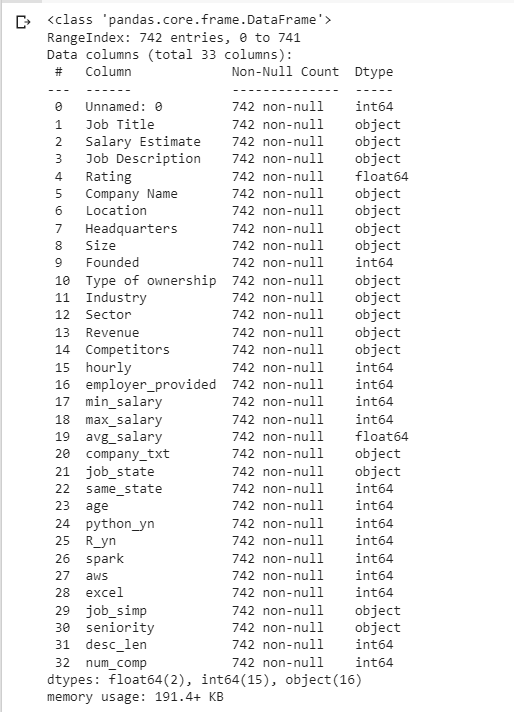
MINI PROJECT

**Objective:**

To analyse and perform various plotting technique in Salary Prediction Dataset.

**Data Description:**

df.info()



**Program:**

from google.colab import drive

drive.mount('/content/drive')

import pandas as pd

import numpy as np

df=pd.read\_csv("/content/drive/MyDrive/eda\_data.csv")

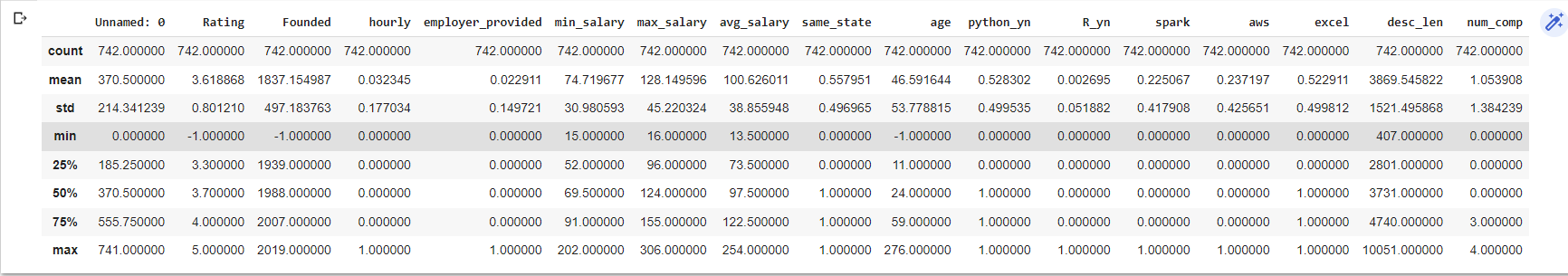
df.head()



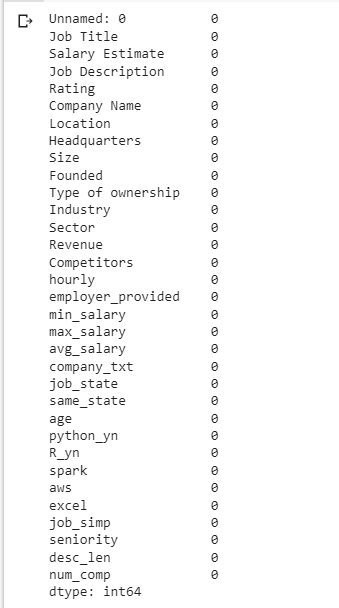
df.shape



df.describe()



df.isnull().sum()

