

```
In [4]: !pip install numpy  
!pip install pandas  
!pip install matplotlib # install the packages(python)  
!pip install seaborn
```

Requirement already satisfied: numpy in c:\users\sai\anaconda3\lib\site-packages (1.2.4.3)
Requirement already satisfied: pandas in c:\users\sai\anaconda3\lib\site-packages (2.0.3)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\sai\anaconda3\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\sai\anaconda3\lib\site-packages (from pandas) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in c:\users\sai\anaconda3\lib\site-packages (from pandas) (2023.3)
Requirement already satisfied: numpy>=1.21.0 in c:\users\sai\anaconda3\lib\site-packages (from pandas) (1.24.3)
Requirement already satisfied: six>=1.5 in c:\users\sai\anaconda3\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
Requirement already satisfied: matplotlib in c:\users\sai\anaconda3\lib\site-packages (3.7.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\sai\anaconda3\lib\site-packages (from matplotlib) (1.0.5)
Requirement already satisfied: cycler>=0.10 in c:\users\sai\anaconda3\lib\site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\sai\anaconda3\lib\site-packages (from matplotlib) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\sai\anaconda3\lib\site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: numpy>=1.20 in c:\users\sai\anaconda3\lib\site-packages (from matplotlib) (1.24.3)
Requirement already satisfied: packaging>=20.0 in c:\users\sai\anaconda3\lib\site-packages (from matplotlib) (23.1)
Requirement already satisfied: pillow>=6.2.0 in c:\users\sai\anaconda3\lib\site-packages (from matplotlib) (10.0.1)
Requirement already satisfied: pyparsing<3.1,>=2.3.1 in c:\users\sai\anaconda3\lib\site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\sai\anaconda3\lib\site-packages (from matplotlib) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\users\sai\anaconda3\lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.16.0)
Requirement already satisfied: seaborn in c:\users\sai\anaconda3\lib\site-packages (0.12.2)
Requirement already satisfied: numpy!=1.24.0,>=1.17 in c:\users\sai\anaconda3\lib\site-packages (from seaborn) (1.24.3)
Requirement already satisfied: pandas>=0.25 in c:\users\sai\anaconda3\lib\site-packages (from seaborn) (2.0.3)
Requirement already satisfied: matplotlib!=3.6.1,>=3.1 in c:\users\sai\anaconda3\lib\site-packages (from seaborn) (3.7.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\sai\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.0.5)
Requirement already satisfied: cycler>=0.10 in c:\users\sai\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\sai\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\sai\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\users\sai\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (23.1)
Requirement already satisfied: pillow>=6.2.0 in c:\users\sai\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (10.0.1)
Requirement already satisfied: pyparsing<3.1,>=2.3.1 in c:\users\sai\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\sai\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (2.8.2)

```
Requirement already satisfied: pytz>=2020.1 in c:\users\sai\anaconda3\lib\site-packages (from pandas>=0.25->seaborn) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in c:\users\sai\anaconda3\lib\site-packages (from pandas>=0.25->seaborn) (2023.3)
Requirement already satisfied: six>=1.5 in c:\users\sai\anaconda3\lib\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.1->seaborn) (1.16.0)
```

```
In [8]: import numpy as nh
import pandas as ts
import matplotlib.pyplot as plt
import seaborn as skt
```

```
In [9]: import pandas as ts
df1=ts.read_excel("D:\Dataset\KAG1.xls") # Importing Pandas and Loading dataset
df1
```

```
Out[9]:
```

	ad_id	xyz_campaign_id	fb_campaign_id	age	gender	interest	Impressions	Clicks	Sp
0	708746	916	103916	30-34	M	15	7350	1	1.4300
1	708749	916	103917	30-34	M	16	17861	2	1.8200
2	708771	916	103920	30-34	M	20	693	0	0.0000
3	708815	916	103928	30-34	M	28	4259	1	1.2500
4	708818	916	103928	30-34	M	28	4133	1	1.2900
...
1138	1314410	1178	179977	45-49	F	109	1129773	252	358.1899
1139	1314411	1178	179978	45-49	F	110	637549	120	173.8800
1140	1314412	1178	179979	45-49	F	111	151531	28	40.2899
1141	1314414	1178	179981	45-49	F	113	790253	135	198.7100
1142	1314415	1178	179982	45-49	F	114	513161	114	165.6099

1143 rows × 11 columns

```
In [10]: df1.head()
```

Out[10]:

	ad_id	xyz_campaign_id	fb_campaign_id	age	gender	interest	Impressions	Clicks	Spent	Total
0	708746	916	103916	30-34	M	15	7350	1	1.43	
1	708749	916	103917	30-34	M	16	17861	2	1.82	
2	708771	916	103920	30-34	M	20	693	0	0.00	
3	708815	916	103928	30-34	M	28	4259	1	1.25	
4	708818	916	103928	30-34	M	28	4133	1	1.29	

◀ ➡

In [11]: df1.tail()

	ad_id	xyz_campaign_id	fb_campaign_id	age	gender	interest	Impressions	Clicks	Spent	Total
1138	1314410	1178	179977	45-49	F	109	1129773	252	358.1899	
1139	1314411	1178	179978	45-49	F	110	637549	120	173.8800	
1140	1314412	1178	179979	45-49	F	111	151531	28	40.2899	
1141	1314414	1178	179981	45-49	F	113	790253	135	198.7100	
1142	1314415	1178	179982	45-49	F	114	513161	114	165.6099	

◀ ➡

In [12]: df1.describe()

	ad_id	xyz_campaign_id	fb_campaign_id	interest	Impressions	Clicks
count	1.143000e+03	1143.000000	1143.000000	1143.000000	1.143000e+03	1143.000000
mean	9.872611e+05	1067.382327	133783.989501	32.766404	1.867321e+05	33.390201
std	1.939928e+05	121.629393	20500.308622	26.952131	3.127622e+05	56.892438
min	7.087460e+05	916.000000	103916.000000	2.000000	8.700000e+01	0.000000
25%	7.776325e+05	936.000000	115716.000000	16.000000	6.503500e+03	1.000000
50%	1.121185e+06	1178.000000	144549.000000	25.000000	5.150900e+04	8.000000
75%	1.121804e+06	1178.000000	144657.500000	31.000000	2.217690e+05	37.500000
max	1.314415e+06	1178.000000	179982.000000	114.000000	3.052003e+06	421.000000

◀ ➡

```
In [18]: df1.isnull().sum()
```

```
Out[18]: ad_id          0  
xyz_campaign_id      0  
fb_campaign_id        0  
age                  0  
gender               0  
interest              0  
Impressions          0  
Clicks                0  
Spent                 0  
Total_Conversion      0  
Approved_Conversion   0  
dtype: int64
```

```
In [19]: df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 1143 entries, 0 to 1142  
Data columns (total 11 columns):  
 #   Column           Non-Null Count  Dtype    
---  --    
 0   ad_id            1143 non-null    int64  
 1   xyz_campaign_id  1143 non-null    int64  
 2   fb_campaign_id   1143 non-null    int64  
 3   age              1143 non-null    object  
 4   gender            1143 non-null    object  
 5   interest          1143 non-null    int64  
 6   Impressions       1143 non-null    int64  
 7   Clicks            1143 non-null    int64  
 8   Spent              1143 non-null    float64  
 9   Total_Conversion  1143 non-null    int64  
 10  Approved_Conversion 1143 non-null    int64  
dtypes: float64(1), int64(8), object(2)  
memory usage: 98.4+ KB
```

```
In [25]: df1.drop(1139)
```

