Marwadi University	**Marwadi University** **Faculty of Engineering and Technology** **Department of Information and Communication Technology**	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

Aim: Exploratory Data Analysis (EDA) using Python

IDE: Tableau

Theory:

Exploratory Data Analysis (EDA) is a critical phase in the data analysis process that involves examining and understanding data to gain insights, detect patterns, and identify relationships. As an essential precursor to formal statistical modeling, EDA helps analysts familiarize themselves with the data's characteristics, uncover potential anomalies or errors, and inform subsequent analytical decisions. At its core, EDA emphasizes visualization techniques, statistical summaries, and intuitive exploration to extract meaningful information from raw datasets.

One of the fundamental principles of EDA is data visualization, which enables analysts to visually inspect the distribution, structure, and relationships within the dataset. Graphical representations such as histograms, scatter plots, box plots, and heatmaps provide intuitive insights into the data's central tendencies, variability, and distributions across different variables. Through visual exploration, analysts can identify outliers, understand data patterns, and detect potential relationships between variables, laying the groundwork for further analysis.

Statistical summaries are another integral aspect of EDA, providing quantitative insights into the dataset's key characteristics. Measures such as mean, median, mode, variance, and standard deviation offer valuable summaries of central tendency, dispersion, and shape for numerical variables. For categorical variables, frequency tables, proportions, and percentages illuminate the distribution and prevalence of different categories. These statistical summaries complement visualizations by providing numerical context and facilitating comparisons across variables and subsets of data.

EDA also involves exploring relationships and dependencies between variables to uncover underlying patterns and associations. Correlation analysis examines the strength and direction of relationships between pairs of numerical variables, highlighting potential dependencies and predictive insights. Additionally, cross-tabulations and contingency tables assess the association between categorical variables, revealing patterns and dependencies within the data. Exploring these relationships is crucial for understanding the underlying structure of the data and informing subsequent modeling decisions.

Another key aspect of EDA is outlier detection, which involves identifying observations that deviate significantly from the overall pattern or distribution of the data. Outliers can arise due to measurement errors, data entry mistakes, or genuine anomalies in the underlying process being studied. Through visual inspection, statistical tests, and domain knowledge, analysts can identify and assess the impact of outliers on the data analysis process. Handling outliers appropriately is essential to ensure the robustness and validity of subsequent analyses and modeling efforts.

Marwadi University	**Marwadi University** **Faculty of Engineering and Technology** **Department of Information and Communication Technology**	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

In addition to these techniques, EDA encompasses data transformation and feature engineering to prepare the data for further analysis. This may involve handling missing values, transforming variables to meet modeling assumptions, and creating new variables or features to capture additional information. By iteratively exploring, visualizing, and transforming the data, analysts can gain a comprehensive understanding of its structure, patterns, and characteristics, laying the foundation for informed decision-making and hypothesis generation.

Exploratory Data Analysis is a critical step in the data analysis process, enabling analysts to explore, visualize, and understand data to extract meaningful insights and inform subsequent analytical decisions. Through a combination of visualization techniques, statistical summaries, and intuitive exploration, analysts can uncover patterns, detect anomalies, and identify relationships within the data, laying the groundwork for further analysis and modeling. EDA serves as a powerful tool for hypothesis generation, data preparation, and decision-making, driving value and actionable insights from raw datasets.

Pre Lab Exercise:

Write the use-case/applications for following graphs:

1. Histogram:

2. Barplot:

3. Pie Chart:

4. Scatterplot:

5. BoxPlot:

Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	Aim: Exploratory Data Analysis (EDA) using Python	
Experiment No: 12	Date: 30-03-2024	Enrollment No: 92200133030

Tasks:

Perform the following tasks:

1) Perform the EDA analysis using Python over Superstore dataset.

❖ **Import Necessary Library and Read The Dataset :-**

Code :-

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns

Dataset = pd.read_excel("./Sample - Superstore.xlsx", "Orders")
print(Dataset)
```

Results :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python> & "C:/Program Files/Python312/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Data Visulization and Dashboards/Lab Manual/Exp-12 EDA Using Python/EDA.py"
      Row ID        Order ID Order Date  Month  ...     Sales Quantity Discount    Profit
0          1  CA-2016-152156 2016-11-08     11  ...  261.9600        2     0.00   41.9136
1          2  CA-2016-152156 2016-11-08     11  ...  731.9400        3     0.00  219.5820
2          3  CA-2016-138688 2016-06-12      6  ...   14.6200        2     0.00    6.8714
3          4  US-2015-108966 2015-10-11     10  ...  957.5775        5     0.45 -383.0310
4          5  US-2015-108966 2015-10-11     10  ...   22.3680        2     0.20    2.5164
...      ...             ...         ...    ...  ...       ...      ...      ...       ...
9989    9990  CA-2014-110422 2014-01-21      1  ...   25.2480        3     0.20    4.1028
9990    9991  CA-2017-121258 2017-02-26      2  ...   91.9600        2     0.00   15.6332
9991    9992  CA-2017-121258 2017-02-26      2  ...  258.5760        2     0.20   19.3932
9992    9993  CA-2017-121258 2017-02-26      2  ...   29.6000        4     0.00   13.3200
9993    9994  CA-2017-119914 2017-05-04      5  ...  243.1600        2     0.00   72.9480

[9994 rows x 23 columns]
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python>
```

❖ **To get the shape of your Dataset :-**

Code :-

```
print(Dataset.shape)
```

Result :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python> & "C:/Program Files/Python312/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Data Visulization and Dashboards/Lab Manual/Exp-12 EDA Using Python/EDA.py"
(9994, 23)
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python>
```

Marwadi University	**Marwadi University** **Faculty of Engineering and Technology** **Department of Information and Communication Technology**	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

❖ **To get the columns of your dataset :-**

Code :-

```
print(Dataset.columns)
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python> & "C:/Program Files/Python312/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Data Visulization and Dashboards/Lab Manual/Exp-12 EDA Using Python/EDA.py"
Index(['Row ID', 'Order ID', 'Order Date', 'Month', 'Year', 'Ship Date',
       'Ship Mode', 'Customer ID', 'Customer Name', 'Segment', 'Country',
       'City', 'State', 'Postal Code', 'Region', 'Product ID', 'Category',
       'Sub-Category', 'Product Name', 'Sales', 'Quantity', 'Discount',
       'Profit'],
      dtype='object')
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python>
```

❖ **To Get the Information of all the columns of your dataset :-**

Code :-

```
print(Dataset.info())
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python> & "C:/Program Files/Python312/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Data Visulization and Dashboards/Lab Manual/Exp-12 EDA Using Python/EDA.py"
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 23 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Row ID         9994 non-null   int64
 1   Order ID       9994 non-null   object
 2   Order Date     9994 non-null   datetime64[ns]
 3   Month          9994 non-null   int64
 4   Year           9994 non-null   int64
 5   Ship Date      9994 non-null   datetime64[ns]
 6   Ship Mode      9994 non-null   object
 7   Customer ID    9994 non-null   object
 8   Customer Name  9994 non-null   object
 9   Segment        9994 non-null   object
 10  Country        9994 non-null   object
 11  City           9994 non-null   object
 12  State          9994 non-null   object
 13  Postal Code    9994 non-null   int64
 14  Region         9994 non-null   object
 15  Product ID     9994 non-null   object
 16  Category       9994 non-null   object
 17  Sub-Category   9994 non-null   object
 18  Product Name   9994 non-null   object
 19  Sales          9994 non-null   float64
 20  Quantity       9994 non-null   int64
 21  Discount       9994 non-null   float64
 22  Profit         9994 non-null   float64
dtypes: datetime64[ns](2), float64(3), int64(5), object(13)
memory usage: 1.8+ MB
None
```

Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	Aim: Exploratory Data Analysis (EDA) using Python	
Experiment No: 12	Date: 30-03-2024	Enrollment No: 92200133030

❖ **To Get the top rows of Datset**

Code :-

```
print(Dataset.head())
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python> & "C:/Program Files/Python312/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Data Visulization and Dashboards/Lab Manual/Exp-12 EDA Using Python/EDA.py"
   Row ID        Order ID Order Date  Month  ...     Sales Quantity Discount    Profit
0       1  CA-2016-152156 2016-11-08     11  ...  261.9600        2     0.00   41.9136
1       2  CA-2016-152156 2016-11-08     11  ...  731.9400        3     0.00  219.5820
2       3  CA-2016-138688 2016-06-12      6  ...   14.6200        2     0.00    6.8714
3       4  US-2015-108966 2015-10-11     10  ...  957.5775        5     0.45 -383.0310
4       5  US-2015-108966 2015-10-11     10  ...   22.3680        2     0.20    2.5164

[5 rows x 23 columns]
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python>
```

❖ **To Get the Bottom rows of the Dataset**

Code :-

```
print(Dataset.tail())
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python> & "C:/Program Files/Python312/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Data Visulization and Dashboards/Lab Manual/Exp-12 EDA Using Python/EDA.py"
      Row ID        Order ID Order Date  Month  ...    Sales Quantity Discount   Profit
9989    9990  CA-2014-110422 2014-01-21      1  ...   25.248        3      0.2   4.1028
9990    9991  CA-2017-121258 2017-02-26      2  ...   91.960        2      0.0  15.6332
9991    9992  CA-2017-121258 2017-02-26      2  ...  258.576        2      0.2  19.3932
9992    9993  CA-2017-121258 2017-02-26      2  ...   29.600        4      0.0  13.3200
9993    9994  CA-2017-119914 2017-05-04      5  ...  243.160        2      0.0  72.9480

[5 rows x 23 columns]
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python>
```

❖ **To Get Know if there is any null values**

Code :-

```
print(Dataset.isnull())
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python> & "C:/Program Files/Python312/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Data Visulization and Dashboards/Lab Manual/Exp-12 EDA Using Python/EDA.py"
      Row ID  Order ID  Order Date  Month   Year  Ship Date  ...  Sub-Category  Product Name  Sales  Quantity  Discount  Profit
0      False     False       False  False  False      False  ...         False         False  False     False     False   False
1      False     False       False  False  False      False  ...         False         False  False     False     False   False
2      False     False       False  False  False      False  ...         False         False  False     False     False   False
3      False     False       False  False  False      False  ...         False         False  False     False     False   False
4      False     False       False  False  False      False  ...         False         False  False     False     False   False
...      ...       ...         ...    ...    ...        ...  ...           ...           ...    ...       ...       ...     ...
9989   False     False       False  False  False      False  ...         False         False  False     False     False   False
9990   False     False       False  False  False      False  ...         False         False  False     False     False   False
9991   False     False       False  False  False      False  ...         False         False  False     False     False   False
9992   False     False       False  False  False      False  ...         False         False  False     False     False   False
9993   False     False       False  False  False      False  ...         False         False  False     False     False   False

[9994 rows x 23 columns]
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python>
```

Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	Aim: Exploratory Data Analysis (EDA) using Python	
Experiment No: 12	Date: 30-03-2024	Enrollment No: 92200133030

❖ **To Get the No of Null Values in the Columns**

Code :-

```
print(Dataset.isnull().sum())
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python> & "C:/Program Files/Python312/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Data Visulization and Dashboards/Lab Manual/Exp-12 EDA Using Python/EDA.py"
Row ID           0
Order ID         0
Order Date       0
Month            0
Year             0
Ship Date        0
Ship Mode        0
Customer ID      0
Customer Name    0
Segment          0
Country          0
City             0
State            0
Postal Code      0
Region           0
Product ID       0
Category         0
Sub-Category     0
Product Name     0
Sales            0
Quantity         0
Discount         0
Profit           0
dtype: int64
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python>
```

❖ **To get the Total No of Null Values in the Dataset**

Code :-

```
print(Dataset.isnull().sum().sum())
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python> & "C:/Program Files/Python312/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Data Visulization and Dashboards/Lab Manual/Exp-12 EDA Using Python/EDA.py"
0
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python>
```

Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	Aim: Exploratory Data Analysis (EDA) using Python	
Experiment No: 12	Date: 30-03-2024	Enrollment No: 92200133030

❖ **To Get the No of Unique Values in each Columns**

Code :-

```
print(Dataset.nunique())
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python> & "C:/Program Files/Python312/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Data Visulization and Dashboards/Lab Manual/Exp-12 EDA Using Python/EDA.py"
Row ID          9994
Order ID        5009
Order Date      1237
Month             12
Year               4
Ship Date       1334
Ship Mode          4
Customer ID      793
Customer Name    793
Segment            3
Country            1
City             531
State             49
Postal Code      631
Region             4
Product ID      1862
Category           3
Sub-Category      17
Product Name    1850
Sales           6144
Quantity          14
Discount          12
Profit          7545
dtype: int64
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python>
```

❖ **To Drop The Column from a Dataset.**

Code :-

```
Dataset = Dataset.drop(columns=["Postal Code"], axis=1)
print(Dataset.columns)
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python> & "C:/Program Files/Python312/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Data Visulization and Dashboards/Lab Manual/Exp-12 EDA Using Python/EDA.py"
Index(['Row ID', 'Order ID', 'Order Date', 'Month', 'Year', 'Ship Date',
       'Ship Mode', 'Customer ID', 'Customer Name', 'Segment', 'Country',
       'City', 'State', 'Region', 'Product ID', 'Category', 'Sub-Category',
       'Product Name', 'Sales', 'Quantity', 'Discount', 'Profit'],
      dtype='object')
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python>
```

Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	Aim: Exploratory Data Analysis (EDA) using Python	
Experiment No: 12	Date: 30-03-2024	Enrollment No: 92200133030

❖ **To Get the State Count for Entire Dataset :-**

Code :-

```
State_Wise_Counts = Dataset["State"].value_counts(ascending=True)
print(State_Wise_Counts)
```

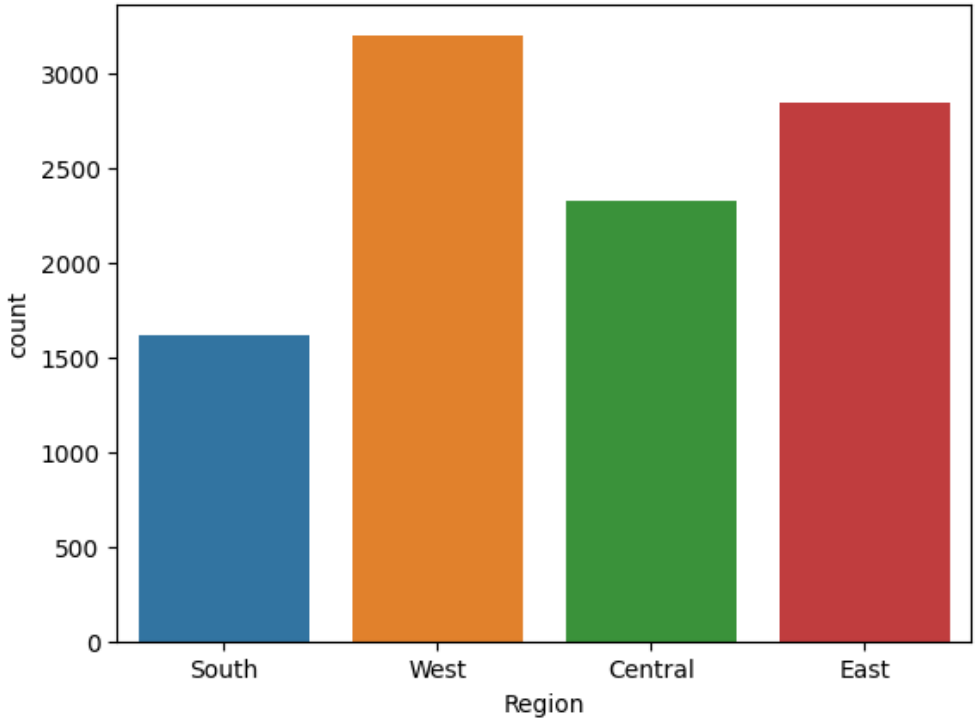
Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Data Visulization and Dashboards\Lab Manual\Exp-12 EDA Using Python> & "C:/Program Files/Python312/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Data Visulization and Dashboards/Lab Manual/Exp-12 EDA Using Python/EDA.py"
State
Wyoming                 1
West Virginia           4
North Dakota            7
Maine                   8
District of Columbia    10
Vermont                 11
South Dakota            12
Montana                 15
Idaho                   21
Kansas                  24
New Hampshire           27
Iowa                    30
New Mexico              37
Nebraska                38
Nevada                  39
Louisiana               42
South Carolina          42
Mississippi             53
Utah                    53
Rhode Island            56
Arkansas                60
Alabama                 61
Missouri                66
```

Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	Aim: Exploratory Data Analysis (EDA) using Python	
Experiment No: 12	Date: 30-03-2024	Enrollment No: 92200133030

