

# E-commerce Superstore



**Karan Vibhute**

**BI Analyst, VS Group**

**24.01.2025**

# AGENDA

01

Introduction

02

New Database Structure

03

Tableau Dashboard - EDA

04

KPI

05

K - Means Clustering

06

3D Model Analysis

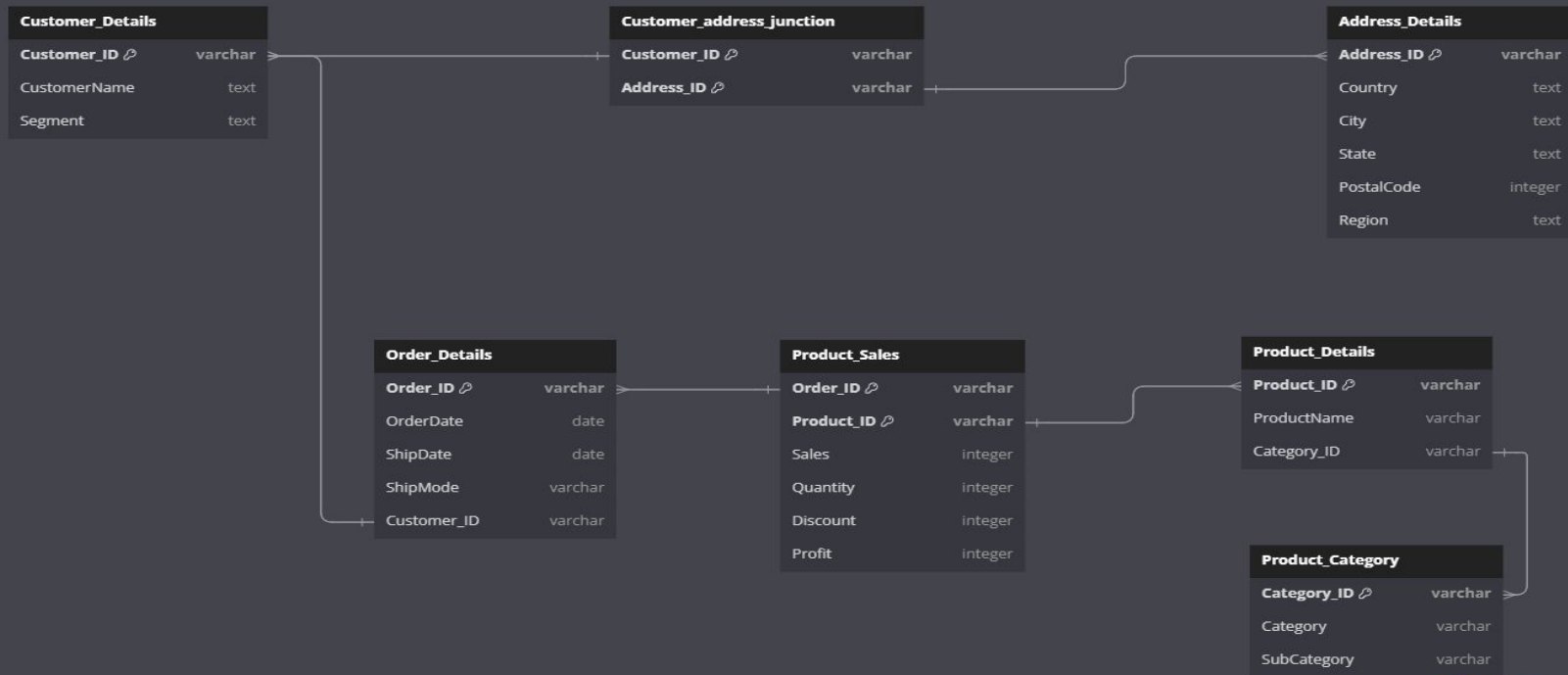
07

Conclusion

# Introduction

- ❖ VS Group is a U.S. company with ambitions to expand its E-Commerce business within the USA
- ❖ Currently Working with CSV File having:
  - 20 fields and 9994 Entries
    - Fields like order Id, order date, Customer ID, customer name, address house no., road no etc.
- ❖ Outdated database structure - want sustainable SQL database solution to enhance operational efficiency and scalability
- ❖ What are the Strength and Weakness and Future Business strategy

# New Database structure



# SQL Dashboard - EDA

## Product

- Highest Selling Category is Binders
- Highest Profit making Category is Binders
- Highest Discount giving Category is Tables
- Tables Category due to highest discount is in loss
- Highest Loss making Category is Machine

## Sales, Profit and Loss

- Highest sales States are California, New York, Texas, Pennsylvania, Washington
- Highest Profit States are California, New York, Washington, Michigan, Burgenia
- Highest Loss making States are Texas, Ohio, Pennsylvania

