Introduction to Data Management PROJECT REPORT

(Project Semester August-December 2021)

PROJECT REPORT

ON

online store's customer success

Submitted by

M. Vishnu Vardhan

11908323

Programme: B.Tech CSE(Hons.)

Section: KM001

Course Code: INT217

Under the Guidance of

SandeepKaur:23614

Assistant Professor

Discipline of CSE/IT

Lovely School of Computer Science & Engineering

Lovely Professional University, Phagwara

DECLARATION

I, M. Vishnu Vardhan, 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.

M. Vishnu Vardhan

Registration No: 11908323

ACKNOWLEDGEMENT

Primarily I'd thank god for being able to complete my project with success. Then I'd like to thank my mentor **Ms. Sandeep Kaur**, 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.

Last but not least 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 aspects of life.

M. Vishnu Vardhan

CONTENTS:

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 Online store's customer success data
- This Online store's customer success dataset contains 10 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:

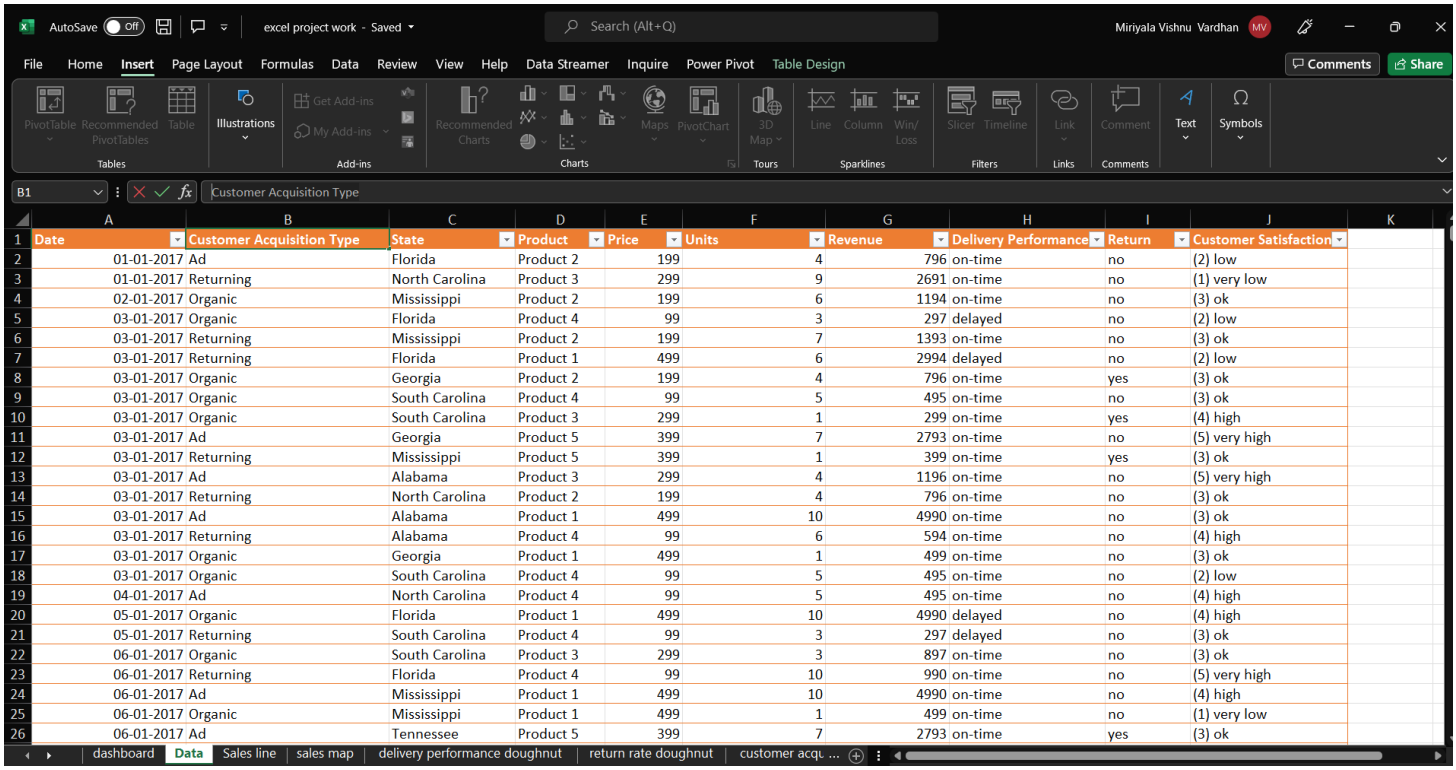
- Total revenues earned by online store in each year
- Total deliveries are on time
- Total count of revenue to the customer acquisition
- customers statsfaction to count of revenue
- sum of revenue to the STATES
- Total number of returns to the count of revenue

DATASET:

The columns included in the dataset are given below:

- Date
- Customer Acquisition Type
- State
- Product type
- Price
- Units
- Revenue
- Delivery performance
- Return
- Customer satisfaction

Sample of dataset with the dat fields:



Date	Customer Acquisition Type	State	Product	Price	Units	Revenue	Delivery Performance	Return	Customer Satisfaction
01-01-2017	Ad	Florida	Product 2	199	4	796	on-time	no	(2) low
01-01-2017	Returning	North Carolina	Product 3	299	9	2691	on-time	no	(1) very low
02-01-2017	Organic	Mississippi	Product 2	199	6	1194	on-time	no	(3) ok
03-01-2017	Organic	Florida	Product 4	99	3	297	delayed	no	(2) low
03-01-2017	Returning	Mississippi	Product 2	199	7	1393	on-time	no	(3) ok
03-01-2017	Returning	Florida	Product 1	499	6	2994	delayed	no	(2) low
03-01-2017	Organic	Georgia	Product 2	199	4	796	on-time	yes	(3) ok
03-01-2017	Organic	South Carolina	Product 4	99	5	495	on-time	no	(3) ok
03-01-2017	Organic	South Carolina	Product 3	299	1	299	on-time	yes	(4) high
03-01-2017	Ad	Georgia	Product 5	399	7	2793	on-time	no	(5) very high
03-01-2017	Returning	Mississippi	Product 5	399	1	399	on-time	yes	(3) ok
03-01-2017	Ad	Alabama	Product 3	299	4	1196	on-time	no	(5) very high
03-01-2017	Returning	North Carolina	Product 2	199	4	796	on-time	no	(3) ok
03-01-2017	Ad	Alabama	Product 1	499	10	4990	on-time	no	(3) ok
03-01-2017	Returning	Alabama	Product 4	99	6	594	on-time	no	(4) high
03-01-2017	Organic	Georgia	Product 1	499	1	499	on-time	no	(3) ok
03-01-2017	Organic	South Carolina	Product 4	99	5	495	on-time	no	(2) low
04-01-2017	Ad	North Carolina	Product 4	99	5	495	on-time	no	(4) high
05-01-2017	Organic	Florida	Product 1	499	10	4990	delayed	no	(4) high
05-01-2017	Returning	South Carolina	Product 4	99	3	297	delayed	no	(3) ok
06-01-2017	Organic	South Carolina	Product 3	299	3	897	on-time	no	(3) ok
06-01-2017	Returning	Florida	Product 4	99	10	990	on-time	no	(5) very high
06-01-2017	Ad	Mississippi	Product 1	499	10	4990	on-time	no	(4) high
06-01-2017	Organic	Mississippi	Product 1	499	1	499	on-time	no	(1) very low
06-01-2017	Ad	Tennessee	Product 5	399	7	2793	on-time	yes	(3) ok

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.