❖ **To Plot The Graph For the Same**

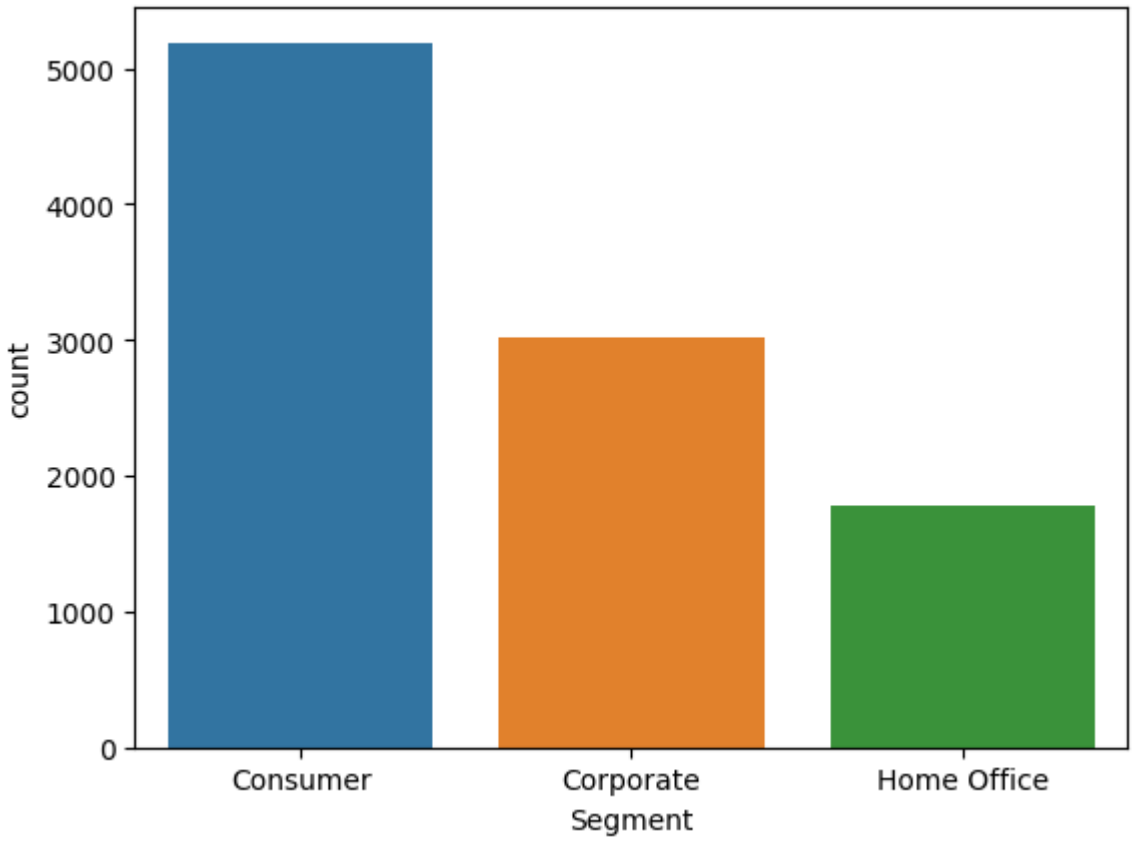
```
sns.countplot(x=Dataset["Region"], hue=Dataset["Region"])
plt.show()
```



❖ **To Plot The Count Plot for the Segment :-**

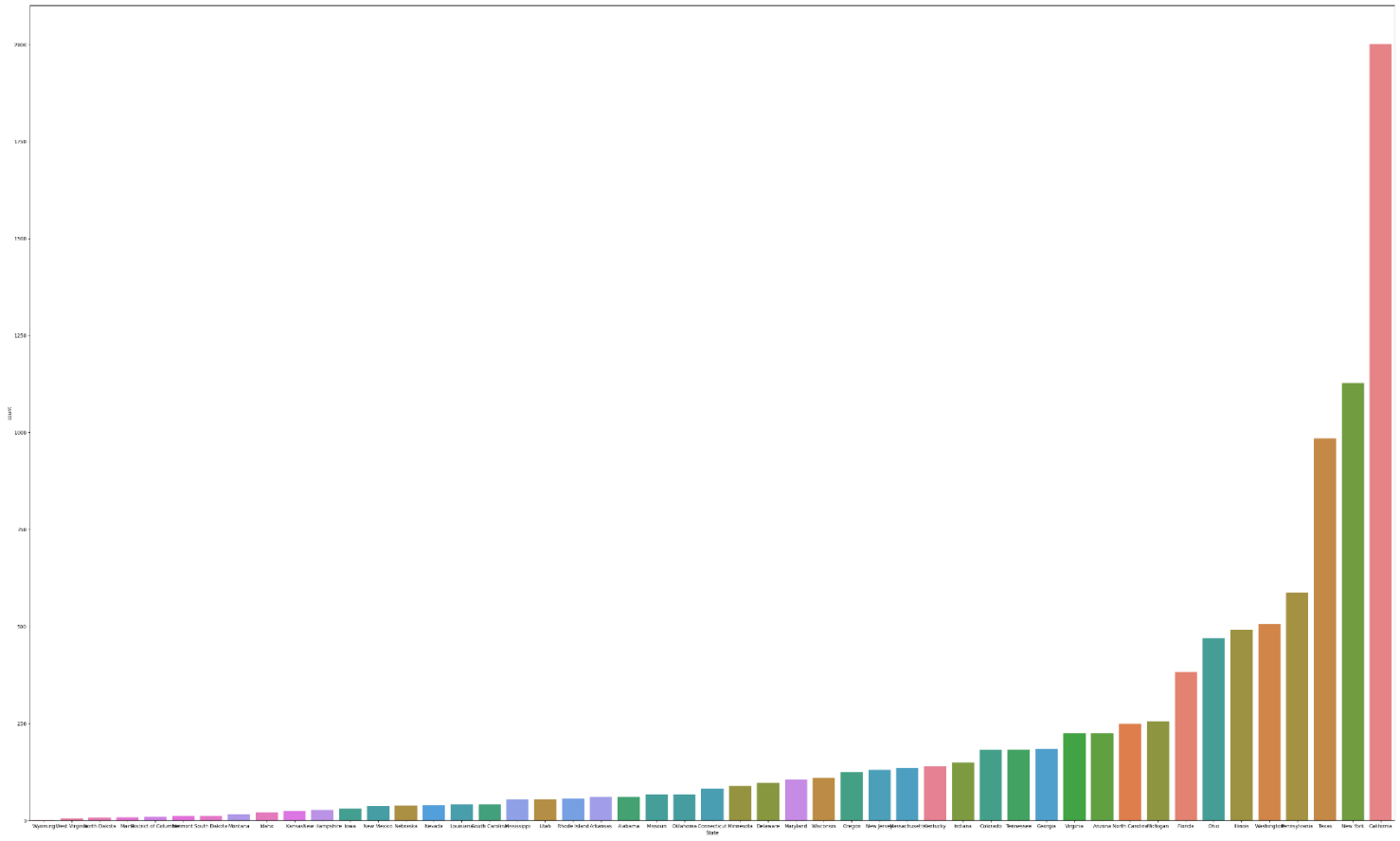
```
sns.countplot(x=Dataset["Segment"], hue=Dataset["Segment"])
plt.show()
```

Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	Aim: Exploratory Data Analysis (EDA) using Python	
Experiment No: 12	Date: 30-03-2024	Enrollment No: 92200133030



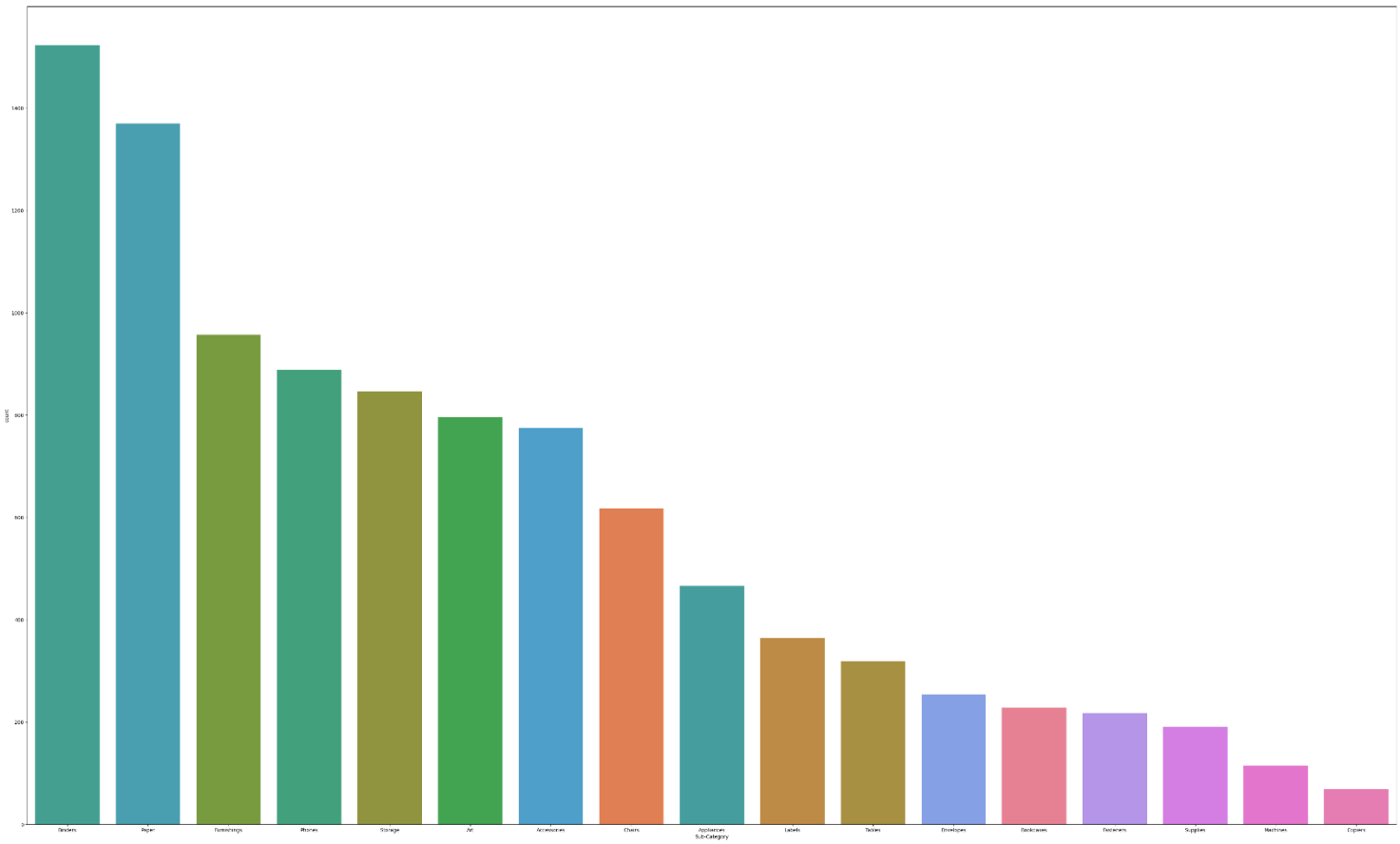

```
plt.figure(figsize=(50, 30))
sns.countplot(
   x=Dataset["State"],
   hue=Dataset["State"],
   order=Dataset["State"].value_counts(ascending=True).index,
)
plt.show()
```

Marwadi University	**Marwadi University** **Faculty of Engineering and Technology** **Department of Information and Communication Technology**	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**



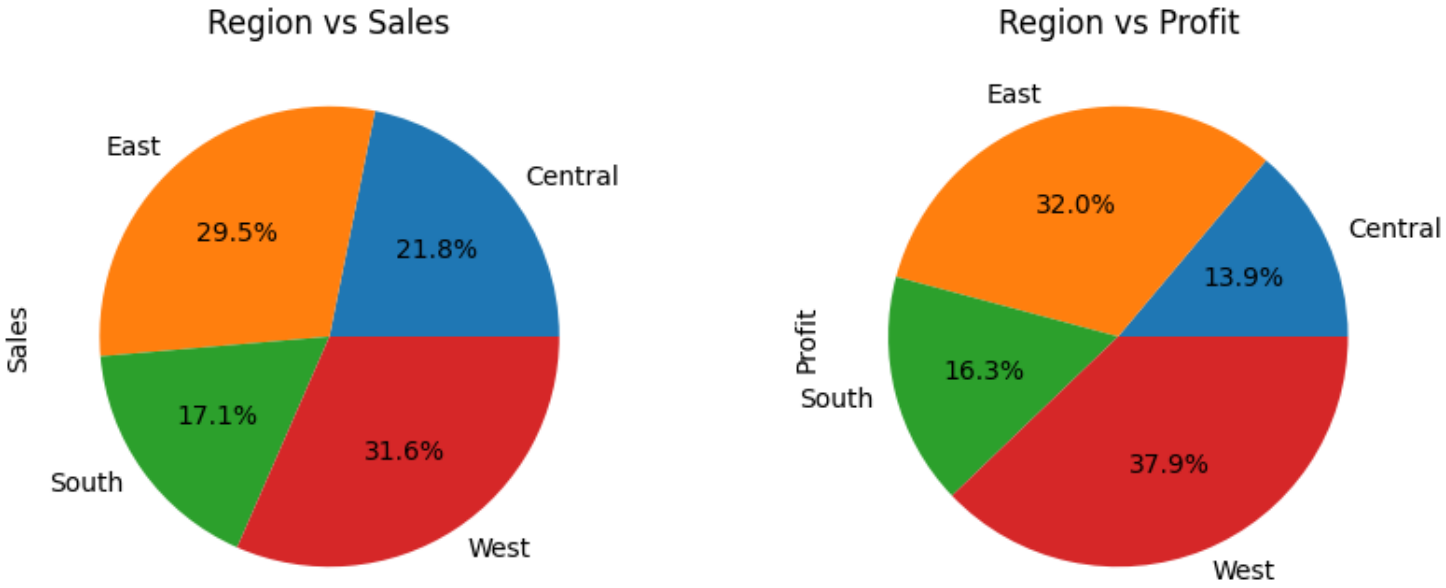
```
plt.figure(figsize=(50, 30))
sns.countplot(
    data=Dataset,
    x="Sub-Category",
    hue="Sub-Category",
    order=Dataset["Sub-Category"].value_counts().index,
)
plt.show()
```

Marwadi University	**Marwadi University** **Faculty of Engineering and Technology** **Department of Information and Communication Technology**	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**



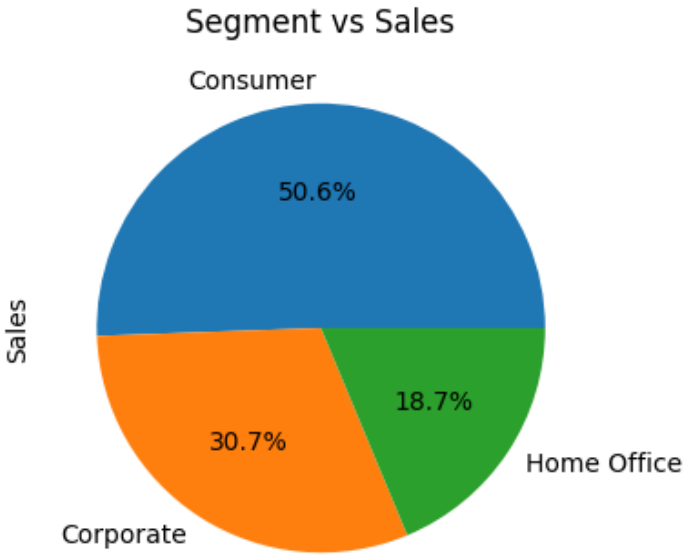
```
fig = plt.figure(figsize=(10, 4))
plt.subplot(1, 2, 1)
Dataset.groupby("Region")["Sales"].sum().plot.pie(autopct="%1.01f%%")
plt.title("Region vs Sales")
plt.subplot(1, 2, 2)
Dataset.groupby("Region")["Profit"].sum().plot.pie(autopct="%1.01f%%")
plt.title("Region vs Profit")
plt.show()
```

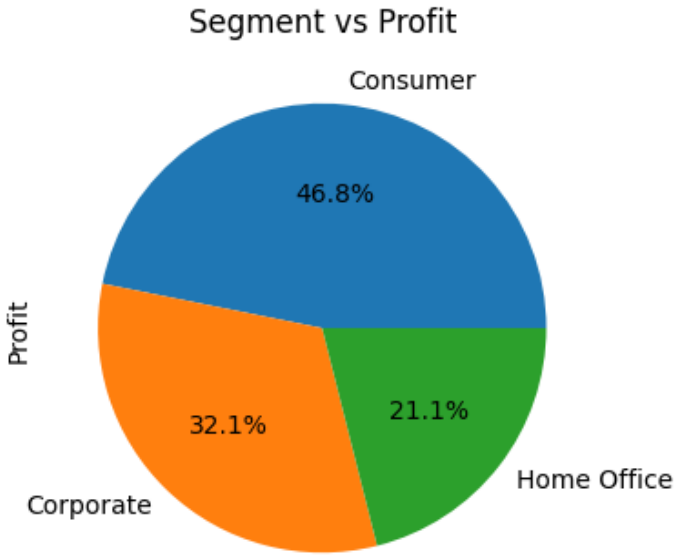
Marwadi University	**Marwadi University** **Faculty of Engineering and Technology** **Department of Information and Communication Technology**	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**



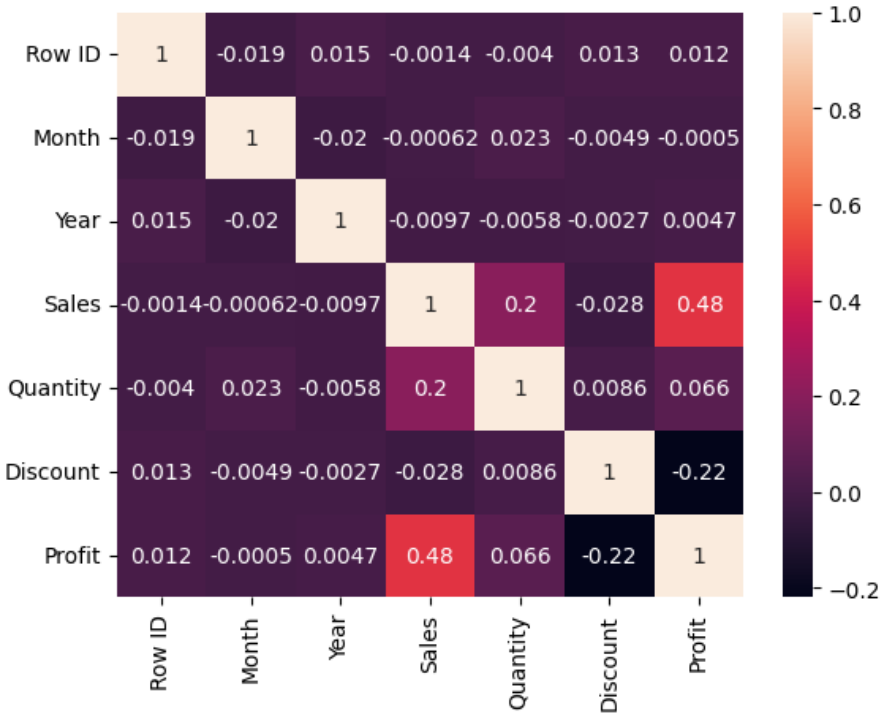

```
fig = plt.figure(figsize=(10, 4))
plt.subplot(1, 2, 1)
Dataset.groupby("Segment")["Sales"].sum().plot.pie(autopct="%1.01f%%")
plt.title("Segment vs Sales")
plt.subplot(1, 2, 2)
Dataset.groupby("Segment")["Profit"].sum().plot.pie(autopct="%1.01f%%")
plt.title("Segment vs Profit")
plt.show()
```