Out[25]:

	ad_id	xyz_campaign_id	fb_campaign_id	age	gender	interest	Impressions	Clicks	Sp
0	708746	916	103916	30-34	M	15	7350	1	1.4300
1	708749	916	103917	30-34	M	16	17861	2	1.8200
2	708771	916	103920	30-34	M	20	693	0	0.0000
3	708815	916	103928	30-34	M	28	4259	1	1.2500
4	708818	916	103928	30-34	M	28	4133	1	1.2900
...
1137	1314409	1178	179976	45-49	F	108	212508	33	47.6900
1138	1314410	1178	179977	45-49	F	109	1129773	252	358.1899
1140	1314412	1178	179979	45-49	F	111	151531	28	40.2899
1141	1314414	1178	179981	45-49	F	113	790253	135	198.7100
1142	1314415	1178	179982	45-49	F	114	513161	114	165.6099

1142 rows × 11 columns

In [26]: `df1.dropna()`

Out[26]:

	ad_id	xyz_campaign_id	fb_campaign_id	age	gender	interest	Impressions	Clicks	Spent
0	708746	916	103916	30-34	M	15	7350	1	1.4300
1	708749	916	103917	30-34	M	16	17861	2	1.8200
2	708771	916	103920	30-34	M	20	693	0	0.0000
3	708815	916	103928	30-34	M	28	4259	1	1.2500
4	708818	916	103928	30-34	M	28	4133	1	1.2900
...
1138	1314410	1178	179977	45-49	F	109	1129773	252	358.1899
1139	1314411	1178	179978	45-49	F	110	637549	120	173.8800
1140	1314412	1178	179979	45-49	F	111	151531	28	40.2899
1141	1314414	1178	179981	45-49	F	113	790253	135	198.7100
1142	1314415	1178	179982	45-49	F	114	513161	114	165.6099

1143 rows × 11 columns



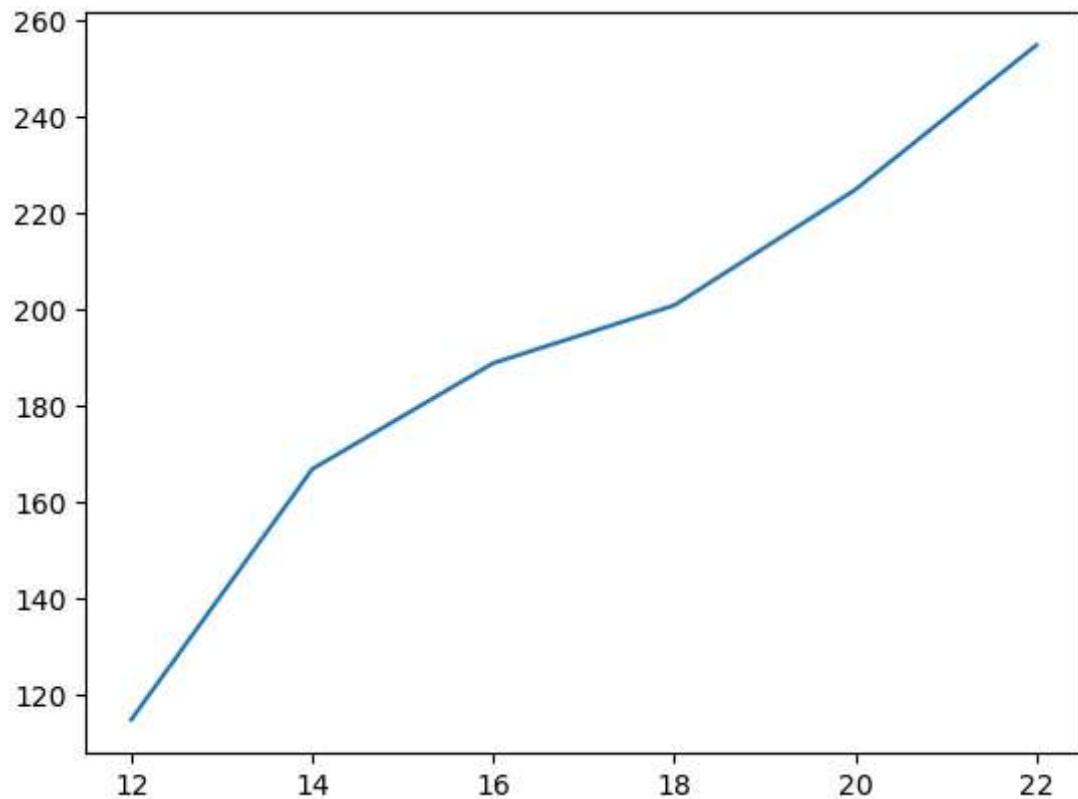
In [28]:	df1.columns
Out[28]:	Index(['ad_id', 'xyz_campaign_id', 'fb_campaign_id', 'age', 'gender', 'interest', 'Impressions', 'Clicks', 'Spent', 'Total_Conversion', 'Approved_Conversion'], dtype='object')
In [33]:	df1.shape
Out[33]:	(1143, 11)
In [35]:	df1.nunique()

```
Out[35]: ad_id          1143
xyz_campaign_id      3
fb_campaign_id       691
age                  4
gender                2
interest              40
Impressions         1130
Clicks                183
Spent                 869
Total_Conversion     32
Approved_Conversion   16
dtype: int64
```

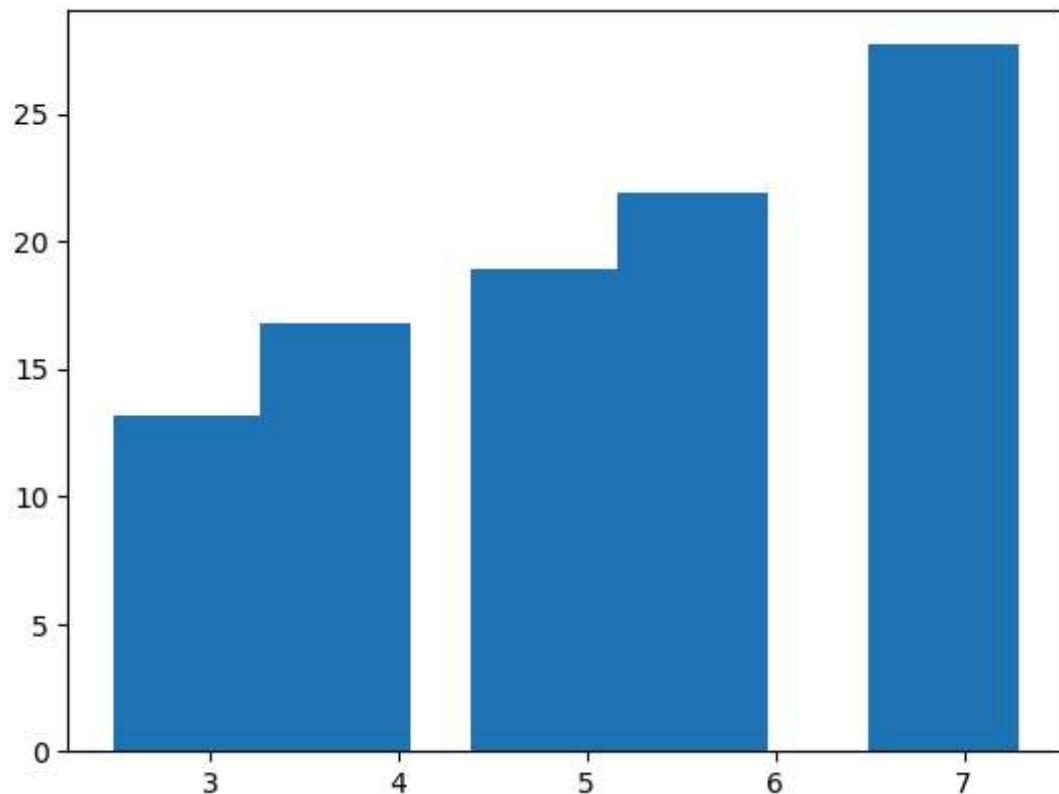
```
In [5]: import pandas as ts
df1=ts.read_excel("D:\\Dataset\\KAG1.xls") # for index Loading the dataset and import p
df1.index
```

```
Out[5]: RangeIndex(start=0, stop=1143, step=1)
```

