diwali

October 15, 2024

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import plotly.express as px
```

[2]:
```
df=pd.read_csv("Diwali_Sales_Data.csv",encoding = 'unicode_escape')
```

[3]:
```
df.head()
```

[3]:

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupation	Product_Category	Orders	Amount	Status	unnamed1
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtra	Western	Healthcare	Auto	1	23952.0	NaN	NaN
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Pradesh	Southern	Govt	Auto	3	23934.0	NaN	NaN
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	Central	Automobile	Auto	3	23924.0	NaN	NaN
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southern	Construction	Auto	2	23912.0	NaN	NaN
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	Western	Food Processing	Auto	2	23877.0	NaN	NaN

[ ]:
```

```

1 data processing

[4]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11251 entries, 0 to 11250
Data columns (total 15 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   User_ID           11251 non-null  int64
 1   Cust_name         11251 non-null  object
 2   Product_ID        11251 non-null  object
 3   Gender            11251 non-null  object
 4   Age Group         11251 non-null  object
 5   Age               11251 non-null  int64
 6   Marital_Status    11251 non-null  int64
 7   State             11251 non-null  object
 8   Zone              11251 non-null  object
 9   Occupation        11251 non-null  object
 10  Product_Category  11251 non-null  object
 11  Orders            11251 non-null  int64
 12  Amount            11239 non-null  float64
 13  Status            0 non-null      float64
 14  unnamed1          0 non-null      float64
dtypes: float64(3), int64(4), object(8)
memory usage: 1.3+ MB
```

[5]:
```
df.rename(columns= {'Marital_Status':'Married'},inplace = True)
```

[11]:
```
df['Amount'] = df['Amount'].astype(int)
```

[6]:
```
df=df.drop(columns=['Status','unnamed1'],axis=1)
```

[12]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 11239 entries, 0 to 11250
Data columns (total 13 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   User_ID           11239 non-null  int64
 1   Cust_name         11239 non-null  object
 2   Product_ID        11239 non-null  object
 3   Gender            11239 non-null  object
 4   Age Group         11239 non-null  object
 5   Age               11239 non-null  int64
 6   Married           11239 non-null  int64
 7   State             11239 non-null  object
```

```
 8   Zone                11239 non-null  object
 9   Occupation          11239 non-null  object
 10  Product_Category    11239 non-null  object
 11  Orders              11239 non-null  int64
 12  Amount              11239 non-null  int32
dtypes: int32(1), int64(4), object(8)
memory usage: 1.2+ MB
```

[13]:
```
df['Married']=df['Married'].astype(str)
```

[15]:
```
df['Married']=df['Married'].str.replace('0','not_Married').str.
 ↪replace('1','Married')
df=df.dropna()
```

[16]:
```
df.isna().sum()
```

[16]:
```
User_ID             0
Cust_name           0
Product_ID          0
Gender              0
Age Group           0
Age                 0
Married             0
State               0
Zone                0
Occupation          0
Product_Category    0
Orders              0
Amount              0
dtype: int64
```

[17]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 11239 entries, 0 to 11250
Data columns (total 13 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   User_ID           11239 non-null  int64
 1   Cust_name         11239 non-null  object
 2   Product_ID        11239 non-null  object
 3   Gender            11239 non-null  object
 4   Age Group         11239 non-null  object
 5   Age               11239 non-null  int64
 6   Married           11239 non-null  object
 7   State             11239 non-null  object
 8   Zone              11239 non-null  object
 9   Occupation        11239 non-null  object
```

```
 10  Product_Category  11239 non-null  object
 11  Orders            11239 non-null  int64
 12  Amount            11239 non-null  int32
dtypes: int32(1), int64(3), object(9)
memory usage: 1.2+ MB
```

[57]:
```
df.sample(1)
```

[57]:
```
      User_ID Cust_name Product_ID Gender Age Group  Age      Married  \
6321  1000134  Bickford  P00337942      F     26-35   30  not_Married

               State      Zone Occupation       Product_Category  Orders  \
6321  Andhra Pradesh  Southern   Aviation  Electronics & Gadgets       3

      Amount
6321    7865
```

