EDS Theory Activity 1

- **NAME : Bhoomi Bharat Tiple**
- **ROLL NO : CS5-88**
- **DIVISION :CS5**
- **PRN :202401100027**

Dataset Name : Sales Dataset

Problem 1: Load the dataset and display the first 5 rows.

```
import pandas as pd
```

```
path="/content/drive/MyDrive/Product Sales Data.csv"
df = pd.read_csv(path)
print(df.head(5))
```

```
  Month  Cream  Detergent  Moisturizer  Sanitizer  Shampoo  Soap  Total Units  \
0   Jan   5100       9600         1525       1550     1200  2510        21485
1   Feb   5000       6500         1225       1250     2100  2640        18715
2   Mar   4450       9950         1365       1390     3550  2150        22855
3   Apr   5770       9270         1155       1180     1870  3410        22655
4   May   4460       8160         1765       1790     1560  3610        21345

   Total Profit
0        211500
1        183800
2        225200
3        223200
4        210100
```

Problem 2: Calculate the total profit for the entire year.

```
import pandas as pd
path="/content/drive/MyDrive/Product Sales Data.csv"
df = pd.read_csv(path)
total_profit = df['Total Profit'].sum()
print("Total Profit for the Year:", total_profit)
```

```
Total Profit for the Year: 3129300
```

Problem 3: Find the month with the highest total units sold.

```
import pandas as pd
path="/content/drive/MyDrive/Product Sales Data.csv"
df = pd.read_csv(path)
max_units_month = df.loc[df['Total Units'].idxmax(), 'Month']
print("Month with Highest Units Sold:", max_units_month)
```

```
Month with Highest Units Sold: Dec
```

Problem 4: Calculate the average units sold for each product.

```
import pandas as pd
path="/content/drive/MyDrive/Product Sales Data.csv"
df = pd.read_csv(path)
products = ['Cream', 'Detergent', 'Moisturizer', 'Sanitizer', 'Shampoo', 'Soap']
avg_units = df[products].mean()
print("Average Units Sold per Product:\n", avg_units)
```

```
Average Units Sold per Product:
 Cream          5725.833333
Detergent      9900.833333
Moisturizer    1567.916667
Sanitizer      1592.916667
Shampoo        2117.500000
Soap           2883.333333
dtype: float64
```

Problem 5: Determine the product with the highest total sales for the year.

```
import pandas as pd
path="/content/drive/MyDrive/Product Sales Data.csv"
df = pd.read_csv(path)
products = ['Cream', 'Detergent', 'Moisturizer', 'Sanitizer', 'Shampoo', 'Soap']
total_sales = df[products].sum()
highest_selling_product = total_sales.idxmax()
print("Highest Selling Product:", highest_selling_product)
```

```
Highest Selling Product: Detergent
```

Problem 6: Find the month with the lowest profit.

```
import pandas as pd
path="/content/drive/MyDrive/Product Sales Data.csv"
df = pd.read_csv(path)
min_profit_month = df.loc[df['Total Profit'].idxmin(), 'Month']
print("Month with Lowest Profit:", min_profit_month)
```

```
Month with Lowest Profit: Feb
```

Problem 7: Calculate the standard deviation of units sold for each product.

```
import pandas as pd
path="/content/drive/MyDrive/Product Sales Data.csv"
df = pd.read_csv(path)
products = ['Cream', 'Detergent', 'Moisturizer', 'Sanitizer', 'Shampoo', 'Soap']
std_dev = df[products].std()
print("Standard Deviation of Units Sold:\n", std_dev)
```

```
Standard Deviation of Units Sold:
 Cream          1242.032486
Detergent      2348.095779
Moisturizer     316.733745
Sanitizer       316.733745
Shampoo         617.724931
Soap            584.595172
dtype: float64
```

Problem 8: Identify the product with the most consistent sales (lowest standard deviation).

```
import pandas as pd
path="/content/drive/MyDrive/Product Sales Data.csv"
df = pd.read_csv(path)
products = ['Cream', 'Detergent', 'Moisturizer', 'Sanitizer', 'Shampoo', 'Soap']
most_consistent_product = std_dev.idxmin()
print("Most Consistent Product:", most_consistent_product)
```

```
Most Consistent Product: Moisturizer
```

Problem 9: Compute the correlation between total units sold and total profit.

```
import pandas as pd
path="/content/drive/MyDrive/Product Sales Data.csv"
df = pd.read_csv(path)
correlation = df['Total Units'].corr(df['Total Profit'])
print("Correlation between Units Sold and Profit:", correlation)
```

```
Correlation between Units Sold and Profit: 0.7680890631902555
```

Problem 10: Find the month where shampoo sales were the highest.

```
import pandas as pd
path="/content/drive/MyDrive/Product Sales Data.csv"
df = pd.read_csv(path)
products = ['Cream', 'Detergent', 'Moisturizer', 'Sanitizer', 'Shampoo', 'Soap']
max_shampoo_month = df.loc[df['Shampoo'].idxmax(), 'Month']
print("Month with Highest Shampoo Sales:", max_shampoo_month)
```

```
Month with Highest Shampoo Sales: Mar
```

