**---DIWALI\_SALES\_ANALYSIS---**

**import** numpy **as** np

**import** pandas **as** pd

**import** matplotlib.pyplot **as** plt

**import** seaborn **as** sns

**BASIC DATA CLEANING----**

***# method1:-*** *use r before address to avoid special character error and encoding='unicode\_escape'*

df**=**pd**.**read\_csv(r'C:\Users\rohit\Desktop\py projects\Diwali Sales Data.csv',encoding**=**'unicode\_escape')

df**.**shape

**#method2:-**

df**=**pd**.**read\_csv('Diwali Sales Data.csv',encoding**=**'unicode\_escape')

***# get top data------***

df**.**head()

***# info about data---***

df**.**info()

***# to check if data is null or not---***

pd**.**isnull(df)

***# to remove unrealted or blank columns---***

df**.**drop(['status','unnamed1'],axis**=**1,inplace**=True**)

***# to check for null values, to get total no of null values available---***

pd**.**isnull(df)**.**sum()

***# to delete null values---***

df**.**dropna(inplace**=True**)

***# UNDERSTANDING inplace=True-----***

data\_test**=**[['madhav',11],['kesgav',],['gopi',12],['lalita',16]]

***# create data set----***

df\_test**=**pd**.**DataFrame(data\_test,columns**=**['Name','Age'])

df\_test

df\_test.dropna()

:

*# here we will get keshav data again,concept is we have not saved the file after editing.so we have to use “inplace=True” to save file or we can save it in another dataframe so that any change will saved in that file and we can use it.But it is not advised as it can create multiple files*

df\_test**.**dropna(inplace**=True**)

***# Change data type----***

df['Amount']**=**df['Amount']**.**astype('int')

df['Amount']**.**dtypes

***# rename colums----***

df**.**columns

df**.**rename(columns**=**{'Marital\_Status':'Marital status'})

***# describe()—returns description of numerical values in df i.e mean,count,std***

df**.**describe()

***# to get specific colums ----***

df[['Age']]**.**describe()

# OBJECTIVE----

* *Improve customer experience by analysing sales data.*
* *Increase sales revenue.*
* *Identify sales pattern by state, gender, occupation.*
* *Identify the top and bottom product sale.*

***-----EXPLORATORY DATA ANALYSIS or EDA-----***

**import** numpy **as** np

**import** pandas **as** pd

**import** matplotlib.pyplot **as** plt

**import** seaborn **as** sns

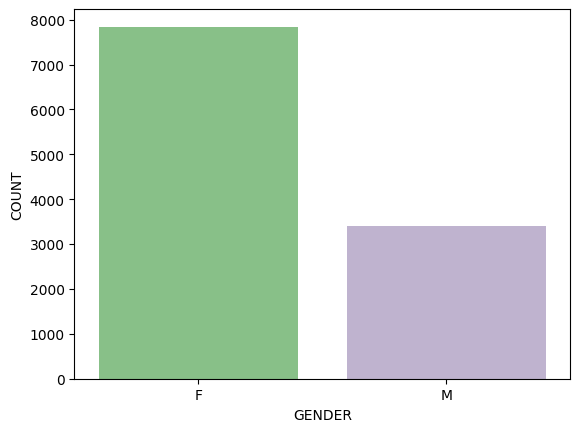
df**=**pd**.**read\_csv(r'C:\Users\rohit\Desktop\py projects\Diwali Sales Data.csv',encoding**=**'unicode\_escape')

cnt**=**sns**.**countplot(x**=**'Gender',data**=**df,palette**=**'Accent')

plt**.**xlabel("GENDER")

plt**.**ylabel("COUNT")

plt**.**show()



**GENDER BASED ANALYSIS----**

In [14]:

***# to get count on each bars---***

cnt**=**sns**.**countplot(x**=**'Gender',data**=**df,palette**=**'Accent')