Marwadi University	**Marwadi University** **Faculty of Engineering and Technology** **Department of Information and Communication Technology**	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**



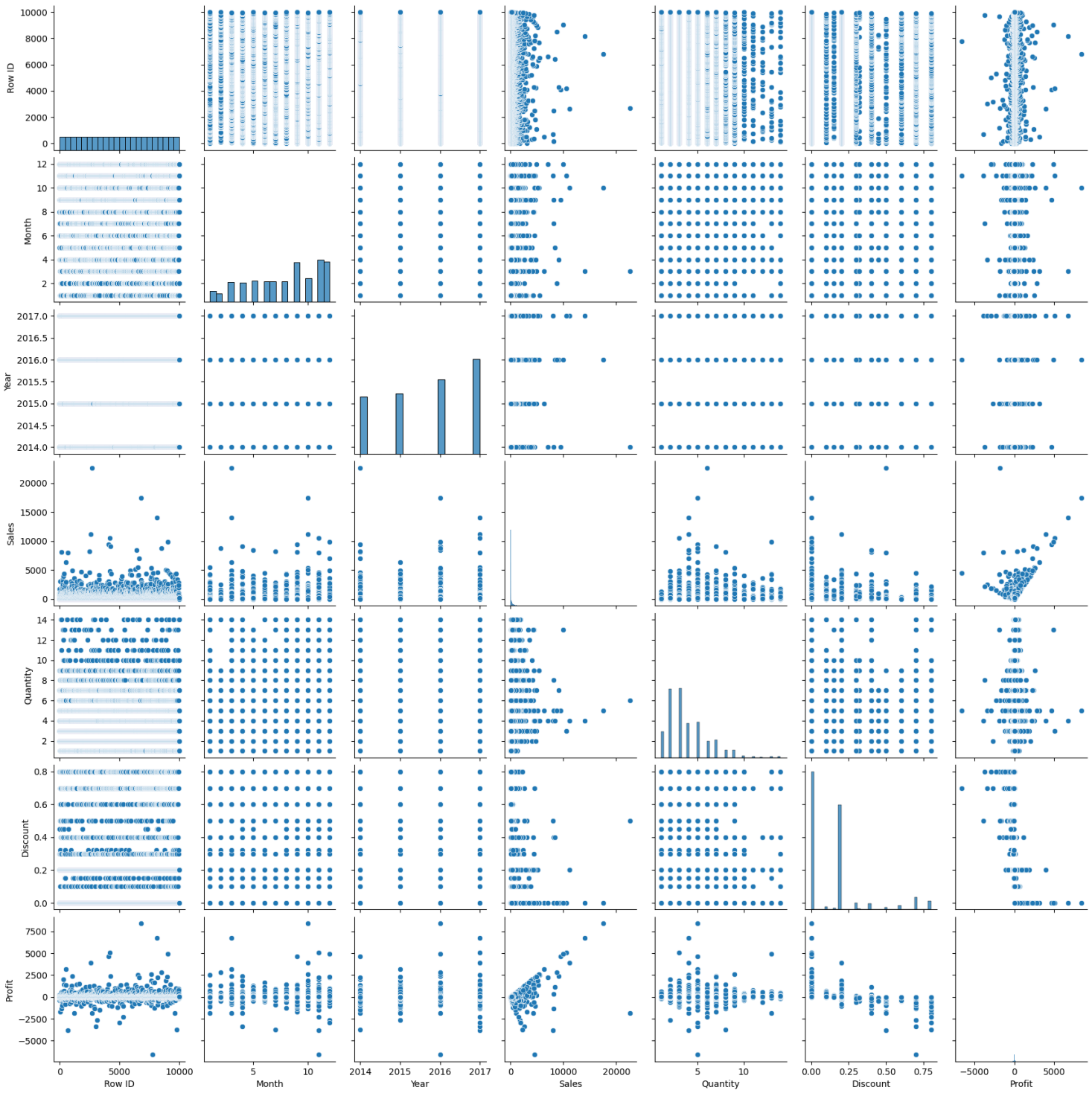



```
non_numeric_columns = Dataset.select_dtypes(exclude=["number"]).columns
Dataset_numeric = Dataset.drop(columns=non_numeric_columns)
sns.heatmap(Dataset_numeric.corr(), annot=True)
```



Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

```
fig = plt.figure(figsize=(10, 4))
sns.pairplot(Dataset)
plt.show()
```



Marwadi University	**Marwadi University** **Faculty of Engineering and Technology** **Department of Information and Communication Technology**	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

Observation and Result Analysis:

Write your inference corresponding to each of the analysis

Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

Post Lab Exercise:

Exercise 1: Perform the following EDA analysis using Python over the Book Shop dataset.

Pre-Requisites:-

Import Dataset :-

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

Book = pd.read_excel('./Bookshop.xlsx' , 'Book')
Author = pd.read_excel("./Bookshop.xlsx", "Author")
Author["Full_Name"] = Author["First Name"] + " " + Author["Last Name"]
Info = pd.read_excel('./Bookshop.xlsx' , 'Info')
Info['BookID'] = Info['BookID1'].astype(str) + Info['BookID2'].astype(str)
Award = pd.read_excel('./Bookshop.xlsx' , 'Award')
Checkouts = pd.read_excel('./Bookshop.xlsx' , 'Checkouts')
Edition = pd.read_excel('./Bookshop.xlsx' , 'Edition')
Publisher = pd.read_excel('./Bookshop.xlsx' , 'Publisher')
Ratings = pd.read_excel('./Bookshop.xlsx' , 'Ratings')
Series = pd.read_excel('./Bookshop.xlsx' , 'Series')
Sales_Q1 = pd.read_excel('./Bookshop.xlsx' , 'Sales Q1')
Sales_Q2 = pd.read_excel('./Bookshop.xlsx' , 'Sales Q2')
Sales_Q3 = pd.read_excel('./Bookshop.xlsx' , 'Sales Q3')
Sales_Q4 = pd.read_excel("./Bookshop.xlsx", "Sales Q4")
```

1. Who is the most popular or famous author (a. based on ratings; b. based on sales)

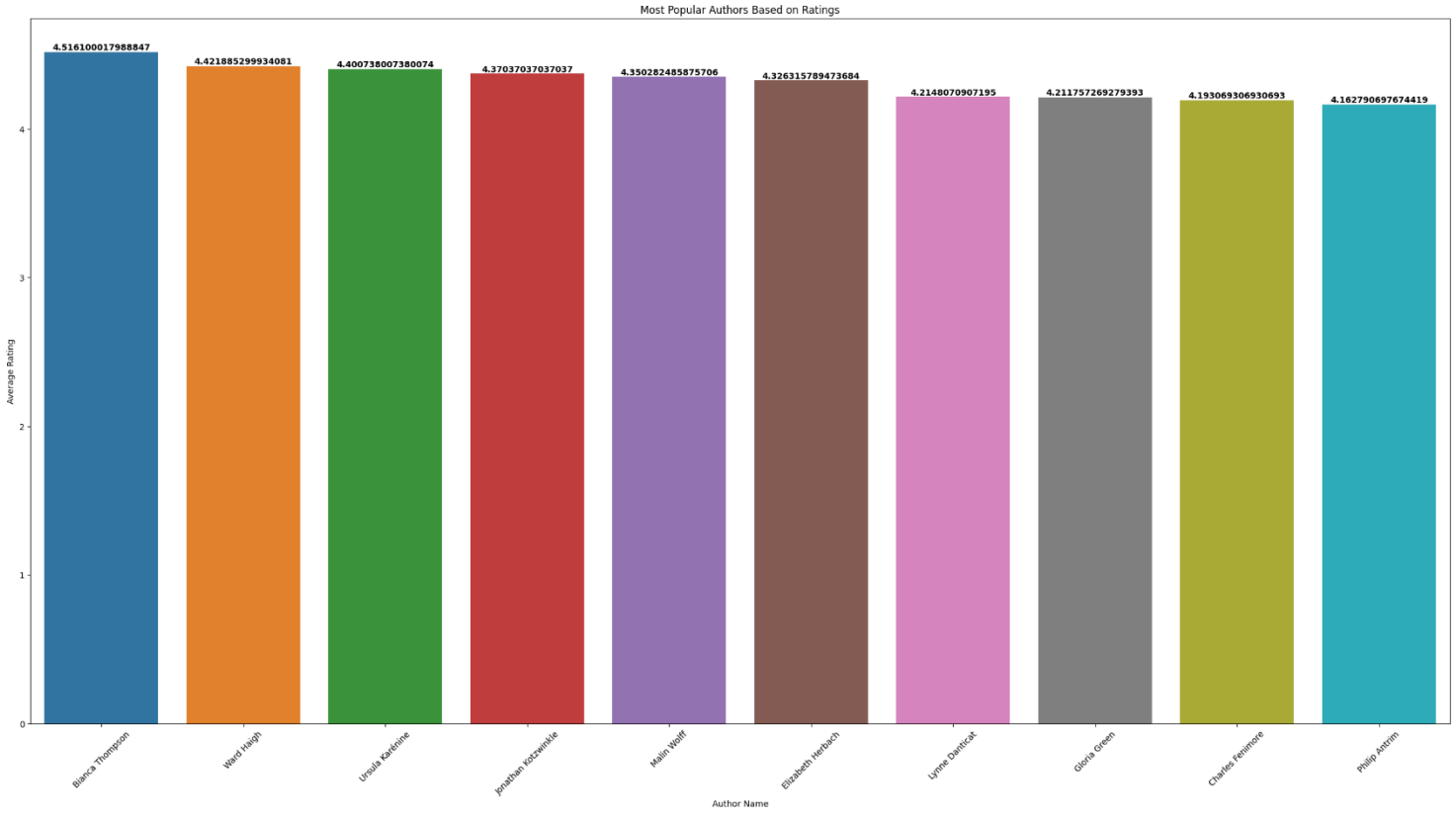
Code :-

```
# Based on Ratings
DS1 = pd.merge(Author, Book, on="AuthID")
DS1 = pd.merge(DS1, Ratings, on="BookID")
DS1["Full_Name"] = DS1["First Name"] + " " + DS1["Last Name"]
df1 = (DS1.groupby("Full_Name")["Rating"].mean().reset_index().sort_values(by="Rating", ascending=False))[:10]
plt.figure(figsize=(30, 15))
sns.barplot(x=df1["Full_Name"], y=df1["Rating"] , hue = df1['Full_Name'])
```

Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	Aim: Exploratory Data Analysis (EDA) using Python	
Experiment No: 12	Date: 30-03-2024	Enrollment No: 92200133030

```
for i, value in enumerate(df1['Rating']):
    plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")

plt.xlabel("Author Name")
plt.ylabel("Average Rating")
plt.title("Most Popular Authors Based on Ratings")
plt.xticks(rotation=45)
plt.show()
```

Output :-



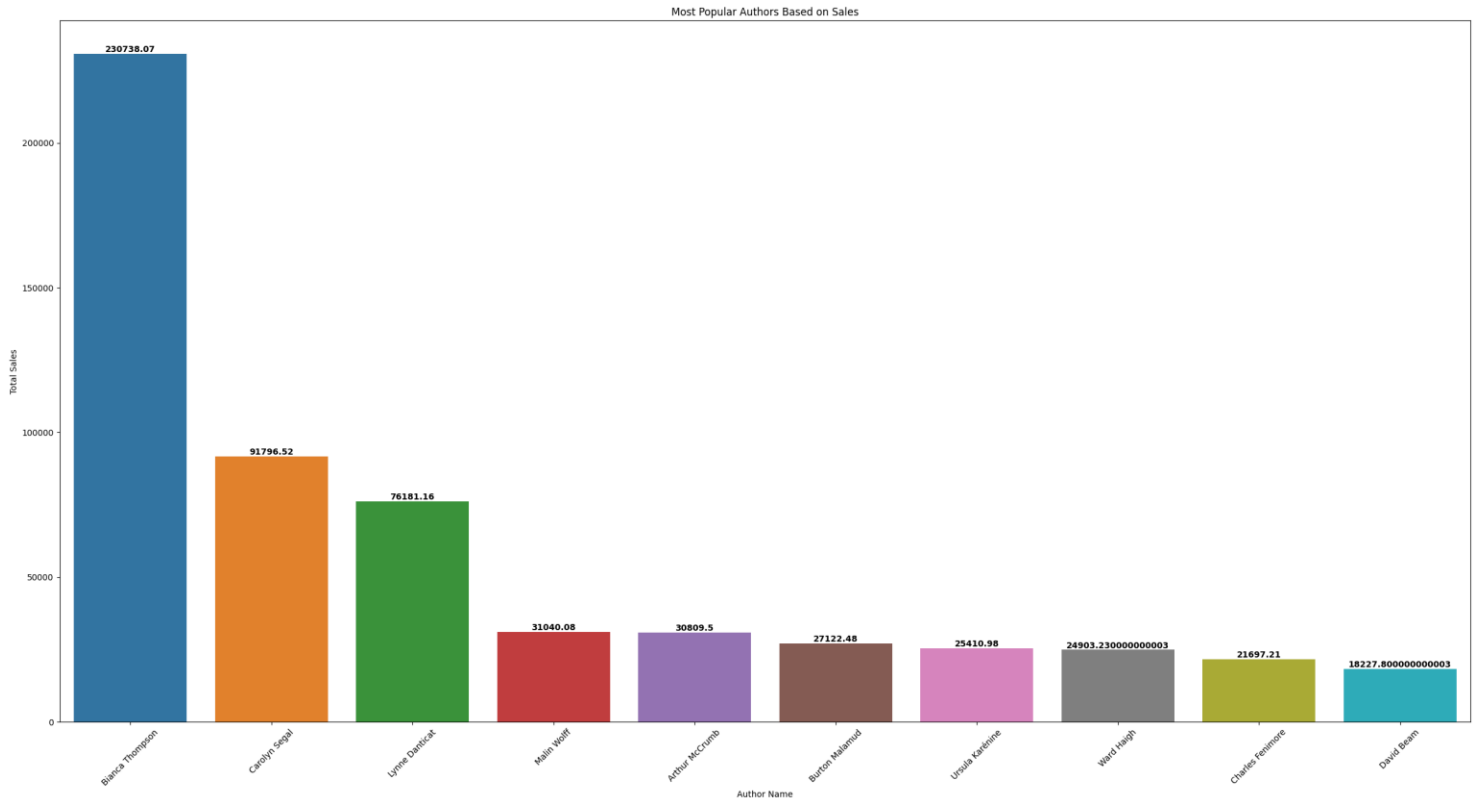
Code :-

```
Sales =
    pd.concat([Sales_Q1.value_counts("ISBN").reset_index(),Sales_Q2.value_counts("ISBN").reset_index(),Sales_
    Q3.value_counts("ISBN").reset_index(),Sales_Q4.value_counts('ISBN').reset_index(),],ignore_index=True,)
Sales = pd.merge(Sales , Edition , on="ISBN")
Sales['Total Price'] = Sales['count'] * Sales['Price']
Sales = Sales[['ISBN' , 'Total Price']].sort_values(by='Total Price' , ascending=False)
```

Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	Aim: Exploratory Data Analysis (EDA) using Python	
Experiment No: 12	Date: 30-03-2024	Enrollment No: 92200133030