Steps taken to clean dataset thorough ETL process

Step 1

using tableau I have done this process, open tableau connect the excel sheet and click on plus (+) icon, go to clean step

step 2:

Changed data type will change the type of data fields such **as** product and units converted from text into number, Revenue converted from text into currency

Step 3:

checking for null value sand deleting them

step 4 :

calculating year from date using calculated field

step 5:

click on the plus icon and click output

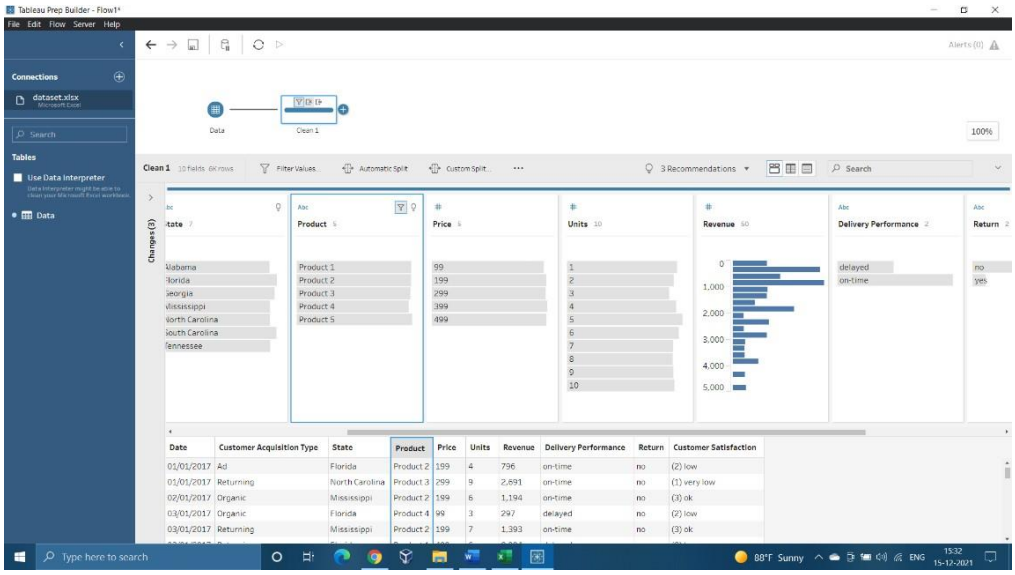


Tableau Prep Builder - Flow1*
File Edit Flow Server Help
Connections
dataset.xlsx
Microsoft Excel
Search
Tables
Use Data Interpreter
Data
Data
Clean 1
100%
Clean 1 10 Fields 6K rows
Filter Values...
Automatic Split
Custom Split...
3 Recommendations
Search
Changes (3)
State 7
Product 5
Price 5
Units 10
Revenue 10
Delivery Performance 2
Return 2
Alabama
Florida
Georgia
Mississippi
North Carolina
South Carolina
Tennessee
Product 1
Product 2
Product 3
Product 4
Product 5
99
199
299
399
499
1
2
3
4
5
6
7
8
9
10
0
1,000
2,000
3,000
4,000
5,000
delayed
on-time
no
yes
Date
Customer Acquisition Type
State
Product
Price
Units
Revenue
Delivery Performance
Return
Customer Satisfaction
01/01/2017 Ad Florida Product 2 199 4 796 on-time no (2) low
01/01/2017 Returning North Carolina Product 3 299 9 2,691 on-time no (1) very low
02/01/2017 Organic Mississippi Product 2 199 6 1,194 on-time no (3) ok
03/01/2017 Organic Florida Product 4 99 3 297 delayed no (2) low
03/01/2017 Returning Mississippi Product 2 199 7 1,393 on-time no (3) ok
Type here to search
88°F Sunny
ENG
15:32
15-12-2021

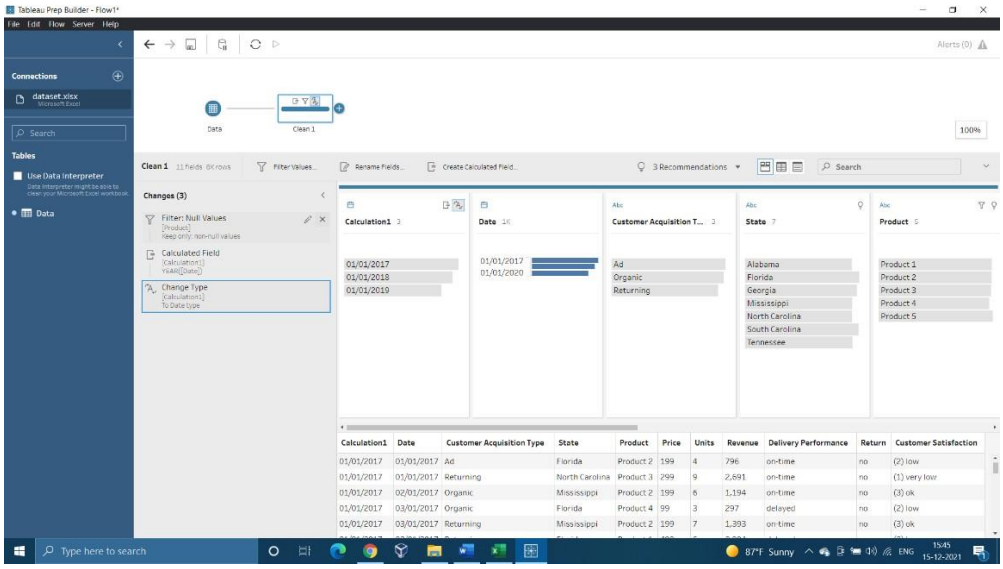
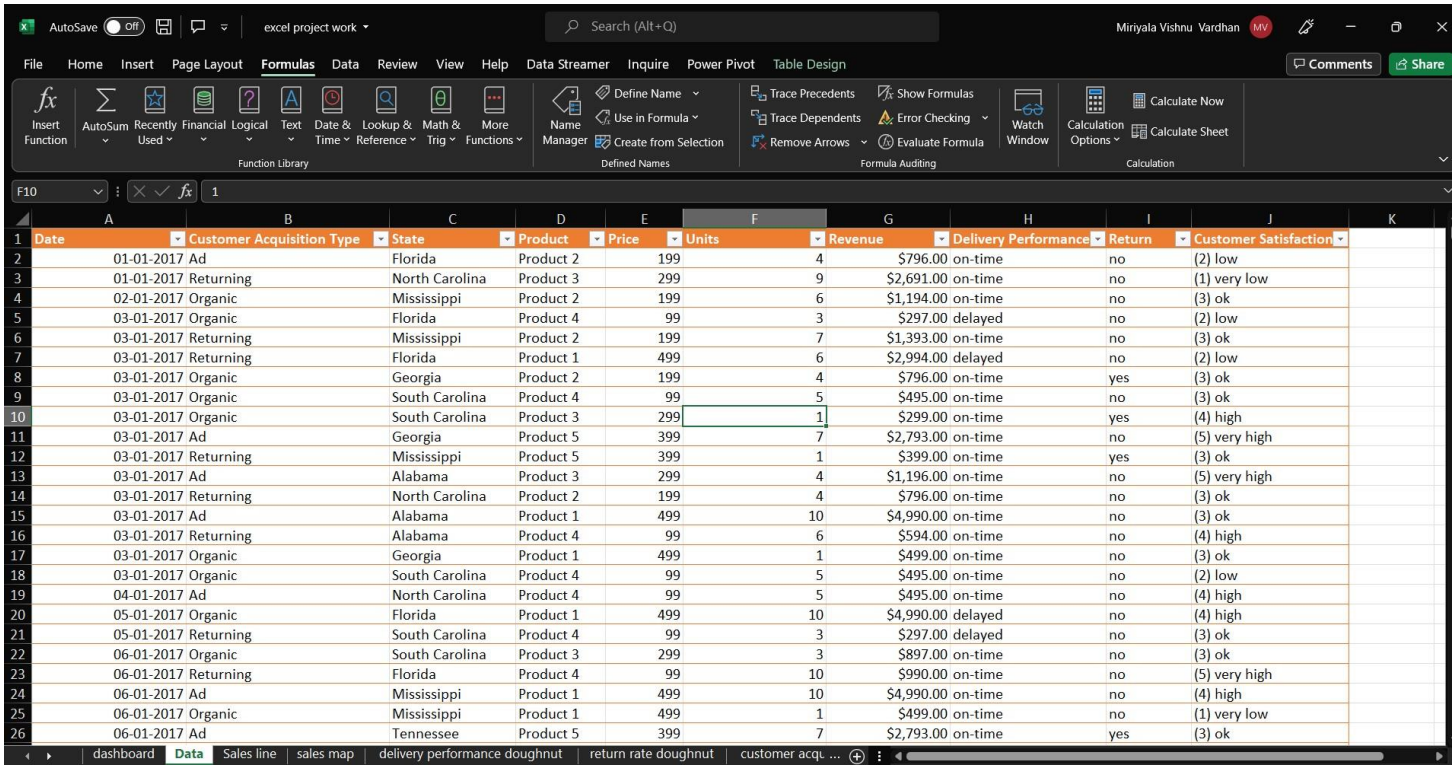


