

Practical 9

AIM:

Use the inbuilt dataset 'titanic' as used in previous experiment . Plot a box plot for distribution of age with respect to each gender along with the information about whether they survived or not.(column name 'sex' and 'age')

REQUIREMENT:

AnacondaInstaller

Windows10OS

Linux

JupyterNotebook

THEORY:

What is Data Visualization:

- Data visualization is the **graphical representation of data** using charts, graphs, and maps.
 - It helps to **understand trends, patterns, and outliers easily**.
 - It is important in **Big Data** to analyze large datasets and make decisions.
-

Uses of Data Visualization:

- Helps to **model complex events**
 - Used to visualize things that cannot be seen directly like:
 - Weather patterns
 - Medical data
 - Mathematical relationships
-

Data Visualization Technique: Box and Whisker Plot (Box Plot):

- Used to show **statistical summary of data**
 - Displays:
 - **Minimum and Maximum values (whiskers)**
 - **Quartiles (25%, 50%, 75%)**
 - Helps to identify **outliers** (values outside whiskers)
 - Shows **Interquartile Range (IQR)** for data spread
-

Implementation:

- Box plot is available in **Seaborn library**
- Used for **univariate analysis** (one variable)
- Helps to study the effect of one feature on another (e.g., feature vs class)

LibrariesUsed:

Seaborn

Conclusion

in this experiment we have studied about the data visualization technique and implemented data visualization on the built in data set in seaborn laibrary i.e., titanic dataset