

#### Introduction to Data Management PROJECT REPORT

(Project Semester August-December 2021)

**PROJECT REPORT ON**

**Superstore Data 2011-15**

Submitted by

#### GUNDUKA SRINIVAS

#### 11910285

Program: Bachelor of Technology Section: KM005

Course Code: INT217

Under the Guidance of **Komal Arora: 17783 Assistant Professor**

#### Discipline of CSE/IT

**Lovely School of Computer Science & Engineering Lovely Professional University, Phagwara**

**DECLARATION**

*I, Gunduka Srinivas, student of Computer Science & Engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.*

#### Date: 16/12/2021

**Gunduka Srinivas**

**Registration No: 11910285**

**Signature:** 

**ACKNOWLEDGEMENT**

*Primarily I'd thank God for being able to complete my project with success. Then I'd like to thank my mentor* ***Ms. Komal Arora****, whose valuable guidance has been the ones that helped me patch this project and make it full proof success in contribution towards the completion of this project.*

*Finally, I'd rather thanks to* ***Lovely Professional University,*** *and my parent’s inspiration, who gave me this golden opportunity to learn many new things, to learn another aspect of life.*

#### -Gunduka Srinivas

**CONTENTS:**

|  |  |  |
| --- | --- | --- |
| ***Sr No*.** | ***Title*** | ***Page No.*** |
| *1* | *Introduction* | *5* |
| *2* | *Objectives/Scope of the Analysis* | *6* |
| *3* | *Source of dataset* | *7-8* |
| *4* | *ETL Process* | *9-18* |
| *5* | *Analysis of dataset* | *19-29* |
| *6* | *List of Analysis with results* | *30-35* |
| *7* | *Final Dashboard* | *36* |
| *8* | *Bibliography* | *37* |

**INTRODUCTION**

*Global Superstore is a global online retailer, boasting a broad product catalogue and aiming to be a one-stop-shop for its customers. Global Superstore’s clientele, hailing from 147 different countries, can browse through an endless offering with more than 10,000 products. This large selection consists of three main product categories: office supplies (e.g., staples), furniture (e.g., chairs), and technology (e.g., smartphone).*

*Tableau is a data analysis and visualization tool which is commonly used in today's industry. Many organizations still find it important for the research relevant to data science. The ease of use of Tableau is due to it providing a drag and drop interface. This feature helps to perform tasks like sorting, comparing, and analyzing, very easily and fast. Tableau is also compatible with multiple sources, including Excel, SQL Server, and cloud-based data repositories which makes it an excellent choice for Data Scientists.*

# OBJECTIVES/SCOPE OF ANALYSIS

*After analysis of the dataset, the aim of this project is to give answer of given objectives in easy way:*

* *Region wise distribution of Sales and Profits*
* *Discounts based on Category and Sub-category*
* *Sales and Profits based on Category and Sub-category*
* *Sales and profits Trend over time*
* *Top 5 Customer & Products5*

# SOURCE OF DATASET:

#### Source of dataset:

https://www.kaggle.com/jr2ngb/superstore-data

*Kaggle is an online community for data scientists and machine learners, developed by Google. Kaggle allows users to find and publish data sets, explore, and build models in a web- based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges. Kaggle got its start by offering machine learning competitions and now also offers a public data platform, a cloud- based workbench for data science, and short form AI education. On 8 March 2017, Google announced that they were acquiring Kaggle.*

*This data science project analyzes the Superstore 2011-2015 dataset. It was created for the (B. Tech CSE fifth semester Introduction to Data Science course) project.*

*Every part of the dataset consists of multiple of punctuation errors which is cleaned in the ETL process.*

#### *Sample of dataset with data fields is given below:*

#### *Here we can see multiple columns with names such as Row ID, Order Id, Order Date, Ship Date, Ship Mode, customer ID, Customer Name……………so on*

Table

Description automatically generated with medium confidence

**ETL PROCESS:**

*ETL is defined as a process that extracts the data from different RDBMS source systems, then transforms the data (like applying calculations, concatenations, etc.) and finally loads the data into the Data Warehouse system. ETL full form is Extract, Transform and Load.*

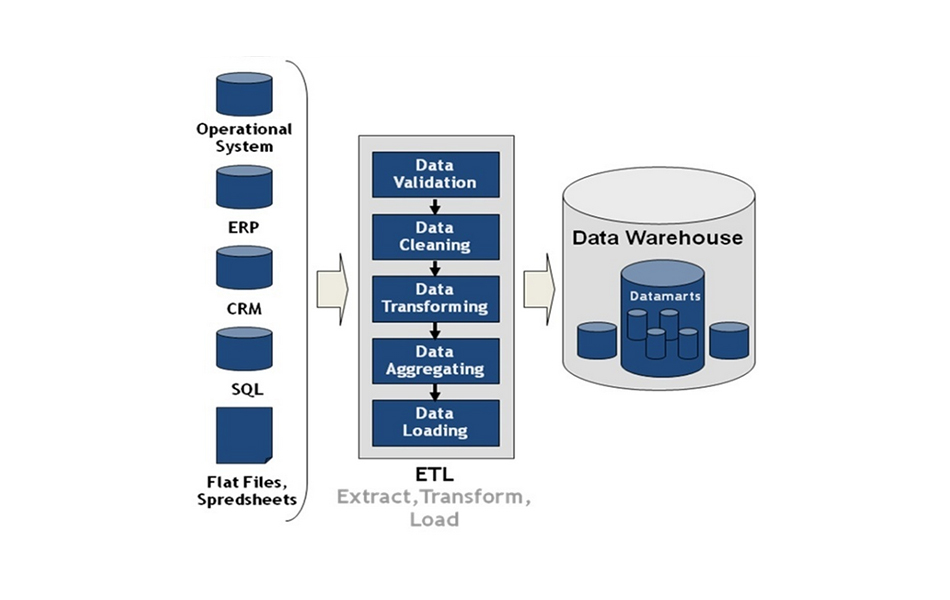
*It's tempting to think a creating a Data warehouse is simply extracting data from multiple sources and loading into database of a Data warehouse. This is far from the truth and requires a complex ETL process. The ETL process requires active inputs from various stakeholders including developers, analysts, testers, top executives and is technically challenging.*

**Need of ETL Process**

* *ETL process allows sample data comparison between the source and the target system.*
* *ETL is a predefined process for accessing and manipulating source data into the target database.*
* *Allow verification of data transformation, aggregation, and calculations rules*.

*When it comes to the implementation of the ETL process, the itinerary of tasks can be divvied up into the full form of its acronym*.