Tableau Prep Builder - Flow1*
File Edit Flow Server Help
Connections
dataset.xlsx
Microsoft Excel
Search
Tables
Use Data Interpreter
Data
Data
Clean 1
100%
Clean 1 11 Fields 6K rows
Filter Values...
Rename Fields...
Create Calculated Field...
3 Recommendations
Search
Changes (3)
Filter: Null Values
[Product]
Keep only: non-null values
Calculated Field
[Calculation1]
YEAR([Date])
Change Type
[Calculation1]
To Date Type
Calculation1 3
Date 1K
Customer Acquisition T... 3
State 7
Product 5
01/01/2017
01/01/2018
01/01/2019
01/01/2017
01/01/2020
Ad
Organic
Returning
Alabama
Florida
Georgia
Mississippi
North Carolina
South Carolina
Tennessee
Product 1
Product 2
Product 3
Product 4
Product 5
Calculation1
Date
Customer Acquisition Type
State
Product
Price
Units
Revenue
Delivery Performance
Return
Customer Satisfaction
01/01/2017 01/01/2017 Ad Florida Product 2 199 4 796 on-time no (2) low
01/01/2017 01/01/2017 Returning North Carolina Product 3 299 9 2,691 on-time no (1) very low
01/01/2017 02/01/2017 Organic Mississippi Product 2 199 6 1,194 on-time no (3) ok
01/01/2017 03/01/2017 Organic Florida Product 4 99 3 297 delayed no (2) low
01/01/2017 03/01/2017 Returning Mississippi Product 2 199 7 1,393 on-time no (3) ok
Type here to search
87°F Sunny
ENG
15:45
15-12-2021

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



Date	Customer Acquisition Type	State	Product	Price	Units	Revenue	Delivery Performance	Return	Customer Satisfaction
01-01-2017	Ad	Florida	Product 2	199	4	$796.00	on-time	no	(2) low
01-01-2017	Returning	North Carolina	Product 3	299	9	$2,691.00	on-time	no	(1) very low
02-01-2017	Organic	Mississippi	Product 2	199	6	$1,194.00	on-time	no	(3) ok
03-01-2017	Organic	Florida	Product 4	99	3	$297.00	delayed	no	(2) low
03-01-2017	Returning	Mississippi	Product 2	199	7	$1,393.00	on-time	no	(3) ok
03-01-2017	Returning	Florida	Product 1	499	6	$2,994.00	delayed	no	(2) low
03-01-2017	Organic	Georgia	Product 2	199	4	$796.00	on-time	yes	(3) ok
03-01-2017	Organic	South Carolina	Product 4	99	5	$495.00	on-time	no	(3) ok
03-01-2017	Organic	South Carolina	Product 3	299	1	$299.00	on-time	yes	(4) high
03-01-2017	Ad	Georgia	Product 5	399	7	$2,793.00	on-time	no	(5) very high
03-01-2017	Returning	Mississippi	Product 5	399	1	$399.00	on-time	yes	(3) ok
03-01-2017	Ad	Alabama	Product 3	299	4	$1,196.00	on-time	no	(5) very high
03-01-2017	Returning	North Carolina	Product 2	199	4	$796.00	on-time	no	(3) ok
03-01-2017	Ad	Alabama	Product 1	499	10	$4,990.00	on-time	no	(3) ok
03-01-2017	Returning	Alabama	Product 4	99	6	$594.00	on-time	no	(4) high
03-01-2017	Organic	Georgia	Product 1	499	1	$499.00	on-time	no	(3) ok
03-01-2017	Organic	South Carolina	Product 4	99	5	$495.00	on-time	no	(2) low
04-01-2017	Ad	North Carolina	Product 4	99	5	$495.00	on-time	no	(4) high
05-01-2017	Organic	Florida	Product 1	499	10	$4,990.00	delayed	no	(4) high
05-01-2017	Returning	South Carolina	Product 4	99	3	$297.00	delayed	no	(3) ok
06-01-2017	Organic	South Carolina	Product 3	299	3	$897.00	on-time	no	(3) ok
06-01-2017	Returning	Florida	Product 4	99	10	$990.00	on-time	no	(5) very high
06-01-2017	Ad	Mississippi	Product 1	499	10	$4,990.00	on-time	no	(4) high
06-01-2017	Organic	Mississippi	Product 1	499	1	$499.00	on-time	no	(1) very low
06-01-2017	Ad	Tennessee	Product 5	399	7	$2,793.00	on-time	yes	(3) ok

Analysis on dataset

Total revenues earned by online store in each year

1. Total revenues earned by online store in each year

Introduction

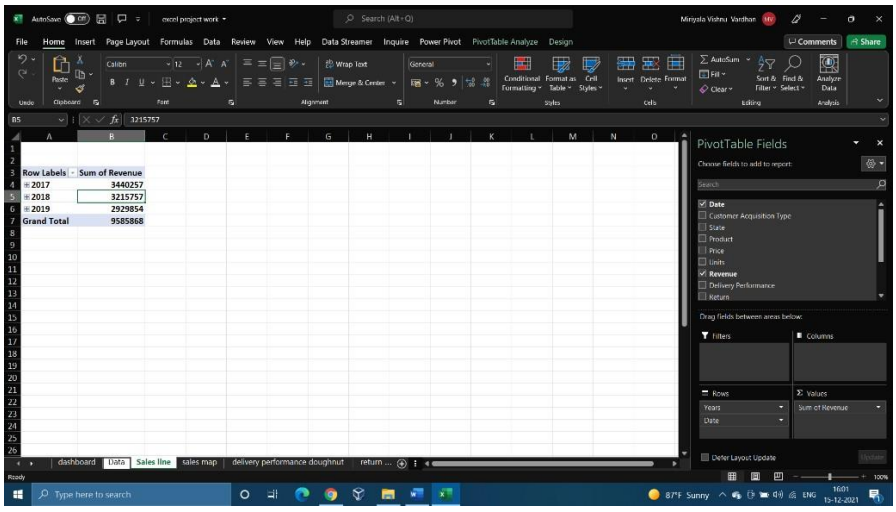
- By performing this analysis, Total revenues earn by the store by each year (months) .