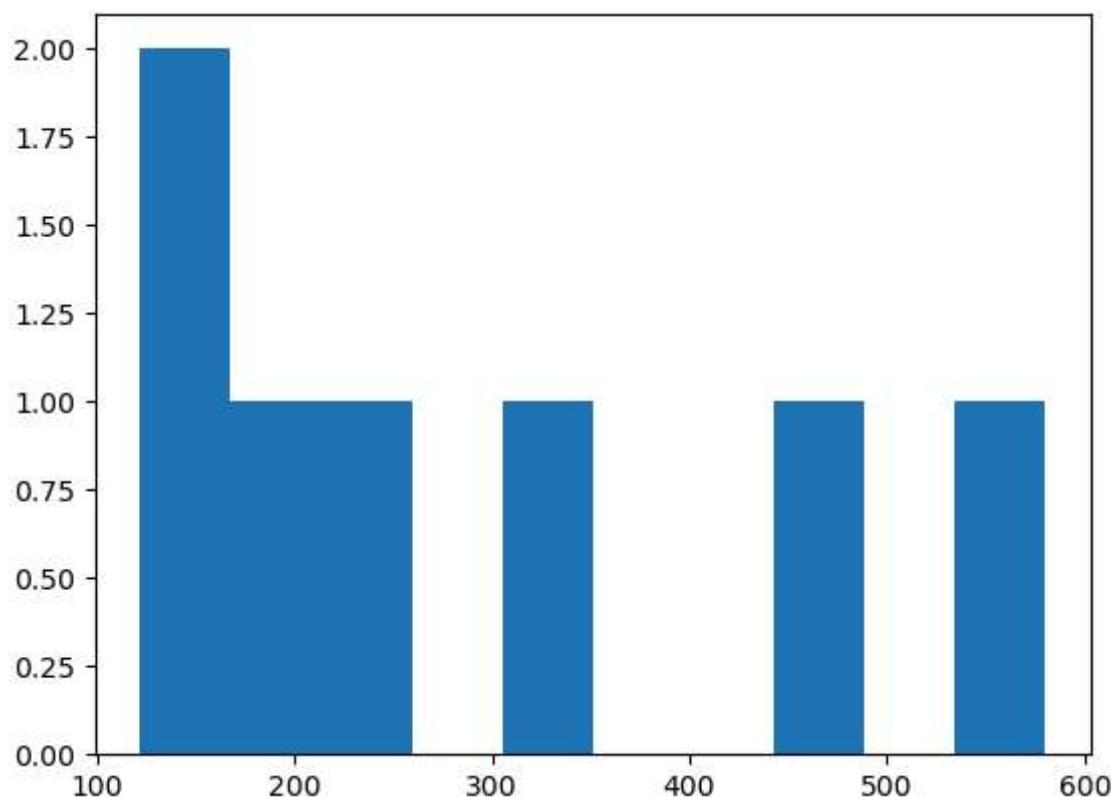
```
In [6]: import matplotlib.pyplot as plt
X = [12,14,16,18,20,22]
Y = [115,167,189,201,225,255]    # Line plot using matplotlib
plt.plot(X,Y)
plt.show()
```



```
In [8]: import matplotlib.pyplot as plt
X = [2.89,3.67,4.78,5.56,6.89]
Y = [13.2,16.8,18.9,21.9,27.7]    # matplotlib using barplot
plt.bar(X,Y)
plt.show()
```

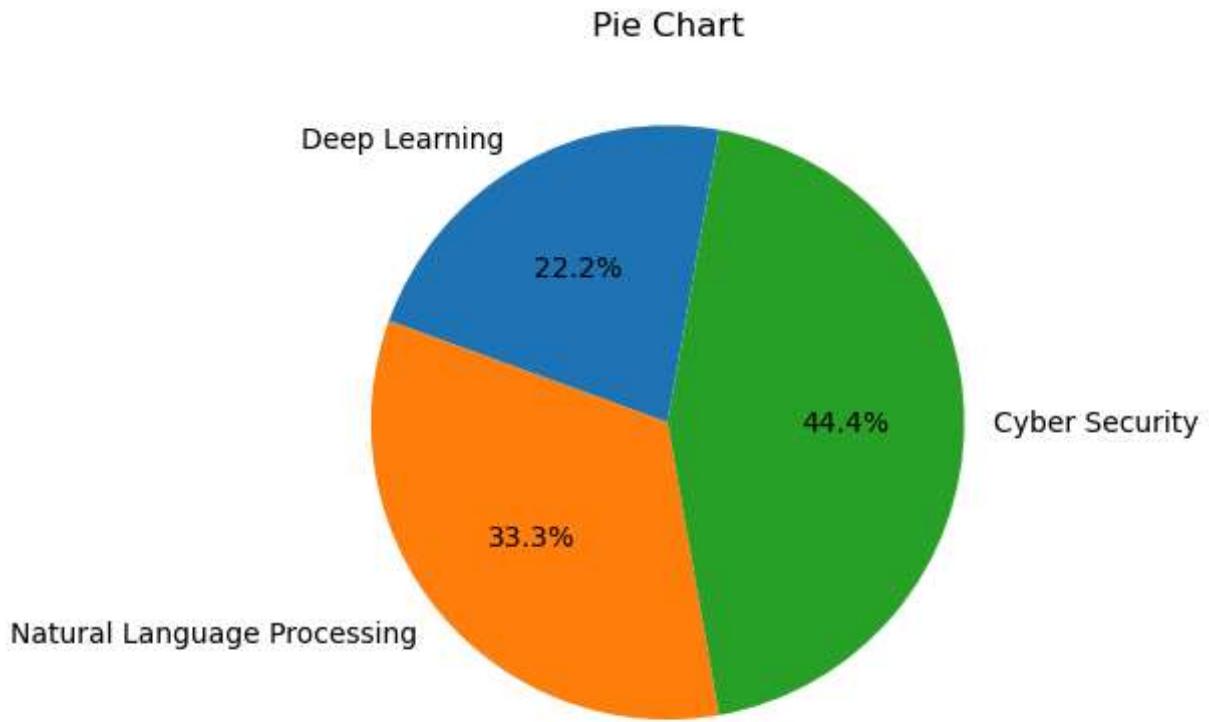


```
In [9]: import matplotlib.pyplot as plt
KN1 = [122,145,189,256,339,459,580]
plt.hist(KN1)                                     #matplotlib using hist and KN1 MEANS Variable
plt.show()
```