* 1. **E – Extraction**
  2. **T – Transformation**
  3. **L – Loading**

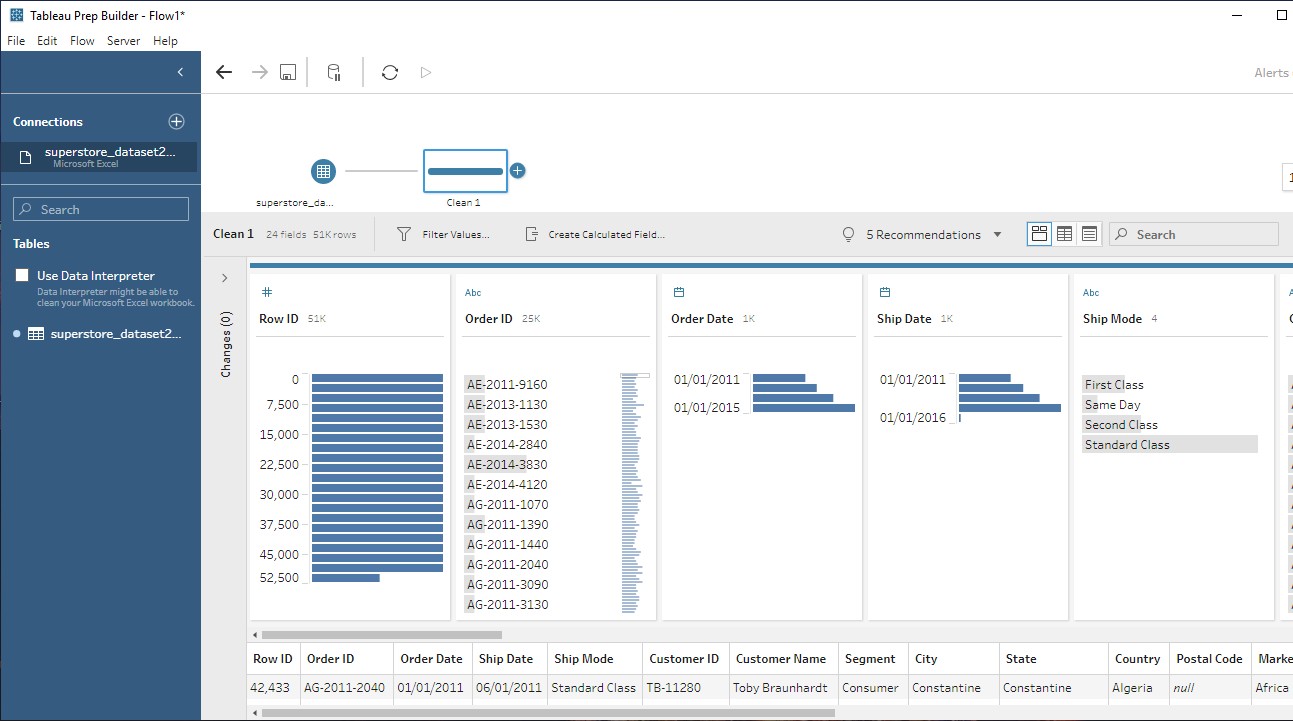


# ETL Process used in Project

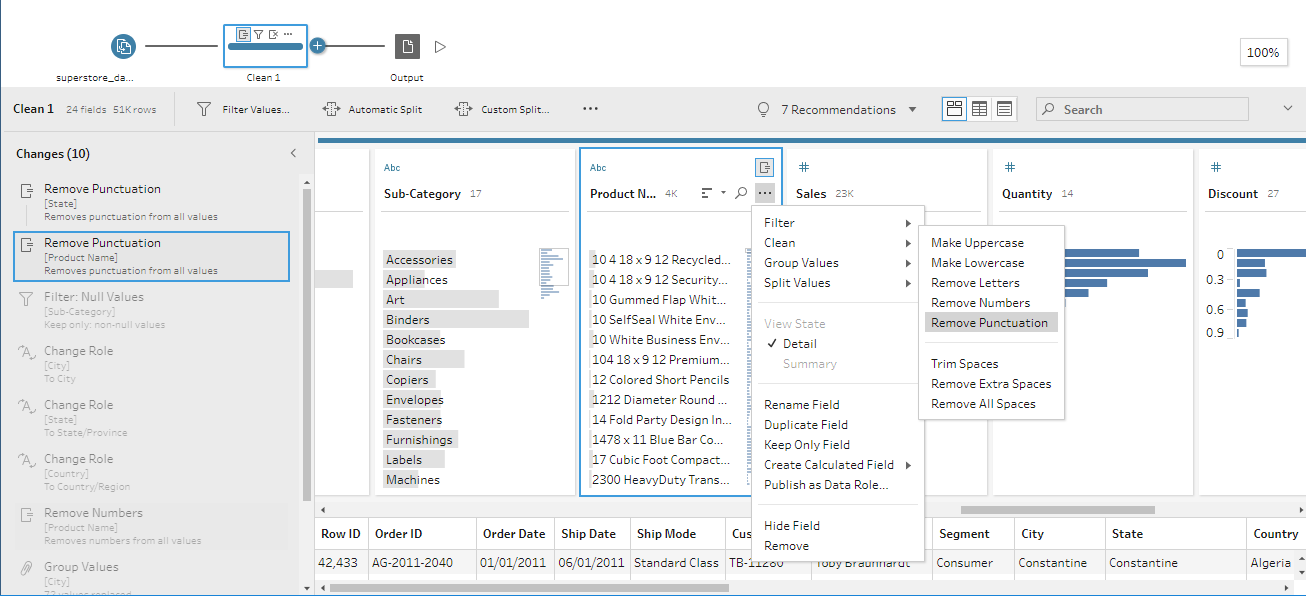
**Extraction**

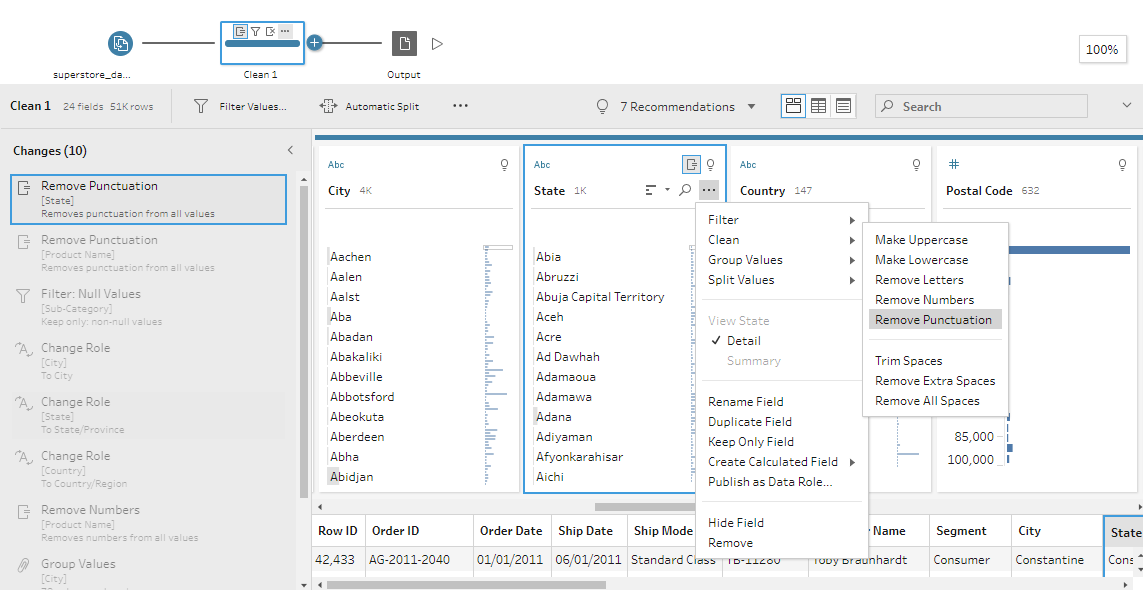
*Extracting the dataset from PC to Tableau for removing the unwanted characters, fields, spelling errors etc.*

