PRN- 202201040210

ROLL NO.- 477

DIVISION- D

Batch- D4

Code-

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
import numpy as np
from sklearn.linear_model import LinearRegression
df=pd.read_csv("D:\\Hotels.csv")
print(df)
print("displaying hotels booked for 3 days")
z1=df[df['days']==3]
print(z1['name'])
print("Displaying hotel name whose price range in between 700 to 1000")
z1=df[(df['total'] >700) & (df['total'] < 1000)]
print(z1['name'])
print("Displaying the number of hotels region wise")
aa=df.groupby('place')
print(aa['place'].value_counts())
ff = df[df['place'] == 'Salvador (BH)']
print(ff)
print("Displaying the average total at hotel 'Salvador'")
total = ff['total'].sum()
count1 = ff['total'].value_counts()
```

```
av=total / count1
average = ff['total'].mean()
print(average)
print("plotting the data")
print("plotting the horizontal bar chat for days VS total ")
days=df['days']
total=df['total']
# Creating the bar graph
plt.bar(total, days)

# Adding labels and title
plt.xlabel('days')
plt.ylabel('total')
plt.title('Bar Graph Example')

# Displaying the graph
plt.show()
# Creating the bar graph
plt.bar(days, total)
# Adding labels and title
plt.xlabel('days')
plt.ylabel('total')
plt.title('Bar Graph Example')

# Displaying the graph
plt.show()
s = df['price']
x=df['place']
print("plotting the  chart average vs price and place ")
# Creating a bar graph
plt.bar(x, s)
```

```
# Calculating the average
average = s.mean()

# Adding the average line
plt.axhline(y=average, color='r', linestyle='--', label='Average')

# Adding labels and title
plt.xlabel('Place')
plt.ylabel('Price')
plt.title('Bar Graph with Average')

# Adding legend
plt.legend()

# Displaying the graph
plt.show()



   # Extract the desired feature for clustering
X = df[['days','total']].values

   # Specify the number of clusters
k = 5

   # Create a KMeans object
kmeans = KMeans(n_clusters=k)

   # Fit the model to the data
kmeans.fit(X)
```

```
    # Predict the cluster labels
labels = kmeans.predict(X)

