Supermarket Sales Prediction

The purpose of the project ! 👁

◉ **Problem**: Supermarket owners want to predict the prices of goods and merchandise in order to know which branches make more profits and which ones should be developed in order to determine more profit.
◉ **Solving the Problem**: With the development we are witnessing from artificial intelligence, machine learning models can be used and then trained on a set of training data, then tested on a set of test data, and predicts prices and incomes based on the data to be entered.

Data Features! ✨ ✨ ✨

Invoice id: Computer generated sales slip invoice identification number

Branch: Branch of supercenter (3 branches are available identified by A, B and C).

City: Location of supercenters

Customer type: Type of customers, recorded by Members for customers using member card and Normal for without member card.

Gender: Gender type of customer

Product line: General item categorization groups - Electronic accessories, Fashion accessories, Food and beverages, Health and beauty, Home and lifestyle, Sports and travel

Unit price: Price of each product in $

Quantity: Number of products purchased by customer

Tax: 5% tax fee for customer buying

Total: Total price including tax

Date: Date of purchase (Record available from January 2019 to March 2019)

Time: Purchase time (10am to 9pm)

Payment: Payment used by customer for purchase (3 methods are available – Cash, Credit card and Ewallet)

COGS: Cost of goods sold

Gross margin percentage: Gross margin percentage

Gross income: Gross income

Rating: Customer stratification rating on their overall shopping experience (On a scale of 1 to 10)

Importing libraries ⏳

In [1]:
```
#Reading data
import pandas as pd


#Fixings warnings
import warnings
warnings.filterwarnings('ignore')


#For mathematical operations
import numpy as np


#Visualisation
import seaborn as sns
import plotly.express as px
import matplotlib.pyplot as plt

#Data preprocessing
from sklearn.preprocessing import LabelEncoder


#Data spliting
from sklearn.model_selection import train_test_split
```

Reading data 📖

In [2]:
```
data=pd.read_csv('F:\Sales forcasting\Sales Kaggle New\supermarket_sales - Sheet1.csv')

#head() for display the first 5 rows
data.head().style.set_properties(**{'background-color': '#873600','color': '#E2EEF3'}) #for colored output
```

Out[2]:

	Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Time	Payment	cogs	gr mai percent
0	750-67-8428	A	Yangon	Member	Female	Health and beauty	74.690000	7	26.141500	548.971500	1/5/2019	13:08	Ewallet	522.830000	4.761
1	226-31-3081	C	Naypyitaw	Normal	Female	Electronic accessories	15.280000	5	3.820000	80.220000	3/8/2019	10:29	Cash	76.400000	4.761
2	631-41-3108	A	Yangon	Normal	Male	Home and lifestyle	46.330000	7	16.215500	340.525500	3/3/2019	13:23	Credit card	324.310000	4.761
3	123-19-1176	A	Yangon	Member	Male	Health and beauty	58.220000	8	23.288000	489.048000	1/27/2019	20:33	Ewallet	465.760000	4.761
4	373-73-7910	A	Yangon	Normal	Male	Sports and travel	86.310000	7	30.208500	634.378500	2/8/2019	10:37	Ewallet	604.170000	4.761

Statistical information 🔍

- **A visual and mathematical portrayal of information is statistics.**
- **Data science is all about making calculations with data.**
- **We make decisions based on that data using mathematical conditions known.**

In [3]:
```
data.shape
```

Out[3]: (1000, 17)

Observations 📋

- Here **1000 rows** , **17 coulmns**

In [4]:
```
data.size
```

Out[4]: 17000

In [5]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 17 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   Invoice ID              1000 non-null   object 
 1   Branch                  1000 non-null   object 
 2   City                    1000 non-null   object 
 3   Customer type           1000 non-null   object 
 4   Gender                  1000 non-null   object 
 5   Product line            1000 non-null   object 
 6   Unit price              1000 non-null   float64
 7   Quantity                1000 non-null   int64  
 8   Tax 5%                  1000 non-null   float64
 9   Total                   1000 non-null   float64
 10  Date                    1000 non-null   object 
 11  Time                    1000 non-null   object 
 12  Payment                 1000 non-null   object 
 13  cogs                    1000 non-null   float64
 14  gross margin percentage  1000 non-null   float64
 15  gross income            1000 non-null   float64
 16  Rating                  1000 non-null   float64
dtypes: float64(7), int64(1), object(9)
memory usage: 132.9+ KB
```

Observations 📋

- There are **9** columns are **string** the rest are **numeric** in terms of datatype.
- There aren't **null** values

In [6]:
```
#for statistical info
data.describe().style.background_gradient(cmap='Oranges') #for colored output
```

Out[6]:

	Unit price	Quantity	Tax 5%	Total	cogs	gross margin percentage	gross income	Rating
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	55.672130	5.510000	15.379369	322.966749	307.587380	4.761905	15.379369	6.972700
std	26.494628	2.923431	11.708825	245.885335	234.176510	0.000000	11.708825	1.718580
min	10.080000	1.000000	0.508500	10.678500	10.170000	4.761905	0.508500	4.000000
25%	32.875000	3.000000	5.924875	124.422375	118.497500	4.761905	5.924875	5.500000
50%	55.230000	5.000000	12.088000	253.848000	241.760000	4.761905	12.088000	7.000000
75%	77.935000	8.000000	22.445250	471.350250	448.905000	4.761905	22.445250	8.500000
max	99.960000	10.000000	49.650000	1042.650000	993.000000	4.761905	49.650000	10.000000

In [7]:
```
#for statistical info including string values
data.describe(include='O').style.set_properties(**{'background-color': '#873600',
                                                    'color': '#E2EEF3'})
```

Out[7]:

	Invoice ID	Branch	City	Customer type	Gender	Product line	Date	Time	Payment
count	1000	1000	1000	1000	1000	1000	1000	1000	1000
unique	1000	3	3	2	2	6	89	506	3
top	750-67-8428	A	Yangon	Member	Female	Fashion accessories	2/7/2019	19:48	Ewallet
freq	1	340	340	501	501	178	20	7	345

Exploratory Data Analysis **(EDA)** 📊

In [8]:
```
data.columns #for show names of columns
```

Out[8]:
```
Index(['Invoice ID', 'Branch', 'City', 'Customer type', 'Gender',
       'Product line', 'Unit price', 'Quantity', 'Tax 5%', 'Total', 'Date',
       'Time', 'Payment', 'cogs', 'gross margin percentage', 'gross income',
       'Rating'],
      dtype='object')
```

In [9]:
```
data.nunique() #for number of values of columns
```

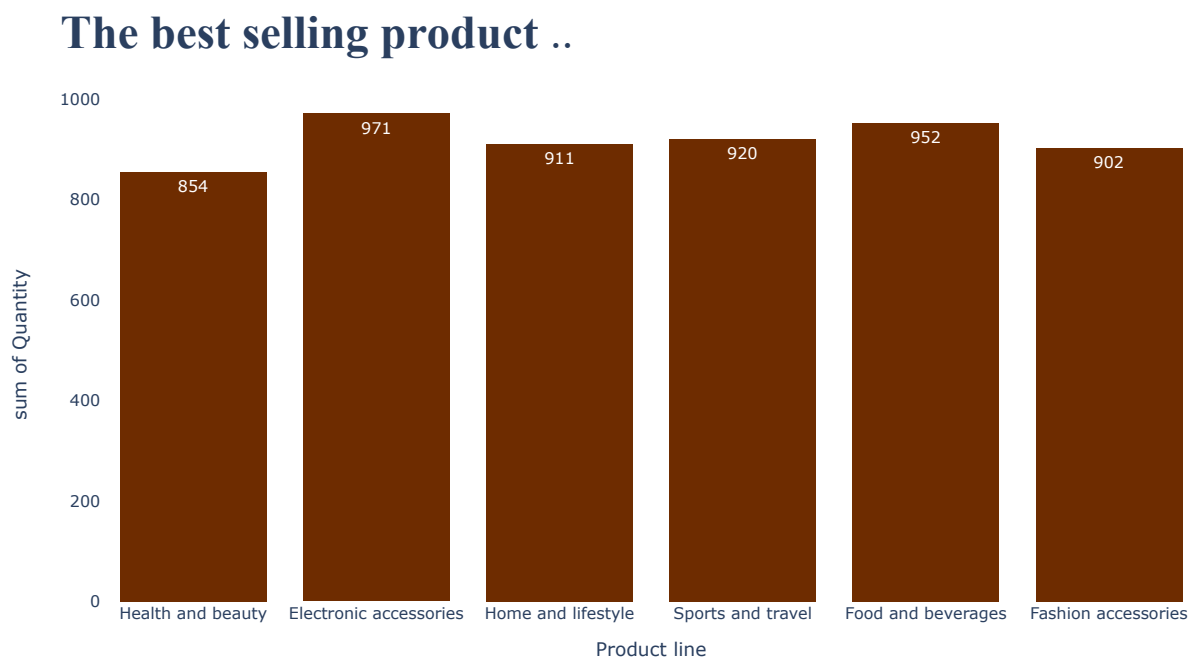
Out[9]:
```
Invoice ID                1000
Branch                       3
City                         3
Customer type                2
Gender                       2
Product line                 6
Unit price                 943
Quantity                    10
Tax 5%                     990
Total                      990
Date                        89
Time                       506
Payment                      3
cogs                       990
gross margin percentage      1
gross income               990
Rating                      61
dtype: int64
```

In [10]:
```
data[['Product line','Quantity']].groupby(['Product line']).mean().sort_values(by='Quantity',ascending=False).style.backgrou
```

Out[10]:

Product line	Quantity
Electronic accessories	5.711765
Home and lifestyle	5.693750
Health and beauty	5.618421
Sports and travel	5.542169
Food and beverages	5.471264
Fashion accessories	5.067416

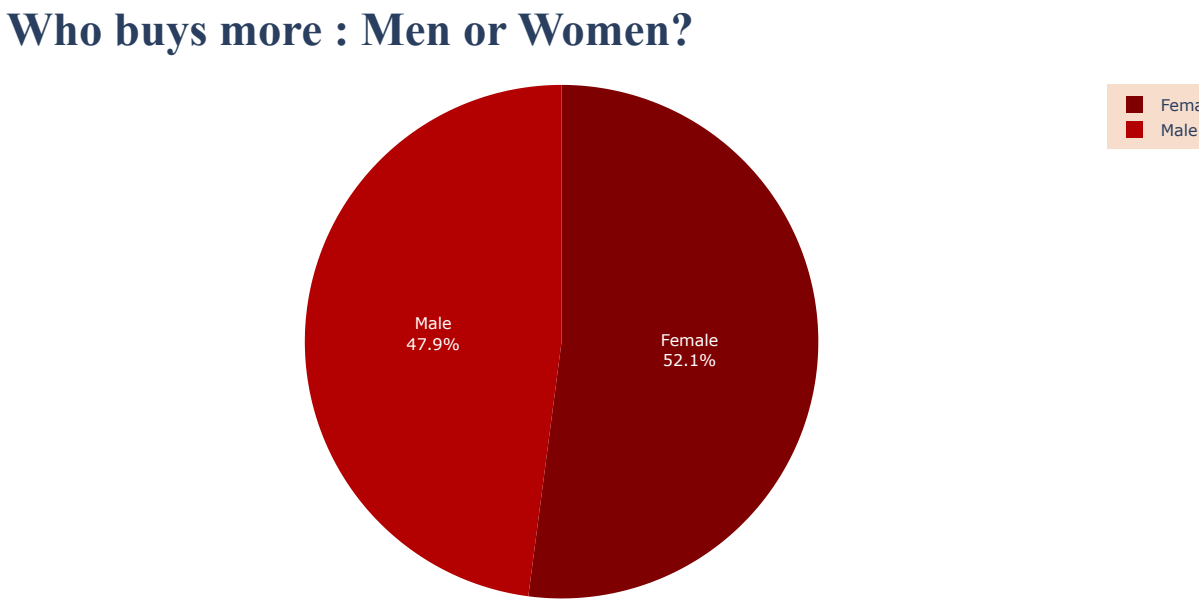
In [11]:
```
fig=px.histogram(data,x='Product line',y='Quantity',
                 color_discrete_sequence=['#6E2C00'],
                 text_auto=True)


