Vaishnavi Pawar (vpawar@iu.edu)

Fall 2024

Structured Data Exercise

Vaishnavi Vishwas Pawar

Structured Data Excercise Assignment- VP
📊 Project

Share   Bookmark   **Launch workspace**

Overview   Activity   Insights   Discussion   Settings

Overview

DESCRIPTION   Edit

Global Superstore data to explore sales trends, product performance, & customer behavior for operational insight

SUMMARY   Edit

Project Summary: I535 Fall 2024 - Global Superstore Analysis

This assignment examines the Global Superstore dataset, which includes sales and order data for a global retail company. The database consists of details on products, customers, orders, and sales from various markets and regions.

The database consists of two main datasets: the Global Superstore dataset, this dataset includes the product category, city, country, customer_id, etc, and the Global Superstore_login_Dataset, this dataset contains the user_id, market information, and login date of the customer. Our purpose is to gather insights into the company's global operations by researching sales trends, product performance, and consumer behavior.

Additionally, I have gathered supplementary data from the web, which includes additional data about more products from various locations. I have integrated data into the project

About this project   Edit

SHARED WITH   🔒 No one

CREATED   7 hours ago by @vaishnavipawar09

TAGS   + Add tags

Queries (4)   See all

Average Discount by Segment
👥 @vaishnavipawar09 · 17 minutes ago

Sales Performance by Region
👥 @vaishnavipawar09 · 1 hour ago

Top 10 Cities Where Most Pro...
👥 @vaishnavipawar09 · 1 hour ago

Total Sales by Category
👥 @vaishnavipawar09 · 24 minutes ago

Data sources

Project files
3 files, 0 tables

2918diy/Global Superstore

Project Summary

Structured Data Excercise Assignment- VP

Project directory

Project Summary: I535 Fall 2024 - Global Superstore Analysis

This assignment examines the Global Superstore dataset, which includes sales and order data for a global retail company. The database consists of details on products, customers, orders, and sales from various markets and regions.

The database consists of two main datasets: the Global Superstore dataset, this dataset includes the product category, city, country, customer_id, etc, and the Global Superstore_login_Dataset, this dataset contains the user_id, market information, and login date of the customer. Our purpose is to gather insights into the company's global operations by researching sales trends, product performance, and consumer behavior.

Additionally, I have gathered supplementary data from the web, which includes additional data about more products from various locations. I have integrated data into the project and compared the data between the two.

Using all of the datasets, I developed three queries and chart representations for each of them to better visualize them with bar and point charts. Here are the insights I've plotted.

- Average Discount by Segment

visualize them with bar and point charts. Here are the insights I've plotted.

- Average Discount by Segment
- Sales performance by Region
- Total Sales by Category

Data Description:

The Global Superstore dataset appears to be a comprehensive record of sales transactions for a large retail company that operates across multiple countries and regions. It includes the following key information:

Order details: Order ID, Order Date, Ship Date, Ship Mode

Customer information: Customer ID, Customer Name, Segment

Product details: Product ID, Product Name, Category, Sub-Category

Geographic data: Country, City, State, Region, Market

Financial information: Sales, Profit, Shipping Cost, Discount

Other: Quantity, Order Priority, Year, Week Number

Data Dictionary

Share

Project directory + Add

Jump to a file or query (Ctrl + K)

Pr... | Da... × | Glo... × | Sal... × | Glo... × | Tot... × | Sal... × | Sal... × | Pr... × | pac... × | Pr... × | Sal... × | Tot... × | To... × | Av... × | Av... ×

Data dictionary

Home
Project summary
Data dictionary

PROJECT FILES
Average Discount by segment.p...
Sales performance by region.png
Total sales by category.png

CONNECTED DATASETS
Global Superstore (6)
@2918diy · View dataset

QUERIES
Average Discount by Segment
Sales Performance by Region
Top 10 Cities Where Most Prod...
Total Sales by Category

INSIGHTS
No insights...yet.

vaishnavipawar09/structured-data-excercise-assignment-vp
PRIVATE   Updated 9/9/2024