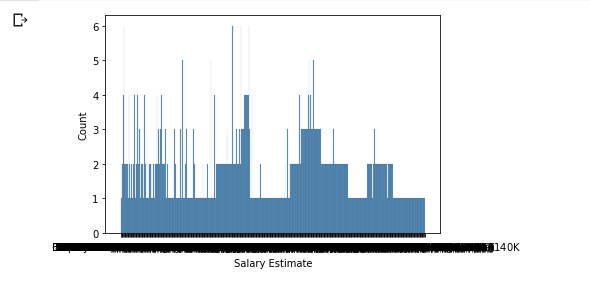


import seaborn as sns

import matplotlib.pyplot as plt

sns.histplot(x='Salary Estimate', data=df, )

plt.show()

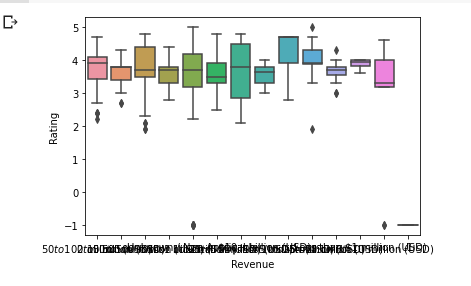


**Observation:**

* The count of 1 member earning 140k.
* The count of 6 member earning 100k.

sns.boxplot( x="Revenue", y='Rating', data=df, )

plt.show()



**Observation:**

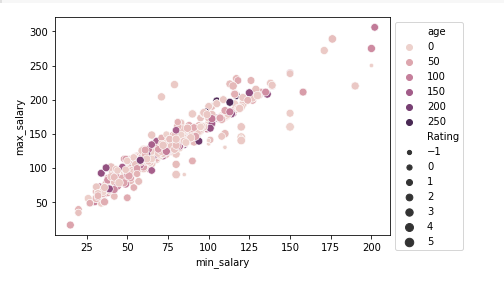
* The average mean of rating with revenue of $50 to $100 million(USD) is 4.

sns.scatterplot( x="min\_salary", y='max\_salary', data=df,

                hue='age', size='Rating')

plt.legend(bbox\_to\_anchor=(1, 1), loc=2)

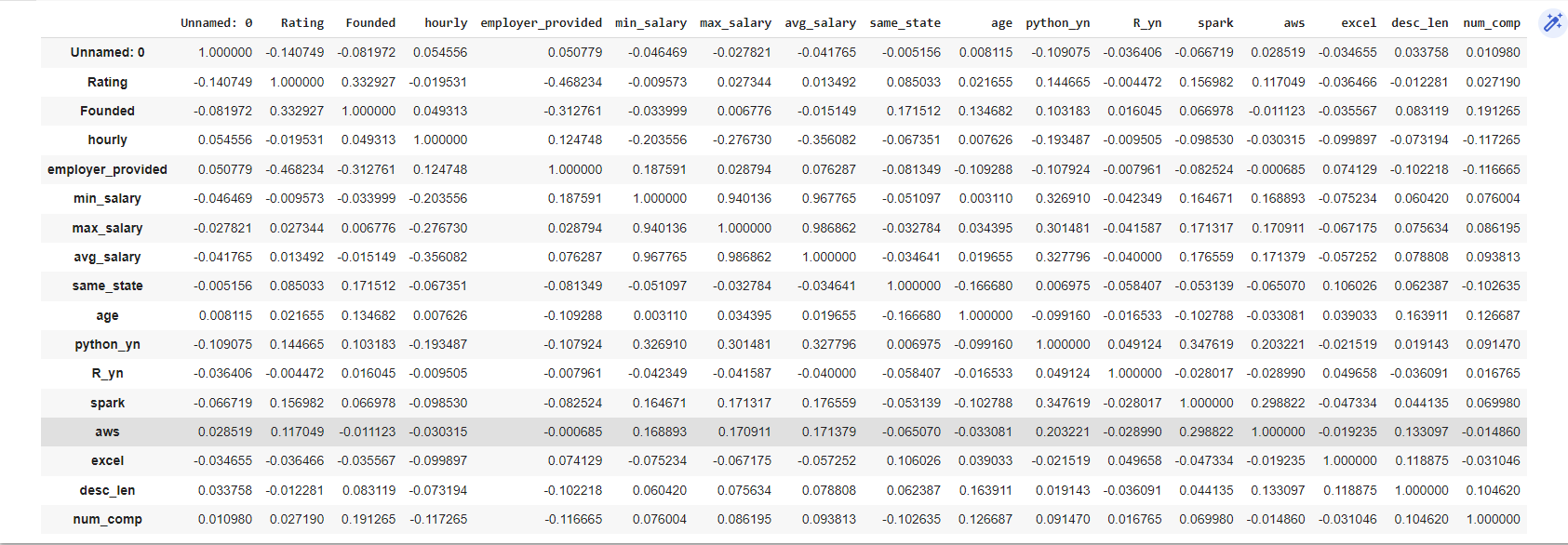
plt.show()