Description

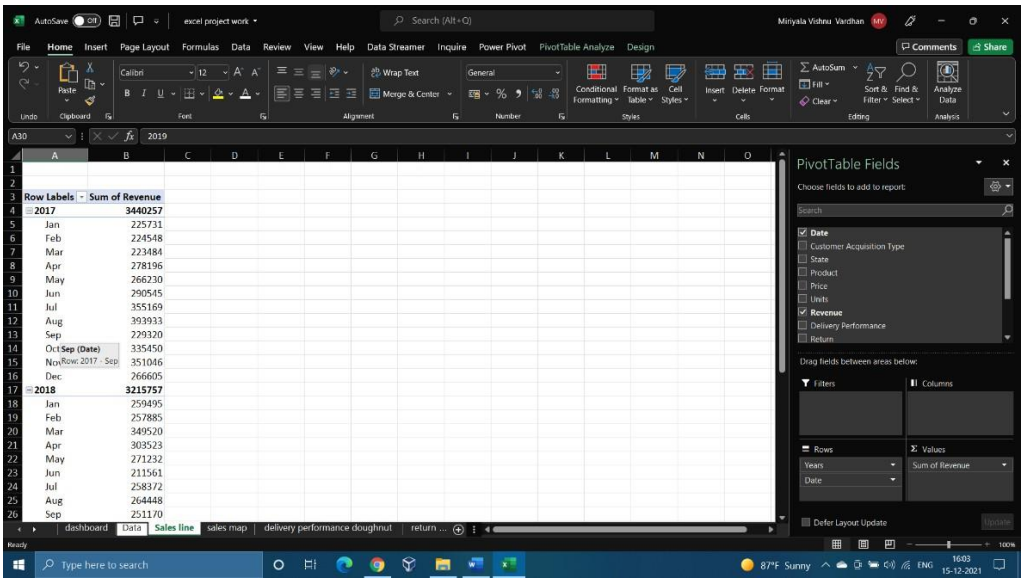
- The analysis is based on date, sum of revenue .

Analysis results

- Pivot table

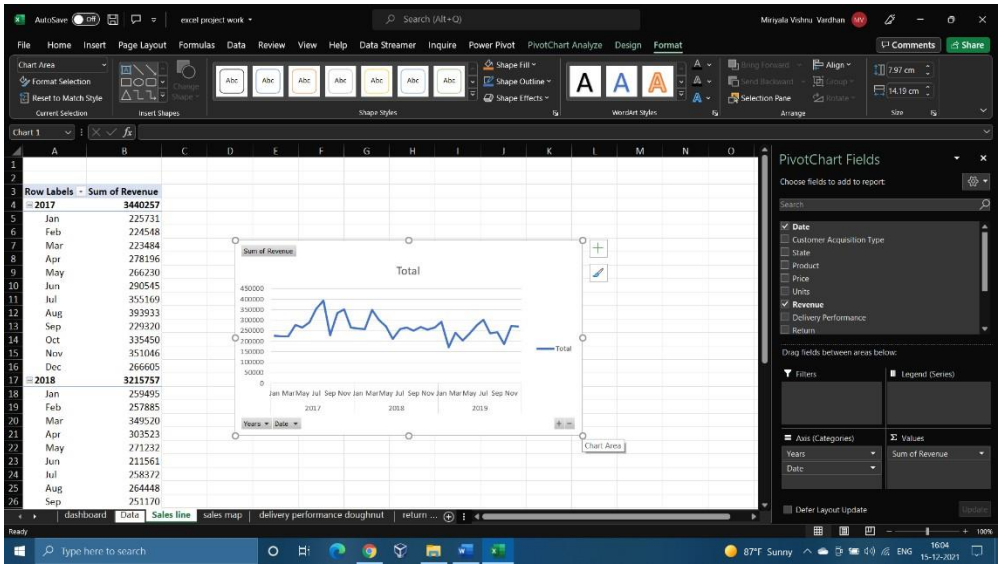


- When the date with years and months



Visualization

- Sales line



2. Total deliveries are in time

Introduction

- By performing this analysis, we will get to know that how many deliveries or on time

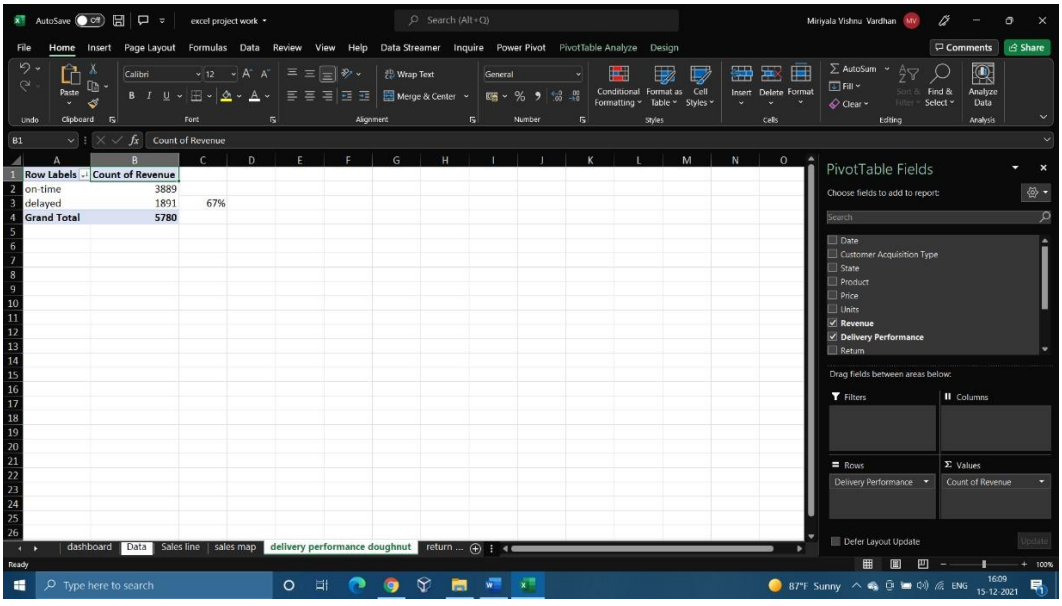
Description

- The analysis based on the delivery performance , revenue of the dataset

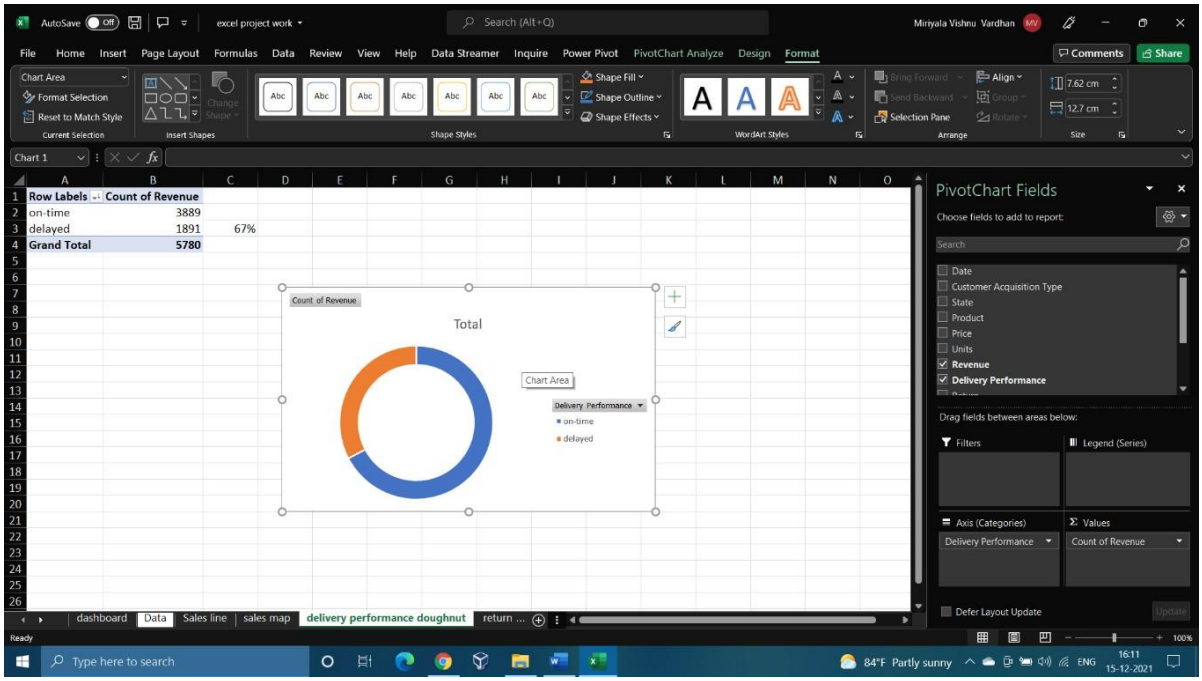
Specific requirements, functions and formulas

