

DATA MANAGEMENT PROJECT REPORT

(Project Semester: August-December 2020)



L OVELY
P ROFESSIONAL
U NIVERSITY

Shoes Companies Data Analysis

Submitted by

Chirag Agarwal

11804188

Programe and Section: B.Tech (CSE), KM073

Course Code: INT217

Under the Guidance of

Miss. Ashu U.id 23631 Assistant Proffessor

Discipline of CSE/IT

Lovely School of Computer Science & Engineering

DECLARATION

I, Chirag Agarwal, student of B.Tech CSE 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:

Registration No.: 11804188

Chirag Agarwal

ACKNOWLEDGEMENT

I would like to express my special thanks of gratitude to my teacher Miss. Ashu who gave me the golden opportunity to do this wonderful project of analysis of the data of a Shoes Companies in India which also helped me in doing a lot of research and I came to know about so many new things. I am thankful to them. Secondly, I would also like to thank my parents and friends who helped me a lot in finalizing this project within the limited time frame.

CONTENTS OF THE REPORT

1. Introduction
2. Objectives/Scope of the Analysis
3. Source of dataset
4. ETL process
5. Analysis on dataset (for each analysis)
 - i. Introduction
 - ii. General Description
 - iii. Specific Requirements, functions and formulas
 - iv. Analysis results
 - v. Visualization
6. List of Analysis with results
7. References
8. Bibliography

INTRODUCTION

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. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, while being used in different business, science, and social science domains.

The Companies would love to check their performance through data analysis of each and every month, year leading to a well-organized and fruitful information. My analysis contains data on Companies which produce and sell Shoes all over the India and are most successful.

Insights of data:

Shoes Companies in India contains the following data fields: -

- Company : In this data field the name of top shoes companies in India which are:
 - ✓ Nike
 - ✓ Adidas
 - ✓ Clarks
 - ✓ Seeandwear
 - ✓ Woodland
 - ✓ Lee Copper
 - ✓ Fila
 - ✓ Puma
 - ✓ Reebok
 - ✓ Allen Cooper
 - ✓ Timberland
- State : In this data field the name of state from where the shoes are buyied.
- Units : In this data field the number of shoes sold.
- Price: In this data field the price of shoes is provided.
- Total revenue: In this data field the total revenue of the company is provided.
- Offline/Online shopping: In this data field how buyers are buying the shoes i.e. through online mode or offline mode.
- Payment Mode: In this data field by which mode of payment the customer had purchased the Shoes.
- Month: In this data field the month in which the units are sold by each company are provided.
- Year: In this data field the year in which the units are sold by each company are provided.

SCOPE OF ANALYSIS

This project on Shoes companies in India provides the overall Statistics details of the units sold, state, revenue generated of Different companies from the year 2012 to 2016.

Objectives of this project:

- To good hand on excel.
- To use different feature and get friendly with the excel.
- How to use pivot table and pivot chart.
- Learn to make dashboard in excel.
- To make different type of graphs in excel.

Aim of this project is to answer the above objectives in the form of visualization by creating a dashboard to convey the answers effectively and efficiently.

SOURCE OF DATASET

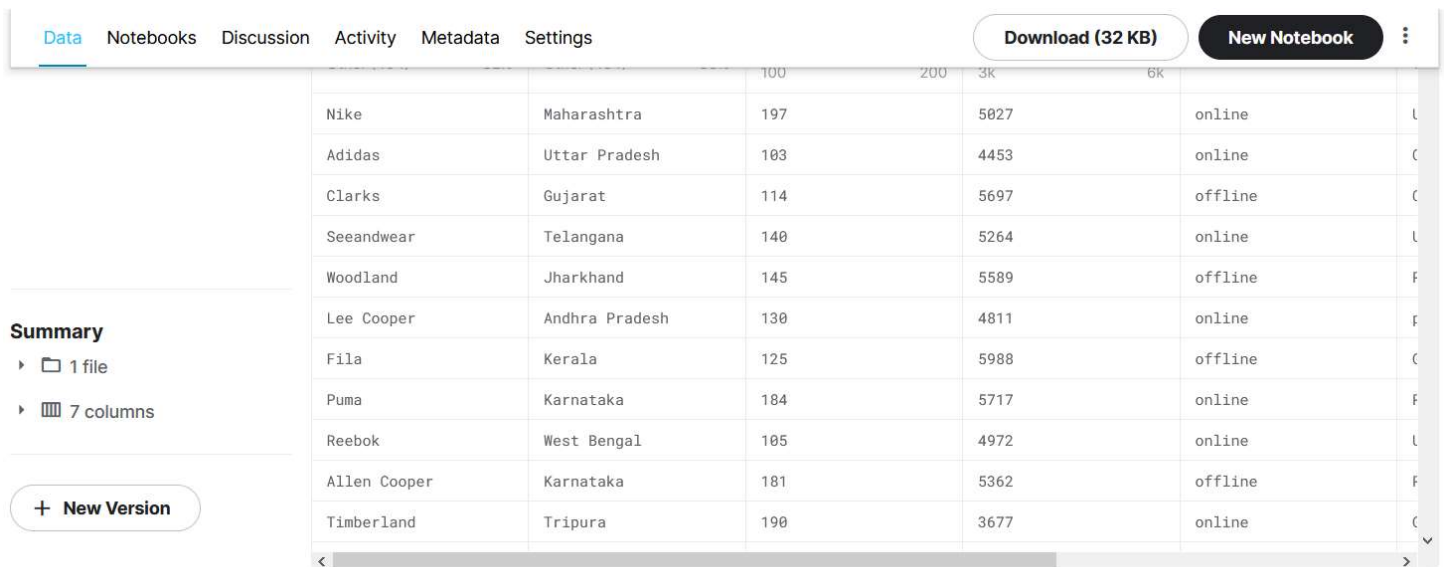
The data is being taken from the Kaggle .Kaggle is an AirBnBfor Data Scientists –this is where they spend their nights and weekends. It's a crowd-sourced platform to attract, nurture, train and challenge data scientists from all around the world to solve data science, machine learning and predictive analytics problems. It has over 536,000 active members from 194 countries and it receives close to 150,000 submissions per month. Started from Melbourne, Australia Kaggle moved to Silicon Valley in 2011, raised some 11 million dollars from the likes of Hal Varian (Chief Economist at Google), Max Levchin (Paypal), Index and Khosla Ventures and then ultimately been acquired by the Google in March of 2017. Kaggle is the number one stop for data science enthusiasts all around the world who compete for prizes and boost theirKaggle rankings. There are only 94 Kaggle Grandmasters in the world to this date.

