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:

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

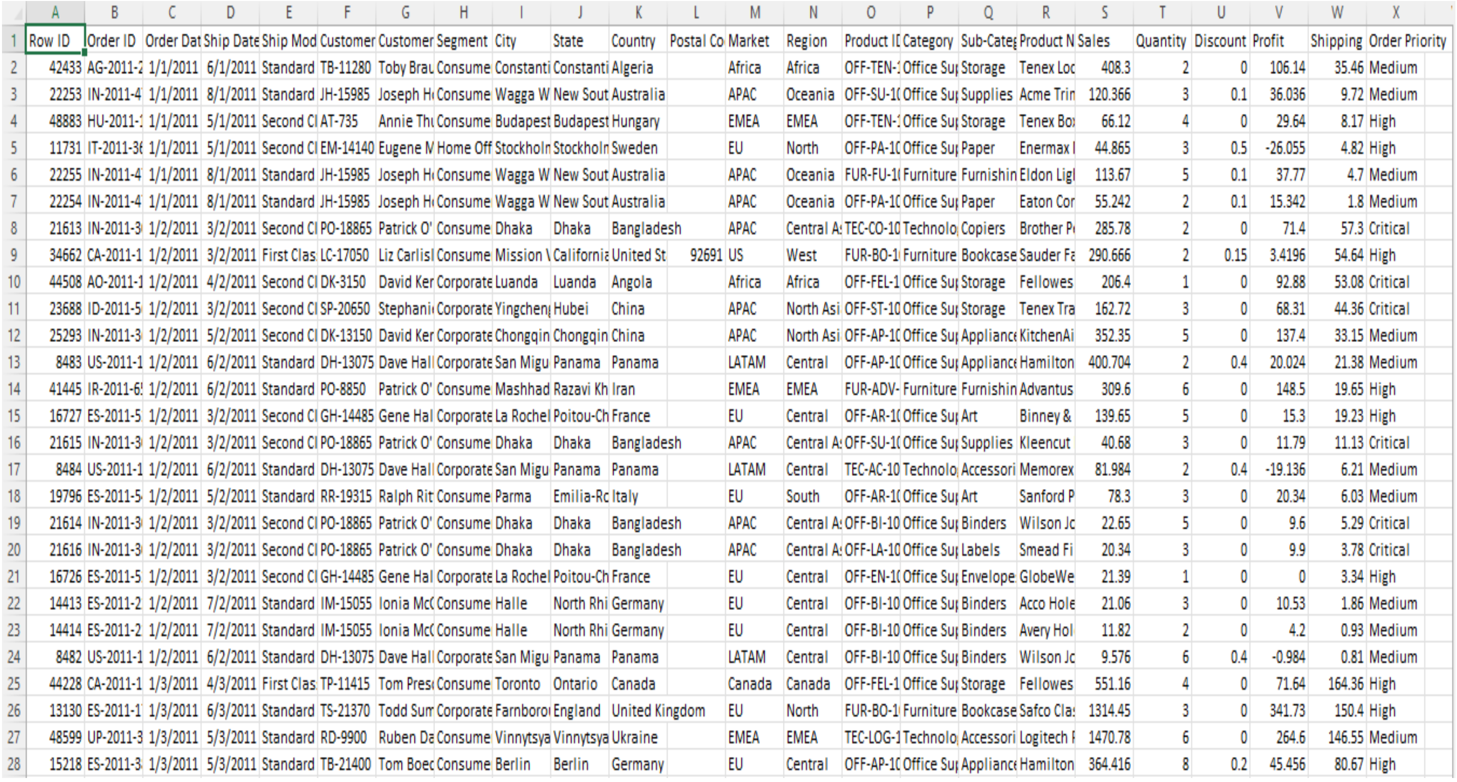
Kaggle is an online community for data scientists and machine learners, developed by Google. Kaggle allows users to find and publish data sets, explore, and build models in a web- based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to 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



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	Row ID	Order ID	Order Dat	Ship Date	Ship Mod	Customer	Customer	Segment	City	State	Country	Postal Co	Market	Region	Product ID	Category	Sub-Categ	Product N	Sales	Quantity	Discount	Profit	Shipping	Order Priority
2	42433	AG-2011-2	1/1/2011	6/1/2011	Standard	TB-11280	Toby Brau	Consume	Constanti	Constanti	Algeria		Africa	Africa	OFF-TEN-1	Office Sup	Storage	Tenex Loc	408.3	2	0	106.14	35.46	Medium
3	22253	IN-2011-4	1/1/2011	8/1/2011	Standard	JH-15985	Joseph H	Consume	Wagga W	New Sout	Australia		APAC	Oceania	OFF-SU-10	Office Sup	Supplies	Acme Trin	120.366	3	0.1	36.036	9.72	Medium
4	48883	HU-2011-1	1/1/2011	5/1/2011	Second Cl	AT-735	Annie Thu	Consume	Budapest	Budapest	Hungary		EMEA	EMEA	OFF-TEN-1	Office Sup	Storage	Tenex Box	66.12	4	0	29.64	8.17	High
5	11731	IT-2011-36	1/1/2011	5/1/2011	Second Cl	EM-14140	Eugene M	Home Off	Stockholn	Stockholn	Sweden		EU	North	OFF-PA-10	Office Sup	Paper	Enermax I	44.865	3	0.5	-26.055	4.82	High
6	22255	IN-2011-4	1/1/2011	8/1/2011	Standard	JH-15985	Joseph H	Consume	Wagga W	New Sout	Australia		APAC	Oceania	FUR-FU-10	Furniture	Furnishin	Eldon Ligl	113.67	5	0.1	37.77	4.7	Medium
7	22254	IN-2011-4	1/1/2011	8/1/2011	Standard	JH-15985	Joseph H	Consume	Wagga W	New Sout	Australia		APAC	Oceania	OFF-PA-10	Office Sup	Paper	Eaton Cor	55.242	2	0.1	15.342	1.8	Medium
8	21613	IN-2011-3	1/2/2011	3/2/2011	Second Cl	PO-18865	Patrick O'	Consume	Dhaka	Dhaka	Bangladesh		APAC	Central As	TEC-CO-10	Technolo	Copiers	Brother P	285.78	2	0	71.4	57.3	Critical
9	34662	CA-2011-1	1/2/2011	3/2/2011	First Clas	LC-17050	Liz Carlisl	Consume	Mission V	California	United St	92691	US	West	FUR-BO-10	Furniture	Bookcase	Sauder Fa	290.666	2	0.15	3.4196	54.64	High
10	44508	AO-2011-1	1/2/2011	4/2/2011	Second Cl	DK-3150	David Ker	Corporate	Luanda	Luanda	Angola		Africa	Africa	OFF-FEL-1	Office Sup	Storage	Fellowes	206.4	1	0	92.88	53.08	Critical
11	23688	ID-2011-5	1/2/2011	3/2/2011	Second Cl	SP-20650	Stephanie	Corporate	Yingcheng	Hubei	China		APAC	North Asi	OFF-ST-10	Office Sup	Storage	Tenex Tra	162.72	3	0	68.31	44.36	Critical
12	25293	IN-2011-3	1/2/2011	5/2/2011	Second Cl	DK-13150	David Ker	Corporate	Chongqin	Chongqin	China		APAC	North Asi	OFF-AP-10	Office Sup	Appliance	KitchenAi	352.35	5	0	137.4	33.15	Medium
13	8483	US-2011-1	1/2/2011	6/2/2011	Standard	DH-13075	Dave Hall	Corporate	San Migu	Panama	Panama		LATAM	Central	OFF-AP-10	Office Sup	Appliance	Hamilton	400.704	2	0.4	20.024	21.38	Medium
14	41445	IR-2011-65	1/2/2011	6/2/2011	Standard	PO-8850	Patrick O'	Consume	Mashhad	Razavi Kh	Iran		EMEA	EMEA	FUR-ADV-	Furniture	Furnishin	Advantus	309.6	6	0	148.5	19.65	High
15	16727	ES-2011-5	1/2/2011	3/2/2011	Second Cl	GH-14485	Gene Hal	Corporate	La Rochel	Poitou-Ch	France		EU	Central	OFF-AR-10	Office Sup	Art	Binney &	139.65	5	0	15.3	19.23	High
16	21615	IN-2011-3	1/2/2011	3/2/2011	Second Cl	PO-18865	Patrick O'	Consume	Dhaka	Dhaka	Bangladesh		APAC	Central As	OFF-SU-10	Office Sup	Supplies	Kleencut	40.68	3	0	11.79	11.13	Critical
17	8484	US-2011-1	1/2/2011	6/2/2011	Standard	DH-13075	Dave Hall	Corporate	San Migu	Panama	Panama		LATAM	Central	TEC-AC-10	Technolo	Accessori	Memorex	81.984	2	0.4	-19.136	6.21	Medium
18	19796	ES-2011-5	1/2/2011	5/2/2011	Standard	RR-19315	Ralph Rit	Consume	Parma	Emilia-Ro	Italy		EU	South	OFF-AR-10	Office Sup	Art	Sanford P	78.3	3	0	20.34	6.03	Medium
19	21614	IN-2011-3	1/2/2011	3/2/2011	Second Cl	PO-18865	Patrick O'	Consume	Dhaka	Dhaka	Bangladesh		APAC	Central As	OFF-BI-10	Office Sup	Binders	Wilson Jo	22.65	5	0	9.6	5.29	Critical
20	21616	IN-2011-3	1/2/2011	3/2/2011	Second Cl	PO-18865	Patrick O'	Consume	Dhaka	Dhaka	Bangladesh		APAC	Central As	OFF-LA-10	Office Sup	Labels	Smead Fi	20.34	3	0	9.9	3.78	Critical
21	16726	ES-2011-5	1/2/2011	3/2/2011	Second Cl	GH-14485	Gene Hal	Corporate	La Rochel	Poitou-Ch	France		EU	Central	OFF-EN-10	Office Sup	Envelope	GlobeWe	21.39	1	0	0	3.34	High
22	14413	ES-2011-2	1/2/2011	7/2/2011	Standard	IM-15055	Ionia McG	Consume	Halle	North Rhi	Germany		EU	Central	OFF-BI-10	Office Sup	Binders	Acco Hole	21.06	3	0	10.53	1.86	Medium
23	14414	ES-2011-2	1/2/2011	7/2/2011	Standard	IM-15055	Ionia McG	Consume	Halle	North Rhi	Germany		EU	Central	OFF-BI-10	Office Sup	Binders	Avery Hol	11.82	2	0	4.2	0.93	Medium
24	8482	US-2011-1	1/2/2011	6/2/2011	Standard	DH-13075	Dave Hall	Corporate	San Migu	Panama	Panama		LATAM	Central	OFF-BI-10	Office Sup	Binders	Wilson Jo	9.576	6	0.4	-0.984	0.81	Medium
25	44228	CA-2011-1	1/3/2011	4/3/2011	First Clas	TP-11415	Tom Pres	Consume	Toronto	Ontario	Canada		Canada	Canada	OFF-FEL-1	Office Sup	Storage	Fellowes	551.16	4	0	71.64	164.36	High
26	13130	ES-2011-1	1/3/2011	6/3/2011	Standard	TS-21370	Todd Sum	Corporate	Farnboro	England	United Kingdom		EU	North	FUR-BO-10	Furniture	Bookcase	Safco Clas	1314.45	3	0	341.73	150.4	High
27	48599	UP-2011-3	1/3/2011	5/3/2011	Standard	RD-9900	Ruben Da	Consume	Vinnytsya	Vinnytsya	Ukraine		EMEA	EMEA	TEC-LOG-1	Technolo	Accessori	Logitech P	1470.78	6	0	264.6	146.55	Medium
28	15218	ES-2011-3	1/3/2011	5/3/2011	Standard	TB-21400	Tom Boec	Consume	Berlin	Berlin	Germany		EU	Central	OFF-AP-10	Office Sup	Appliance	Hamilton	364.416	8	0.2	45.456	80.67	High

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.

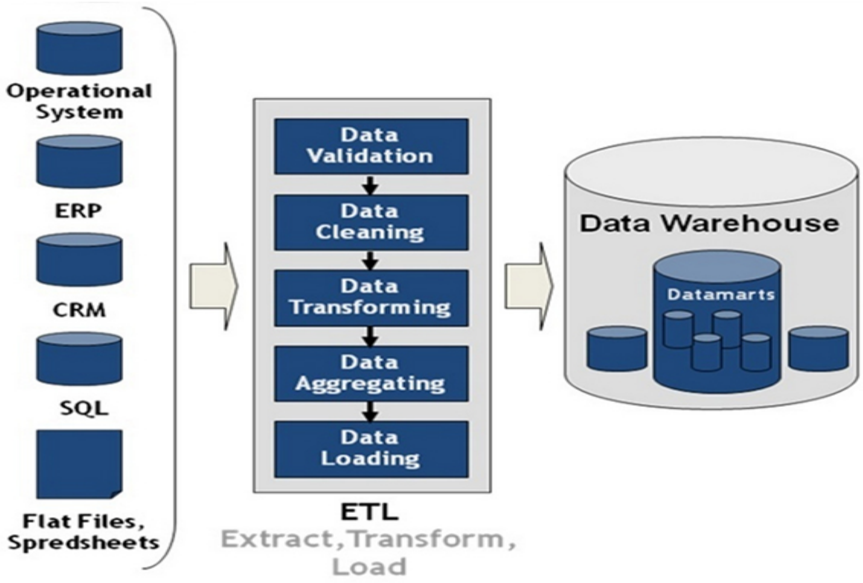
It's tempting to think a creating a Data warehouse is simply extracting data from multiple sources and loading into database of a Data warehouse. This is far from the truth and requires a complex ETL process. The ETL process requires active inputs from various stakeholders including developers, analysts,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