fig.update_layout(title='<b>The best selling product </b>..',
                  title_font={'size':35,'family': 'Serif'},
                  paper_bgcolor='#F6DDCC',
                  plot_bgcolor='#F6DDCC')



fig.update_yaxes(showgrid=False)

fig.show()
```



Observations 📋

- Obviously, the highest percentage of sales is **Electronic accessories**.

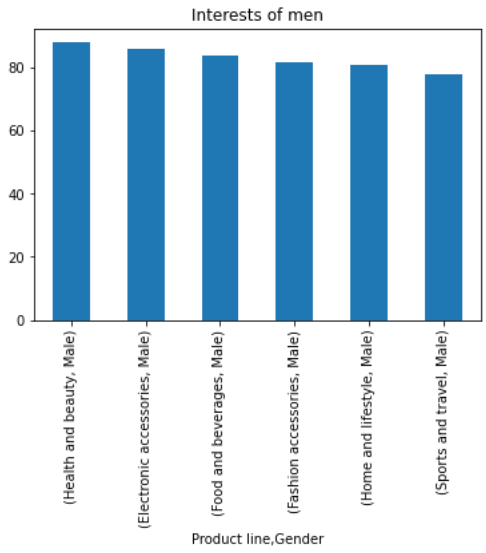
In [12]:
```
fig = px.pie(data,values='Quantity',names='Gender',
             hover_data=['Quantity','Gender'],
             labels={'Gender':'Gender'},
             color_discrete_sequence=px.colors.sequential.OrRd_r)


fig.update_traces(textposition='inside',
                  textinfo='percent+label')


fig.update_layout(title='<b> Who buys more : Men or Women?<b>',
                  titlefont={'size': 35,'family': 'Serif'},
                  showlegend=True,
                  paper_bgcolor='#F6DDCC',
                  plot_bgcolor='#F6DDCC')



fig.show()
```



Observations 📋

- It is clear that **women** buy more than men

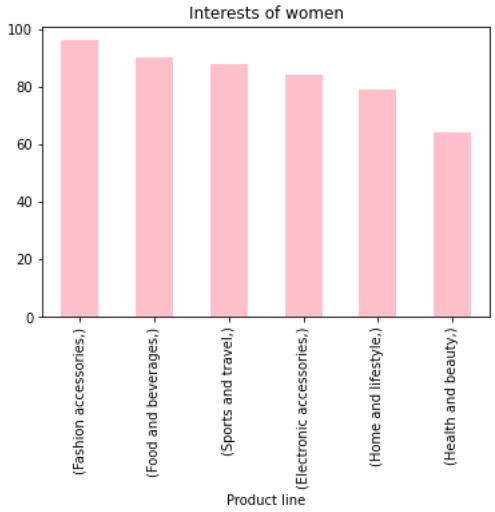
In [13]:
```
data[['Product line','Gender']][(data['Gender']=='Male')].value_counts().plot(kind='bar',title='Interests of men')
plt.show()
```



Observations 📋

- It is clear that the most important interests of men is **health and beauty**.

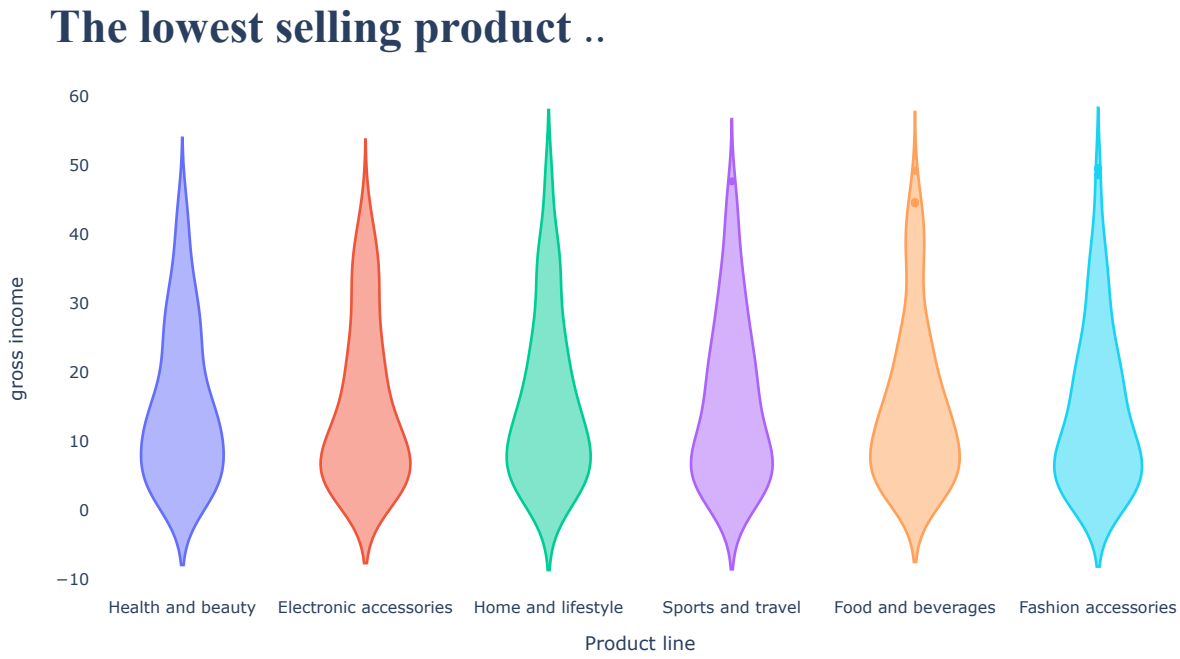
In [14]:
```
data[['Product line']][(data['Gender']=='Female')].value_counts().plot(kind='bar',color='pink',title='Interests of women')
plt.show()
```



Observations 📋

- It is clear that the most important interests of women is **Fashion accessories**.

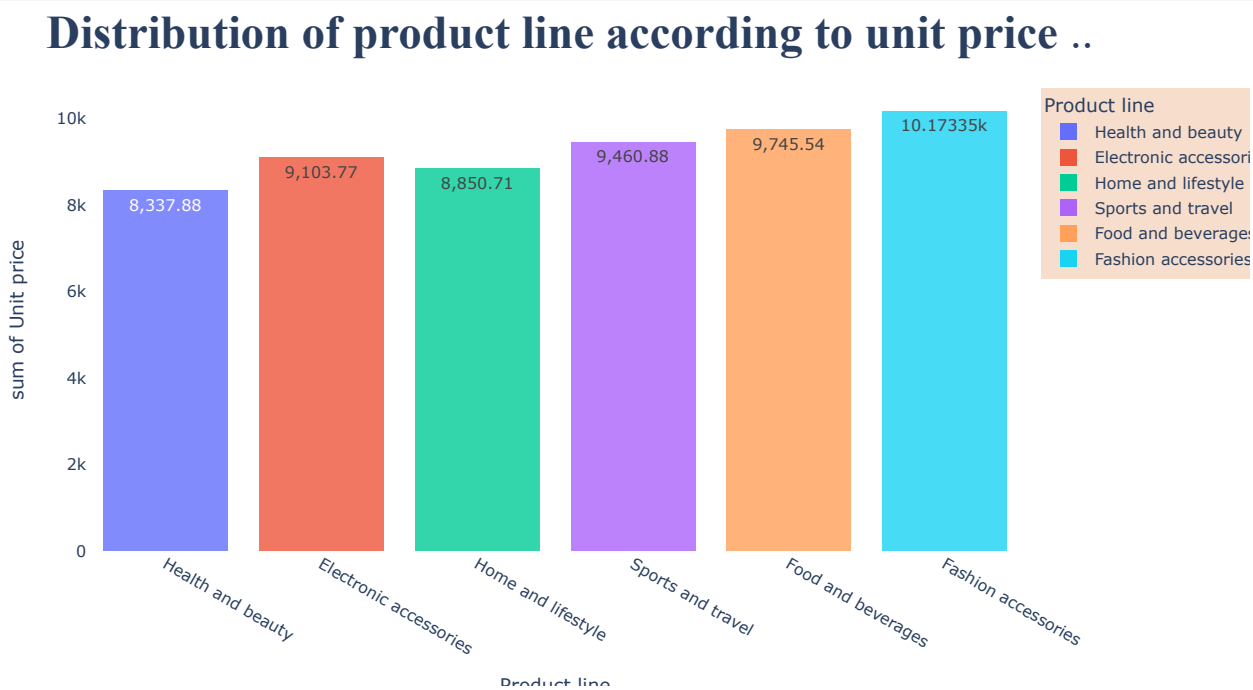
In [15]:
```
fig=px.violin(data,x='Product line',y='gross income',
              color='Product line')


