| import numpy as np import pandas as pd import matplotlib.pyplot as plt # visualizing data %matplotlib inline import seaborn as sns [2]: df = pd.read_csv(r'/Users/adithyachigiri/Documents/Data Analysis/Diwali Sales analysis project/Python_Diwali_Sales_Analysis/Diwali Sales Data.csv', encoding= 'unicode_escape') The dataset contains information about customers, their demographics, and details about the products they purchased, & below is the analysis of Data in accordence with the available Data | • full |
|--|--------|
| | |
| 11246 100695 Manning PO0396942 M 18-25 19 1 Maharashtra Western Chemical Office 4 370 11248 1001099 Reichenbach P00171342 F 38-45 40 0 Mafflya Pradiesh Central Tecilie Office 4 213 11249 1004023 Noonan P00099442 M 36-45 37 0 Karnataka Southern Agriculture Office 3 206 11250 1002744 Brumley P002811742 F 18-25 19 0 Maharashtra Western Healthcare Office 3 188 11239 rows x 13 columns In [5]: df.info() <cl> <cl> <cl> 4. January (1.1. a) (1.1. a</cl></cl></cl> | |
| 1251 nor-null object o | |
| <pre>class 'pandas.core.frame.DataFrame'> Rangefindes: 11254 entries, 0 to 11256 Data columns (total 13 columns):</pre> | |
| Marital_Status | |
| | |
| Marke Mark | |
| Exploratory Data Analysis [27]: | |
| Tender To conder To conder To conder To conder and it's count at a sins.countplot(x = 'Gender', data = df') for bars in ax.containers: | |
| df.groupby(['Gender'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False) ### plotting a bar chart for gender vs total amount ### sales_gen = df.groupby(['Gender'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False) #### sales_gen = df.groupby(['Gender'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False) ################################### | |
| From above graphs we can see that most of the buyers are females and even the purchasing power of females are greater than men | |
| [32]: df.columns | |
| 26-35 0-17 18-25 51-55 46-50 55+ 36-45 Age Group 1 [34]: ax = sns.countplot(data = df, x = 'Age Group', hue = 'Gender') for bars in ax.containers: ax.bar_label(bars) Gender 1 [300 - 1000 | |
| 26-35 0-17 18-25 51-55 46-50 55+ 36-45 Age Group 4000 2000 2000 2000 2000 1000 2000 2000 1000 2000 2000 1000 2000 | |
| From above graphs we can see that most of the buyers are of age group between 26-35 yrs female | |
| State Stat | |
| Uttar Pradesh Maharashtra Karnataka Delhi Madhya Pradesh Andhra Pradesh Himachal Pradesh Kerala Haryana Gujarat State # total amount/sales from top 10 states sales_state = df.groupby(['State'], as_index=False)['Amount'].sum().sert_values(by='Amount', ascending=False).head(19) sns.set(ro=['figure_figisze':(15,5)]) sns.barplot(data = sales_state, x = 'State', y= 'Amount') **Axes: xlabel='Amount'> 167 1.75 1.50 | |
| From above graphs we can see that most of the orders & total sales/amount are from Uttar Pradesh, Maharashtra and Karnataka respectively [48] ax = sns.countplat(data = df, x = 'Marital_Status') sns.set(res('figure-figsize':(7,5))) for bars in ax.containers: ax = sns.countplat(data) ax = sns.containers: ax = sns.countplat(data) ax = sns.containers: ax | |
| 6518 4721 4000 2000 1000 0 Marital_Status | |
| sales_state = of _roroupty[[Marital_Status', 'Gender'], as_index=False]['Amount'].sum().sort_values(by='Amount', ascending=False) sns.set{rec_('rigure-figsize':(6,5))} sns.larplut(data = sales_state, x = 'Marital_Status',y = 'Amount') lef | |
| From above graphs we can see that most of the buyers are married (women) and they have high purchasing power In [1]: # Occupation I [42]: sns.set(re=(*Irgure.flgsize*; (28,5))) ax = sns.countainers: ax.bar_label(bars) 1600 1400 1408 1583 1600 1000 1 | |
| Healthcare Govt Automobile Construction Food Processing Lawyer Media Banking Occupation [43]: sales_state = df.groupby(['occupation'], as_Index=False)['Amount'].sum().sort_values(by='Amount', ascending=False) sns.set(rc=['figure.figsize':(20,5)]) sns.laptpic(data = sales_state, x = 'occupation', y= 'Amount') [43]: Axes: xlabel='Occupation' , ylabel='Amount'> 14 12 10 08 08 | |
| From above graphs we can see that most of the buyers are working in IT, Healthcare and Aviation sector From above graphs we can see that most of the buyers are working in IT, Healthcare and Aviation sector In [1: # Product Category It [44]: sns.set(rc={figure.figsize*:(20,5)}) ax = sns.countplot(data = off, x = Product_Category*) for bars in ax.containers: ax.bur_label(bars) 2655 | |
| 1500 1000 | |
| 1.5 1.0 1.0 1.5 1.0 1.0 1.5 1.0 1.0 1.5 1.0 1.0 1.5 1.0 1.0 1.5 1.0 1.0 1.5 1.0 1.0 1.5 1.0 1.0 1.5 1.0 1.0 1.5 1.0 1.0 1.5 1.0 1.0 1.5 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 | |
| sales_state = df.groupby(['Product_ID'], as_index=False)['Orders'].sum().sort_values(by='Orders', ascending=False).head(10) sns.set(ro='('Igure-figsize':(20,5)}) sns.barplot(data = sales_state, x = 'Product_ID', y= 'Orders') *Axes: xlabel='Product_ID', ylabel='Orders'> 120 80 40 20 0 | |
| P00265242 P00110942 P00237542 P00184942 P00114942 P0017942 P0017942 P00145042 P00044442 P00110842 [47]: # top 18 mast said products (same thing as above) | |
| Conclusion: Married women age group 26-35 yrs from UP, Maharastra and Karnataka working in IT, Healthcare and Aviation are more likely to buy products from Food, Clothing and Electronics category | , |