ETL PROCESS

In computing, extract, transform, load (ETL) is a process in database usage to prepare data for analysis, especially in data warehousing. Data extraction involves extracting data from homogeneous or heterogeneous sources, while data transformation processes data by transforming them into a proper storage format/structure for the purposes of querying and analysis; finally, data loading describes the insertion of data into the final target database such as an operational data store, a data mart, or a data warehouse. A properly designed ETL system extracts data from the source systems, enforces data quality and consistency standards, confirms data so that separate sources can be used together, and finally delivers data in a presentation-ready format so that application developers can build applications and end users can make decisions.

Precisely, 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 stands for Extract, Transform and Load.

Before ETL, the dataset looked like this. **This data is taken from Kaggle.**



The screenshot shows a Kaggle dataset page with a table of shoe brands and their sales data. The table has 7 columns: Brand, State, Sales, and three unlabeled columns. The data is as follows:

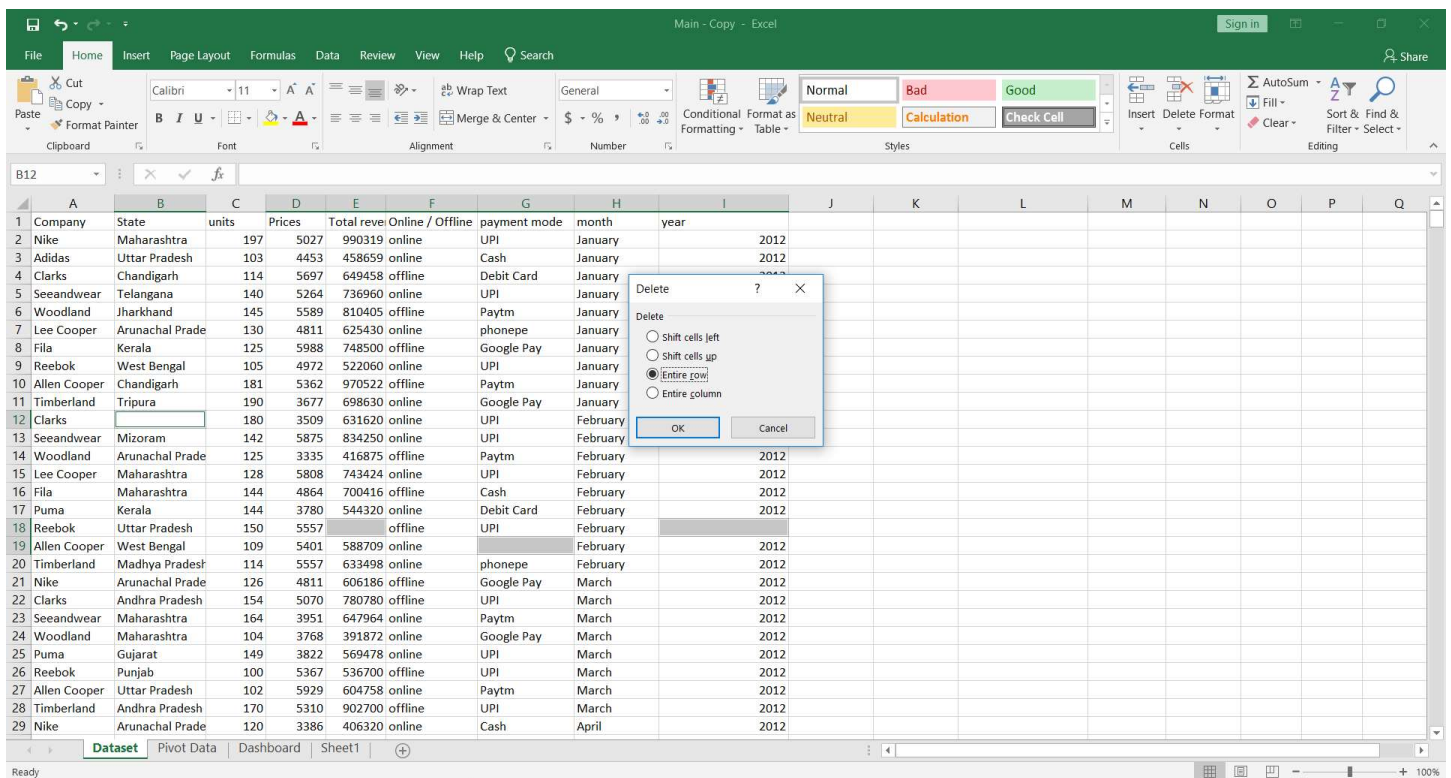
Brand	State	Sales				
Nike	Maharashtra	197		5027	online	
Adidas	Uttar Pradesh	103		4453	online	
Clarks	Gujarat	114		5697	offline	
Seeandwear	Telangana	140		5264	online	
Woodland	Jharkhand	145		5589	offline	
Lee Cooper	Andhra Pradesh	130		4811	online	
Fila	Kerala	125		5988	offline	
Puma	Karnataka	184		5717	online	
Reebok	West Bengal	105		4972	online	
Allen Cooper	Karnataka	181		5362	offline	
Timberland	Tripura	190		3677	online	

