

imagine you are a data science student and you have learnt sql, python(pandas,numpy ,matplotlib,seaborn ), excel , powe bi and r langedue in school during your second semester, you realise that as time goes you will forget what you have been taught in all those fields thus you intend to you a project approach(doen personal data analysis) the projects should in increase in complexity as time goes and also the projects can i either be revolving around actual real world datasets or small datasets that derive that are a section of those real datas. now let us begin with project number 1, chatgpt i need you to generate for me a dataset meeting the rules above or extract it from the internet, then go ahead and generate for me 15 similar questions that are solvable in all those fields i have studied or 15

#### Memory Updated

To help reinforce your knowledge in SQL, Python (Pandas, NumPy, Matplotlib, Seaborn), Excel, Power BI, and R, let's start with a beginner-friendly project using a real-world dataset. The **Global Superstore** dataset is an excellent choice, as it encompasses various aspects of a retail business, including sales, customer information, and product details. This dataset is commonly used for practice in data analysis and visualization. HackerNoon

#### Project 1: Sales Analysis of Global Superstore

**Objective:** Analyze the sales performance of Global Superstore to identify trends and insights that can inform business decisions.

#### Dataset Overview:

The Global Superstore dataset includes the following variables: HackerNoon

- **Order ID:** Unique identifier for each order. HackerNoon
- **Order Date:** Date when the order was placed. HackerNoon +3
- **Ship Date:** Date when the order was shipped. HackerNoon
- **Ship Mode:** Method used to ship the order (e.g., Standard Class, First Class). HackerNoon
- **Customer ID:** Unique identifier for each customer. HackerNoon
- **Customer Name:** Full name of the customer. HackerNoon
- **Segment:** Customer segment (e.g., Consumer, Corporate). HackerNoon +1
- **Country:** Country where the customer resides. HackerNoon
- **City:** City where the customer resides. HackerNoon
- **State:** State where the customer resides.
- **Postal Code:** Postal code of the customer's location.
- **Region:** Region where the customer resides.
- **Product ID:** Unique identifier for each product. Medium
- **Category:** Product category (e.g., Furniture, Office Supplies). 365 Data Science
- **Sub-Category:** Product sub-category (e.g., Chairs, Binders).
- **Product Name:** Name of the product.
- **Sales:** Sales amount for the order.
- **Quantity:** Number of units sold.
- **Discount:** Discount applied to the order.
- **Profit:** Profit earned from the order.

#### Tasks:

1. **Data Cleaning and Preparation:**
  - Load the dataset into your analysis tool of choice (SQL, Python, Excel, Power BI, or R).
  - Check for and handle missing or duplicate data.
  - Ensure data types are appropriate for analysis.
2. **Descriptive Statistics:**
  - Calculate summary statistics for numerical columns (e.g., mean, median, standard deviation of Sales and Profit).
  - Determine the distribution of categorical variables (e.g., count of orders by Region, Category).
3. **Data Visualization:**

### 3. Data Visualization:

- Create visualizations to represent sales trends over time.
- Develop bar charts to show sales and profit by Category and Sub-Category.
- Use geographical maps to display sales performance across different regions and states.

### 4. Customer Analysis:

- Identify the top 10 customers based on total sales.
- Analyze purchasing patterns across different customer segments.

### 5. Product Performance:

- Determine which products are the best and worst sellers.
- Analyze the impact of discounts on sales and profitability.

### 6. Shipping Analysis:

- Evaluate the delivery time by calculating the difference between Order Date and Ship Date.
- Analyze how Ship Mode affects delivery time and customer satisfaction.

HackerNoon

### Tools and Techniques:

- **SQL:** Perform data extraction and complex queries to summarize sales data.
- **Python:** Use Pandas for data manipulation, NumPy for numerical operations, and Matplotlib/Seaborn for data visualization.
- **Excel:** Utilize pivot tables, charts, and formulas to analyze and visualize data.
- **Power BI:** Create interactive dashboards to present sales insights.
- **R:** Employ data manipulation packages like dplyr and visualization packages like ggplot2 to analyze and visualize data.

### Deliverables:

- A comprehensive report detailing your findings, including visualizations and interpretations.
- An interactive dashboard (if using Power BI or Excel) showcasing key metrics and insights.
- Well-documented code/scripts used for analysis (if using SQL, Python, or R).

### Next Steps:

As you progress, consider increasing the complexity of your projects by incorporating predictive analytics, machine learning models, or integrating multiple datasets. This approach will help solidify your understanding and application of the tools and techniques you've learned.

Remember to document your process thoroughly and reflect on your findings to gain deeper insights and improve your analytical skills.

W W Sources

i can not find the dataset kindly give another link to that dataset

k Global Super Store Dataset - Kaggle  
kaggle.com

I understand you're having trouble accessing the Global Superstore dataset. Here are some alternative sources where you can download it:

#### 1. Kaggle:

- [Global Super Store Dataset by Apoorvaappz](#): This dataset includes comprehensive sales data suitable for analysis.
- [Superstore Dataset by Vivek](#): Another version of the Superstore dataset, focusing on sales and profits.

Kaggle

#### 2. GitHub:

- [Global Superstore Dataset by yannie28](#): A CSV file of the dataset is available for direct download.
- [Global Superstore Dataset by andrewmanuel](#): Provides the dataset in CSV, XLSX, and SQL formats.

GitHub

GitHub

#### 3. Tableau Community:




- [Sample - Superstore Sales \(Excel\).xls](#): A sample dataset provided by the Tableau community, commonly used for practice. Tableau Community +1

**Instructions to Access the Dataset:**

- **Kaggle:**
  - Visit the provided Kaggle link.
  - Sign in to your Kaggle account (registration is free).
  - Navigate to the "Data" section of the dataset page.
  - Click on the download icon to download the dataset in your preferred format.
- **GitHub:**
  - Visit the provided GitHub link.
  - Click on the dataset file (e.g., [Global\\_Superstore\(CSV\).csv](#)). GitHub
  - Click on the "Download" button to save the file to your local machine.
- **Tableau Community:**
  - Visit the provided Tableau Community link. Tableau Community +1
  - Look for the download link for the sample dataset.
  - Click on the link to download the Excel file.

If you encounter any further issues accessing these datasets, please let me know, and I'll be happy to assist you further.



Do you like this personality   

Ask anything

