Q1.) Explain Exploratory Data Analysis.

ANS.
- Exploratory Data Analysis (EDA) is a process of Analyzing and summarzing data sets to understand their main characteristics, often using visual Methods.
- It helps in identifying patterns, trends, relationship and anomalies in data before applying any modelling techniqu
- EDA involve both graphical and non-graphical techniques
- Common graphical tools include histogram, boxplots, scatter
- plots and bar charts, while non-graphical methods involves statistics like mean, median, mode, standard deviation and correlation.
- The main goal of EDA is to make sense of data, detect outliers or missing values, and check assumptions required for further Analysis.
- It plays a crucial role in the data science workflow by guiding data cleaning, Feature selection, and model choice.
- EDA is often the first step in any data project and serves as a foundation for building more accurrate and meaningful Models

Q2.) Explain Univariate Analysis.

ANS. Univariate Analysis is the examination of a single variable to understand its distribution and charatteristic.

. It Focuses on summarizing data and identifying patterns within that one variable.

. This type of analysis doesn't consider relationships with other variable.

. For Numerical variables, common tools include histograms, box plots, and summary statistics such as mean, median, variance and standard deviation.

. For Categorical variances, bar plots and Frequency tables are often used.

SPPU-TE-COMP-CONTENT – KSKA Git

- The Gool is to understand the central tendency (where the data is centered), dispersion (how spread out it is), and the presence of Outliers.
- Univariate Analysis is often the first step in data Exploration, as it gives a clear picture of each variable individually, which is essential for detecting errors, prepar for further analysis, and building intuition about the dataset.

Q3.) What is Multivariate Anaplysis? Explain.

ANS.

- Multivariate Analysis involves the examination of more thar two variables simultaneously to understand relationships and interactions among them.
- It is used to identify pattern, trends, and correlations that are not visible in uni-variate or bi-variate Analysis.
- Techniques used in multivariate analysis include multipl regression, principal component Analysis (PCA), factor analysis, cluster analysis, and MANOVA (Multivariate Analysis of Variance).
- These methods helps in understanding how moltiple variables impact a particular outcome or how they group together.
- For Example, in customer segmentation, Multivariate Analysis might be used to find patterns in age, income, and spending habits simultaneously.
- Visualization tools like scatter plot matrices and heat-maps also aid in understanding multi-variate relationships
- It's especially useful in complex datasets with many variables, allowing analysts to reduce dimensionality, detect hidden structures, and make better decesions based on depper insights.

SPPU-TE-COMP-CONTENT – KSKA Git

Q4.> How is Distplot and Boxplot created. Explain Each.

ANS. 1. DISPLOT :-

- Displot is a visualization that combines a histogram with a kernel density estimate (KDE) to show the distribution of a numerical variable.
- Its commonly created using Seaborn's sns.displot() or the older sns.displot().
- It helps in visualizing the shape of the distribution (e.g., normal, skewed) and identifying the presence of multiple nodes
- For Example:- (Python code)

```
import seaborn as sns
sns.displot(data['age'], kde = True)
```

- This will show how values are spread and concentrated across different intervals.

2. BOXPLOT :-

- Boxplot is used to display the distribution of a variable based on five summary statistics : minimum first quartile, median , third quartile, and maximum.
- It also identifies outliers.
- Created using Seaborn's sns.boxplot() or Matplotlib's boxplot() function, its ideal for comparing distributions across different groups.
- For Example :- (Python Code)

```
import seaborn as sns
sns.boxplot(x ='gender', y='income', data = data)
```

· This shows how income varies between genders ,highlighting medians and potential outliers.