#### Step 1: *Opening the dataset in single table.*

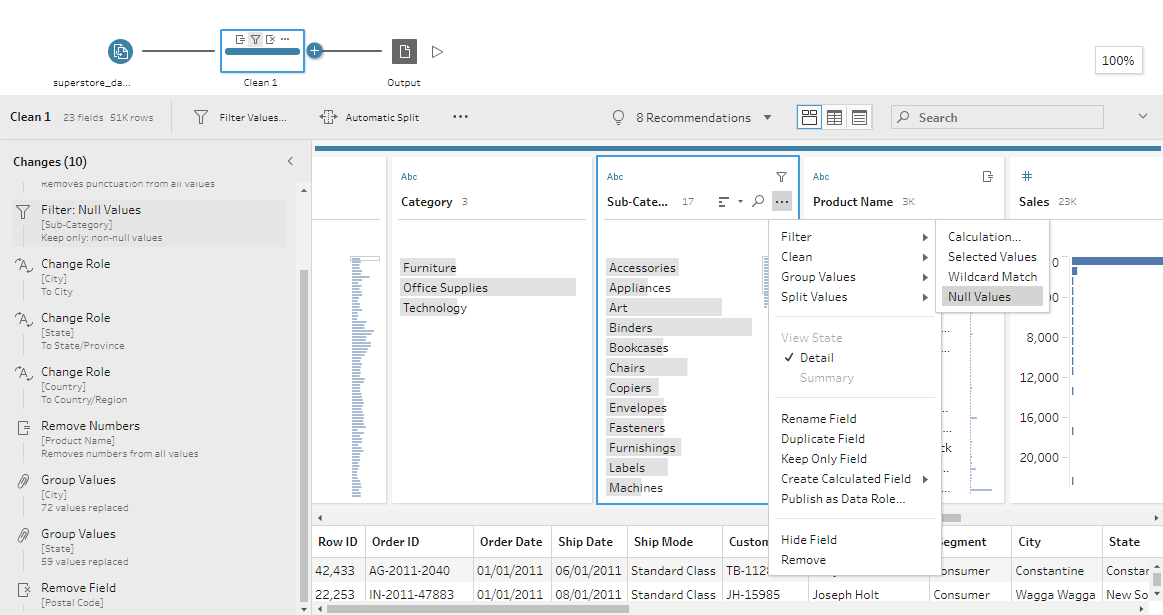


**Step 2*:*** *Removal of punctuations in Product Name and state column*

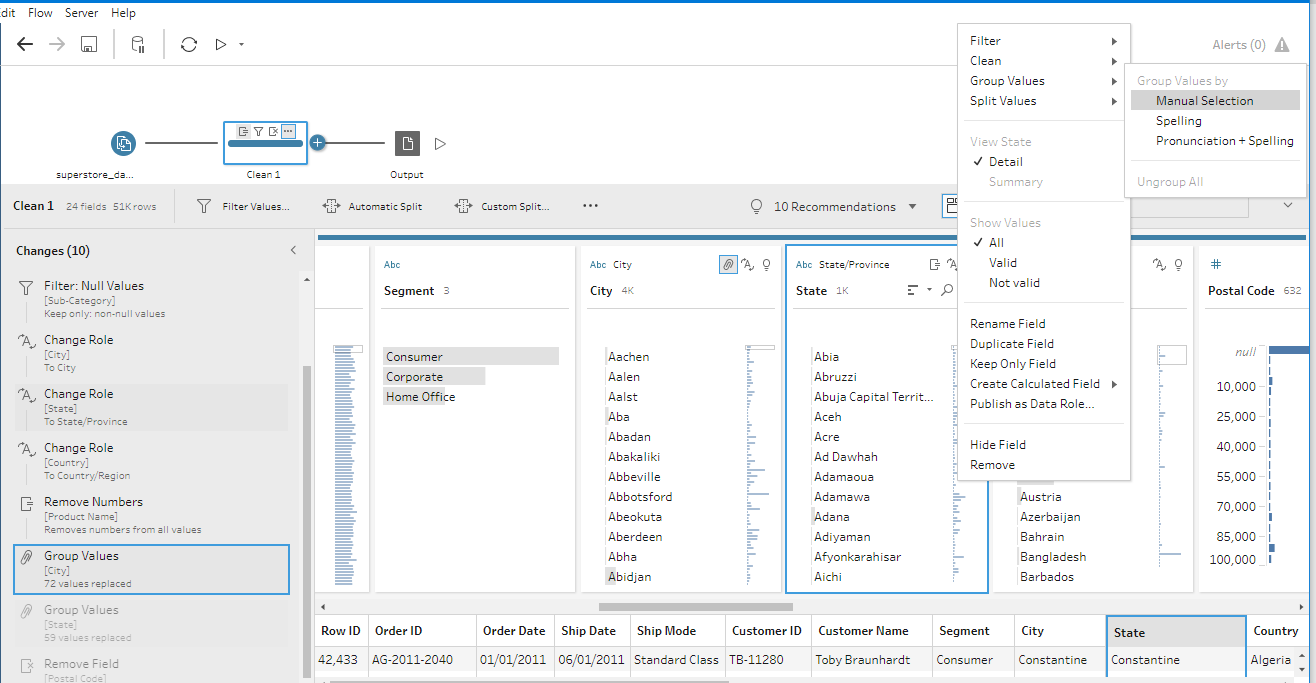


