Assignment on Store\_sales dataset

# Load the Dataset

In [2]: **import** pandas **as** pd

ss**=**pd**.**read\_csv('store\_sales.csv')

# Calculate the total sales for each store across all months.

In [7]: ss['Total\_Sales']**=**ss**.**iloc[:,2:]**.**sum(axis**=**1) print(ss[['store\_id','Total\_Sales']])

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | store\_id | | Total\_Sales | |
| 0 | S\_1 | | 178 | |
| 1 | S\_2 | | 174 | |
| 2 | S\_3 | | 181 | |
| 3 | S\_4 | | 135 | |
| 4 | S\_5 | | 159 | |
| .. | ... | | ... | |
| 95 | S\_96 | | 185 | |
| 96 | S\_97 | | 154 | |
| 97 | S\_98 | | 168 | |
| 98 | S\_99 | | 173 | |
| 99 | S\_100 | | 180 | |
| [100 | | rows x | 2 | columns] |

In [9]: ss

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Out[9]: | **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 × 15 columns

# Find the average sales for each month across all stores

In [18]: average\_sales\_per\_month **=** ss**.**iloc[:, 2:14]**.**mean() # Columns 2 to 13 for 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 [40]: highest\_total\_sales**=**ss**.**loc[ss['Total\_Sales']**.**idxmax()] print(highest\_total\_sales[['store\_id','Total\_Sales']])

store\_id S\_62

Total\_Sales 214

Name: 61, dtype: object

# Calculate the total sales for each city.

In [46]: city\_sales**=** ss**.**groupby('city')['Total\_Sales']**.**sum() city\_sales

Out[46]: city

Arizona 3951

California 7522

Texas 5939

Name: Total\_Sales, dtype: int64

# List stores with total sales greater than 200

In [54]: greater\_than\_200**=**ss[ss['Total\_Sales']**>**200] greater\_than\_200[['store\_id','Total\_Sales']]

|  |  |  |  |
| --- | --- | --- | --- |
| Out[54]: | **37** | **store\_id**  S\_38 | **Total\_Sales**  207 |
|  | **39** | S\_40 | 204 |
|  | **50** | S\_51 | 210 |
|  | **61** | S\_62 | 214 |
|  | **69** | S\_70 | 206 |
|  | **72** | S\_73 | 213 |
|  | **77** | S\_78 | 211 |
|  | **85** | S\_86 | 214 |
|  | **86** | S\_87 | 203 |
|  | **92** | S\_93 | 204 |

# Which month had the highest average sales across all stores?

In [68]: average\_sales\_per\_month**.**idxmax() Out[68]: 'Aug'

In [66]: average\_sales\_per\_month**.**max()

Out[66]: 15.81

# Which city generated the most revenue?

In [72]: city\_sales**.**idxmax() Out[72]: 'California'

In [74]: city\_sales**.**max()

Out[74]: 7522