Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

Department of Computer Science and Engineering (Data Science)

Lab Manual

Subject: Foundations of Data Analysis Laboratory

(DJ19DSL303) Semester: III

Experiment 4

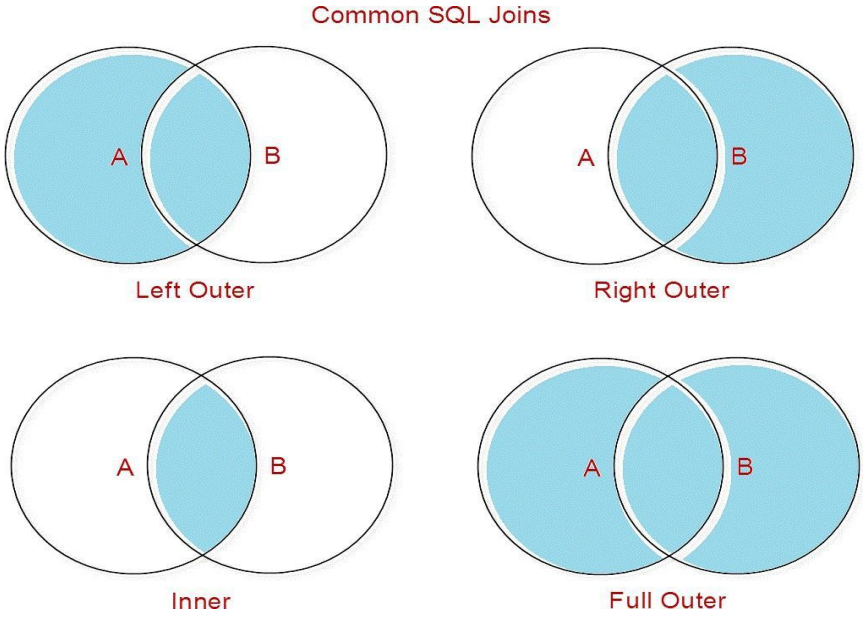
(Data Visualization)

Name:-Rishabh Patil
Sap ID:-60009200056
Div/Batch:-K/K3

Aim: Perform joins, blends and create dual axis charts.

Theory:

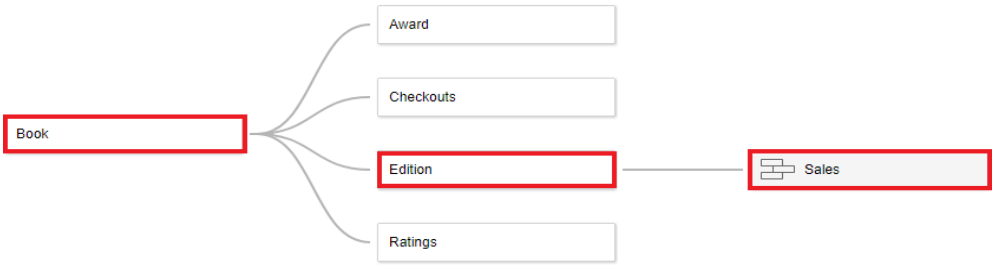
Data joining is a common requirement in any data analysis. You may need to join data from different tables in a single source or join data from multiple sources. A join means combining columns from one or more tables in a relational database. It also creates a set that can be saved as a table, or it can be used as it is. Joins are a concept taken from relational databases like SQL and may be very useful in the future.



Types of joins.

A blend is a smart join on the go. It is always a left join. Blends work for databases of different sources as well. Tableau may automatically suggest/use blends if fields are having the same name. If not, fields can be blended manually. Tableau Relationship

Relationships are a flexible way to combine data for multi-table analysis in Tableau. Think of a relationship as a contract between two tables. When you are building a viz with fields from these tables, Tableau brings in data from these tables using that contract to build a query with the appropriate joins.



When to use Joins & Blends:

Joins: Combining data at row level

Blends: When data sources have different level of granularity or when data sources are from different sources

Relationships: Tableau connects the data at the right level of aggregation. It is a sort of smart on the fly method of connecting tables. It's more flexible than a join. Whenever in doubt, use a relationship. Use a Join or blend if 100% sure that it is required.