Summary: 1 file, 7 columns. + New Version

Through the process of ETL, we are going to clean the dataset and bring all the entities to their proper data format.

Step 1: Removing the blank cells from the dataset.

For this, select the whole dataset. Go to Find and Select in the Home tab of excel. Select Go to Special from the drop-down menu and then tick the blank option. All the blank cells will be selected. Then go to Delete option in the home tab again and select Delete Rows from the drop-down menu. This will remove any rows with blank cells.



Step 2: Removing columns which are not properly defined or not crucial to our analysis.

For this we will columns which are redundant like the column with just the index numbers. For this we will select that particular column and then go to delete option in the home tag and then select Delete Columns from the drop-down menu.

Step 3: Giving proper and appropriate column names.

The dataset does not have proper columns so our next step would be to give proper column names to the columns wherever required.

Company	State	units	Prices	Total reve	Online / Offline	payment mode	month	year
Nike	Maharashtra	197	5027	990319	online	UPI	January	2012
Adidas	Uttar Pradesh	103	4453	458659	online	Cash	January	2012
Clarks	Chandigarh	114	5697	649458	offline	Debit Card	January	2012
Seeandwear	Telangana	140	5264	736960	online	UPI	January	2012
Woodland	Jharkhand	145	5589	810405	offline	Paytm	January	2012
Lee Cooper	Arunachal Prade	130	4811	625430	online	phonepe	January	2012
Fila	Kerala	125	5988	748500	offline	Google Pay	January	2012
Reebok	West Bengal	105	4972	522060	online	UPI	January	2012
Allen Cooper	Chandigarh	181	5362	970522	offline	Paytm	January	2012
Timberland	Tripura	190	3677	698630	online	Google Pay	January	2012
Clarks	Maharashtra	180	3509	631620	online	UPI	February	2012
Seeandwear	Mizoram	142	5875	834250	online	UPI	February	2012
Woodland	Arunachal Prade	125	3335	416875	offline	Paytm	February	2012
Lee Cooper	Maharashtra	128	5808	743424	online	UPI	February	2012
Fila	Maharashtra	144	4864	700416	offline	Cash	February	2012
Puma	Kerala	144	3780	544320	online	Debit Card	February	2012
Reebok	Uttar Pradesh	150	5557	833550	offline	UPI	February	2012
Allen Cooper	West Bengal	109	5401	588709	online	Paytm	February	2012
Timberland	Madhya Pradesh	114	5557	633498	online	phonepe	February	2012
Nike	Arunachal Prade	126	4811	606186	offline	Google Pay	March	2012
Clarks	Andhra Pradesh	154	5070	780780	offline	UPI	March	2012
Seeandwear	Maharashtra	164	3951	647964	online	Paytm	March	2012
Woodland	Maharashtra	104	3768	391872	online	Google Pay	March	2012
Puma	Gujarat	149	3822	569478	online	UPI	March	2012
Reebok	Punjab	100	5367	536700	offline	UPI	March	2012
Allen Cooper	Uttar Pradesh	102	5929	604758	online	Paytm	March	2012
Timberland	Andhra Pradesh	170	5310	902700	offline	UPI	March	2012

ANALYSIS OF DATASET

1. Companies which have highest number of unit Sold:

Description:

In this objective, we will be finding the number of units sold by each company. For this we will be using the total number of Companies and the units of shoes sold by them. It will result in top 5 shoes companies in India have highest number of shoes sold and preferred by customers.

Specific function and requirements

We have to create a pivot table to determine the difference in units of shoes sold by the companies and then visualize it on graph.