fig.update_layout(title='<b>The lowest selling product </b>..',
                  title_font={'size':35,'family': 'Serif'},
                  showlegend=False,
                  paper_bgcolor='#F6DDCC',
                  plot_bgcolor='#F6DDCC')



fig.update_xaxes(showgrid=False)
fig.update_yaxes(showgrid=False)
fig.show()
```



Observations 📋

- It is clear that the most important interests of men is **Sports and travel**.

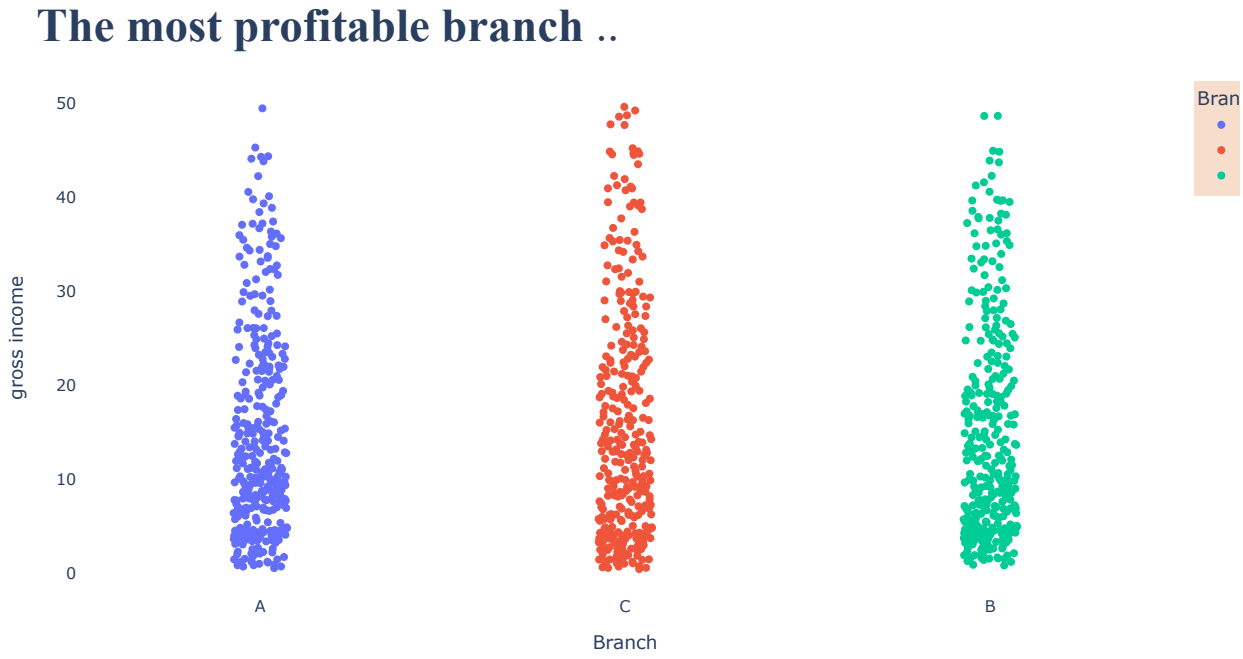
In [16]:
```
fig=px.histogram(data,x='Product line',y='Unit price',
                 color='Product line',nbins=5,
                  text_auto=True,opacity=.8)

fig.update_layout(title='<b>Distribution of product line according to unit price </b>..',
                   title_font={'size':35,'family': 'Serif'},

                   paper_bgcolor='#F6DDCC',
                   plot_bgcolor='#F6DDCC')



fig.update_xaxes(showgrid=False)
fig.update_yaxes(showgrid=False)
fig.show()
```



Observations 📋

- It is clear that the most expensive products are **Fashion accessories** and the cheapest are **Health and beauty**.

In [17]:
```
fig=px.strip(data,x='Branch',y='gross income',
             color='Branch')


fig.update_layout(title='<b>The most profitable branch </b>..',
                  title_font={'size':35,'family': 'Serif'},
                  paper_bgcolor='#F6DDCC',
                  plot_bgcolor='#F6DDCC')