Dual Axis Charts:

A dual axis chart is a great way to easily illustrate the relationship between two different variables. They illustrate a lot of information with limited space and allow you to discover trends you may have otherwise missed if you're switching between graphs. Synchronizing Dual Axis charts is very important in most cases.

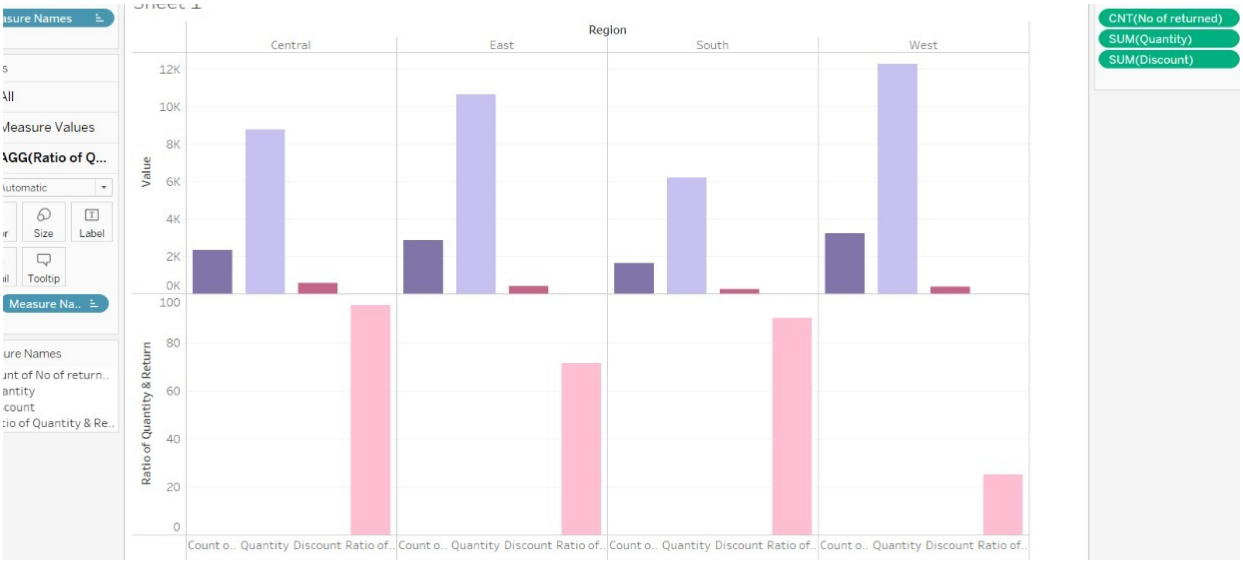
Examples of Dual Axis Charts

Use the given datasets and perform the following tasks:

Datasets:

https://community.tableau.com/s/question/0D54T00000CWeX8SAL/sample-superstore-sales-excelxls https://data.world/2918diy/coffee-chain/workspace/file?filename=Coffee+Chain.txt

a) The company wants to analyze region wise return of products from superstore dataset. Is there any relationship of return of product with the discounts offered on it?