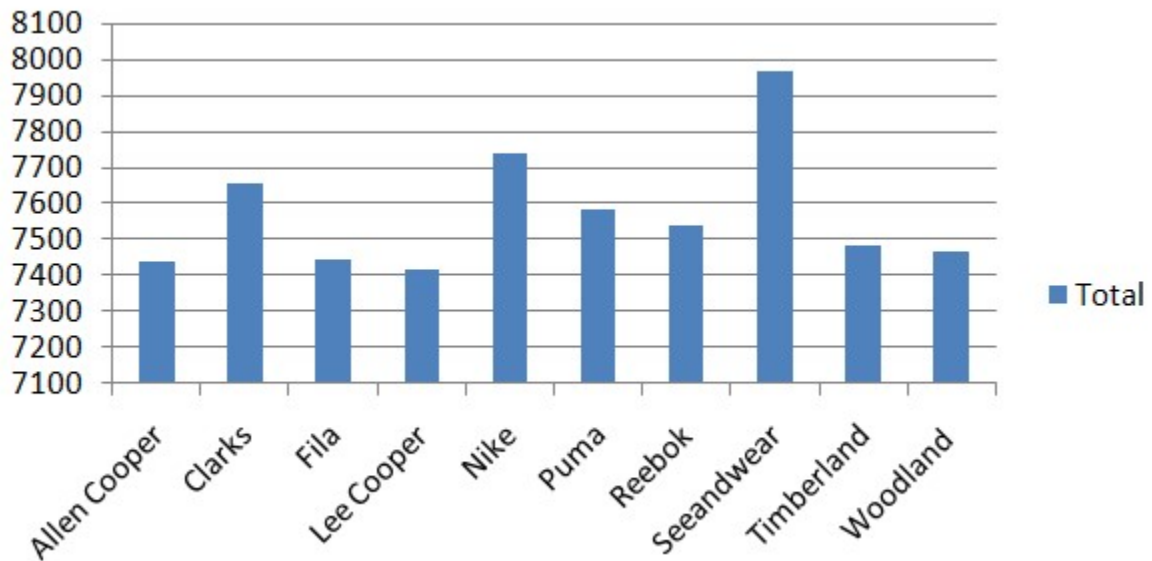
Result:

Company	Units sold
Seeandwear	7965
Nike	7741
Clarks	7655
Puma	7584
Reebok	7541
Timberland	7486
Woodland	7466
Fila	7442
Allen Cooper	7440
Lee Cooper	7415
Grand Total	75735

According to the analysis the units sold by the Seeandwear is the highest with total of 7965 unit from 2012 to 2016. On Second the Nike is preferred by the people of India with 7741 units sold during 2012 to 2016. On Third Clark shoes are worn by the Indian People with the 7655 units sold. On forth Puma is proffered with around 7584 units sold. On fifth Reebok is chosen by the people of India with 7541 unit sold during 2012 to 2016 and so on.

Visualization:

The results are then visualized with the help of graph between Companies and total number of units sold.



2. States where people purchase highest number of units shoes In India during 2012 and 2016.

Description:

India is a vast country with 33 states and 7 union territories. In this objective we will be finding the total number of shoes purchase by different people in different states of India.

Specific function and requirements

In this we need to Create pivot table between the state and units sold and then visualize on the graph.

Result:

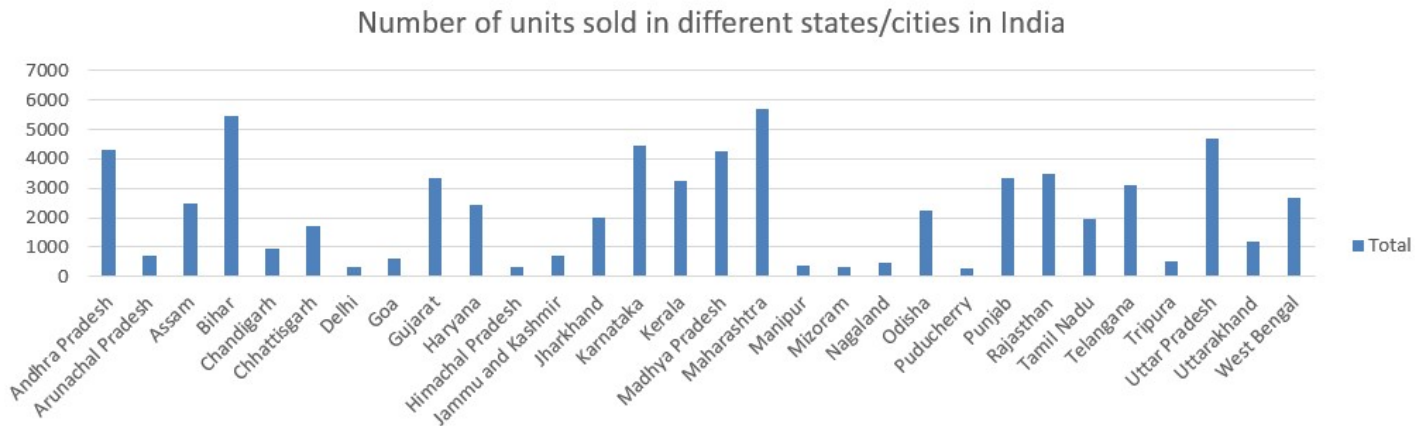
Pivot table between state and units sold

State	Units Sold
Andhra Pradesh	5708
Arunachal Pradesh	116
Assam	2905
Bihar	7044
Chandigarh	196
Chhattisgarh	2184
Delhi	331
Goa	528
Gujarat	4809
Haryana	2600
Himachal Pradesh	309
Jammu and Kashmir	696
Jharkhand	2745
Karnataka	5278
Kerala	3805
Madhya Pradesh	5159
Maharashtra	6713
Manipur	489
Meghalaya	110
Mizoram	306
Nagaland	476
Odisha	2689
Puducherry	259
Punjab	4779
Rajasthan	4295
Tamil Nadu	3082
Telangana	4099
Tripura	799
Uttar Pradesh	5816
Uttarakhand	1343

West Bengal	3347
(blank)	
Grand Total	83015

Visualization:

The results are then visualized with the help of graph between State and total number of units sold.