Average Discount by segment.png   Edit
+ Add a description

Sales performance by region.png   Edit
+ Add a description

Total sales by category.png   Edit
+ Add a description

2918diy/global-superstore
OPEN   Updated 5/29/2020

2918diy/global-superstore
OPEN   Updated 5/29/2020

Global Superstore.txt
Request more info

COLUMN NAME	TYPE	DESCRIPTION
category	string	
city	string	
country	string	
customer_id	string	
customer_name	string	
discount	decimal	
market	string	
ji_lu_shu	integer	
order_date	datetime	
order_id	string	
order_priority	string	
product_id	string	

data.world just got better: Load the latest version

---

product_id	string	
product_name	string	
profit	decimal	
quantity	integer	
region	string	
row_id	integer	
sales	integer	
segment	string	
ship_date	datetime	
ship_mode	string	
shipping_cost	decimal	
state	string	
sub_category	string	
year	year	
market2	string	
weeknum	integer	

oject directory   + Add

Jump to a file or query (Ctrl + K)

Home
Project summary
Data dictionary

OJECT FILES
Average Discount by segment.p...
Sales performance by region.png
Total sales by category.png

NNECTED DATASETS
Global Superstore (6)
@2918diy · View dataset

ERIES
Average Discount by Segment
Sales Performance by Region
Top 10 Cities Where Most Prod...
Total Sales by Category

HGHTS
insights...yet.

	ship_date ⓘ	datetime
	ship_mode ⓘ	string
#	shipping_cost ⓘ	decimal
	state ⓘ	string
	sub_category ⓘ	string
	year ⓘ	year
	market2 ⓘ	string
#	weeknum ⓘ	integer

Global Superstore_login_Data.txt

Request more info

COLUMN NAME	TYPE	DESCRIPTION
user_id ⓘ	string	
market ⓘ	string	
market_1 ⓘ	string	
log_in_date ⓘ	datetime	

SQL Queries

Average Discount by Segment SHARED

```sql
SELECT
    Segment,
    AVG(Discount) AS Average_Discount
FROM
    Global_Superstore
GROUP BY
    Segment
ORDER BY
```

3 query results (0.66 seconds) View log

Download ⌄   Ope

Segment	Average_Discount
Consumer	0.1436
Corporate	0.1429
Home Office	0.1409

Project directory + Add

Home
Project summary
Data dictionary

PROJECT FILES
Average Discount by segment.p...
Sales performance by region.png
Total sales by category.png

CONNECTED DATASETS
Global Superstore (6)
@2918diy · View dataset

QUERIES
Average Discount by Segment
Sales Performance by Region
Top 10 Cities Where Most Prod...
Total Sales by Category

Total Sales by Category SHARED

```sql
SELECT
    Category,
    SUM(Sales) AS Total_Sales
FROM
    Global_Superstore
GROUP BY
    Category
ORDER BY
    Total_Sales DESC;
```

3 query results (0.41 seconds) View log

Category	Total_Sales
Technology	4744691
Furniture	4110884
Office Supplies	3787330

Project directory + Add

Home
Project summary
Data dictionary

PROJECT FILES
Average Discount by segment.p...
Sales performance by region.png
Total sales by category.png

CONNECTED DATASETS
Global Superstore (6)
@2918diy · View dataset

QUERIES
Average Discount by Segment
Sales Performance by Region
Top 10 Cities Where Most Prod...
Total Sales by Category