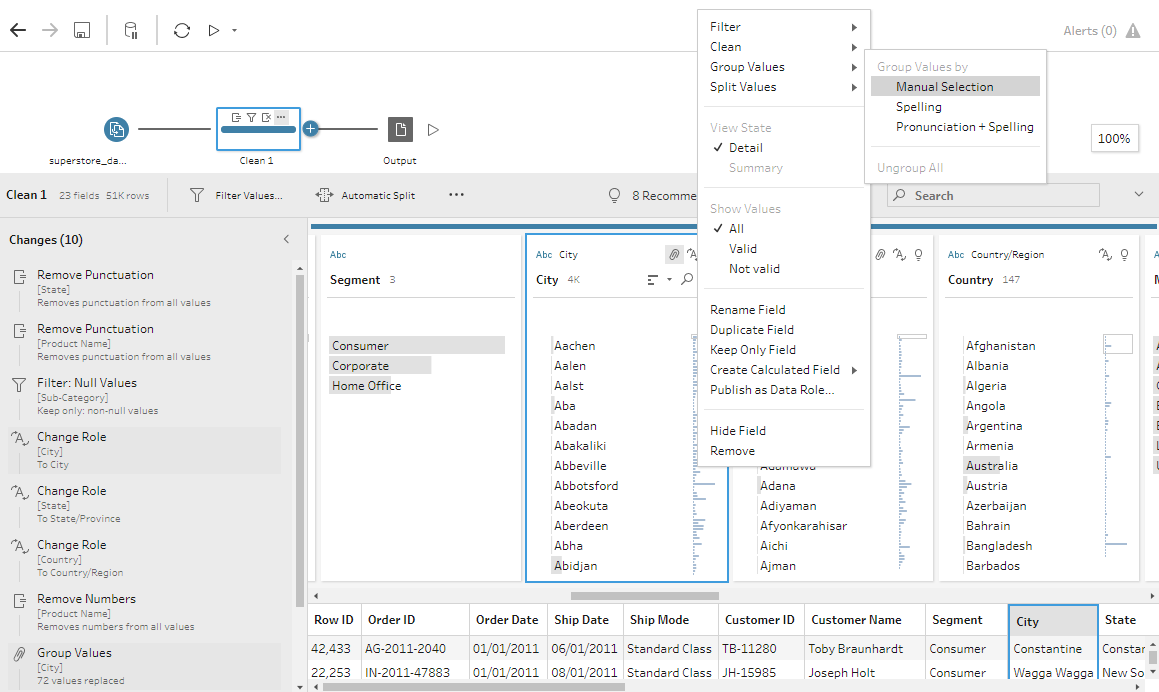


**Step 3***: Removal of Null values*

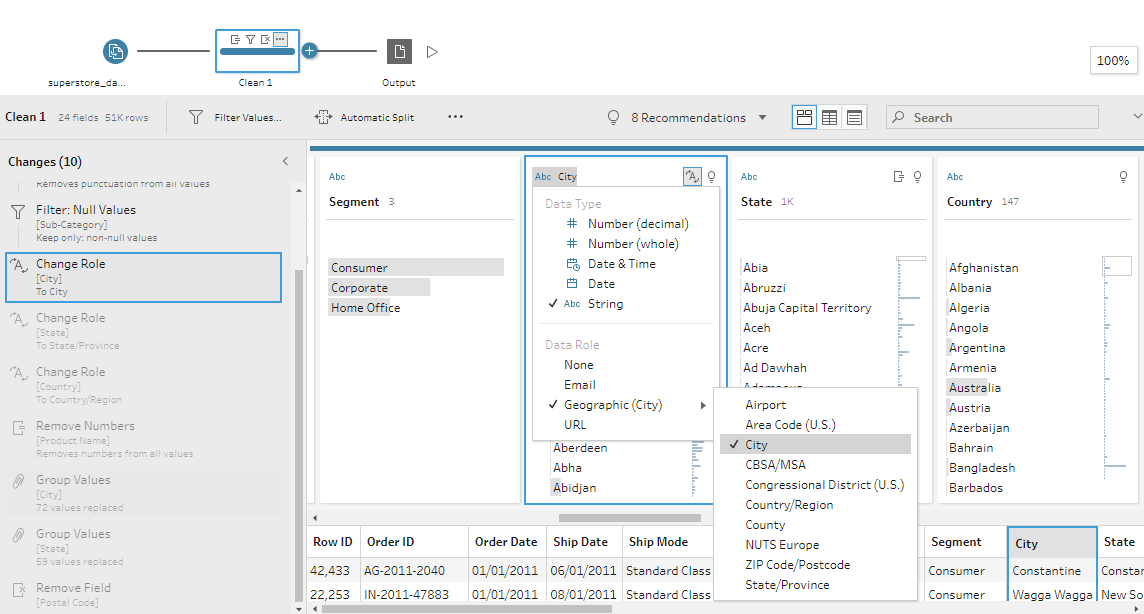


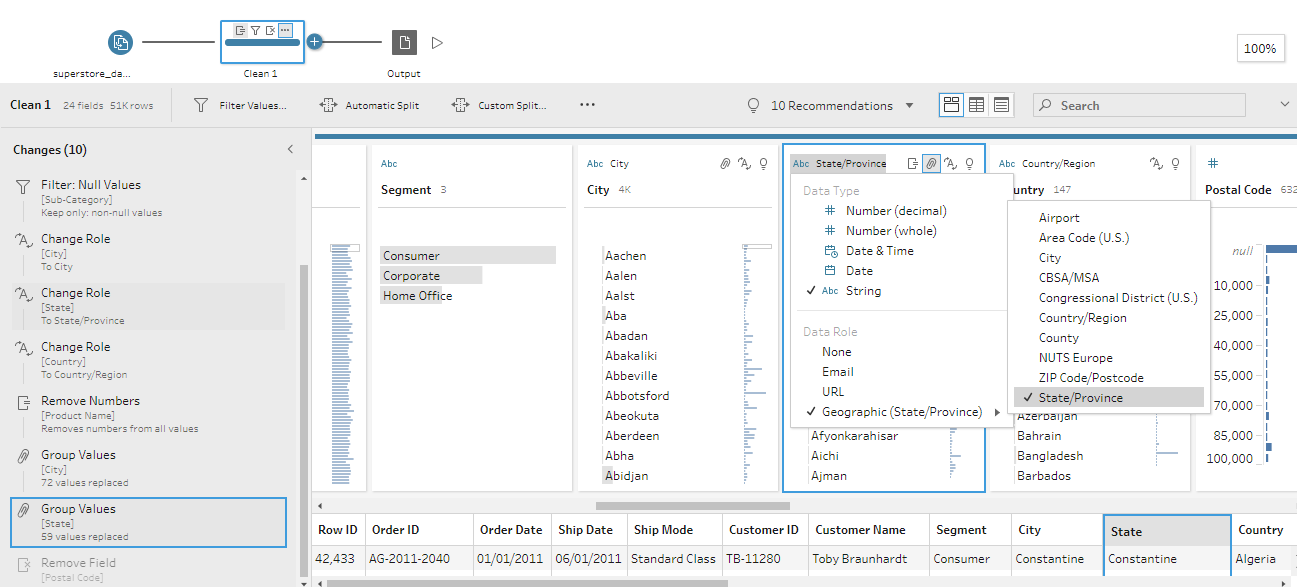
**Step 4***: Grouping values in “State” and “City “columns*