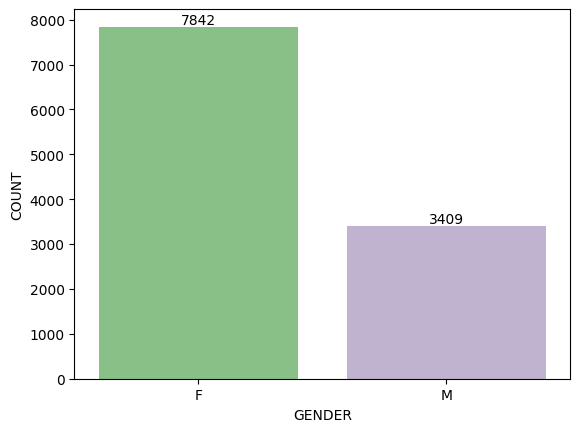
**for** bars **in** cnt**.**containers:

cnt**.**bar\_label(bars)

plt**.**xlabel("GENDER")

plt**.**ylabel("COUNT")

plt.show()

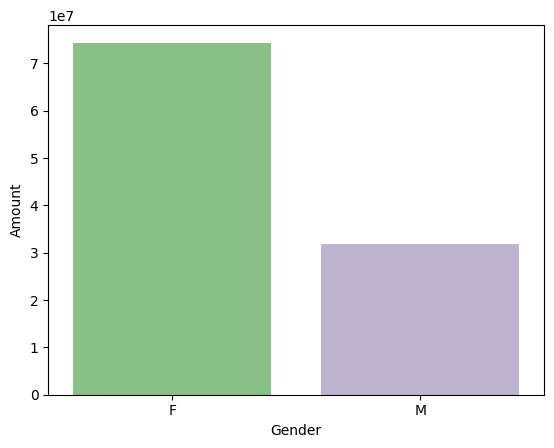


***# to group data based on gender ,sum it based on amount and then sort based on amount---***

sales**=**df**.**groupby(['Gender'],as\_index**=False**)['Amount']**.**sum()**.**sort\_values(by**=**'Amount',ascending**=False**)

sns**.**barplot(x**=**'Gender',y**=**'Amount',data**=**sales,palette**=**'Accent')

plt**.**show()



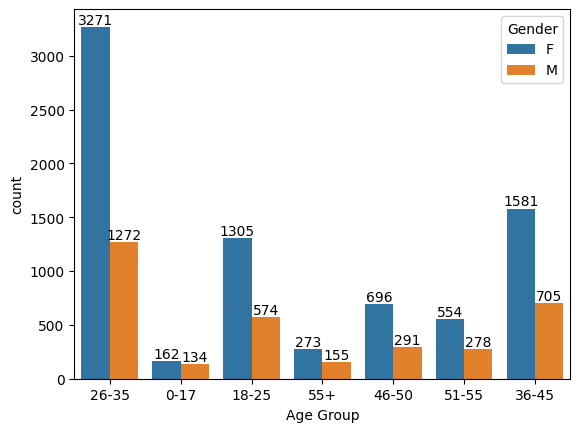
**From above graph it can be interpreted that most buyers are Females and their purchasing power is more than Male.**

**AGE BASED ANALYSIS----**

ag**=**sns**.**countplot(x**=**'Age Group',hue**=**'Gender',data**=**df)

**for** bars **in** ag**.**containers:

ag**.**bar\_label(bars)

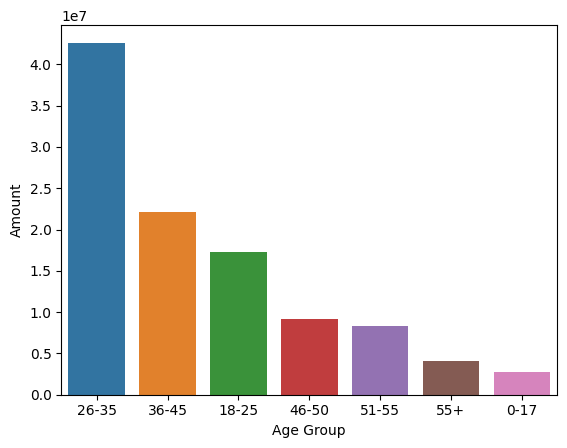


***# total sales vs age group----***

sales\_ag**=**df**.**groupby(['Age Group'],as\_index**=False**)['Amount']**.**sum()**.**sort\_values(by**=**'Amount',ascending**=False**)

sns**.**barplot(x**=**'Age Group',y**=**'Amount',data**=**sales\_ag,hue**=**'Age Group')

plt**.**show()