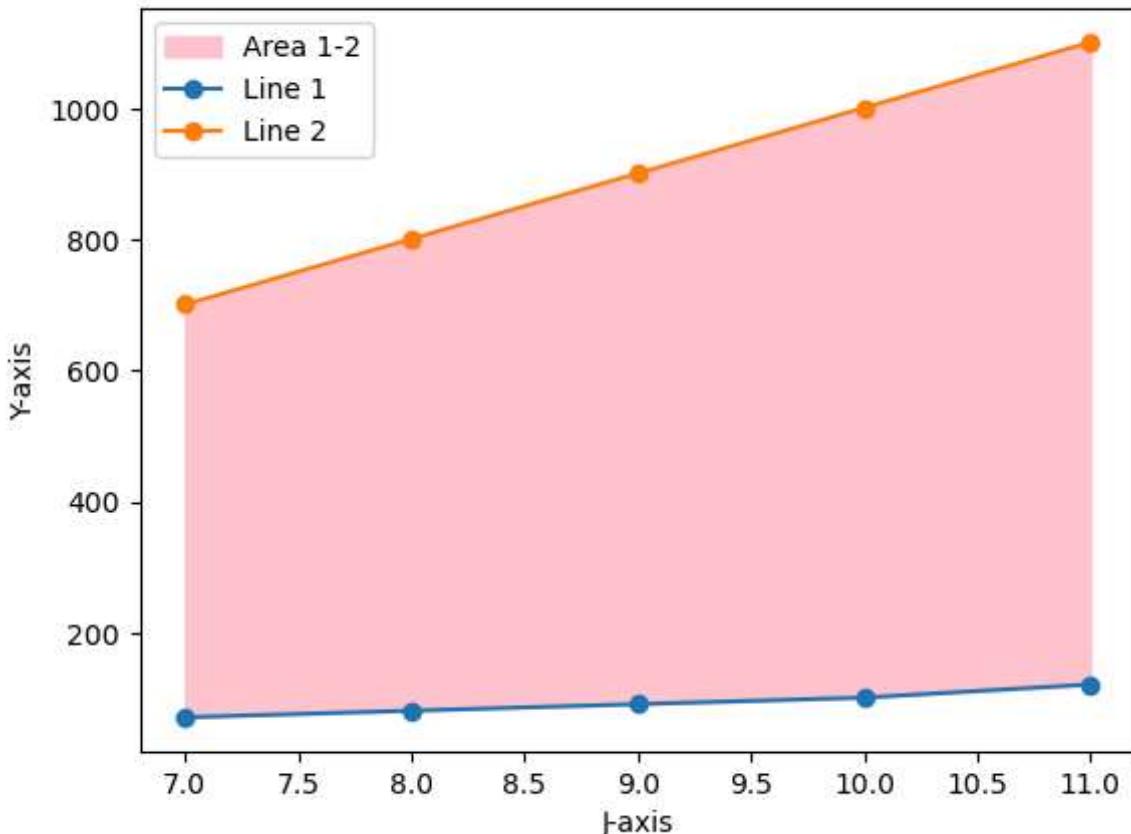
```
In [10]: import matplotlib.pyplot as plt
labels = ["Deep Learning","Natural Language Processing","Cyber Security"]
```

```
sizes = [60,90,120] #matplotlib
plt.pie(sizes,labels=labels,autopct="%1.1f%%",startangle=80)
plt.title("Pie Chart")
plt.show()
```



```
In [16]: import matplotlib.pyplot as plt
J = [7,8,9,10,11]
Y1,Y2 = [71,81,91,101,121],[701,801,901,1001,1101]
plt.fill_between(J,Y1,Y2,color = "Pink",alpha = 0.9,label = 'Area 1-2') #Importing Matplotlib
plt.plot(J,Y1,label='Line 1',marker='o')
plt.plot(J,Y2,label= 'Line 2',marker='o')
plt.xlabel("J-axis"),plt.ylabel("Y-axis"),plt.title("Area Plot")
plt.legend(),plt.show()
```

Area Plot



Out[16]: (`<matplotlib.legend.Legend at 0x19018746dd0>`, None)

In [17]: `import pandas as ts
import seaborn as skt
df1=ts.read_excel("D:\\Dataset\\KAG1.xls") #importing pandas and seaborn and Load the
df1`

Out[17]:

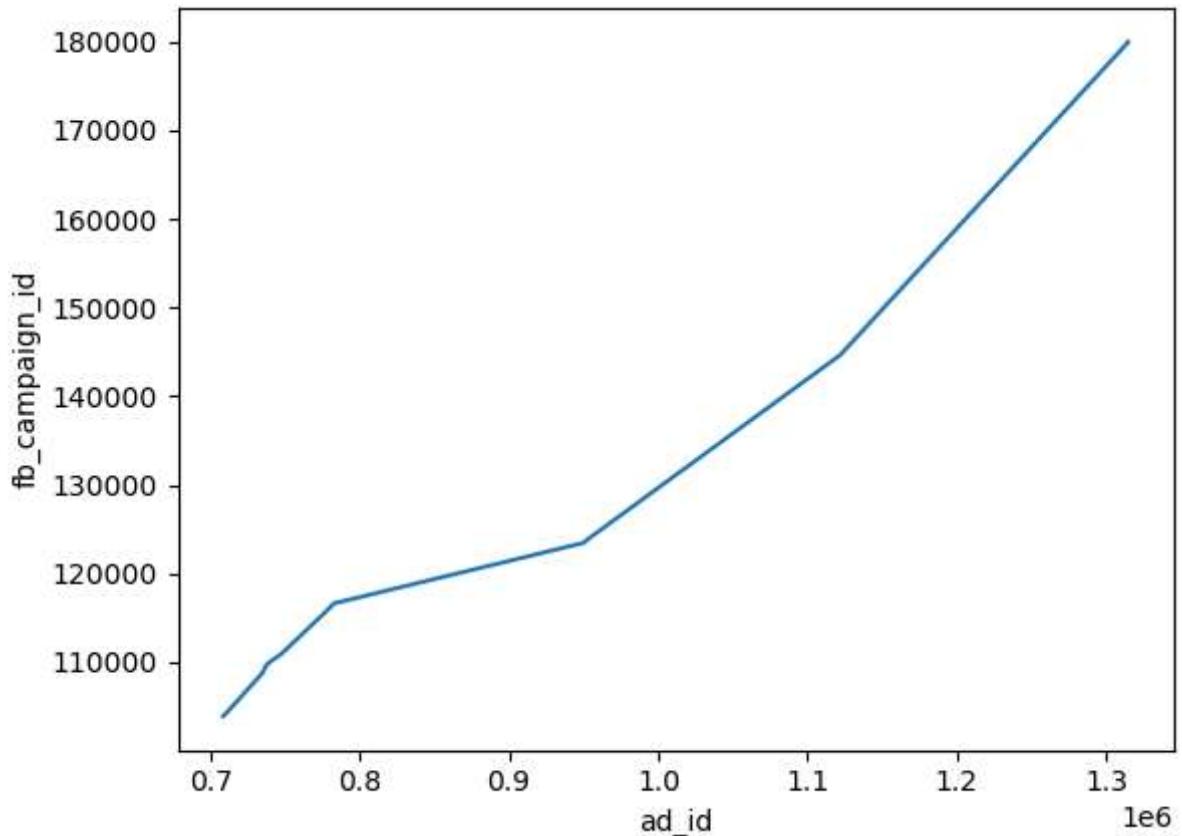
	ad_id	xyz_campaign_id	fb_campaign_id	age	gender	interest	Impressions	Clicks	Sp
0	708746	916	103916	30-34	M	15	7350	1	1.4300
1	708749	916	103917	30-34	M	16	17861	2	1.8200
2	708771	916	103920	30-34	M	20	693	0	0.0000
3	708815	916	103928	30-34	M	28	4259	1	1.2500
4	708818	916	103928	30-34	M	28	4133	1	1.2900
...									
1138	1314410	1178	179977	45-49	F	109	1129773	252	358.1899
1139	1314411	1178	179978	45-49	F	110	637549	120	173.8800
1140	1314412	1178	179979	45-49	F	111	151531	28	40.2899
1141	1314414	1178	179981	45-49	F	113	790253	135	198.7100
1142	1314415	1178	179982	45-49	F	114	513161	114	165.6099