- Pivot table is used for the analysis.
- count function is used in pivot table for the count of the revenues in the pivot table
- =GETPIVOTDATA("Revenue",A1,"Delivery Performance","on-time")/GETPIVOTDATA("Revenue",A1)

Analysis results



Visualization



3. Customer acquisition to the count of revenue

Introduction

- By performing this analysis, we will get to know how man are Oraganic, Ad and Returning

Description

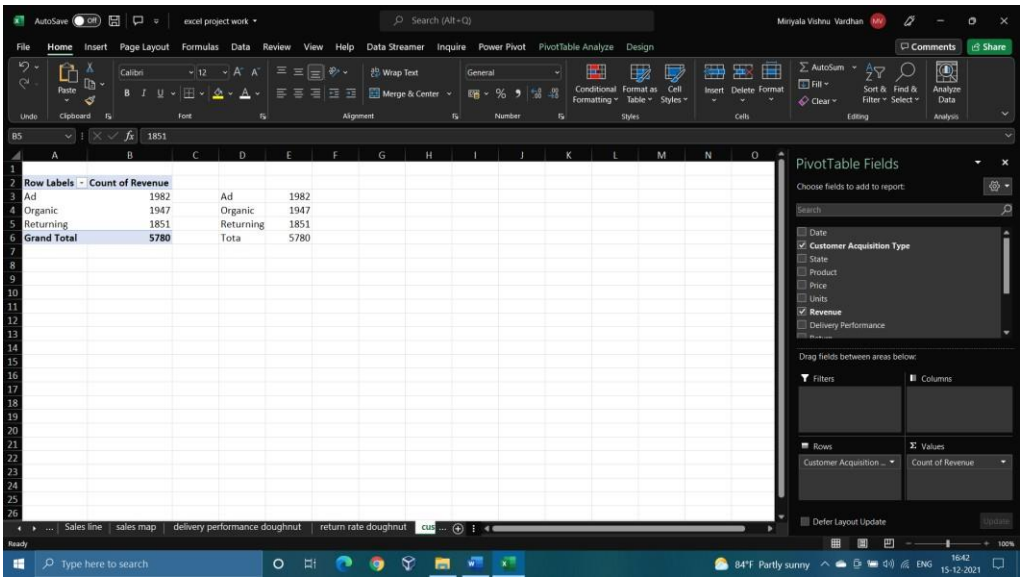
- The analysis based on the customer acquisation, count of revenue

Specific requirements, functions and formulas

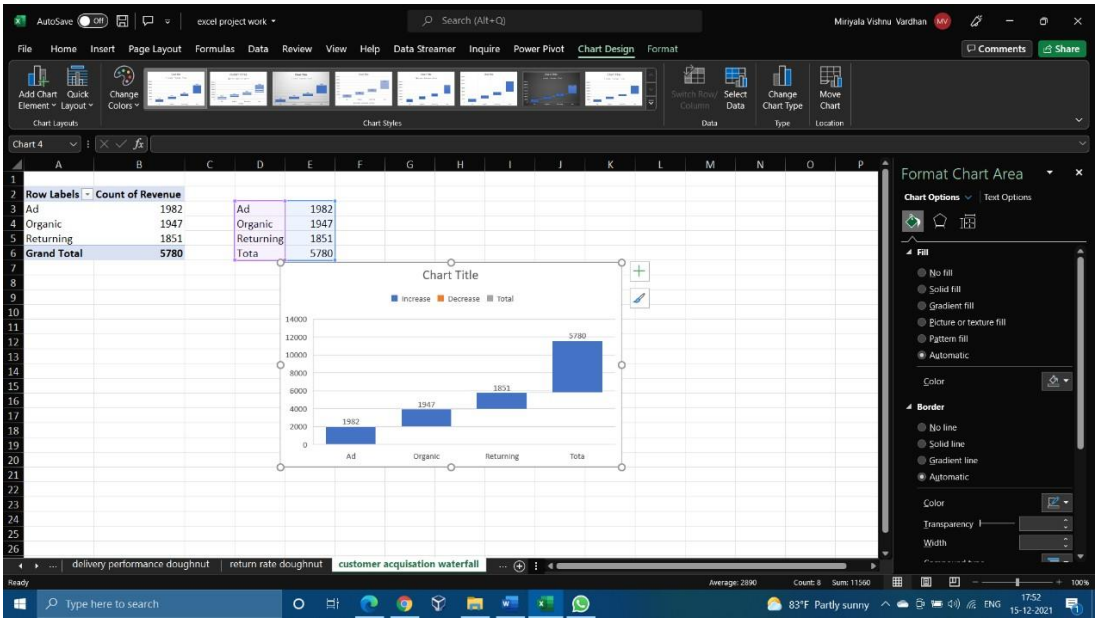
- Pivot table is used for the analysis.

- Count function is used in pivot table for the counting of the title in the pivot table

Analysis results



Visualization



4. Customer satisfaction to count of revenue

Introduction

- By performing this analysis, we will to know that how many liked the product according to product type

