INTRODUCTION TO DATA MANAGEMENT PROJECT REPORT

(Project Semester August-December 2021)

SUPERSTORE SALES ANALYSIS

Submitted by

SHIVAM THAKUR

11910626

Programme: Bachelor of Technology

Section: KM003

Course Code: INT217

Under the Guidance of

Ashu: 23631

Discipline of CSE/IT

Lovely School of Computer Science & Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that Shivam Thakur bearing Registration no. 11910626 has completed INT217 project titled, **"Superstore sales analysis"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Mrs. ASHU
School of computer science & engineering
Lovely Professional University
Phagwara, Punjab.

Date: 13/12/2021

DECLARATION

I, Shivam Thakur 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: 13/12/2021

Registration No. 11910626   Shivam Thakur

ACKNOWLEDGEMENT

Primarily I'd thank God for being able to complete my project with success. Then I'd like to thank my mentor Mrs. Ashu, 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.

Shivam Thakur

CONTENTS

Sr No.	Title	Page No.
1	Introduction	5
2	Objectives/Scope of the Analysis	6
3	Source of dataset	7
4	ETL Process	10
5	Analysis of dataset	16
6	Final Dashboard	33
7	Bibliography	34

INTRODUCTION

- Data management is important because the data your organization creates is a very valuable resource.

- The last thing you want to do is spend time and resources collecting data and business intelligence, only to lose or misplace that information.

- In that case, you would then have to spend time and resources again to get that same business intelligence you already had.

- And on that data analysis is carried out which show visualization of our problems in efficient way.

- Data Analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing

- conclusions, and supporting decision- making.

- This project is based on such data analysis on SUPER-STORE data from 2014 to 2017

- This dataset contains the information related to customer Name, Order id, Order date, Segment, Country, City, Product name, Product category, Product Sub-category, sales, quantity, discount, profit and prefer mode for delivering.

- This Super-Store dataset contains 21 data fields.

OBJECTIVES/SCOPE OF ANALYSIS

After analysis of the dataset, the aim of this project is to give answer of given

objectives in easy way:

- Total profit over sales in various segment

- sub-category wise profit over quantity

- Most favorite products in different sub-category

- State wise sales Trend using map

- Top 5 customers with most orders

- Average sales in different region

- Top 5 Cities with max profit & max loss

SOURCE OF DATASET

Source of Dataset:
https://docs.google.com/spreadsheets/d/1HnaEEz_K9Ny0XLvOreSBZMyRkT6iMS6X/edit?us
p=sharing&ouid=103749046478552787016&rtpof=true&sd=true

The dataset is based on Super Store details from 2014 to 2017.

The columns included in the dataset are given below:

- Row Id
    - Product rank order
- Order Id
    - The Id of the Product
- Order date
    - Date of the order
- Ship date
    - Shipping Date of the Product
- Ship mode
    - Shipping Mode of the Product(class)
- Customer Id
    - Id of the Customer
- Customer Name
    - Customer Name
- Segment
    - segment related to customer

- Country
  - Country of customer
- City
  - city of customer
- Region
  - city of customer
- State
  - State of customer
- Postal code
  - Postal code of customer
- Product Id
  - Id designated to each product
- Category
  - Category of product
- Sub-Category
  - Sub-Category of product
- Name
  - Name of the product
- Sales
  - Sales of the product
- Profit
  - Sales of the product
- Discount
  - Discount on the product
- Quantity
  - the number of product sold

Sample of dataset with data fields is given below:

ETL PROCESS

- **ETL** is a process that extracts the data from different source systems, then transforms the data (like applying calculations, concatenations, etc.) and finally loads the data into the Data Warehouse system.
- Full form of ETL is Extract, Transform and Load.
- The triple combination of ETL provides crucial functions that are many times combined into a single application or suite of tools that help in the following areas:
    - Enhances Business Intelligence solutions for decision making.
    - Allows verification of data transformation, aggregation, and calculations rules.
    - Allows sample data comparison between source and target system.
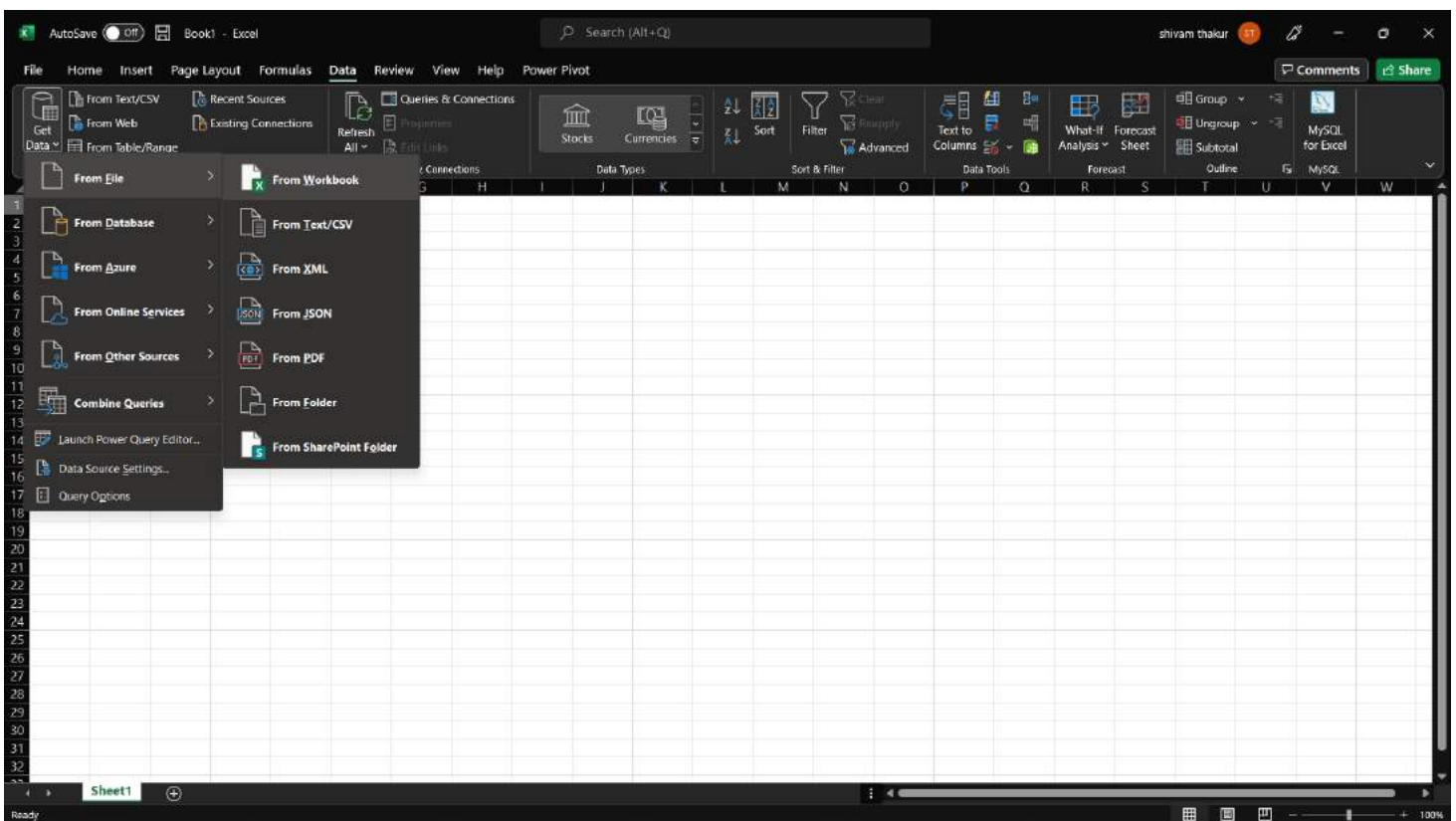   - Helps to improve productivity as it codifies and reuses without additional technical skills.