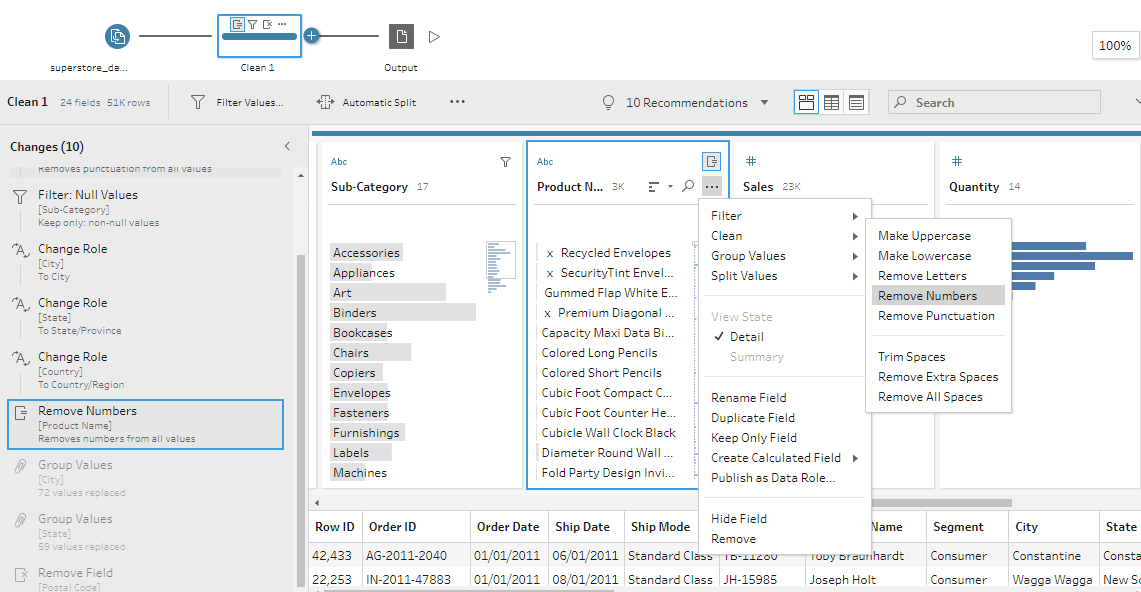
**Step 5:** *Change data type of city, state, and country*