Problem 11: Calculate the percentage contribution of each product to total units sold.

```
import pandas as pd
path="/content/drive/MyDrive/Product Sales Data.csv"
df = pd.read_csv(path)
total_units = df['Total Units'].sum()
percentage_contribution = (df[products].sum() / total_units) * 100
print("Percentage Contribution:\n", percentage_contribution)
```

```
Percentage Contribution:
 Cream          24.069922
Detergent      41.620542
Moisturizer     6.591116
Sanitizer       6.696210
Shampoo         8.901422
Soap           12.120788
dtype: float64
```

Problem 12: Find the month with the highest profit-to-units ratio.

```
import pandas as pd
path="/content/drive/MyDrive/Product Sales Data.csv"
df = pd.read_csv(path)
df['Profit-to-Units Ratio'] = df['Total Profit'] / df['Total Units']
max_ratio_month = df.loc[df['Profit-to-Units Ratio'].idxmax(), 'Month']
print("Month with Highest Profit-to-Units Ratio:", max_ratio_month)
```

```
Month with Highest Profit-to-Units Ratio: Aug
```

Problem 13: Calculate the median units sold for each product.

```
import pandas as pd
path="/content/drive/MyDrive/Product Sales Data.csv"
df = pd.read_csv(path)
median_units = df[products].median()
print("Median Units Sold:\n", median_units)
```

```
Median Units Sold:
 Cream          5430.0
Detergent      9490.0
Moisturizer    1552.5
Sanitizer      1577.5
Shampoo        1995.0
Soap           2840.0
dtype: float64
```

Problem 14: Identify the product with the highest sales in December.

```
import pandas as pd
path="/content/drive/MyDrive/Product Sales Data.csv"
df = pd.read_csv(path)
dec_sales = df[df['Month'] == 'Dec'][products].iloc[0]
highest_dec_product = dec_sales.idxmax()
print("Highest Selling Product in December:", highest_dec_product)
```

```
Highest Selling Product in December: Detergent
```

Problem 15: Compute the total profit for the first quarter (Jan-Mar).

```
import pandas as pd
path="/content/drive/MyDrive/Product Sales Data.csv"
df = pd.read_csv(path)
first_quarter_profit = df[df['Month'].isin(['Jan', 'Feb', 'Mar'])]['Total Profit'].sum()
print("First Quarter Profit:", first_quarter_profit)
```

```
First Quarter Profit: 620500
```

Problem 16: Find the month where detergent sales were above 10,000 units.

```
import pandas as pd
path="/content/drive/MyDrive/Product Sales Data.csv"
df = pd.read_csv(path)
high_detergent_months = df[df['Detergent'] > 10000]['Month']
print("Months with Detergent Sales > 10,000:", high_detergent_months.tolist())
```

```
Months with Detergent Sales > 10,000: ['Aug', 'Oct', 'Nov', 'Dec']
```

Problem 17: Calculate the cumulative profit over the months.

```
import pandas as pd
path="/content/drive/MyDrive/Product Sales Data.csv"
df = pd.read_csv(path)
df['Cumulative Profit'] = df['Total Profit'].cumsum()
print("Cumulative Profit:\n", df[['Month', 'Cumulative Profit']])
```

```
Cumulative Profit:
    Month  Cumulative Profit
0     Jan             211500
1     Feb             395300
2     Mar             620500
3     Apr             843700
4     May            1053800
5     Jun            1255700
6     Jul            1551700
7     Aug            1913600
8     Sep            2148100
9     Oct            2415300
10    Nov            2828600
11    Dec            3129300
```

Problem 18: Determine the average profit per unit sold for each month.

```
import pandas as pd
path="/content/drive/MyDrive/Product Sales Data.csv"
df = pd.read_csv(path)
df['Profit per Unit'] = df['Total Profit'] / df['Total Units']
print("Average Profit per Unit:\n", df[['Month', 'Profit per Unit']])
```

```
Average Profit per Unit:
    Month  Profit per Unit
0     Jan         9.844077
1     Feb         9.820999
2     Mar         9.853424
3     Apr         9.852130
4     May         9.843055
5     Jun         9.836784
6     Jul        13.998581
7     Aug        14.156073
8     Sep         9.859155
9     Oct         9.876178
10    Nov        13.811195
11    Dec         9.889821
```

Problem 19: Find the product with the highest sales variance across months.

```
import pandas as pd
path="/content/drive/MyDrive/Product Sales Data.csv"
df = pd.read_csv(path)
sales_variance = df[products].var()
highest_variance_product = sales_variance.idxmax()
print("Product with Highest Sales Variance:", highest_variance_product)
```

```
Product with Highest Sales Variance: Detergent
```

Problem 20: Identify the month with the highest sales for the most profitable product.

```
import pandas as pd
path="/content/drive/MyDrive/Product Sales Data.csv"
df = pd.read_csv(path)
most_profitable_product = df[products].sum().idxmax()
highest_sales_month = df.loc[df[most_profitable_product].idxmax(), 'Month']
print(f"Month with Highest Sales for {most_profitable_product}:", highest_sales_month)
```

```
Month with Highest Sales for Detergent: Dec
```