**From graph it can be interpreted most buyers are in the age group:26-35.**

**STATE BASED ANALYSIS----**

***#sorting based on state,summing on orders and then sorting on based total no of orders-----***

st**=**df**.** groupby([‘State’], as\_index**=False**) [‘Orders’]**.** sum()**.**sort\_values(by**=**’Orders’,ascending**=False**)**.**head(10)

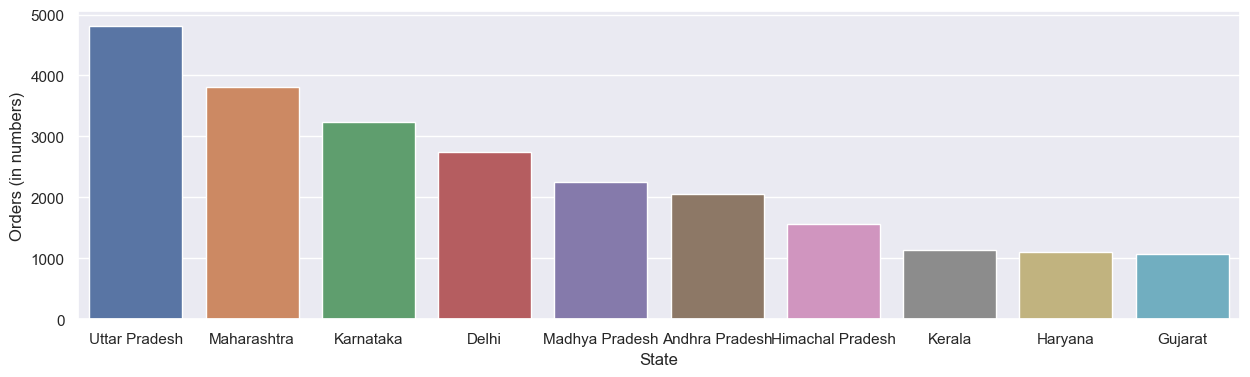
***# setting size of bars to avoid overlapping----'figure.figsize’:(width,height)***

sns**.**set(rc**=**{'figure.figsize':(15,4)})

sns**.**barplot(x**=**'State',y**=**'Orders',data**=**st,hue**=**'State')

plt**.**ylabel('Orders (in numbers)')

plt**.**show()



***#sorting based on state,summing on orders and then sorting on based total no of orders-----***

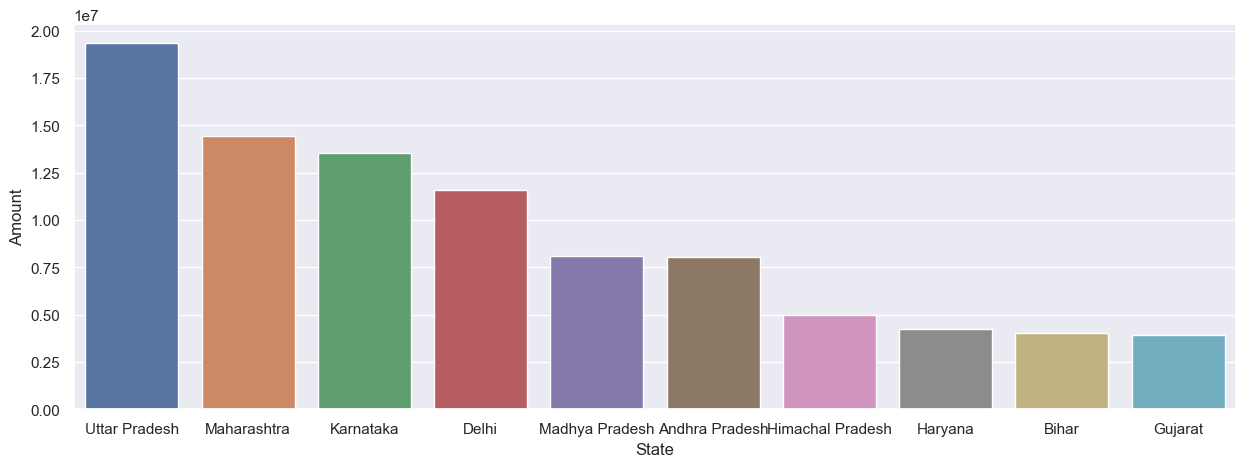
st**=**df**.**groupby(['State'],as\_index**=False**)['Amount']**.**sum()**.**sort\_values(by**=**'Amount',ascending**=False**)**.**head(10)