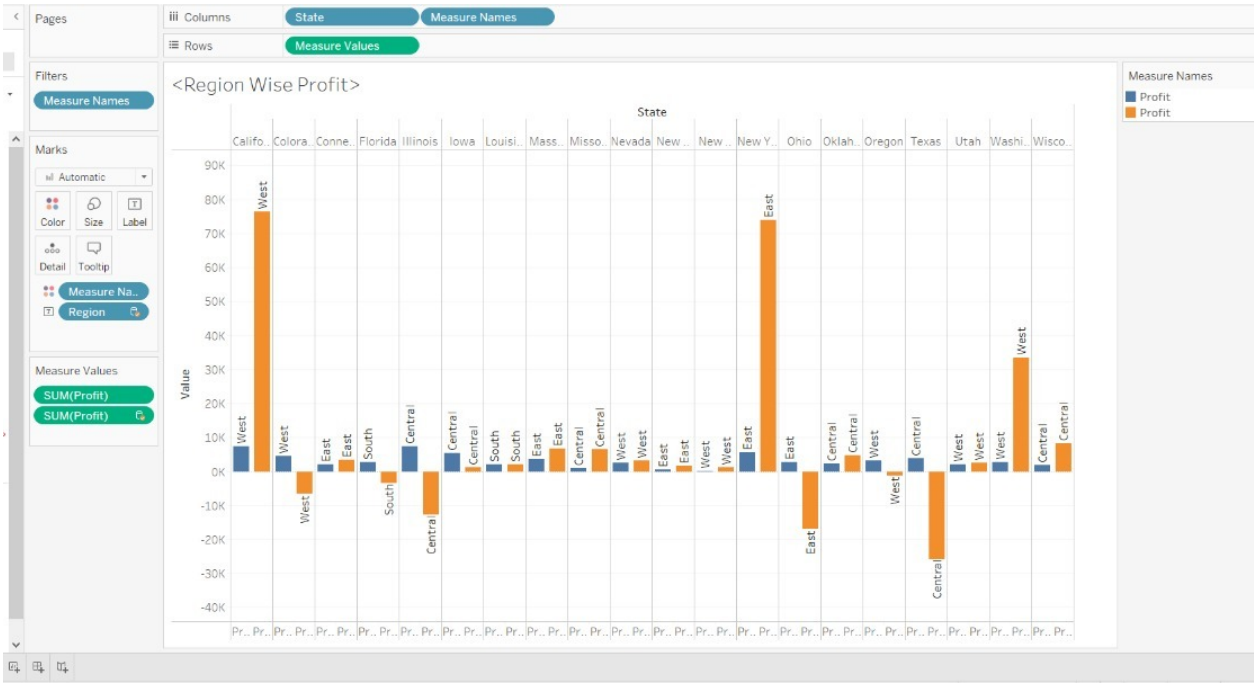
left joining orders and returns of the superstore dataset.

There was a calculated field of ratio of (quantity and returned) to analyze the region wise data.

West region has the highest quantity sold and the ratio is very low which means the count of returned quantity is more.

Central region has the highest ratio which means the quantity sold is much higher then the quantity returned.

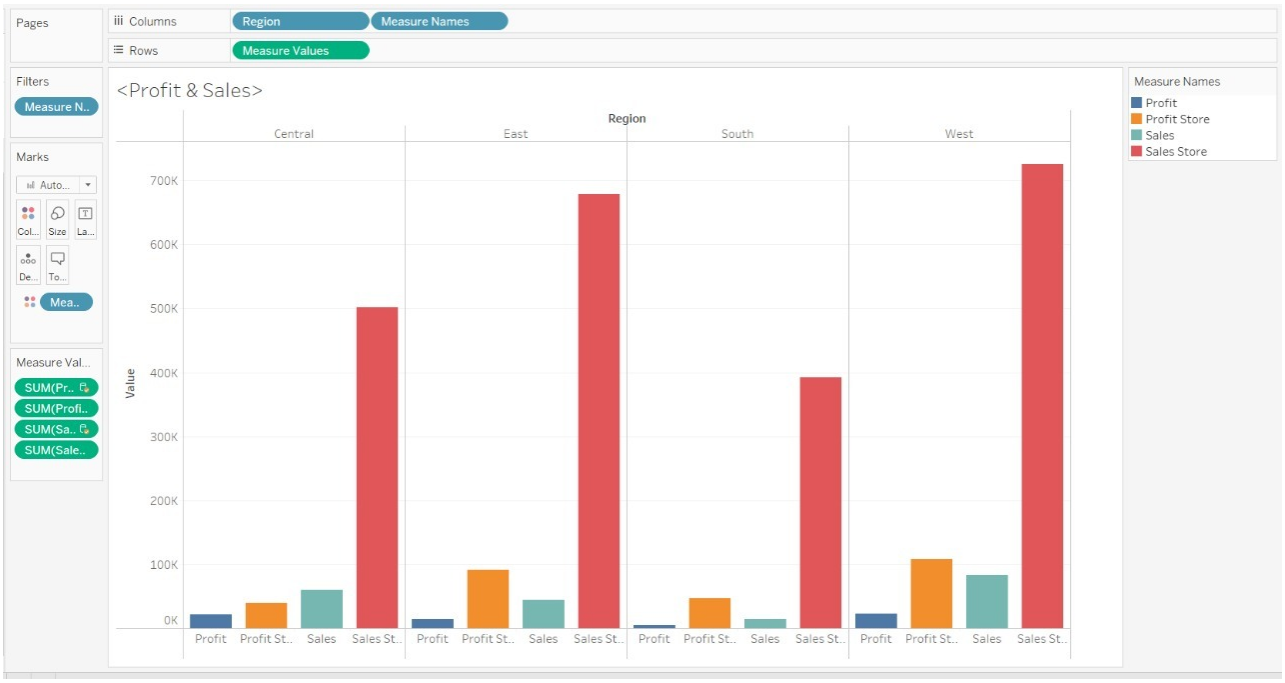
b) Compare the region wise profits of superstore and coffee chain (Market means region). Find which region has performed well for both.