[18]:
```
df['Married'].value_counts()
```

[18]:
```
Married
not_Married    6518
Married        4721
Name: count, dtype: int64
```

2 EDA

2.0.1 1-Sales Trends Analysis

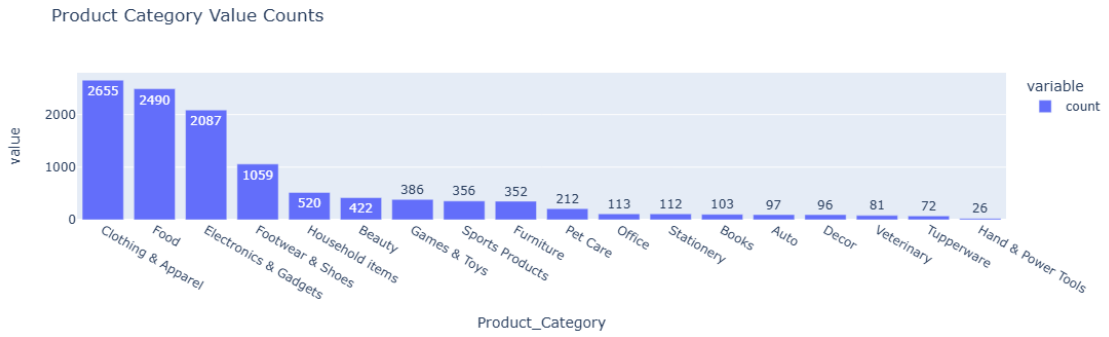
[20]:
```
df['Product_Category'].value_counts()
```

[20]:
```
Product_Category
Clothing & Apparel       2655
Food                     2490
Electronics & Gadgets    2087
Footwear & Shoes         1059
Household items           520
Beauty                    422
Games & Toys              386
Sports Products           356
Furniture                 352
Pet Care                  212
Office                    113
Stationery                112
Books                     103
Auto                       97
Decor                      96
Veterinary                 81
Tupperware                 72
```

```
Hand & Power Tools         26
Name: count, dtype: int64
```

[ ]:

[28]:
```
px.bar(df['Product_Category'].value_counts(), title='Product Category Value Counts',text_auto=True)
```



2.0.2 2-Category-wise Performance

[29]:
```
df.sample()
```

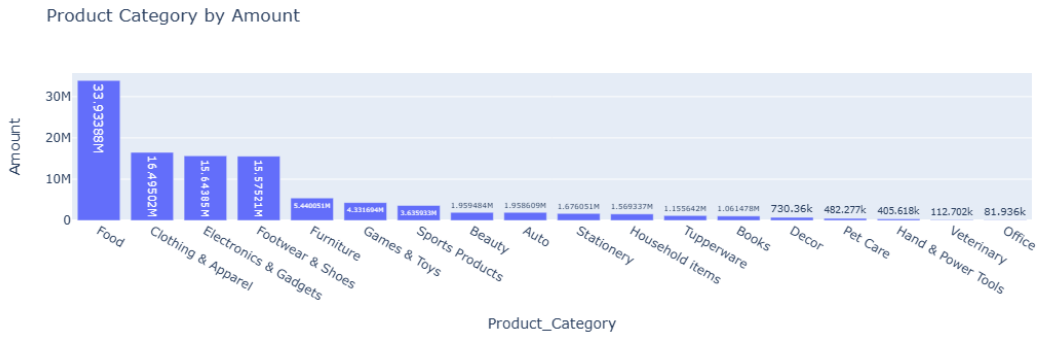
[29]:
```
      User_ID Cust_name Product_ID Gender Age Group  Age      Married  State  \
4749  1000798    Amisha  P00284042      M     26-35   35  not_Married  Delhi

         Zone Occupation    Product_Category  Orders  Amount
4749  Central    Banking  Clothing & Apparel       1    8873
```

[33]:
```
data=df.groupby('Product_Category',as_index=False)['Amount'].sum()
data=data.sort_values(by='Amount',ascending=False)
data
```