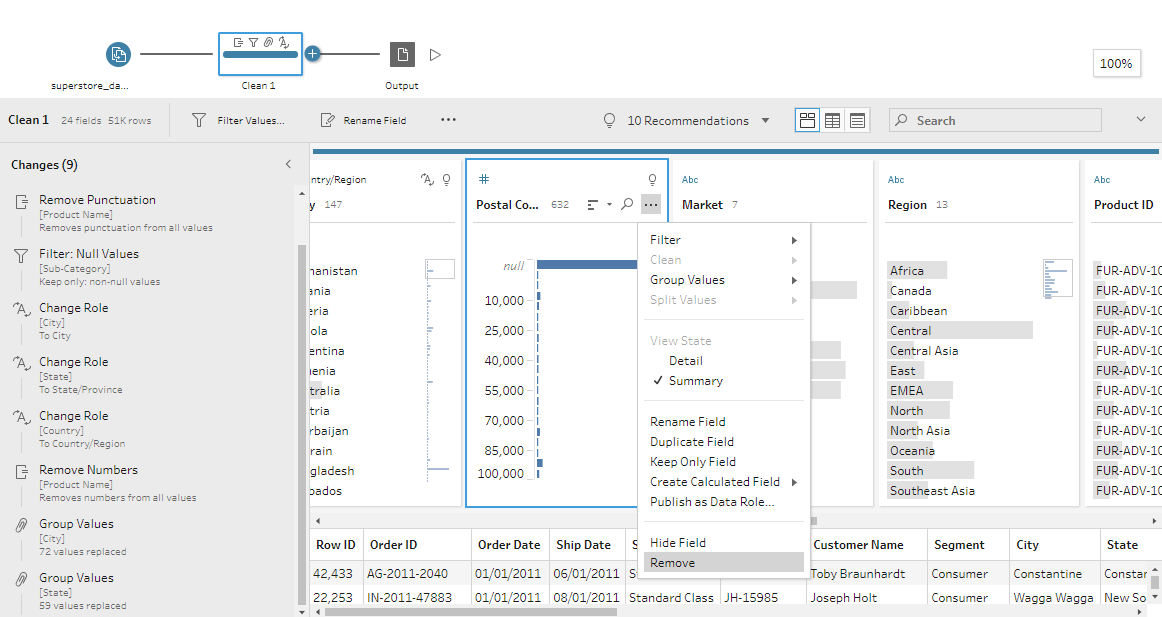




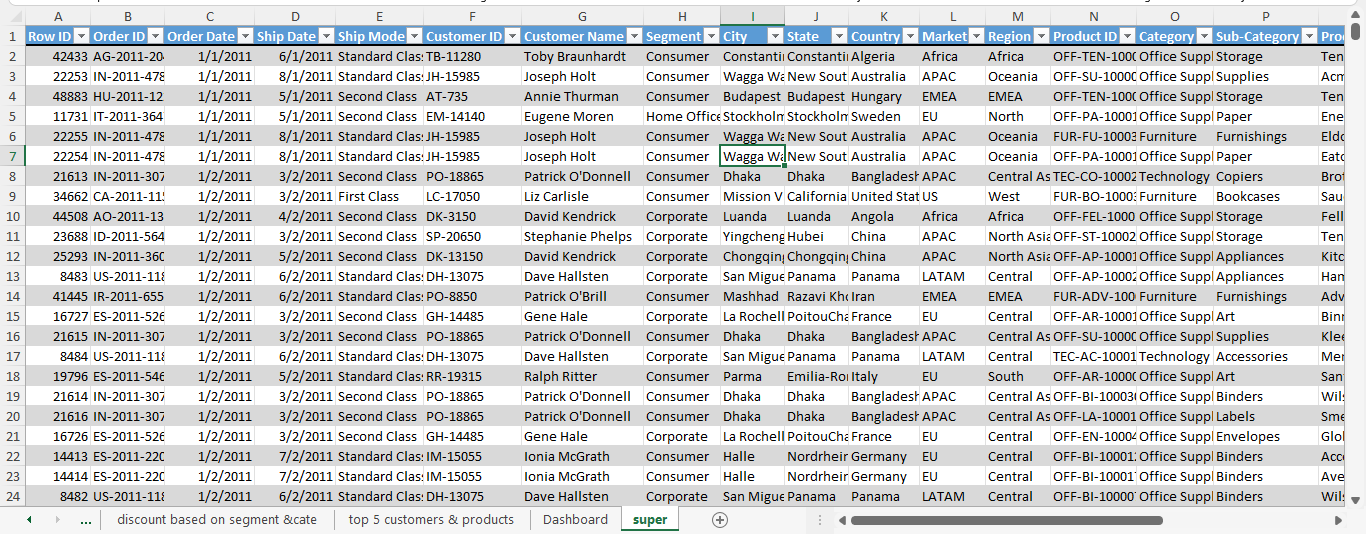
**Step 6***: Apply cleaning process of removal of numbers in “Product name” table*.



**Step-7**: *Removal of Postal Code column*.



#### *Finally, after cleaning the data, the final dataset sample is shown below:*



**Analysis on dataset**

1. ***Region wise distribution of Sales and Profits***

***Introduction***

* *By performing this analysis, we will get Region wise distribution of sales and profits*.

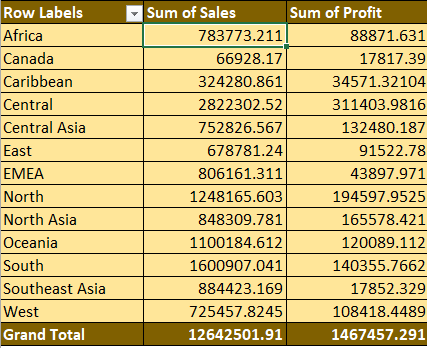
#### *Description:*

*The It is customary to see the rate of growth in sales for a mature region begin to decline and then settle into a relatively tight range over time. The sales trend for a new region is highly dependent on the buildout of a distribution system, retail stores, and/or a regional sales force*.

#### *Specific requirements, functions, and formulas:*

#### *For Grand Total of sales and profit we can use sum function: =SUM()*

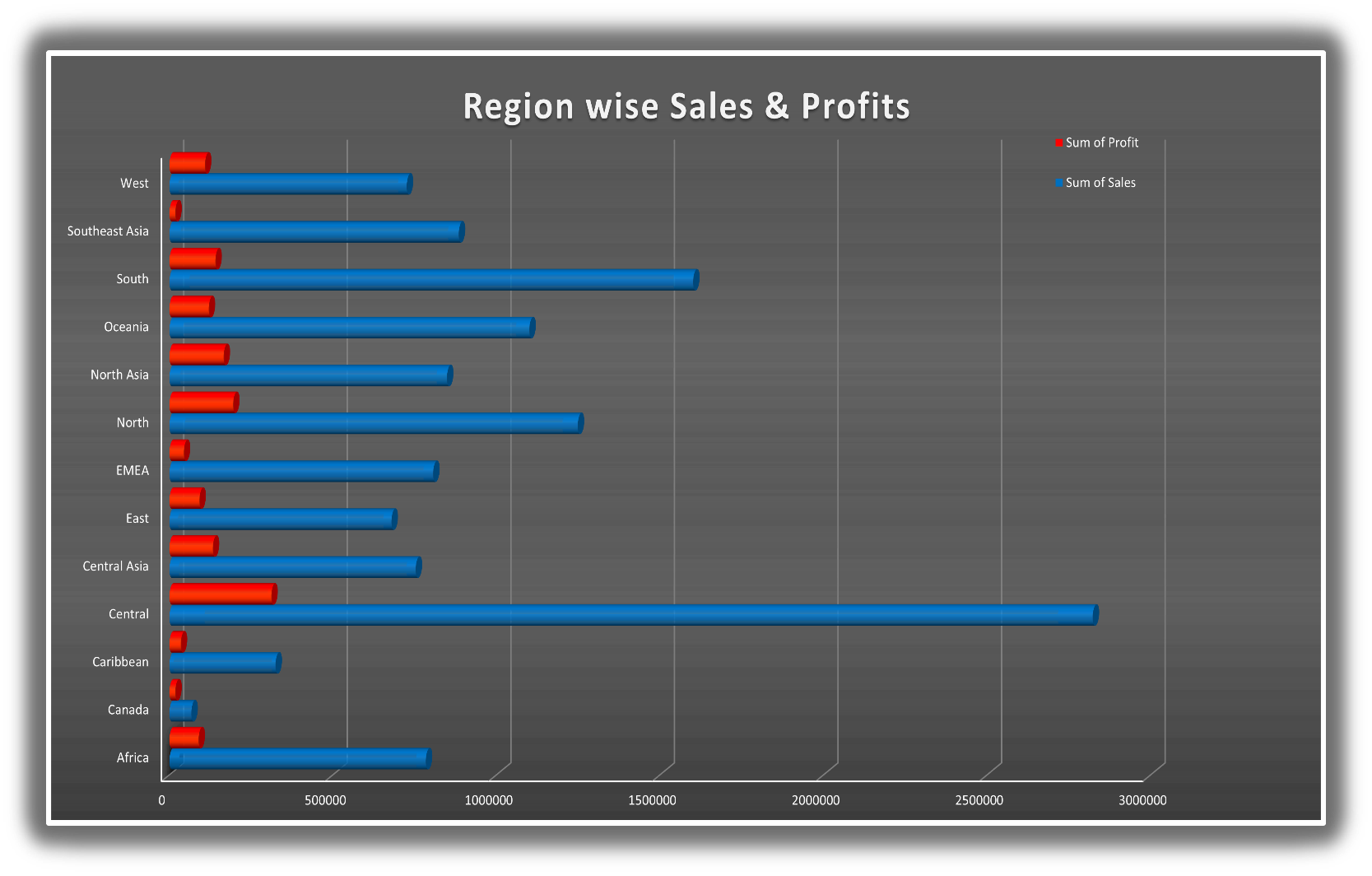
***Analysis results:*** *South region has the most sales while Central region has the most profits on compared with sales*.