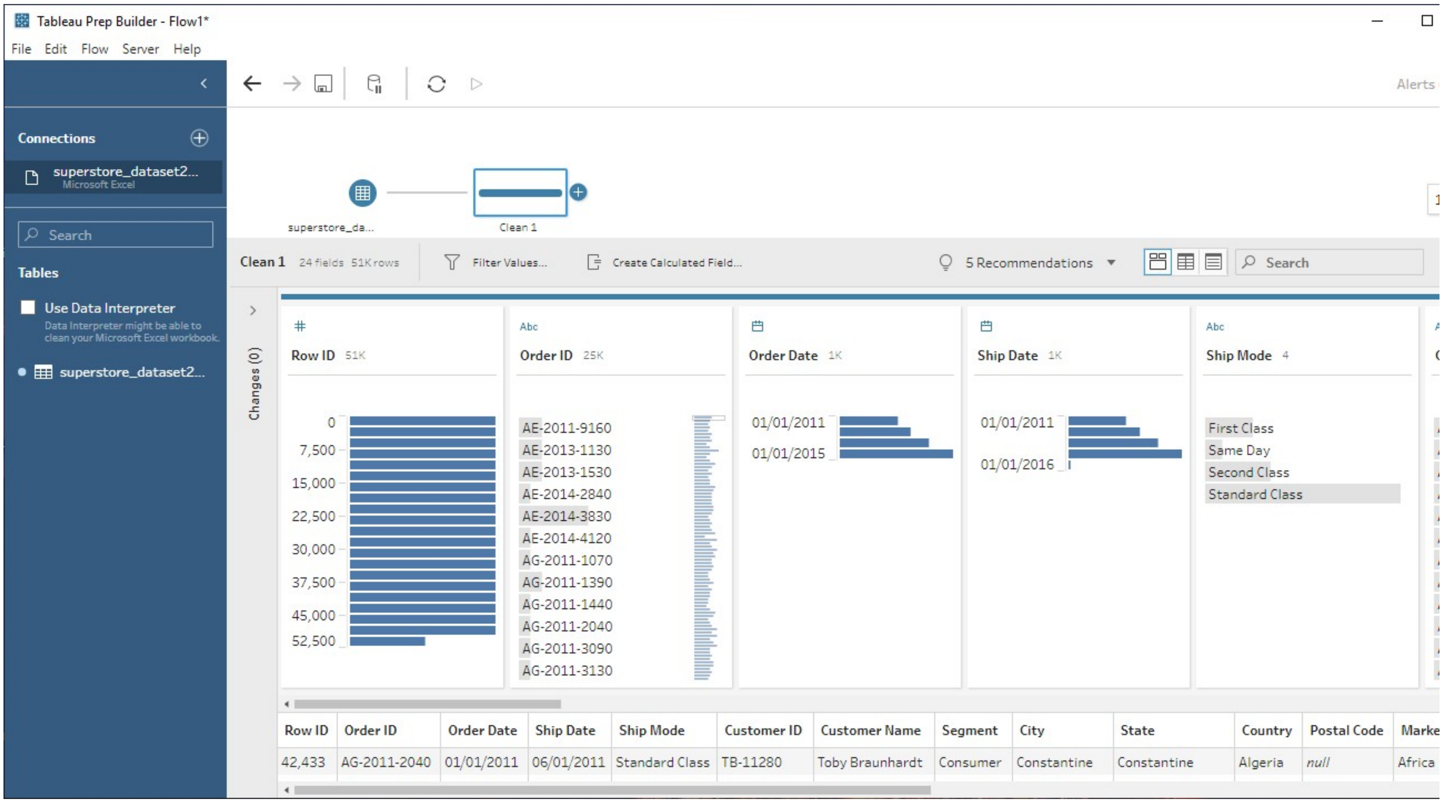


ETL Process used in Project

Extraction

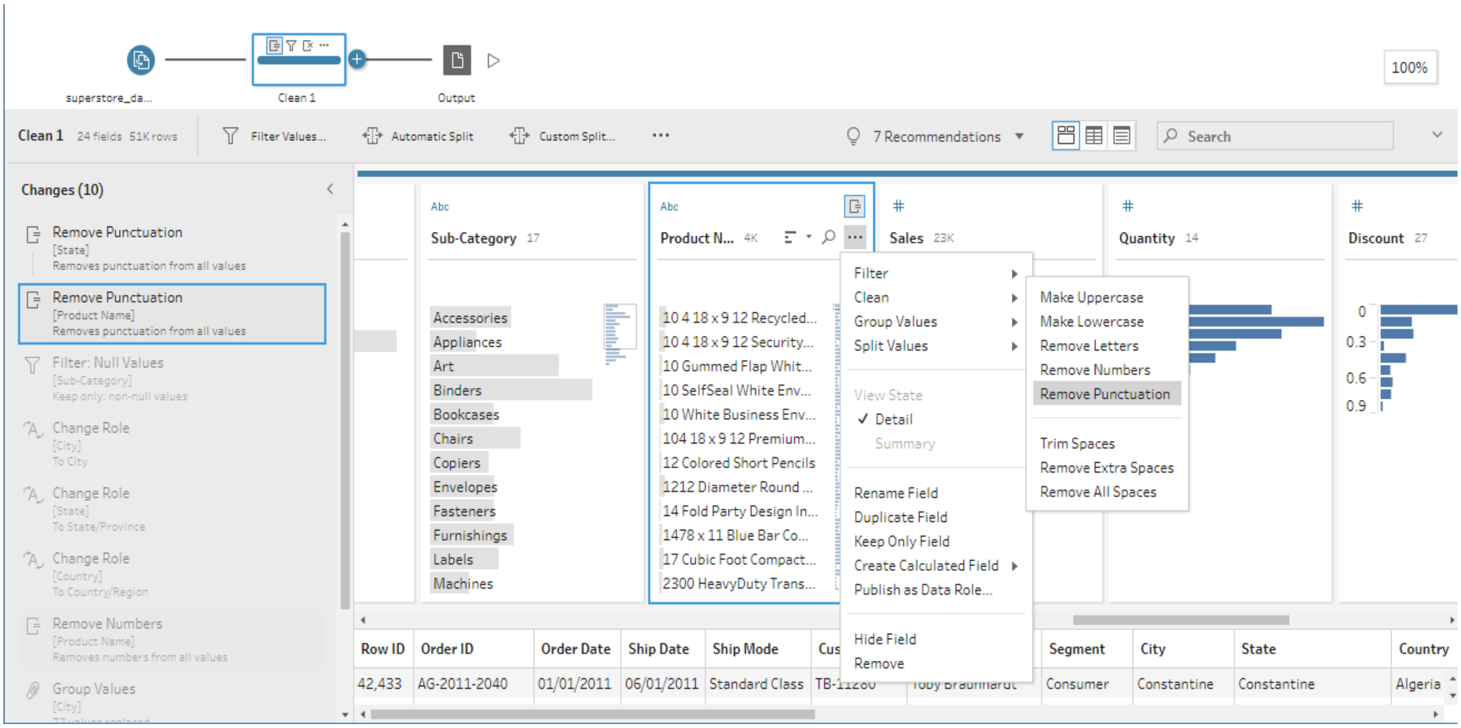
Extracting the dataset from PC to Tableau for removing the unwanted characters, fields,spelling errors etc.

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



Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State	Country	Postal Code	Marke
42,433	AG-2011-2040	01/01/2011	06/01/2011	Standard Class	TB-11280	Toby Braunhardt	Consumer	Constantine	Constantine	Algeria	*null*	Africa

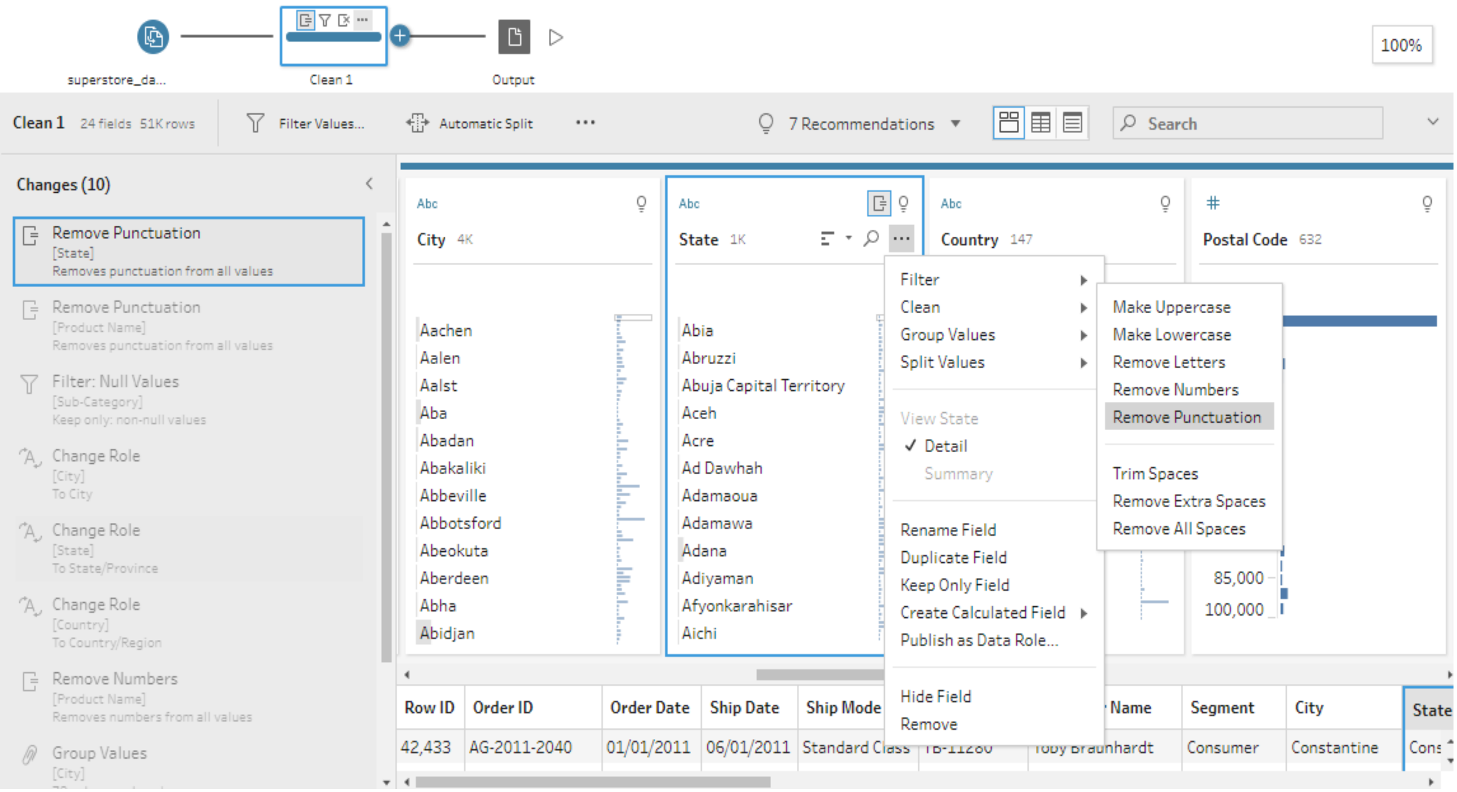
Step 2: *Removal of punctuations in Product Name and state column*



Changes (10)

- Remove Punctuation
  [State]
  Removes punctuation from all values
- Remove Punctuation
  [Product Name]
  Removes punctuation from all values
- Filter: Null Values
  [Sub-Category]
  Keep only: non-null values
- Change Role
  [City]
  To City
- Change Role
  [State]
  To State/Province
- Change Role
  [Country]
  To Country/Region
- Remove Numbers
  [Product Name]
  Removes numbers from all values
- Group Values
  [City]

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Cus		Segment	City	State	Country
42,433	AG-2011-2040	01/01/2011	06/01/2011	Standard Class	TB-11280	Toby Braunhardt	Consumer	Constantine	Constantine	Algeria

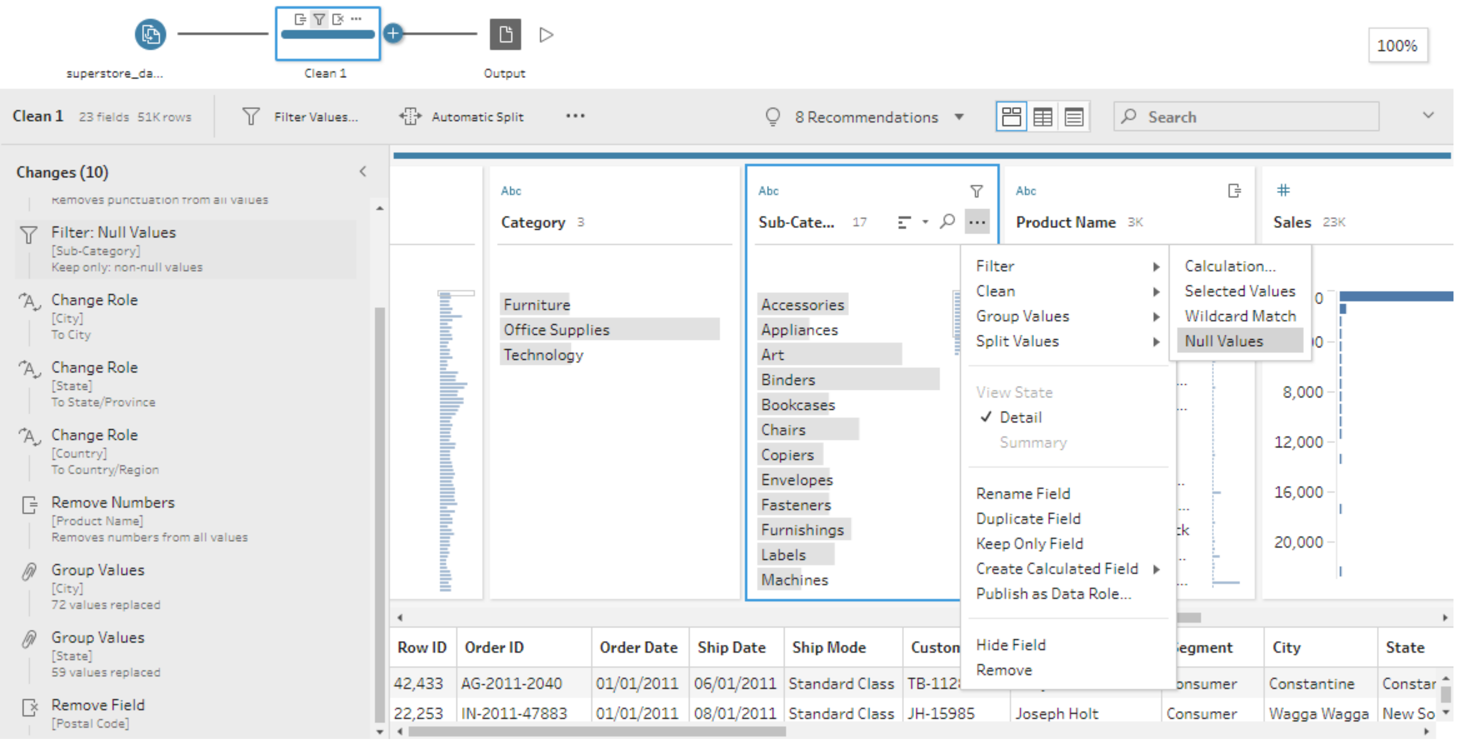


Changes (10)

- Remove Punctuation
  [State]
  Removes punctuation from all values
- Remove Punctuation
  [Product Name]
  Removes punctuation from all values
- Filter: Null Values
  [Sub-Category]
  Keep only: non-null values
- Change Role
  [City]
  To City
- Change Role
  [State]
  To State/Province
- Change Role
  [Country]
  To Country/Region
- Remove Numbers
  [Product Name]
  Removes numbers from all values
- Group Values
  [City]

Row ID	Order ID	Order Date	Ship Date	Ship Mode		Name	Segment	City	State
42,433	AG-2011-2040	01/01/2011	06/01/2011	Standard Class	TB-11280	Toby Braunhardt	Consumer	Constantine	Cons

Step 3: *Removal of Null values*



Changes (10)

Removes punctuation from all values

Filter: Null Values
[Sub-Category]
Keep only: non-null values

Change Role
[City]
To City

Change Role
[State]
To State/Province

Change Role
[Country]
To Country/Region

Remove Numbers
[Product Name]
Removes numbers from all values

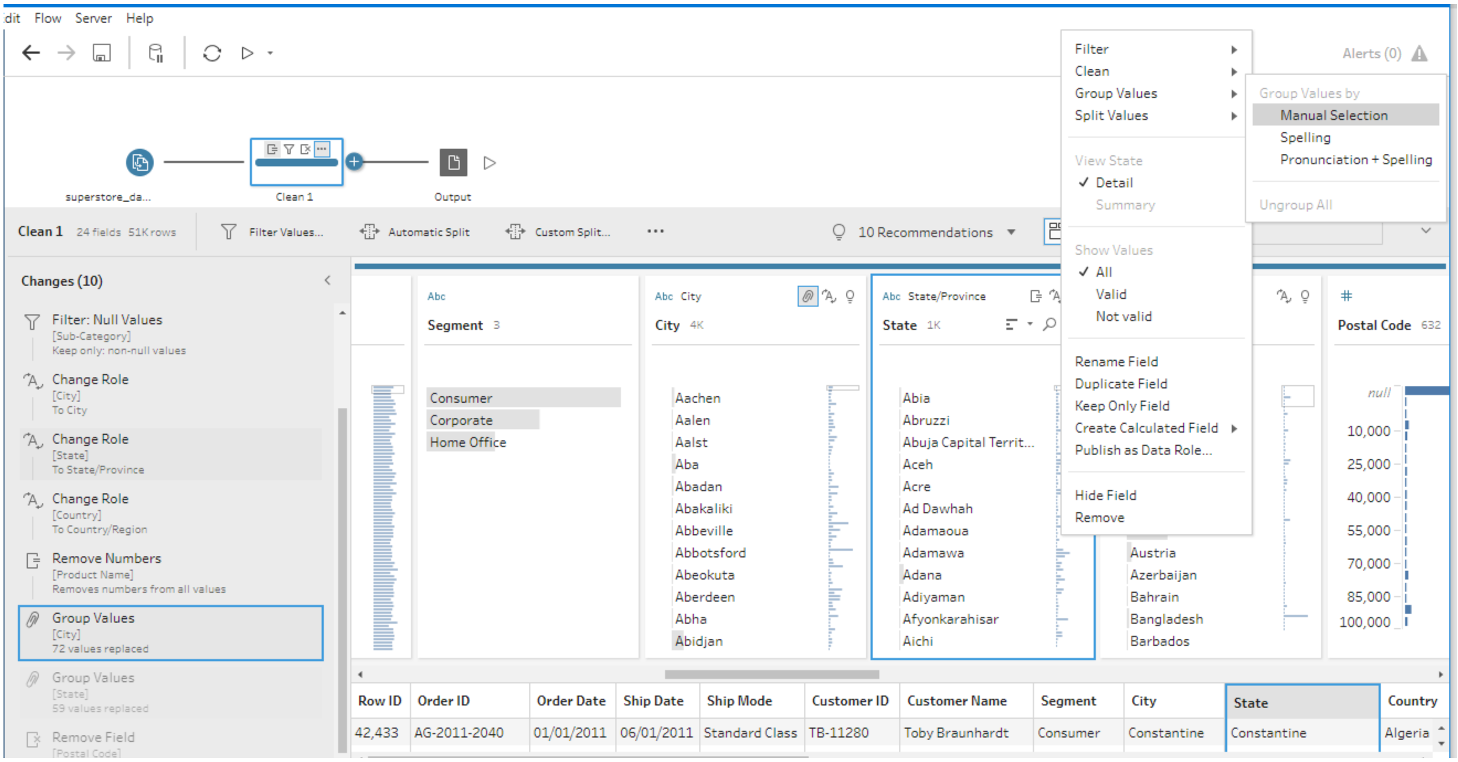
Group Values
[City]
72 values replaced

Group Values
[State]
59 values replaced

Remove Field
[Postal Code]