[33]:
```
          Product_Category    Amount
6                     Food  33933883
3      Clothing & Apparel  16495019
5   Electronics & Gadgets  15643846
7        Footwear & Shoes  15575209
8               Furniture   5440051
9           Games & Toys   4331694
14       Sports Products   3635933
1                 Beauty   1959484
0                   Auto   1958609
15           Stationery   1676051
11       Household items   1569337
```

```
16            Tupperware   1155642
2                  Books   1061478
4                  Decor    730360
13             Pet Care     482277
10    Hand & Power Tools    405618
17            Veterinary    112702
12                Office     81936
```

[37]:
```
px.bar(data_frame=data,x='Product_Category',y='Amount',title='Product Category␣
↪by Amount',text_auto=True)
```



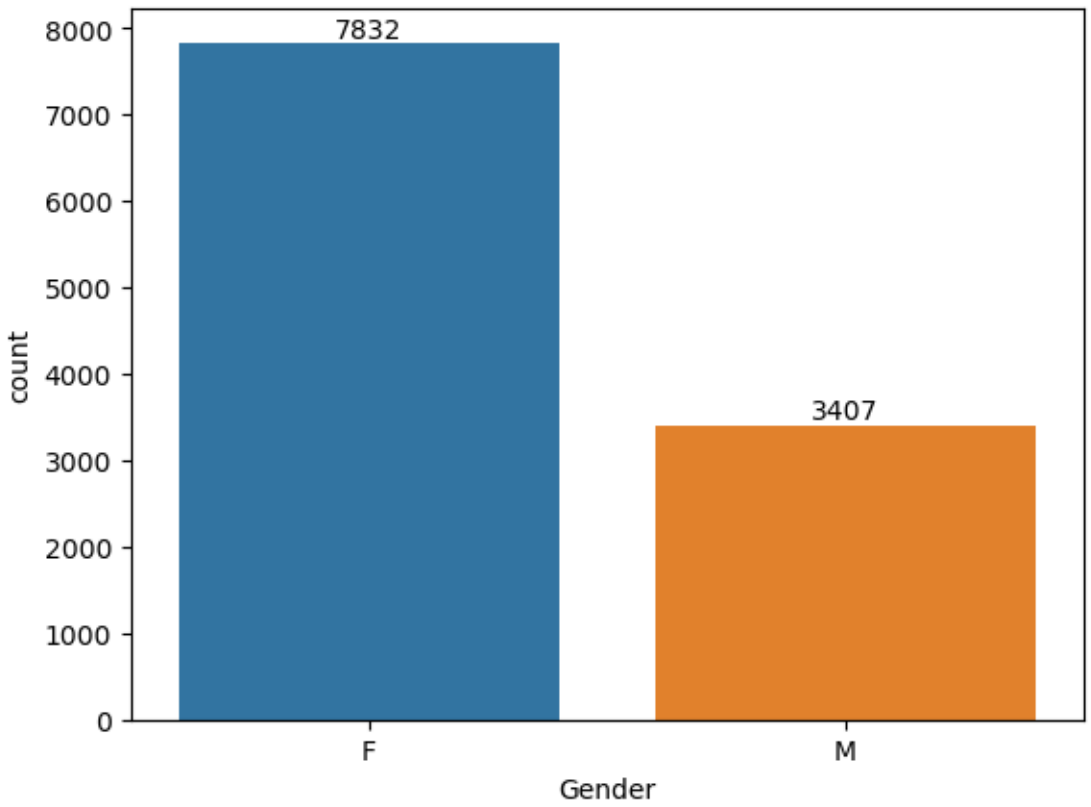
[ ]:
```

```

2.0.3 3- Customer Demographics

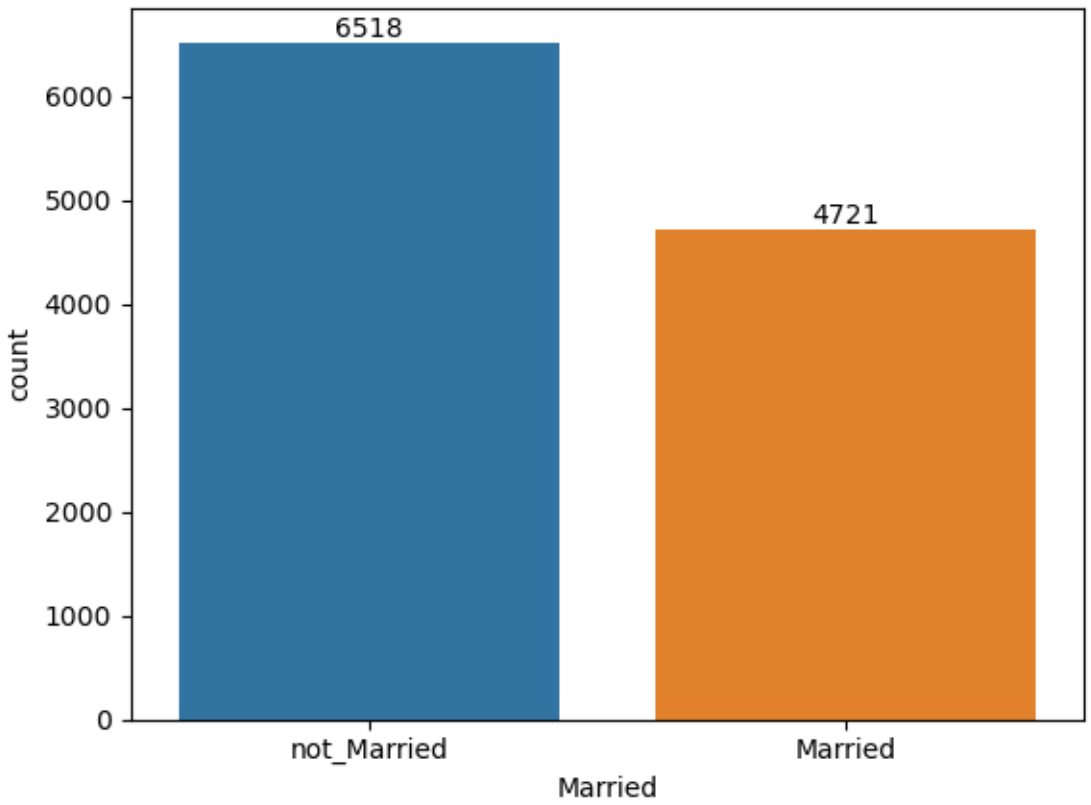
[38]:
```
df.sample()
```

[38]:
```
      User_ID Cust_name Product_ID Gender Age Group  Age      Married   State  \
8691  1002010    Hunter  P00000742      F     18-25   21  not_Married  Kerala

          Zone Occupation    Product_Category  Orders  Amount
8691  Southern    Banking  Clothing & Apparel       4    5324
```

[41]:
```
ax = sns.countplot(data=df,x='Gender')
for bars in ax.containers:
    ax.bar_label(bars)
```



[42]:
```
ax = sns.countplot(data=df,x='Married')
for bars in ax.containers:
    ax.bar_label(bars)
```

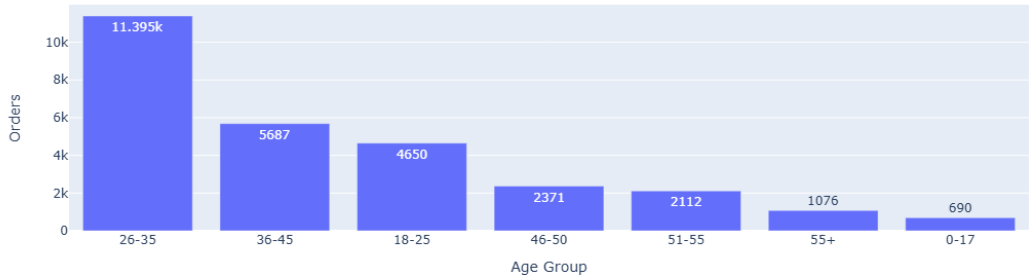


[52]:
```
data=df.groupby('Age Group',as_index=False)['Orders'].sum()
data=data.sort_values(by='Orders',ascending=False)
data
```