#### *Slicer:*

#### 

#### *Visualization:*



1. ***Discounts based on Category and Segment***

***Introduction***

* *By performing this analysis, we will get discount based on category and segment*

#### *Description:*

*The analysis is based on how much discount will get for category which is part of segment*

#### *Specific requirements, functions, and formulas:*

*Formula =MAX (number1, [number2], …) Number1 and number2 are the arguments used for the function, where Number1 is required, and the subsequent values are optional.*

***Analysis results:***

*Home office and corporate has the highest discount and technology has lowest discount in every category.*

Table

Description automatically generated

***Slicer:***

Graphical user interface, text, application

Description automatically generated

***Visualization:***

A screenshot of a computer

Description automatically generated with low confidence

**3. *Sales and Profits based on Category and Sub-category***

***Introduction***

* *By performing this analysis, we will get sales and profits based on category and sun-category*

#### *Description:*

*The analysis based on about sales of sub-category as well as profits of sub-category which is present in category*

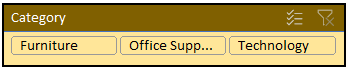
#### *Specific requirements, functions, and formulas:*

#### *Select the cell below the given Quantity and apply the formula ‘=Sum ()’. This function will add the numbers in a range of cells*

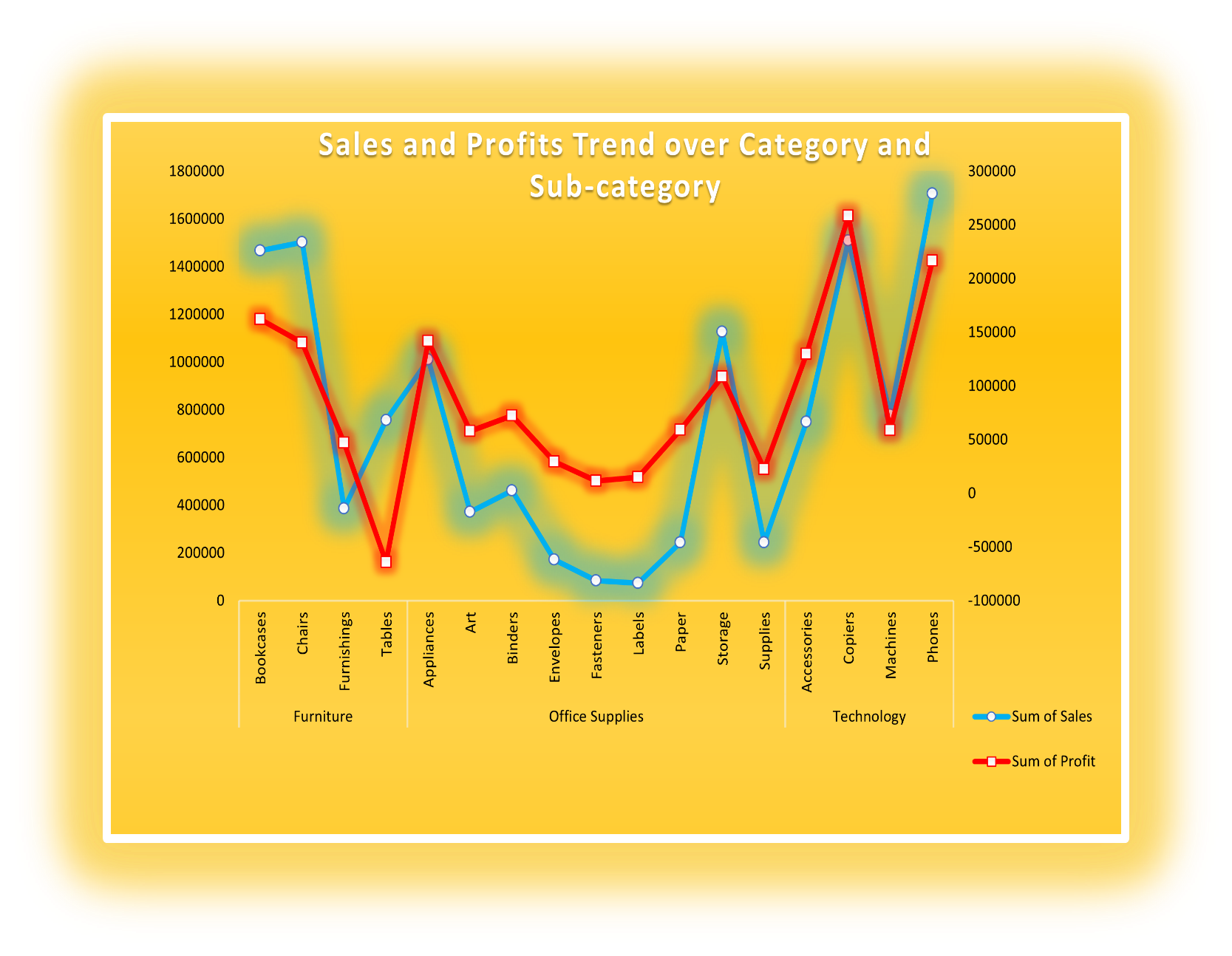
#### *Analysis results:*

#### Table Description automatically generated

***Slicer:***



***Visualization:***



**4.** ***Sales and profits Trend over time***

***Introduction***

* *By performing this analysis, we will get sales and profits in different years.*

#### *Description:*

*Sales trend over time also helps to determine are we meeting our sales goals. it provides easy, measurable way to track our progress. it will inform the increase in sales at what percentage from last year or over the year*.

