ESSENTIAL OF DATA SCIENCE

Theory Activity No. 1

Name – Sarvesh Barale

Div – CS8

Roll No. – CS8-44

PRN – 202401120071

- 20 problem statements for Kaggel Text Classification Dataset using Numpy and Pandas.
- Kaggle Link - https://www.kaggle.com/datasets/vinothkannaece/sales-dataset

10 Problem Statements Using NumPy:

1. **Calculate the total sales amount.**
2. **Find the average number of units sold.**
3. **Determine the standard deviation of the sales amount.**
4. **Identify the maximum unit price.**
5. **Identify the minimum unit cost.**
6. **Count how many transactions had a discount greater than 10%.**
7. **Calculate the median of unit prices.**
8. **Compute total revenue as unit price multiplied by quantity**

sold.
9. Count the number of sales transactions above the average sales amount.
10. Calculate the total profit using NumPy operations.

- **Solution:-**

```
total_sales = np.sum(sales_amount_np)
average_quantity = np.mean(quantity_np)
std_sales_amount = np.std(sales_amount_np)
max_unit_price = np.max(unit_price_np)
min_unit_cost = np.min(unit_cost_np)
high_discount_count = np.sum(discount_np > 0.1)
median_unit_price = np.median(unit_price_np)
total_revenue = np.sum(unit_price_np * quantity_np)
sales_above_avg = np.sum(sales_amount_np >
np.mean(sales_amount_np))
total_profit_numpy = np.sum((unit_price_np - unit_cost_np) *
quantity_np)
```

```
Total sales amount: 5019265.2299999995
Average quantity sold: 25.355
Standard deviation of sales amount: 2845.3663745785966
Maximum unit price: 5442.15
Minimum unit cost: 60.28
Number of transactions with >10% discount: 650
Median unit price: 2696.4
Total revenue: 70329940.71
Number of sales above average: 500
Total profit using NumPy: 6487847.07
```

10 Problem Statements Using Pandas:

1.Find the total sales amount using Pandas.
2. Calculate the average unit cost for each product category.
3. Find the total quantity sold by region.
4. Count the number of transactions by each payment method.
5. Identify the top 3 sales representatives by revenue.
6. Compute the average discount given per sales channel.
7. Find the correlation between unit price and unit cost.
8. Determine the highest single sales transaction.
9. Compare the average sales amount for new vs returning customers.
10. Calculate the total profit (you must first create the Profit column)

- **Solution:-**

```
total_sales_pandas = df['Sales_Amount'].sum()
avg_unit_cost_by_category = df.groupby('Product_Category')['Unit_Cost'].mean()
quantity_sold_by_region = df.groupby('Region')['Quantity_Sold'].sum()
transactions_by_payment_method = df['Payment_Method'].value_counts()
top_3_reps_by_revenue = df.groupby('Sales_Rep')['Sales_Amount'].sum().nlargest(3)
avg_discount_by_channel = df.groupby('Sales_Channel')['Discount'].mean()
price_cost_correlation = df['Unit_Price'].corr(df['Unit_Cost'])
max_sale_transaction = df['Sales_Amount'].max()
avg_sales_by_customer_type = df.groupby('Customer_Type')['Sales_Amount'].mean()
```

total_profit_pandas = df['Profit'].sum()

```
Total Sales Amount: $5,019,265.23

Average Unit Cost by Product Category:
Product_Category
Clothing       2470.587313
Electronics    2545.330081
Food           2407.998938
Furniture      2472.416115
Name: Unit_Cost, dtype: float64

Quantity Sold by Region:
Region
East     6356
North    6705
South    5808
West     6486
Name: Quantity_Sold, dtype: int64

Transactions by Payment Method:
Payment_Method
Credit Card      345
Bank Transfer    342
Cash             313
Name: count, dtype: int64

Top 3 Sales Representatives by Revenue:
Sales_Rep
David    1141737.36
Bob      1080990.63
Eve       970183.99
Name: Sales_Amount, dtype: float64

Average Discount by Sales Channel:
Sales_Channel
Online    0.152561
Retail    0.152227
Name: Discount, dtype: float64

Price-Cost Correlation: 1.00
```

S

```
Price-Cost Correlation: 1.00

Maximum Sales Transaction: $9,989.04

Average Sales by Customer Type:
Customer_Type
New          4972.734722
Returning    5066.546230
Name: Sales_Amount, dtype: float64

Total Profit: $-58,822,828.41
PS C:\Users\Admin\.vscode\extensions\ms-vscode.cpptools-1.22.11-win32-x64\ui\New folder (2)>
```