|  |
| --- |
| **Machine Learning/Data Science** |

**1.Findthe Top ten most viewed movies with their movies Name (Ascending or Descending order)and output the result in a graph as shown below.**

**Python code:**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

#Import Datasets into python

movies\_data=pd.read\_csv("C:\Users\Rajesh-QCC\Desktop\Fragma\movies.csv")

ratings\_data=pd.read\_csv("C:\Users\Rajesh-QCC\Desktop\Fragma\m\_rating.csv")

user\_data=pd.read\_csv("C:\Users\Rajesh-QCC\Desktop\Fragma\users.csv")

#Merging tables by common column

rate\_movie=pd.merge(ratings\_data,movies\_data,on="MovieID")

count\_rate\_movie=pd.value\_counts(rate\_movie['Title'].values,sort=True)

#copy top values in New dataframe

most\_viewed\_movies=count\_rate\_movie[:10,]

#Dataframe

most\_viewed\_movies=pd.DataFrame({'movies':["Star Wars","Star Trek","American Beauty","Mission","Jurassic Par ",”Saving Private Ryan","Terminator 2","Austin Powers","Matrix","Back to the Future"],'COUNT':[11114,5347,3428,2840,2672,2653,2649,2639,2590,2583]})

most\_viewed\_movies.set\_index('MOVIES',inplace=True) #set Movie column as index