***# setting size of bars to avoid overlapping----'figure.figsize':(width,height)***

sns**.**set(rc**=**{'figure.figsize':(15,5)})

sns**.**barplot(x**=**'State',y**=**'Amount',data**=**st,hue**=**'State')

plt**.**show()



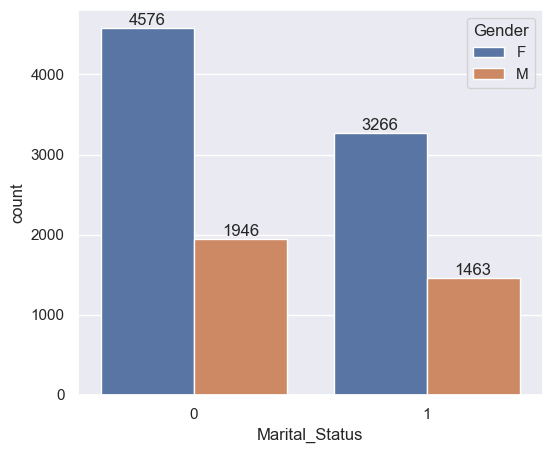
**From above graph it can be interpreted that most of orders and total sales/amount are from Uttar Pradesh, Maharashtra and Karnataka respectively.**

**MARITAL STATUS BASED ANALYSIS----**

ms**=**sns**.**countplot(x**=**'Marital\_Status',data**=**df,hue**=**'Gender')

**for** bars **in** ms**.**containers:

ms**.**bar\_label(bars)



mss**=**df**.** groupby(['Marital\_Status','Gender'], as\_index**=False**) ['Amount']**.**sum()**.**sort\_values(by**=**'Amount',ascending**=False**)

***# setting bars size in plot---***

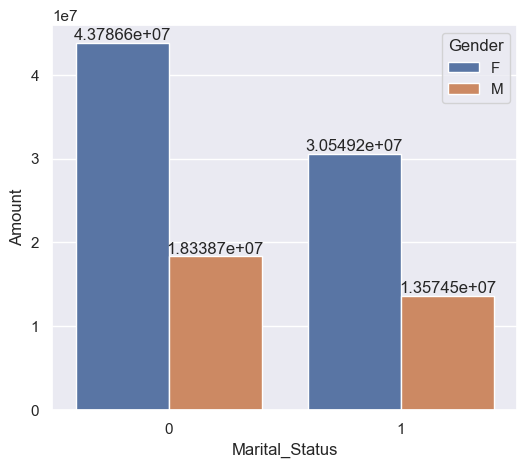
sns**.**set(rc**=**{'figure.figsize':(6,5)})

mvs**=**sns**.**barplot(x**=**'Marital\_Status',y**=**'Amount',data**=**mss,hue**=**'Gender')

**for** bars **in** mvs**.**containers:

mvs**.**bar\_label(bars)

plt**.**show()



**OCCUPATION BASED ANALYSIS----**

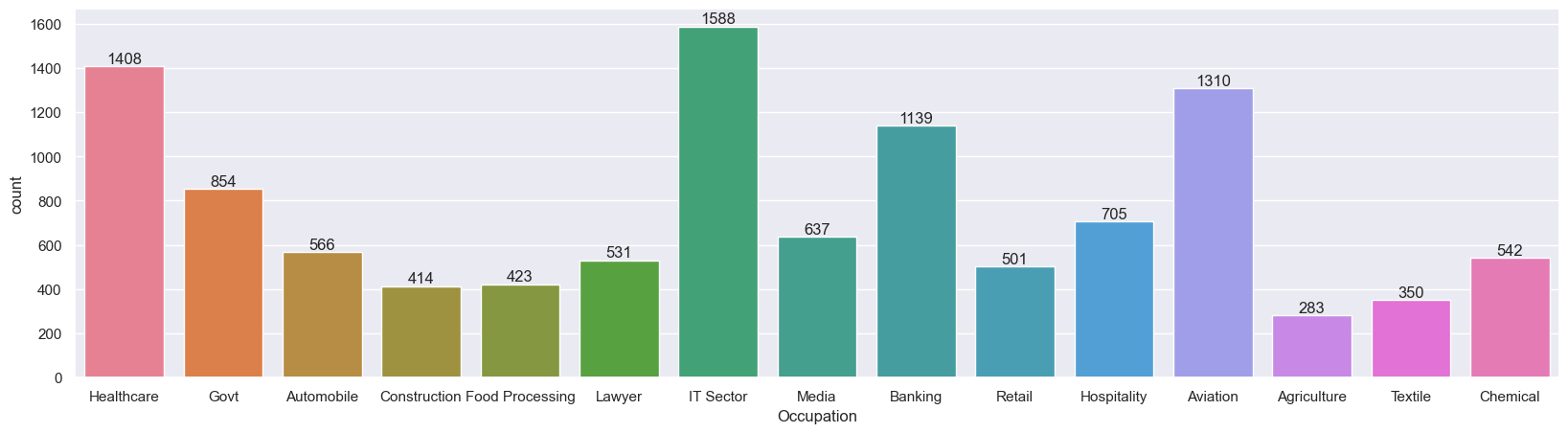
In [83]:

sns**.**set(rc**=**{'figure.figsize':(20,5)})

oc**=**sns**.**countplot(x**=**'Occupation',data**=**df,hue**=**'Occupation')

**for** bars **in** oc**.**containers:

oc**.**bar\_label(bars)



In [15]:

**import** pandas **as** pd

**import** matplotlib.pyplot **as** plt

**import** seaborn **as** sns

df**=**pd**.**read\_csv(r'C:\Users\rohit\Desktop\py projects\Diwali Sales Data.csv',encoding**=**'unicode\_escape')

occ**=**df**.**groupby(['Occupation'],as\_index**=False**)['Amount']**.**sum()**.**sort\_values(by**=**'Amount',ascending**=False**)**.**head(10)

***# set bar size ----***

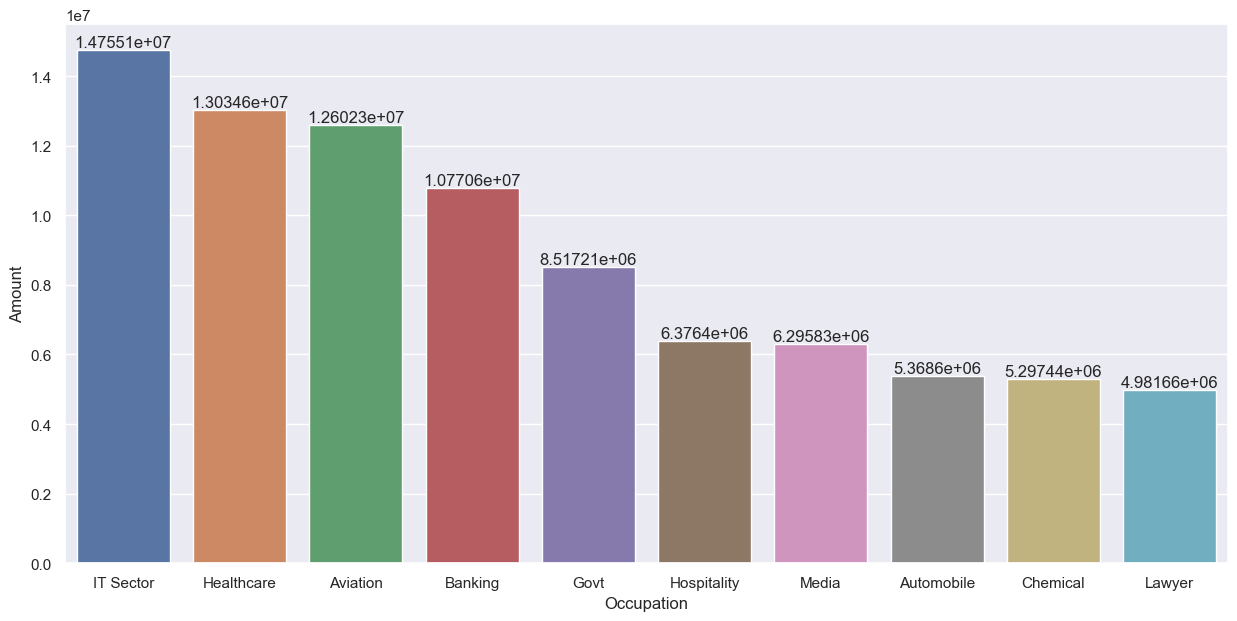
sns**.**set(rc**=**{'figure.figsize':(15,7)})

ocs**=**sns**.**barplot(x**=**'Occupation',y**=**'Amount',data**=**occ,hue**=**'Occupation')