Initially, the raw dataset was arranged as shown in given picture:

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Na	Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category	Pro
1	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United State	Henderson	Kentucky	42420	South	FUR-BO-1000	Furniture	Bookcases	Bu
2	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United State	Henderson	Kentucky	42420	South	FUR-CH-1000	Furniture	Chairs	Ho
3	CA-2016-138688	6/12/2016	6/16/2016	Second Class	DV-13045	Darrin Van H	Corporate	United State	Los Angeles	California	90036	West	OFF-LA-1000(	Office Suppli	Labels	Sel
4	US-2015-108966	10/11/2015	10/18/2015	Standard Cla	SO-20335	Sean O'Donn	Consumer	United State	Fort Lauderd	Florida	33311	South	FUR-TA-1000	Furniture	Tables	Bre
5	US-2015-108966	10/11/2015	10/18/2015	Standard Cla	SO-20335	Sean O'Donn	Consumer	United State	Fort Lauderd	Florida	33311	South	OFF-ST-1000(	Office Suppli	Storage	Eld
6	CA-2014-115812	6/9/2014	6/14/2014	Standard Cla	BH-11710	Brosina Hoff	Consumer	United State	Los Angeles	California	90032	West	FUR-FU-1000	Furniture	Furnishings	Eld
7	CA-2014-115812	6/9/2014	6/14/2014	Standard Cla	BH-11710	Brosina Hoff	Consumer	United State	Los Angeles	California	90032	West	OFF-AR-1000	Office Suppli	Art	Ne
8	CA-2014-115812	6/9/2014	6/14/2014	Standard Cla	BH-11710	Brosina Hoff	Consumer	United State	Los Angeles	California	90032	West	TEC-PH-1000	Technology	Phones	Mi
9	CA-2014-115812	6/9/2014	6/14/2014	Standard Cla	BH-11710	Brosina Hoff	Consumer	United State	Los Angeles	California	90032	West	OFF-BI-10003	Office Suppli	Binders	DX
10	CA-2014-115812	6/9/2014	6/14/2014	Standard Cla	BH-11710	Brosina Hoff	Consumer	United State	Los Angeles	California	90032	West	OFF-AP-1000	Office Suppli	Appliances	Be
11	CA-2014-115812	6/9/2014	6/14/2014	Standard Cla	BH-11710	Brosina Hoff	Consumer	United State	Los Angeles	California	90032	West	FUR-TA-1000	Furniture	Tables	Ch
12	CA-2014-115812	6/9/2014	6/14/2014	Standard Cla	BH-11710	Brosina Hoff	Consumer	United State	Los Angeles	California	90032	West	TEC-PH-1000	Technology	Phones	Ko
13	CA-2017-114412	4/15/2017	4/20/2017	Standard Cla	AA-10480	Andrew Alle	Consumer	United State	Concord	North Caroli	28027	South	OFF-PA-1000	Office Suppli	Paper	Xe
14	CA-2016-161389	12/5/2016	12/10/2016	Standard Cla	IM-15070	Irene Maddo	Consumer	United State	Seattle	Washington	98103	West	OFF-BI-10003	Office Suppli	Binders	Fel
15	US-2015-118983	11/22/2015	11/26/2015	Standard Cla	HP-14815	Harold Pawla	Home Office	United State	Fort Worth	Texas	76106	Central	OFF-AP-1000	Office Suppli	Appliances	Ho
16	US-2015-118983	11/22/2015	11/26/2015	Standard Cla	HP-14815	Harold Pawla	Home Office	United State	Fort Worth	Texas	76106	Central	OFF-BI-1000(	Office Suppli	Binders	Sto
17	CA-2014-105893	11/11/2014	11/18/2014	Standard Cla	PK-19075	Pete Kriz	Consumer	United State	Madison	Wisconsin	53711	Central	OFF-ST-1000(	Office Suppli	Storage	Stu
18	CA-2014-167164	5/13/2014	5/15/2014	Second Class	AG-10270	Alejandro Gr	Consumer	United State	West Jordan	Utah	84084	West	OFF-ST-1000(	Office Suppli	Storage	Fel
19	CA-2014-143336	8/27/2014	9/1/2014	Second Class	ZD-21925	Zuschuss Dor	Consumer	United State	San Francisc	California	94109	West	OFF-AR-1000	Office Suppli	Art	Ne
20	CA-2014-143336	8/27/2014	9/1/2014	Second Class	ZD-21925	Zuschuss Dor	Consumer	United State	San Francisc	California	94109	West	TEC-PH-1000	Technology	Phones	Cis
21	CA-2014-143336	8/27/2014	9/1/2014	Second Class	ZD-21925	Zuschuss Dor	Consumer	United State	San Francisc	California	94109	West	OFF-BI-10002	Office Suppli	Binders	Wi
22	CA-2016-137330	12/9/2016	12/13/2016	Standard Cla	KB-16585	Ken Black	Corporate	United State	Fremont	Nebraska	68025	Central	OFF-AR-1000	Office Suppli	Art	Ne
23	CA-2016-137330	12/9/2016	12/13/2016	Standard Cla	KB-16585	Ken Black	Corporate	United State	Fremont	Nebraska	68025	Central	OFF-AP-1000	Office Suppli	Appliances	Acc
24	US-2017-156909	7/16/2017	7/18/2017	Second Class	SF-20065	Sandra Flana	Consumer	United State	Philadelphia	Pennsylvania	19140	East	FUR-CH-1000	Furniture	Chairs	Glo
25	CA-2015-106320	9/25/2015	9/30/2015	Standard Cla	EB-13870	Emily Burns	Consumer	United State	Orem	Utah	84057	West	FUR-TA-1000	Furniture	Tables	Bre
26	CA-2016-121755	1/16/2016	1/20/2016	Second Class	EH-13945	Eric Hoffman	Consumer	United State	Los Angeles	California	90049	West	OFF-BI-10001	Office Suppli	Binders	Wi
27	CA-2016-121755	1/16/2016	1/20/2016	Second Class	EH-13945	Eric Hoffman	Consumer	United State	Los Angeles	California	90049	West	TEC-AC-1000	Technology	Accessories	Im
28	US-2015-150630	9/17/2015	9/21/2015	Standard Cla	TB-21520	Tracy Blumst	Consumer	United State	Philadelphia	Pennsylvania	19140	East	FUR-BO-100(	Furniture	Bookcases	Riv
29	US-2015-150630	9/17/2015	9/21/2015	Standard Cla	TB-21520	Tracy Blumst	Consumer	United State	Philadelphia	Pennsylvania	19140	East	OFF-BI-1000(	Office Suppli	Binders	Av