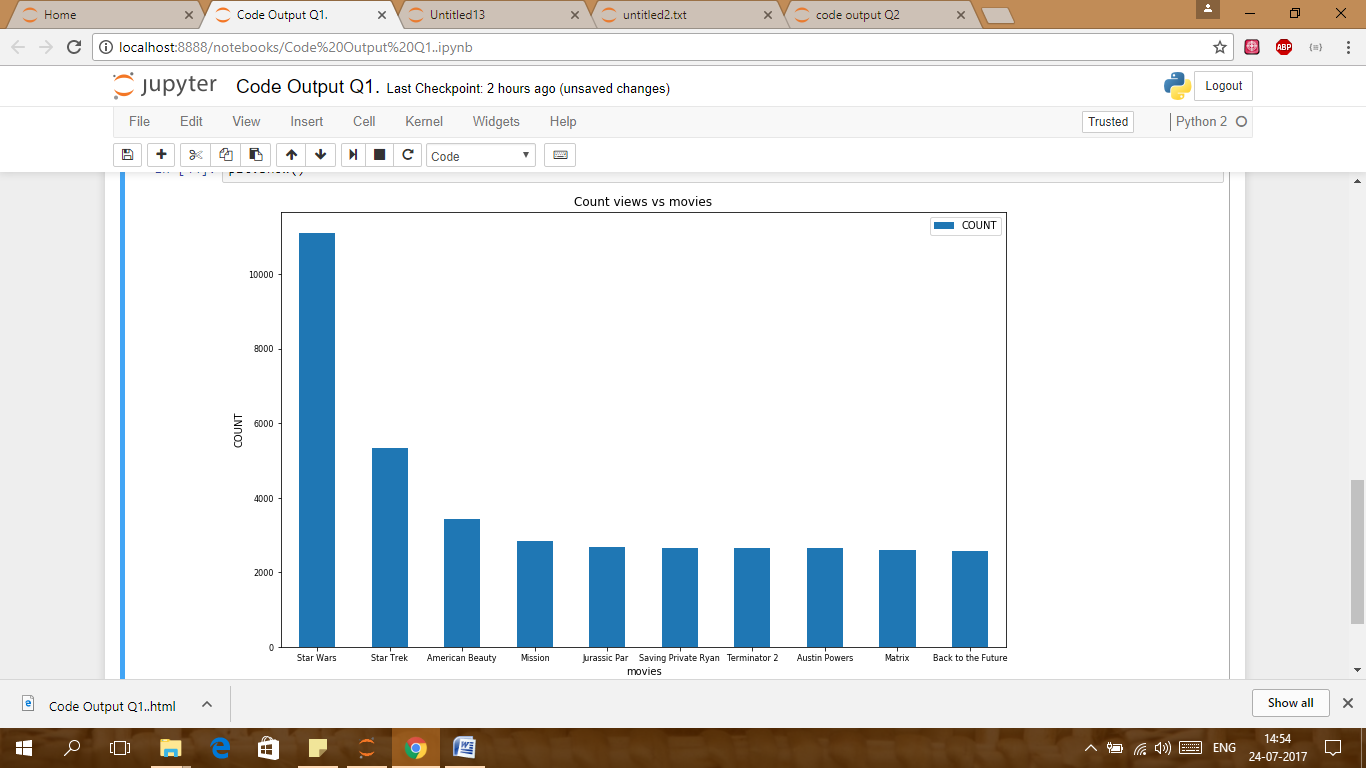
#Bar Graph

bargraph1=most\_viewed\_movies[['COUNT']].plot(kind='bar',title="Count views vs movies",figsize=(15,10),legend=True,fontsize=12)

bargraph1.ylabel('COUNT)

plt.show()

**Output1:**



----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Q2. Write a module to plot a chart for the age wise distribution of the count of views for the movie American beauty**

**Code:**

**#Subset dataset**

am=rate\_movie\_users[rate\_movie\_users['Title']=='American Beauty (1999)']

am1=pd.value\_counts(am['AGE BIN'])

#Converted series into dataframe and make the AGE BIN as index

am1=pd.DataFrame({'AGE BIN':['Under 18','56+','50-55','45-49','35-44','18-24','25-34'],'COUNT':[92,184,248,258,597,715,1334]})

am1.set\_index('AGE BIN',inplace=True)