3. Revenue generated through online/offline shopping

Description:

Now a days online shopping is one of the important thing. Everyone who want to buy anything, then he/she can check price and even buy. Sometimes during festival's time, many companies gives great offers and coupons to the customers. Which benefits the customers a lot, so it becomes very important to keep track on the revenue generated and units sold and region covered through online and offline shopping.

Specific function and requirements

In this we need to Create pivot table between the revenue generated and mode of shopping i.e. through online or offline shopping.

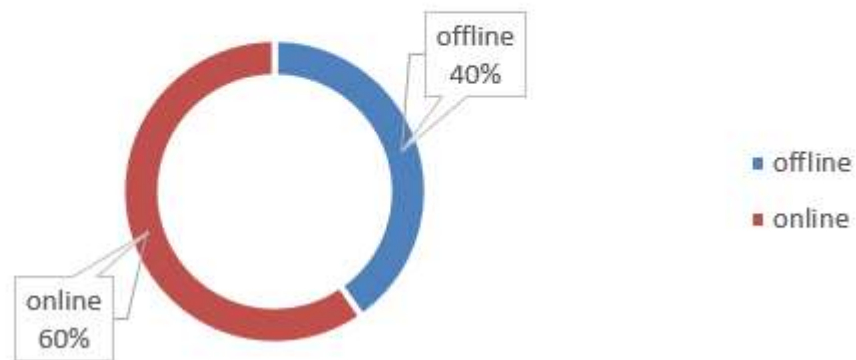
Pivot table between revenue generated and online and offline mode of shopping

Online / Offline Shopping	Revenue Generated
offline	827379
online	1225070
Grand Total	2052449

Visualization:

The results are then visualized with the help of graph between Offline and online shopping and total revenue generated through the online and offline mode of shopping.

Revenue Generated from Online / Offline Shopping



4. Share of revenue generated by companies from 2012-2016 in percent,

Description:

In India there are many shoes company and it is very important to keep track on which company is on top as per the revenue generated. So In this we will be analyzing the Share of revenue generated by company.

Specific function and requirements

In this we will be creating the pivot table between companies and share of total revenue generated by companies between 2012 to 2016.

Formula used:

Total revenue = Unit sold x Price of each unit

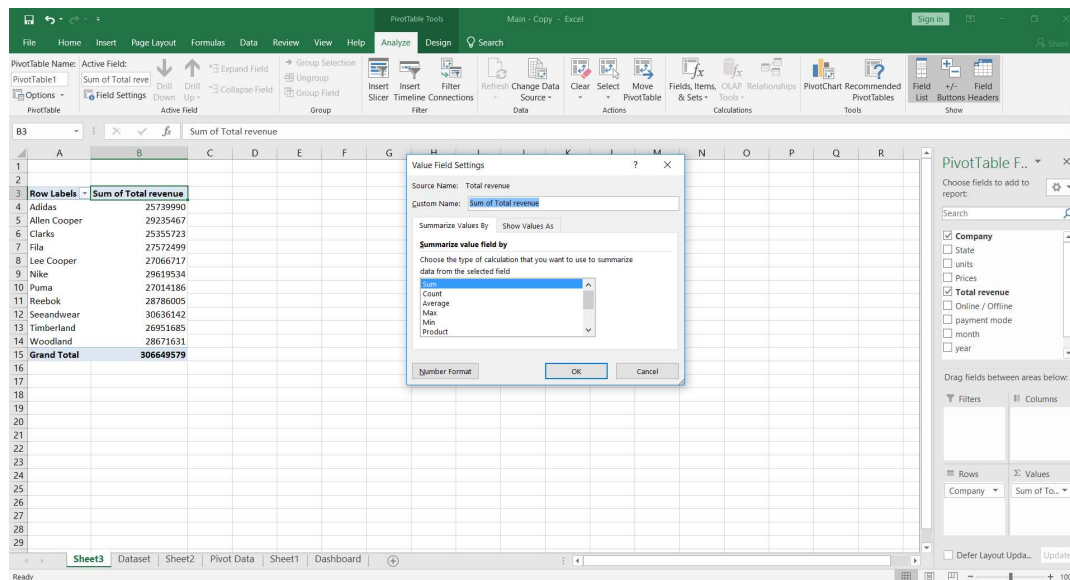
To calculate the share of revenue generated by company

Step 1:

Create the pivot table between companies and revenue generated

Step 2:

Double click on the header of revenue generated then the another option of window field settings will open



Step 3

After double clicking click on the show value as tab.

Value Field Settings ? X

Source Name: Total revenue

Custom Name: Sum of Total revenue

Summarize Values By Show Values As

Summarize value field by

Choose the type of calculation that you want to use to summarize data from the selected field

Sum
Count
Average
Max
Min
Product

Number Format OK Cancel

Step4:

After click show value as select % of total column by clicking on the dropdown show value as.