fig.update_yaxes(showgrid=False)
fig.show()
```



Observations

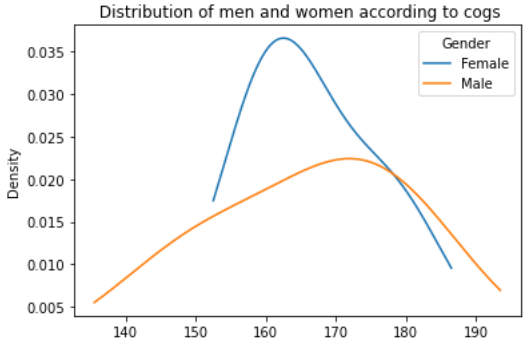
- It is clear that The most profitable branch is **c** .

In [18]:
```
data.pivot_table(index='Branch',columns='Gender',values='cogs',aggfunc='count').style.background_gradient(cmap='Oranges')
```

Out[18]:

Gender	Female	Male
Branch		
A	161	179
B	162	170
C	178	150

In [19]:
```
data.pivot_table(index='Branch',columns='Gender',values='cogs',aggfunc='count').plot(kind='kde',title=' Distribution of men a
plt.show()
```

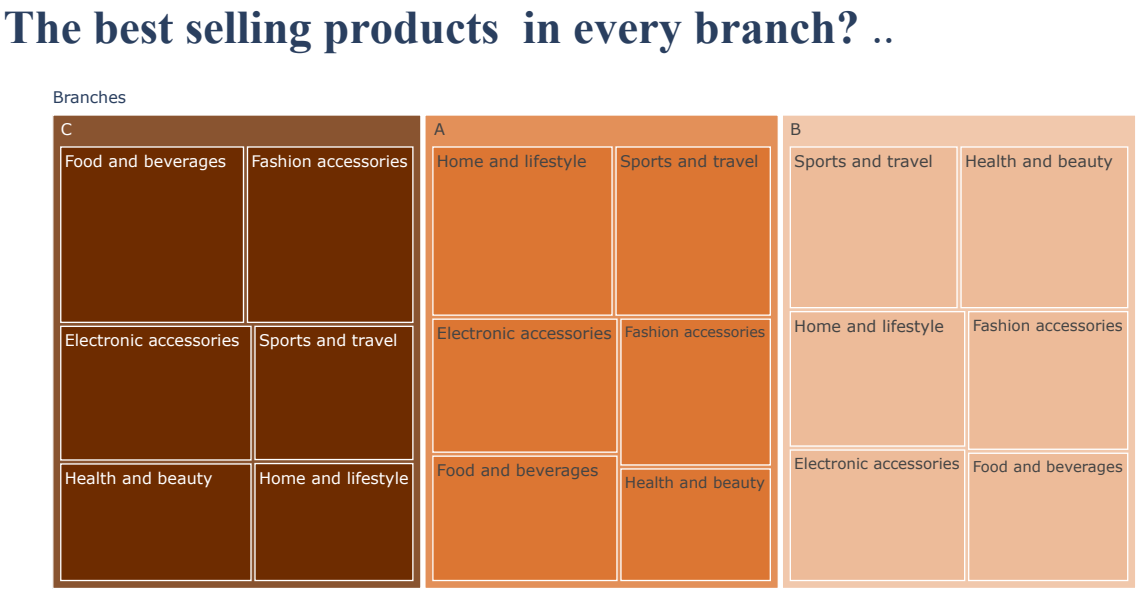


In [21]:
```
data.pivot_table(index='Product line',columns='Branch',values='cogs',aggfunc='max').style.background_gradient(cmap='Oranges')
```

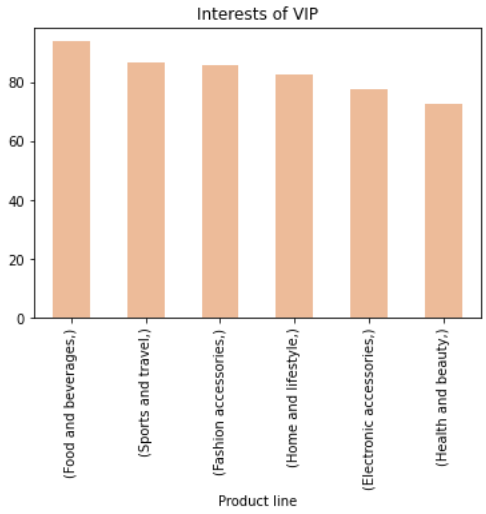
Out[21]:

Product line	A	B	C
Electronic accessories	886.700000	897.570000	823.400000
Fashion accessories	989.800000	832.500000	993.000000
Food and beverages	887.940000	846.300000	985.200000
Health and beauty	716.800000	878.700000	905.000000
Home and lifestyle	906.500000	973.800000	975.000000
Sports and travel	882.810000	899.640000	954.400000

In [22]:
```
fig=px.treemap(data,path=[px.Constant('Branches'),'Branch','Product line'],
               values='cogs',
               color_discrete_sequence=['#6E2C00','#DC7633','#EDBB99'])


fig.update_layout(title='<b>The best selling products  in every branch?</b> ..',
                  titlefont={'size': 35,'family': 'Serif'})



fig.show()
```



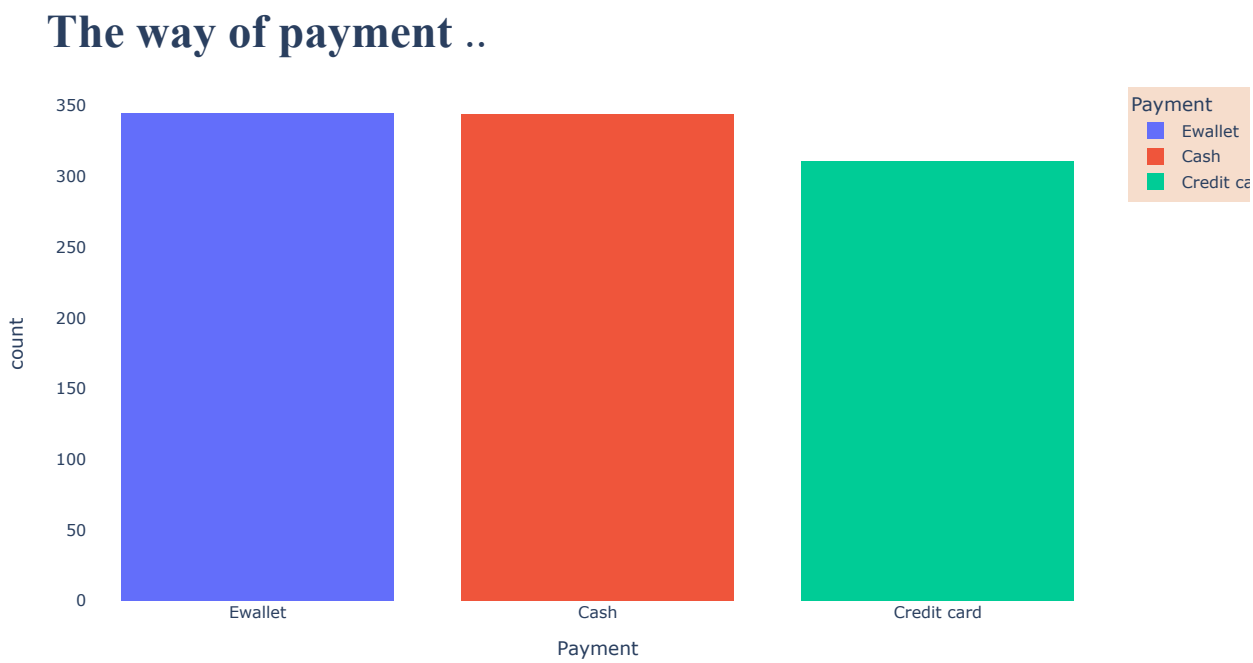
In [23]:
```
data[['Product line']][(data['Customer type']=='Member')].value_counts().plot(kind='bar',color='#EDBB99',title='Interests of
plt.show()
```



Observations

- It is clear that **Members** are more interested in **food and beverages**.

In [24]:
```
fig=px.histogram(data,x='Payment',color='Payment')

fig.update_layout(title='<b>The way of payment </b>..',
                  title_font={'size':35,'family': 'Serif'},
                  paper_bgcolor='#F6DDCC',
                  plot_bgcolor='#F6DDCC')



fig.update_yaxes(showgrid=False)
fig.show()
```



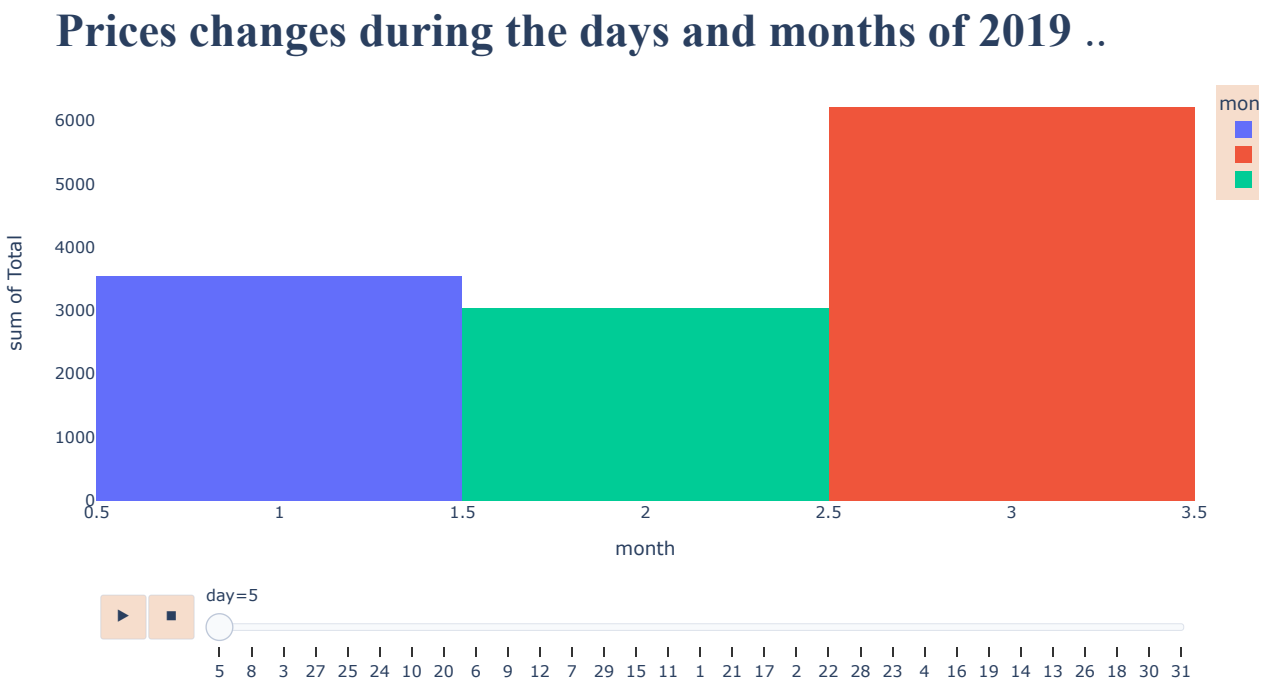
Observations 📋

- It is clear that The most available method is **the wallet**.

In [25]:
```
data['Date']=pd.to_datetime(data['Date'])
```

In [26]:
```
data.insert(10,'month',data.Date.dt.month)
data.insert(11,'day',data.Date.dt.day)
```

In [27]:
```
fig=px.histogram(data,x='month',y='Total',
                 color='month',animation_frame='day',
                 hover_data=['Product line','Gender'])