**for** bars **in** ocs**.**containers:

ocs**.**bar\_label(bars)

plt**.**show()



**From graph it can be interpreted that most of the buyers are working in IT,AVIATION AND HEALTHCARE sector.**

***----PRODUCT CATEGORY BASED ANALYSIS----***

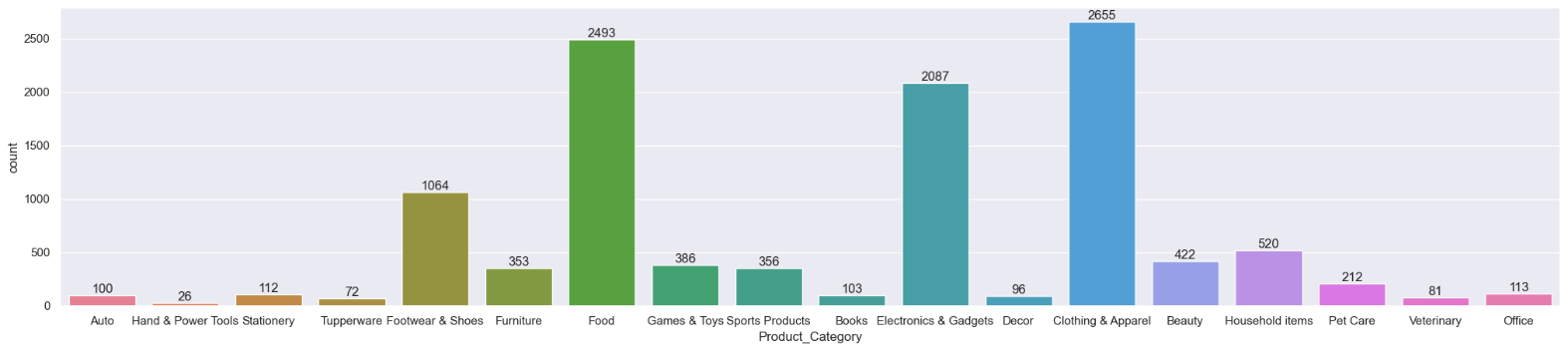
df**=**pd**.**read\_csv(r'C:\Users\rohit\Desktop\py projects\Diwali Sales Data.csv',encoding**=**'unicode\_escape')

sns**.**set(rc**=**{'figure.figsize':(25,5)})

pc**=**sns**.**countplot(x**=**'Product\_Category',data**=**df,hue**=**'Product\_Category')

**for** bars **in** pc**.**containers:

pc**.**bar\_label(bars)



pcs**=**df**.**groupby(['Product\_Category'],as\_index**=False**)['Amount']**.**sum()**.**sort\_values(by**=**'Amount',ascending**=False**)**.**head(10)

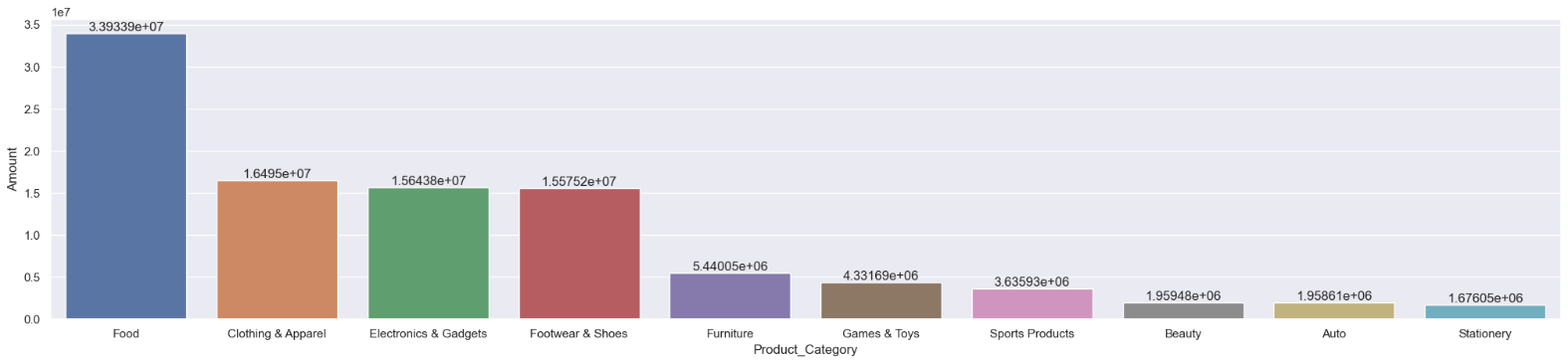
sns**.**set(rc**=**{'figure.figsize':(25,5)})