**#Function For Bar graph**

**def bargraph(x,y,tick\_label,m,n,z):**

**plt.bar(x,y,color='b',width=0.8)**

**tick\_label=plt.xticks(x,tick\_label,rotation='horizontal')**

**m=plt.xlabel(m)**

**n=plt.ylabel(n)**

**z=plt.title(z)**

**plt.show()**

**AGE = [1,2,3,4,5,6,7]**

**COUNT =[92,184,248,258,597,715,1334]**

**tick\_label=('Under 18','56+','50-55','45-49','35-44','18-24','25-34')**

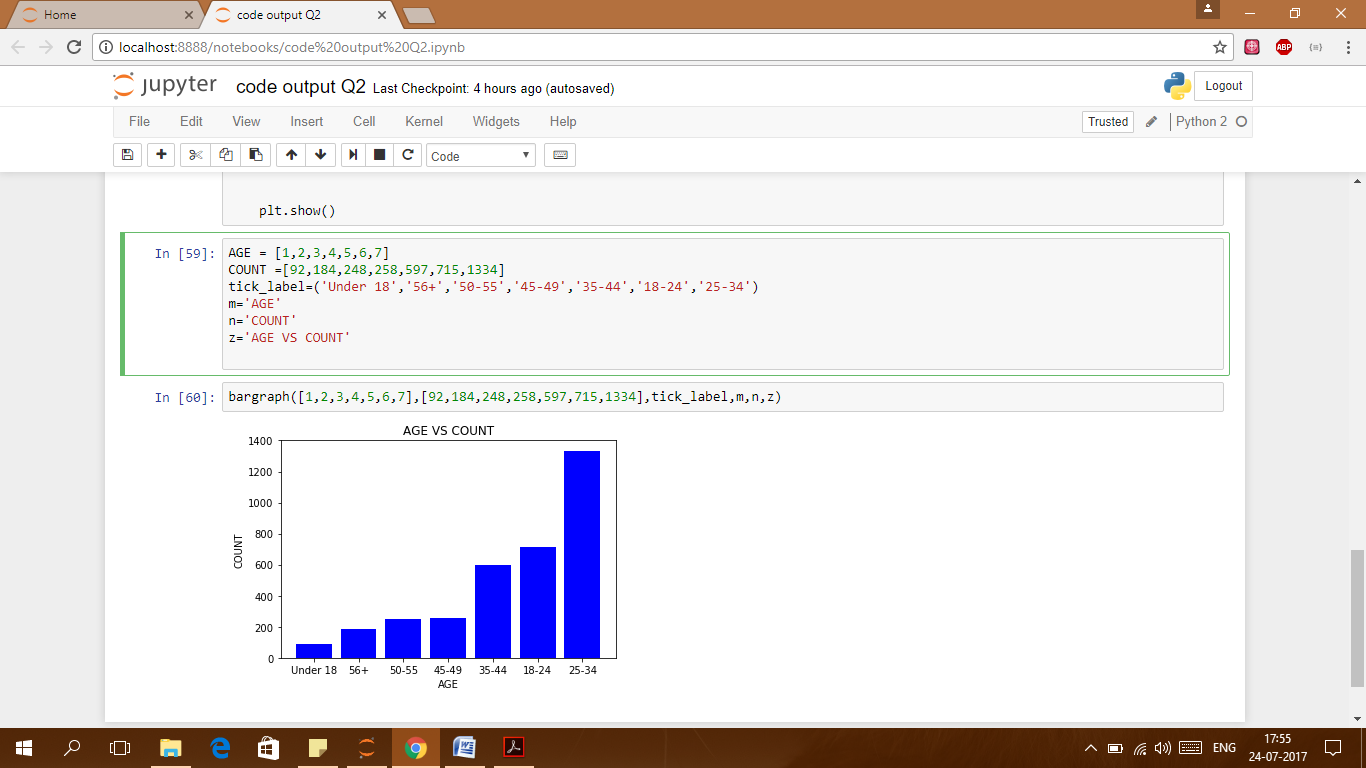
**m='AGE'**

**n='COUNT'**

**z='AGE VS COUNT'**

bargraph([1,2,3,4,5,6,7],[92,184,248,258,597,715,1334],tick\_label,m,n,z)

**Output:**



--------------------------------------------------------------------------------------------------------------------------------

**3.Display how the genres are ranked for each profession. You need to formulate results in following table:**

**Output:**

**#import the datasets**

**movies\_data=pd.read\_csv("C:\Users\Rajesh-QCC\Desktop\Fragma\movies.csv")**

**ratings\_data=pd.read\_csv("C:\Users\Rajesh-QCC\Desktop\Fragma\mm\_rating.csv")**

**user\_data=pd.read\_csv("C:\Users\Rajesh-QCC\Desktop\Fragma\users11.csv")**

**#Merged all three datasets**

**rate\_movie=pd.merge(ratings\_data,movies\_data,on="MovieID")**

**rate\_movie\_users=pd.merge(rate\_movie,user\_data,on="UserID")**

**#Used pivot table to get desired result**

**Rank=pd.pivot\_table(rate\_movie\_users,index=['Occupation','Genres'],**

**values=['Rating1'],aggfunc=np.mean,fill\_value=0)**

**Rank.head(10)**

**#created function to convert Avg rate into Rank**

def add(Rating1):

if Rating1 <=1: return 'Rank1'

elif 1 < Rating1 <=2: return 'Rank2'

elif 2 < Rating1 <=3: return 'Rank3'

elif 3 < Rating1 <=4: return 'Rank4'

elif 4 < Rating1 <=5: return 'Rank5'