```
DS1A = pd.merge(Book, Edition, on="BookID")
DS1A = pd.merge(DS1A, Author, on="AuthID")
DS1A = pd.merge(DS1A, Sales , on="ISBN")
DS1A = DS1A.groupby('Full_Name')['Total Price'].sum().reset_index().sort_values(by="Total Price",
    ascending=False)[:10]
df1a = DS1A

plt.figure(figsize=(30, 15))
sns.barplot(x=df1a["Full_Name"], y=df1a["Total Price"], hue=df1a["Full_Name"])

for i, value in enumerate(df1a["Total Price"]):
   plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")

plt.xlabel("Author Name")
plt.ylabel("Total Sales")
plt.title("Most Popular Authors Based on Sales")
plt.xticks(rotation=45)
plt.show()
```

Output :-



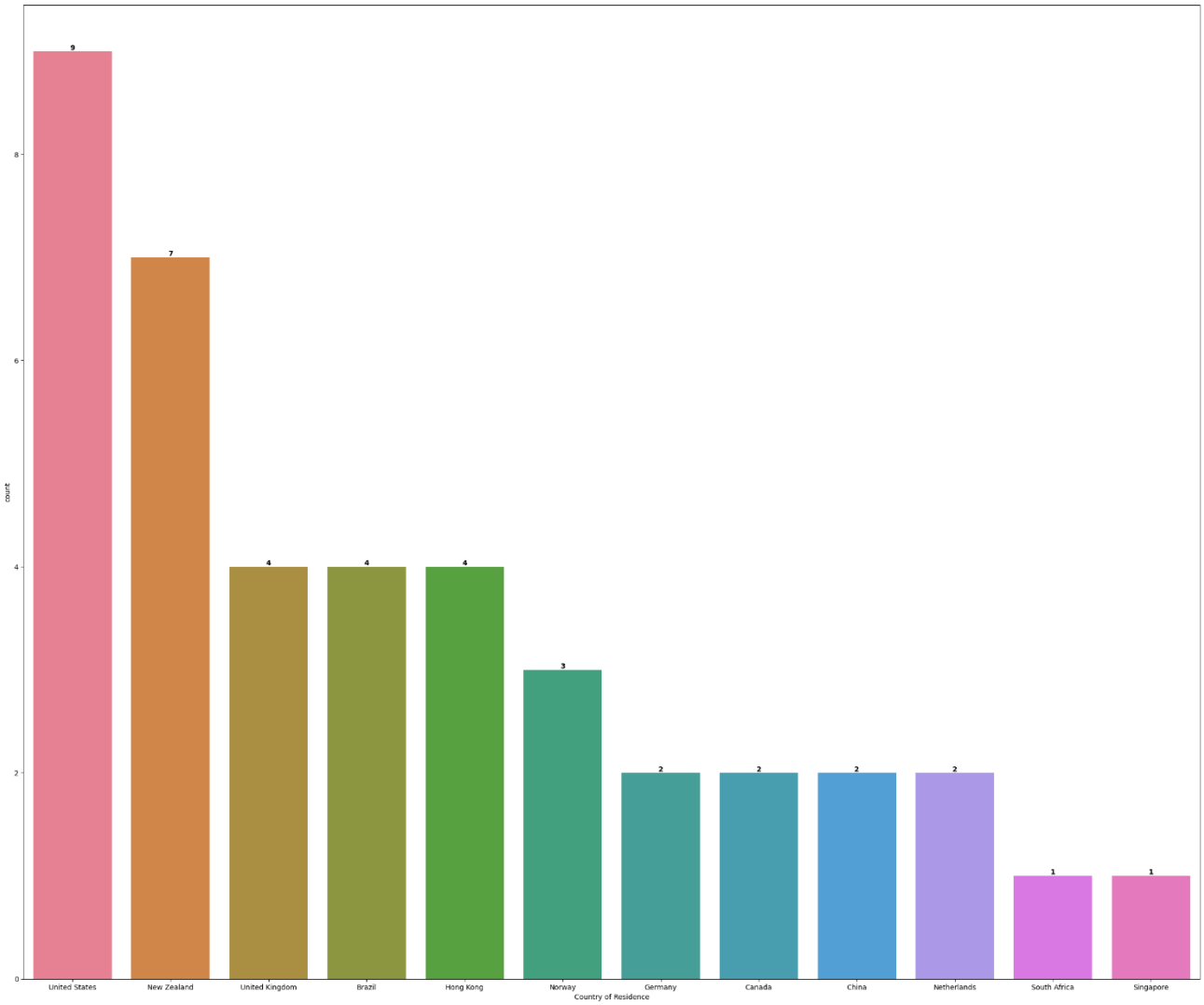
Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

2. Top-3 countries with the highest number of authors

Code :-

```
df2 = Author['Country of Residence'].value_counts().reset_index()
plt.figure(figsize=(30, 25))
sns.barplot(data=df2, x="Country of Residence", y = "count" , hue="Country of Residence")
for i, value in enumerate(Author["Country of Residence"].value_counts()):
    plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")
plt.show()
```

Output :-



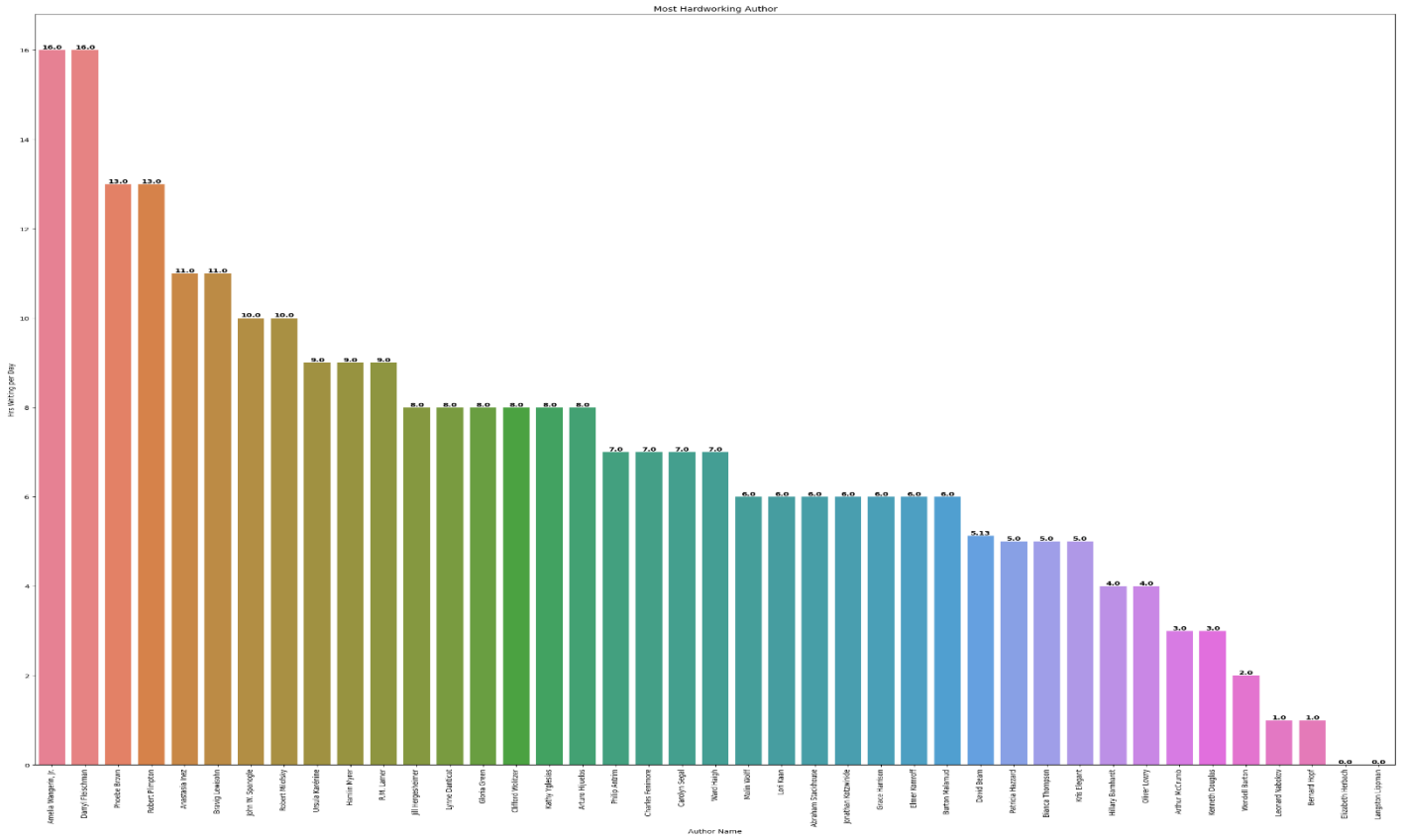
Marwadi University	**Marwadi University** **Faculty of Engineering and Technology** **Department of Information and Communication Technology**	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

3. Who is the hardworking author in terms of working per day

Code :-

```
df3 = Author.sort_values(by="Hrs Writing per Day" , ascending=False)
df3 = df3[['Full_Name' , "Hrs Writing per Day"]]

plt.figure(figsize=(30, 25))
sns.barplot(data=df3, x="Full_Name", y="Hrs Writing per Day", hue="Full_Name")

for i, value in enumerate(df3["Hrs Writing per Day"]):
    plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")

plt.xlabel("Author Name")
plt.ylabel("Hrs Writing per Day")
plt.title("Most Hardworking Author")
plt.xticks(rotation=90)
plt.show()
```

Output :-



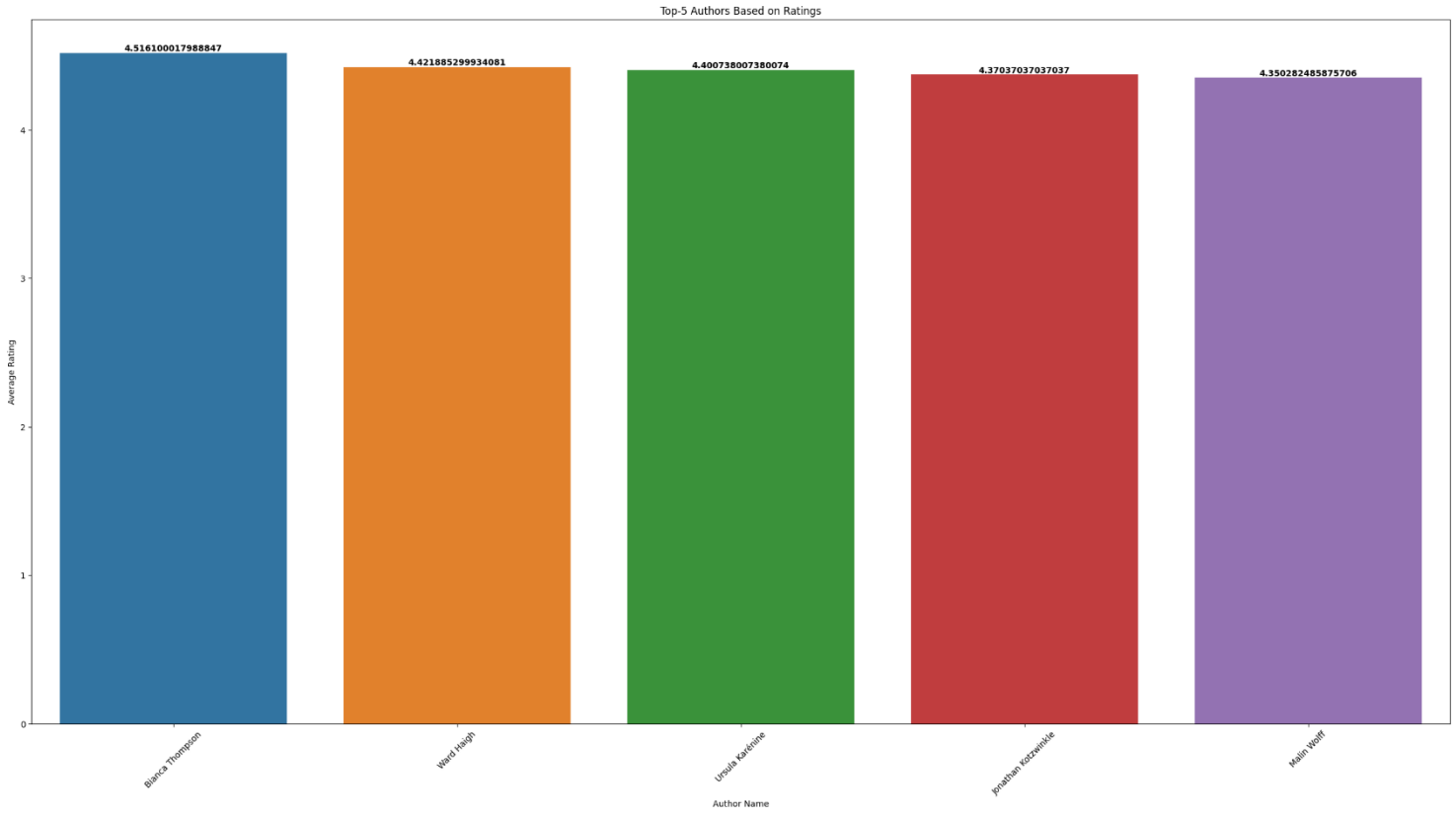
Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	Aim: Exploratory Data Analysis (EDA) using Python	
Experiment No: 12	Date: 30-03-2024	Enrollment No: 92200133030

4. Top 5 authors having the highest average ratings

Code :-

```
DS4 = pd.merge(Book , Ratings , on = "BookID")
DS4 = pd.merge(DS4 , Author , on = "AuthID")
df_4 = DS4.groupby('Full_Name')['Rating'].mean().sort_values(ascending=False).reset_index()[:5]
plt.figure(figsize=(30, 15))
sns.barplot(x=df_4["Full_Name"], y=df_4["Rating"] , hue = df_4['Full_Name'])
for i, value in enumerate(df_4['Rating']):
    plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")
plt.xlabel("Author Name")
plt.ylabel("Average Rating")
plt.title("Top-5 Authors Based on Ratings")
plt.xticks(rotation=45)
plt.show()
```

Output :-



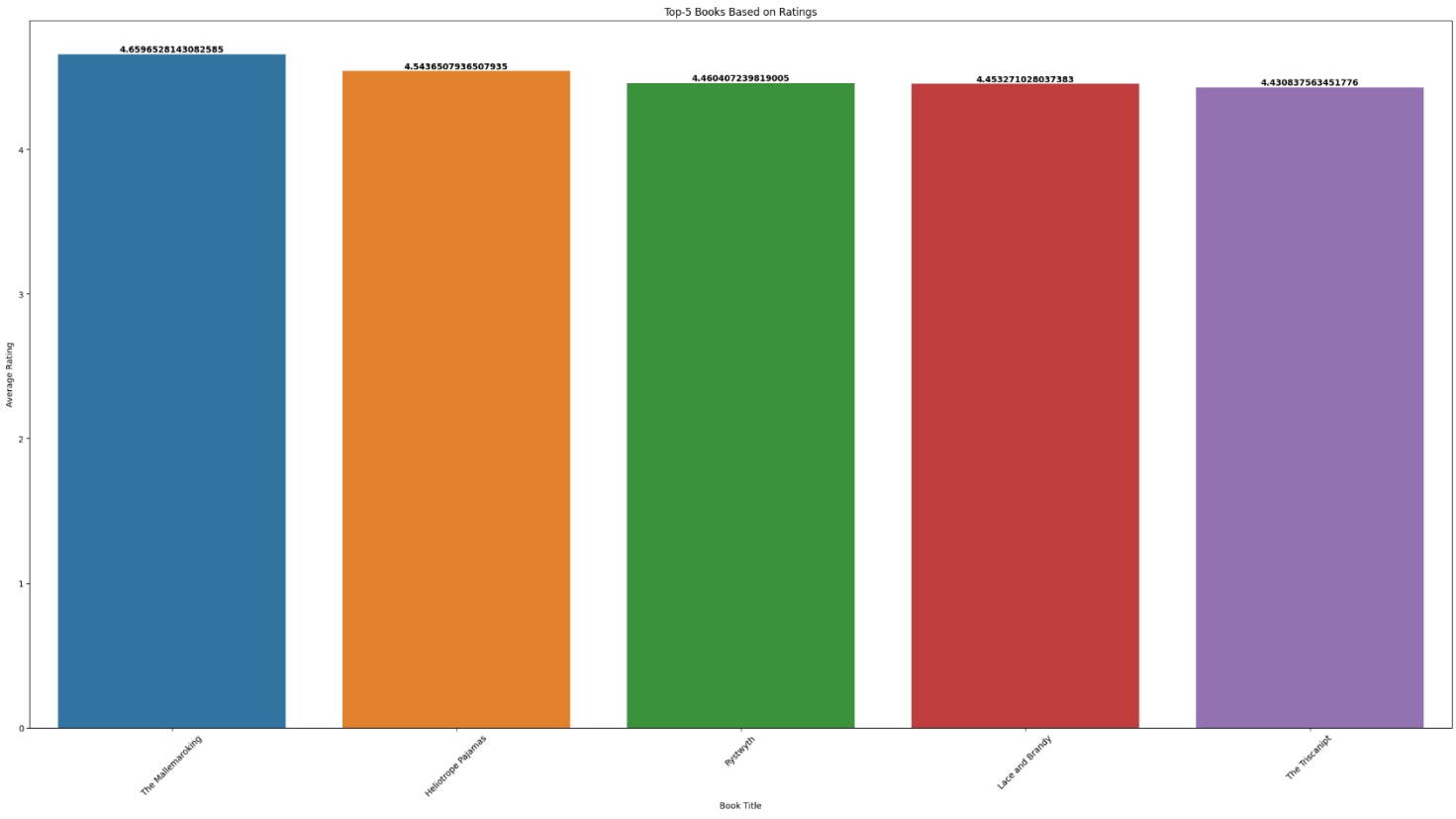
Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	Aim: Exploratory Data Analysis (EDA) using Python	
Experiment No: 12	Date: 30-03-2024	Enrollment No: 92200133030

5. Top 5 Books having the highest average ratings

Code :-