INSIGHTS
No insights...yet.
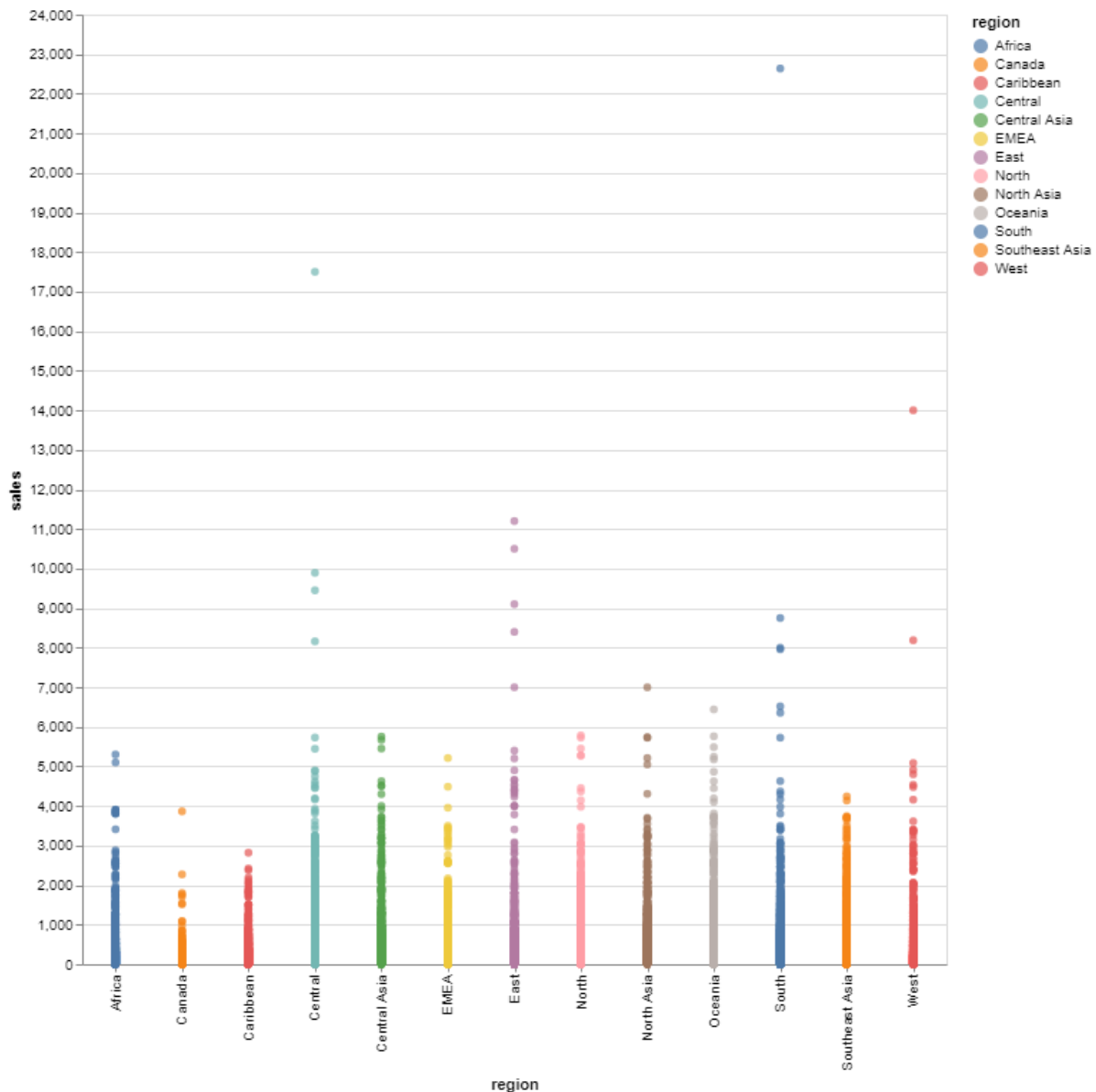
Visualizations