Orders | Returns | People

Steps taken to clean dataset thorough ETL process

Step 1:

Firstly, open blank excel file and go to Data option from tab and select **Get Data** then go for **from file** option and use **from workbook** option

Step 2:

After the first step, navigator will appear use the required sheet

Step 3:

Select transform option for cleaning, and transforming data and applied steps are showed in right side in query setting

Step 4:

After successfully extracting, transforming now time to load the data

Step 5:
Finally, after cleaning the data, the final dataset sample is shown below:

Analysis on Dataset

Segment wise Sales-Profit Trend

Introduction

By performing this analysis, we will get segment wise sales- profit trend.

Description

The analysis is based on order date, sales, profit of the order.

Specific requirements, functions and formulas

- o Select the table dataset and then insert pivot table in new sheet
- o From pivot table fields use order date in rows, segment in columns
  - And sales, profit in values
- o Group date as years and quarters.

Analysis results

years		Consumer		Corporate		Home Office	
		Sum of Sales	Sum of Profit	Sum of Sales	Sum of Profit	Sum of Sales	Sum of Profit
⊟ 2014							
	⊞ Qtr1	21964.975	1860.3705	13991.995	941.2655	38490.826	1009.593
	⊞ Qtr2	47851.5226	5340.7592	27244.643	3523.899	11442.594	2339.411
	⊞ Qtr3	96103.6535	8242.5939	32992.507	2606.0886	14537.0518	1956.0393
	⊟ Qtr4						
	Oct	16246.697	2387.3712	12637.678	746.3934	2569.018	314.4927
	Nov	46571.978	1975.0425	20994.4947	3843.5716	11062.244	3473.5128
	Dec	37357.9865	4513.7131	20573.556	1852.0588	11614.078	2617.798
⊞ 2015		266535.9333	28460.1665	128757.3069	20688.3248	75239.2688	12470.1124
⊞ 2016		296863.8992	35770.9532	207106.3618	30995.169	105235.337	15029.0521
⊞ 2017		331904.6999	45568.2391	241847.8244	26782.3633	159462.7309	21088.6672

Visualization

Area chart is used for visualization of the segment wise sales and profit trend
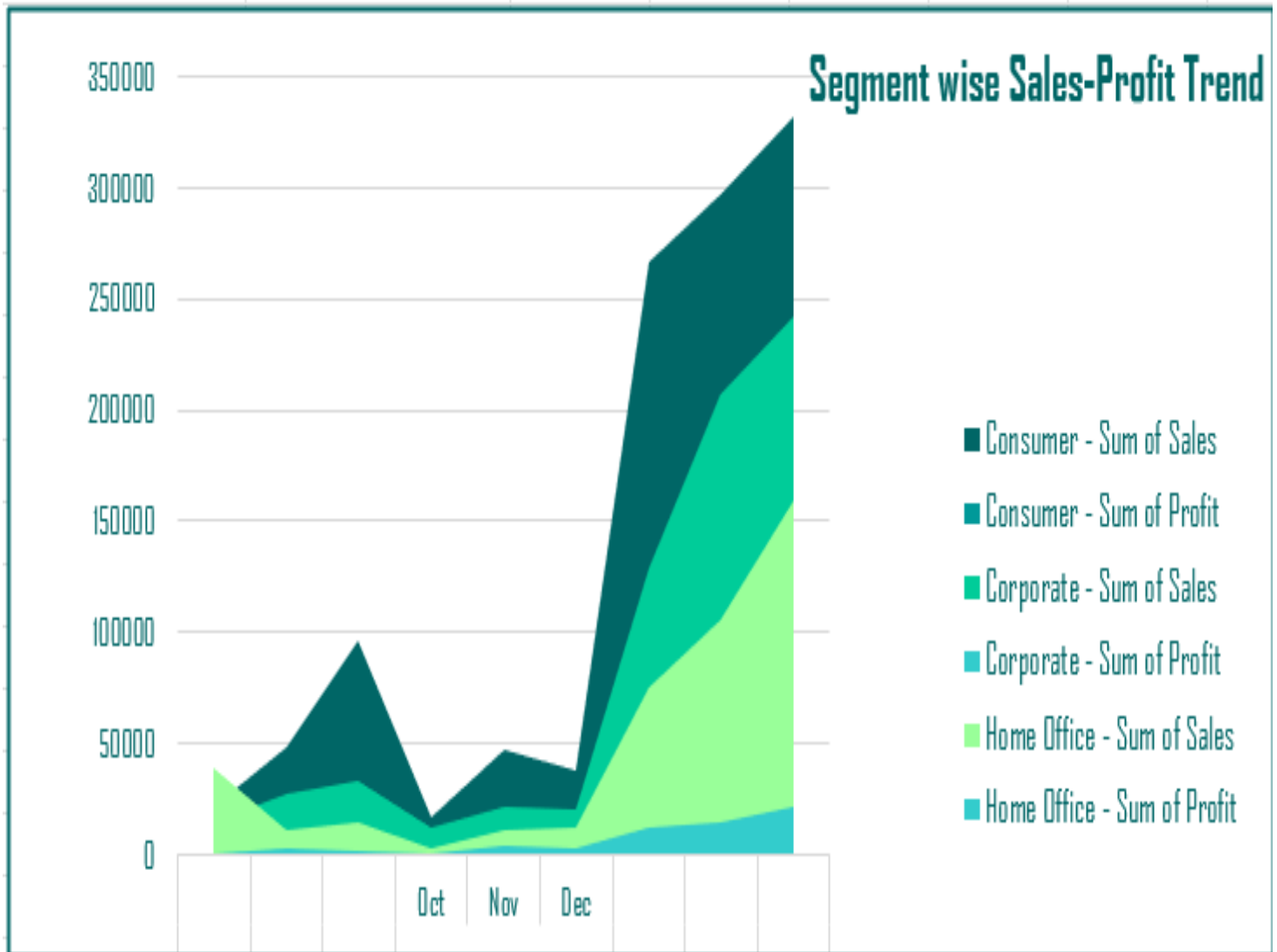
Sub-category wise profit over quantity

Introduction
By performing this analysis, we will get to know about most profitable sub-category.