```
DS5 = pd.merge(Book , Ratings , on='BookID')
df5 = pd.DataFrame(DS5.groupby('Title')['Rating'].mean().reset_index()).sort_values(by = 'Rating' ,
    ascending=False)[:5]
plt.figure(figsize=(30, 15))
sns.barplot(x=df5["Title"], y=df5["Rating"] , hue = df5['Title'])
for i, value in enumerate(df5["Rating"]):
  plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")
plt.xlabel("Book Title")
plt.ylabel("Average Rating")
plt.title("Top-5 Books Based on Ratings")
plt.xticks(rotation=45)
plt.show()
```

Output :-



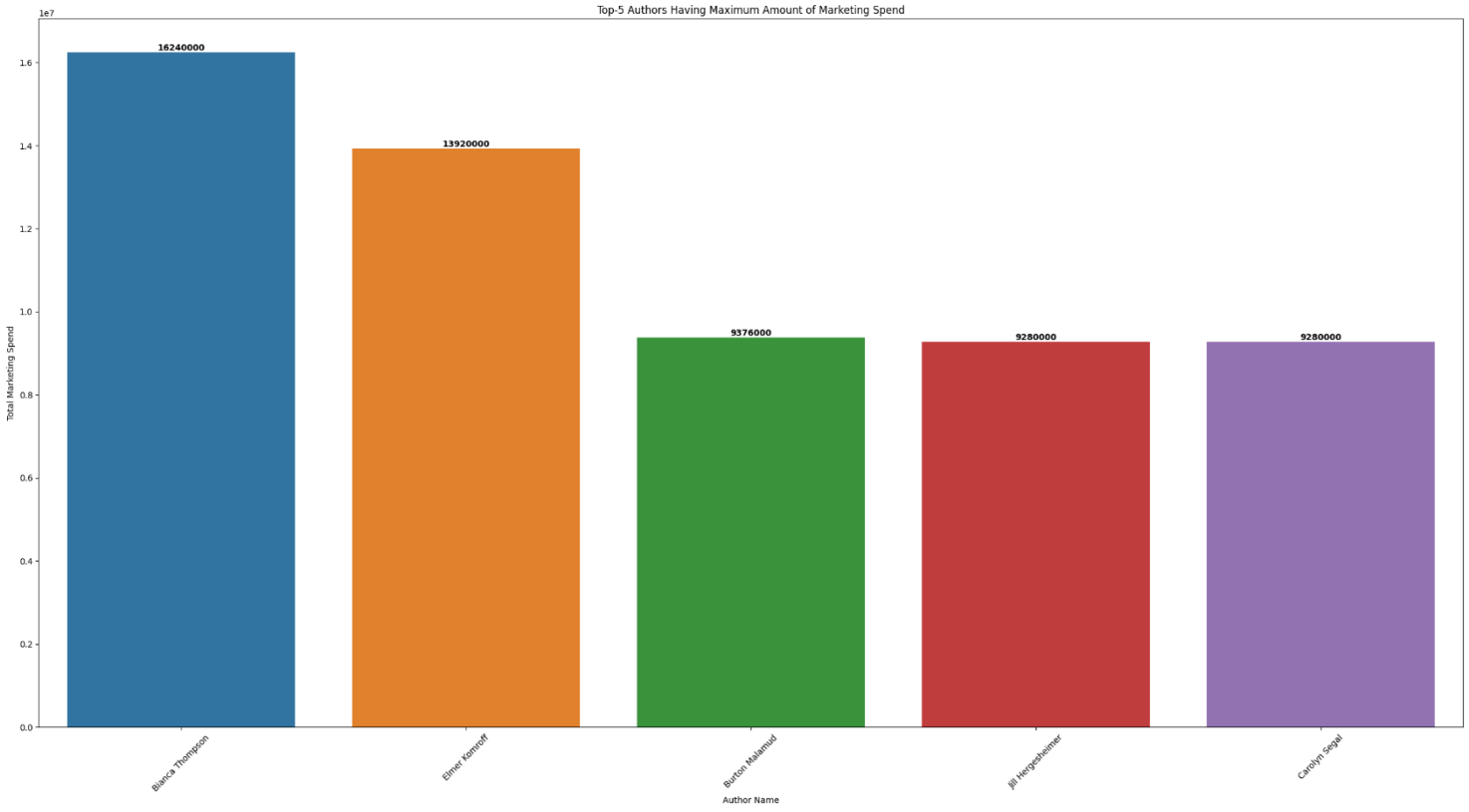
Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

6. Top-5 authors spending maximum amount on marketing

Code :-

```
DS6 = pd.merge(Book, Author, on="AuthID")
DS6 = pd.merge(DS6, Edition, on="BookID")
DS6 = pd.merge(DS6, Publisher, on="PubID")
df6 = pd.DataFrame(DS6.groupby("Full_Name")["Marketing
    Spend"].sum().reset_index()).sort_values(by="Marketing Spend", ascending=False)[:5]
plt.figure(figsize=(30, 15))
sns.barplot(x=df6["Full_Name"], y=df6["Marketing Spend"], hue=df6["Full_Name"])
for i, value in enumerate(df6["Marketing Spend"]):
  plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")
plt.xlabel("Author Name")
plt.ylabel("Total Marketing Spend")
plt.title("Top-5 Authors Having Maximum Amount of Marketing Spend")
plt.xticks(rotation=45)
plt.show()
```

Output :-



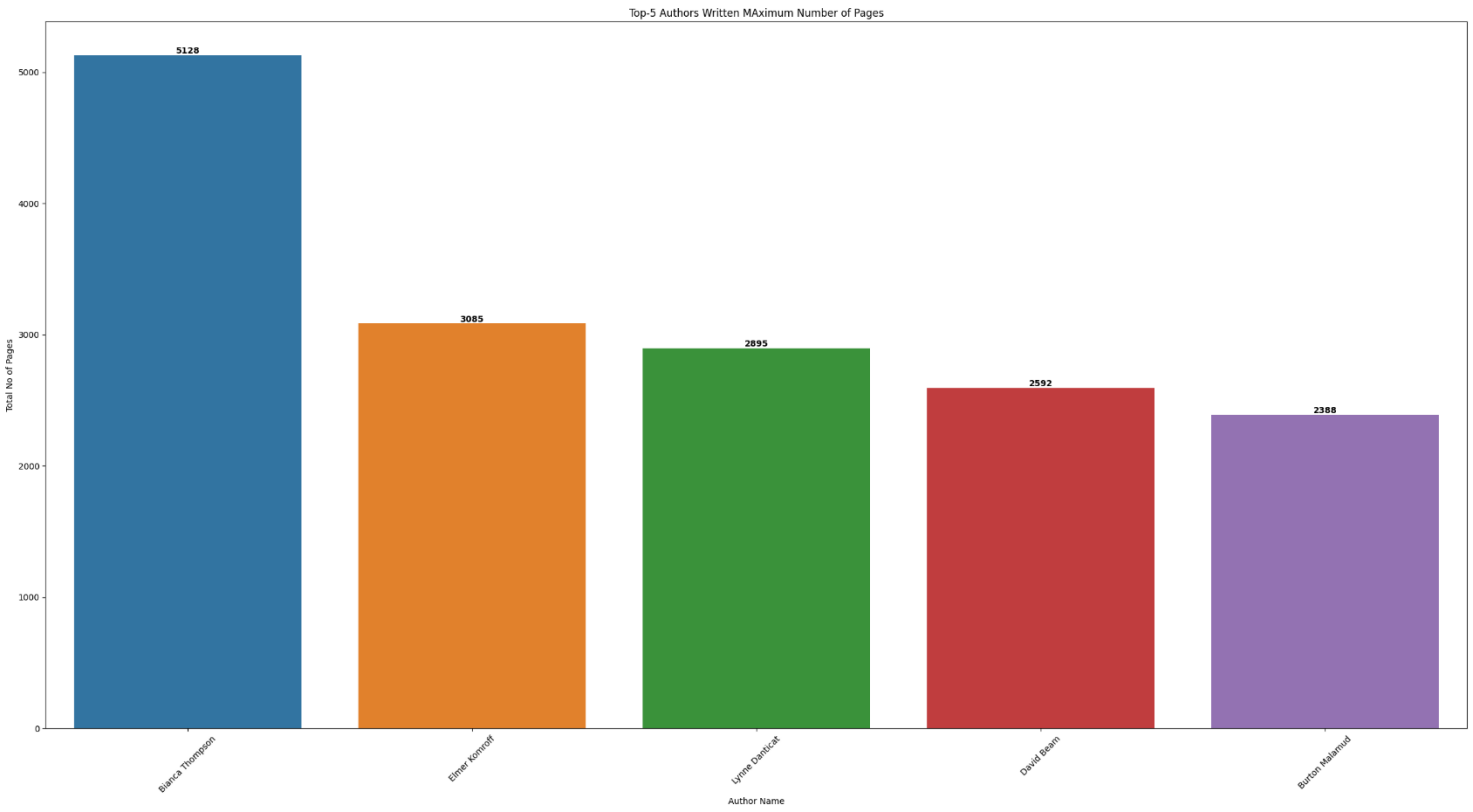
Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

7. Top 5 authors writing the highest total number of pages

Code :-

```
DS7 = pd.merge(Book, Author, on="AuthID")
DS7 = pd.merge(DS7, Edition, on="BookID")
df7 = pd.DataFrame(DS7.groupby("Full_Name")["Pages"].sum().reset_index()).sort_values(
   by="Pages", ascending=False
)[:5]
plt.figure(figsize=(30, 15))
sns.barplot(x=df7["Full_Name"], y=df7["Pages"], hue=df7["Full_Name"])
for i, value in enumerate(df7["Pages"]):
   plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")
plt.xlabel("Author Name")
plt.ylabel("Total No of Pages")
plt.title("Top-5 Authors Written MAximum Number of Pages")
plt.xticks(rotation=45)
plt.show()
```

Output :-



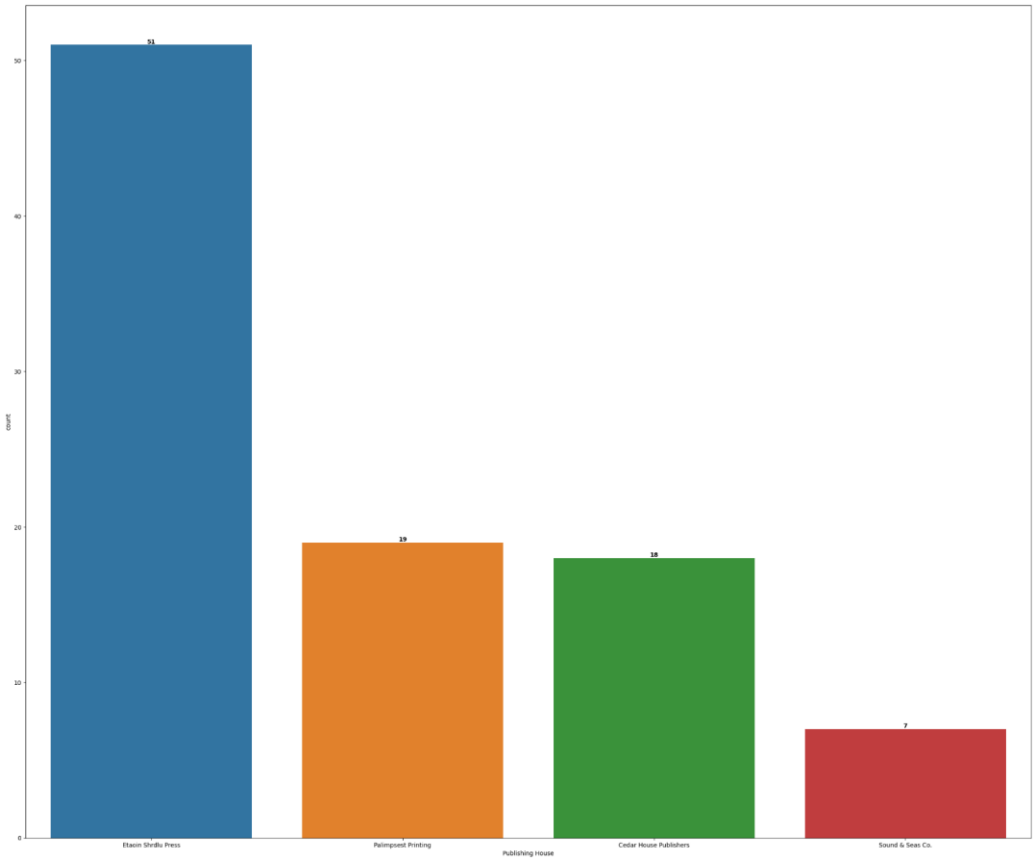
Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

8. Top-5 Publication House (in terms of count of books published)

Code :-

```
DS8 = pd.merge(Edition, Publisher, on="PubID")
df8 = DS8["Publishing House"].value_counts().reset_index()
plt.figure(figsize=(30, 25))
sns.barplot(data=df8, x="Publishing House", y="count", hue="Publishing House")
for i, value in enumerate(DS8["Publishing House"].value_counts()):
    plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")
plt.show()
```

Output :-



9. Top-5 publication house with highest-priced books

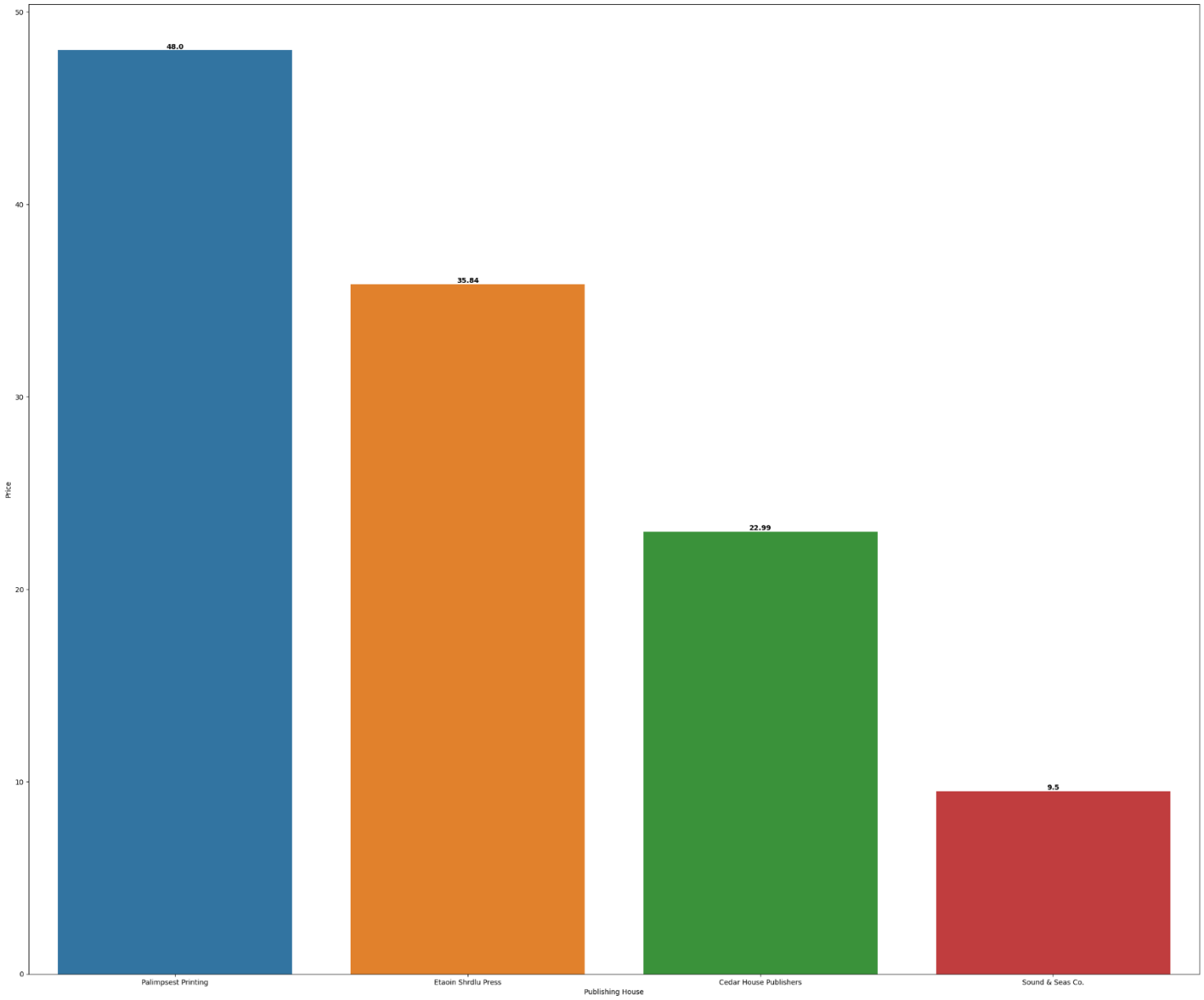
Code :-

```
DS9 = pd.merge(Edition, Publisher, on="PubID")
df9 = pd.DataFrame(
    DS9.groupby("Publishing House")["Price"].max().reset_index()
).sort_values(by="Price" , ascending= False)
```

Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	Aim: Exploratory Data Analysis (EDA) using Python	
Experiment No: 12	Date: 30-03-2024	Enrollment No: 92200133030

```
plt.figure(figsize=(30, 25))
sns.barplot(data= df9, x="Publishing House", y="Price", hue="Publishing House")
for i, value in enumerate(df9["Price"]):
    plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")
plt.show()
```

Output :-



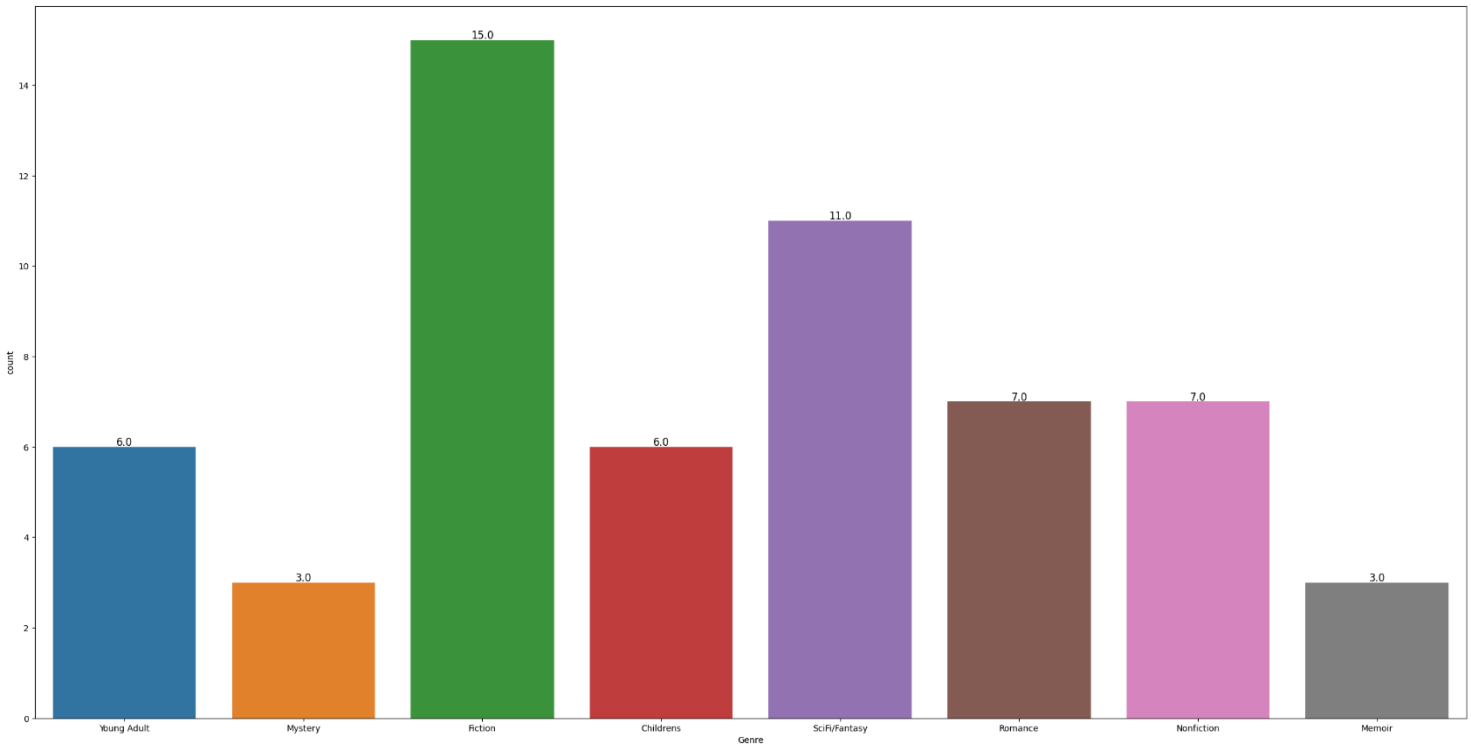
Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

10. Total number of books in each genre

Code :-