Value Field Settings ? X

Source Name: Total revenue

Custom Name: Sum of Total revenue

Summarize Values By Show Values As

Show values as

% of Column Total

Base field:

Company
State
units
Prices
Total revenue
Online / Offline

Base item:

Number Format OK Cancel

Step 5:

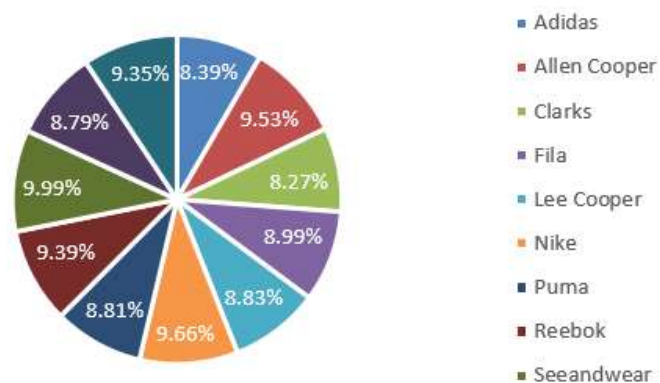
now the revenue generate is converted to the total share.

Company	Share of Total revenue
Adidas	8.39%
Allen Cooper	9.53%
Clarks	8.27%
Fila	8.99%
Lee Cooper	8.83%
Nike	9.66%
Puma	8.81%
Reebok	9.39%
Seeandwear	9.99%
Timberland	8.79%
Woodland	9.35%
Grand Total	100.00%

Visualization:

In this we will be creating graph between Company and share of total revenue generated.

Share of total revenue by Company



5. Unit sold by the companies by year;

Description:

It is very essential to keep track on the total number of units sold by companies by year. From this companies can analyze that in which year the selling was the most and take effective measurements and solution to increase the selling. In this we will be analyze that In which year the selling was the most and compare the selling with other years.

Specific function and requirements

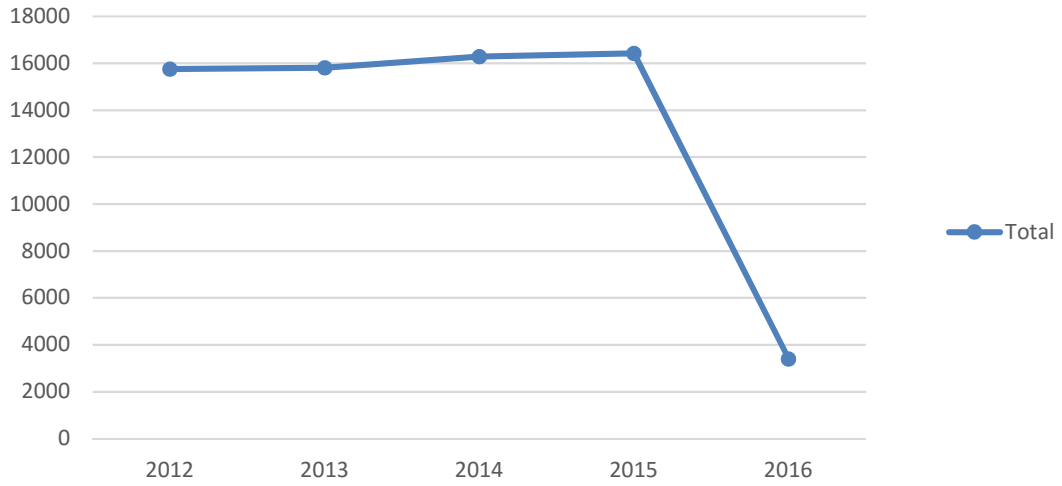
In this we will be creating the pivot table between Unit sold and year.

year	Units Sold
2012	15756
2013	15813
2014	16290
2015	16422
2016	3405
Grand Total	67686

Visualization:

We will analyze the result with the help of graph between unit sold and year

Unit sold from 2012-2016



List of analysis with results

In India the top 5 companies which sold maximum shoes are :

1. seeandwear with 6557 Units sold during 2012-2016
2. Reebok with 6284 Units sold
3. Woodland with 6278 units sold
4. alien cooper with 6187 units sold
5. Clarks with 6181 units sold

Company	Units Sold
Seeandwear	6557
Reebok	6284
Woodland	6278
Allen Cooper	6187
Clarks	6181
Grand Total	31487



Pair of shoes sold between 2012-2016

During 2015 the number of shoes sold is maximum with total selling of 16422 while least in 2016 with total sell 3405.

