Scikit-learn (sklearn) ek Python library hai jo machine learning ke algorithms aur data analysis tools ke liye use hoti hai.

Matplotlib

ek Python library hai jo data visualization ke liye graphs aur charts banane mein madad karti hai.

`pyplot`

ek Python library hai jo Matplotlib ka part hai aur data visualization ke liye graphs aur charts banane mein madad karta hai.

Seaborn

ek Python library hai jo statistical data visualization ke liye tools provide karti hai, aur yeh Matplotlib ke upar built hai.

No. of pattern and plots for data visualition

Cell par click karke do baar d ko dabao cell delete hoga

Pandas

ek Python library hai jo data analysis, cleaning aur data manipulation ke liye use hoti hai.

Pandas ek Python ka tool hai jo data ko aasani se samajhne aur badalne mein madad karta hai.

NumPy

ek Python library hai jo numerical computing ke liye high-performance multidimensional arrays aur functions provide karti hai.

Array k liye

**inplace=True**: Ye parameter original DataFrame df ko modify kar deta hai bina nayi copy banaye.

pd.isnull(df) Pandas mein use hota hai taaki DataFrame mein missing values (NaN) ko check kiya ja sake. Yeh function ek nayi DataFrame return karta hai jahan True hai agar value missing hai, aur False agar value hai.

df.dropna(inplace=True) ka use karke DataFrame (df) se sabhi rows ko hata sakte hain jinmein missing values (NaNs) hain. inplace=True ka matlab hai ki yeh change original DataFrame pe hi ho jayega, na ki ek naya DataFrame banega.

pd.isnull(df).sum() Pandas mein use hota hai DataFrame df mein missing values check karne ke liye.

* pd.isnull(df): Yeh function ek DataFrame return karta hai jismein True hoga jahan missing values (NaN) hain aur False jahan nahi hain.
* .sum(): Jab isko boolean DataFrame pe use karte hain (jahan True aur False hote hain), toh True ko 1 aur False ko 0 mein convert karke sum karta hai. Isse har column mein missing values ki total count milti hai.

df\_test.dropna(inplace=True) command pandas mein use hota hai. Ye command missing values (NaN) wale rows ko remove kar deta hai.

inplace=True: Iska matlab hai ki changes seedha original DataFrame (df\_test) pe kiye jayenge, aur ek naya DataFrame create nahi hoga.

Both are same: -

ex = df\_test.dropna()

df\_test.dropna(inplace=True)

Data type change

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

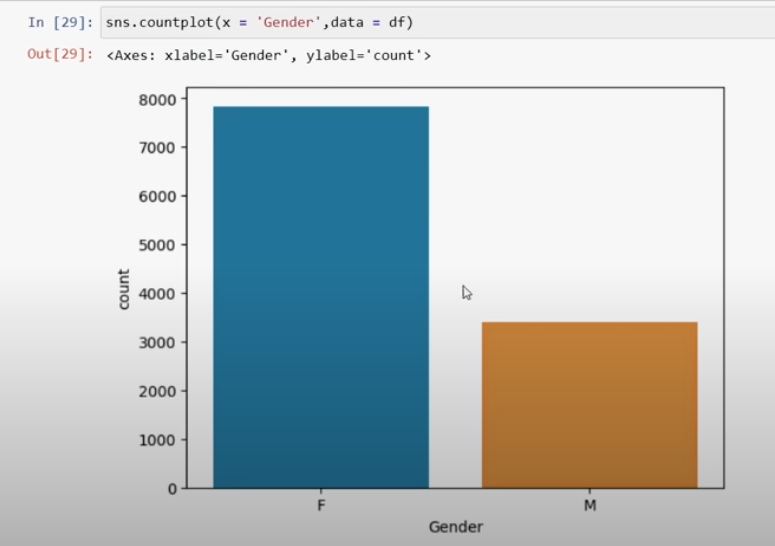
df['Amount'].dtypes -> datatype dikhata h

Df.columns -> columns dikhata h

df.describe() -> saare data ki description dega

df[['Age','Orders','Amount']].describe() -> specific data ki description dega

df.rename(columns = {'Martial\_Status':'Shaadi'}) -> column ko rename karna



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

Aap data ko 'Gender' ke hisaab se group kar rahe hain, fir har group ke 'Amount' ko add kar rahe hain, aur finally unhe descending order me sort kar rahe hain. Isse aapko sabse zyada 'Amount' wale gender pehle milenge. Koi specific detail chahiye?

df.rename(columns={'Qty':'Quantity'}) # rename columns

df[['Qty','Amount']].describe() # use describe for specific columns

# pie chart

B2B\_Check = df['B2B'].value\_counts()

df = pd.read\_csv(r'video\_id\_info.csv', on\_bad\_lines='skip')

jo lines error create karengi, unhe skip kar diya jayega.