```
plt.figure(figsize=(30, 15))
ax = sns.countplot(data=Info, x="Genre", hue="Genre")

for p in ax.patches:
   ax.annotate(f"{p.get_height()}",(p.get_x() + p.get_width() / 2.0,
     p.get_height()),ha="center",va="center",fontsize=12,color="black",xytext=(0, 5),textcoords="offset points",)

plt.show()
```

Output:-



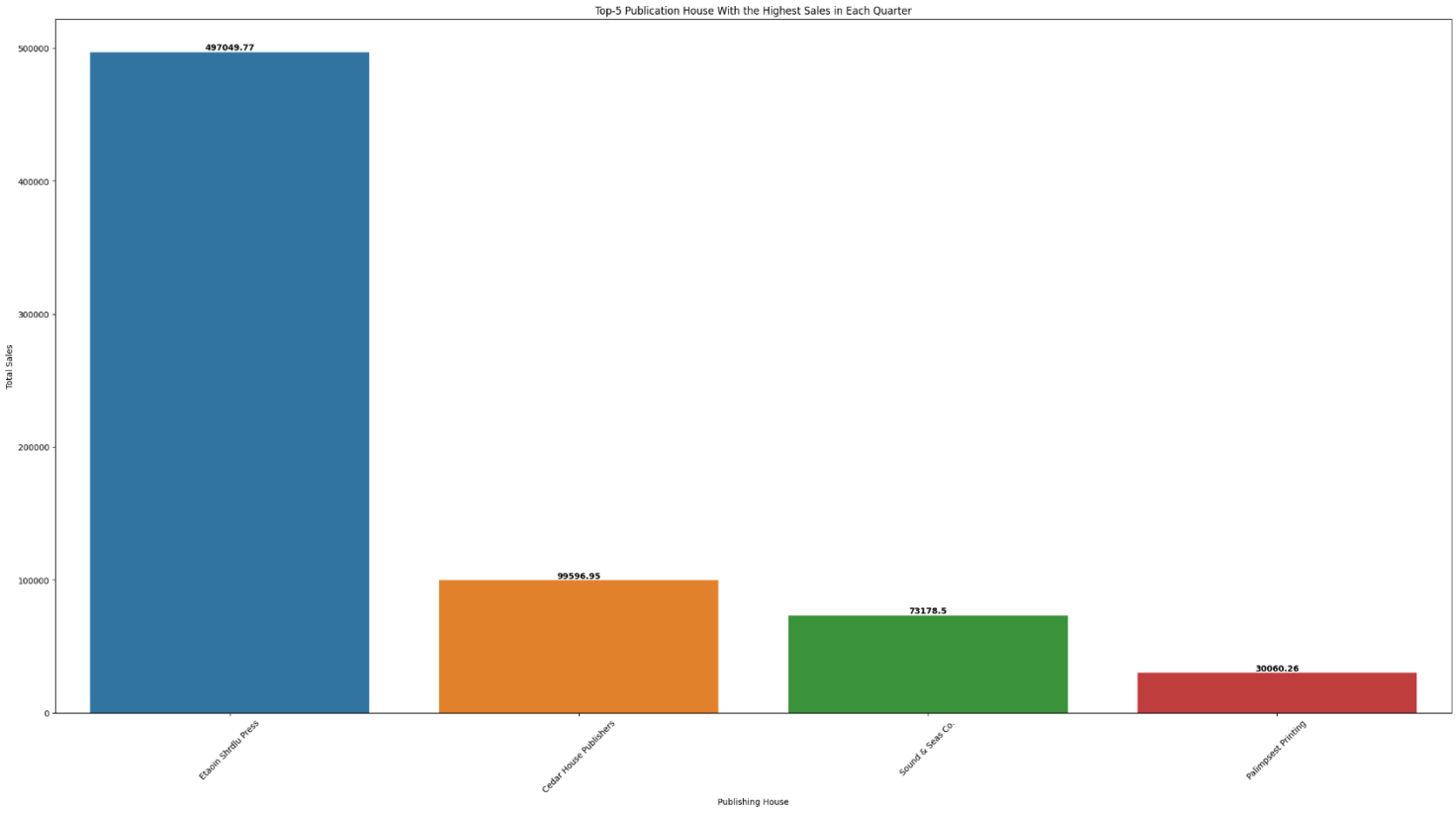
11. Top-5 publication house with the highest sales in each quarter

Code:-

```
DS11 =
pd.concat([Sales_Q1.value_counts("ISBN").reset_index(),Sales_Q2.value_counts("ISBN").reset_index(),Sales_Q3.va
lue_counts("ISBN").reset_index(),Sales_Q4.value_counts("ISBN").reset_index(),],ignore_index=True)
DS11 = pd.merge(DS11, Edition, on="ISBN")
DS11["Total Price"] = DS11["count"] * DS11["Price"]
DS11 = pd.merge(DS11, Publisher, on="PubID")
df11 = pd.DataFrame(DS11.groupby("Publishing House")["Total Price"].sum()).reset_index().sort_values(by = "Total
```

```
Price" , ascending= False)

plt.figure(figsize=(30, 15))
sns.barplot(
   x=df11["Publishing House"], y=df11["Total Price"], hue=df11["Publishing House"]
)

for i, value in enumerate(df11["Total Price"]):
   plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")

plt.xlabel("Publishing House")
plt.ylabel("Total Sales")
plt.title("Top-5 Publication House With the Highest Sales in Each Quarter")
plt.xticks(rotation=45)
plt.show()
```

Output :-



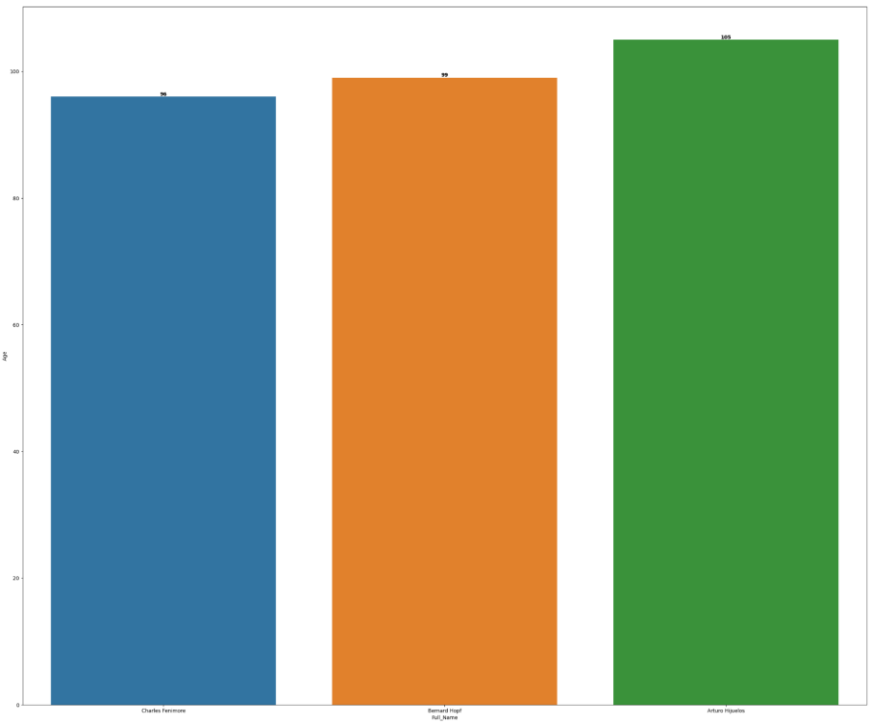
Marwadi University	**Marwadi University** **Faculty of Engineering and Technology** **Department of Information and Communication Technology**	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

12. Top-3 youngest authors

Code :-

```
from datetime import datetime
df12 = Author
def calculate_age(birthdate):
    current_date = datetime.now()
    age = (current_date.year- birthdate.year- ((current_date.month, current_date.day) <
      (birthdate.month,birthdate.day)))
    return age

df12["Age"] = (df12["Birthday"].apply(calculate_age))*-1
df12 = df12[["Full_Name", "Age"]]
df12 = df12.sort_values(by="Age" , ascending=True)[:3]

plt.figure(figsize=(30, 25))
sns.barplot(data=df12, x="Full_Name", y="Age", hue="Full_Name")
for i, value in enumerate(df12["Age"]):
    plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")
plt.show()
```

Output :-



Marwadi University	**Marwadi University** **Faculty of Engineering and Technology** **Department of Information and Communication Technology**	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

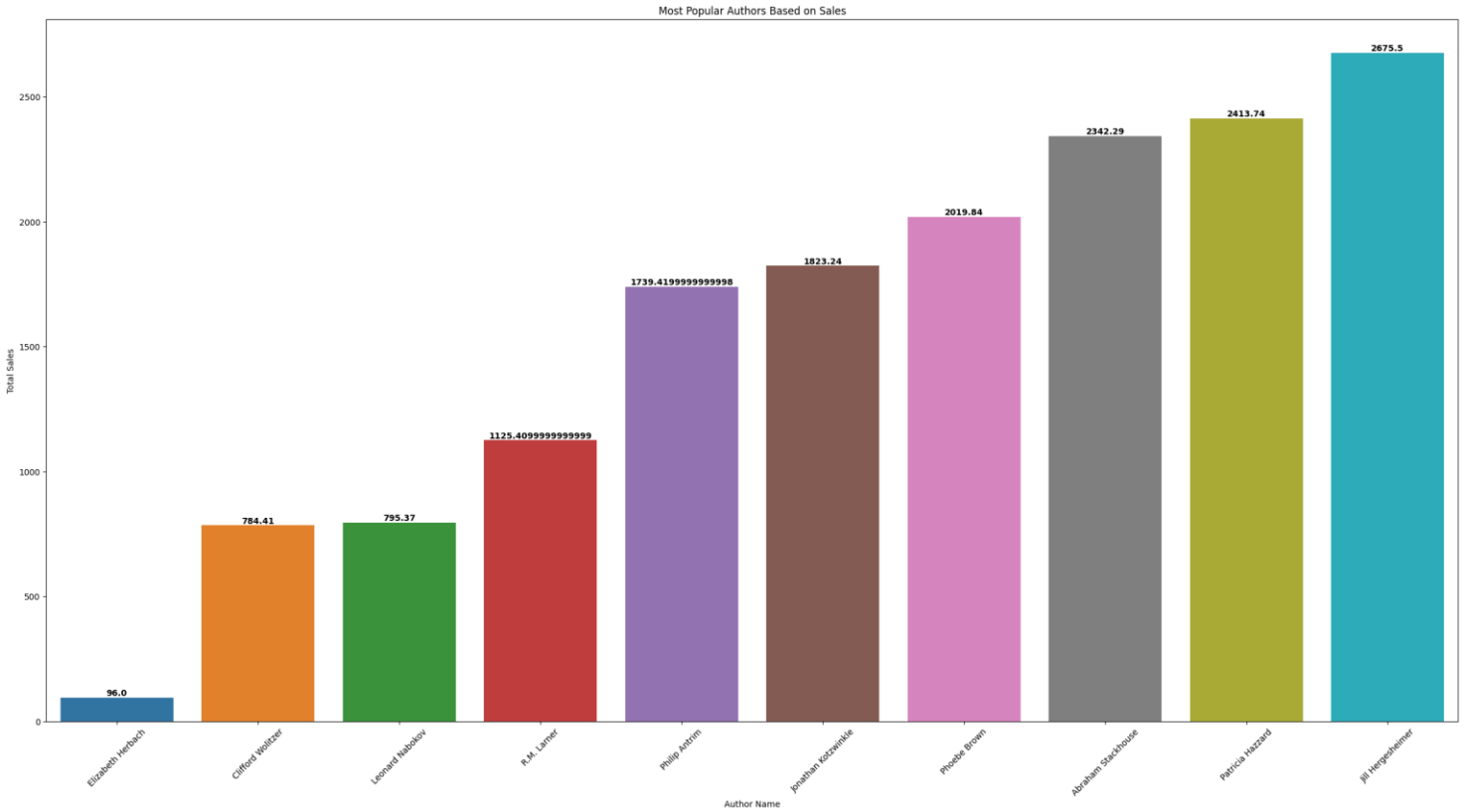
13. Whose (Author) book is least read by the readers

Code :-

```
Sales =
    pd.concat([Sales_Q1.value_counts("ISBN").reset_index(),Sales_Q2.value_counts("ISBN").reset_index(),Sales
    _Q3.value_counts("ISBN").reset_index(),Sales_Q4.value_counts('ISBN').reset_index(),],ignore_index=True,)
Sales = pd.merge(Sales , Edition , on="ISBN")
Sales['Total Price'] = Sales['count'] * Sales['Price']
Sales = Sales[['ISBN' , 'Total Price']].sort_values(by='Total Price' , ascending=False)

DS1A = pd.merge(Book, Edition, on="BookID")
DS1A = pd.merge(DS1A, Author, on="AuthID")
DS1A = pd.merge(DS1A, Sales , on="ISBN")
DS1A = DS1A.groupby('Full_Name')['Total Price'].sum().reset_index().sort_values(by="Total Price",
    ascending=True)[:10]
df1a = DS1A

plt.figure(figsize=(30, 15))
sns.barplot(x=df1a["Full_Name"], y=df1a["Total Price"], hue=df1a["Full_Name"])

for i, value in enumerate(df1a["Total Price"]):
   plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")

plt.xlabel("Author Name")
plt.ylabel("Total Sales")
plt.title("Most Popular Authors Based on Sales")
plt.xticks(rotation=45)
plt.show()
```

Marwadi University	**Marwadi University** **Faculty of Engineering and Technology** **Department of Information and Communication Technology**	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

Output :-



14. Average price of the books published by top-5 authors (in terms of their published books)

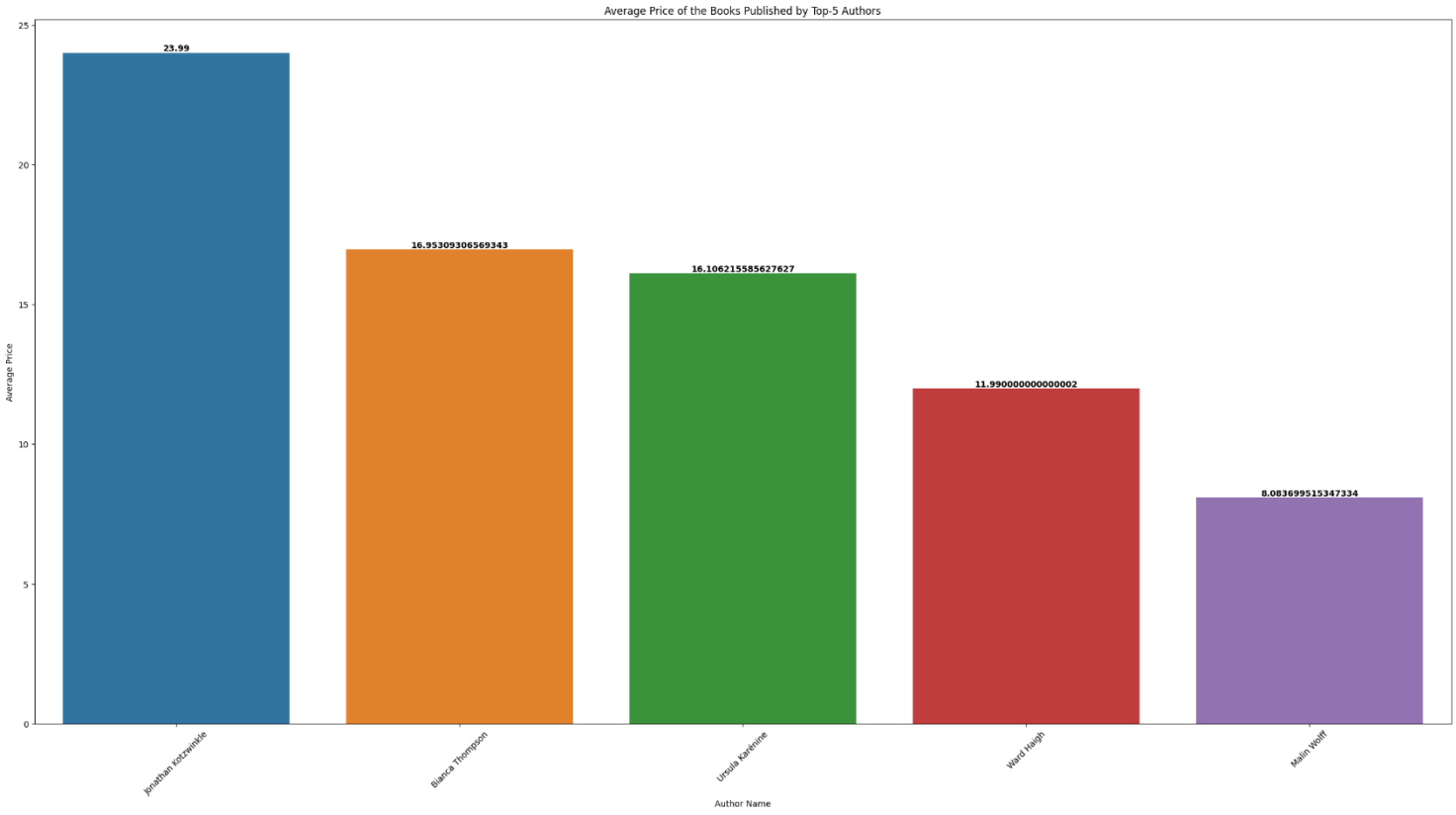
Code :-