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Custon		egment	City	State
42,433	AG-2011-2040	01/01/2011	06/01/2011	Standard Class	TB-112		onsumer	Constantine	Constar
22,253	IN-2011-47883	01/01/2011	08/01/2011	Standard Class	JH-15985	Joseph Holt	Consumer	Wagga Wagga	New So

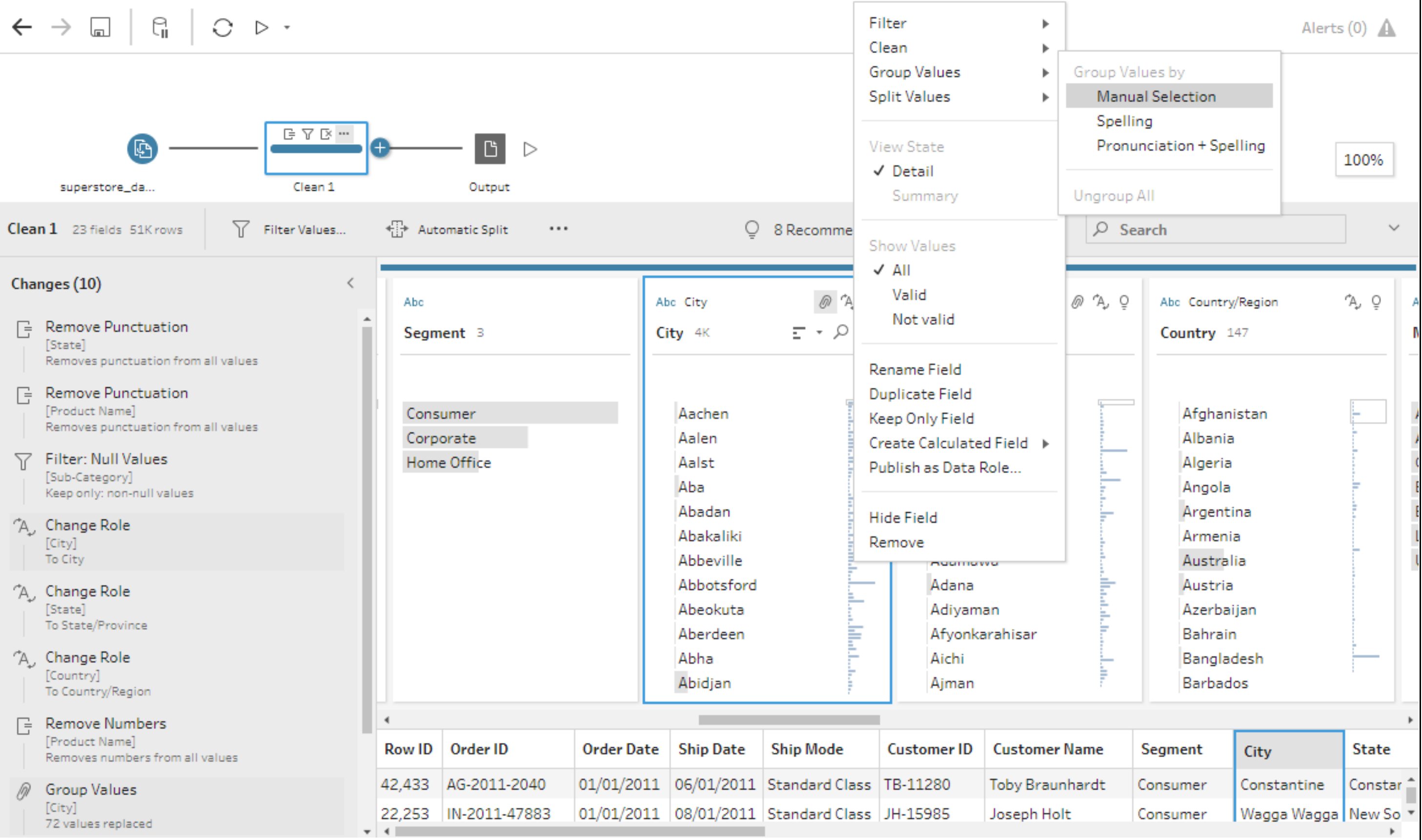
Step 4: *Grouping values in "State" and "City "columns*



Changes (10)

- Filter: Null Values [Sub-Category] Keep only: non-null values
- Change Role [City] To City
- Change Role [State] To State/Province
- Change Role [Country] To Country/Region
- Remove Numbers [Product Name] Removes numbers from all values
- Group Values [City] 72 values replaced
- Group Values [State] 59 values replaced
- Remove Field [Postal Code]

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State	Country
42,433	AG-2011-2040	01/01/2011	06/01/2011	Standard Class	TB-11280	Toby Braunhardt	Consumer	Constantine	Constantine	Algeria

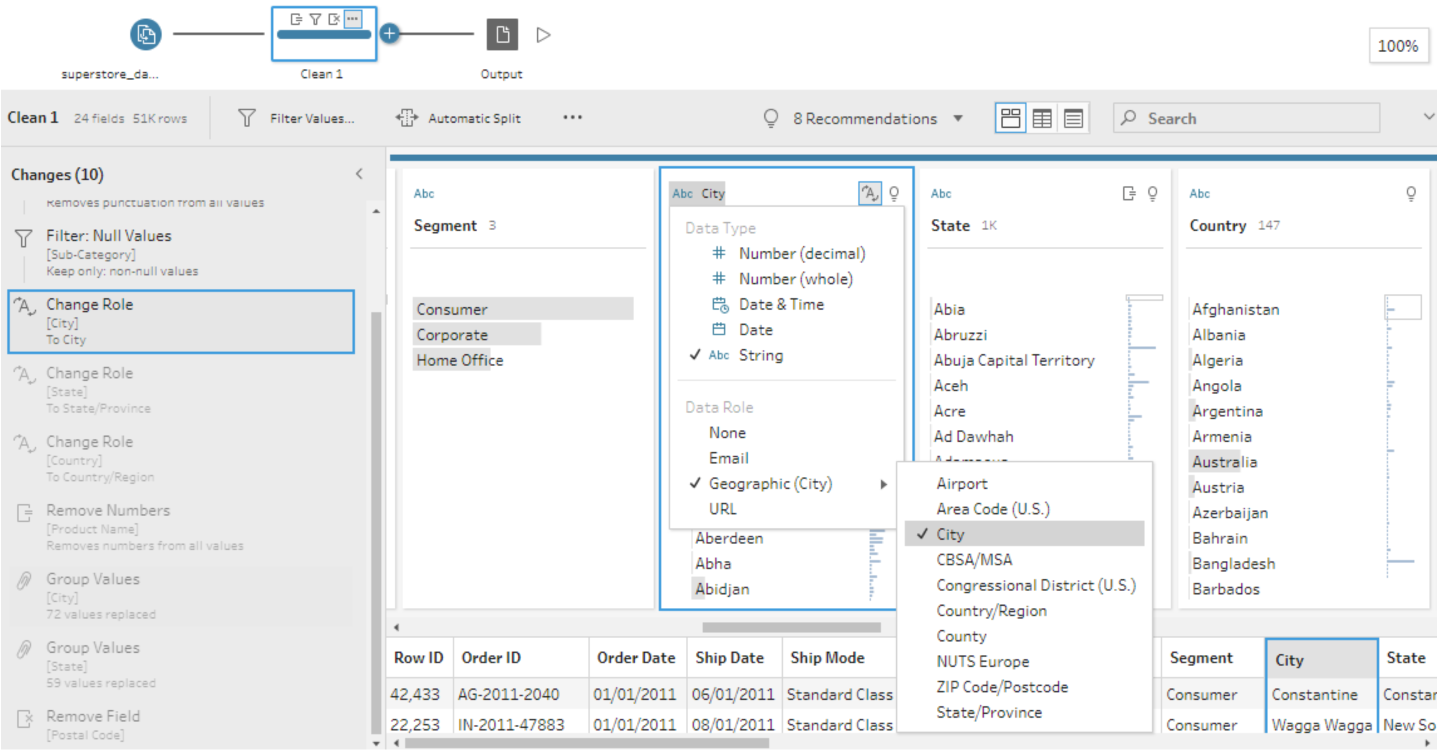


Changes (10)

- Remove Punctuation [State] Removes punctuation from all values
- Remove Punctuation [Product Name] Removes punctuation from all values
- Filter: Null Values [Sub-Category] Keep only: non-null values
- Change Role [City] To City
- Change Role [State] To State/Province
- Change Role [Country] To Country/Region
- Remove Numbers [Product Name] Removes numbers from all values
- Group Values [City] 72 values replaced

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State
42,433	AG-2011-2040	01/01/2011	06/01/2011	Standard Class	TB-11280	Toby Braunhardt	Consumer	Constantine	Constar
22,253	IN-2011-47883	01/01/2011	08/01/2011	Standard Class	JH-15985	Joseph Holt	Consumer	Wagga Wagga	New So

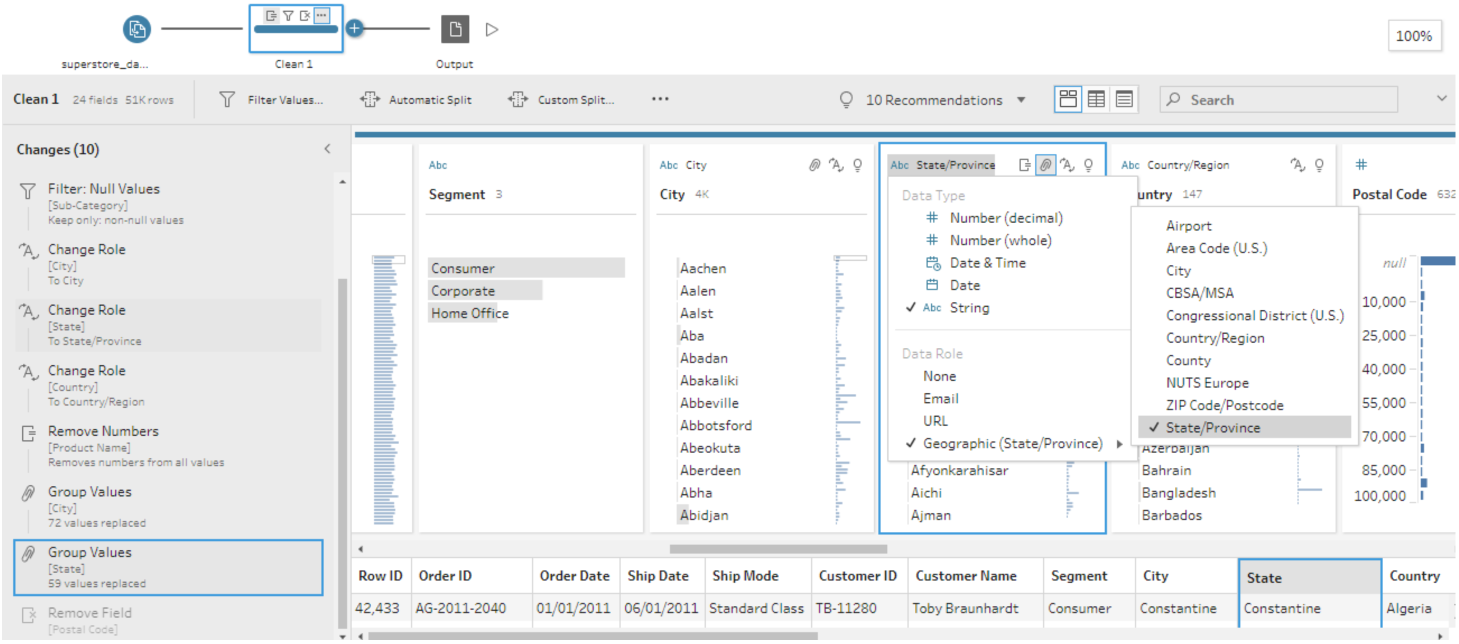
Step 5: *Change data type of city, state, and country*



Changes (10)

- Removes punctuation from all values
- Filter: Null Values — [Sub-Category] — Keep only: non-null values
- Change Role — [City] — To City
- Change Role — [State] — To State/Province
- Change Role — [Country] — To Country/Region
- Remove Numbers — [Product Name] — Removes numbers from all values
- Group Values — [City] — 72 values replaced
- Group Values — [State] — 59 values replaced
- Remove Field — [Postal Code]

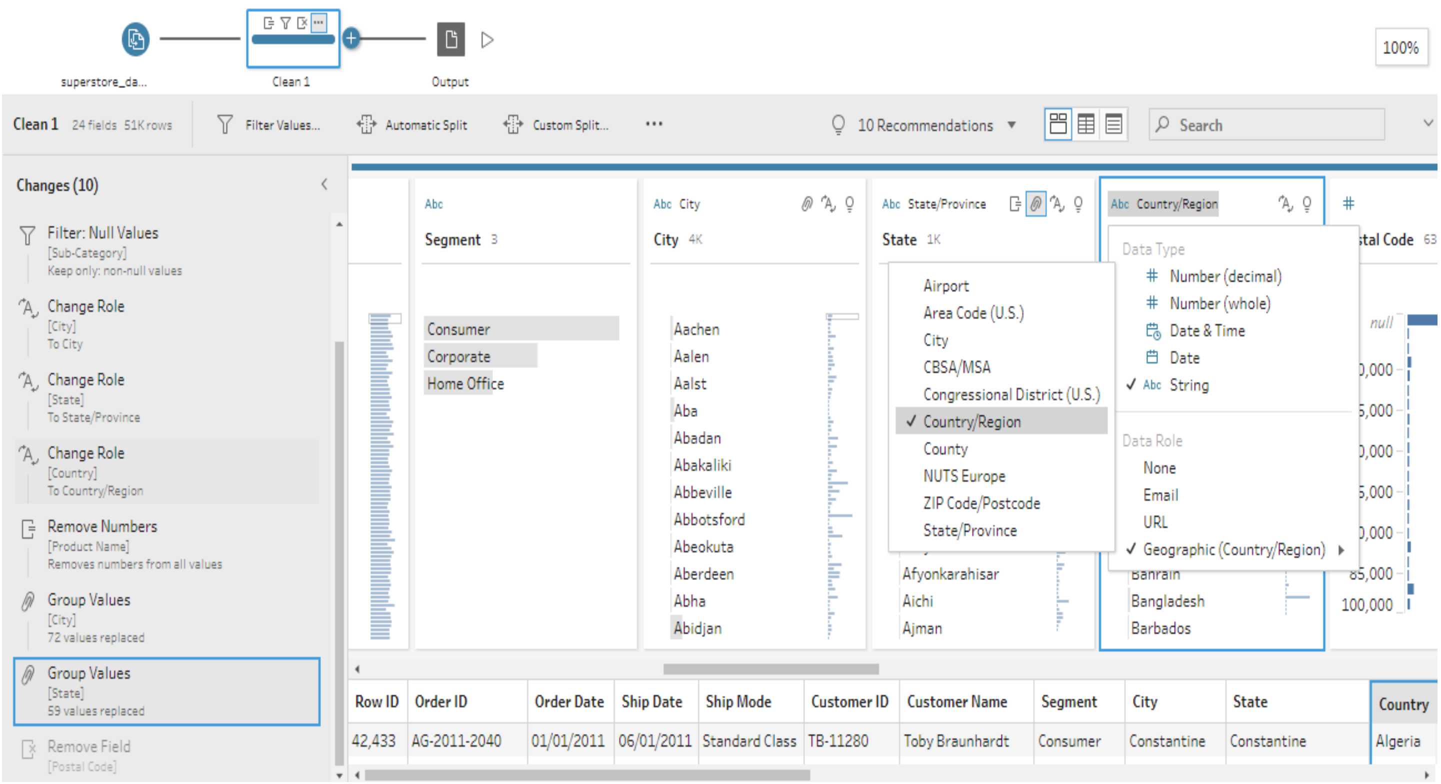
Row ID	Order ID	Order Date	Ship Date	Ship Mode	Segment	City	State
42,433	AG-2011-2040	01/01/2011	06/01/2011	Standard Class	Consumer	Constantine	Constar
22,253	IN-2011-47883	01/01/2011	08/01/2011	Standard Class	Consumer	Wagga Wagga	New So



