Assignment on Store_sales dataset

Load the Dataset

In [1]:
```
import pandas as pd
df=pd.read_csv('store_sales.csv')
df
```

Out[1]:

	store_id	city	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec
0	S_1	Texas	8	20	13	21	17	20	24	17	16	9	7	6
1	S_2	California	12	19	15	15	11	19	7	15	10	11	21	19
2	S_3	California	16	16	14	19	23	6	13	13	15	14	24	8
3	S_4	Texas	8	18	13	10	14	14	6	8	8	18	7	11
4	S_5	Texas	19	5	24	9	5	24	10	5	24	15	6	13
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
95	S_96	Texas	7	10	20	20	10	15	15	21	15	7	23	22
96	S_97	California	13	6	7	15	22	10	21	23	10	6	12	9
97	S_98	Texas	16	9	6	14	20	13	11	10	8	22	17	22
98	S_99	Arizona	18	16	9	5	12	22	11	13	21	17	19	10
99	S_100	California	5	23	17	24	15	21	19	10	12	20	5	9

100 rows × 14 columns

Calculate the total sales for each store across all months.

In [4]:
```
import pandas as pd

df = pd.read_csv('store_sales.csv')

total_sales_per_store = df.drop(['store_id', 'city'], axis=1).sum(axis=1)

df['total_sales'] = total_sales_per_store

df[['store_id', 'city', 'total_sales']]
```

Out[4]:

	store_id	city	total_sales
0	S_1	Texas	178
1	S_2	California	174
2	S_3	California	181
3	S_4	Texas	135
4	S_5	Texas	159
...	...	...	...
95	S_96	Texas	185
96	S_97	California	154
97	S_98	Texas	168
98	S_99	Arizona	173
99	S_100	California	180

100 rows × 3 columns

Find the average sales for each month across all stores

In [5]:
```
import pandas as pd

df = pd.read_csv('store_sales.csv')


average_sales_per_month = df.drop(['store_id', 'city'], axis=1).mean()


print(average_sales_per_month)
```

```
Jan     14.46
Feb     15.09
Mar     14.56
Apr     14.57
May     13.56
Jun     13.80
July    14.38
Aug     15.81
Sep     14.91
Oct     14.10
Nov     15.31
Dec     13.57
dtype: float64
```

Identify the store with the highest total sales.

In [6]:
```
import pandas as pd


df = pd.read_csv('store_sales.csv')


total_sales_per_store = df.drop(['store_id', 'city'], axis=1).sum(axis=1)


df['total_sales'] = total_sales_per_store


df[['store_id', 'city', 'total_sales']]

store_with_highest_sales = df.loc[df['total_sales'].idxmax()]


print(store_with_highest_sales[['store_id', 'city', 'total_sales']])
```

```
store_id             S_62
city           California
total_sales           214
Name: 61, dtype: object
```

Calculate the total sales for each city.

In [8]:
```
import pandas as pd

df = pd.read_csv('store_sales.csv')

total_sales_per_city = df.drop(['store_id'], axis=1).groupby('city').sum().s


print(total_sales_per_city)
```

```
city
Arizona       3951
California    7522
Texas         5939
dtype: int64
```

List stores with total sales greater than 200

In [9]:
```
import pandas as pd


df = pd.read_csv('store_sales.csv')

total_sales_per_store = df.drop(['store_id', 'city'], axis=1).sum(axis=1)

df['total_sales'] = total_sales_per_store


stores_above_200 = df[df['total_sales'] > 200]


print(stores_above_200[['store_id', 'city', 'total_sales']])
```

```
   store_id        city  total_sales
37     S_38  California          207
39     S_40       Texas          204
50     S_51     Arizona          210
61     S_62  California          214
69     S_70       Texas          206
72     S_73  California          213
77     S_78     Arizona          211
85     S_86  California          214
86     S_87       Texas          203
92     S_93  California          204
```

Which month had the highest average sales across all stores?

In [11]:
```
import pandas as pd

df = pd.read_csv('store_sales.csv')


monthly_sales = df.drop(['store_id', 'city'], axis=1).sum()

month_with_highest_sales = monthly_sales.idxmax()
highest_sales_value = monthly_sales.max()

print(f"The month with the highest sales is {month_with_highest_sales} with
```

```
The month with the highest sales is Aug with total sales of 1581.
```

Which city generated the most revenue?

In [12]:
```
import pandas as pd


df = pd.read_csv('store_sales.csv')

total_sales_per_city = df.drop(['store_id'], axis=1).groupby('city').sum().s


city_with_highest_sales = total_sales_per_city.idxmax()
highest_sales_value = total_sales_per_city.max()

print(f"The city that generated the most revenue is {city_with_highest_sales
```

```
The city that generated the most revenue is California with total sales of
7522.
```

In [ ]:
```

```