```
DS14 = pd.merge(Author, Book, on="AuthID")
DS14 = pd.merge(DS14, Ratings, on="BookID")
DS14 = pd.merge(DS14, Edition, on="BookID")
DS14["Full_Name"] = DS14["First Name"] + " " + DS14["Last Name"]

Top_Authors = list(
  (
    DS14.groupby("Full_Name")["Rating"]
    .mean()
    .reset_index()
    .sort_values(by="Rating", ascending=False)
  )["Full_Name"][:5]
)

DS14 = DS14[DS14["Full_Name"].isin(Top_Authors)]
df14 = pd.DataFrame(DS14.groupby("Full_Name")['Price'].mean()).reset_index().sort_values(by = 'Price' ,
    ascending=False)
```

Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

```
plt.figure(figsize=(30, 15))
sns.barplot(x=df14["Full_Name"], y=df14["Price"], hue=df14["Full_Name"])

for i, value in enumerate(df14["Price"]):
    plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")

plt.xlabel("Author Name")
plt.ylabel("Average Price")
plt.title("Average Price of the Books Published by Top-5 Authors")
plt.xticks(rotation=45)
plt.show()
```

Output :-



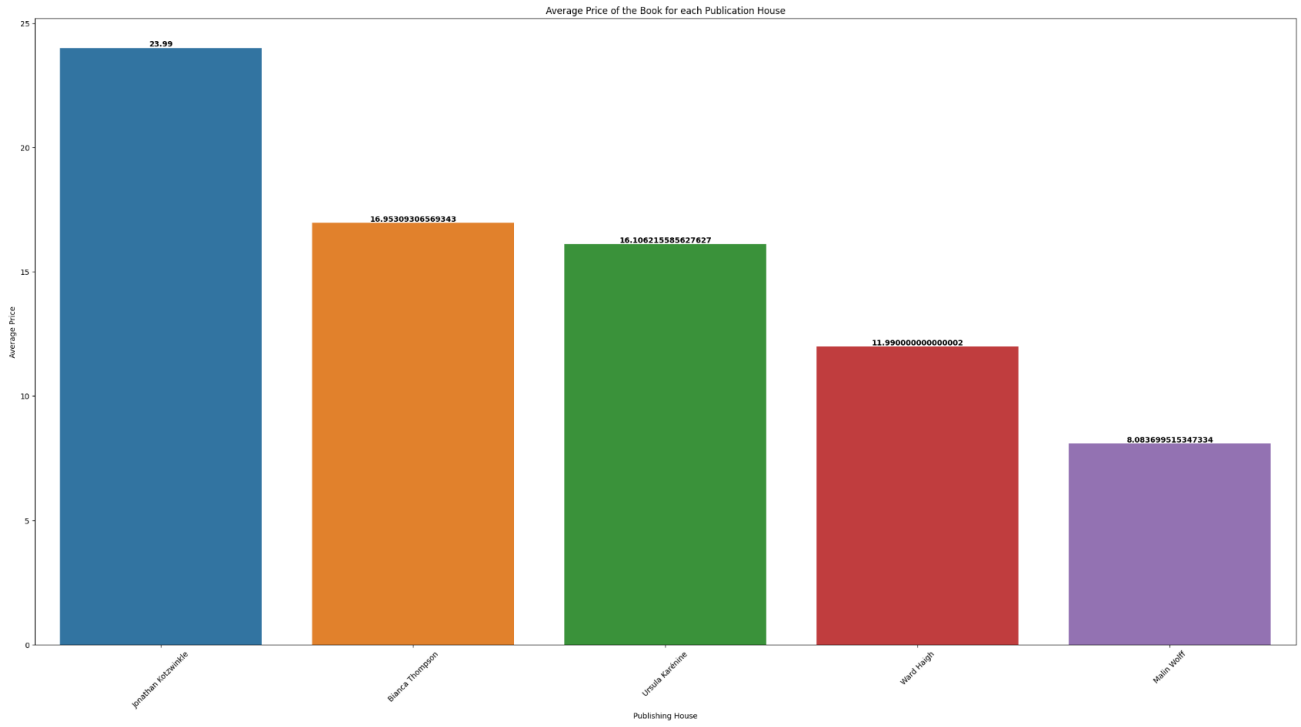
Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

15. Average price of the books for each publication house

Code :-

```
DS15 = pd.merge(Book  ,Edition , on = "BookID")
DS15 = pd.merge(DS15, Publisher, on="PubID")
df15 = (pd.DataFrame(DS15.groupby("Publishing House")["Price"].mean()).reset_index().sort_values(by="Price",
    ascending=False))

plt.figure(figsize=(30, 15))
sns.barplot(x=df14["Full_Name"], y=df14["Price"], hue=df14["Full_Name"])

for i, value in enumerate(df14["Price"]):
  plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")

plt.xlabel("Publishing House")
plt.ylabel("Average Price")
plt.title("Average Price of the Book for each Publication House")
plt.xticks(rotation=45)
plt.show()
```

Output :-



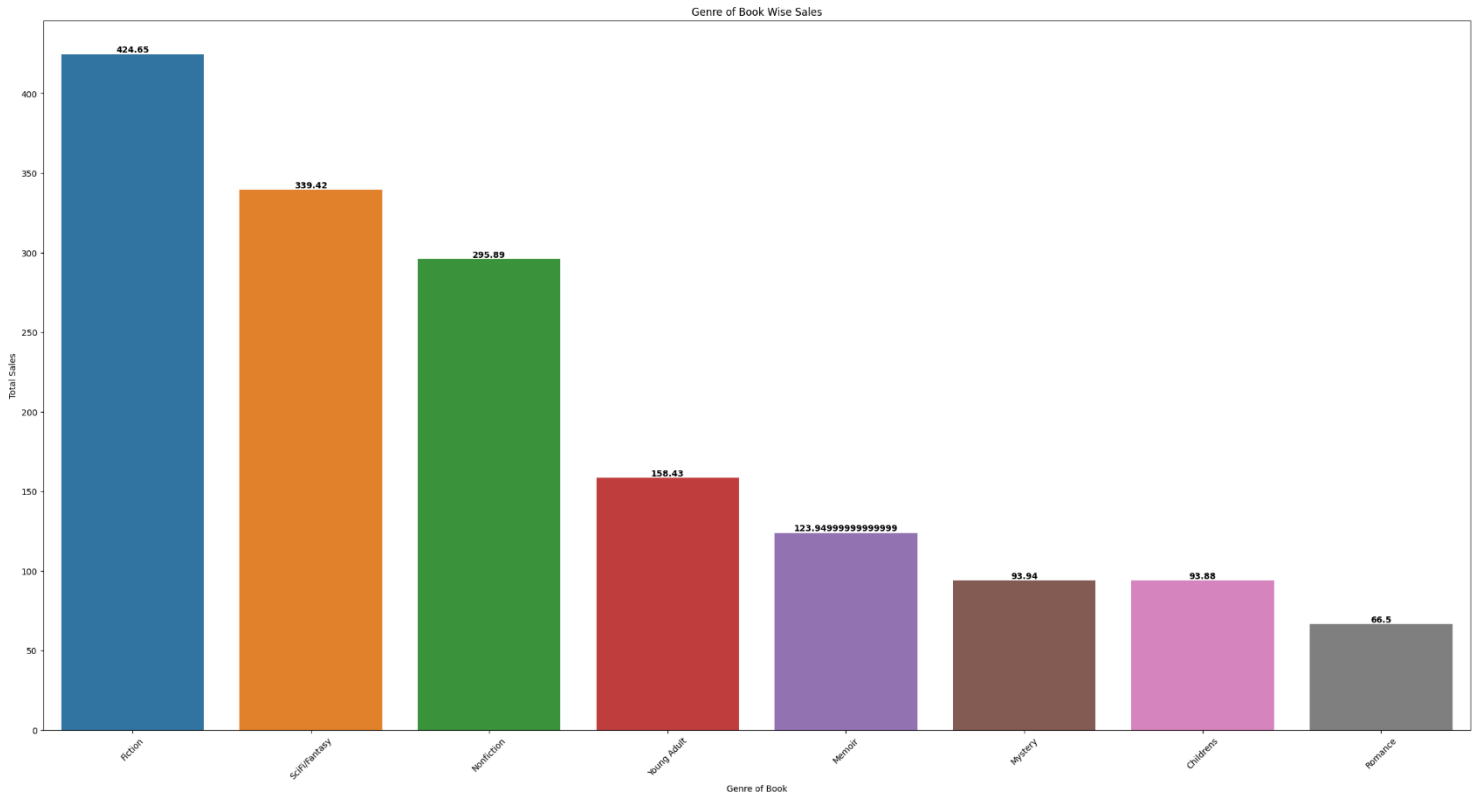
Marwadi University	**Marwadi University** **Faculty of Engineering and Technology** **Department of Information and Communication Technology**	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

16. Which genre of the book has the highest sales

Code :-

```
DS16 = pd.merge(Info, Edition, on="BookID")
df16 = (pd.DataFrame(DS16.groupby("Genre")["Price"].sum()).reset_index().sort_values(by="Price" , ascending=
    False))

plt.figure(figsize=(30, 15))
sns.barplot(x=df16["Genre"], y=df16["Price"], hue=df16["Genre"])

for i, value in enumerate(df16["Price"]):
  plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")

plt.xlabel("Genre of Book")
plt.ylabel("Total Sales")
plt.title("Genre of Book Wise Sales")
plt.xticks(rotation=45)
plt.show()
```

Output :-



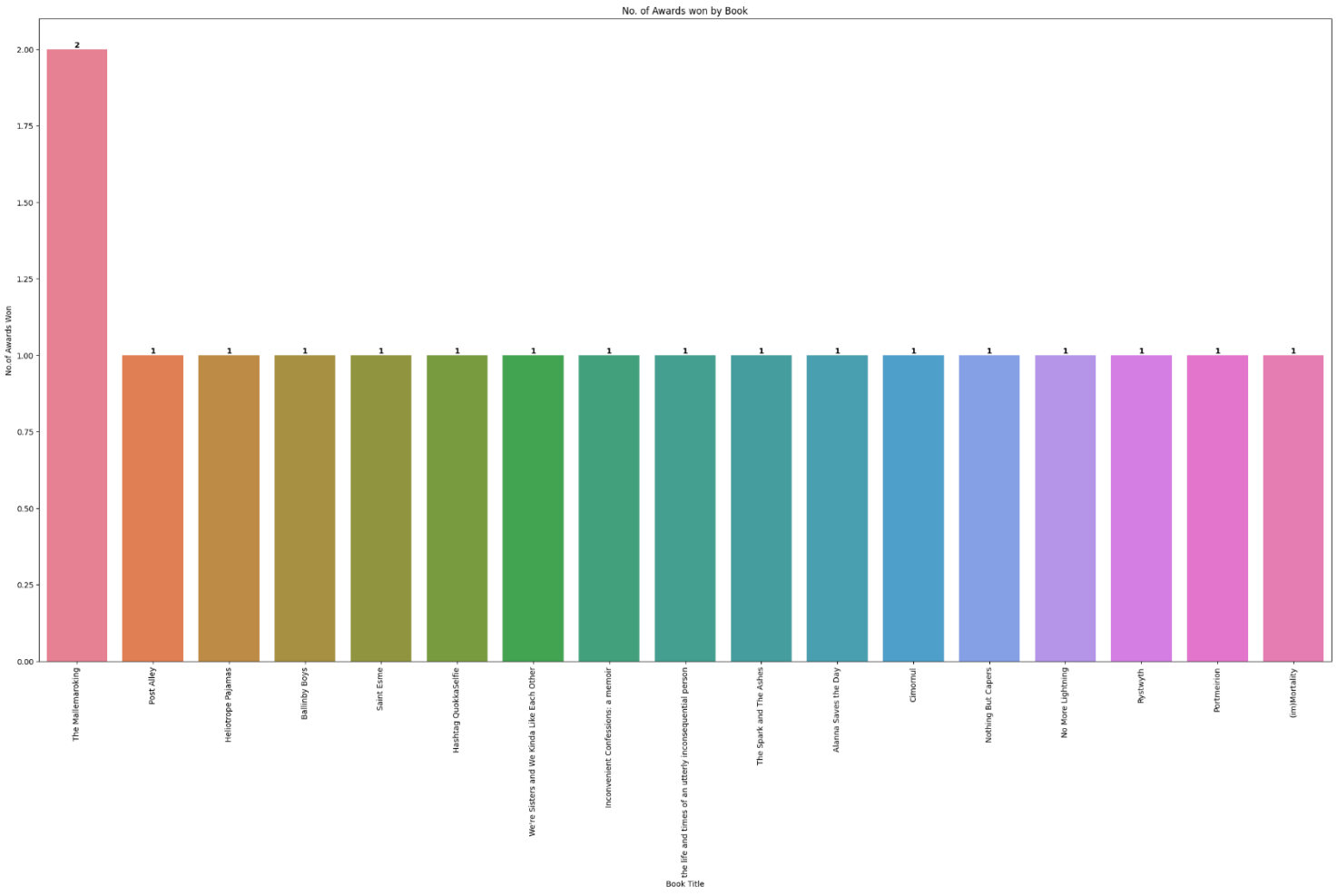
Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

17. Top 5 books having won the maximum number of awards

Code :-

```
DS17 = Award["Title"].value_counts().reset_index()
plt.figure(figsize=(30, 15))
sns.barplot(x=DS17["Title"], y=DS17["count"], hue=DS17["Title"])
for i, value in enumerate(DS17["count"]):
   plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")

plt.xlabel("Book Title")
plt.ylabel("No.of Awards Won")
plt.title("No. of Awards won by Book")
plt.xticks(rotation=90)
plt.show()
```

Output :-



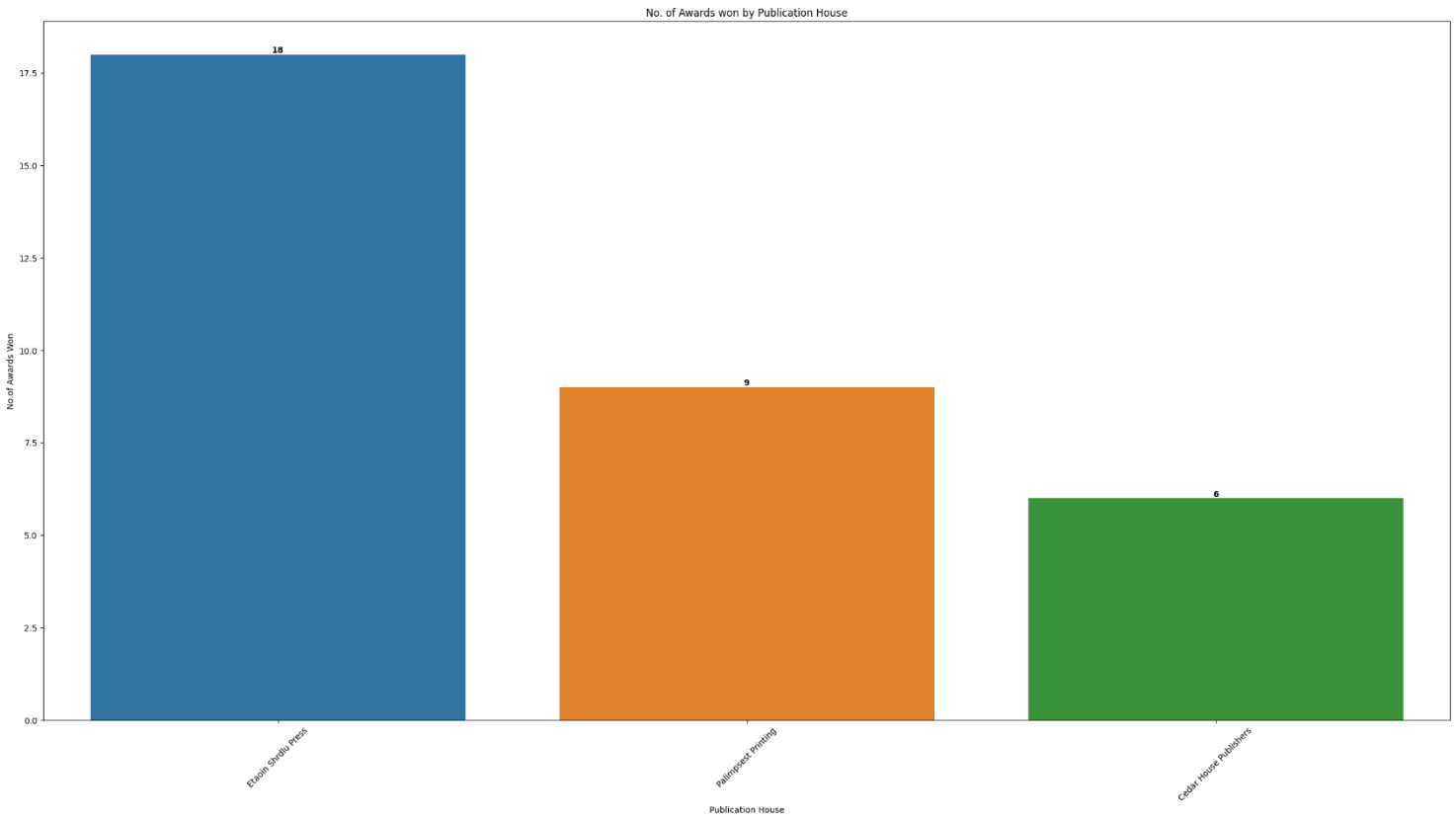
Marwadi University	**Marwadi University** **Faculty of Engineering and Technology** **Department of Information and Communication Technology**	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

18. Top-3 publication houses having won the maximum number of awards

Code :-