Sales performance by region

The above visualization shows the complete Global store dataset, indicating the sales according to their region. This scatter plot depicts sales performance by region, displaying a wide variety of sales numbers from various worldwide regions. The West and Canada have the highest individual sales points, whereas other regions have a wider dispersed range of sales, demonstrating that performance varies across geographic areas.
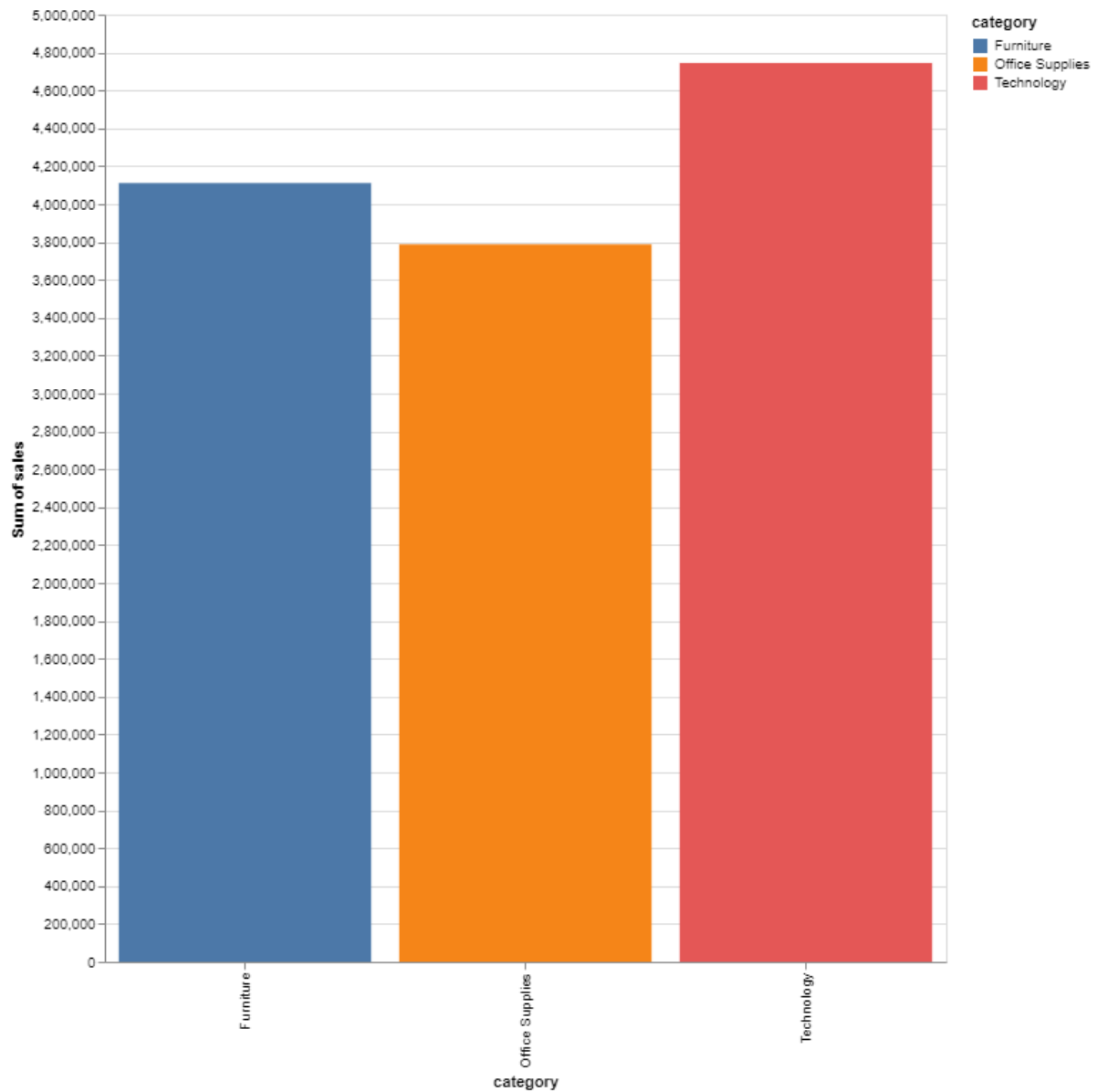
Total sales by Category

This bar chart shows the total sales for three categories: furniture, office supplies, and technology. The Technology category has the highest overall sales, followed by Furniture and Office Supplies, showing that technology-related products are in more demand or more profitable than other categories.

a. What is your impression of this platform? What do you see as its advantages and disadvantages?

Platform Impression: Overall, my experience with data.world was positive. The platform has a user-friendly interface that enables smooth data exploration, manipulation, and collaboration. One significant feature is its SQL query builder, which streamlines the process of querying datasets straight from the platform, removing the need for other tools. The capacity to integrate databases, visualize outcomes, and share discoveries with partners is a major advantage.

Advantages:

- User-friendly Interface: The platform is simple to use, making it suitable for users with varied degrees of technical ability.
- Integrated SQL Query Builder: Users can write and run SQL queries directly from the platform, eliminating the requirement for an external tool.
- Data Visualization Tools: Provides built-in functionality for creating visualizations from datasets, allowing users to easily generate insights.

Disadvantages:

- Performance with Large Datasets: Working with very large datasets might occasionally cause performance concerns or slow down the platform.
- Limited Visualization Customization: In comparison to standalone tools such as Tableau or Power BI, the built-in visualization tool provides less customization choices.

b. What dataset have you used and how would you describe it using the terms you learned in the lecture and readings?

For this research, I used the Global Superstore dataset, which is a structured collection of sales and order data from a global retail corporation. The dataset includes a variety of elements, including product category, client information, regional data, financial details such as sales and profit, and order specifics. In addition, I integrated the Global Superstore Login dataset, which contains customer login data such as user_id, market information, and login date.

Using the concepts from the lecture, this dataset may be classified as structured data because of its well-organized format in rows and columns, which is similar to relational databases. It is multivariate, meaning it contains several features or variables (such as client groups, product categories, and sales quantities) that may be studied together to better understand various elements of the business. This dataset also enables descriptive analytics, which allows for the analysis of sales patterns, product performance, and customer behavior across many areas. I ran relational analysis on the datasets to see how login activity connects with sales and market success.