Description
The analysis based on the Sub-category, profit, sales of the dataset

Specific requirements, functions and formulas
Pivot table is used for the analysis.
Sum function is used in pivot table for the sum of the profit, sales in the pivot table

Analysis results

Sub-Category	Sum of Quantity	Sum of Profit
Accessories	2,976	$41,937
Appliances	1,729	$18,138
Art	3,000	$6,528
Binders	5,974	$30,222
Bookcases	868	-$3,473
Chairs	2,356	$26,590
Copiers	234	$55,618
Envelopes	906	$6,964
Fasteners	914	$950
Furnishings	3,563	$13,059
Labels	1,400	$5,546
Machines	440	$3,385
Paper	5,178	$34,054
Phones	3,289	$44,516
Storage	3,158	$21,279
Supplies	647	-$1,189
Tables	1,241	-$17,725

Visualization

For visualization area-column combo chart is used

SUB-CATEGORY WISE PROFIT OVER QUANTITY

Values shown: $41,937; $18,138; $6,528; $30,222; $26,590; $55,618; -$3,473; $6,964; $950; $13,059; $5,546; $3,385; $34,054; $44,516; $21,279; -$1,189; -$17,725

Legend: Sum of Quantity, Sum of Profit

X-axis categories: Accessories, Appliances, Art, Binders, Bookcases, Chairs, Copiers, Envelopes, Fasteners, Furnishings, Labels, Machines, Paper, Phones, Storage, Supplies, Tables

Mode of shipment usage Trend

Introduction

By performing this analysis, we will get to know about most preferred by the user shipment mode

Description

The analysis based on the order id, shipment mode, oreder of the dataset

Specific requirements, functions and formulas

Pivot table is used for the analysis.
count function is used in pivot table for count order id for particular shipment mode in the pivot table

Analysis results

Count of Order ID	shipment Mode			
years/quarters	First Class	Same Day	Second Class	Standard Class
2014				
Qtr1	4	1	14	27
Qtr2	15	3	13	78
Qtr3	31	14	27	94
Qtr4	29	12	39	116
2015				
Qtr1	11	4	18	47
Qtr2	23	1	15	84
Qtr3	35	15	36	133
Qtr4	42	16	57	107
2016				
Qtr1	15	7	29	67
Qtr2	33	15	28	126
Qtr3	37	9	41	101
Qtr4	49	16	36	157
2017				
Qtr1	22	5	30	61
Qtr2	45	7	23	106
Qtr3	48	17	60	151
Qtr4	51	13	64	218

Visualization

For visualization line chart is used combo chart is used
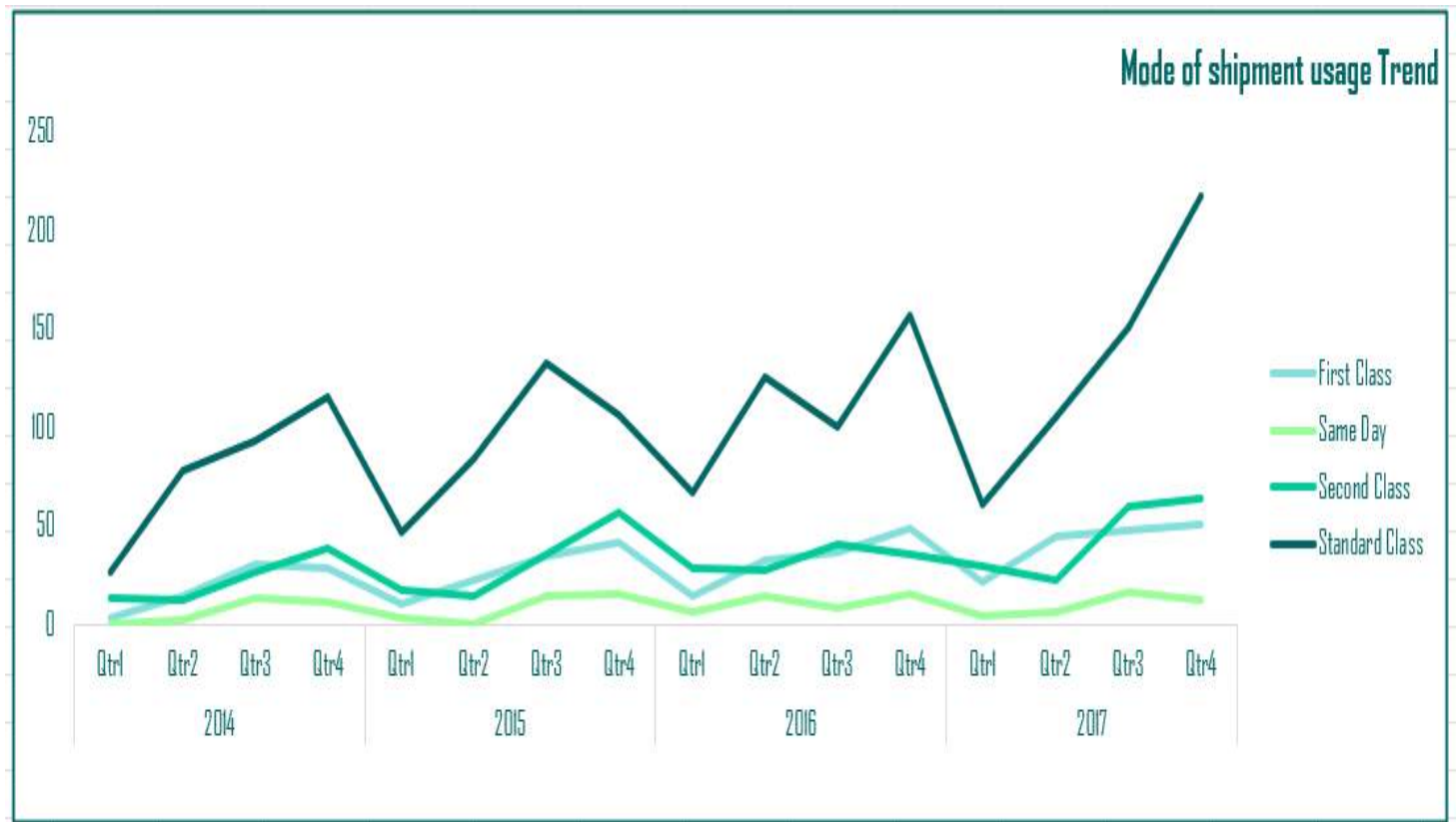
Top 5 most sold product sub-category wise

Introduction
By performing this analysis, we will get to know about most sold products