1143 rows × 11 columns



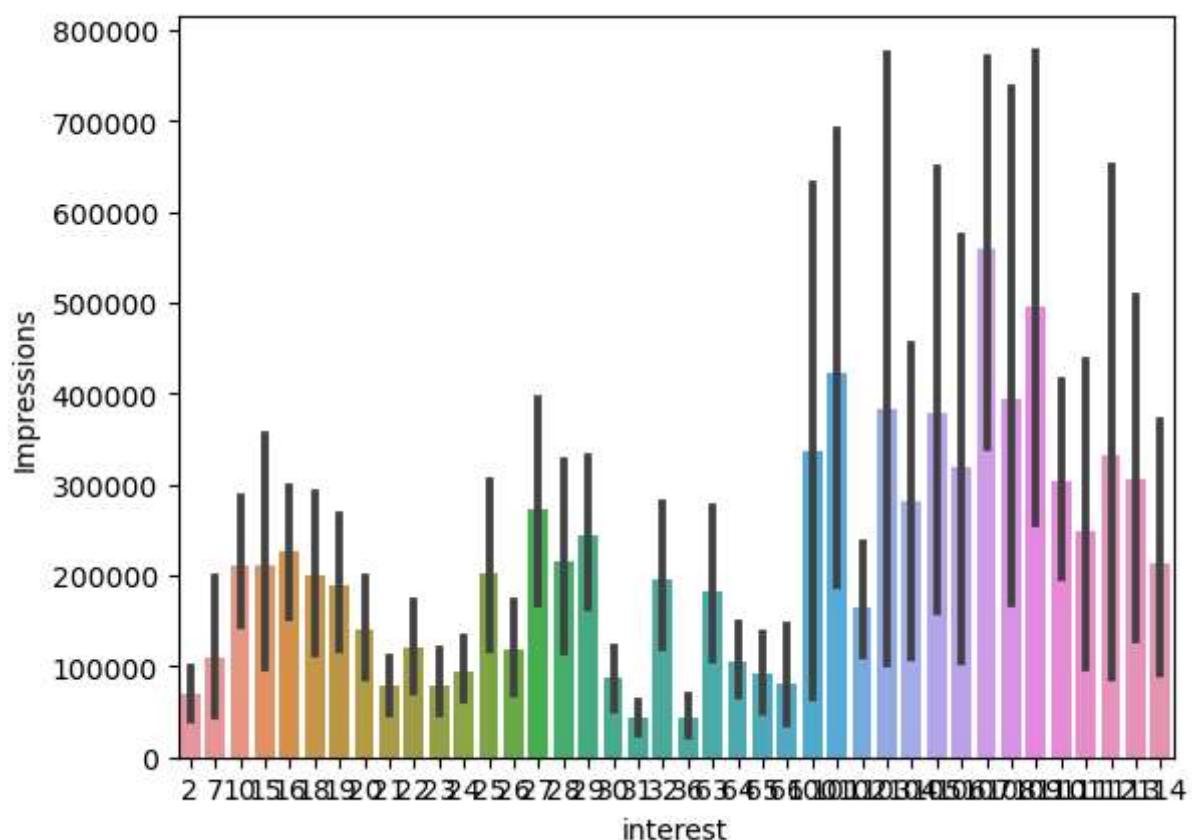
In [18]: `skt.lineplot(x="ad_id",y="fb_campaign_id",data=df1) # lineplot using seaborn(skt)`

Out[18]: `<Axes: xlabel='ad_id', ylabel='fb_campaign_id'>`



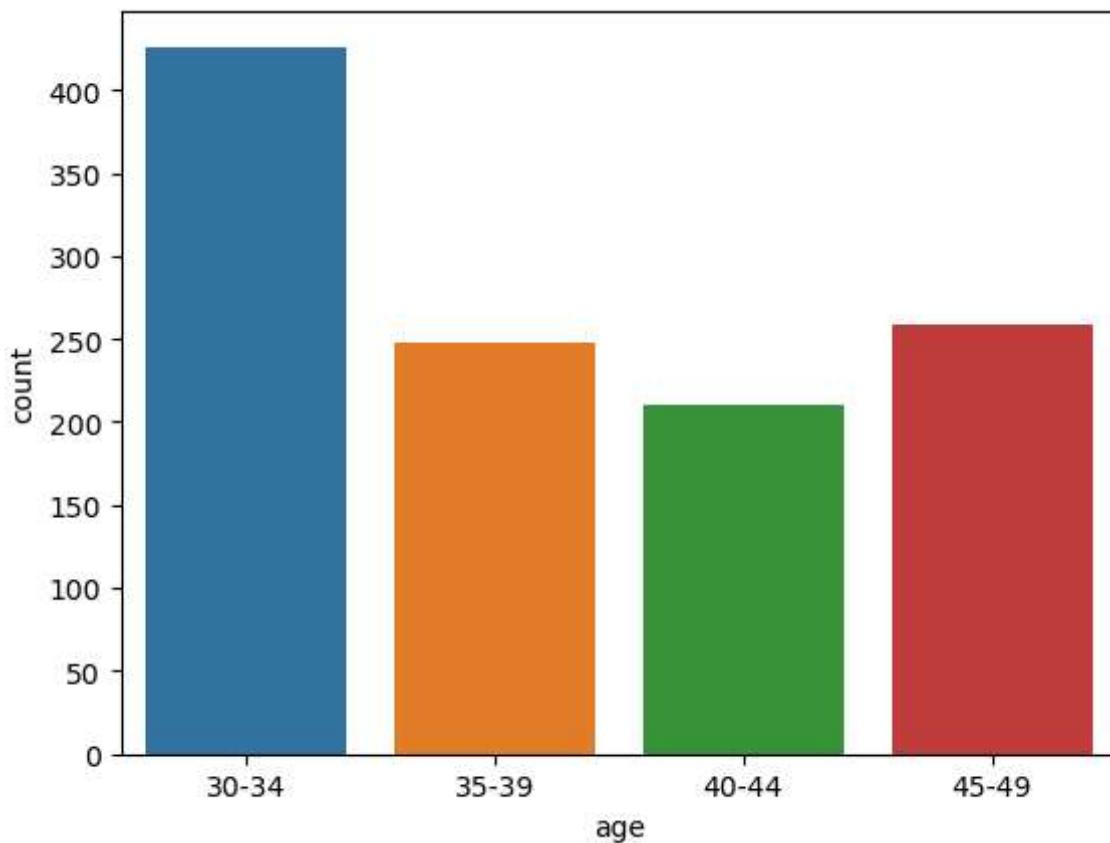
```
In [19]: skt.barplot(x = "interest",y = "Impressions",data=df1) #BarPlot using Seaborn(skt)
```

```
Out[19]: <Axes: xlabel='interest', ylabel='Impressions'>
```