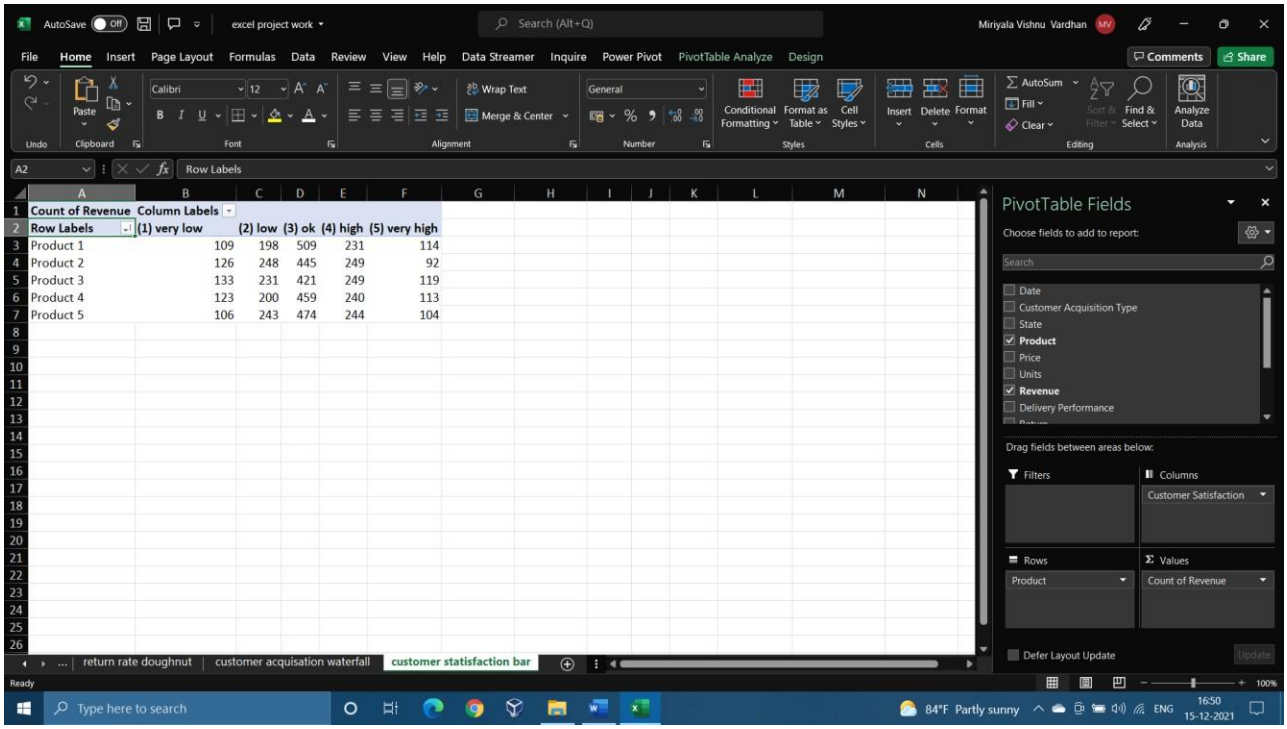
Description

- The analysis is based on customer statisfaction, count of revenue of the dataset

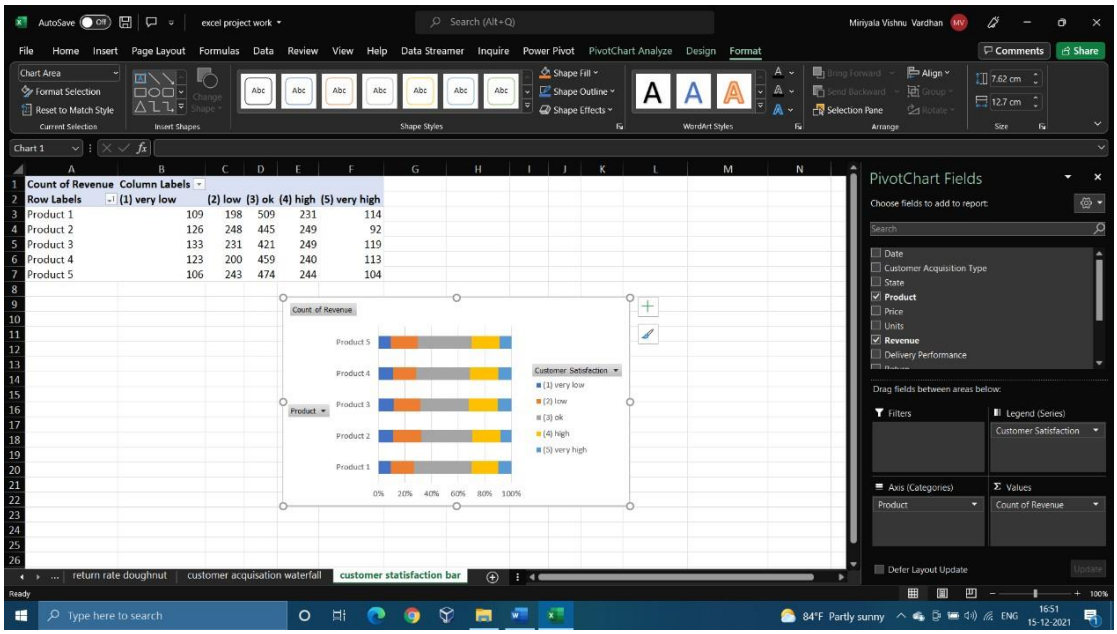
Specific requirements, functions and formulas

- Pivot table is used for the analysis.
- count function is used in pivot table for the count of revenue of the data set

