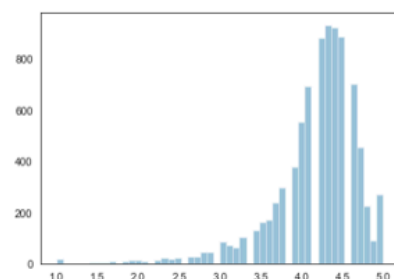(75th Percentile)

OUTLIERS : Values outside the range (Q1 - 1.5 x IQR, Q3 + 1.5 x IQR)

- Then you also learnt how to manipulate the number of bins to obtain more granular results as shown in the figure below:

upGrad and IIITB Machine Learning & AI Program - July 2024

Learn   **Live**   **Discussions**   2
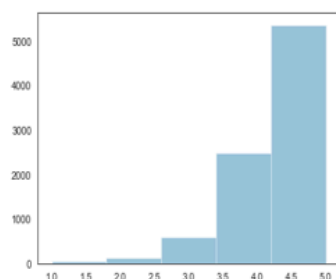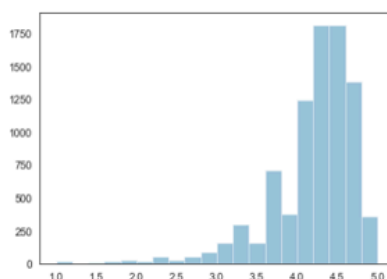
Question 1/1   Mandatory

Top 3 Takeaways

What are your top three takeaways from this session?

Word Count 50	Word Limit 5 - 100

Visualization make data more understandable.

plotting different graphs

you are analysing the total sum which is a numeric variable which is a categorical variable. Hence it will need a bar plot.

you're understanding the spread of a numeric variable by checking the frequency. Hence a histogram will be used here.

Attempt 0 of 1   Submit

In the next session

With this, you have completed the basic data-handling and cleaning tasks. In the next session, you'll begin your journey of the Seaborn library, and use its visual tools to obtain insights and answer the business questions that were asked earlier while discussing the case study objectives.

upGrad and
IIITB Machine
Learning & AI
Program - July
2024

🏠 **Learn**

((o)) **Live**

💬 **Discussions**