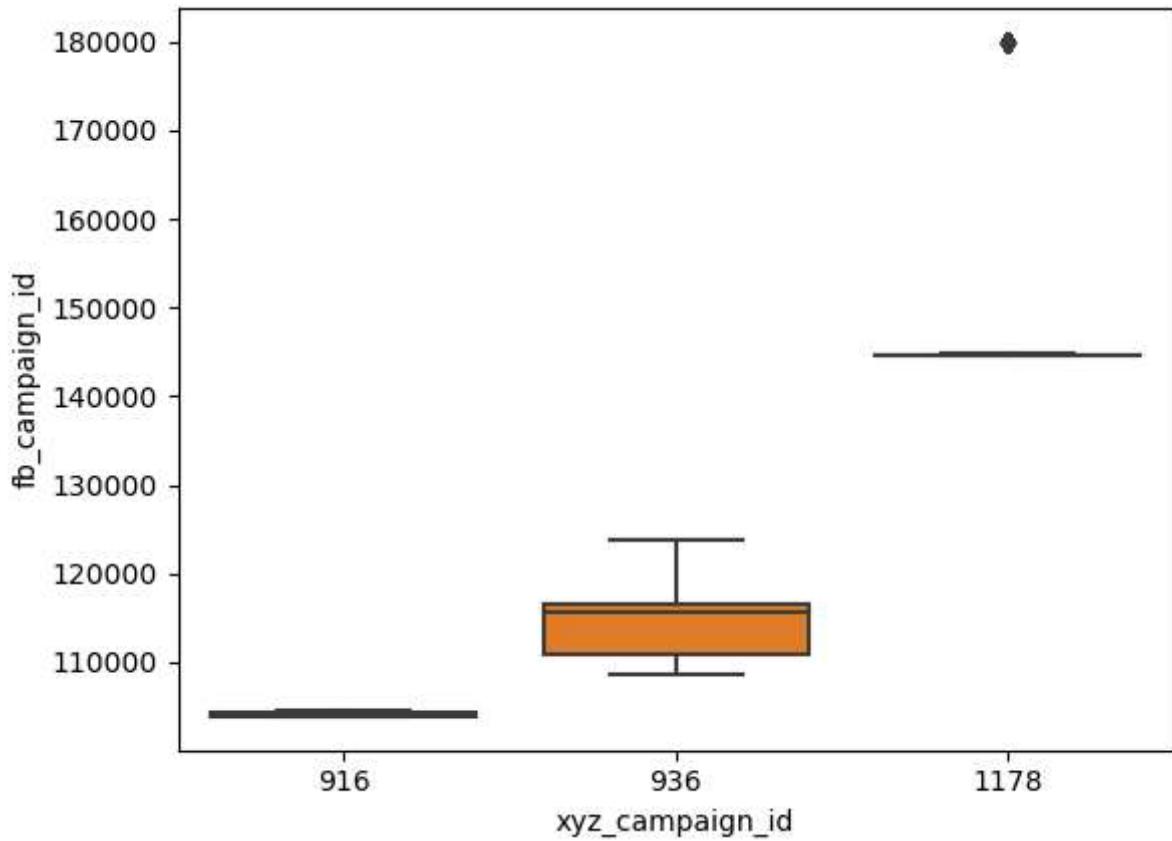
```
In [21]: skt.countplot(x="age",data=df1) # countplot using seaborn(skt)
```

```
Out[21]: <Axes: xlabel='age', ylabel='count'>
```

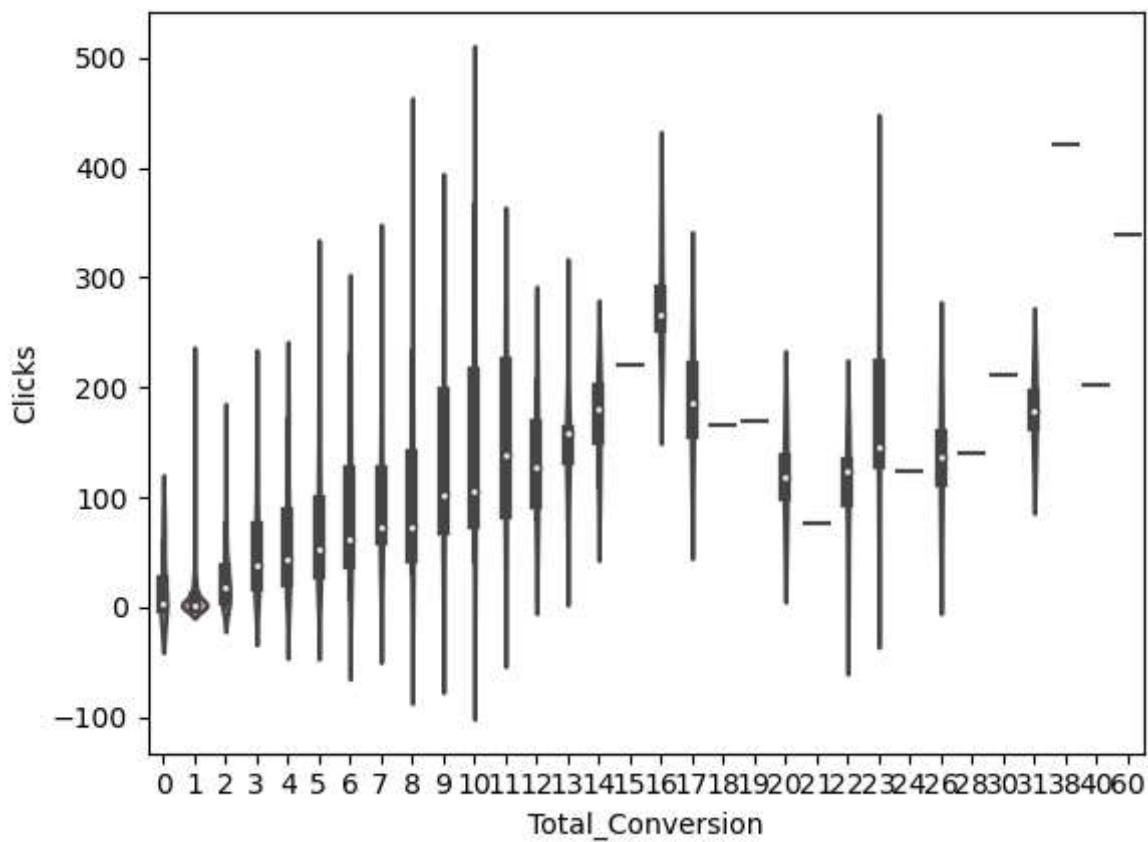


```
In [22]: skt.boxplot(x="xyz_campaign_id",y="fb_campaign_id",data=df1) # boxplot using seaborn
```

```
Out[22]: <Axes: xlabel='xyz_campaign_id', ylabel='fb_campaign_id'>
```