Now the list of shoes sold in different year in India

2012 - 15756

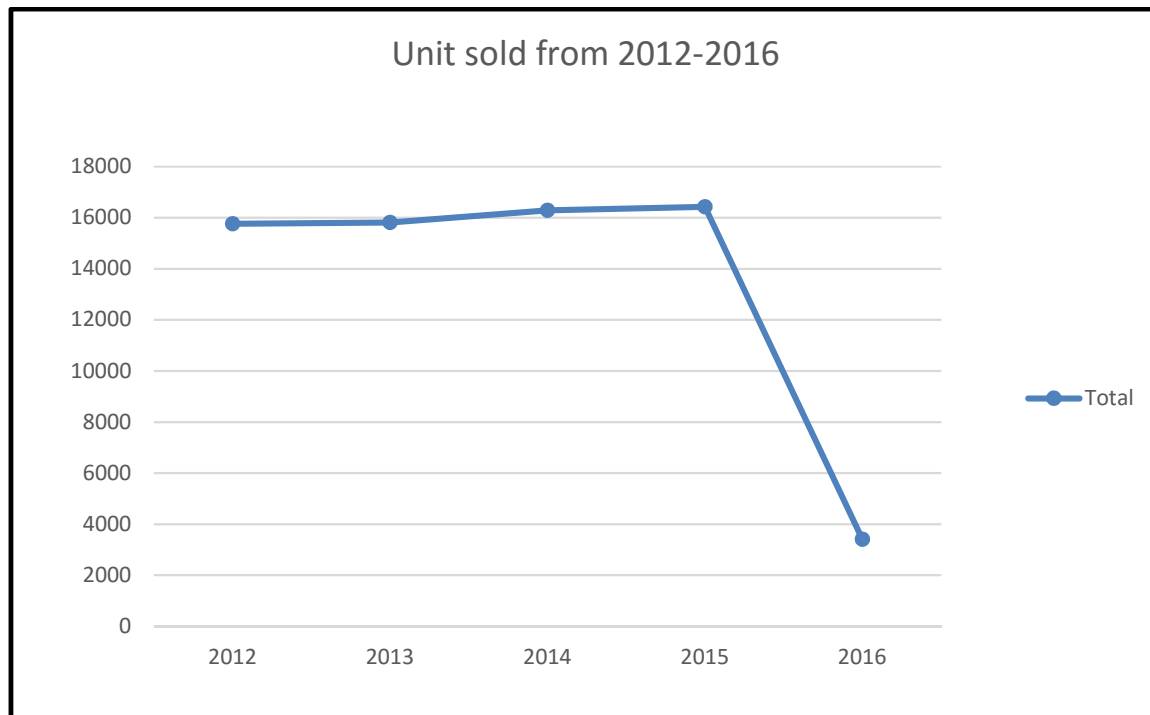
2013 – 15813

2014 – 16290

2015 – 16422

2016 – 3405 (till march)

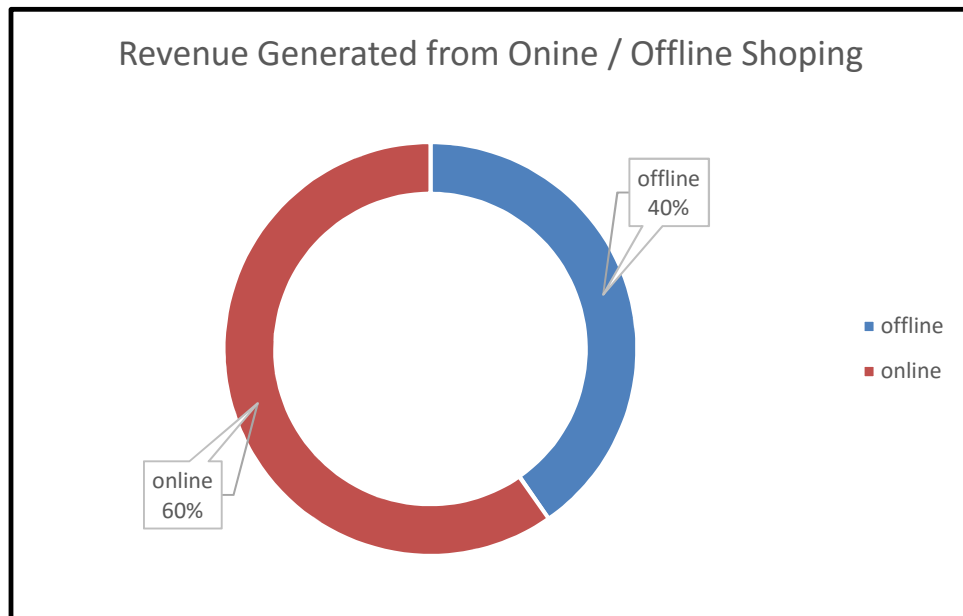
year	Units Sold
2012	15756
2013	15813
2014	16290
2015	16422
2016	3405
Grand Total	67686



Shopping done using online/offline shopping

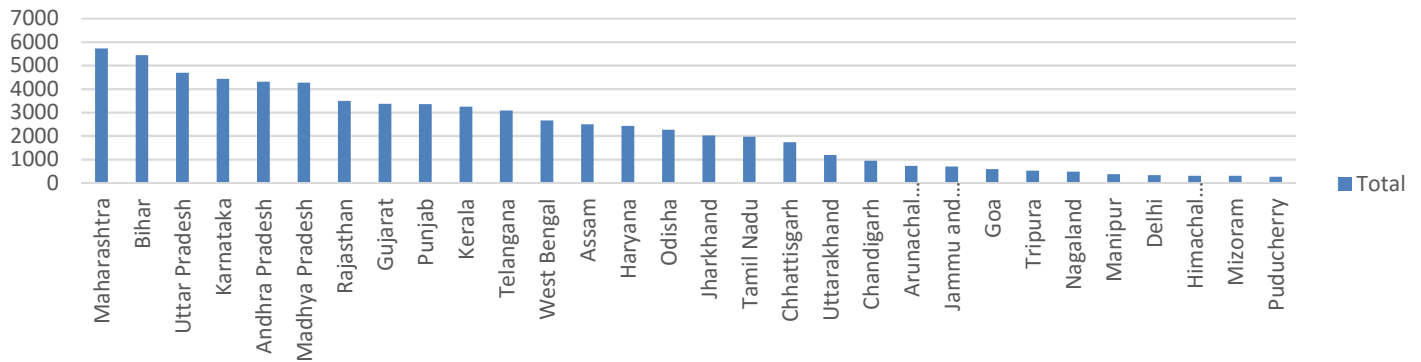
The total revenue generated with online shopping is 1225070 while 827379 by offline shopping