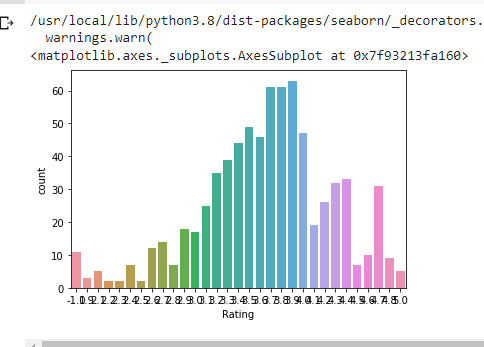
**Observation:**

.The persons with age 250 whose max\_salary is 250 and min\_salary is 125.

df.corr()



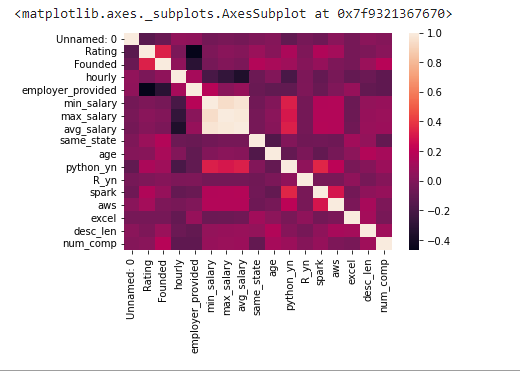
sns.countplot(df['Rating'])



**Observation:**

* Here the highest count is 65 with rating 8.O

sns.heatmap(df.corr())



**Observation:**

* From this map ,it is observed that rating is highly correlated with employer\_provided.

df['Rating'].mean()