fig.update_layout(title='<b>Prices changes during the days and months of 2019 </b>..',
                  title_font={'size':35,'family': 'Serif'},
                  paper_bgcolor='#F6DDCC',
                  plot_bgcolor='#F6DDCC')



fig.show()
```



Data preprocessing 🛠️

Data preprocessing refers to the technique of preparing (cleaning and organizing) the raw data to make it suitable for a building and training Machine Learning models.

In [28]:
```
data.isna().sum()
```

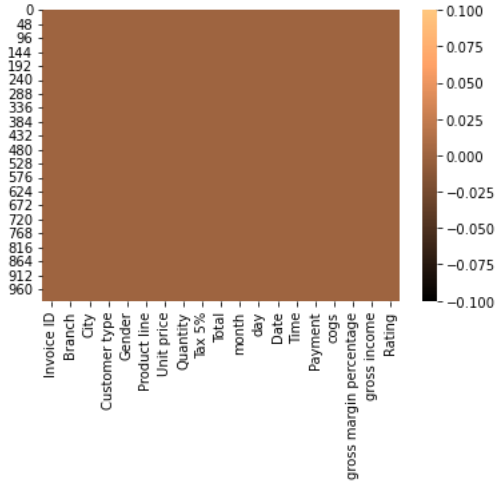
Out[28]:
```
Invoice ID                0
Branch                    0
City                      0
Customer type             0
Gender                    0
Product line              0
Unit price                0
Quantity                  0
Tax 5%                    0
Total                     0
month                     0
day                       0
Date                      0
Time                      0
Payment                   0
cogs                      0
gross margin percentage   0
gross income              0
Rating                    0
dtype: int64
```

Observations 📋

- There is no null values

In [29]:
```
sns.heatmap(data.isna(),cmap='copper')
```

Out[29]: <AxesSubplot:>



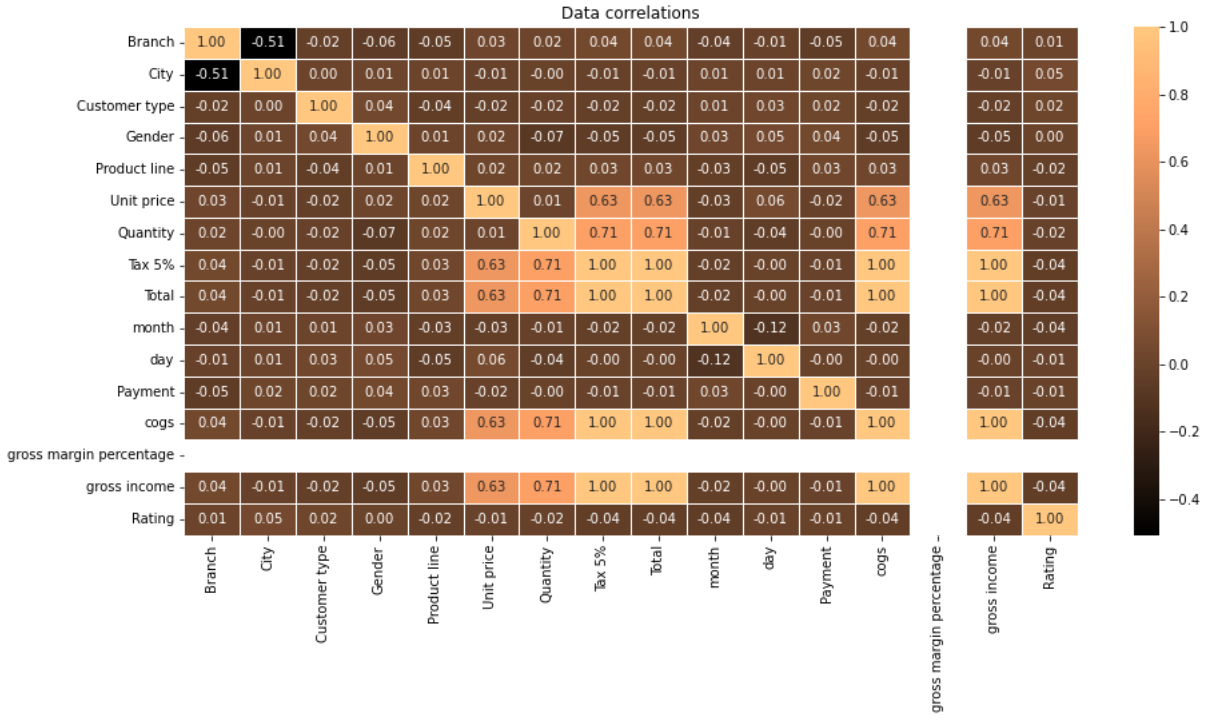
In [30]:
```
data=data.drop(['Invoice ID','Time','Date'],axis=1)
data.sample(3).style.set_properties(**{'background-color': '#873600',
                                       'color': '#E2EEF3'})
```

Out[30]:

	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	month	day	Payment	cogs	gross margin percentage	gro inco
462	C	Naypyitaw	Normal	Female	Sports and travel	73.950000	4	14.790000	310.590000	2	3	Cash	295.800000	4.761905	14.790(
758	A	Yangon	Member	Male	Food and beverages	18.850000	10	9.425000	197.925000	2	27	Ewallet	188.500000	4.761905	9.425(
870	A	Yangon	Member	Male	Food and beverages	24.820000	7	8.687000	182.427000	2	16	Credit card	173.740000	4.761905	8.687(

In [31]:
```
LE=LabelEncoder()

categories=['Branch', 'City', 'Customer type', 'Gender', 'Product line', 'Payment']
for label in categories:
    data[label]=LE.fit_transform(data[label])
```

In [32]:
```
correlation=data.corr()  #To show how interconnected the data is
plt.figure(figsize=(15,7))
sns.heatmap(correlation,annot=True,fmt='.2f',annot_kws={'size': 10},linewidths=0.5,cmap='copper')
plt.title("Data correlations")
plt .show()
```



Data Spliting:-

In [33]:
```
x=data.drop('Total',axis=1)
y=data['Total']



print('The dimensions of x is : ',x.shape)
print('The dimensions of y is : ',y.shape)
```

```
The dimensions of x is :  (1000, 15)
The dimensions of y is :  (1000,)
```

In [34]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=.33,random_state=42,shuffle=True)
```

In [35]:
```
print("x train dimensions :",x_train.shape)
print("x test dimensions: ",x_test.shape)
print("y train dimensions :",y_train.shape)
print("y test dimensions :",y_test.shape)
```

```
x train dimensions : (670, 15)
x test dimensions:  (330, 15)
y train dimensions : (670,)
y test dimensions : (330,)
```