Online / Offline Shopping	Revenue Generated
offline	827379
online	1225070
Grand Total	2052449

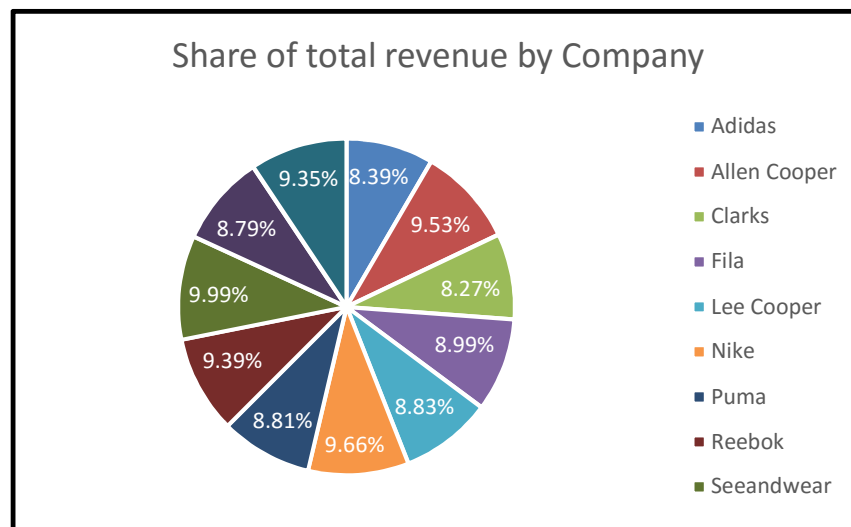


State/city with Highest number of shoes sold is Maharashtra and the state/city with least shoes sold is Puducherry

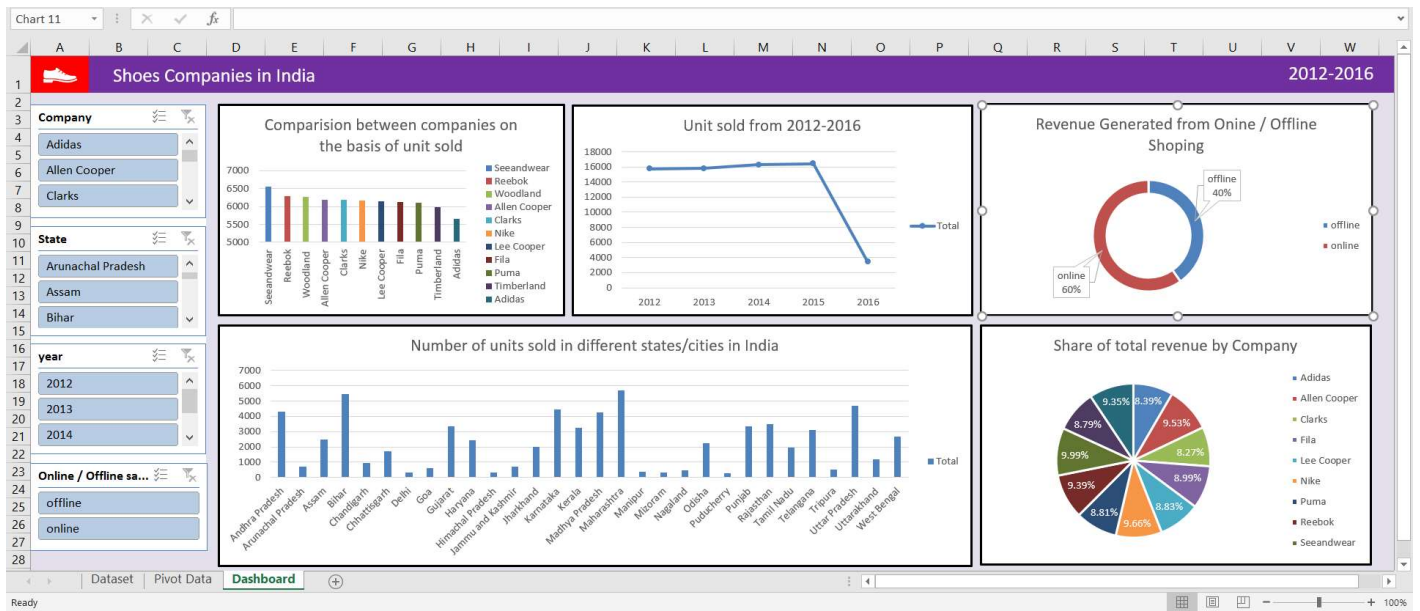
Number of units sold in different states/cities in India



Share of total revenue by company



Final Dashboard



Future Scope

In India Shoes Industries is one of the fastest growing industries. In India the majority of people like to buy new shoes every month as the taste changes. So to meet the quality and the taste of Indian shoes consumers, many companies comes up with their products. The price is also the one of the major factor which affects the consumer. The main thing which consumer need is good quality in budget.

And coming to the future, it has great scope and potential and companies taking care of it so well it does not seem the company is going lose any shine in near future.

Reference and Bibliography

- Youtube
- Kaggle
- Google.com