Yellow=profit of superstore
Blue=profit of coffee store

The WEST region has performed the best in terms of profits for the superstore.
The CENTRAL region has performed the best in terms of profit for coffee store.
The **WEST REGION** has fairly performed well for both Superstore and Coffee Store.

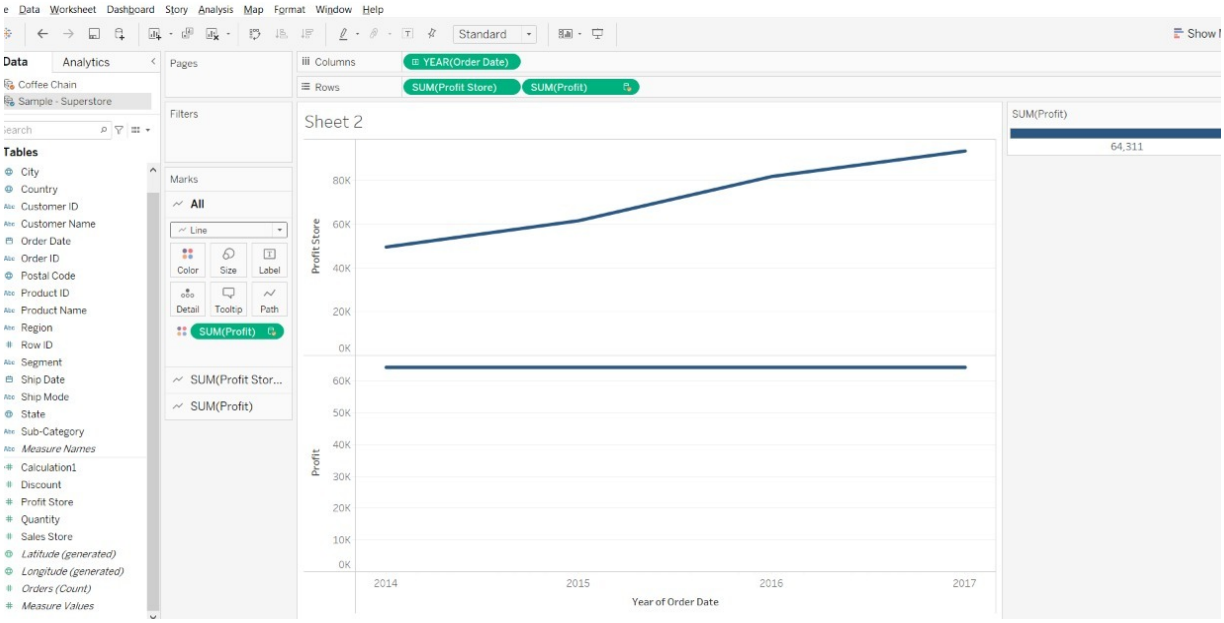
c) Show sales and profit together for both the datasets choosing an appropriate visualization.



Y= Region and values
X= measure values (profit store, profit coffee, sales coffee, sales store)

The **WEST Region** has performed best in sales and profit of both coffee and superstore.

d)Analyze year wise profit for both the datasets and illustrate the limitations of blending.



The profit of the coffee store has remained constant throughout the four years and the profit of the store has increased every year.

The limitations for using blending are that the data of profit for the store is not available from 2010 to 2014.
Therefore we can only visualize the data from 2014 -2017.