In [36]:
```
pip install pycaret
```

```
Collecting pycaretNote: you may need to restart the kernel to use updated packages.

  Using cached pycaret-3.1.0-py3-none-any.whl (483 kB)
Requirement already satisfied: ipywidgets>=7.6.5 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (7.6.5)
Requirement already satisfied: yellowbrick>=1.4 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (1.5)
Requirement already satisfied: schemdraw==0.15 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (0.15)
Requirement already satisfied: numba>=0.55.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (0.55.1)
Requirement already satisfied: scikit-learn<1.3.0,>=1.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (1.
0.2)
Requirement already satisfied: pandas<2.0.0,>=1.3.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (1.4.2)
Requirement already satisfied: sktime!=0.17.1,!=0.17.2,!=0.18.0,<0.22.0,>=0.16.1 in c:\users\hp_9046\anaconda3\lib\site-pac
kages (from pycaret) (0.21.1)
Collecting plotly-resampler>=0.8.3.1
  Using cached plotly_resampler-0.9.1-py3-none-any.whl (73 kB)
Requirement already satisfied: requests>=2.27.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (2.27.1)
Requirement already satisfied: importlib-metadata>=4.12.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret)
(6.8.0)
Requirement already satisfied: pmdarima!=1.8.1,<3.0.0,>=1.8.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycare
t) (2.0.3)
Requirement already satisfied: tqdm>=4.62.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (4.64.0)
Requirement already satisfied: scipy~=1.10.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (1.10.1)
Requirement already satisfied: nbformat>=4.2.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (5.3.0)
Collecting category-encoders>=2.4.0
  Using cached category_encoders-2.6.2-py2.py3-none-any.whl (81 kB)
Requirement already satisfied: tbats>=1.1.3 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (1.1.3)
Collecting imbalanced-learn>=0.8.1
  Using cached imbalanced_learn-0.11.0-py3-none-any.whl (235 kB)
Requirement already satisfied: joblib>=1.2.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (1.3.2)
Requirement already satisfied: scikit-plot>=0.3.7 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (0.3.7)
Collecting deprecation>=2.1.0
  Using cached deprecation-2.1.0-py2.py3-none-any.whl (11 kB)
Requirement already satisfied: psutil>=5.9.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (5.9.5)
Requirement already satisfied: markupsafe>=2.0.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (2.0.1)
Requirement already satisfied: numpy<1.24,>=1.21 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (1.21.5)
Requirement already satisfied: jinja2>=1.2 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (2.11.3)
Requirement already satisfied: matplotlib>=3.3.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (3.5.1)
Requirement already satisfied: ipython>=5.5.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (8.2.0)
Requirement already satisfied: xxhash in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (3.4.1)
Requirement already satisfied: cloudpickle in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (2.0.0)
Collecting kaleido>=0.2.1
  Using cached kaleido-0.2.1-py2.py3-none-win_amd64.whl (65.9 MB)
Requirement already satisfied: statsmodels>=0.12.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (0.13.2)
Collecting lightgbm>=3.0.0
  Using cached lightgbm-4.1.0-py3-none-win_amd64.whl (1.3 MB)
Requirement already satisfied: pyod>=1.0.8 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (1.1.0)
Requirement already satisfied: plotly>=5.0.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from pycaret) (5.6.0)
Requirement already satisfied: patsy>=0.5.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from category-encoders>=2.4.0-
>pycaret) (0.5.2)
Requirement already satisfied: packaging in c:\users\hp_9046\anaconda3\lib\site-packages (from deprecation>=2.1.0->pycaret)
(21.3)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from imbalanced-learn>
=0.8.1->pycaret) (2.2.0)
Requirement already satisfied: zipp>=0.5 in c:\users\hp_9046\anaconda3\lib\site-packages (from importlib-metadata>=4.12.0->
pycaret) (3.7.0)
Requirement already satisfied: colorama in c:\users\hp_9046\anaconda3\lib\site-packages (from ipython>=5.5.0->pycaret) (0.
4.4)
Requirement already satisfied: jedi>=0.16 in c:\users\hp_9046\anaconda3\lib\site-packages (from ipython>=5.5.0->pycaret)
(0.18.1)
Requirement already satisfied: prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.0.0 in c:\users\hp_9046\anaconda3\lib\site-packages
(from ipython>=5.5.0->pycaret) (3.0.20)
Requirement already satisfied: pickleshare in c:\users\hp_9046\anaconda3\lib\site-packages (from ipython>=5.5.0->pycaret)
(0.7.5)
Requirement already satisfied: matplotlib-inline in c:\users\hp_9046\anaconda3\lib\site-packages (from ipython>=5.5.0->pyca
ret) (0.1.2)
Requirement already satisfied: pygments>=2.4.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from ipython>=5.5.0->pycare
t) (2.11.2)
Requirement already satisfied: backcall in c:\users\hp_9046\anaconda3\lib\site-packages (from ipython>=5.5.0->pycaret) (0.
2.0)
Requirement already satisfied: decorator in c:\users\hp_9046\anaconda3\lib\site-packages (from ipython>=5.5.0->pycaret) (5.
1.1)
Requirement already satisfied: stack-data in c:\users\hp_9046\anaconda3\lib\site-packages (from ipython>=5.5.0->pycaret)
(0.2.0)
Requirement already satisfied: traitlets>=5 in c:\users\hp_9046\anaconda3\lib\site-packages (from ipython>=5.5.0->pycaret)
(5.1.1)
Requirement already satisfied: setuptools>=18.5 in c:\users\hp_9046\anaconda3\lib\site-packages (from ipython>=5.5.0->pycar
et) (61.2.0)
Requirement already satisfied: widgetsnbextension~=3.5.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from ipywidgets>=
7.6.5->pycaret) (3.5.2)
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from ipywidgets>=
7.6.5->pycaret) (1.0.0)
Requirement already satisfied: ipython-genutils~=0.2.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from ipywidgets>=7.
6.5->pycaret) (0.2.0)
Requirement already satisfied: ipykernel>=4.5.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from ipywidgets>=7.6.5->py
caret) (6.9.1)
Requirement already satisfied: nest-asyncio in c:\users\hp_9046\anaconda3\lib\site-packages (from ipykernel>=4.5.1->ipywidg
ets>=7.6.5->pycaret) (1.5.5)
Requirement already satisfied: debugpy<2.0,>=1.0.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from ipykernel>=4.5.1->
ipywidgets>=7.6.5->pycaret) (1.5.1)
Requirement already satisfied: tornado<7.0,>=4.2 in c:\users\hp_9046\anaconda3\lib\site-packages (from ipykernel>=4.5.1->ip
ywidgets>=7.6.5->pycaret) (6.1)
Requirement already satisfied: jupyter-client<8.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from ipykernel>=4.5.1->i
pywidgets>=7.6.5->pycaret) (6.1.12)
```

```
Requirement already satisfied: parso<0.9.0,>=0.8.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from jedi>=0.16->ipytho
n>=5.5.0->pycaret) (0.8.3)
Requirement already satisfied: python-dateutil>=2.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from jupyter-client<8.
0->ipykernel>=4.5.1->ipywidgets>=7.6.5->pycaret) (2.8.2)
Requirement already satisfied: jupyter-core>=4.6.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from jupyter-client<8.0
->ipykernel>=4.5.1->ipywidgets>=7.6.5->pycaret) (4.9.2)
Requirement already satisfied: pyzmq>=13 in c:\users\hp_9046\anaconda3\lib\site-packages (from jupyter-client<8.0->ipykerne
l>=4.5.1->ipywidgets>=7.6.5->pycaret) (22.3.0)
Requirement already satisfied: pywin32>=1.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from jupyter-core>=4.6.0->jupy
ter-client<8.0->ipykernel>=4.5.1->ipywidgets>=7.6.5->pycaret) (302)
Requirement already satisfied: cycler>=0.10 in c:\users\hp_9046\anaconda3\lib\site-packages (from matplotlib>=3.3.0->pycare
t) (0.11.0)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from matplotlib>=3.3.0->py
caret) (3.0.4)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from matplotlib>=3.3.0->p
ycaret) (4.25.0)
Requirement already satisfied: pillow>=6.2.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from matplotlib>=3.3.0->pycar
et) (9.0.1)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from matplotlib>=3.3.0->p
ycaret) (1.3.2)
Requirement already satisfied: fastjsonschema in c:\users\hp_9046\anaconda3\lib\site-packages (from nbformat>=4.2.0->pycare
t) (2.15.1)
Requirement already satisfied: jsonschema>=2.6 in c:\users\hp_9046\anaconda3\lib\site-packages (from nbformat>=4.2.0->pycar
et) (4.4.0)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in c:\users\hp_9046\anaconda3\lib\site-package
s (from jsonschema>=2.6->nbformat>=4.2.0->pycaret) (0.18.0)
Requirement already satisfied: attrs>=17.4.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from jsonschema>=2.6->nbforma
t>=4.2.0->pycaret) (21.4.0)
Requirement already satisfied: llvmlite<0.39,>=0.38.0rc1 in c:\users\hp_9046\anaconda3\lib\site-packages (from numba>=0.55.