Description
The analysis based on the Sub-category, Quantity of the dataset

Specific requirements, functions, and formulas
Pivot table is used for the analysis.
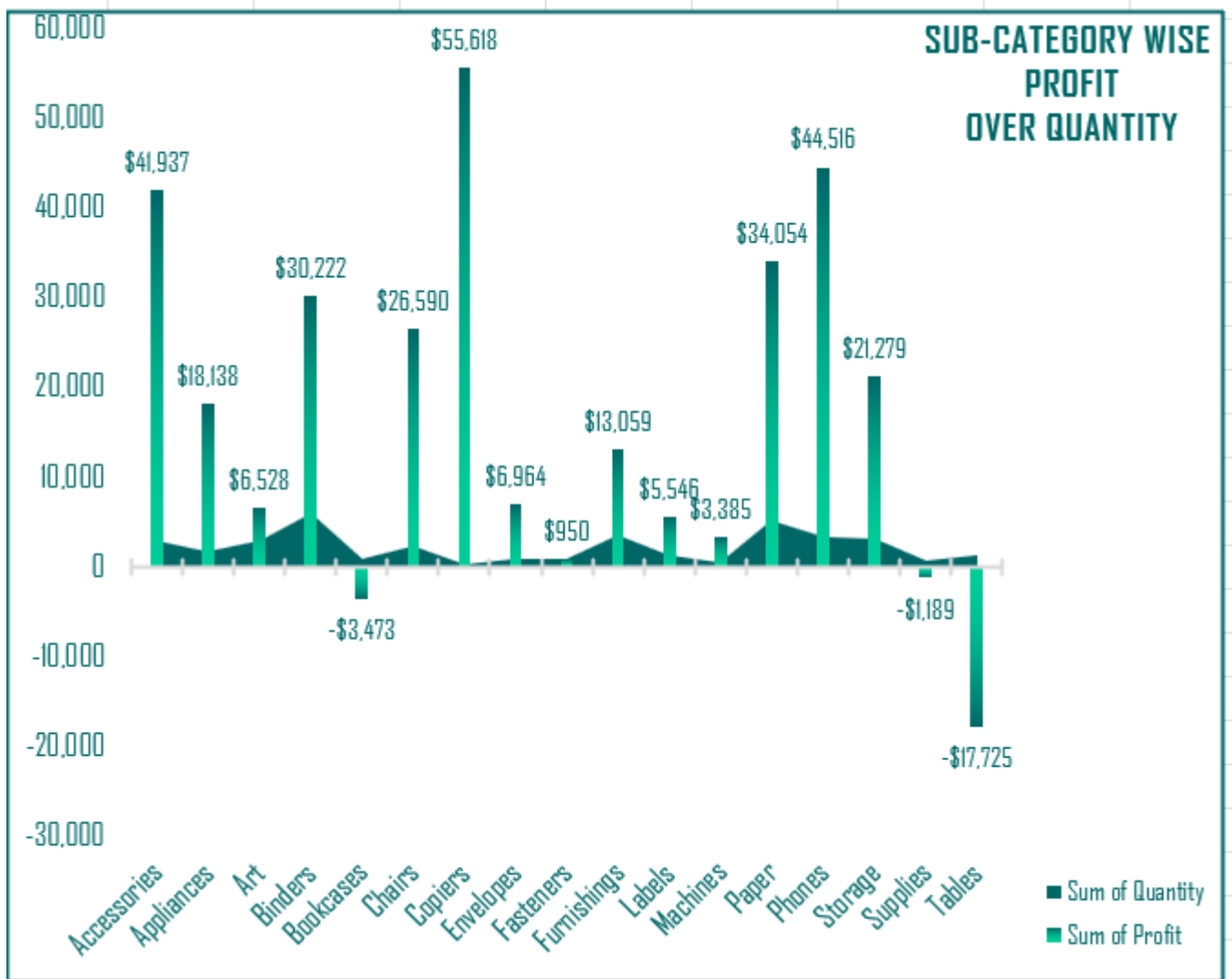Sum function is used in pivot table for the sum of the quantity in the pivot table

Analysis results

Visualization

For visualization area-column combo chart is used

Top 5 most sold product sub-category wise

Top 5 customers

Introduction
By performing this analysis, we will get to know about Top 5 customers.

Description
The analysis based on the customer's name, order id of the dataset

Specific requirements, functions, and formulas
Pivot table is used for the analysis.
count function is used in pivot table for the count of order id in the pivot table

Analysis results

Customer Name ⊤	Count of Order ID
Chloris Kastensmidt	32
Edward Hooks	32
John Lee	34
Jonathan Doherty	32
Matt Abelman	34
Paul Prost	34
Seth Vernon	32
William Brown	37

Visualization

For visualization pie chart is used

TOP 5 CUSTOMER

William Brown 14%

Chloris Kastens midt 12%

Edward Hooks 12%

Seth Vernon 12%

John Lee 12%

Paul Prost 13%

Jonathan Doherty 12%

Matt Abelman 13%

Average sales Trend region wise

Introduction

By performing this analysis, we will get to know about average sales Trend in each region

Description

The analysis based on the sales, order date and region of the dataset

Specific requirements, functions, and formulas

Pivot table is used for the analysis.

average function is used in pivot table for calculating average sales in each region in the pivot table

Analysis results

Average of Sales Years	Quarters	Region Central	East	South	West
⊟ 2014	Qtr1	$122.87	$143.03	$491.80	$197.44
	Qtr2	$181.32	$193.25	$300.32	$228.07
	Qtr3	$380.79	$201.47	$200.76	$247.31
	Qtr4	$182.93	$344.87	$201.91	$211.72
⊟ 2015	Qtr1	$178.31	$214.32	$382.43	$330.89
	Qtr2	$214.10	$184.58	$184.71	$216.43
	Qtr3	$231.00	$231.86	$228.29	$193.86
	Qtr4	$222.09	$295.97	$149.58	$204.87
⊟ 2016	Qtr1	$288.74	$209.96	$362.63	$300.21
	Qtr2	$202.83	$258.25	$185.68	$239.60
	Qtr3	$168.59	$202.40	$180.30	$216.75
	Qtr4	$327.30	$254.62	$231.71	$224.29
⊟ 2017	Qtr1	$270.20	$152.98	$252.63	$286.07
	Qtr2	$171.40	$174.10	$212.50	$216.29
	Qtr3	$171.79	$237.31	$194.10	$236.23
	Qtr4	$167.25	$283.30	$276.18	$202.55

Visualization
For visualization line chart is used

Average sales Trend Region wise

State wise sales Trend

Introduction
By performing this analysis, we will get to know about state wise sales trend .