pcs**=**sns**.**barplot(x**=**'Product\_Category',y**=**'Amount',data**=**pcs,hue**=**'Product\_Category')

**for** bars **in** pcs**.**containers:

pcs**.**bar\_label(bars)

plt**.**show()



**From above it is interpreted that most sold products are from Food,Clothing and Electronics category.**

pcs**=**df**.**groupby(['Product\_ID'],as\_index**=False**)['Orders']**.**sum()**.**sort\_values(by**=**'Orders',ascending**=False**)**.**head(10)

sns**.**set(rc**=**{'figure.figsize':(25,5)})

pcs**=**sns**.**barplot(x**=**'Product\_ID',y**=**'Orders',data**=**pcs,hue**=**'Product\_ID')

**for** bars **in** pcs**.**containers:

pcs**.**bar\_label(bars)

plt**.**show()

***# top 10 sold products using nlargest------***

fig1,ax1**=**plt**.**subplots(figsize**=**(12,7))

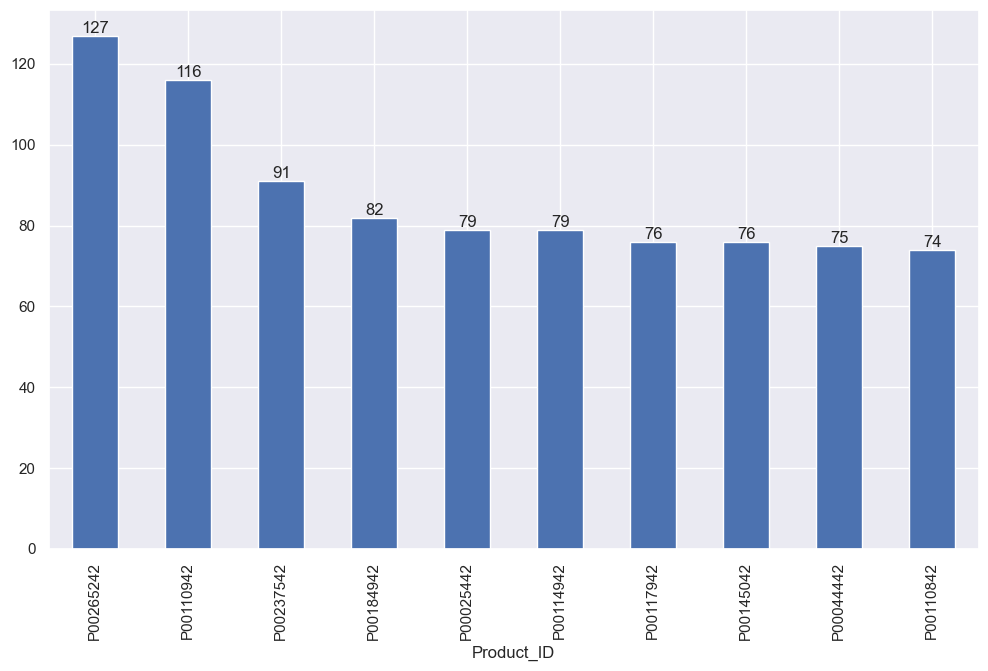
pcs1**=**df**.**groupby(['Product\_ID'])['Orders']**.**sum()**.**nlargest(10)**.**sort\_values(ascending**=False**)**.**plot(kind**=**'bar')

***# count on bar----***

**for** bars **in** pcs1**.**containers:

pcs1**.**bar\_label(bars)

plt**.**show()



**CONCLUSION-----**

Married women in the age group 26-35 from UP, MAHARASHTRA and KARNATAKA working in IT, AVIATION and HEALTHCARE are more likely buying products from FOOD, CLOTHING and ELECTRONICS category.