0->pycaret) (0.38.0)
Requirement already satisfied: pytz>=2020.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from pandas<2.0.0,>=1.3.0->pyc
aret) (2021.3)
Requirement already satisfied: six in c:\users\hp_9046\anaconda3\lib\site-packages (from patsy>=0.5.1->category-encoders>=
2.4.0->pycaret) (1.16.0)
Requirement already satisfied: tenacity>=6.2.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from plotly>=5.0.0->pycare
t) (8.0.1)
Requirement already satisfied: dash<3.0.0,>=2.11.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from plotly-resampler>=
0.8.3.1->pycaret) (2.14.0)
Requirement already satisfied: orjson<4.0.0,>=3.8.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from plotly-resampler>
=0.8.3.1->pycaret) (3.9.8)
Requirement already satisfied: trace-updater>=0.0.8 in c:\users\hp_9046\anaconda3\lib\site-packages (from plotly-resampler>
=0.8.3.1->pycaret) (0.0.9.1)
Requirement already satisfied: tsdownsample==0.1.2 in c:\users\hp_9046\anaconda3\lib\site-packages (from plotly-resampler>=
0.8.3.1->pycaret) (0.1.2)
Requirement already satisfied: dash-core-components==2.0.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from dash<3.0.
0,>=2.11.0->plotly-resampler>=0.8.3.1->pycaret) (2.0.0)
Requirement already satisfied: dash-html-components==2.0.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from dash<3.0.
0,>=2.11.0->plotly-resampler>=0.8.3.1->pycaret) (2.0.0)
Requirement already satisfied: retrying in c:\users\hp_9046\anaconda3\lib\site-packages (from dash<3.0.0,>=2.11.0->plotly-r
esampler>=0.8.3.1->pycaret) (1.3.4)
Requirement already satisfied: ansi2html in c:\users\hp_9046\anaconda3\lib\site-packages (from dash<3.0.0,>=2.11.0->plotly-
resampler>=0.8.3.1->pycaret) (1.8.0)
Requirement already satisfied: typing-extensions>=4.1.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from dash<3.0.0,>=
2.11.0->plotly-resampler>=0.8.3.1->pycaret) (4.1.1)
Requirement already satisfied: dash-table==5.0.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from dash<3.0.0,>=2.11.0-
>plotly-resampler>=0.8.3.1->pycaret) (5.0.0)
Requirement already satisfied: Flask<2.3.0,>=1.0.4 in c:\users\hp_9046\anaconda3\lib\site-packages (from dash<3.0.0,>=2.11.
0->plotly-resampler>=0.8.3.1->pycaret) (1.1.2)
Requirement already satisfied: Werkzeug<2.3.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from dash<3.0.0,>=2.11.0->pl
otly-resampler>=0.8.3.1->pycaret) (2.0.3)
Requirement already satisfied: click>=5.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from Flask<2.3.0,>=1.0.4->dash<
3.0.0,>=2.11.0->plotly-resampler>=0.8.3.1->pycaret) (8.0.4)
Requirement already satisfied: itsdangerous>=0.24 in c:\users\hp_9046\anaconda3\lib\site-packages (from Flask<2.3.0,>=1.0.4
->dash<3.0.0,>=2.11.0->plotly-resampler>=0.8.3.1->pycaret) (2.0.1)
Requirement already satisfied: urllib3 in c:\users\hp_9046\anaconda3\lib\site-packages (from pmdarima!=1.8.1,<3.0.0,>=1.8.0
->pycaret) (1.26.9)
Requirement already satisfied: Cython!=0.29.18,!=0.29.31,>=0.29 in c:\users\hp_9046\anaconda3\lib\site-packages (from pmdar
ima!=1.8.1,<3.0.0,>=1.8.0->pycaret) (0.29.28)
Requirement already satisfied: wcwidth in c:\users\hp_9046\anaconda3\lib\site-packages (from prompt-toolkit!=3.0.0,!=3.0.1,
<3.1.0,>=2.0.0->ipython>=5.5.0->pycaret) (0.2.5)
Requirement already satisfied: idna<4,>=2.5 in c:\users\hp_9046\anaconda3\lib\site-packages (from requests>=2.27.1->pycare
t) (3.3)
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from requests>=2.
27.1->pycaret) (2.0.4)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\hp_9046\anaconda3\lib\site-packages (from requests>=2.27.1->p
ycaret) (2021.10.8)
Requirement already satisfied: deprecated>=1.2.13 in c:\users\hp_9046\anaconda3\lib\site-packages (from sktime!=0.17.1,!=0.
17.2,!=0.18.0,<0.22.0,>=0.16.1->pycaret) (1.2.14)
Requirement already satisfied: scikit-base<0.6.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from sktime!=0.17.1,!=0.1
7.2,!=0.18.0,<0.22.0,>=0.16.1->pycaret) (0.5.2)
Requirement already satisfied: wrapt<2,>=1.10 in c:\users\hp_9046\anaconda3\lib\site-packages (from deprecated>=1.2.13->skt
ime!=0.17.1,!=0.17.2,!=0.18.0,<0.22.0,>=0.16.1->pycaret) (1.12.1)
Requirement already satisfied: notebook>=4.4.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from widgetsnbextension~=3.
5.0->ipywidgets>=7.6.5->pycaret) (6.4.8)
Requirement already satisfied: Send2Trash>=1.8.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from notebook>=4.4.1->wid
getsnbextension~=3.5.0->ipywidgets>=7.6.5->pycaret) (1.8.0)
Requirement already satisfied: terminado>=0.8.3 in c:\users\hp_9046\anaconda3\lib\site-packages (from notebook>=4.4.1->widg
etsnbextension~=3.5.0->ipywidgets>=7.6.5->pycaret) (0.13.1)
Requirement already satisfied: prometheus-client in c:\users\hp_9046\anaconda3\lib\site-packages (from notebook>=4.4.1->wid
getsnbextension~=3.5.0->ipywidgets>=7.6.5->pycaret) (0.13.1)
Requirement already satisfied: argon2-cffi in c:\users\hp_9046\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnb
extension~=3.5.0->ipywidgets>=7.6.5->pycaret) (21.3.0)
```

```
Requirement already satisfied: nbconvert in c:\users\hp_9046\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnbex
tension~=3.5.0->ipywidgets>=7.6.5->pycaret) (6.4.4)
Requirement already satisfied: pywinpty>=1.1.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from terminado>=0.8.3->note
book>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.6.5->pycaret) (2.0.2)
Requirement already satisfied: argon2-cffi-bindings in c:\users\hp_9046\anaconda3\lib\site-packages (from argon2-cffi->note
book>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.6.5->pycaret) (21.2.0)
Requirement already satisfied: cffi>=1.0.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from argon2-cffi-bindings->argo
n2-cffi->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.6.5->pycaret) (1.15.0)
Requirement already satisfied: pycparser in c:\users\hp_9046\anaconda3\lib\site-packages (from cffi>=1.0.1->argon2-cffi-bin
dings->argon2-cffi->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.6.5->pycaret) (2.21)
Requirement already satisfied: mistune<2,>=0.8.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from nbconvert->notebook>
=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.6.5->pycaret) (0.8.4)
Requirement already satisfied: jupyterlab-pygments in c:\users\hp_9046\anaconda3\lib\site-packages (from nbconvert->noteboo
k>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.6.5->pycaret) (0.1.2)
Requirement already satisfied: entrypoints>=0.2.2 in c:\users\hp_9046\anaconda3\lib\site-packages (from nbconvert->notebook
>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.6.5->pycaret) (0.4)
Requirement already satisfied: bleach in c:\users\hp_9046\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4.1->wid
getsnbextension~=3.5.0->ipywidgets>=7.6.5->pycaret) (4.1.0)
Requirement already satisfied: defusedxml in c:\users\hp_9046\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4.1-
>widgetsnbextension~=3.5.0->ipywidgets>=7.6.5->pycaret) (0.7.1)
Requirement already satisfied: beautifulsoup4 in c:\users\hp_9046\anaconda3\lib\site-packages (from nbconvert->notebook>=4.
4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.6.5->pycaret) (4.11.1)
Requirement already satisfied: nbclient<0.6.0,>=0.5.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from nbconvert->note
book>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.6.5->pycaret) (0.5.13)
Requirement already satisfied: pandocfilters>=1.4.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from nbconvert->notebo
ok>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.6.5->pycaret) (1.5.0)
Requirement already satisfied: testpath in c:\users\hp_9046\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4.1->w
idgetsnbextension~=3.5.0->ipywidgets>=7.6.5->pycaret) (0.5.0)
Requirement already satisfied: soupsieve>1.2 in c:\users\hp_9046\anaconda3\lib\site-packages (from beautifulsoup4->nbconver
t->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.6.5->pycaret) (2.3.1)
Requirement already satisfied: webencodings in c:\users\hp_9046\anaconda3\lib\site-packages (from bleach->nbconvert->notebo
ok>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.6.5->pycaret) (0.5.1)
Requirement already satisfied: executing in c:\users\hp_9046\anaconda3\lib\site-packages (from stack-data->ipython>=5.5.0->
pycaret) (0.8.3)
Requirement already satisfied: asttokens in c:\users\hp_9046\anaconda3\lib\site-packages (from stack-data->ipython>=5.5.0->
pycaret) (2.0.5)
Requirement already satisfied: pure-eval in c:\users\hp_9046\anaconda3\lib\site-packages (from stack-data->ipython>=5.5.0->
pycaret) (0.2.2)
Installing collected packages: plotly-resampler, lightgbm, kaleido, imbalanced-learn, deprecation, category-encoders, pycar
et
Successfully installed category-encoders-2.6.2 deprecation-2.1.0 imbalanced-learn-0.11.0 kaleido-0.2.1 lightgbm-4.1.0 plotl
y-resampler-0.9.1 pycaret-3.1.0
```