## Orders and States

- Most Purchase in month September, November and December (Q4)
- Over the Year Number of Orders increased by 25 % to 30 %
- Highest sales and Highest Orders state is California
- Lowest Sales and Lowest Order State is Wyoming

# SQL Dashboard - KPI

## Superstore Dashboard - United Sate Sales KPI

**Sales**  
2,295,510

**Profit**  
286,014

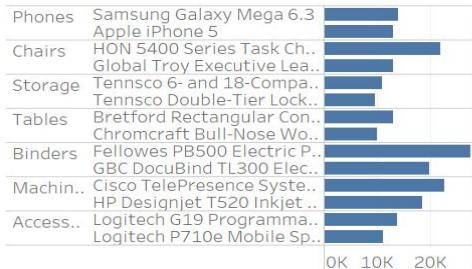
**Number of Orders**  
5,009

**Number of customers**  
793

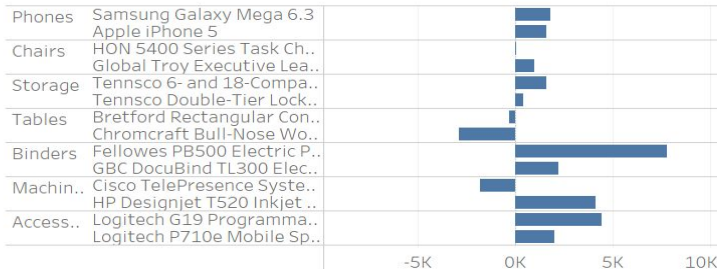
Top

2

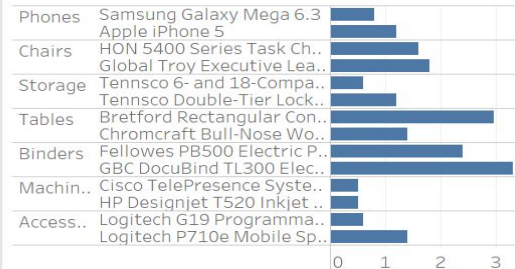
### Top 2 Products Per Categories - Sales



### Top 2 Products Per Categories - Profit



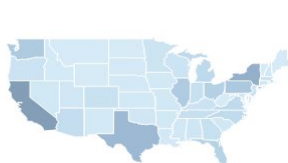
### Top 2 Products Per Categories - Discount