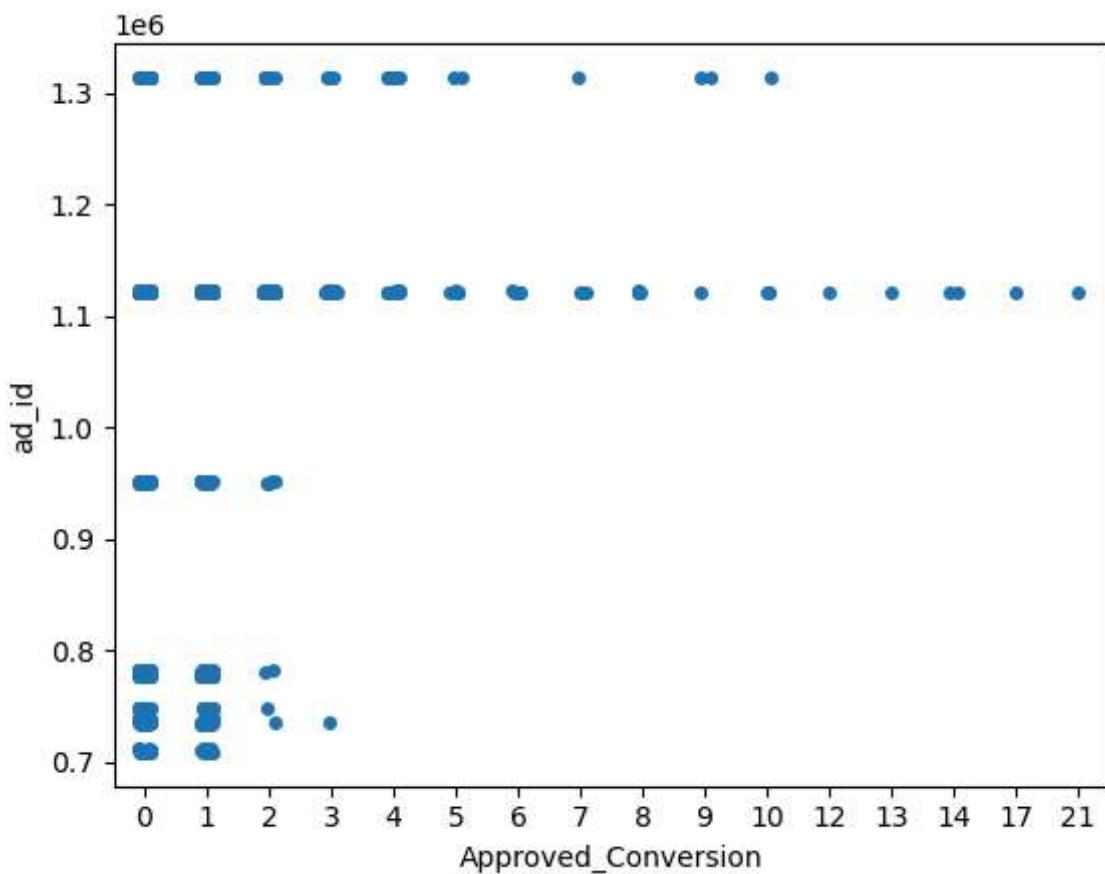


```
In [25]: skt.violinplot(x="Total_Conversion",y="Clicks",data=df1)  
# violin plot using seak  
Out[25]: <Axes: xlabel='Total_Conversion', ylabel='Clicks'>
```



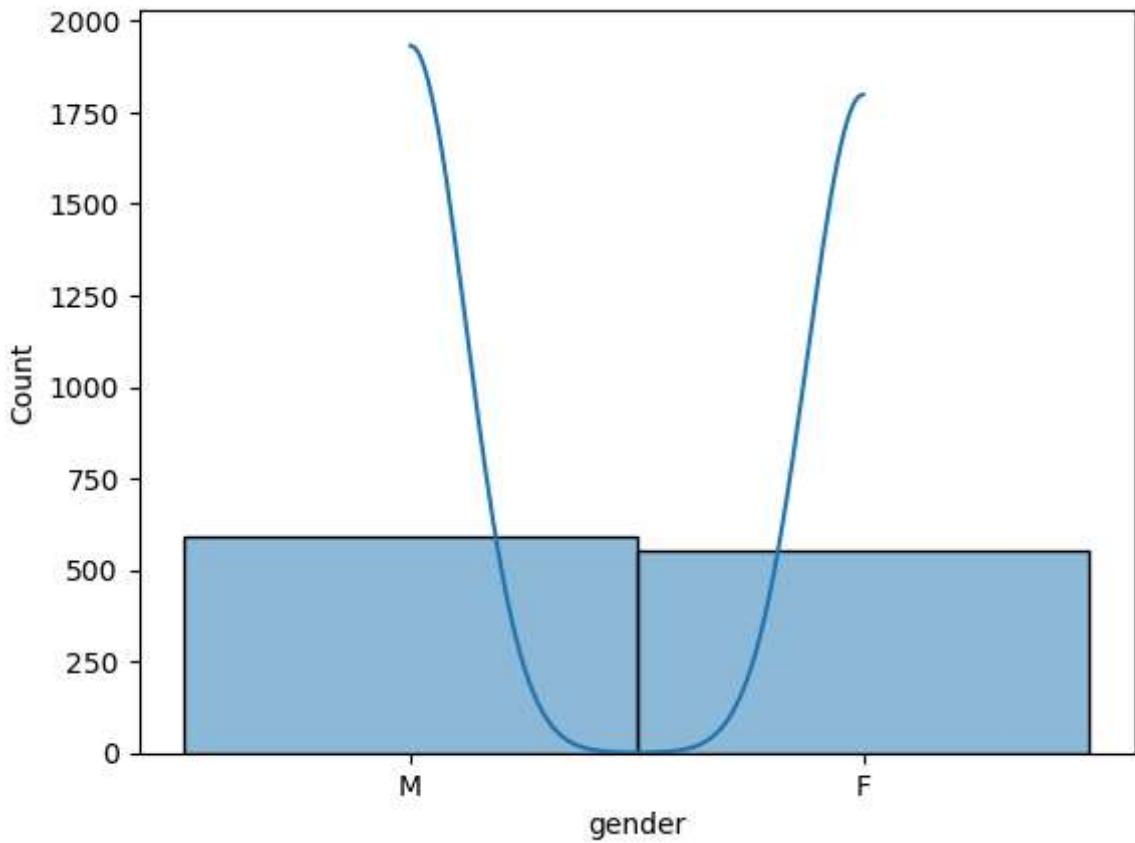
```
In [26]: skt.stripplot(x="Approved_Conversion",y="ad_id",data=df1)      # stripplot using seaborn
```

```
Out[26]: <Axes: xlabel='Approved_Conversion', ylabel='ad_id'>
```



```
In [27]: skt.histplot(data=df1,x="gender",kde=True)    # histplot using seaborn(skt)
```

```
Out[27]: <Axes: xlabel='gender', ylabel='Count'>
```