Analysis results



Visualization



5. Sum of the revenue to the states

Introduction

- By performing this analysis, we will get which states has more revenue

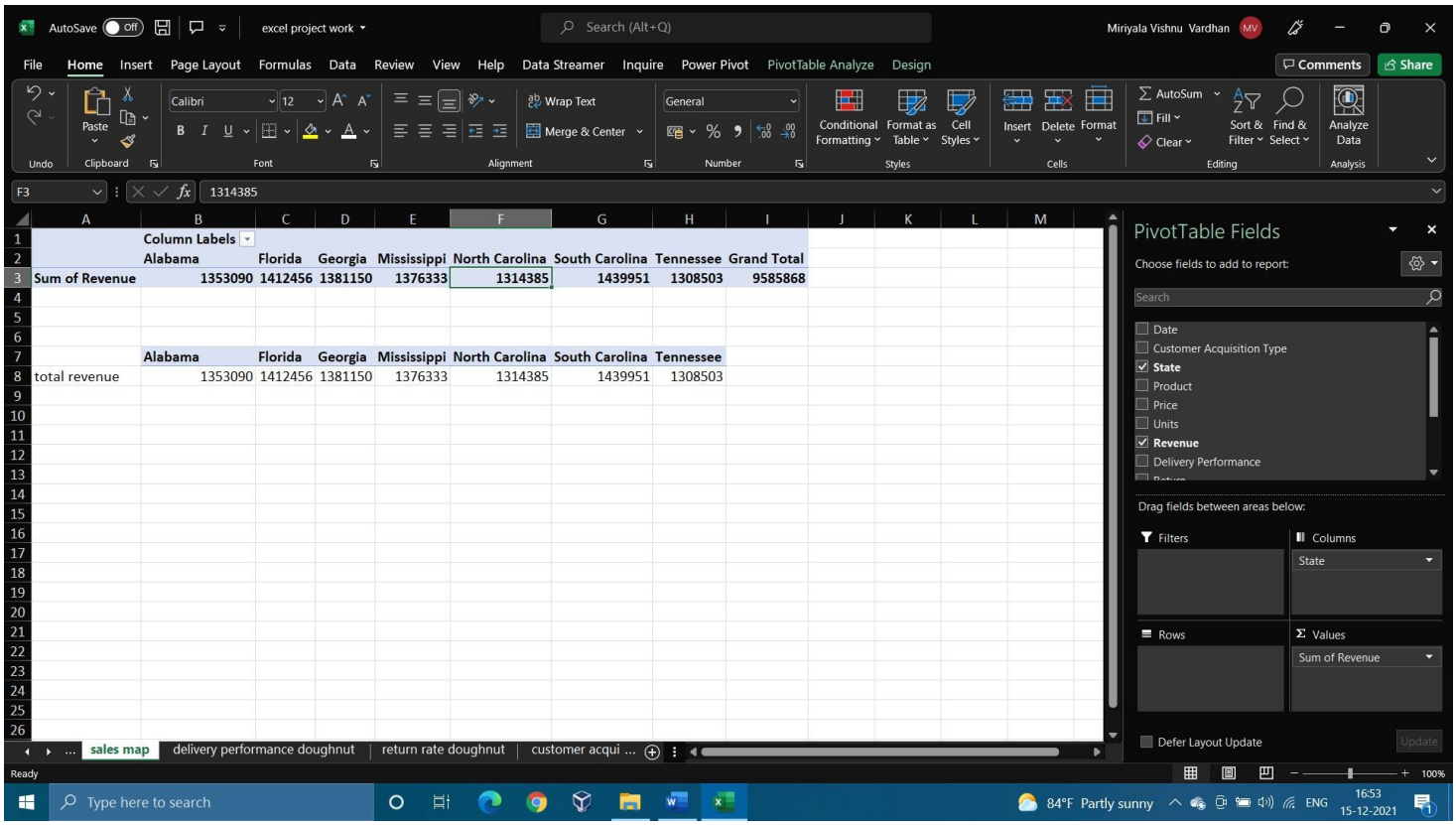
Description

- The analysis is based on state,sum of the revenue

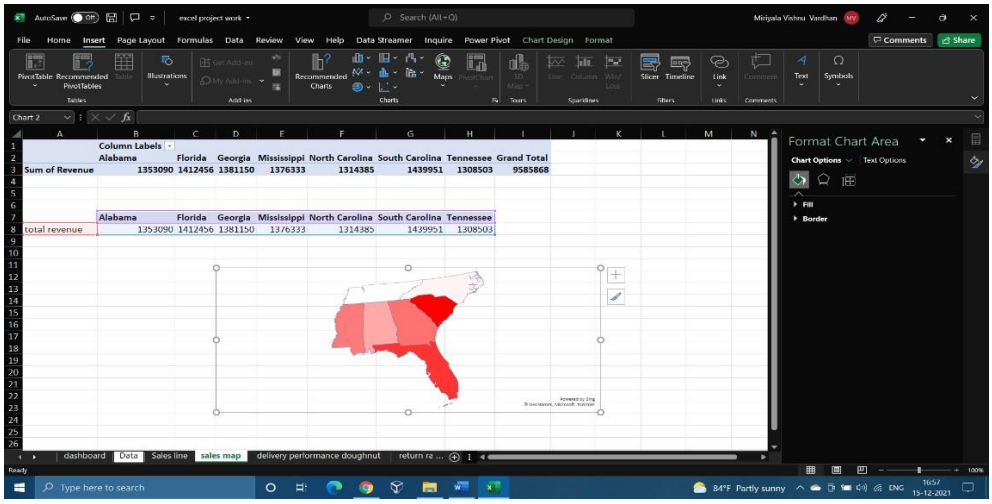
Specific requirements, functions and formulas

- Pivot table is used for the analysis.
- Sum function is used in pivot table for the sum of revenue.

