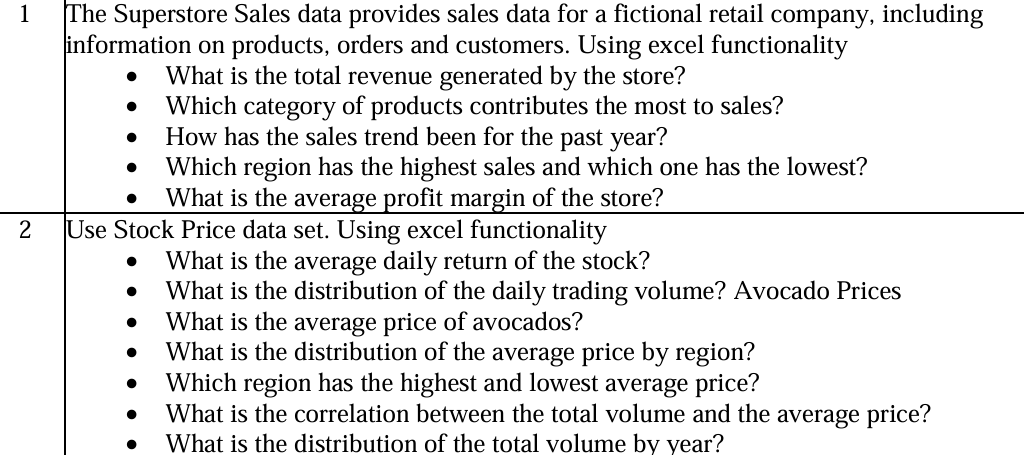
****

**The Superstore Sales data provides sales data for a fictional retail company, including information on products, orders and customers.**

• What is the total revenue generated by the store?

-> =sum(sales range)

• Which category of products contributes the most to sales?

-> collect all data go to insert-> select pivot table -> category insert in row and sales in values

• How has the sales trend been for the past year?

-> =YEAR(range)

• Which region has the highest sales and which one has the lowest?

**-> collect all data go to insert-> select pivot table -> region insert in row and sales in values**

• What is the average profit margin of the store

**-> average(Sum(profit)/sum(sales))**

**Use Stock Price data set.**

**1)• What is the average daily return of the stock?**

Average(sum(close-open)

**2)• What is the distribution of the daily trading volume? Avocado Prices**

Select volue column -> insert->chart -> historgram

**3)• What is the average price of avocados?**

**4)• What is the distribution of the average price by region?**

Create bar char average price vs region

**5)• Which region has the highest and lowest average price?**

Pivot table row=region,value=average price;

**6)• What is the correlation between the total volume and the average price?**

Create bar char of average price and total volume

**7)• What is the distribution of the total volume by year?**

Create pivote table .row =year,value=volume by sum

**Superstore Sales Data Analysis using Excel:**

1. **What is the total revenue generated by the store?**
   * Use the formula: =SUM(Sales Range)  
     (Replace "Sales Range" with the actual range of your sales column.)
2. **Which category of products contributes the most to sales?**
   * Steps:
     1. Select all your data.
     2. Go to **Insert** -> **PivotTable**.
     3. In the PivotTable:
        + Drag **Category** to Rows.
        + Drag **Sales** to Values (set to Sum if needed).
     4. Identify the category with the highest sales.
3. **How has the sales trend been for the past year?**
   * Add a new column with formula: =YEAR(Date Range)
   * Create a PivotTable:
     1. Rows: Year (from the new column)
     2. Values: Sum of Sales
   * Insert a line chart to visualize the trend.
4. **Which region has the highest sales and which one has the lowest?**
   * Steps:
     1. Select all data.
     2. Insert a **PivotTable**.
     3. In the PivotTable:
        + Drag **Region** to Rows.
        + Drag **Sales** to Values.
     4. Identify the region with the highest and lowest total sales.
5. **What is the average profit margin of the store?**
   * Formula: =SUM(Profit Range)/SUM(Sales Range)  
     (Then, format the cell as a percentage.)

**Stock Price Data Analysis using Excel:**

1. **What is the average daily return of the stock?**
   * Create a new column: =Close - Open (for each day).
   * Then, calculate: =AVERAGE(New Daily Return Column)
2. **What is the distribution of the daily trading volume?**
   * Steps:
     1. Select the **Volume** column. Go to **Insert** -> **Chart** -> **Histogram**.

**Avocado Prices Data Analysis using Excel:**

1. **What is the average price of avocados?**
   * Formula: =AVERAGE(Average Price Range)
2. **What is the distribution of the average price by region?**
   * Create a PivotTable:
     + Rows: Region & Values: Average of Average Price.
   * Insert a **Bar Chart** to display the distribution.
3. **Which region has the highest and lowest average price?**

 Use a **PivotTable** (Region in Rows, Average of Average Price in Values)

 Then use a **Bar Chart** or **Column Chart** to **compare** the average prices across regions clearly.