#### *Specific requirements, functions, and formulas:*

#### *Select the cell below the given Quantity and apply the formula ‘=Sum ()’.*

#### *This function will add the numbers in a range of cells*

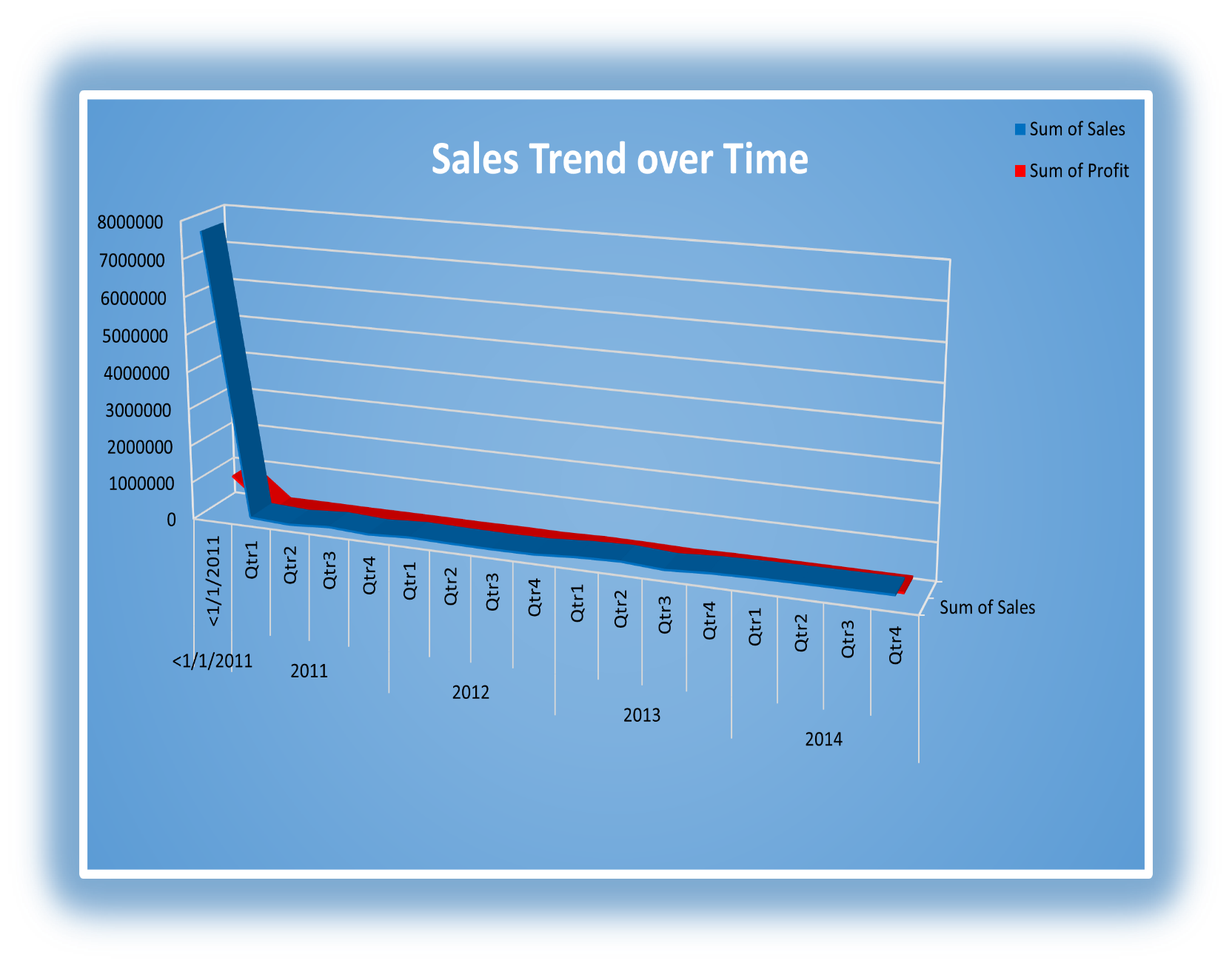
#### *Analysis results:*



#### *Slicer:*

#### 

#### *Visualization:*



**5.** ***Top 5 Costumers & Products***

***Introduction***

* *By performing this analysis, we will get top 5 costumers and products*

#### *Description:*

#### *5 Most Successful Products Ever and What Small Businesses Can Learn from Them 1 Set your business’s next product design on the path to profitability with the valuable lessons*.

#### *Specific requirements, functions, and formulas:*

#### *Select the cell below the given Quantity and apply the formula ‘=Sum ()’.*

#### *This function will add the numbers in a range of cells*

#### *Analysis results:*

Table

Description automatically generated

Table

Description automatically generated

#### *Slicer:*

#### Table Description automatically generated with low confidence

***Visualization:***

Chart, line chart

Description automatically generated

Graphical user interface, chart

Description automatically generated

**List of Analysis with results**

1. ***Top 5 Regions on sales and profits:***

*1.South*

*2.North*

*3.Oceania*

*4.Southeast Asia*

*5.Central Asia*

Chart, bar chart

Description automatically generated

1. ***Top Segments in top category:***

*1.Corporate Furniture*

*2.Home Office Furniture*

*3.Coporate Furniture*

Chart, pie chart

Description automatically generated

## *Top 4 sales in sub-category:*

## *1.Phones*

## *2.Copies*

## *3.Chairs*

## *4.Bookcases*

## Chart, line chart Description automatically generated

.

## *Most Sales & profits Happen in 2014:*

*The rise in sales and profits is visible each year. There was a high incline in the year 2014 by the fourth quarter*.

Chart, surface chart

Description automatically generated

## *Top 5 Costumers & Products:*

## *Costumers:*

## *1.Tom Ashbrook*

## *2.Tamara Chand*

## *3.Grag Tran*

## *4.Christopherconant*

## *5.Sean Miller*

Chart, line chart

Description automatically generated

***Products:***

*1.Apple Smart Phone Full Size*

*2.Cisco Smart Phone Full Size*

*3.Motorola Smart Phone Full Size*

*4.Nokia Smart Phone Full Size*

*5.Canon Image Class Advanced Copier*

Graphical user interface, chart

Description automatically generated

# FINAL DASHBOARD:

Graphical user interface

Description automatically generated

**BIBLIOGRAPHY:**

* *Kaggle*
* *Trending videos of YouTube analysis*
* [*https://www.kaggle.com/datasnaek/youtube-new*](https://www.kaggle.com/datasnaek/youtube-new)
* [*https://www.analyticsvidhya.com/blog/2019/09/7-data-science- projects-github- showcase-your-*](https://www.analyticsvidhya.com/blog/2019/09/7-data-science-%20%20%20%20projects-github-%20%20%20%20showcase-your-)[*skills/*](https://www.analyticsvidhya.com/blog/2019/09/7-data-science-projects-github-showcase-your-skills/)
* [*https://en.wikipedia.org/wiki/YouTube*](https://en.wikipedia.org/wiki/YouTube)
* [*https://www.quora.com/How-does-Youtubes-algorithm-works-in-terms-of-trending-a-video*](https://www.quora.com/How-does-Youtubes-algorithm-works-in-terms-of-trending-a-video)