```
DS18 = pd.merge(Book , Award , on="Title")
DS18 = pd.merge(DS18 , Edition , on="BookID")
DS18 = pd.merge(DS18, Publisher, on="PubID")
df18 = (DS18["Publishing House"].value_counts()).reset_index().sort_values(by = "count"  ,ascending= False)[:3]
plt.figure(figsize=(30, 15))
sns.barplot(x=df18["Publishing House"], y=df18["count"], hue=df18["Publishing House"])
for i, value in enumerate(df18["count"]):
    plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")
plt.xlabel("Publication House")
plt.ylabel("No.of Awards Won")
plt.title("No. of Awards won by Publication House")
plt.xticks(rotation=45)
plt.show()
```

Output :-



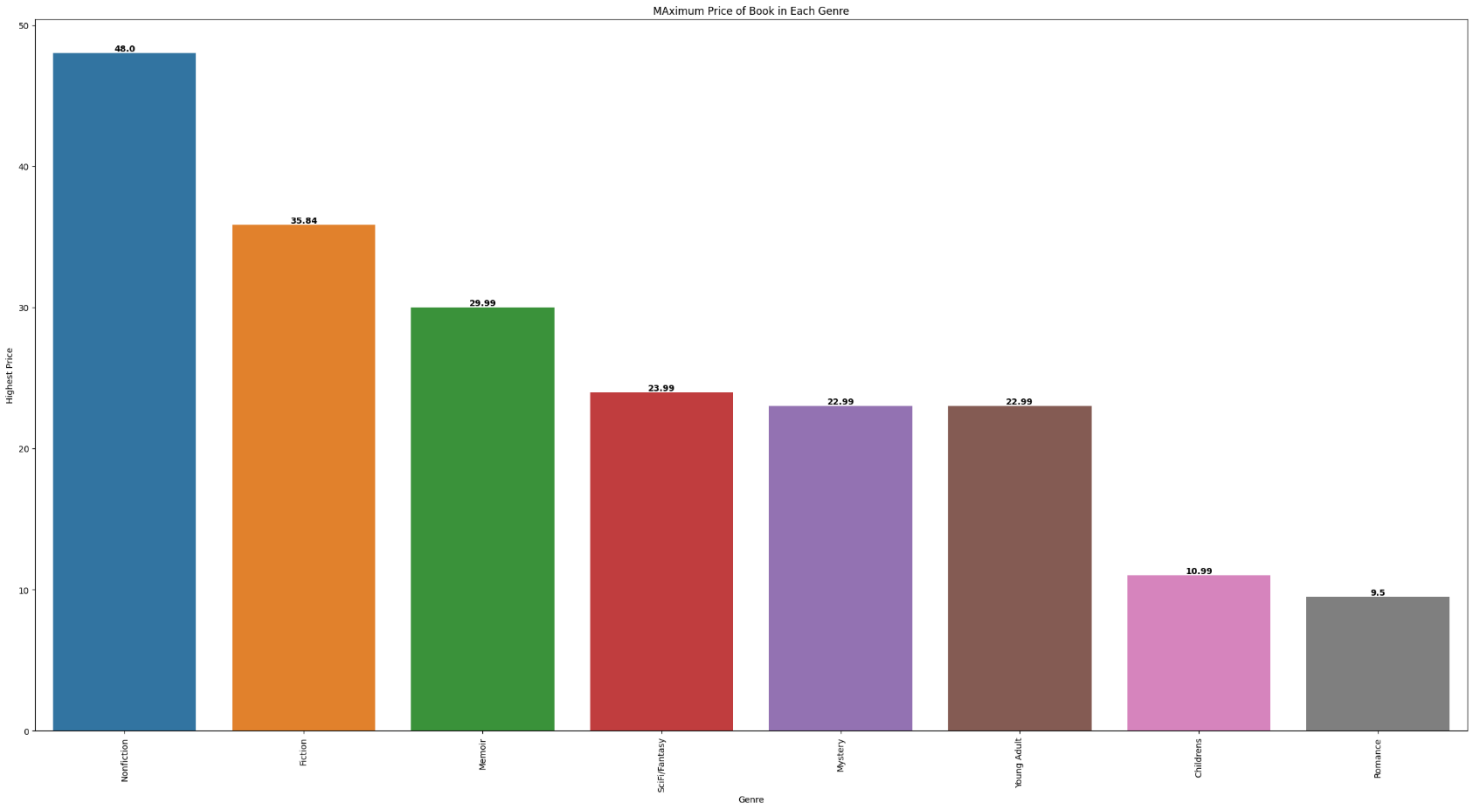
Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

19. Which genre of the books has the highest price

Code :-

```
DS19 = pd.merge(Info, Edition, on="BookID")
df19 = pd.DataFrame(DS19.groupby("Genre")["Price"].max()).sort_values(by="Price" , ascending=
    False).reset_index()

plt.figure(figsize=(30, 15))
sns.barplot(x=df19["Genre"], y=df19["Price"], hue=df19["Genre"])

for i, value in enumerate(df19["Price"]):
  plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")

plt.xlabel("Genre")
plt.ylabel("Highest Price")
plt.title("MAximum Price of Book in Each Genre")
plt.xticks(rotation=90)
plt.show()
```

Output :-



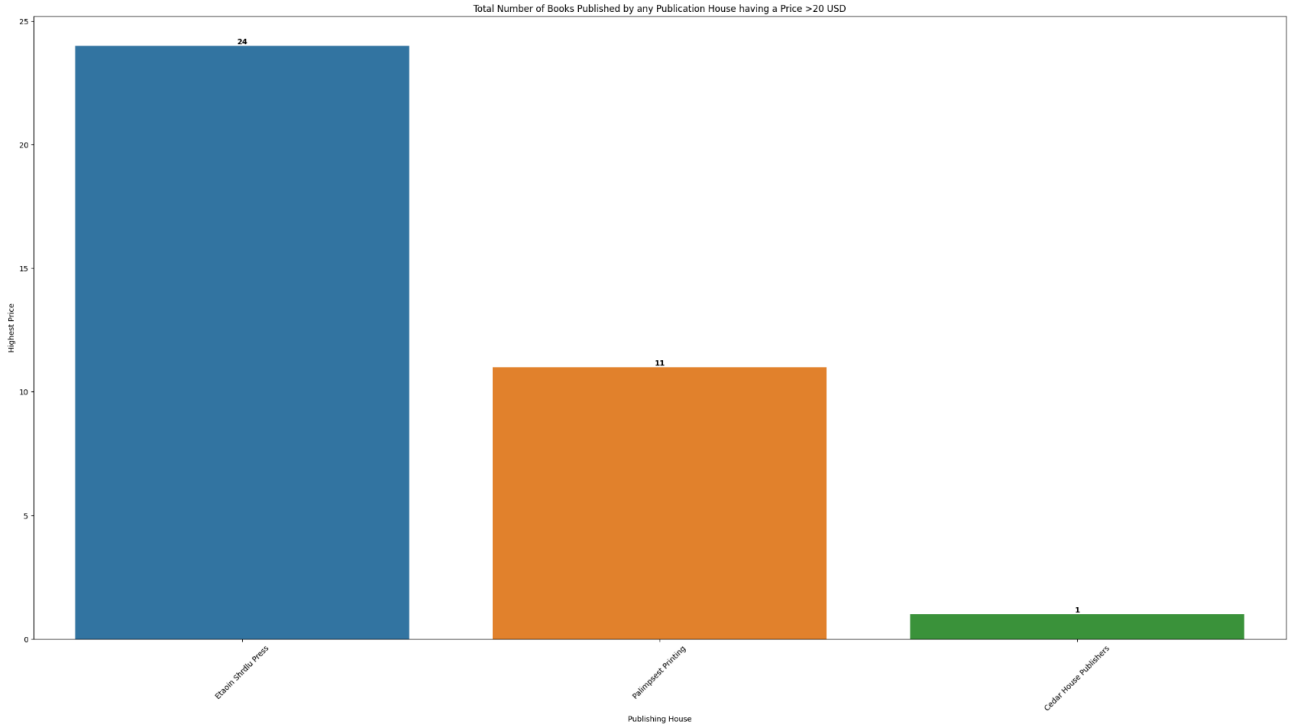
Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

20. Total number of books published by any publication house having a price >20 USD

Code :-

```
DS20 = pd.merge(Edition, Publisher, on="PubID")
DS20 = DS20[DS20['Price'] > 20.00]
df20 = DS20["Publishing House"].value_counts().reset_index()

plt.figure(figsize=(30, 15))
sns.barplot(x=df20["Publishing House"], y=df20["count"], hue=df20["Publishing House"])

for i, value in enumerate(df20["count"]):
  plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")

plt.xlabel("Publishing House")
plt.ylabel("Highest Price")
plt.title("Total Number of Books Published by any Publication House having a Price >20 USD")
plt.xticks(rotation=45)
plt.show()
```

Output :-



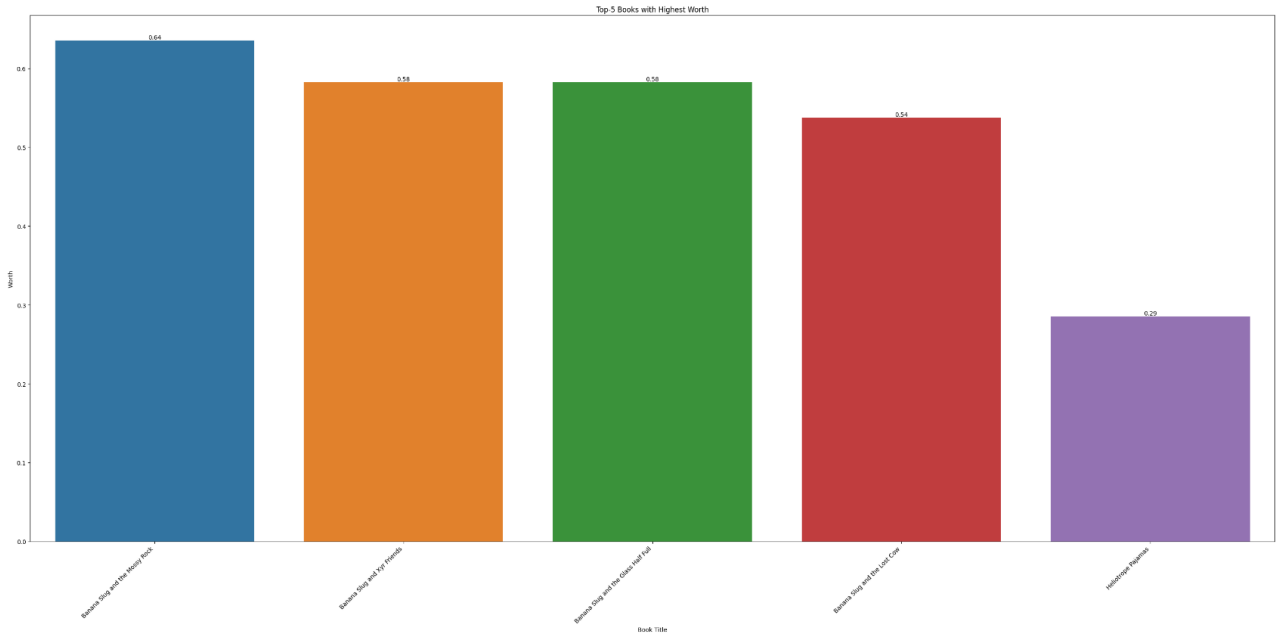
Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

21.Top-5 books with the highest worth (worth=price/number of pages)

Code :-

```
DS21 = pd.merge(Book, Edition, on="BookID")
DS21["Worth"] = DS21["Price"] / DS21["Pages"]
DS21 = DS21.drop_duplicates(subset=["Title"])
DS21 = DS21.sort_values(by="Worth", ascending=False)
df21 = DS21.head(5)

plt.figure(figsize=(30, 15))
sns.barplot(x=df21["Title"], y=df21["Worth"], hue=df21["Title"])

for index, value in enumerate(df21["Worth"]):
    plt.text(index, value, str(round(value, 2)), ha="center", va="bottom", fontsize=10)

plt.xlabel("Book Title")
plt.ylabel("Worth")
plt.title("Top-5 Books with Highest Worth")
plt.xticks(rotation=45, ha="right")
plt.tight_layout()
plt.show()
```

Output :-



Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

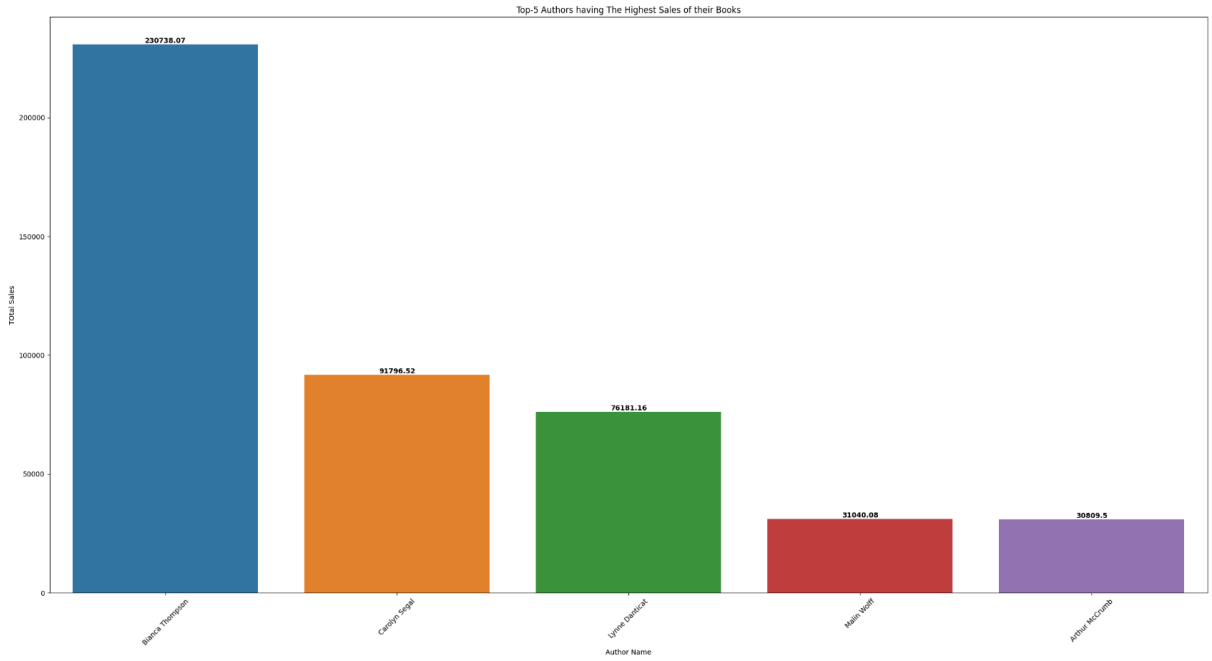
22. Top-5 authors having the highest sales of their books

Code :-

```
DS22 =
    pd.concat([Sales_Q1.value_counts("ISBN").reset_index(),Sales_Q2.value_counts("ISBN").reset_index(),Sales_Q3.
    value_counts("ISBN").reset_index(),Sales_Q4.value_counts("ISBN").reset_index(),],ignore_index=True,)
DS22 = pd.merge(DS22, Edition, on="ISBN")
DS22["Total Price"] = DS22["count"] * DS22["Price"]
DS22 = pd.merge(DS22, Book, on="BookID")
DS22 = pd.merge(DS22, Author, on="AuthID")
DS22 = DS22[["Full_Name", "Total Price"]]
df22 = (
  pd.DataFrame(DS22.groupby("Full_Name")["Total Price"].sum())
  .reset_index()
  .sort_values(by="Total Price" , ascending=False)
)
df22 = df22.head(5)

plt.figure(figsize=(30, 15))
sns.barplot(x=df22["Full_Name"], y=df22["Total Price"], hue=df22["Full_Name"])

for i, value in enumerate(df22["Total Price"]):
  plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")

plt.xlabel("Author Name")
plt.ylabel("TOtal Sales")
plt.title("Top-5 Authors having The Highest Sales of their Books")
plt.xticks(rotation=45)
plt.show()
```

Marwadi University	Marwadi University Faculty of Engineering and Technology Department of Information and Communication Technology	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

Output :-



23. Top-5 books with the highest number of pages

Code :-

```
DS23 = pd.merge(Edition , Book , on = "BookID")
DS23 = pd.merge(DS23, Author, on="AuthID")
DS23 = DS23.drop_duplicates(subset=["Title"])
DS23 = DS23.sort_values(by="Pages" , ascending= False)
df23 = DS23[["Title", "Pages"]].head(5)

plt.figure(figsize=(30, 15))
sns.barplot(x=df23["Title"], y=df23["Pages"], hue=df23["Title"])

for i, value in enumerate(df23["Pages"]):
   plt.text(i, value, str(value), ha="center", va="bottom", weight="bold")

plt.xlabel("Book Title")
plt.ylabel("No.of Pages")
plt.title("Top-5 Books with the Highest Number of Pages")
plt.xticks(rotation=45)
plt.show()
```

Marwadi University	**Marwadi University** **Faculty of Engineering and Technology** **Department of Information and Communication Technology**	
Subject: Data Visualization and Dashboard (01CT0410)	**Aim: Exploratory Data Analysis (EDA) using Python**	
Experiment No: 12	**Date: 30-03-2024**	**Enrollment No: 92200133030**

Output :-