DATA ANALYSIS – DIWALI SALES (PYTHON)

1. Import the file
2. Data Cleaning

* Drop empty columns
* Check for Null values in each column
* Replace null values with a values
* Check and change datatype as required
* Rename column

1. Data Analysis

* Count the number of Male and Female with Labels
* Give the sales by gender and sort in descending order & plot bar graph (with Lables)
* Using group by + using Count plot find the number for each Age groups by their Gender (With Label)
* Sales by Age group with Graph
* Total orders from Top 10 states
* Top 10 sales by states
* Count of Martial Status with graph
* Sales by Martial Status & Gender + graph
* Count of orders by occupation
* Sales by occupation
* Count by prod category
* Sales by prod category

**CODES**

1. IMPORT FILE
2. import pandas as pd
3. df = pd.read\_csv('/content/drive/MyDrive/dataset/Diwali Sales Data.csv', encoding='unicode\_escape')
4. df.head()

2. Dropped blank columns

df.info()

df.drop(columns =['Status', 'unnamed1'], inplace = True)

3. Finding and replacing Nulls

df.isnull().sum()

repvalue = df['Amount'].mean()

df['Amount'].fillna(repvalue, inplace = True)

4. Rename

df.rename(columns = {'Amount': 'Sales'})

1. Check Data Type
2. df['Sales'] = df['Sales'].astype(int)

**Data Analysis**

**1. Count the number of Male and Female with Labels**

G = df.groupby('Gender')['Gender'].count()

G = pd.DataFrame(G)

G = G.rename(columns = {'Gender': 'Gcount'})

G = G.sort\_values(by = ['Gcount'], ascending = [False])

import seaborn as sns

var = sns.countplot(x= df['Gender'], palette ='viridis')

for bars in var.containers:

  var.bar\_label(bars)

**2. Give the sales by gender and sort in descending order & plot bar graph (with Lables)**

xx = df.groupby('Gender')['Sales'].sum().reset\_index()

var1 = sns.barplot(x=xx['Gender'], y = xx['Sales'], palette = 'viridis')

for bars in var1.containers:

  var1.bar\_label(bars)

**3. Using group by + using Count plot find the number for each Age groups by their Gender (With Label)**

yy = df.groupby(['Age Group','Gender'])['Sales'].sum().reset\_index()

import matplotlib.pyplot as plt

plt.figure(figsize=(15,7))

var2 = sns.barplot(x = yy['Age Group'], y =yy['Sales'], hue = yy['Gender'])

for bars in var2.containers:

  var2.bar\_label(bars)

# PER TRANSACTION VALUE SPEND BY GENDER TYPE IN EACH AGE GROUP

plt.figure(figsize=(15,7))

z = sns.barplot(x = df['Age Group'], y =df['Sales'], hue = df['Gender'], palette = 'viridis')

for bars in z.containers:

  z.bar\_label(bars)

1. **Total orders from Top 10 states**
2. zz = df.groupby('State')['Orders'].count().reset\_index()
3. zz = zz.sort\_values(by = 'Orders', ascending = False).head(10)

plt.figure(figsize=(15,7))

var3 = sns.barplot(x = zz['State'], y = zz['Orders'],palette = 'viridis')

for bars in var3.containers:

  var3.bar\_label(bars)

**5.Total orders from Top 10 states**

zzz = df.groupby('State')['Sales'].sum().reset\_index()

zzz = zzz.sort\_values(by = 'Sales', ascending = False).head(10)

plt.figure(figsize=(15,7))

var4 = sns.barplot(x = zzz['State'], y = zzz['Sales'],palette = 'RdYlBu')

for bars in var4.containers:

  var4.bar\_label(bars)

**6.Count of orders by Martial Status with graph**

ww = df.groupby(['Marital\_Status', 'Gender'])['Orders'].sum().reset\_index().sort\_values(by = 'Orders', ascending = False)

var5 = sns.barplot(x = ww['Marital\_Status'], y = ww['Orders'], hue = ww['Gender'], palette = 'Greys')

for bars in var5.containers:

  var5.bar\_label(bars)