Description
The analysis based on the country, state, sales of the dataset

Specific requirements, functions, and formulas
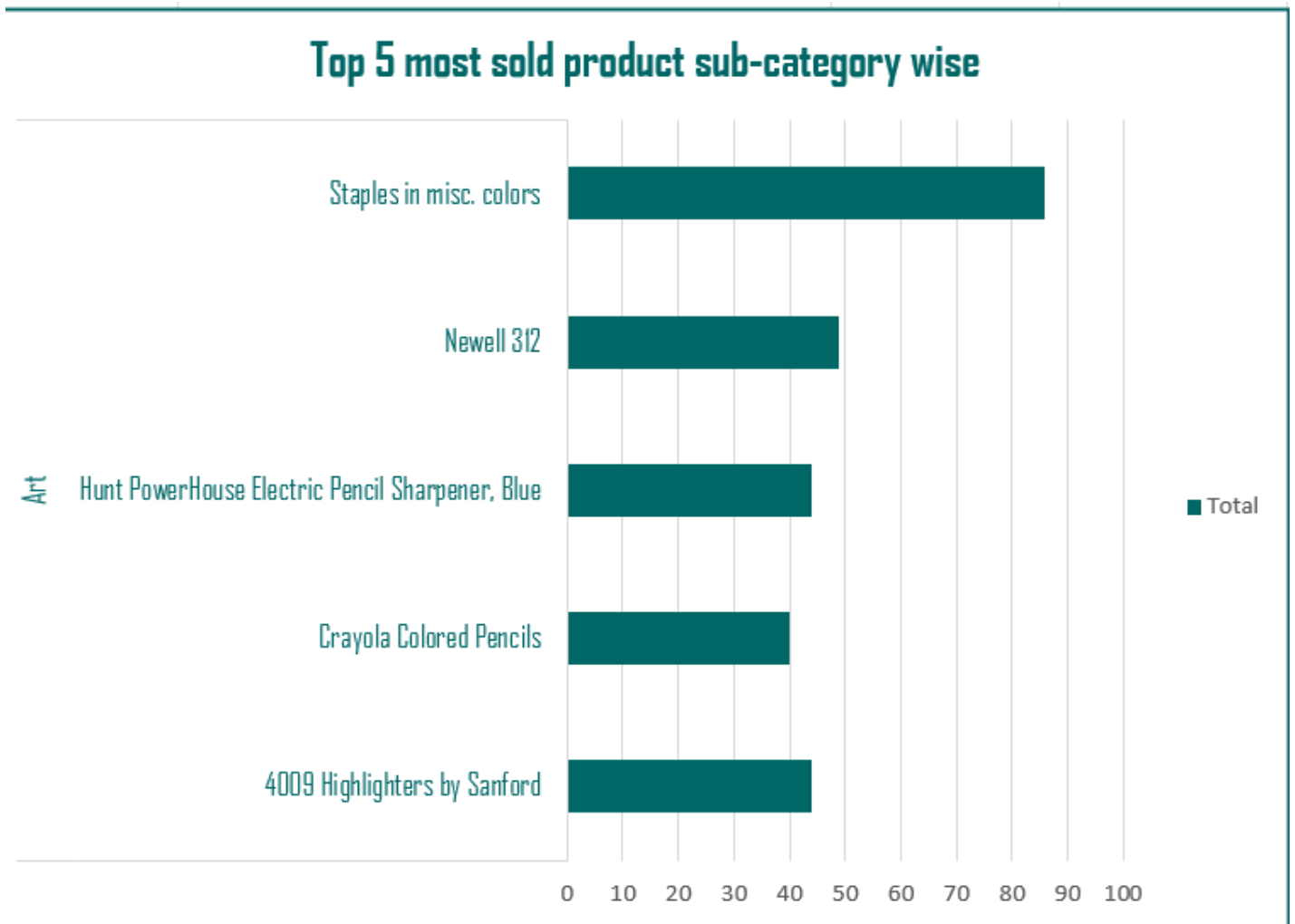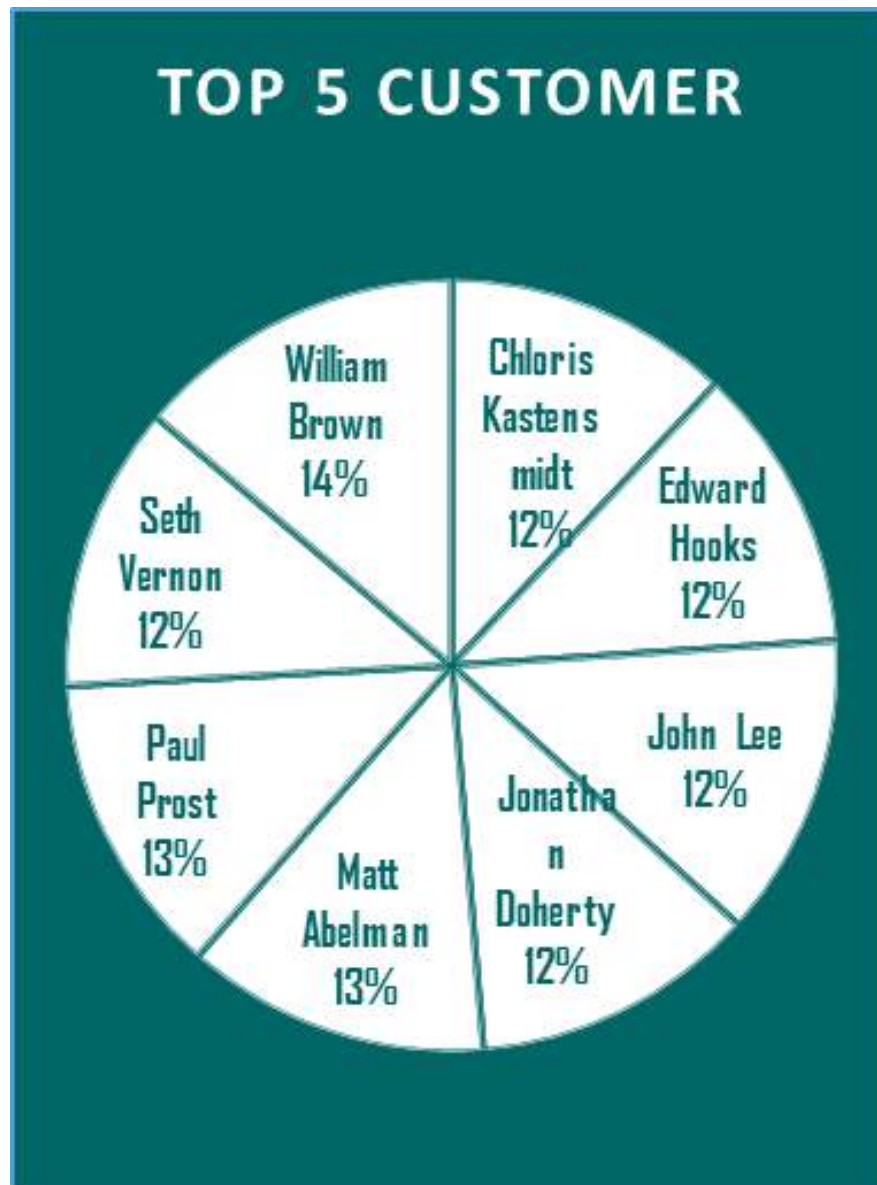Pivot table is used for the analysis.
sum function is used in pivot table for the sum of sales in the pivot table