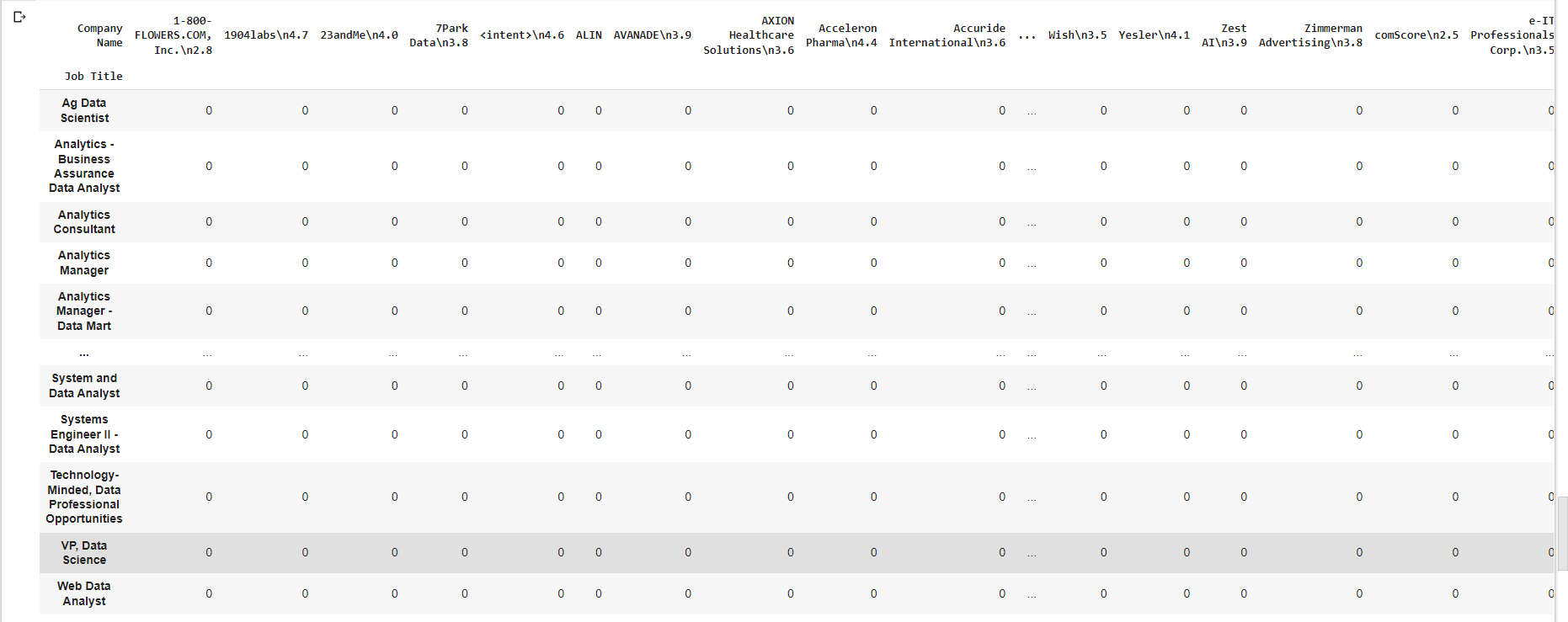


df['Rating'].mode()



contingence=pd.crosstab(index=df["Job Title"],columns=df["Company Name"]

contingence



**Objective:**

To analyse and perform univariate,bivariate and multivariate on time series dataset.

**Program:**

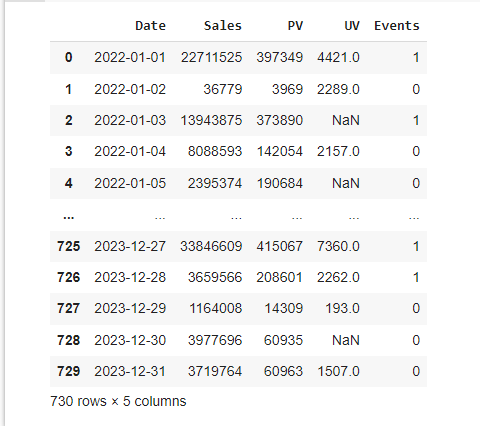
#importing Libraries

import pandas as pd

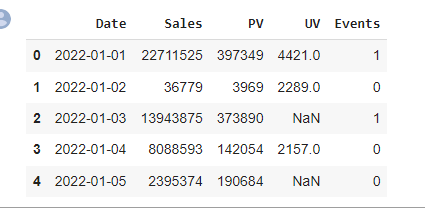
import numpy as np

df=pd.read\_csv('/content/basic2.csv')

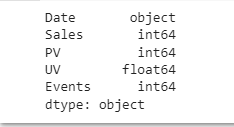
df



df.head()



df.dtypes



import matplotlib.pyplot as plt

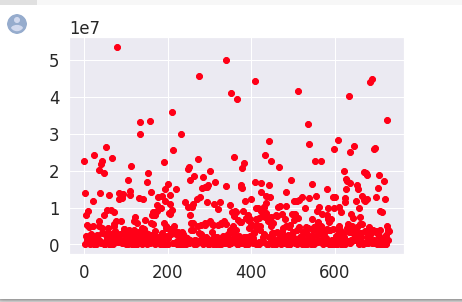
%matplotlib inline

#SEABORN

import seaborn as sns

plt.scatter(df.index,df['Sales'],c='Red')