In [37]:
```
from pycaret.regression import *
s=setup(data,target='Total')
```

	Description	Value
0	Session id	2848
1	Target	Total
2	Target type	Regression
3	Original data shape	(1000, 16)
4	Transformed data shape	(1000, 16)
5	Transformed train set shape	(700, 16)
6	Transformed test set shape	(300, 16)
7	Numeric features	15
8	Preprocess	True
9	Imputation type	simple
10	Numeric imputation	mean
11	Categorical imputation	mode
12	Fold Generator	KFold
13	Fold Number	10
14	CPU Jobs	-1
15	Use GPU	False
16	Log Experiment	False
17	Experiment Name	reg-default-name
18	USI	992b

Data modeling 

LinearRegression Model:-

In [38]:
```
from sklearn.linear_model import LinearRegression
LR=LinearRegression().fit(x_train,y_train)
```

In [39]:
```
print("LR training score :",round(LR.score(x_train,y_train),4)*100)
print("LR testing score :",round(LR.score(x_test,y_test),4)*100)
```

```
LR training score : 100.0
LR testing score : 100.0
```

In [40]:
```
LR_y_pred=LR.predict(x_test)
```

XGBRegressor Model:-

In [41]:
```
pip install xgboost
```

```
Requirement already satisfied: xgboost in c:\users\hp_9046\anaconda3\lib\site-packages (2.0.0)
Requirement already satisfied: scipy in c:\users\hp_9046\anaconda3\lib\site-packages (from xgboost) (1.10.1)
Requirement already satisfied: numpy in c:\users\hp_9046\anaconda3\lib\site-packages (from xgboost) (1.21.5)
Note: you may need to restart the kernel to use updated packages.
```

In [42]:
```
from xgboost import XGBRegressor
xgb=XGBRegressor().fit(x_train,y_train)
```

In [43]:
```
print("xgb training score :",round(xgb.score(x_train,y_train),4)*100)
print("xgb testing score :",round(xgb.score(x_test,y_test),4)*100)
```

```
xgb training score : 100.0
xgb testing score : 99.92999999999999
```

In [44]:
```
xgb_y_pred=xgb.predict(x_test)
```

CatBoostRegressor Model:-

In [45]:
```
pip install catboost
```

```
Requirement already satisfied: catboost in c:\users\hp_9046\anaconda3\lib\site-packages (1.2.2)
Requirement already satisfied: scipy in c:\users\hp_9046\anaconda3\lib\site-packages (from catboost) (1.10.1)
Requirement already satisfied: six in c:\users\hp_9046\anaconda3\lib\site-packages (from catboost) (1.16.0)
Requirement already satisfied: graphviz in c:\users\hp_9046\anaconda3\lib\site-packages (from catboost) (0.20.1)
Requirement already satisfied: numpy>=1.16.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from catboost) (1.21.5)
Requirement already satisfied: plotly in c:\users\hp_9046\anaconda3\lib\site-packages (from catboost) (5.6.0)
Requirement already satisfied: matplotlib in c:\users\hp_9046\anaconda3\lib\site-packages (from catboost) (3.5.1)
Requirement already satisfied: pandas>=0.24 in c:\users\hp_9046\anaconda3\lib\site-packages (from catboost) (1.4.2)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from pandas>=0.24->c
atboost) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from pandas>=0.24->catboost)
(2021.3)
Requirement already satisfied: pillow>=6.2.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from matplotlib->catboost)
(9.0.1)Note: you may need to restart the kernel to use updated packages.
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from matplotlib->catboos
t) (1.3.2)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from matplotlib->catboos
t) (4.25.0)
Requirement already satisfied: cycler>=0.10 in c:\users\hp_9046\anaconda3\lib\site-packages (from matplotlib->catboost) (0.
11.0)
Requirement already satisfied: packaging>=20.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from matplotlib->catboost)
(21.3)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from matplotlib->catboost)
(3.0.4)
Requirement already satisfied: tenacity>=6.2.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from plotly->catboost) (8.
0.1)
```

In [46]:
```
from catboost import CatBoostRegressor
CBR=CatBoostRegressor(verbose=False).fit(x_train,y_train)
```

In [47]:
```
print("CBR training score :",round(CBR.score(x_train,y_train),4)*100)
print("CBR testing score :",round(CBR.score(x_test,y_test),4)*100)
```

```
CBR training score : 100.0
CBR testing score : 99.89
```

In [48]:
```
CBR_y_pred=CBR.predict(x_test)
```

GradientBoostingRegressor Model:-

In [49]:
```
from sklearn.ensemble import GradientBoostingRegressor
GBR = GradientBoostingRegressor().fit(x_train,y_train)
```

In [50]:
```
print("GBR training score :",round(GBR.score(x_train,y_train),3)*100)
print("GBR testing score :",round(GBR.score(x_test,y_test),4)*100)
```

```
GBR training score : 100.0
GBR testing score : 99.99
```

In [51]:
```
GBR_y_pred=GBR.predict(x_test)
```

Models evaluation

In [52]:
```
from sklearn.metrics import mean_absolute_error,mean_squared_error,median_absolute_error
```

In [53]:
```
models_predictions={'LR':LR_y_pred,'xgb':xgb_y_pred,'CBR':CBR_y_pred,'GBR':GBR_y_pred}
metrics={'mean_absolute_error':mean_absolute_error,'mean_squared_error':mean_squared_error,'median_absolute_error':median_abs

for model,y_pred in models_predictions.items():
    print(model,'Model :-','\n')
    for m,metric in metrics.items():
        MetricValue = round(metric(y_test, y_pred),2)
        print( m ,' Value is : ',MetricValue,'\n\n')
```

```
LR Model :-

mean_absolute_error  Value is :  0.0


mean_squared_error  Value is :  0.0


median_absolute_error  Value is :  0.0


xgb Model :-

mean_absolute_error  Value is :  3.01


mean_squared_error  Value is :  46.98


median_absolute_error  Value is :  1.89


CBR Model :-

mean_absolute_error  Value is :  4.62


mean_squared_error  Value is :  68.25


median_absolute_error  Value is :  2.91


GBR Model :-

mean_absolute_error  Value is :  1.64


mean_squared_error  Value is :  4.84


median_absolute_error  Value is :  1.2
```

Interpretation one model

In [54]:
```
pip install shap
```

```
Requirement already satisfied: shap in c:\users\hp_9046\anaconda3\lib\site-packages (0.43.0)
Requirement already satisfied: slicer==0.0.7 in c:\users\hp_9046\anaconda3\lib\site-packages (from shap) (0.0.7)
Requirement already satisfied: pandas in c:\users\hp_9046\anaconda3\lib\site-packages (from shap) (1.4.2)
Requirement already satisfied: tqdm>=4.27.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from shap) (4.64.0)
Requirement already satisfied: numba in c:\users\hp_9046\anaconda3\lib\site-packages (from shap) (0.55.1)
Requirement already satisfied: scikit-learn in c:\users\hp_9046\anaconda3\lib\site-packages (from shap) (1.0.2)
Requirement already satisfied: cloudpickle in c:\users\hp_9046\anaconda3\lib\site-packages (from shap) (2.0.0)
Requirement already satisfied: packaging>20.9 in c:\users\hp_9046\anaconda3\lib\site-packages (from shap) (21.3)
Requirement already satisfied: numpy in c:\users\hp_9046\anaconda3\lib\site-packages (from shap) (1.21.5)
Requirement already satisfied: scipy in c:\users\hp_9046\anaconda3\lib\site-packages (from shap) (1.10.1)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in c:\users\hp_9046\anaconda3\lib\site-packages (from packaging>20.
9->shap) (3.0.4)
Requirement already satisfied: colorama in c:\users\hp_9046\anaconda3\lib\site-packages (from tqdm>=4.27.0->shap) (0.4.4)
Requirement already satisfied: llvmlite<0.39,>=0.38.0rc1 in c:\users\hp_9046\anaconda3\lib\site-packages (from numba->shap)
(0.38.0)
Requirement already satisfied: setuptools in c:\users\hp_9046\anaconda3\lib\site-packages (from numba->shap) (61.2.0)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from pandas->shap)
(2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\hp_9046\anaconda3\lib\site-packages (from pandas->shap) (2021.3)
Requirement already satisfied: six>=1.5 in c:\users\hp_9046\anaconda3\lib\site-packages (from python-dateutil>=2.8.1->panda
s->shap) (1.16.0)
Requirement already satisfied: joblib>=0.11 in c:\users\hp_9046\anaconda3\lib\site-packages (from scikit-learn->shap) (1.3.
2)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\hp_9046\anaconda3\lib\site-packages (from scikit-learn->sha
p) (2.2.0)
Note: you may need to restart the kernel to use updated packages.
```

In [55]:
```
import shap

explainer=shap.Explainer(GBR)
shape_values=explainer(x_test)
shap.plots.bar(shape_values,max_display=15)
```