Analysis results

Row Labels	Sum of Sales
☐ United States	**2297200.86**
Alabama	19510.64
Arizona	35282.001
Arkansas	11678.13
California	457687.6315
Colorado	32108.118
Connecticut	13384.357
Delaware	27451.069
District of Columbia	2865.02
Florida	89473.708
Georgia	49095.84
Idaho	4382.486
Illinois	80166.101
Indiana	53555.36
Iowa	4579.76
Kansas	2914.31
Kentucky	36591.75
Louisiana	9217.03
Maine	1270.53
Maryland	23705.523
Massachusetts	28634.434
Michigan	76269.614
Minnesota	29863.15
Mississippi	10771.34
Missouri	22205.15
Montana	5589.352
Nebraska	7464.93
Nevada	16729.102
New Hampshire	7292.524
New Jersey	35764.312
New Mexico	4783.522
New York	310876.271
North Carolina	55603.164
North Dakota	919.91
Ohio	78258.136
Oklahoma	19683.39
Oregon	17431.15
Pennsylvania	116511.914
Rhode Island	22627.956
South Carolina	8481.71
South Dakota	1315.56
Tennessee	30661.873
Texas	170188.0458
Utah	11220.056
Vermont	8929.37
Virginia	70636.72
Washington	138641.27
West Virginia	1209.824
Wisconsin	32114.61
Wyoming	1603.136

Visualization

For visualization map chart is used

State wise sales trend

Sum of Sales

2297200.86

919.91

138641.27
5589.352
919.91
29863.15
1270...
17431.15
4382.486
32114.61
310876.271
1315.56
762...
116511.914
1603.136
4579.76
78258.1...
16729.102
7464.93
80166.... 5355...
12...
11220.056
32108.118
70636...
457687...
2914.31
22205.15
36591.75
35282.001
4783.522
19683.39
30661.873
55603.164
11678.13
8481.71
1077... 19510.... 49095.84
170188.0458
9217.03
89...

Powered by Bing
© GeoNames, Microsoft, TomTom

Top 5 cities with max profit and max loss

Introduction

By performing this analysis, we will get to know about state wise sales trend.

Description

The analysis based on the country, state, sales of the dataset

Specific requirements, functions, and formulas

Pivot table is used for the analysis.

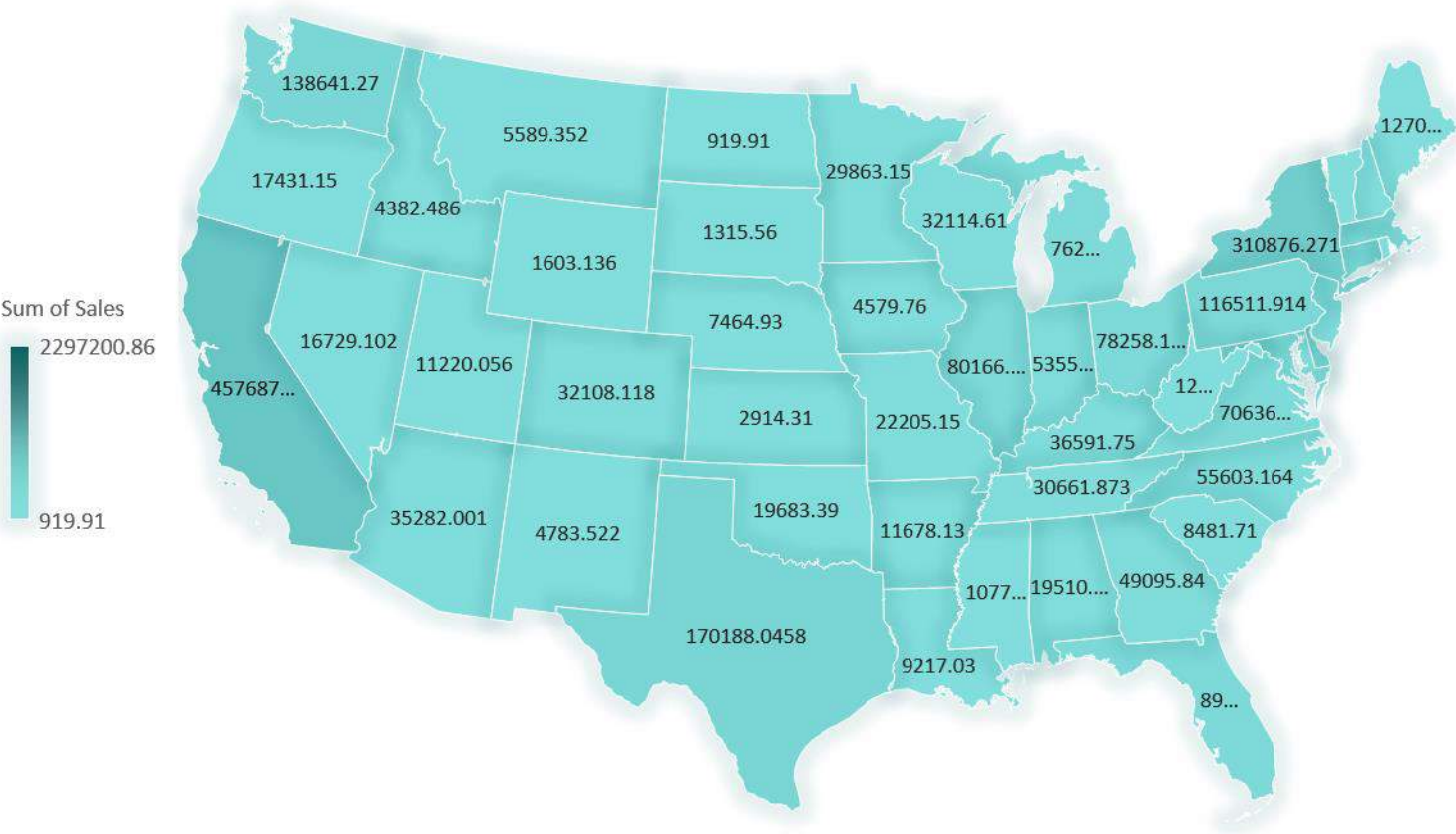sum function is used in pivot table for the sum of sales in the pivot table

Analysis results

CITY NAME	PROFIT
Detroit	$13,182
Los Angeles	$30,441
New York City	$62,037
San Francisco	$17,507
Seattle	$29,156

CITY NAME	LOSS
Philadelphia	-$13,838
Houston	-$10,154
San Antonio	-$7,299
Lancaster	-$7,239
Chicago	-$6,655

Visualization

For visualization pie chart and doughnut char is used respectively.