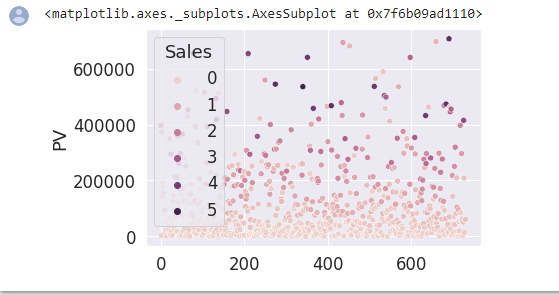
plt.gray()



**Observation:**

* Here the index with 100 is having high sales of 5.

sns.scatterplot (x=df.index, y=df['PV'], hue=df['Sales'])



**Observation**:

The PV of range between 400000 and 600000 ,sales is high.

plt.figure (figsize=(6,6))

plt.title('Line plot of Sales')

plt.xlabel('index', fontsize=20)

plt.ylabel('Sales', fontsize=20)

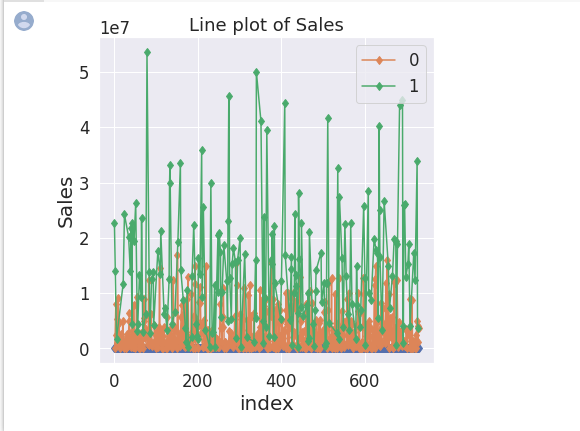
plt.plot(df.index, df ['Events'], markevery=1, marker='d')

for name, group in df.groupby('Events'):

    plt.plot(group.index, group['Sales'], label=name, markevery=1,marker='d')

plt.legend()

plt.show()

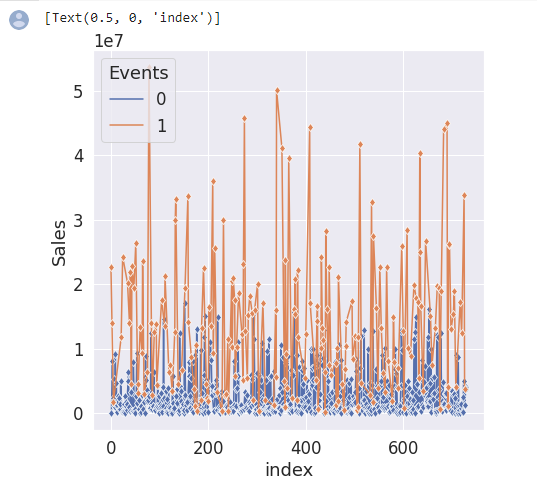


sns.set(rc={'figure.figsize': (7,7)})

sns.set (font\_scale=1.5)

fig=sns.lineplot (x=df.index, y=df['Sales'], markevery=1, marker='d',data=df, hue=df ['Events'])

fig.set(xlabel='index')



**Observation:**

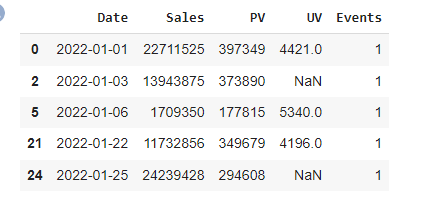
At event 1 ,the sales is high.