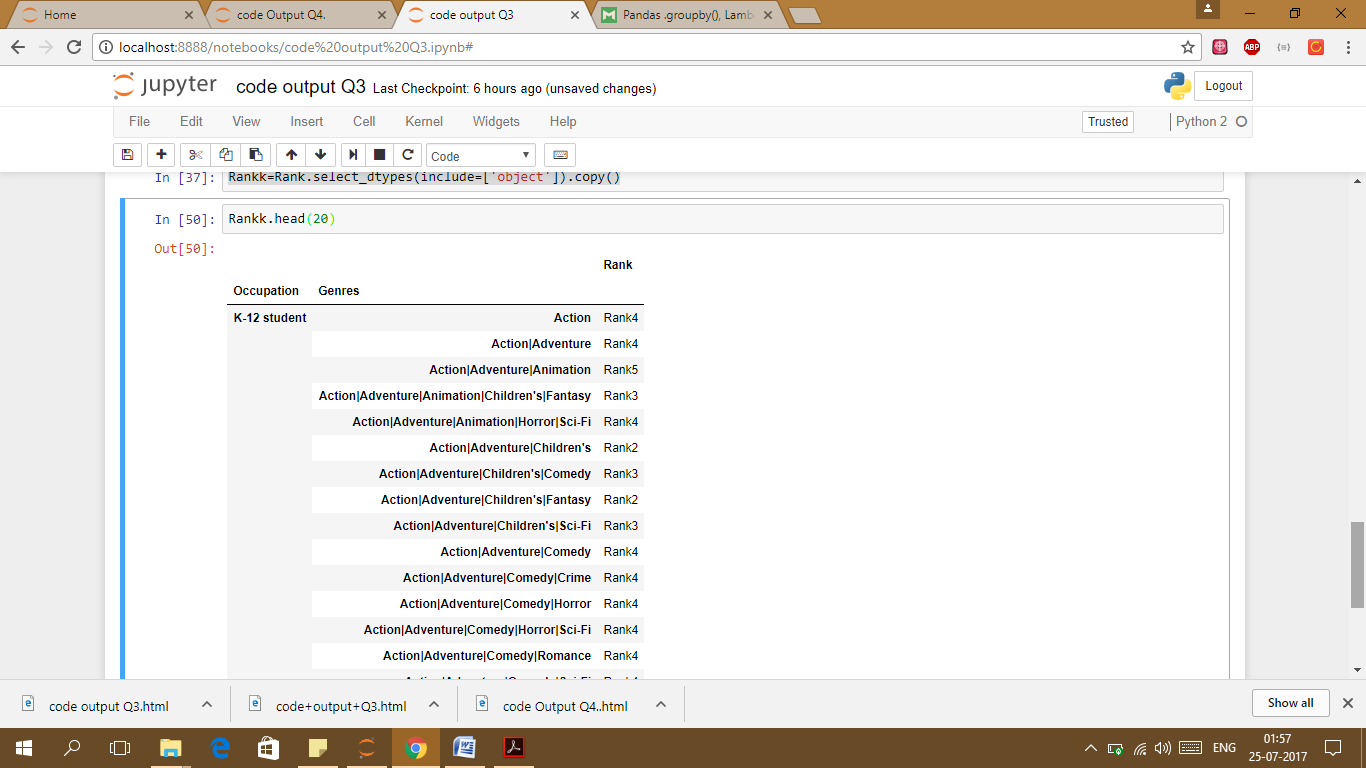
else: return 'None'

Rank["Rank"]=Rank['Rating1'].map(add)

**##only selected object variables**

Rankk=Rank.select\_dtypes(include=['object']).copy()

**Output**



--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**4. And For Programmer occupation show frequency distribution of the top 5 Ranked Genre**

**Code:**

**pr=rate\_movie\_users[rate\_movie\_users['Occupation']==12]**

**pr1=pd.value\_counts(pr['Genres'])**

**pr1=pr1.head(5)**

**pr1**

**Comedy 6309**

**Drama 5308**

**Comedy|Romance 2250**

**Comedy|Drama 2141**

**Action|Thriller 1687**

**Name: Genres, dtype: int64**

#Desired Dataframe

**prog=pd.DataFrame({'Genres':["Comedy","Drama","Comedy|Romance","Comedy|Drama","Action|Thriller"], 'COUNT':[6309,5308,2250,2141,1687]})**

#Bargraph

**prog.set\_index('Genres',inplace=True)**

**PLTT=prog[['COUNT']].plot(kind='bar',title="Count vs Genres)", figsize=(12,7),legend=True,fontsize=12,rot='horizontal')**

**plt.ylabel('COUNT')**

**plt.show()**

**Output:**