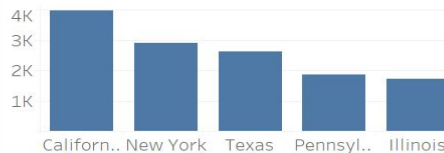
### Sales by Month



### Sales by Country



### Orders Per state



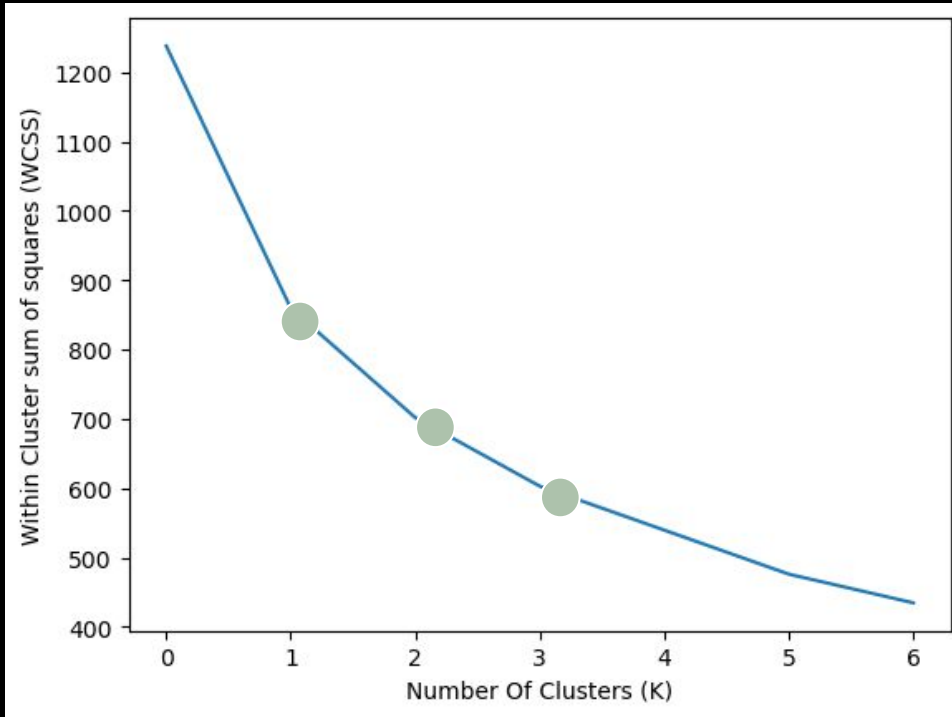
### Year

2014	2014
2015	2015
2016	2016
2017	2017

# K - Means Clustering

- ❖ Created 3 Parameter based on database
  - Amount - total amount of sale for each customer ID
  - Frequency - frequency of Purchase for each customer ID
  - Recency - no of days from last purchase for each customer ID
- ❖ Remove Outliers - K-Means algorithm is susceptible to outliers.
- ❖ Feature Scaling - Scale the data between 0 to 1
- ❖ Building the model - finding Optimum Value of K with two methods
  - Elbow Method
  - Silhouette Analysis

## Elbow Method



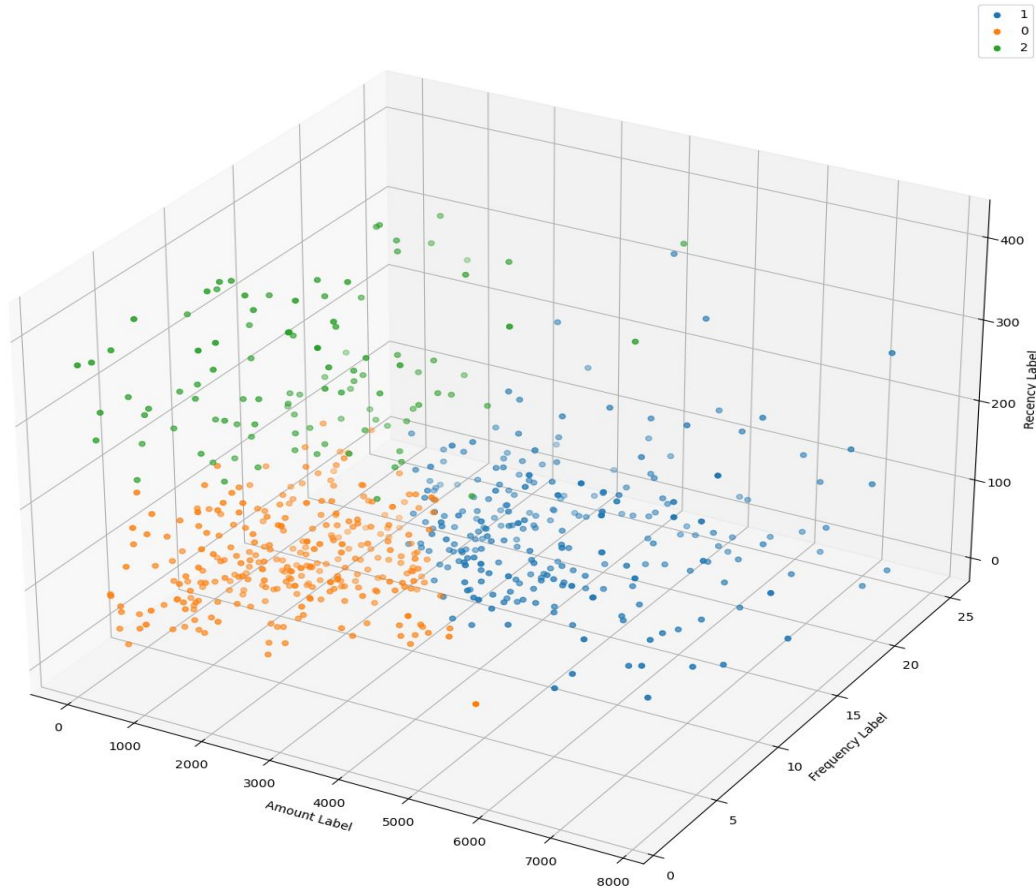
**Optimal K value = 3**

## Silhouette Analysis

- clusters=2, the silhouette score is 0.34127785935046806
- clusters=3, the silhouette score is 0.36377534980419834
- clusters=4, the silhouette score is 0.3186376015115111
- clusters=5, the silhouette score is 0.317844770727942
- clusters=6, the silhouette score is 0.30979346292869453



# 3D analysis of K-means Clustering



- ❖ **Premium Customer** - High amount and High Frequency
- ❖ **Stable Customer** - Medium amount and Medium Frequency
- ❖ **Deal seekers** - Low amount and High Frequency



# Conclusion

01

Provided new Database, where having unique entries and take less storage place

02

New Tableau Dashboard is created with which :

❖ Can be check EDA Based on:

- Product
- Sales
- Profit
- Discount

❖ We Can track KPI

- Monthly
- quarterly
- yearly

03

As Per Clustering we should focus on Group 2 - Deal Seekers Customers



**THANK  
YOU**