Changes (10)

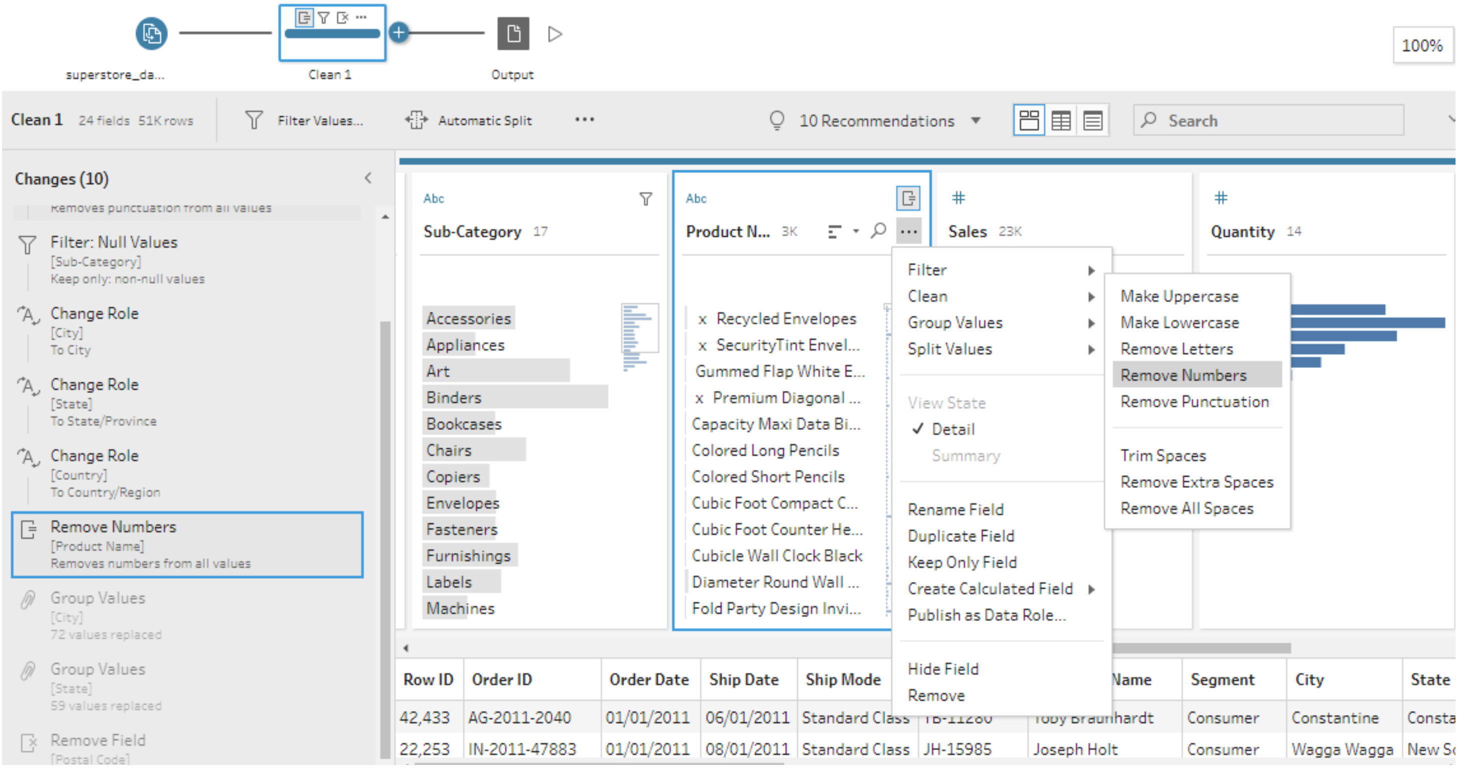
- Filter: Null Values — [Sub-Category] — Keep only: non-null values
- Change Role — [City] — To City
- Change Role — [State] — To State/Province
- Change Role — [Country] — To Country/Region
- Remove Numbers — [Product Name] — Removes numbers from all values
- Group Values — [City] — 72 values replaced
- Group Values — [State] — 59 values replaced
- Remove Field — [Postal Code]

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State	Country
42,433	AG-2011-2040	01/01/2011	06/01/2011	Standard Class	TB-11280	Toby Braunhardt	Consumer	Constantine	Constantine	Algeria



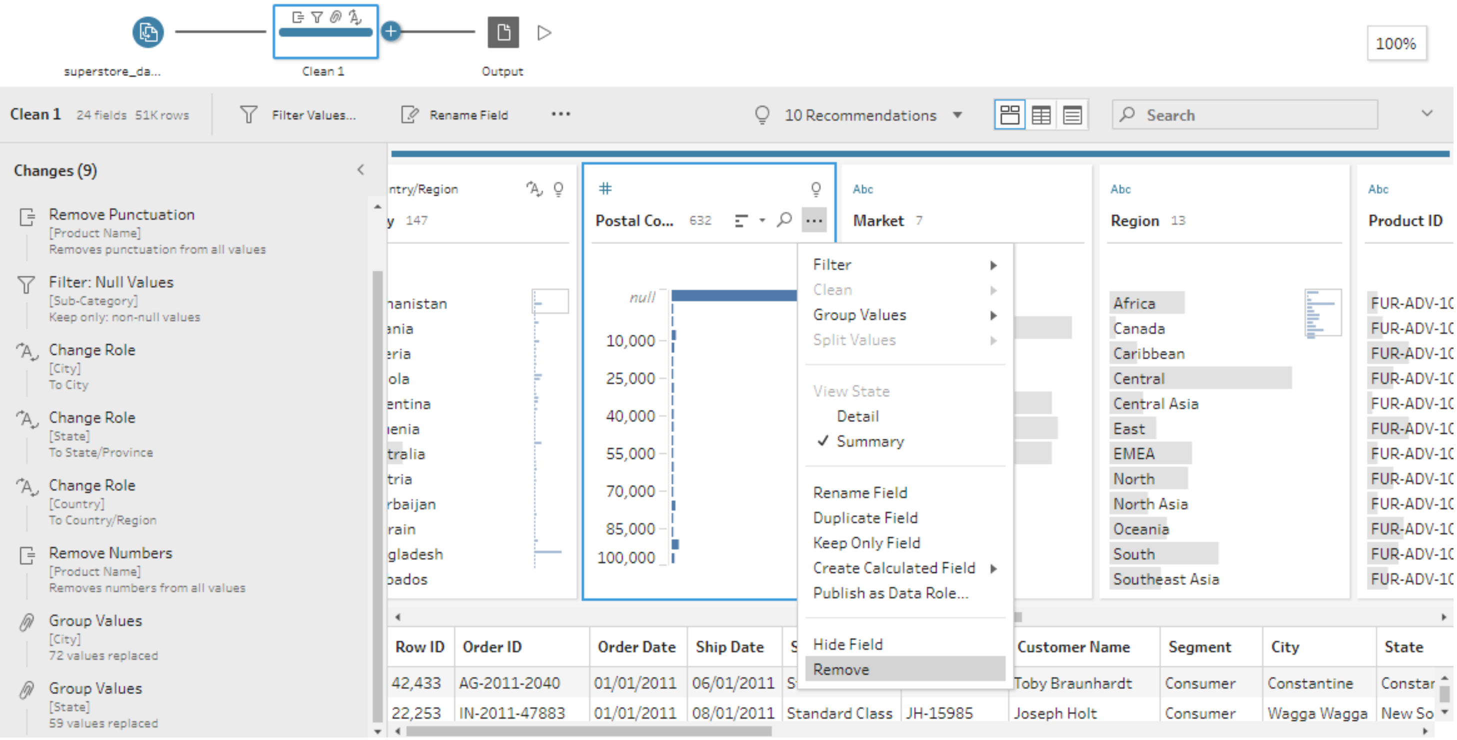
Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State	Country
42,433	AG-2011-2040	01/01/2011	06/01/2011	Standard Class	TB-11280	Toby Braunhardt	Consumer	Constantine	Constantine	Algeria

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



Row ID	Order ID	Order Date	Ship Date	Ship Mode			Name	Segment	City	State
42,433	AG-2011-2040	01/01/2011	06/01/2011	Standard Class	TB-11280	Toby Braunhardt		Consumer	Constantine	Consta
22,253	IN-2011-47883	01/01/2011	08/01/2011	Standard Class	JH-15985	Joseph Holt		Consumer	Wagga Wagga	New S

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



Changes (9)

Remove Punctuation
[Product Name]
Removes punctuation from all values

Filter: Null Values
[Sub-Category]
Keep only: non-null values

Change Role
[City]
To City

Change Role
[State]
To State/Province

Change Role
[Country]
To Country/Region

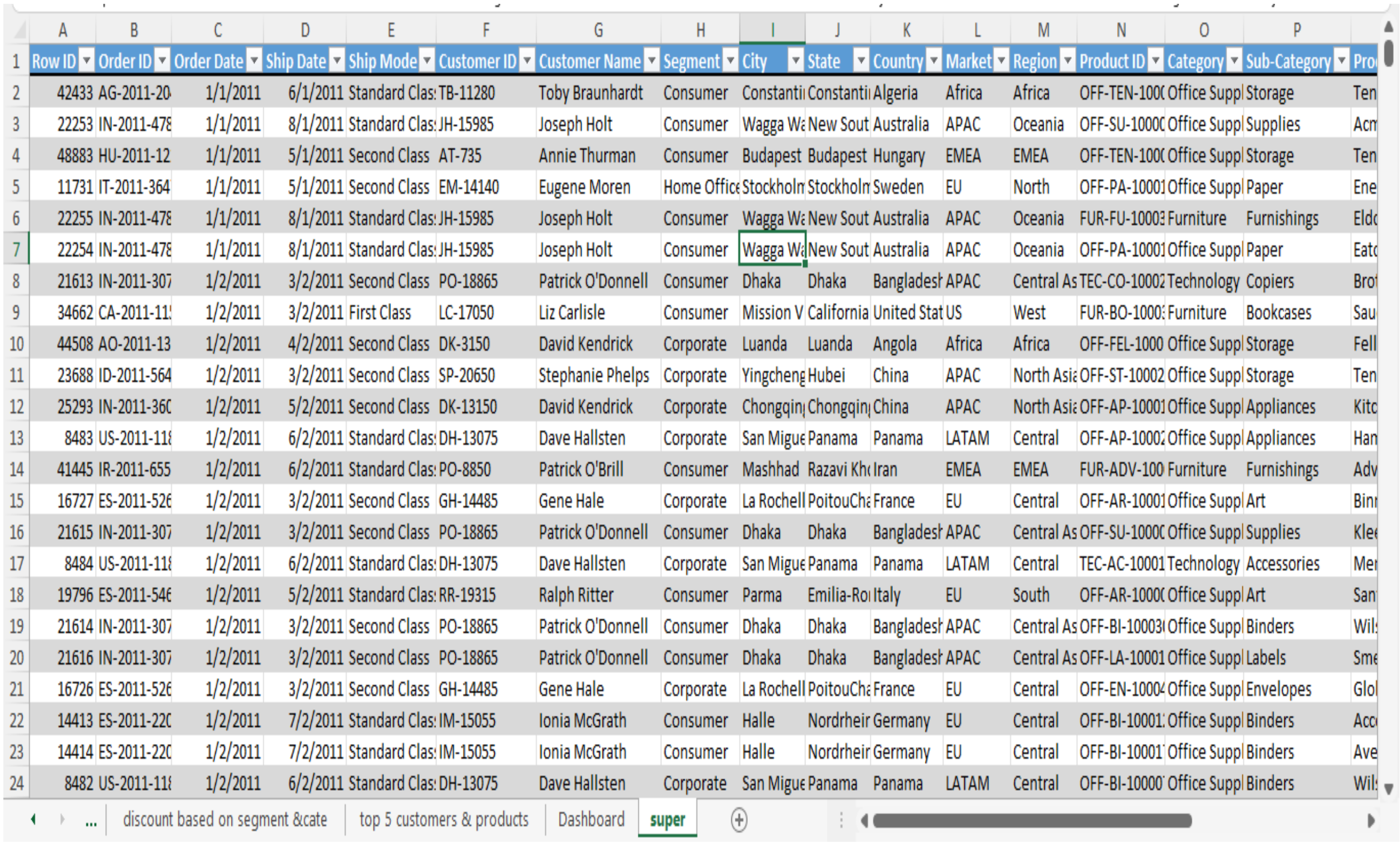
Remove Numbers
[Product Name]
Removes numbers from all values

Group Values
[City]
72 values replaced

Group Values
[State]
59 values replaced

Row ID	Order ID	Order Date	Ship Date	S		Customer Name	Segment	City	State
42,433	AG-2011-2040	01/01/2011	06/01/2011	S		Toby Braunhardt	Consumer	Constantine	Constar
22,253	IN-2011-47883	01/01/2011	08/01/2011	Standard Class	JH-15985	Joseph Holt	Consumer	Wagga Wagga	New So

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



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State	Country	Market	Region	Product ID	Category	Sub-Category	Pro
2	42433	AG-2011-20	1/1/2011	6/1/2011	Standard Clas	TB-11280	Toby Braunhardt	Consumer	Constanti	Constanti	Algeria	Africa	Africa	OFF-TEN-100	Office Suppl	Storage	Ten
3	22253	IN-2011-478	1/1/2011	8/1/2011	Standard Clas	JH-15985	Joseph Holt	Consumer	Wagga Wa	New Sout	Australia	APAC	Oceania	OFF-SU-1000	Office Suppl	Supplies	Acm
4	48883	HU-2011-12	1/1/2011	5/1/2011	Second Class	AT-735	Annie Thurman	Consumer	Budapest	Budapest	Hungary	EMEA	EMEA	OFF-TEN-100	Office Suppl	Storage	Ten
5	11731	IT-2011-364	1/1/2011	5/1/2011	Second Class	EM-14140	Eugene Moren	Home Office	Stockholm	Stockholm	Sweden	EU	North	OFF-PA-10001	Office Suppl	Paper	Ene
6	22255	IN-2011-478	1/1/2011	8/1/2011	Standard Clas	JH-15985	Joseph Holt	Consumer	Wagga Wa	New Sout	Australia	APAC	Oceania	FUR-FU-10003	Furniture	Furnishings	Eldo
7	22254	IN-2011-478	1/1/2011	8/1/2011	Standard Clas	JH-15985	Joseph Holt	Consumer	Wagga Wa	New Sout	Australia	APAC	Oceania	OFF-PA-10001	Office Suppl	Paper	Eato
8	21613	IN-2011-307	1/2/2011	3/2/2011	Second Class	PO-18865	Patrick O'Donnell	Consumer	Dhaka	Dhaka	Banglades	APAC	Central As	TEC-CO-10002	Technology	Copiers	Bro
9	34662	CA-2011-11	1/2/2011	3/2/2011	First Class	LC-17050	Liz Carlisle	Consumer	Mission V	California	United Stat	US	West	FUR-BO-1000	Furniture	Bookcases	Sau
10	44508	AO-2011-13	1/2/2011	4/2/2011	Second Class	DK-3150	David Kendrick	Corporate	Luanda	Luanda	Angola	Africa	Africa	OFF-FEL-1000	Office Suppl	Storage	Fell
11	23688	ID-2011-564	1/2/2011	3/2/2011	Second Class	SP-20650	Stephanie Phelps	Corporate	Yingcheng	Hubei	China	APAC	North Asi	OFF-ST-10002	Office Suppl	Storage	Ten
12	25293	IN-2011-360	1/2/2011	5/2/2011	Second Class	DK-13150	David Kendrick	Corporate	Chongqin	Chongqin	China	APAC	North Asi	OFF-AP-10001	Office Suppl	Appliances	Kitc
13	8483	US-2011-11	1/2/2011	6/2/2011	Standard Clas	DH-13075	Dave Hallsten	Corporate	San Migue	Panama	Panama	LATAM	Central	OFF-AP-10002	Office Suppl	Appliances	Han
14	41445	IR-2011-655	1/2/2011	6/2/2011	Standard Clas	PO-8850	Patrick O'Brill	Consumer	Mashhad	Razavi Kh	Iran	EMEA	EMEA	FUR-ADV-100	Furniture	Furnishings	Adv
15	16727	ES-2011-526	1/2/2011	3/2/2011	Second Class	GH-14485	Gene Hale	Corporate	La Rochell	PoitouCh	France	EU	Central	OFF-AR-10001	Office Suppl	Art	Bin
16	21615	IN-2011-307	1/2/2011	3/2/2011	Second Class	PO-18865	Patrick O'Donnell	Consumer	Dhaka	Dhaka	Banglades	APAC	Central As	OFF-SU-1000	Office Suppl	Supplies	Kle
17	8484	US-2011-11	1/2/2011	6/2/2011	Standard Clas	DH-13075	Dave Hallsten	Corporate	San Migue	Panama	Panama	LATAM	Central	TEC-AC-10001	Technology	Accessories	Me
18	19796	ES-2011-546	1/2/2011	5/2/2011	Standard Clas	RR-19315	Ralph Ritter	Consumer	Parma	Emilia-Ro	Italy	EU	South	OFF-AR-1000	Office Suppl	Art	San
19	21614	IN-2011-307	1/2/2011	3/2/2011	Second Class	PO-18865	Patrick O'Donnell	Consumer	Dhaka	Dhaka	Banglades	APAC	Central As	OFF-BI-10003	Office Suppl	Binders	Wil
20	21616	IN-2011-307	1/2/2011	3/2/2011	Second Class	PO-18865	Patrick O'Donnell	Consumer	Dhaka	Dhaka	Banglades	APAC	Central As	OFF-LA-10001	Office Suppl	Labels	Sme
21	16726	ES-2011-526	1/2/2011	3/2/2011	Second Class	GH-14485	Gene Hale	Corporate	La Rochell	PoitouCh	France	EU	Central	OFF-EN-1000	Office Suppl	Envelopes	Glo
22	14413	ES-2011-220	1/2/2011	7/2/2011	Standard Clas	IM-15055	Ionia McGrath	Consumer	Halle	Nordrhein	Germany	EU	Central	OFF-BI-10001	Office Suppl	Binders	Acc
23	14414	ES-2011-220	1/2/2011	7/2/2011	Standard Clas	IM-15055	Ionia McGrath	Consumer	Halle	Nordrhein	Germany	EU	Central	OFF-BI-10001	Office Suppl	Binders	Ave
24	8482	US-2011-11	1/2/2011	6/2/2011	Standard Clas	DH-13075	Dave Hallsten	Corporate	San Migue	Panama	Panama	LATAM	Central	OFF-BI-10000	Office Suppl	Binders	Wil