1. **What is the correlation between the total volume and the average price?**
   * Select the **Total Volume** and **Average Price** columns.
   * Insert a **Scatter Plot**.
   * Optionally, use =CORREL(Total Volume Range, Average Price Range) to find the correlation coefficient numerically.
2. **What is the distribution of the total volume by year?**
   * + Create a new column: =YEAR(Date Range).
     + Create a PivotTable:
       - Rows: Year
       - Values: Sum of Total Volume.
     + Insert a **Column Chart** to visualize.

**For** **Tableau**

**Superstore Sales Data Analysis using Tableau:**

1. **What is the total revenue generated by the store?**
   * Connect your Superstore data to Tableau.
   * Drag **Sales** field into the **Text** area of a **new worksheet**.
   * Tableau will automatically **sum** the Sales and show the **total revenue**.
2. **Which category of products contributes the most to sales?**
   * Drag **Category** to Rows.
   * Drag **Sales** to Columns.
   * Tableau will create a **bar chart** automatically.
   * The longest bar shows the **top contributing category**.
3. **How has the sales trend been for the past year?**
   * Drag **Order Date** to Columns.
   * Drag **Sales** to Rows.
   * Right-click on Order Date -> Select "Year".
   * It will show a **sales trend line year by year** automatically.
4. **Which region has the highest sales and which one has the lowest?**
   * Drag **Region** to Rows.
   * Drag **Sales** to Columns.
   * A **bar chart** will show sales by region.
   * The highest and lowest bars show the regions clearly.
5. **What is the average profit margin of the store?**
   * Create a **calculated field**:
     + Go to Analysis → Create Calculated Field.
     + Name: "Profit Margin"
     + Formula: [Profit] / [Sales]
   * Drag this **Profit Margin** field into the **Text** area to see the **average**.

**Stock Price Data Analysis using Tableau:**

1. **What is the average daily return of the stock?**
   * Create a calculated field:
     + Name: "Daily Return"
     + Formula: [Close] - [Open]
   * Drag "Daily Return" to Text and apply **AVG** aggregation.
2. **What is the distribution of the daily trading volume?**
   * Drag **Volume** to Columns.
   * Click on "Show Me" -> Select **Histogram**.

**Avocado Prices Data Analysis using Tableau:**

1. **What is the average price of avocados?**
   * Drag **Average Price** to the **Text** shelf.
   * Apply **AVG** aggregation to display the average price.
2. **What is the distribution of the average price by region?**
   * Drag **Region** to Rows.
   * Drag **Average Price** to Columns.
   * Use a **Bar Chart** to see the distribution.
3. **Which region has the highest and lowest average price?**
   * Same chart as above.
   * The highest bar = highest average price.
   * The lowest bar = lowest average price.
4. **What is the correlation between the total volume and the average price?**
   * Drag **Total Volume** to Columns.
   * Drag **Average Price** to Rows.
   * Tableau will show a **scatter plot**.
   * You can add a **Trend Line** (Right-click on the scatter plot → Trend Lines → Show Trend Lines) to see correlation visually.
5. **What is the distribution of the total volume by year?**
   * Create a calculated field:
     + Name: "Year"
     + Formula: YEAR([Date])
   * Drag **Year** to Columns.
   * Drag **Total Volume** to Rows.
   * Tableau will create a **bar chart** automatically.

✅ **Summary:**

* PivotTables → **Drag and Drop** Fields in Tableau
* Excel Formulas → **Calculated Fields** in Tableau
* Excel Charts → **Automatic Charts** in Tableau using "Show Me"
* Scatter plot, Bar chart, Line chart → All available easily

**For Power bi**

**Superstore Sales Data Analysis using Power BI:**

1. **What is the total revenue generated by the store?**
   * Import the Superstore dataset into Power BI.
   * Create a **Card visualization**.
   * Drag **Sales** into the card.
   * It will automatically show the **SUM of Sales** (total revenue).
2. **Which category of products contributes the most to sales?**
   * Create a **Bar Chart**.
   * Drag **Category** to Axis (X-axis).
   * Drag **Sales** to Values (Y-axis).
   * The longest bar shows the category with the highest sales.
3. **How has the sales trend been for the past year?**
   * Create a **Line Chart**.
   * Drag **Order Date** to Axis.
   * Drag **Sales** to Values.
   * Click on the date field and set it to **Year** (from the Date Hierarchy).
   * You will see a **trend over years**.
4. **Which region has the highest and lowest sales?**
   * Create a **Bar Chart** again.
   * Drag **Region** to Axis.
   * Drag **Sales** to Values.
   * You can clearly see the highest and lowest sales region.
5. **What is the average profit margin of the store?**
   * Create a **New Measure**:
     + Go to "Modeling" → "New Measure".
     + Formula:

powerbi

CopyEdit

Profit Margin = SUM(Profit) / SUM(Sales)

* + Create a **Card** and place this **Profit Margin** measure on it.
  + Format it as **Percentage**.