    # Get the cluster centers
centers = kmeans.cluster_centers_

    # Visualize the clusters
plt.scatter(X[:, 0], np.zeros_like(X[:, 0]), c=labels, cmap='viridis')
plt.scatter(centers[:, 0], np.zeros_like(centers[:, 0]), marker='x', color='red')
plt.title('K-means Clustering of the duration')
plt.xlabel('days vs total knn data')
plt.show()
confirmed_cases = df['days']
deaths = df['total']
plt.scatter(confirmed_cases, deaths)
X = confirmed_cases.values.reshape(-2, 1)
y = deaths.values.reshape(-1, 1)
regressor = LinearRegression()
regressor.fit(X, y)
y_pred = regressor.predict(X)
plt.plot(X, y_pred, color='red', linewidth=3)
plt.xlabel('DAYS')
plt.ylabel('TOTAL')
plt.title('Linear Regression: Confirmed Cases vs Deaths')
plt.show()
```

csv data-

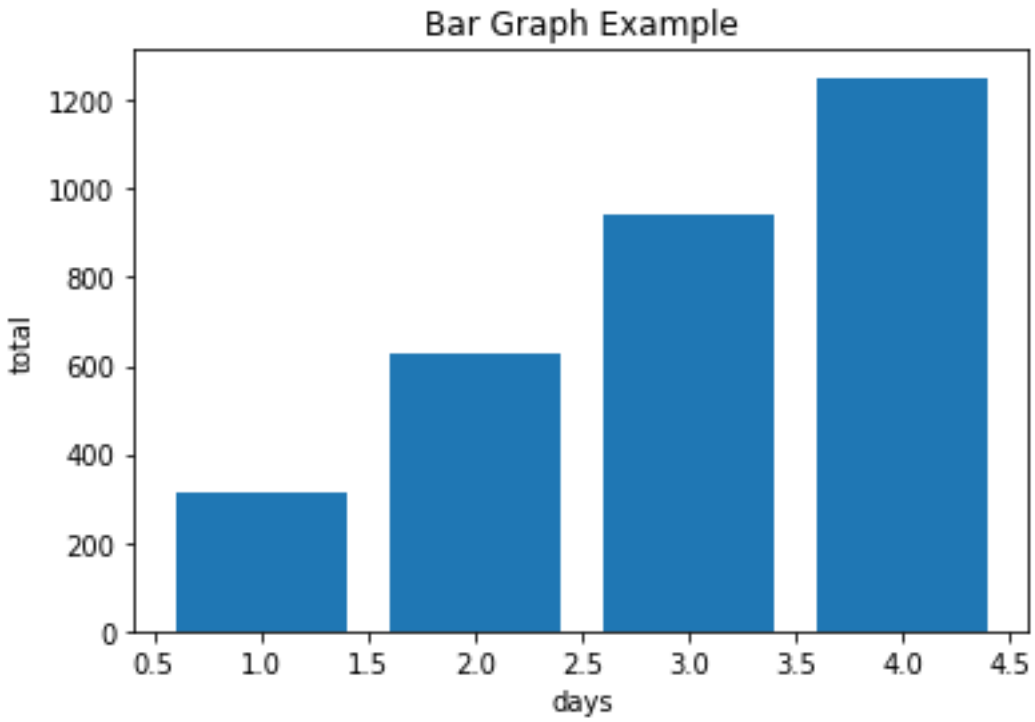
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	travelCod	userCode	name	place	days	price	total	date										
2	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019					1	Display the hotel which are booked for 4 days				
3	2	0	Hotel K	Salvador (BH)	2	263.41	526.82	########						Display hotel name whose price range in between 200				
4	7	0	Hotel K	Salvador (BH)	3	263.41	790.23	11/14/2019					2	Display number of hotels region wise				