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.*

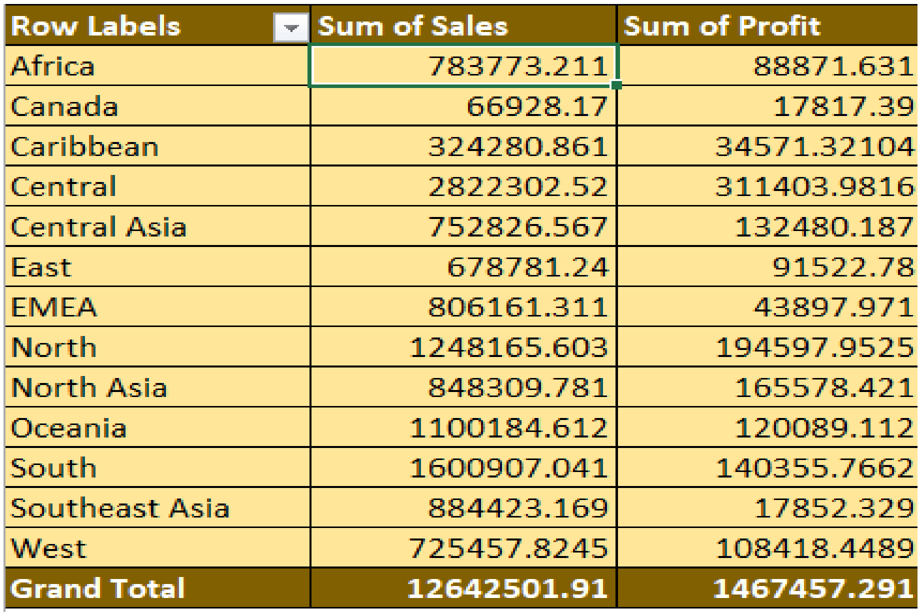
Description:

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:

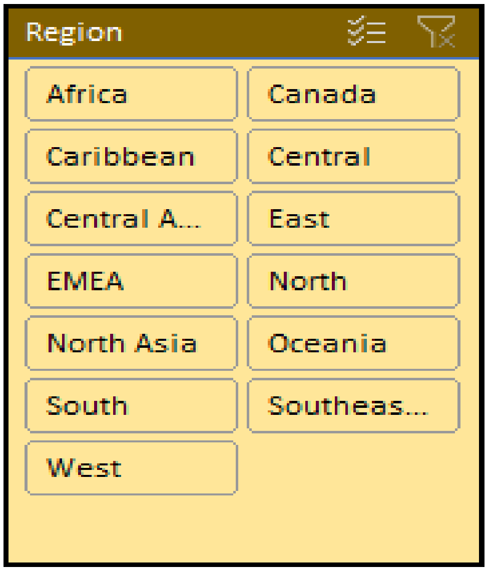
For Grand Total of sales and profit we can use sum function: =SUM()

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

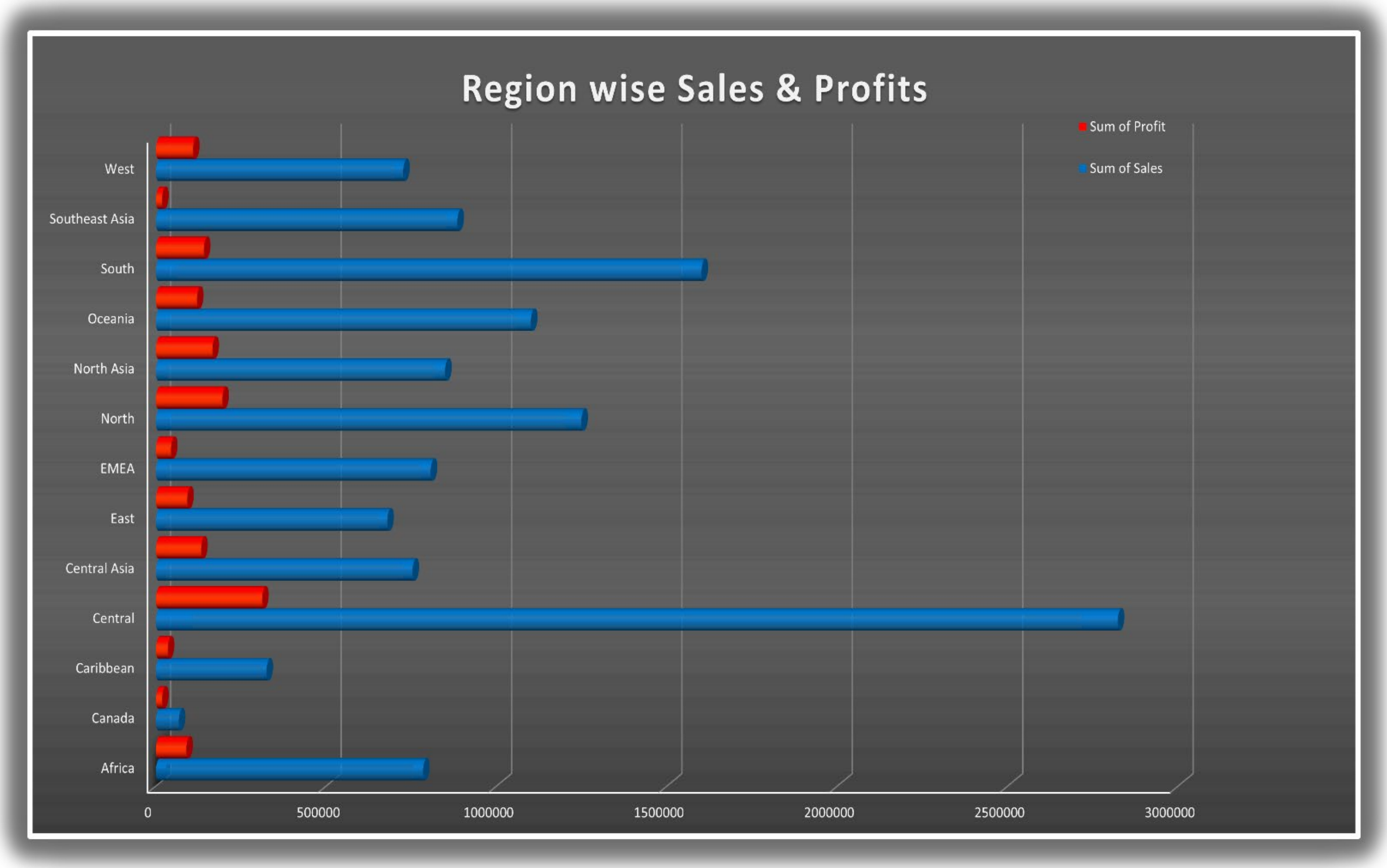


Row Labels	Sum of Sales	Sum of Profit
Africa	783773.211	88871.631
Canada	66928.17	17817.39
Caribbean	324280.861	34571.32104
Central	2822302.52	311403.9816
Central Asia	752826.567	132480.187
East	678781.24	91522.78
EMEA	806161.311	43897.971
North	1248165.603	194597.9525
North Asia	848309.781	165578.421
Oceania	1100184.612	120089.112
South	1600907.041	140355.7662
Southeast Asia	884423.169	17852.329
West	725457.8245	108418.4489
Grand Total	**12642501.91**	**1467457.291**

Slicer:



Visualization:



2. Discounts based on Category and Segment

Introduction

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

Description:

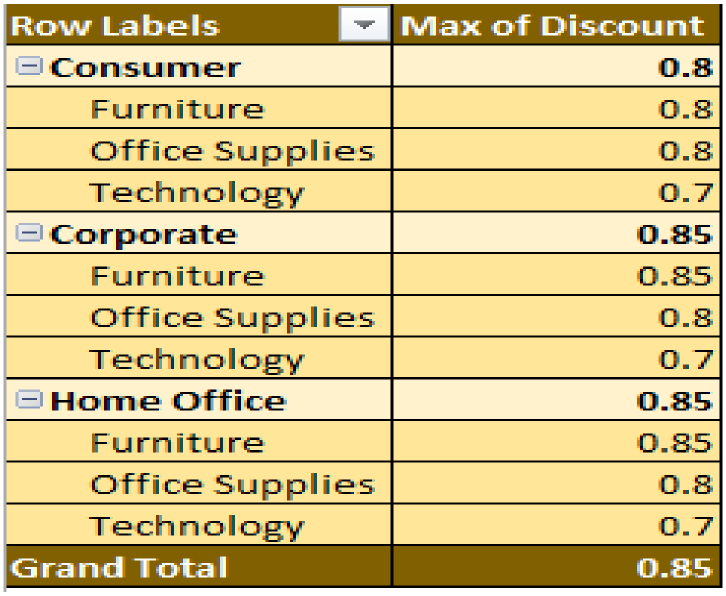
The analysis is based on how much discount will get for category which is part of segment

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.

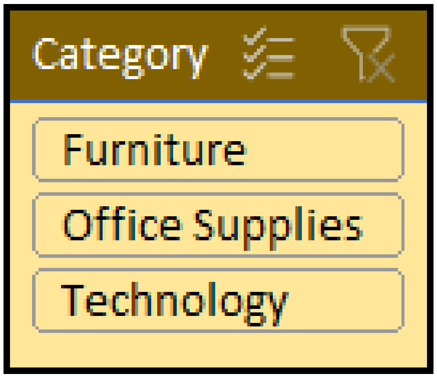
Analysis results:

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



Row Labels	Max of Discount
⊟ **Consumer**	**0.8**
Furniture	0.8
Office Supplies	0.8
Technology	0.7
⊟ **Corporate**	**0.85**
Furniture	0.85
Office Supplies	0.8
Technology	0.7
⊟ **Home Office**	**0.85**
Furniture	0.85
Office Supplies	0.8
Technology	0.7
Grand Total	0.85

Slicer:



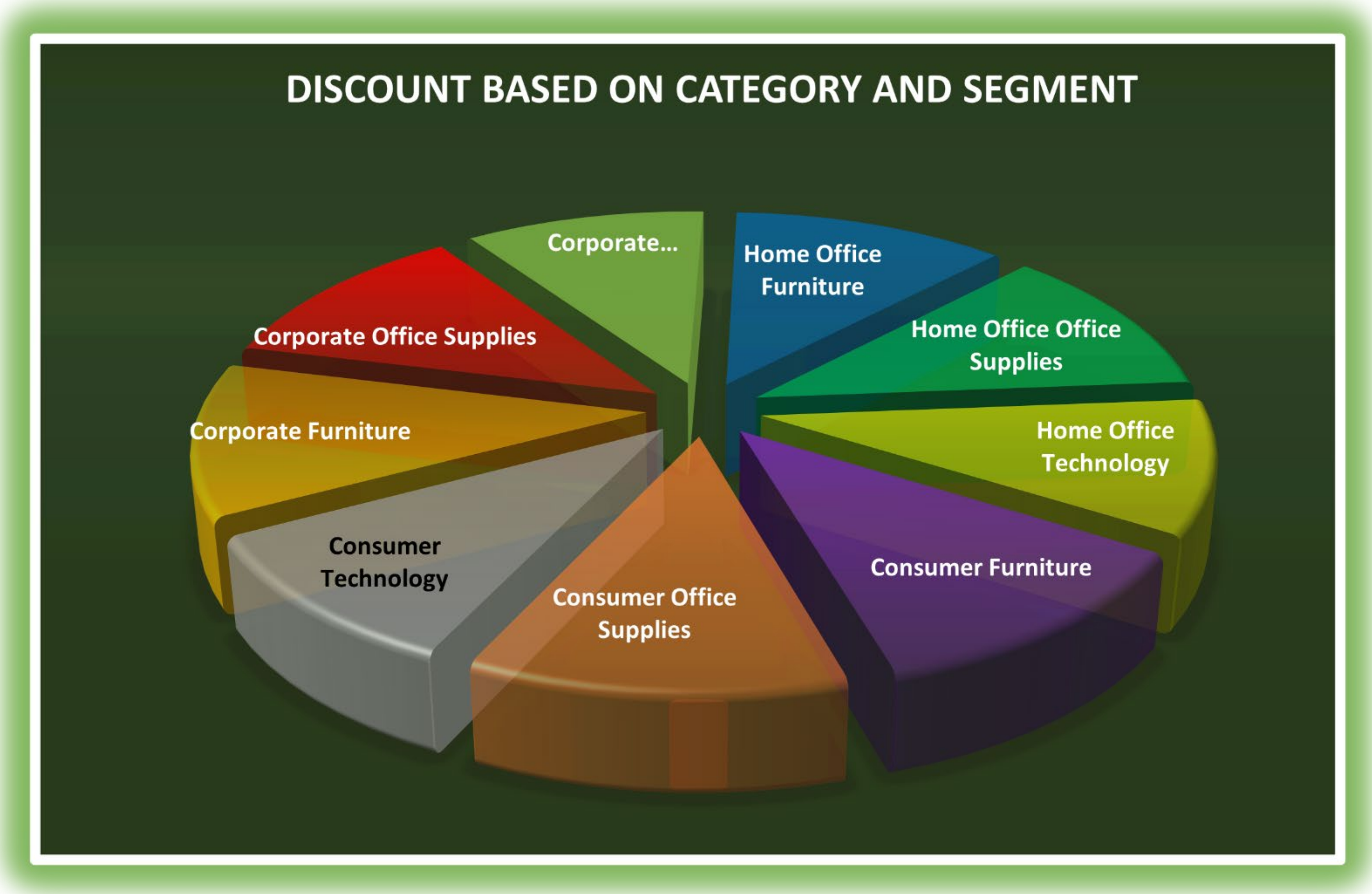
Category

Furniture

Office Supplies

Technology

Visualization:



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*

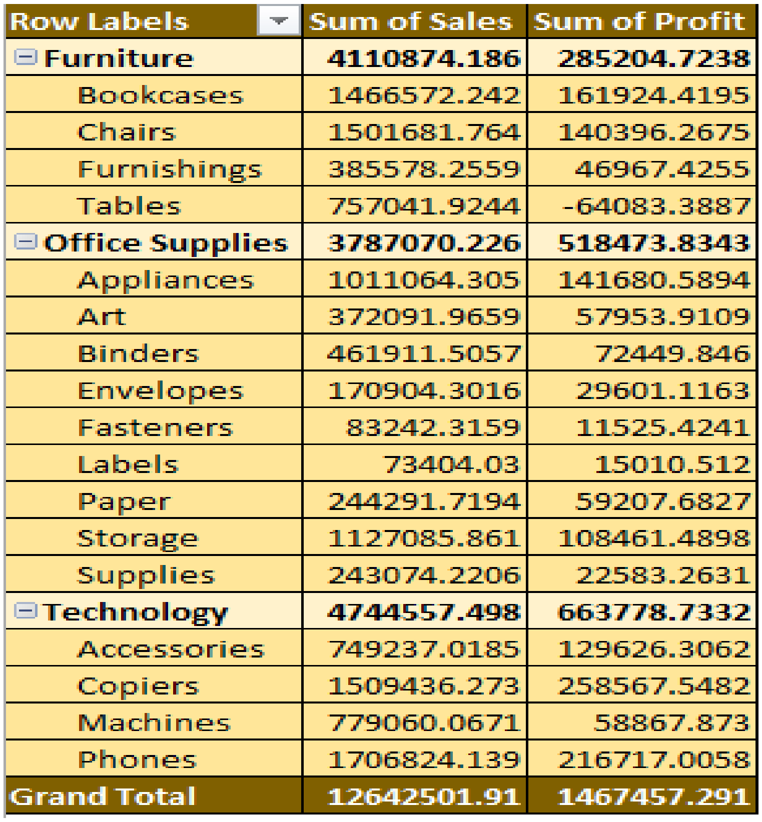
Description:

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:

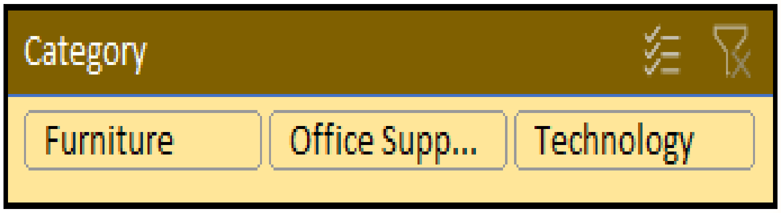
Select the cell below the given Quantity and apply the formula '=Sum ()'. This function will add the numbers in a range of cells

Analysis results:

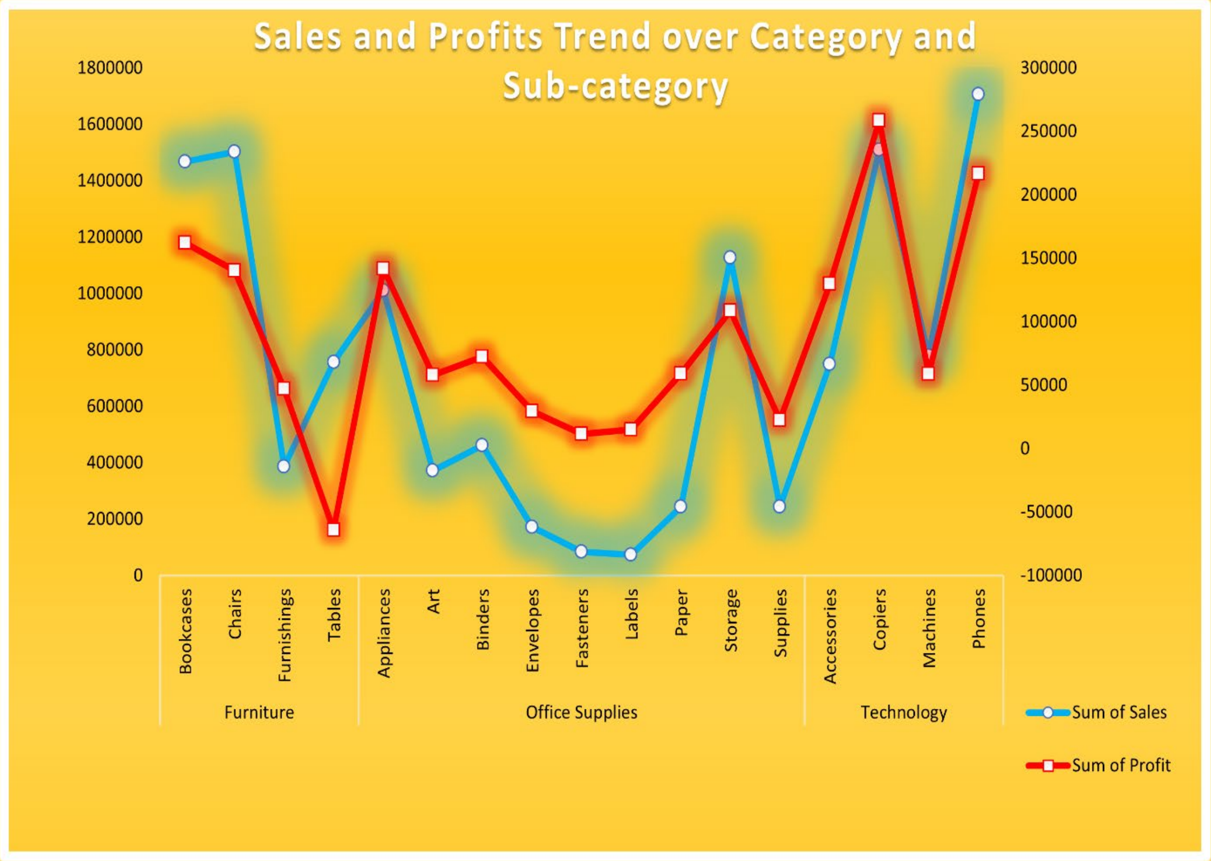


Row Labels	Sum of Sales	Sum of Profit
⊟ **Furniture**	**4110874.186**	**285204.7238**
Bookcases	1466572.242	161924.4195
Chairs	1501681.764	140396.2675
Furnishings	385578.2559	46967.4255
Tables	757041.9244	-64083.3887
⊟ **Office Supplies**	**3787070.226**	**518473.8343**
Appliances	1011064.305	141680.5894
Art	372091.9659	57953.9109
Binders	461911.5057	72449.846
Envelopes	170904.3016	29601.1163
Fasteners	83242.3159	11525.4241
Labels	73404.03	15010.512
Paper	244291.7194	59207.6827
Storage	1127085.861	108461.4898
Supplies	243074.2206	22583.2631
⊟ **Technology**	**4744557.498**	**663778.7332**
Accessories	749237.0185	129626.3062
Copiers	1509436.273	258567.5482
Machines	779060.0671	58867.873
Phones	1706824.139	216717.0058
Grand Total	12642501.91	1467457.291

Slicer:



Visualization:



4. *Sales and profits Trend over time*

Introduction

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

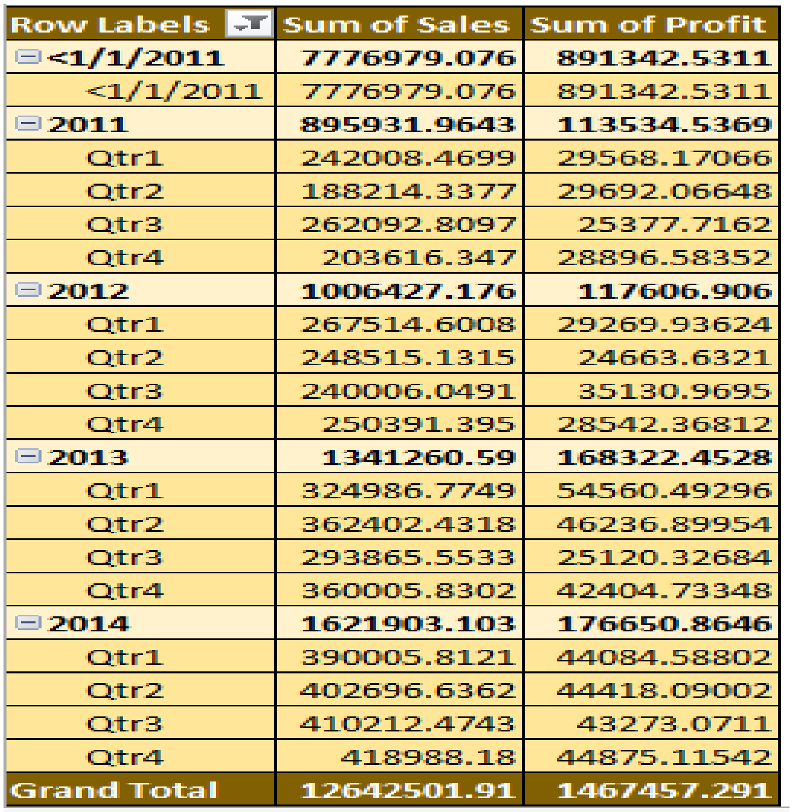
Description:

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:

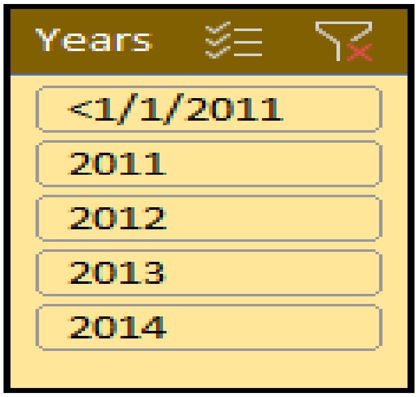
Select the cell below the given Quantity and apply the formula '=Sum ()'.
This function will add the numbers in a range of cells

Analysis results:

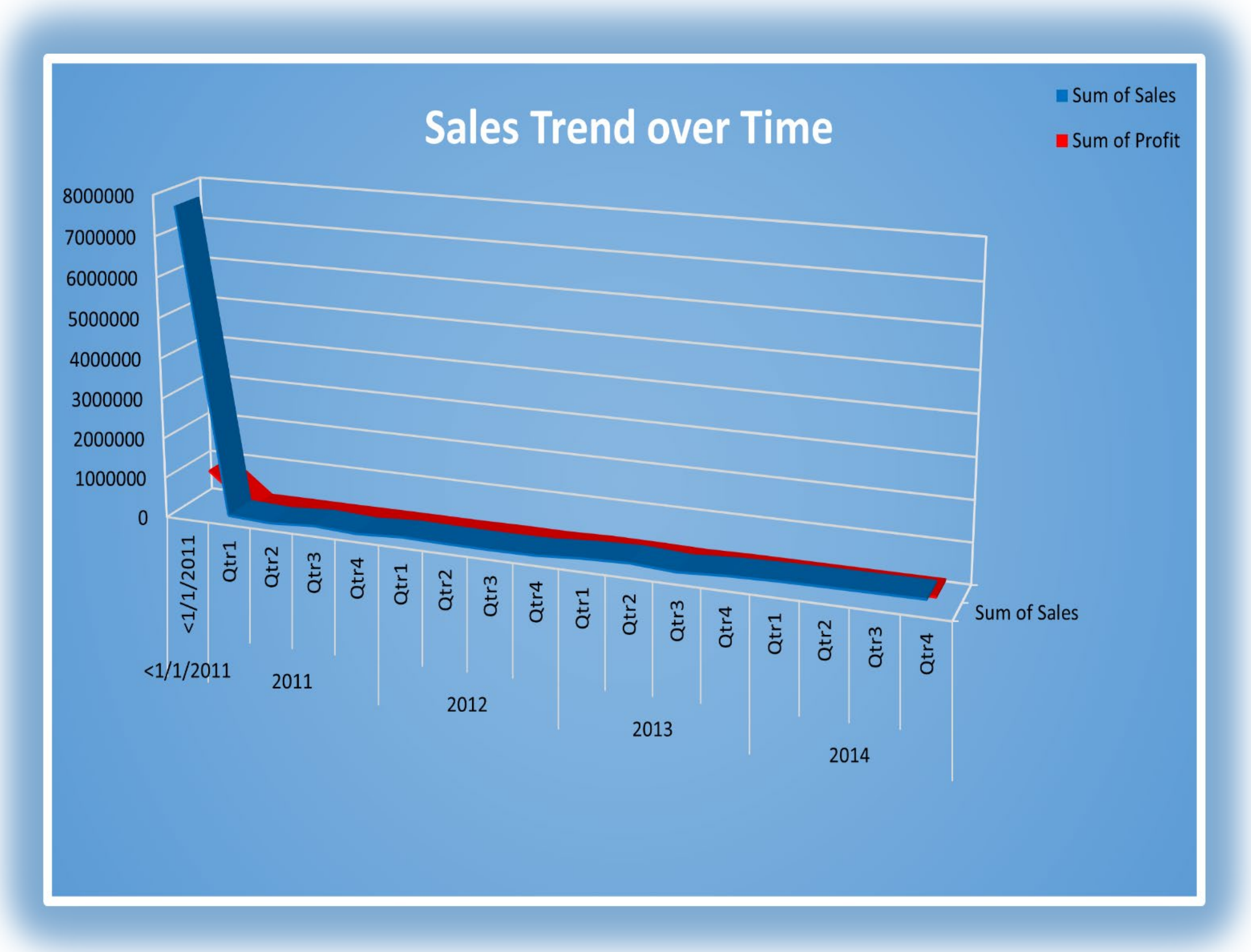


Row Labels	Sum of Sales	Sum of Profit
⊟ <1/1/2011	7776979.076	891342.5311
<1/1/2011	7776979.076	891342.5311
⊟ 2011	895931.9643	113534.5369
Qtr1	242008.4699	29568.17066
Qtr2	188214.3377	29692.06648
Qtr3	262092.8097	25377.7162
Qtr4	203616.347	28896.58352
⊟ 2012	1006427.176	117606.906
Qtr1	267514.6008	29269.93624
Qtr2	248515.1315	24663.6321
Qtr3	240006.0491	35130.9695
Qtr4	250391.395	28542.36812
⊟ 2013	1341260.59	168322.4528
Qtr1	324986.7749	54560.49296
Qtr2	362402.4318	46236.89954
Qtr3	293865.5533	25120.32684
Qtr4	360005.8302	42404.73348
⊟ 2014	1621903.103	176650.8646
Qtr1	390005.8121	44084.58802
Qtr2	402696.6362	44418.09002
Qtr3	410212.4743	43273.0711
Qtr4	418988.18	44875.11542
Grand Total	12642501.91	1467457.291

Slicer:



Visualization:



5. *Top 5 Costumers & Products*

Introduction

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

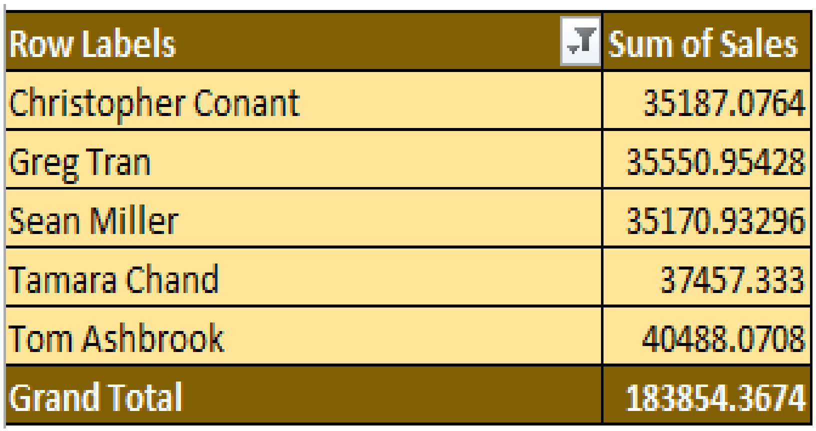
Description:

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.