**UNIVARIATE ANALYSIS**

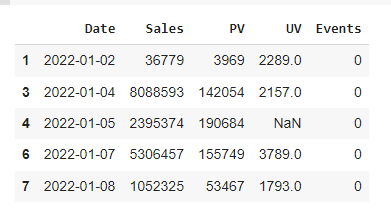
df\_placed = df.loc[df['Events']==1]

df\_Notplaced = df.loc[df['Events']==0]

df\_placed.head()



df\_Notplaced.head()

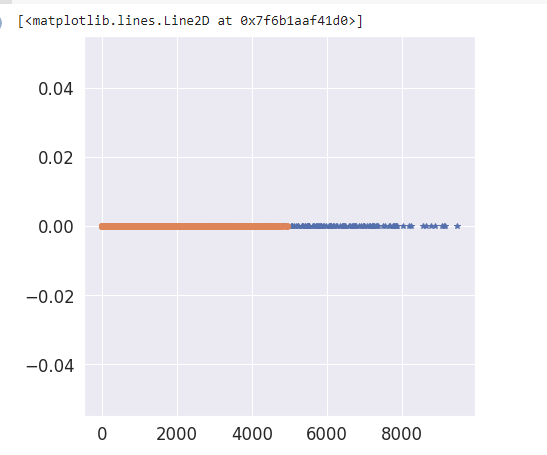


import matplotlib.pyplot as plt

import numpy as np

plt.plot(df\_placed['UV'],np.zeros\_like(df\_placed['Sales']),'\*')

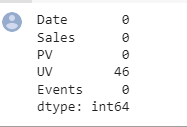
plt.plot(df\_Notplaced['UV'],np.zeros\_like(df\_Notplaced['Sales']),'o')



**BIVARIATE ANALYSIS**

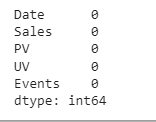
import seaborn as sns

df.isna().sum()



drop = df.dropna()

drop.isna().sum()

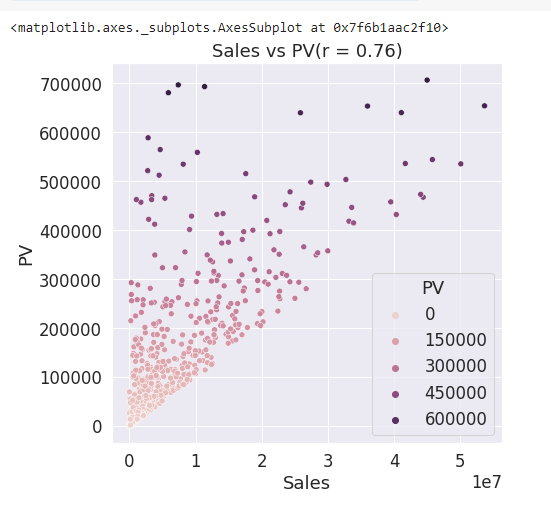


import matplotlib.pyplot as plt

from scipy.stats import pearsonr

plt.title('Sales vs PV(r = {0:0.2f})'.format(pearsonr(drop['Sales'],drop['PV'])[0]))

sns.scatterplot(x="Sales",y="PV",data = df, hue='PV')

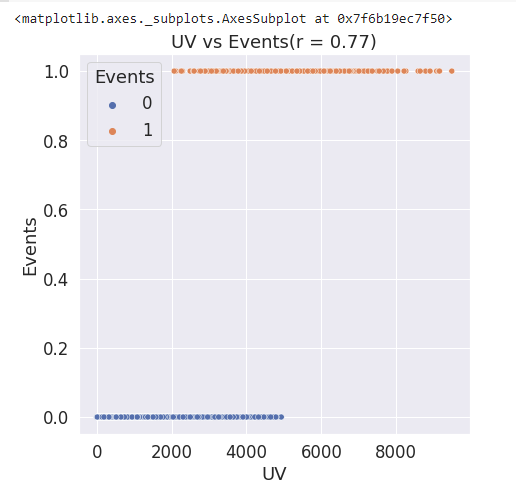


**Observation:**

The Pv of range between 600000 and 700000, the sales is high.

plt.title('UV vs Events(r = {0:0.2f})'.format(pearsonr(drop['UV'],drop['Events'])[0]))

sns.scatterplot(x="UV",y="Events",data = df, hue='Events')