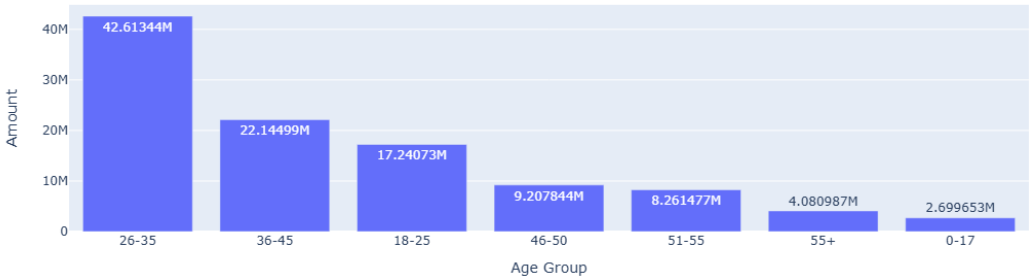
[52]:

	Age Group	Orders
2	26-35	11395
3	36-45	5687
1	18-25	4650
4	46-50	2371
5	51-55	2112
6	55+	1076
0	0-17	690

[53]:
```
px.bar(data_frame=data,x='Age Group',y='Orders',text_auto=True)
```



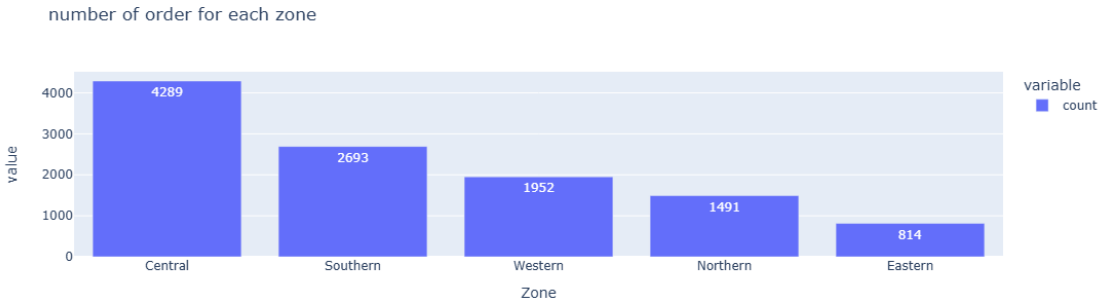
[55]:
```
data=df.groupby('Age Group',as_index=False)['Amount'].sum()
data=data.sort_values(by='Amount',ascending=False)
px.bar(data_frame=data,x='Age Group',y='Amount',text_auto=True)
```



[58]:
```
df['Zone'].value_counts()
```

[58]:
```
Zone
Central     4289
Southern    2693
Western     1952
Northern    1491
Eastern      814
Name: count, dtype: int64
```

[61]:
```
px.bar(df['Zone'].value_counts(), title='number of order for each␣
 ↪zone',text_auto=True)
```



[62]:
```
df.sample()
```

[62]:
```
      User_ID Cust_name Product_ID Gender Age Group  Age      Married   State  \
8863  1002153      Mike  P00196842      F     18-25   24  not_Married  Kerala

          Zone Occupation    Product_Category  Orders  Amount
8863  Southern    Banking  Clothing & Apparel       2    5240
```