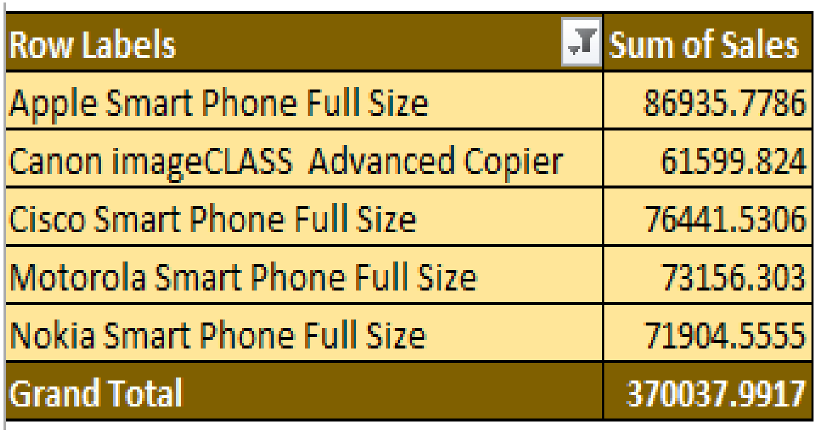
Specific requirements, functions, and formulas:

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:

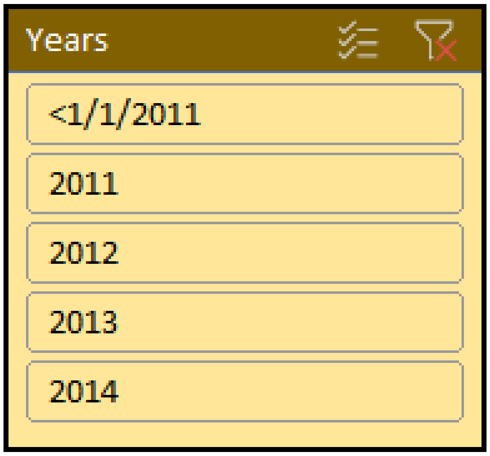


Row Labels	Sum of Sales
Christopher Conant	35187.0764
Greg Tran	35550.95428
Sean Miller	35170.93296
Tamara Chand	37457.333
Tom Ashbrook	40488.0708
Grand Total	183854.3674

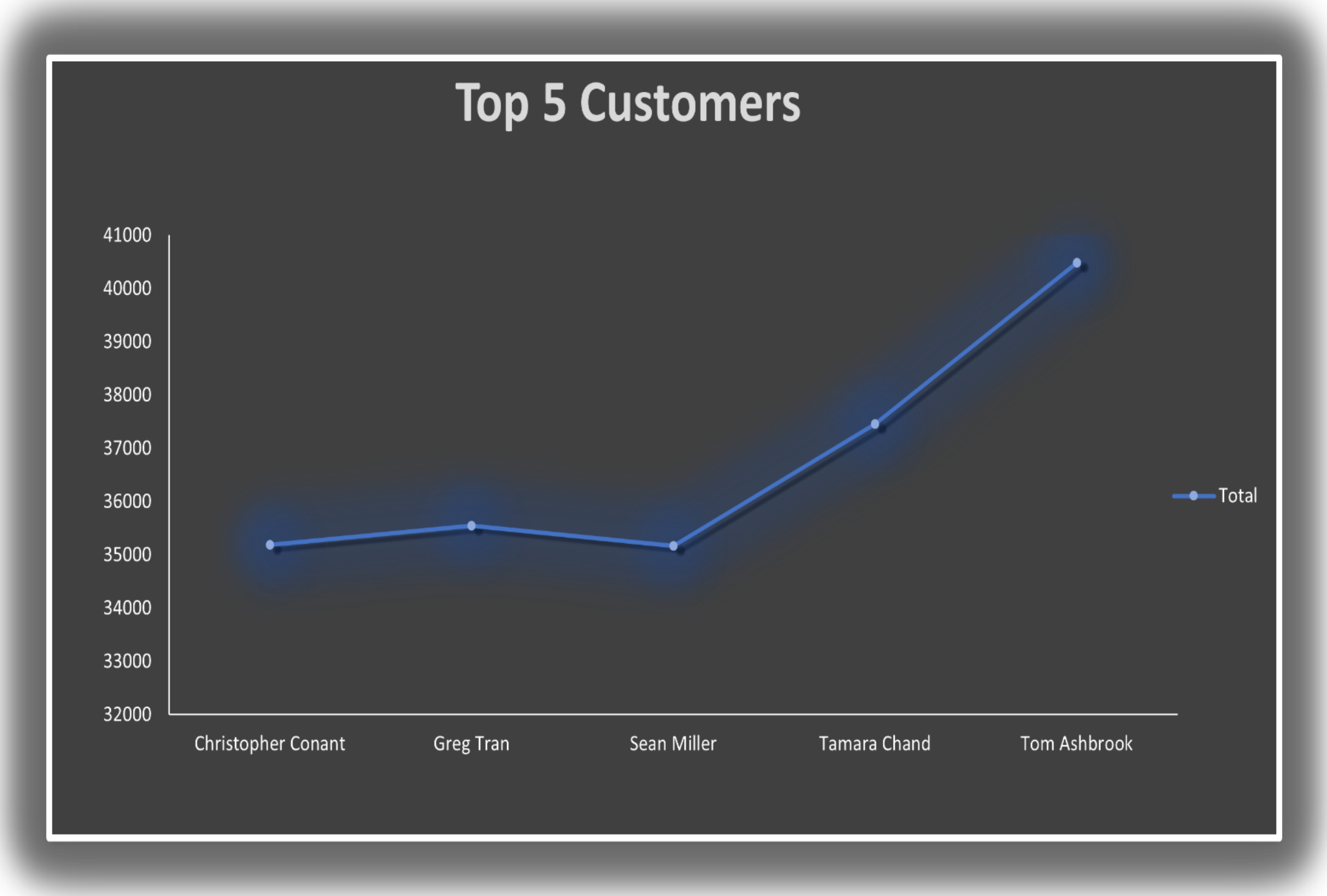


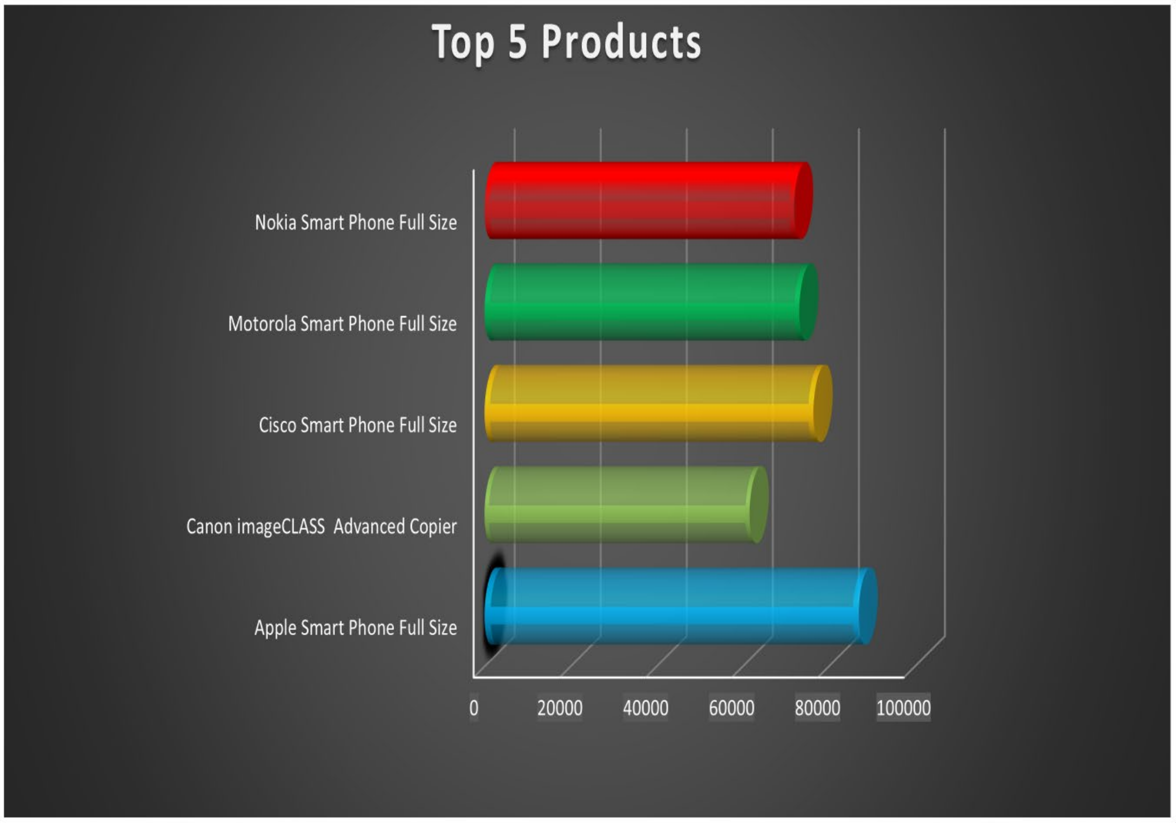
Row Labels	Sum of Sales
Apple Smart Phone Full Size	86935.7786
Canon imageCLASS Advanced Copier	61599.824
Cisco Smart Phone Full Size	76441.5306
Motorola Smart Phone Full Size	73156.303
Nokia Smart Phone Full Size	71904.5555
Grand Total	370037.9917

Slicer:



Visualization:

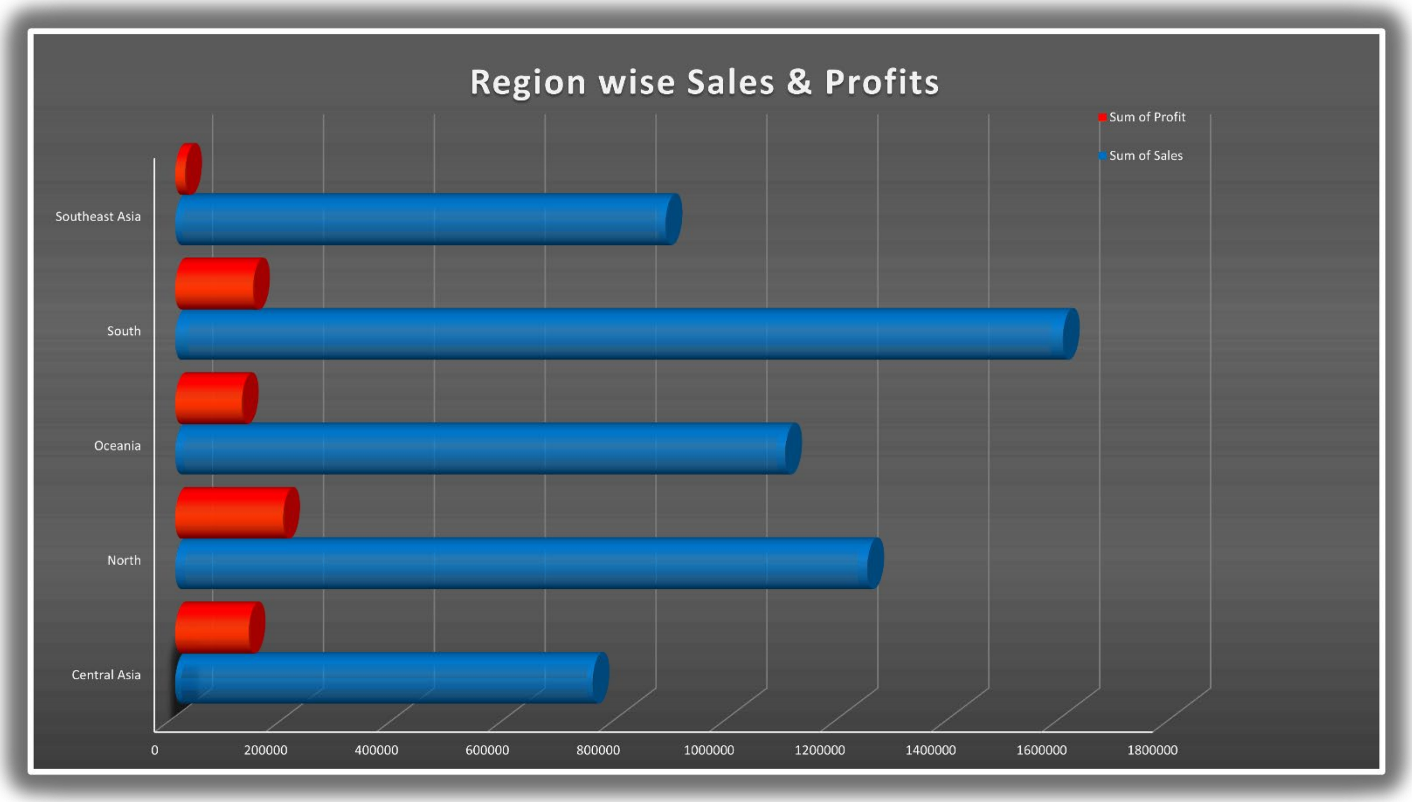



Top 5 Products
Nokia Smart Phone Full Size
Motorola Smart Phone Full Size
Cisco Smart Phone Full Size
Canon imageCLASS Advanced Copier
Apple Smart Phone Full Size
0
20000
40000
60000
80000
100000

List of Analysis with results

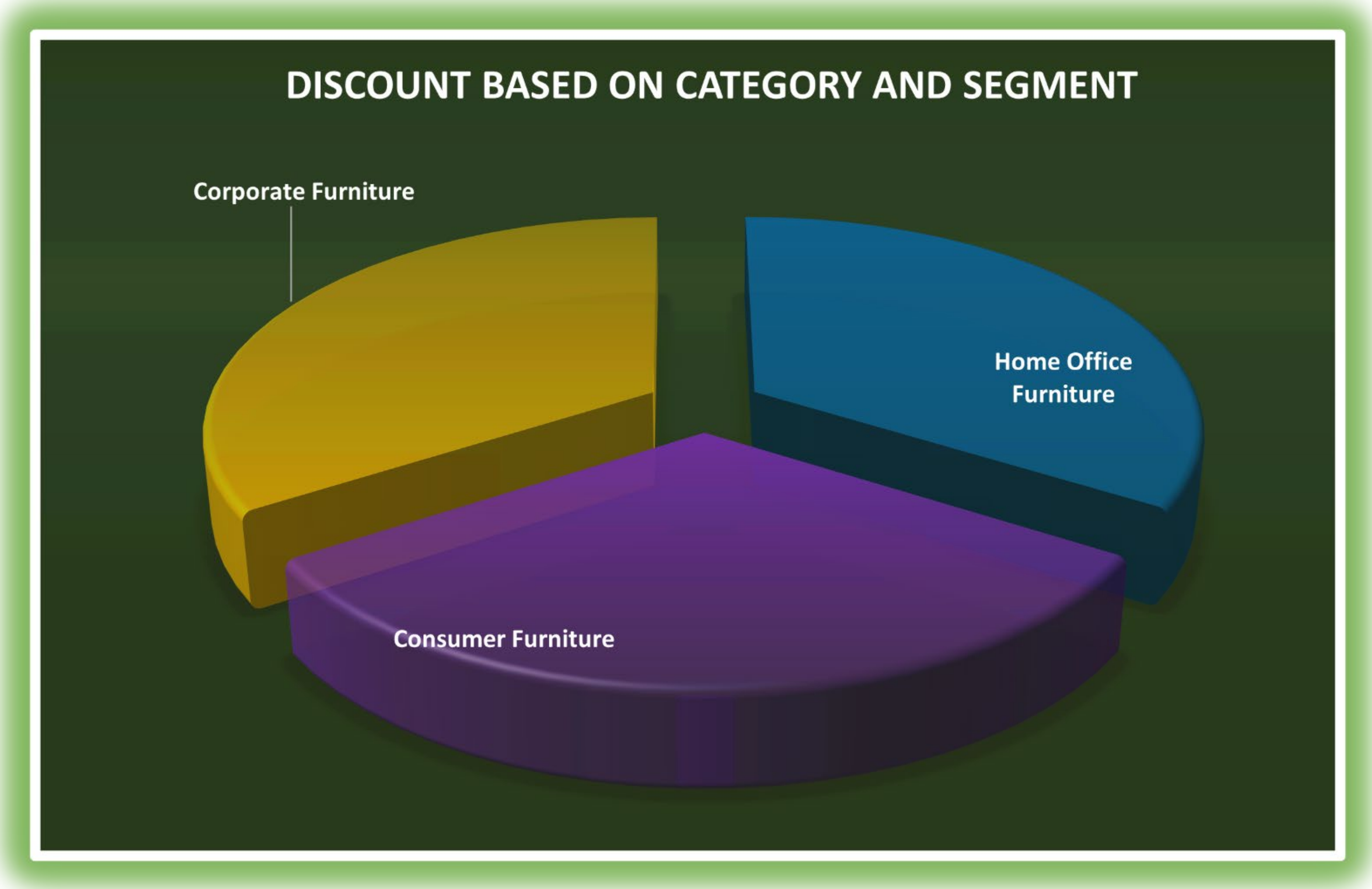
1. *Top 5 Regions on sales and profits:*