TOP 5 CITIES WITH MAX LOSS

- Philadelphia: -$13,838
- Houston: -$10,154
- San Antonio: -$7,299
- Lancaster: -$7,239
- Chicago: -$6,655

TOP 5 CITIES WITH MAX PROFIT

- Detroit: $13,182
- Los Angeles: $30,441
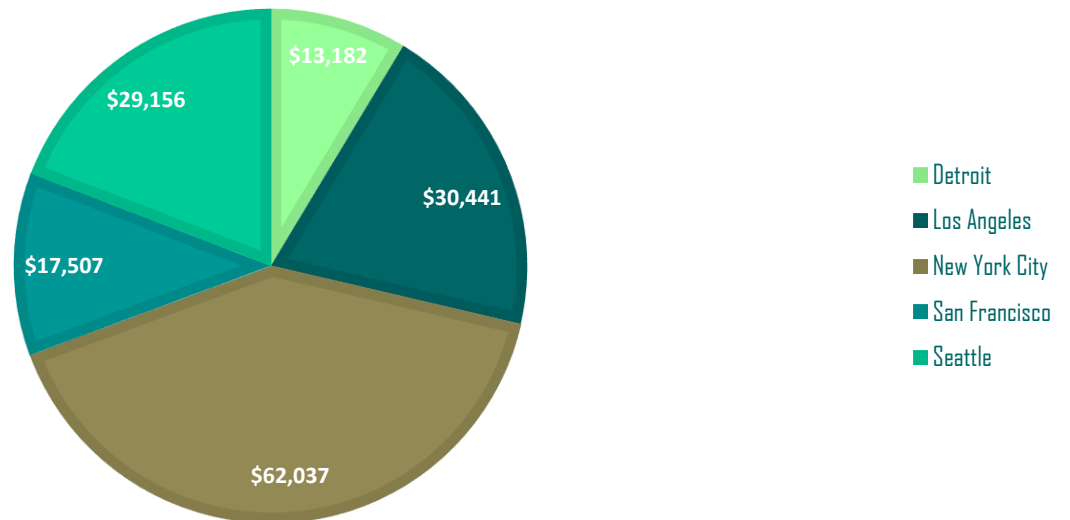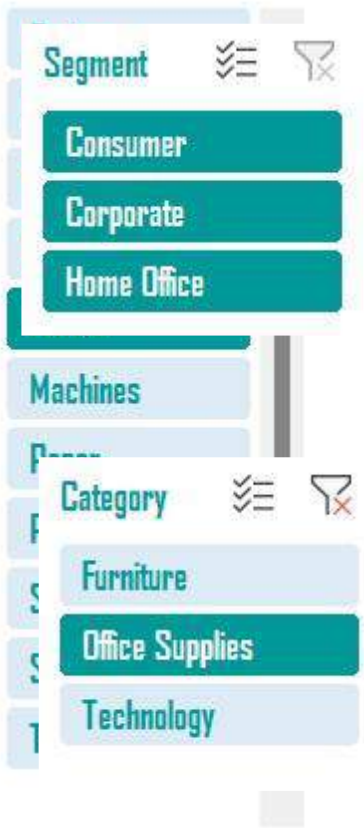- New York City: $62,037
- San Francisco: $17,507
- Seattle: $29,156
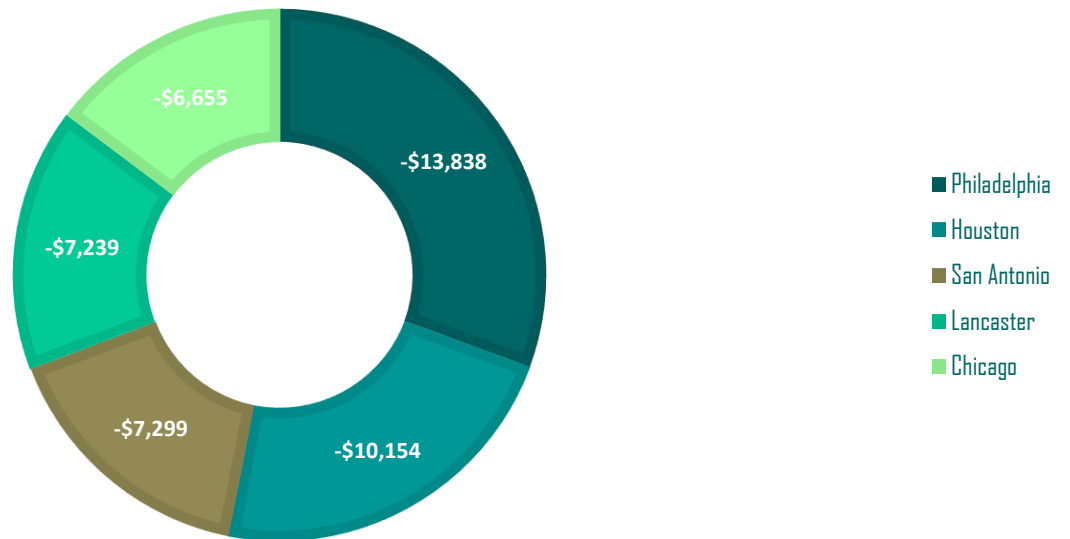
Slicers and Timeline used

For sub-category   For segment   For region

For category

Timeline

Order Date

2014 - 2017   QUARTERS ▾

2014				2015				2016				2017			
Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
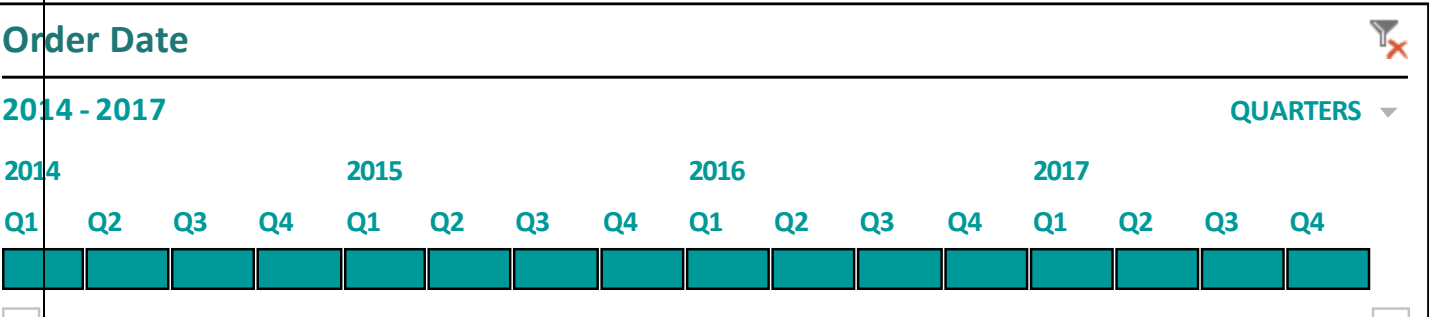
FINAL DASHBOARD

BIBLIOGRAPHY

Dataset source:

> https://docs.google.com/spreadsheets/d/1HnaEEz_K9Ny0XLvOreSBZMyR kT6iMS6X/edit?usp=sharing&ouid=103749046478552787016&rtpof=tr ue&sd=true

Introduction to data management :

> https://www.blue-pencil.ca/what-is-data-management-and-why-it-is- important/