Exploratory Data Analysis using Python

Performing EDA using Juptyer Notebook.

**The process of performing EDA involves:**

* Import necessary libraries such as pandas, NumPy
* Load the dataset
* Data cleaning and preprocessing
* Derive Statistical insights (mean, median, mode)
* Perform EDA (univariate, bivariate, multivariate)
* Extract business insights
* Import necessary libraries

The first step involves importing the necessary libraries

* Load the dataset using pandas
* Load the dataset using the necessary function from the pandas library. If dataset is in excel format, convert it into CSV and load the dataset using read\_csv().
* Display first few rows of the dataset using head()
* Get shape of data and information about the dataset using the shape method and info().
* Data cleaning and preprocessing
* Check for null values using isnull(). Replace the values with ‘Unknown’.
* Dateofbill column has text datatype. Convert it into datetime format using the to\_datetime() function.
* Derive Statistical insights
* The describe() gives statistical insights such as mean, median, mode, and standard deviation about numeric columns from the dataset.
* Perform Exploratory Data Analysis(Visualization)
* Univariate Analysis:

1. Typeofsales: The overall frequency of Sales is greater than the frequency of Return
2. Form1 from Formulation and Department1 have highest frequency.

* Bivariate Analysis: Perform bivariate analysis categorical column using a bar chart concerning Final\_Sales.
  + - * 1. Dept: Department1 has a higher sales rate in terms of Final\_Sales
        2. Formulation: The patent category is the highest selling in terms of Final\_Sales
        3. On the basis of months, May, July and December show higher sales compared with other months in the year 2022.
        4. Most of the quantities sold are between 0-20 items.
* Multivariate Analysis: Perform correlation matrix to find correlation between variables.
* Extract business insights:
* Questionnaire:

1. What is the proportion of sales versus returns in the dataset?

Proportion of Sales: 88.18%

Proportion of Return: 11.82 %

1. Which specialisations exhibit higher return rates?

Specialisation54 has a higher return rate of 42.86% with 6 Return count.

1. Which department shows the highest rate of returns?

Department1 has a higher rate of returns of 15.61 of total returns with a 10760 Return count.

1. Are there any noticeable seasonal variations in returns?

May, August and July are the top 3 months with a higher return rate