1.South
2.North
3.Oceania
4.Southeast Asia
5.Central Asia



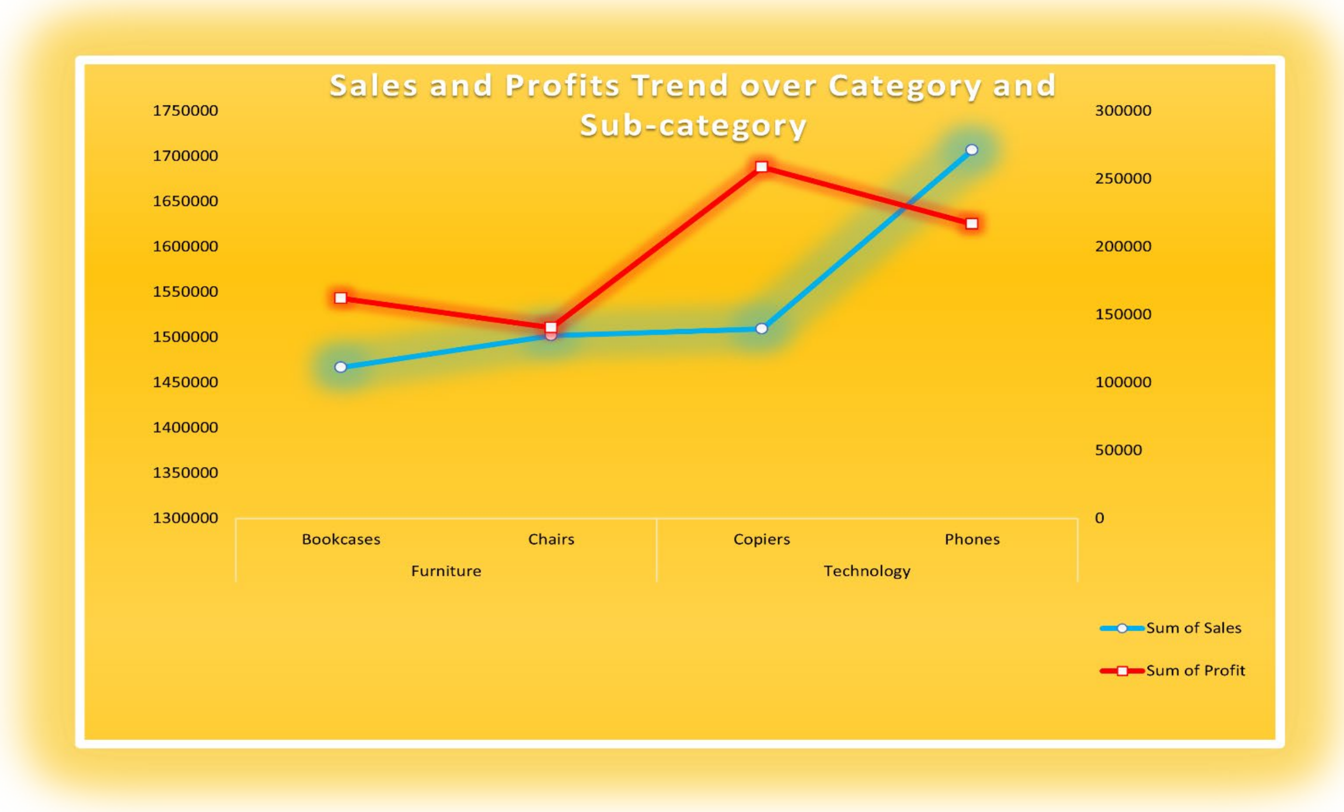
2. ***Top Segments in top category:***

1.Corporate Furniture
2.Home Office Furniture
3.Coporate Furniture



3. *Top 4 sales in sub-category:*

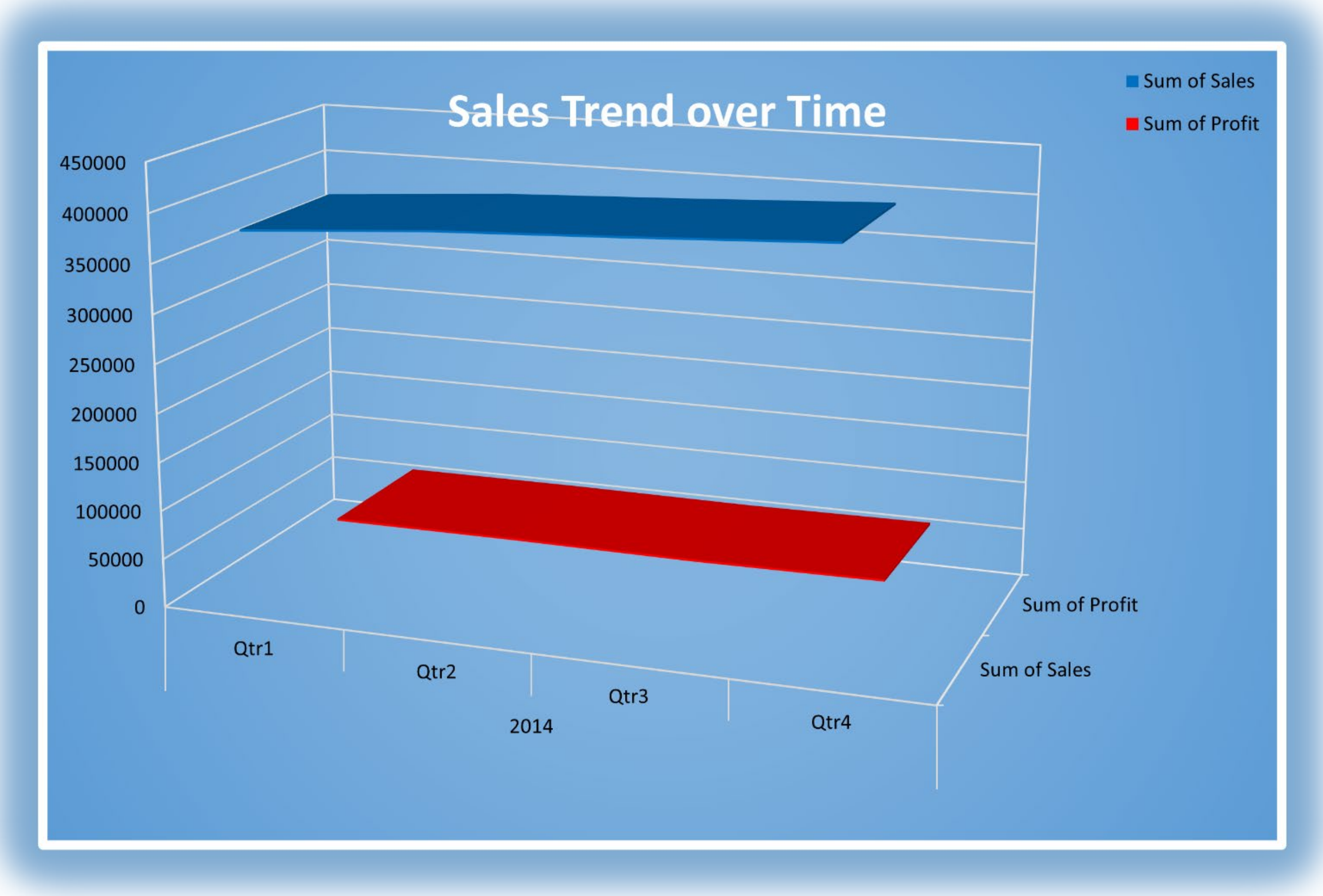
1.Phones
2.Copies
3.Chairs
4.Bookcases



.

4. *Most Sales & profits Happen in 2014:*

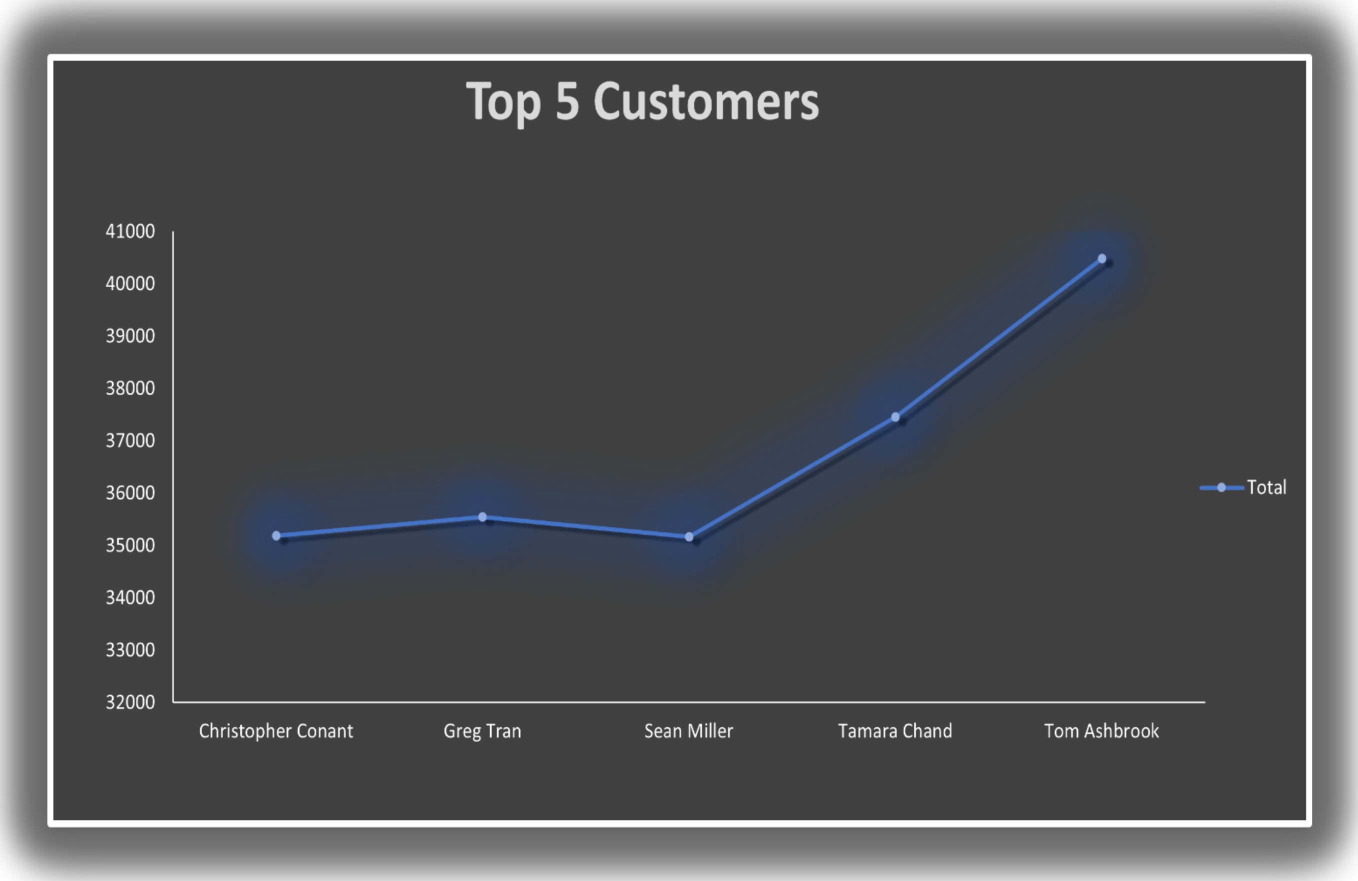
The rise in sales and profits is visible each year. There was a high incline in the year2014 by the fourth quarter.



5. *Top 5 Costumers & Products:*

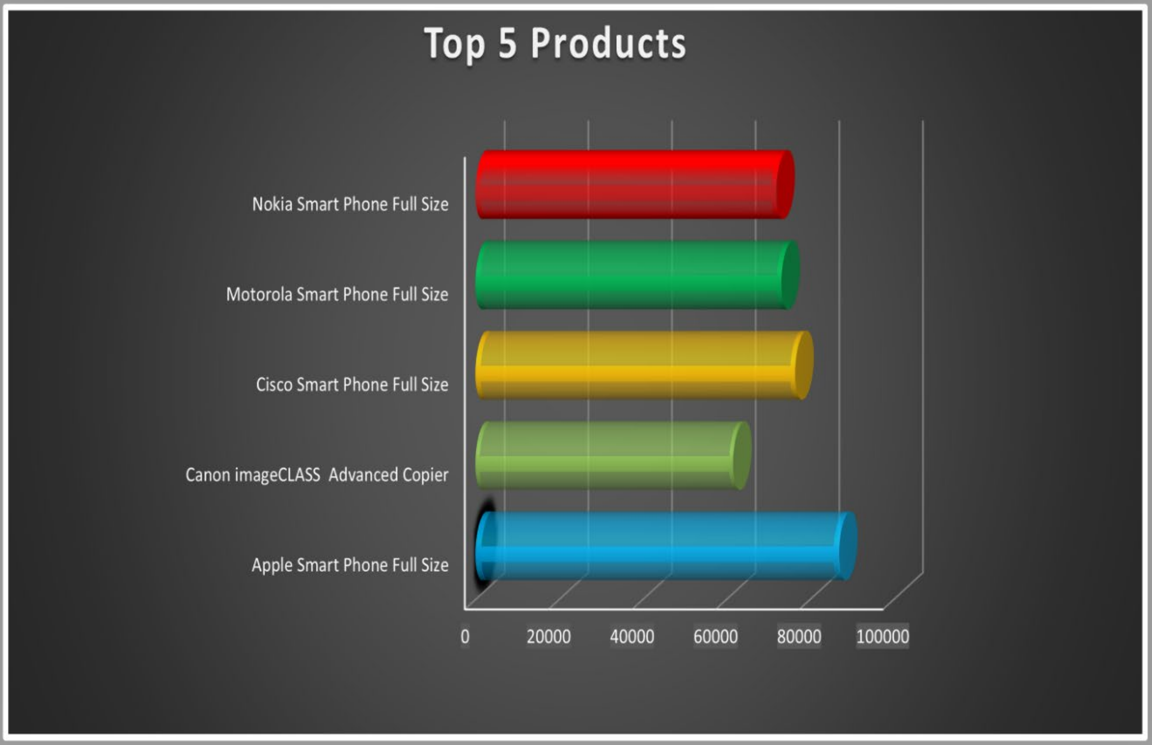
Costumers:

1.Tom Ashbrook
2.Tamara Chand
3.Grag Tran
4.Christopherconant
5.Sean Miller

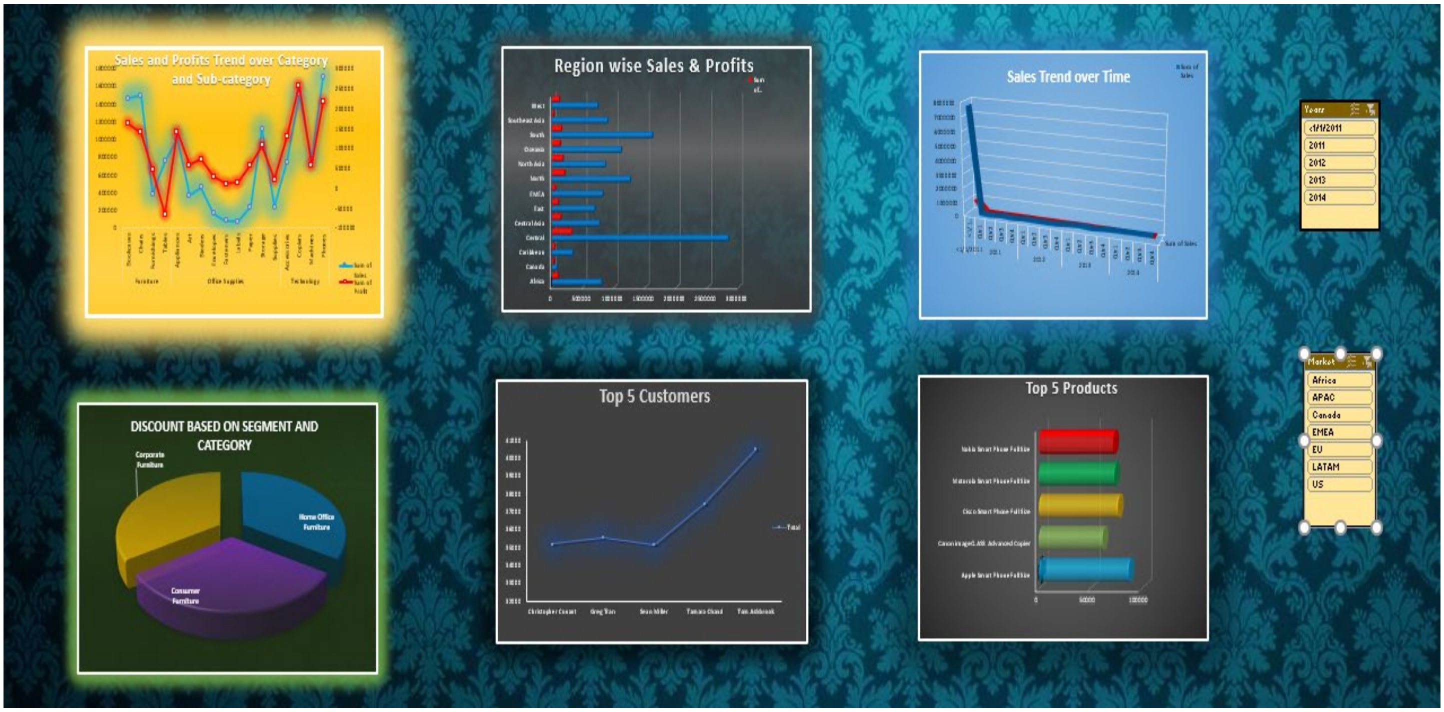


Products:

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



FINAL DASHBOARD:



BIBLIOGRAPHY:

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