**Stock Price Data Analysis using Power BI:**

1. **What is the average daily return of the stock?**
   * Create a **New Column**:
     + Formula:

powerbi

CopyEdit

Daily Return = [Close] - [Open]

* + Then, create a **Card** or use an **Average aggregation** in a Table to show **Average(Daily Return)**.

1. **What is the distribution of the daily trading volume?**
   * Create a **Histogram**:
     + Install "Histogram" from **Visualizations** (Marketplace) if not already available.
     + Drag **Volume** into it to see the distribution.

**Avocado Prices Data Analysis using Power BI:**

1. **What is the average price of avocados?**
   * Create a **Card Visualization**.
   * Drag **AveragePrice** field onto it.
   * Set the aggregation to **Average**.
2. **What is the distribution of the average price by region?**
   * Create a **Bar Chart**.
   * Drag **Region** to Axis and **Average of AveragePrice** to Values.
3. **Which region has the highest and lowest average price?**
   * Same bar chart as above.
   * Sort the chart by **Average Price descending** to quickly spot highest and lowest.
4. **What is the correlation between the total volume and the average price?**
   * Create a **Scatter Plot**.
   * Drag **Total Volume** to X-Axis and **Average Price** to Y-Axis.
   * You can visually observe the correlation (linear trend).
   * **Bonus:** You can also calculate Pearson correlation manually by creating a custom measure if needed.
5. **What is the distribution of the total volume by year?**
   * Create a **New Column**:

powerbi

CopyEdit

Year = YEAR([Date])

* + Create a **Column Chart**.
  + Drag **Year** to Axis and **Sum of Total Volume** to Values.

✅ **Summary:**

* **PivotTable in Excel** → **Visuals in Power BI (Card, Bar Chart, Line Chart)**
* **Excel Formulas** → **New Columns and Measures in Power BI**
* **Charts and Histograms** → Directly available or from the Marketplace in Power BI

**Practical No. 7**

def read\_weather\_data(file\_path):

    temperatures = []

    dew\_points = []

    wind\_speeds = []

    with open(file\_path, 'r') as file:

        lines = file.readlines()

        # Skip the header line

        for line in lines[1:]:

            parts = line.strip().split()

            temp = float(parts[1])

            dew = float(parts[2])

            wind = float(parts[3])

            temperatures.append(temp)

            dew\_points.append(dew)

            wind\_speeds.append(wind)

    return temperatures, dew\_points, wind\_speeds

def calculate\_average(values):

    return sum(values) / len(values)

# Read and process data

temps, dews, winds = read\_weather\_data('weather\_data.txt')

# Calculate averages

avg\_temp = calculate\_average(temps)

avg\_dew = calculate\_average(dews)

avg\_wind = calculate\_average(winds)

# Print results

print(f"Average Temperature: {avg\_temp:.2f}°C")

print(f"Average Dew Point: {avg\_dew:.2f}°C")

print(f"Average Wind Speed: {avg\_wind:.2f} km/h")

**Practical No 7**

# code explanation : https://chatgpt.com/c/68027058-4fac-8002-98cb-96500b535125

# nltk ==natural langurage toolkit

# Step 1: Install Required Libraries

# !pip install nltk scikit-learn

# import nltk

# nltk.download('all')

import nltk

from nltk.tokenize import word\_tokenize

from nltk import pos\_tag

from nltk.corpus import stopwords

from nltk.stem import PorterStemmer, WordNetLemmatizer

from sklearn.feature\_extraction.text import TfidfVectorizer

# Step 2: Sample Document

doc1 = "The farmer uses modern technology in agriculture to improve productivity and save water."

# Step 3: Tokenization

tokens = word\_tokenize(doc1)

print("Tokens:", tokens)

# Step 4: POS Tagging

pos\_tags = pos\_tag(tokens)

print("POS Tags:", pos\_tags)

# Step 5: Stop Words Removal

stop\_words = set(stopwords.words('english'))

filtered\_tokens = [word for word in tokens if word.lower() not in stop\_words]

print("After Stop Words Removal:", filtered\_tokens)

# Step 6: Stemming

stemmer = PorterStemmer()

stemmed = [stemmer.stem(word) for word in filtered\_tokens]

print("After Stemming:", stemmed)

# Step 7: Lemmatization

lemmatizer = WordNetLemmatizer()

lemmatized = [lemmatizer.lemmatize(word.lower()) for word in filtered\_tokens]

print("After Lemmatization:", lemmatized)

# Step 8: TF-IDF( Term Frequency – Inverse Document Frequency) Calculation

# Let's assume we have two documents

documents = [

    "The farmer uses modern technology in agriculture to improve productivity and save water.",

    "Technology helps farmers increase productivity and reduce water usage."

]

vectorizer = TfidfVectorizer()

tfidf\_matrix = vectorizer.fit\_transform(documents)

# Show TF-IDF matrix

print("\nTF-IDF Matrix:")

print(tfidf\_matrix.toarray())

# Show feature names (words)

print("Feature Names:", vectorizer.get\_feature\_names\_out())