5	11	0	Hotel K	Salvador (BH)	4	263.41	1053.64	########						Display the average total at hotel "Salvador"				
6	13	0	Hotel A	Florianopolis (SC)	1	313.02	313.02	12/26/2019					3	Plot the bar chart for days VS total				
7	15	0	Hotel BD	Natal (RN)	2	242.88	485.76	########					4	Plot the suitable graph to display average price Vs Plac				
8	22	0	Hotel Z	Aracaju (SE)	2	208.04	416.08	2/27/2020					5	Apply Linear regrssion on days and total				
9	29	0	Hotel AU	Recife (PE)	4	312.83	1251.32	4/16/2020										
10	32	0	Hotel AF	Sao Paulo (SP)	2	139.1	278.2	########										
11	33	0	Hotel K	Salvador (BH)	4	263.41	1053.64	5/14/2020										
12	34	0	Hotel AF	Sao Paulo (SP)	3	139.1	417.3	5/21/2020										
13	38	0	Hotel BD	Natal (RN)	2	242.88	485.76	6/18/2020										
14	39	0	Hotel K	Salvador (BH)	1	263.41	263.41	6/25/2020										
15	42	0	Hotel BW	Campo Grande (MS)	3	60.39	181.17	7/16/2020										
16	43	0	Hotel K	Salvador (BH)	4	263.41	1053.64	7/23/2020										
17	45	0	Hotel BD	Natal (RN)	1	242.88	242.88	########										
18	51	0	Hotel K	Salvador (BH)	1	263.41	263.41	9/17/2020										
19	53	0	Hotel BW	Campo Grande (MS)	2	60.39	120.78	########										