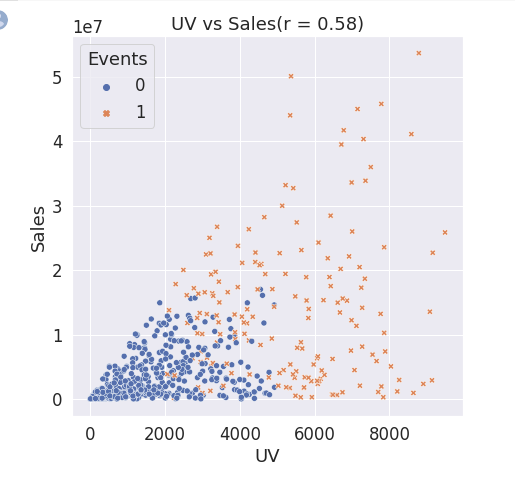
**Observation:**

* The UV range between 0 and 5000 the event 0 occurs.
* The UV range between 2000 and 9000 the event 1 occurs.

plt.title('UV vs Sales(r = {0:0.2f})'.format(pearsonr(drop['UV'],drop['Sales'])[0]))

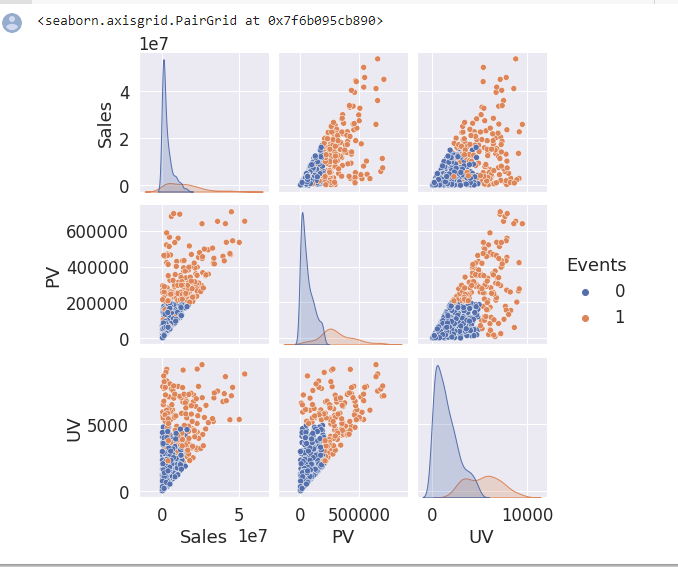
sns.scatterplot(x="UV", y="Sales",hue="Events", style="Events", data=df)

plt.show()



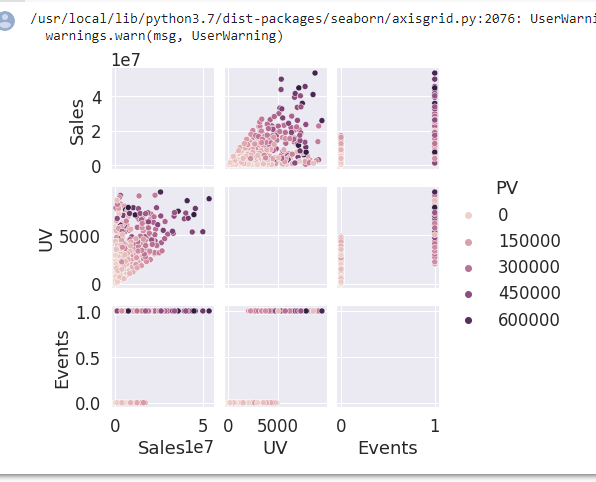
**MULTIVARIATE ANALYSIS**

sns.pairplot(data=df[ ['Sales', 'PV', 'UV', 'Date', 'Events']], hue='Events')



sns.pairplot(df,hue ='PV',size = 2)

plt.show()



**Observation:**

* At sales of 5,the uv is high at 8000 with PV value 600000.
* It shows that the sales and uv are highly correlated.