[63]:
```
data=df[df['Married']=='not_Married']
```

[66]:
```
data.sample(10)
```

[66]:
```
       User_ID  Cust_name Product_ID Gender Age Group  Age      Married  \
10053  1003518      Greer  P00304142      M     18-25   24  not_Married
4577   1000329  Shrichand  P00124642      F     36-45   36  not_Married
3137   1000904     Aditi   P00303042      F     36-45   41  not_Married
8510   1005950   Sharelle  P00139542      F     26-35   30  not_Married
9219   1002898   Kotsonis  P00100942      F     18-25   19  not_Married
437    1001669     George  P00053842      M     26-35   32  not_Married
3397   1005333  MacIntyre  P00003642      M     26-35   30  not_Married
10630  1002895      Carol  P00303142      F     26-35   34  not_Married
1529   1004116     Ramesh  P00346142      F     36-45   44  not_Married
785    1001425     Sudevi  P00244042      M      0-17   17  not_Married

                State      Zone    Occupation       Product_Category  Orders  \
10053          Kerala  Southern  Construction     Clothing & Apparel       4
4577           Kerala  Southern         Media  Electronics & Gadgets       3
3137        Karnataka  Southern      Aviation                   Food       4
8510           Kerala  Southern      Aviation     Clothing & Apparel       2
9219   Madhya Pradesh   Central   Hospitality                 Beauty       3
437    Andhra Pradesh  Southern          Govt                   Food       1
3397          Haryana  Northern    Healthcare                   Food       2
10630       Rajasthan  Northern    Healthcare  Electronics & Gadgets       1
1529      Maharashtra   Western     IT Sector       Footwear & Shoes       2
```

```
785            Kerala  Southern    Healthcare                    Food       2

       Amount
10053    3477
4577     9742
3137    12076
8510     5412
9219     4479
437     19708
3397    11813
10630    2183
1529    16194
785     19228
```

[67]:
```
data['Product_Category'].value_counts()
```

[67]:
```
Product_Category
Clothing & Apparel       1535
Food                     1500
Electronics & Gadgets    1174
Footwear & Shoes          613
Household items           295
Beauty                    242
Sports Products           234
Games & Toys              227
Furniture                 206
Pet Care                  121
Stationery                 65
Office                     63
Auto                       54
Decor                      49
Books                      49
Tupperware                 45
Veterinary                 34
Hand & Power Tools         12
Name: count, dtype: int64
```

[71]:
```
data=df.groupby('Occupation',as_index=False)['Amount'].sum()
data=data.sort_values(by='Amount',ascending=False)
data
```

[71]:
```
         Occupation    Amount
10        IT Sector  14755079
8        Healthcare  13034586
2          Aviation  12602298
3           Banking  10770610
7              Govt   8517212
```

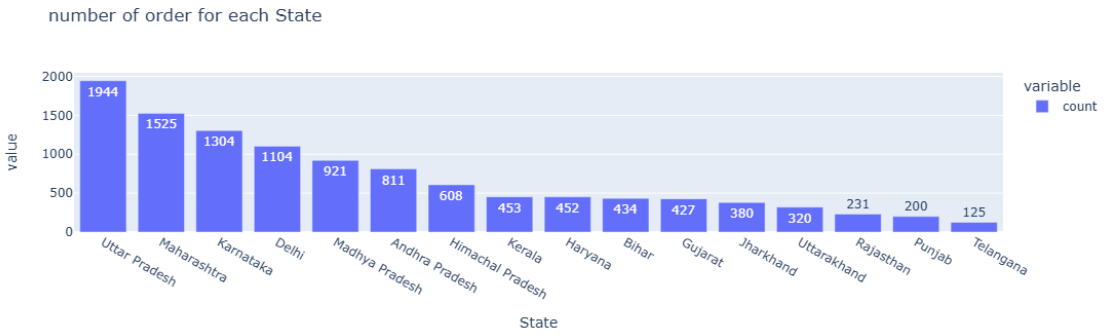
```
9       Hospitality   6376405
12            Media   6295832
1        Automobile   5368596
4          Chemical   5297436
11           Lawyer   4981665
13           Retail   4783170
6   Food Processing   4070670
5      Construction   3597511
14          Textile   3204972
0       Agriculture   2593087
```

[72]:
```
px.bar(data_frame=data,x='Occupation',y='Amount',text_auto=True)
```

[73]:
```
df.State.value_counts()
```

[73]:
```
State
Uttar Pradesh       1944
Maharashtra         1525
Karnataka           1304
Delhi               1104
Madhya Pradesh       921
Andhra Pradesh       811
Himachal Pradesh     608
Kerala               453
Haryana              452
Bihar                434
Gujarat              427
Jharkhand            380
Uttarakhand          320
Rajasthan            231
Punjab               200
Telangana            125
Name: count, dtype: int64
```

[74]:
```
px.bar(df['State'].value_counts(), title='number of order for each␣
↪State',text_auto=True)
```



3 thanks

3.0.1 Mostafa Ashraf

[ ]:
```

```