20	54	0	Hotel AF	Sao Paulo (SP)	3	139.1	417.3	########										

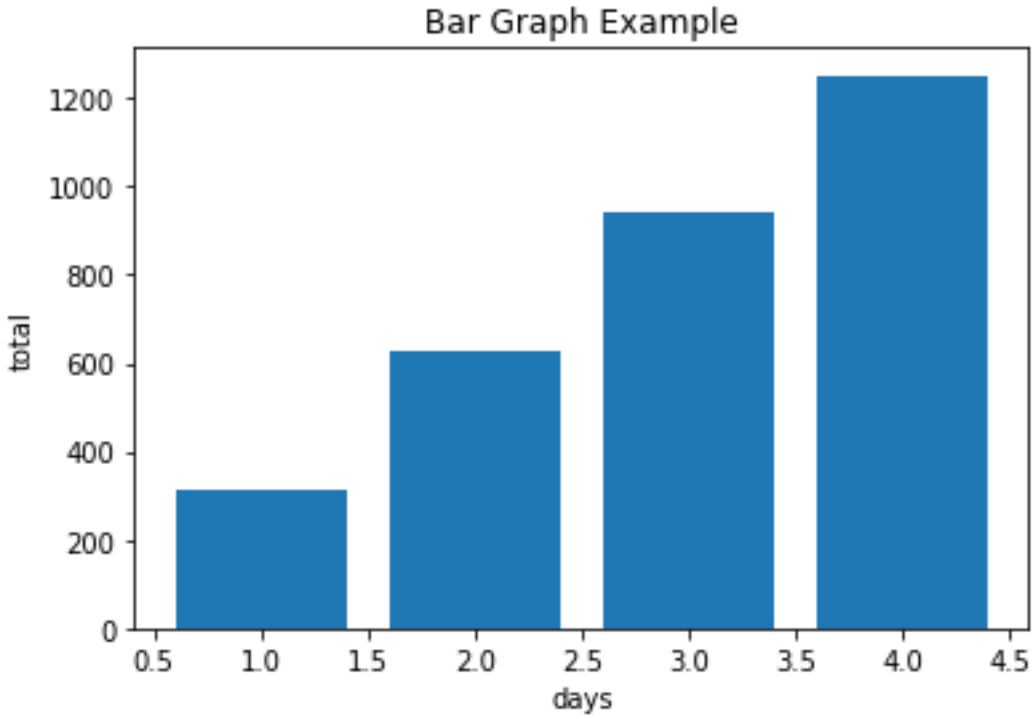
Dataset

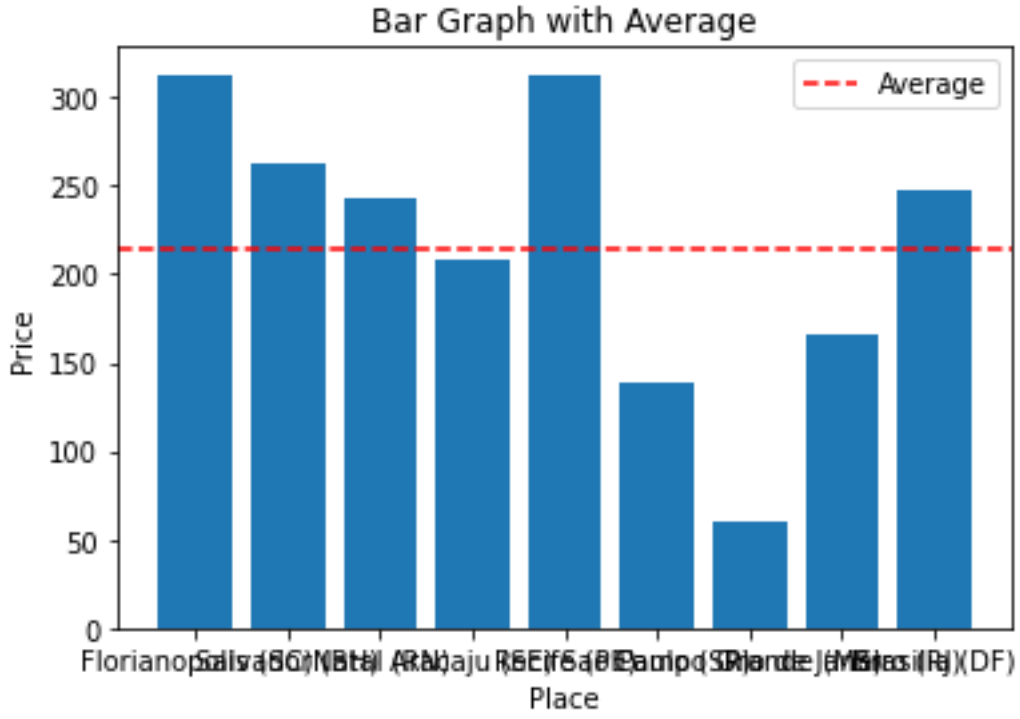
Output screenshot-

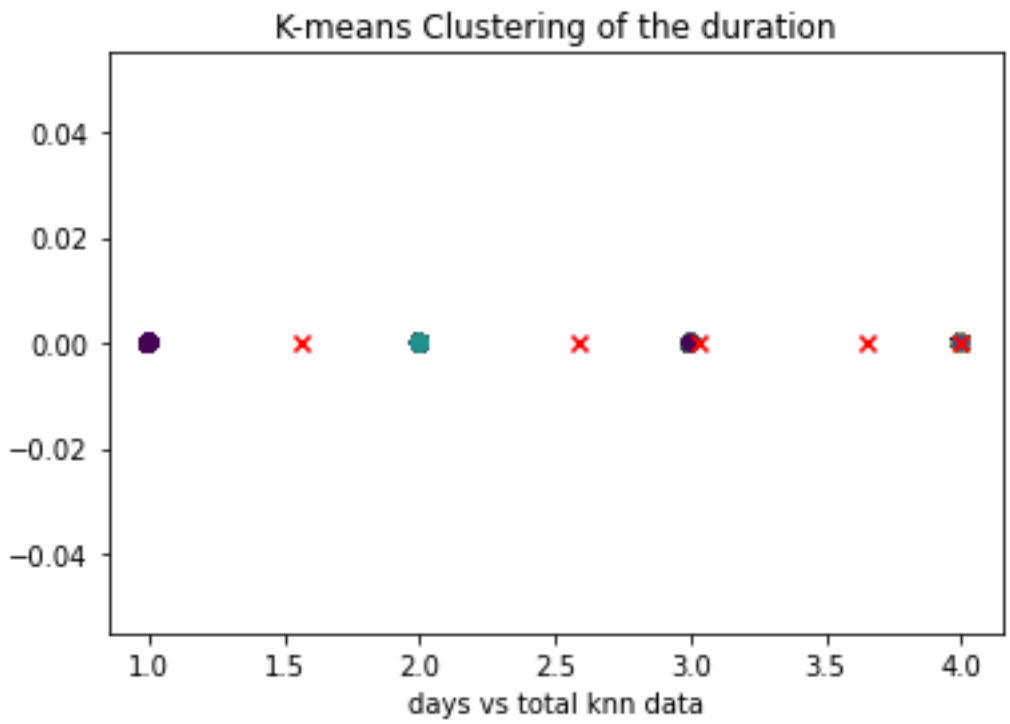
```
In [25]: runfile('C:/Users/Nirmal chaturvedi/Desktop/end sem.py', wdir='C:/Users/Nirmal chaturvedi/Desktop')
       travelCode  ...                                         Unnamed: 13
0               0  ...      Display the hotel which are booked for 4 days
1               2  ...  Display hotel name whose price range in betwee...
2               7  ...                Display number of hotels region wise
3              11  ...      Display the average total at hotel "Salvador"
4              13  ...                Plot the bar chart for days VS total
...           ...  ...                                                 ...
40547      135938  ...                                                 NaN
40548      135939  ...                                                 NaN
40549      135940  ...                                                 NaN
40550      135941  ...                                                 NaN
40551      135942  ...                                                 NaN