Analysis results



Visualization



6. Return of items

Introduction

- By performing this analysis, we will how many returns are happened

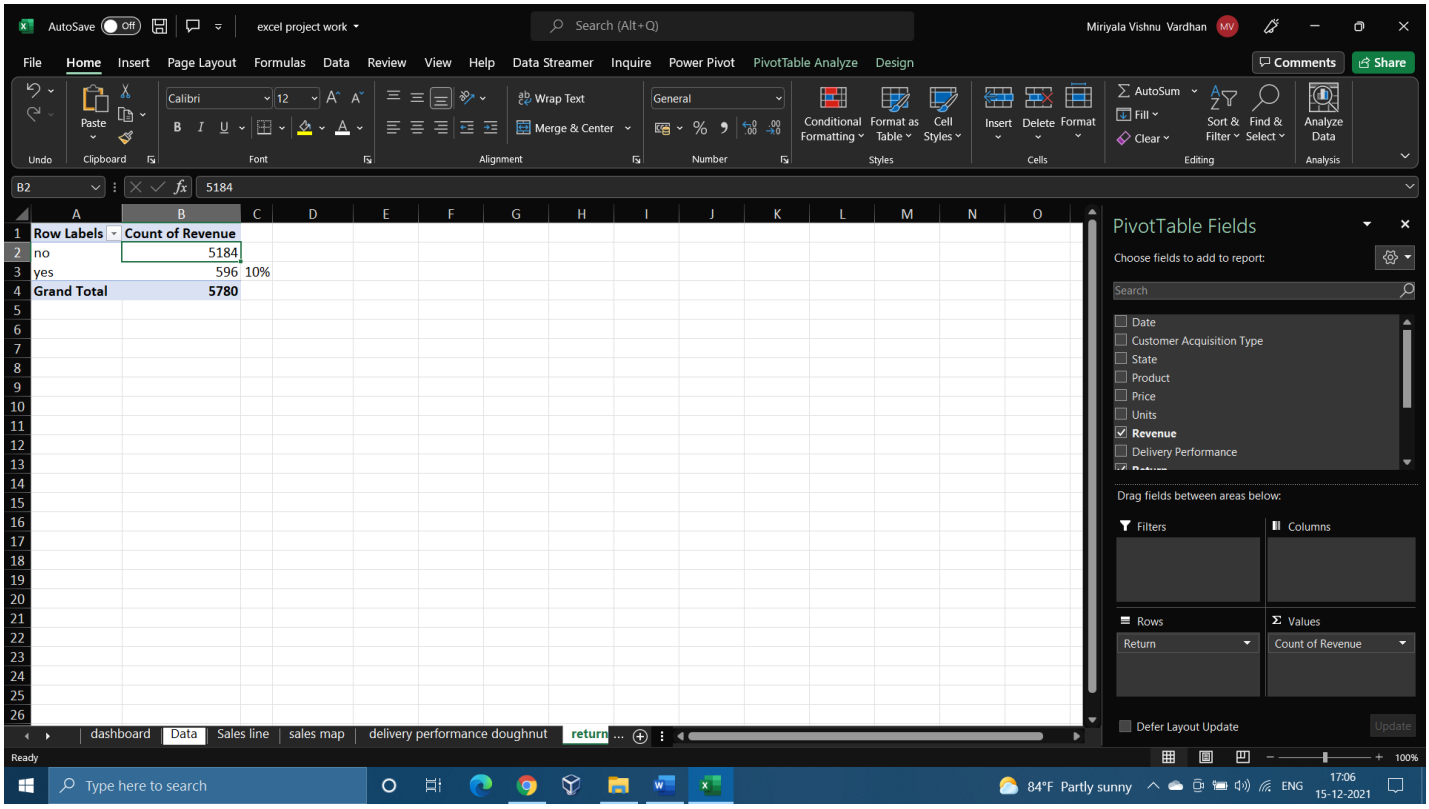
Description

- The analysis is based on return , count of revenue of the dataset

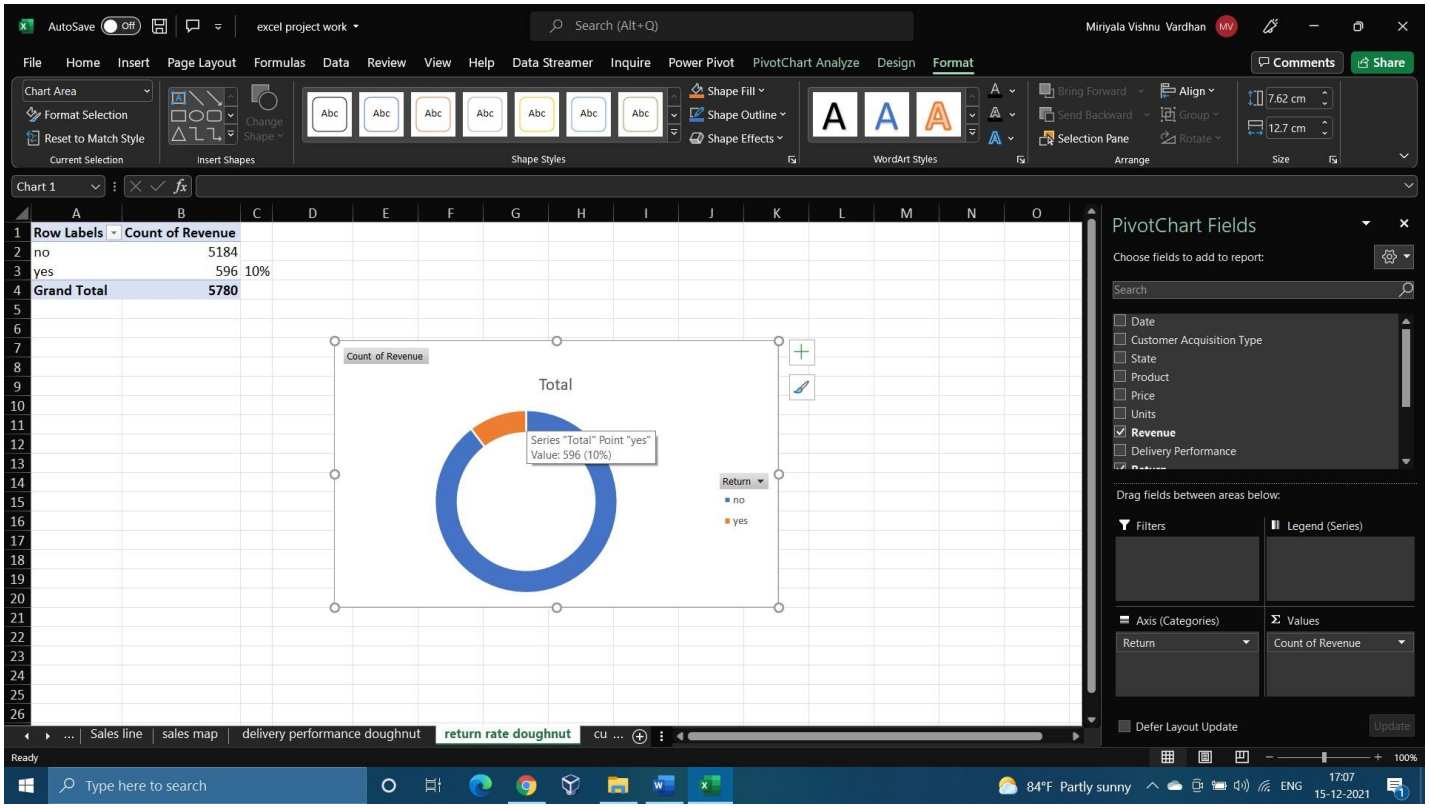
Specific requirements, functions and formulas

- Pivot table is used for the analysis.
- Count function is used in pivot table for the count of the revenue

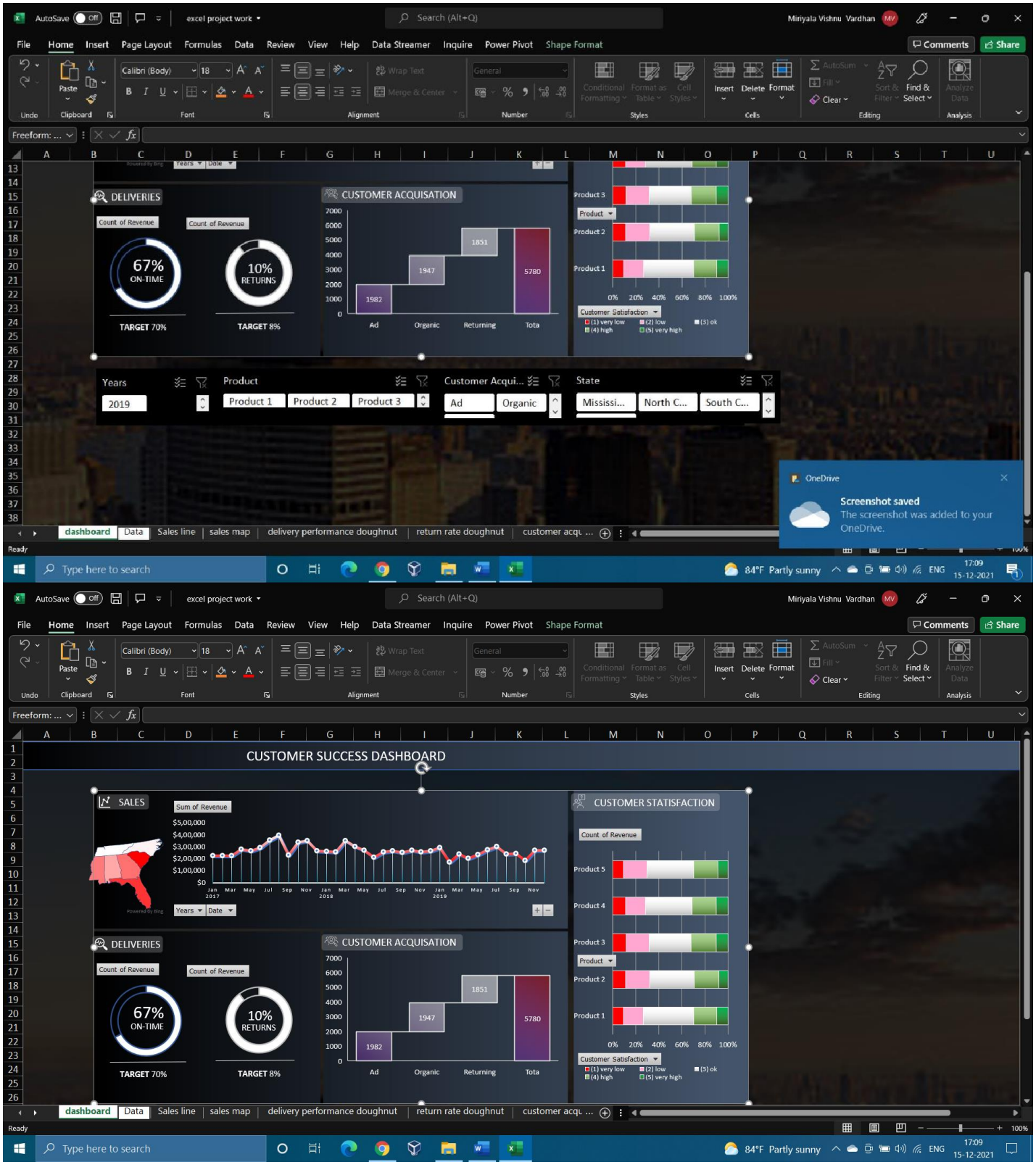
Analysis results



Visualization



Dashboard:


CUSTOMER SUCCESS DASHBOARD
DELIVERIES
67% ON-TIME
10% RETURNS
TARGET 70%
TARGET 8%
CUSTOMER ACQUISATION
SALES
CUSTOMER STATISFACTION
Years
2019
Product
Product 1
Product 2
Product 3
Customer Acqui...
Ad
Organic
State
Mississi...
North C...
South C...
dashboard
Data
Sales line
sales map
delivery performance doughnut
return rate doughnut
customer acqu ...
Ready
OneDrive
Screenshot saved
The screenshot was added to your OneDrive.
Type here to search
84°F Partly sunny
ENG
17:09
15-12-2021

List of analysis:

1. We get know the 67% of the deliveries are on time
2. We get know the 10% of the items are returned
3. The trend of the sales in 3 years
4. Maximum number of sales are in south carolina
5. The result gives the satisfaction of the customers to the products

References

1. www.kaggle.com
2. www.youtube.com
3. www.google.com
4. www.stackoverflow.com
5. www.github.com

Bibliography

1. Microsoft Excel 2016 Bible: The Comprehensive Tutorial Resource by John Walkenbach, Wiley