**7.Sum of Sales by Martial Status with graph**

vv = df.groupby(['Marital\_Status', 'Gender'])['Sales'].sum().reset\_index().sort\_values(by = 'Sales', ascending = False)

var6 = sns.barplot(x = vv['Marital\_Status'], y = vv['Sales'], hue = vv['Gender'], palette = 'PRGn')

for bars in var6.containers:

  var6.bar\_label(bars)

**8. Count of orders by occupation**

uu = df.groupby(['Occupation', 'Gender'])['Orders'].sum().reset\_index().sort\_values(by = 'Orders', ascending = False)

uu

var7 = sns.relplot(x = uu['Occupation'], y = uu['Orders'], kind = 'line', hue = uu['Gender'], palette = 'Dark2', height = 7, aspect = 2.2)

**9. Sum of Sales by occupation**

df1 = df.groupby(['Occupation', 'Gender'])['Sales'].sum().reset\_index()

df1 = df1.sort\_values(by = 'Sales', ascending = False)

plt.figure(figsize = (17,7))

var8 = sns.barplot(x = df1['Occupation'],y = df1['Sales'], hue = df1['Gender'],palette = 'Dark2')

for bars in var8.containers:

  var8.bar\_label(bars)

**10. Count by prod category**

df2 = df.groupby(['Product\_Category', 'Gender'])['Orders'].sum().reset\_index().sort\_values(by = 'Orders', ascending = False )

sns.relplot(x = 'Product\_Category', y = 'Orders', data = df2, kind = 'line', hue = 'Gender', palette = 'Pastel1', height = 8, aspect = 3 )

**11. Count by prod category**

df3 = df.groupby(['Product\_Category', 'Gender'])['Sales'].sum().reset\_index().sort\_values(by = 'Sales', ascending = False )

plt.figure(figsize=(20,8))

sns.barplot(x = 'Product\_Category', y = 'Sales', data = df3, hue = 'Gender', palette = 'Accent')

**Insights from the Data**

1. **Top Spending by Gender**:
   * The average spending by customers shows significant engagement across different genders. Analyzing spending patterns can highlight key gender-based trends.
2. **Age Group Analysis**:
   * The majority of customers fall within the 26-35 age group, indicating that this demographic is highly engaged in Diwali sales. Targeted marketing strategies for this age group can enhance sales further.
3. **Marital Status**:
   * A substantial number of customers are unmarried, suggesting potential for different promotional strategies tailored for singles versus married individuals.
4. **Geographical Insights**:
   * Significant sales data from various states and zones can help in region-specific marketing campaigns. States like Maharashtra and Andhra Pradesh are major contributors.
5. **Occupation-Based Spending**:
   * Occupations such as Healthcare, Govt, and Automobile show high sales numbers, indicating strong engagement from these sectors. Customized offers for these professionals might drive more sales.
6. **Product Category Preferences**:
   * Product categories such as 'Auto' are popular among the customers. Understanding product preferences can guide inventory management and promotional efforts.

**Recommendations**

1. **Targeted Marketing**:
   * Develop targeted marketing campaigns focusing on the 26-35 age group and unmarried individuals, as they represent a significant portion of the customer base.
2. **Gender-Specific Promotions**:
   * Implement gender-specific promotions and discounts to leverage spending patterns observed across different genders.
3. **Regional Campaigns**:
   * Design region-specific promotional strategies for top-performing states like Maharashtra and Andhra Pradesh to maximize regional sales.
4. **Occupation-Based Offers**:
   * Create special offers and discounts for professionals in the Healthcare, Govt, and Automobile sectors to encourage more purchases from these high-spending occupations.
5. **Enhance Product Inventory**:
   * Ensure adequate inventory for popular product categories such as 'Auto' to meet customer demand during peak sales periods.