[40552 rows x 14 columns]
displaying hotels booked for 3 days
2         Hotel K
10       Hotel AF
13       Hotel BW
18       Hotel AF
25       Hotel AF
           ...
40537    Hotel BD
40546     Hotel Z
40547    Hotel BP
40549    Hotel BW
40550    Hotel BW
Name: name, Length: 10108, dtype: object
Displaying hotel name whose price range in between 700 to 1000
2         Hotel K
20        Hotel Z
21        Hotel Z
```

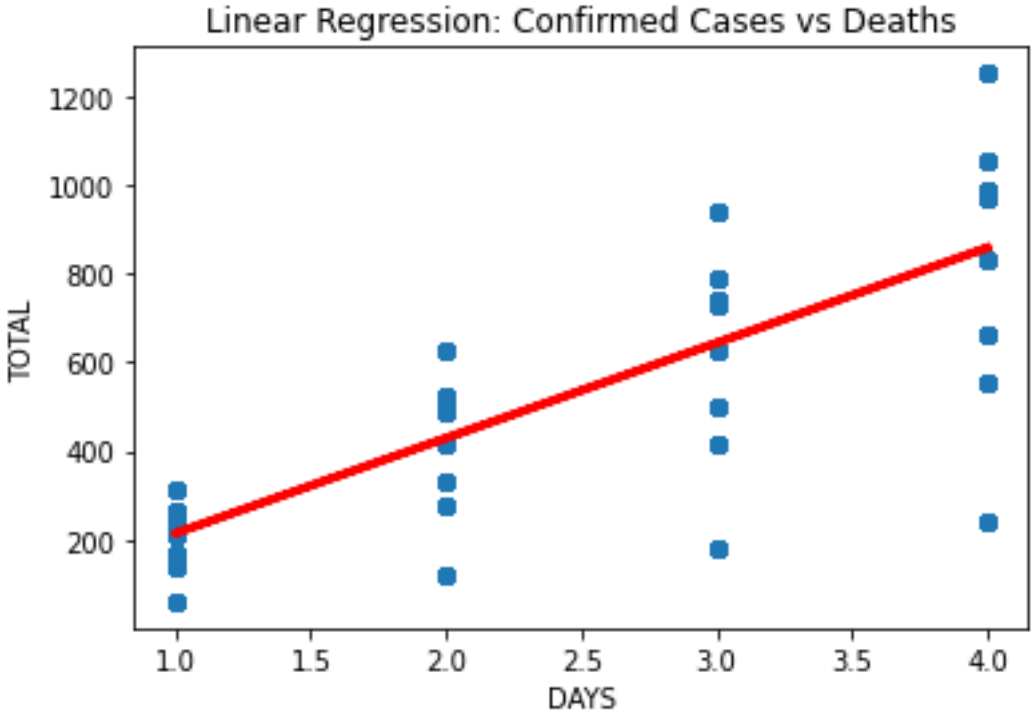
```
40538    Hotel BP
40547    Hotel BP
40551    Hotel BD
Name: name, Length: 8854, dtype: object
Displaying the number of hotels region wise
place
Aracaju (SE)           4205
Brasilia (DF)          4437
Campo Grande (MS)      4333
Florianopolis (SC)     3330
Natal (RN)             4829
Recife (PE)            4467
Rio de Janeiro (RJ)    5029
Salvador (BH)          5094
Sao Paulo (SP)         4828
Name: count, dtype: int64
       travelCode  ...                                     Unnamed: 13
1               2  ...  Display hotel name whose price range in betwee...
2               7  ...                Display number of hotels region wise
3              11  ...      Display the average total at hotel "Salvador"
9              33  ...                                                NaN
12             39  ...                                                NaN
...           ...  ...                                                ...
40518      135840  ...                                                NaN
40529      135866  ...                                                NaN
40530      135867  ...                                                NaN
40539      135903  ...                                                NaN
40544      135926  ...                                                NaN

[5094 rows x 14 columns]
Displaying the average total at hotel 'Salvador'
536.2295129710002
plotting the data
plotting the horizontal bar chat for days VS total
```




Bar Graph Example
total
days
0
200
400
600
800
1000
1200
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5


Bar Graph with Average
Average
Price
Place
0
50
100
150
200
250
300


K-means Clustering of the duration
0.04
0.02
0.00
−0.02
−0.04
1.0
1.5
2.0
2.5
3.0
3.5
4.0
days vs total knn data


Linear Regression: Confirmed Cases vs Deaths
TOTAL
1200
1000
800
600
400
200
1.0
1.5
2.0
2.5
3.0
3.5
4.0
DAYS