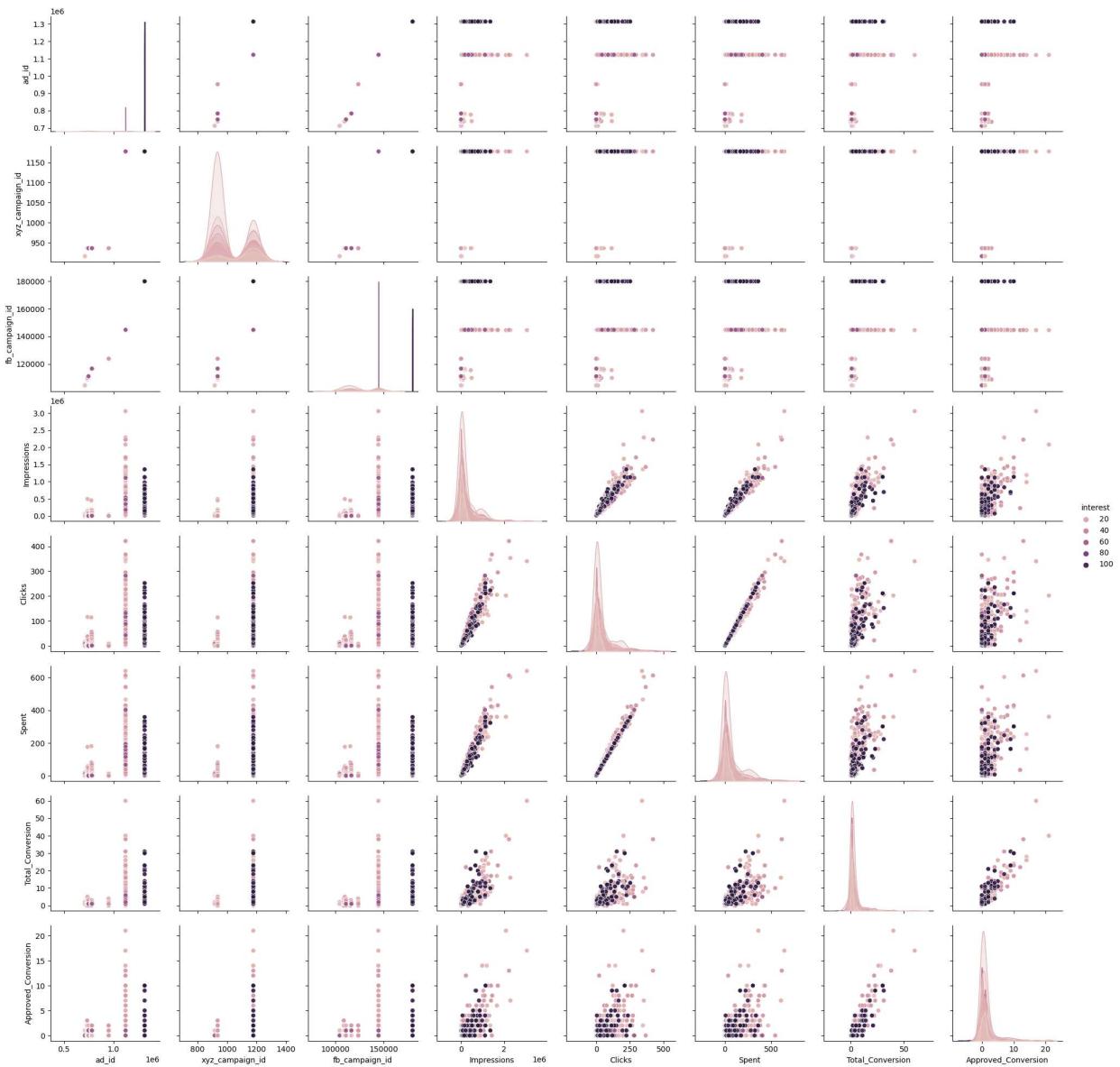
```
In [28]: skt.pairplot(df1,hue="interest") # pairplot using seaborn(skt)
```

```
C:\Users\SAI\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to tight
```

```
    self._figure.tight_layout(*args, **kwargs)
```

```
<seaborn.axisgrid.PairGrid at 0x19020641e10>
```

```
Out[28]:
```



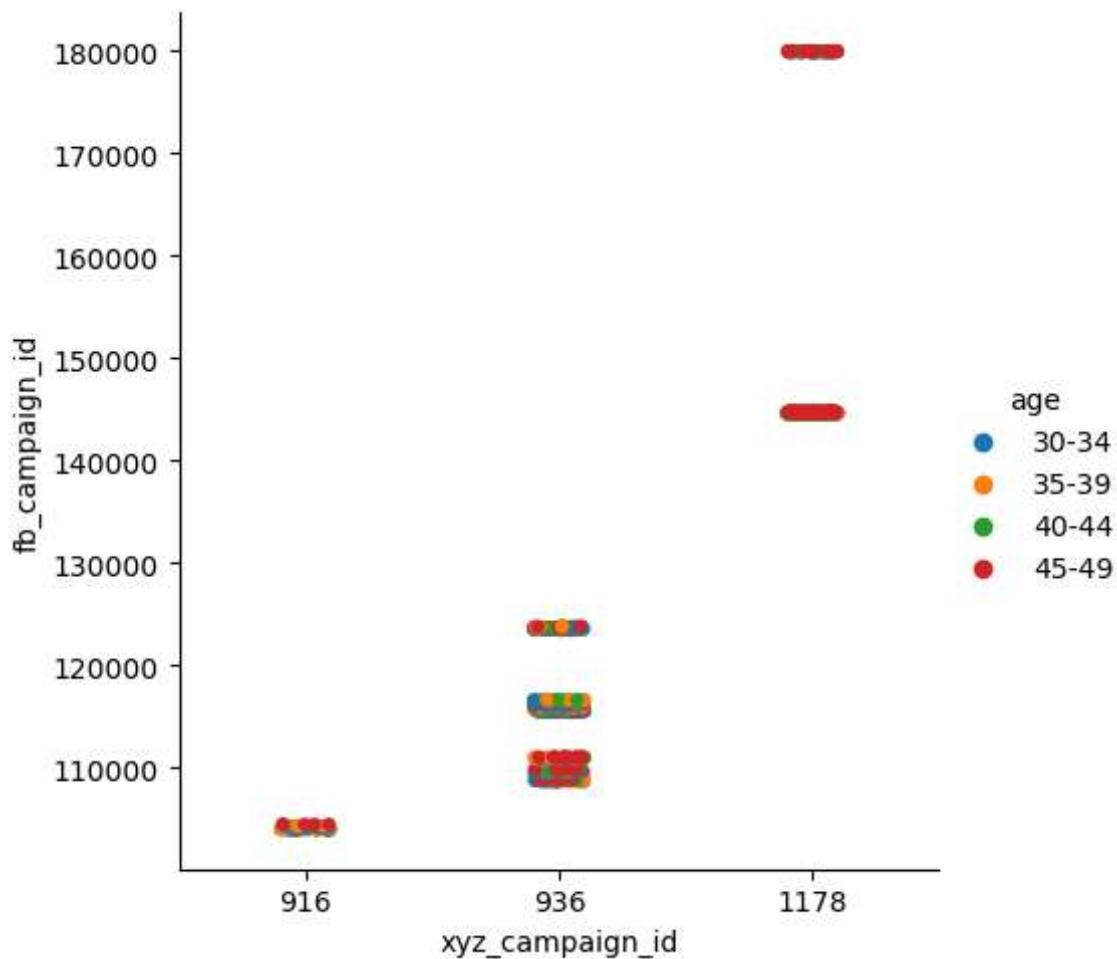
```
In [29]: skt.catplot(x="xyz_campaign_id",y="fb_campaign_id",hue="age",data=df1) #catplot using
```

```
C:\Users\SAI\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to tight
```

```
    self._figure.tight_layout(*args, **kwargs)
```

```
<seaborn.axisgrid.FacetGrid at 0x190254f9090>
```

```
Out[29]:
```



In []: