LOW LEVEL DESIGN DOCUMENT

Amazon sales data analysis

Sagar V Gundale

VERSION: 1.0

15-01-2023

Document Version Control:

Amazon sales data Analysis – Business Intelligence Project

Version	Date	Author	Change
1.0	15/01/2023	Sagar V Gundale	First Version of Complete LLD

Abstract

The e-commerce domain, as a vital part of the overall supply chain, is expected to highly evolve in the upcoming years via the developments, which are taking place on the side of the Future Internet. This analysis presents overall perspective of sales and profit of amazon in the various regions and in various countries within regions, which aims to improve the business by the collaboration of numerous stakeholders belonging to associated business domains, in an effective and flexible manner.

In the world of rising new technology and innovation, the e-commerce is advancing with the role of Data Science and Analytics. Data analysis can help them to understand their business in a quite different manner and helps to improve the quality of the service by identifying the weak areas of the business. This study demonstrates how different analysis help to make better business decisionsand help analyze sales and profit trends in different years by amazon, which can lead to new and better products and services. Different analyses were performed such as Exploratory Data Analysis and Descriptive Analysis on a variety of use cases to get the key insights from this data based on which business decisions will be taken.

This dataset provides a huge amount of information on sales that amazon is involved in various regions by country and by the products. Based on the Information the ultimate goal would be to predict product and country regions which are contributing highest sales and profit to the company for making better decision to meet overall demand of the products and services by supply chain management.

Contents:

Low Level Design (LLD)

1. Introduction:

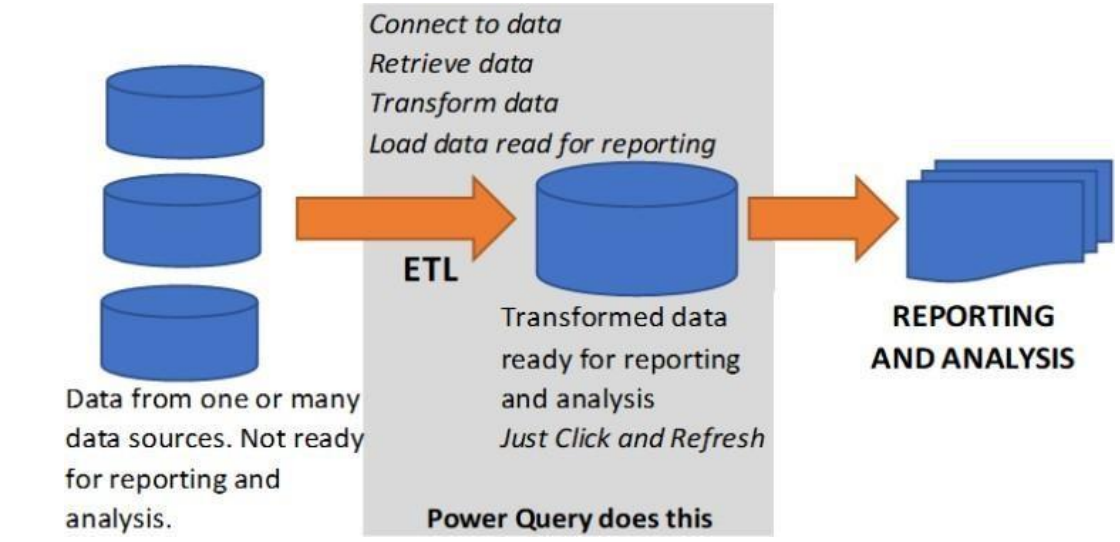
1.1. Why this Low-Level Design Document?

The goal of the LDD or Low-level design document (LLDD) is to give the internal logic design of the actual program code for the Bank Marketing Campaign Analysis. LDD describes the class diagrams with the methods and relations between classes and programs specs. It describes the modules so that the programmer can directly code the program from the document.

1.2. Scope

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

2. Architecture:



.

ETL (extract, transform and load) in Power BI uses preparation of data sets for analysis by removing irregularities in the data. It also involves data visualization to draw meaningful patterns and insights.

Based on the results of ETL, companies also make business decisions, which can have repercussions later.

- If ETL is not done properly then it can damage the business a lot in many ways such as loss of client which we are working for, the decision making will go completely wrong and many more issues.
- If done well, it may improve the efficacy of everything we do next.

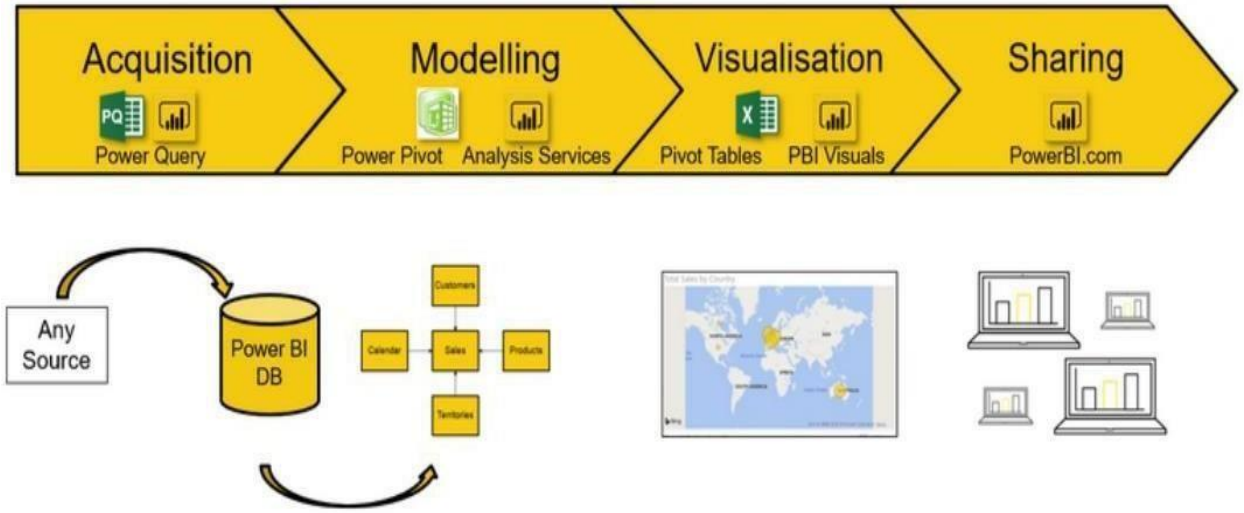
Below are following steps to follow for ETL:

1. Data Sourcing

2. Data Cleaning

3. Data Modelling

4. Data Visualization



3. Architecture Description:

3.1 Data Sourcing:

The dataset is in csv (comma separated values) format. MS Excel is used to study the data.

Citation Request:

This Dataset is publicly available for research, Available at

https://drive.google.com/drive/folders/1FkmFVL8wlJmQWP1z52TD8PlhOJhitTyI?usp=sharing

1. Title: Amazon sales data
2. Source – https://drive.google.com/drive/folders/1FkmFVL8wlJmQWP1z52TD8PlhOJhitTyI?usp=sharing

3. Data Overview –

- The Data includes single .csv file with all examples.
- The Number of Regions -7
- Number of Countries – 75
- Number of Products categories--12

3.2 Data Description –

- **Regions :–**Asia, Australia and oceania, Central America and Caribbean, Europe, Middle east and north Africa, North america, Sub Saharan Africa.

- **Countries :-**
  Albania, Angolia, Australia, Austria, Azerbaijan, Bangladesh, Brunei, Belize, Burkina Faso, Bulgaria, Cape Verde, Cameroon, Costa Rica, Comoros, Costa Rica, Cote d'Ivoire, Djibouti, Democratic Republic of the Congo, East Timor, France, Fiji, Federated States of Micronesia, Gabon, Grenada, Haiti, Honduras, Iran, Iceland, Kiribati, Kenya, Kyrgyzstan, Kuwait, Lesotho, Laos, Lebanon, Libya, Lithuania, Macedonia, Madagascar, Malaysia, Mongolia, Moldova, Mali, Mauritania, Mexico, Monaco, Mozambique, Myanmar, Norway, New Zealand, Niger, Nicaragua, Pakistan, Portugal, Russia, Rwanda, Republic of the Congo, Romania, Republic of the Congo, Sao Tome and Principe, Solomon Islands, Senegal, Sri Lanka, Switzerland, South Sudan, Syria, Slovakia, Saudi Arabia, Sierra Leone, San Marino, Samoa, Spain, Slovenia, Tuvalu, Turkmenistan, The Gambia, Zambia.

- **Sales channel :-** Online and Offline
- **Products:-**

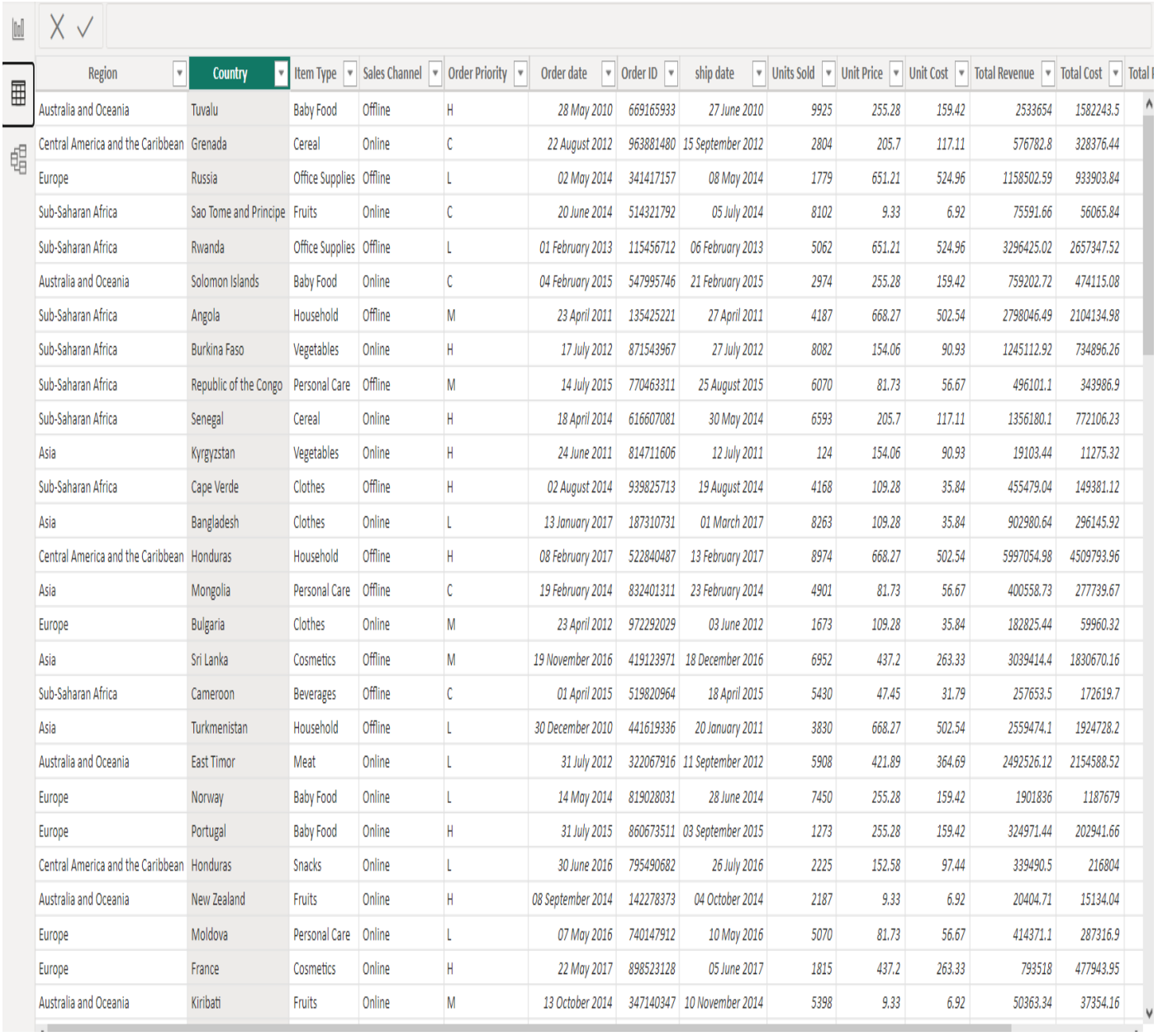
  Baby Food, Beverages, Cereal, Clothes, Cosmetics, Fruits, Household, Meat, Office Supplies, Snacks, Personal Care , Vegetables

- **Order Priority:- 'H','C','L','M'**

3.4 Data loading in Power BI Query Editor

Power Query is the data connectivity and data preparation technology that enables end users to seamlessly import and reshape data from within a wide range of Microsoft products, including Excel, Power BI, Analysis Services, data verse, and more with the following characteristics:

- There can be multiple rows and columns in the data.
- Each row represents a sample of data,
- Each column contains a different variable that describes the samples (rows).
- The data in every column can be a different type of data – e.g. numbers, strings, dates, Boolean etc.



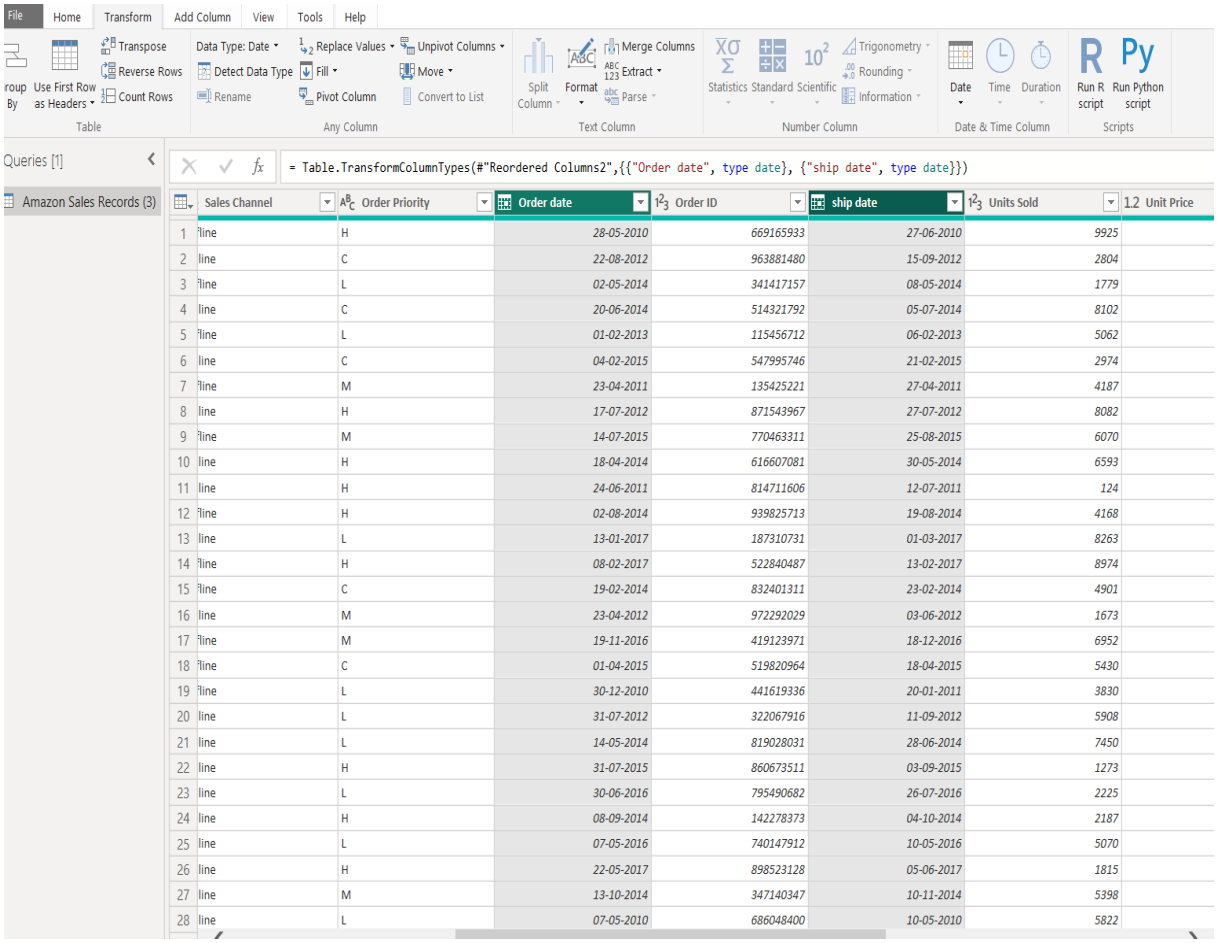
Region	Country	Item Type	Sales Channel	Order Priority	Order date	Order ID	ship date	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total
Australia and Oceania	Tuvalu	Baby Food	Offline	H	28 May 2010	669165933	27 June 2010	9925	255.28	159.42	2533654	1582243.5	
Central America and the Caribbean	Grenada	Cereal	Online	C	22 August 2012	963881480	15 September 2012	2804	205.7	117.11	576782.8	328376.44	
Europe	Russia	Office Supplies	Offline	L	02 May 2014	341417157	08 May 2014	1779	651.21	524.96	1158502.59	933903.84	
Sub-Saharan Africa	Sao Tome and Principe	Fruits	Online	C	20 June 2014	514321792	05 July 2014	8102	9.33	6.92	75591.66	56065.84	
Sub-Saharan Africa	Rwanda	Office Supplies	Offline	L	01 February 2013	115456712	06 February 2013	5062	651.21	524.96	3296425.02	2657347.52	
Australia and Oceania	Solomon Islands	Baby Food	Online	C	04 February 2015	547995746	21 February 2015	2974	255.28	159.42	759202.72	474115.08	
Sub-Saharan Africa	Angola	Household	Offline	M	23 April 2011	135425221	27 April 2011	4187	668.27	502.54	2798046.49	2104134.98	
Sub-Saharan Africa	Burkina Faso	Vegetables	Online	H	17 July 2012	871543967	27 July 2012	8082	154.06	90.93	1245112.92	734896.26	
Sub-Saharan Africa	Republic of the Congo	Personal Care	Offline	M	14 July 2015	770463311	25 August 2015	6070	81.73	56.67	496101.1	343986.9	
Sub-Saharan Africa	Senegal	Cereal	Online	H	18 April 2014	616607081	30 May 2014	6593	205.7	117.11	1356180.1	772106.23	
Asia	Kyrgyzstan	Vegetables	Online	H	24 June 2011	814711606	12 July 2011	124	154.06	90.93	19103.44	11275.32	
Sub-Saharan Africa	Cape Verde	Clothes	Offline	H	02 August 2014	939825713	19 August 2014	4168	109.28	35.84	455479.04	149381.12	
Asia	Bangladesh	Clothes	Online	L	13 January 2017	187310731	01 March 2017	8263	109.28	35.84	902980.64	296145.92	
Central America and the Caribbean	Honduras	Household	Offline	H	08 February 2017	522840487	13 February 2017	8974	668.27	502.54	5997054.98	4509793.96	
Asia	Mongolia	Personal Care	Offline	C	19 February 2014	832401311	23 February 2014	4901	81.73	56.67	400558.73	277739.67	
Europe	Bulgaria	Clothes	Online	M	23 April 2012	972292029	03 June 2012	1673	109.28	35.84	182825.44	59960.32	
Asia	Sri Lanka	Cosmetics	Offline	M	19 November 2016	419123971	18 December 2016	6952	437.2	263.33	3039414.4	1830670.16	
Sub-Saharan Africa	Cameroon	Beverages	Offline	C	01 April 2015	519820964	18 April 2015	5430	47.45	31.79	257653.5	172619.7	
Asia	Turkmenistan	Household	Offline	L	30 December 2010	441619336	20 January 2011	3830	668.27	502.54	2559474.1	1924728.2	
Australia and Oceania	East Timor	Meat	Online	L	31 July 2012	322067916	11 September 2012	5908	421.89	364.69	2492526.12	2154588.52	
Europe	Norway	Baby Food	Online	L	14 May 2014	819028031	28 June 2014	7450	255.28	159.42	1901836	1187679	
Europe	Portugal	Baby Food	Online	H	31 July 2015	860673511	03 September 2015	1273	255.28	159.42	324971.44	202941.66	
Central America and the Caribbean	Honduras	Snacks	Online	L	30 June 2016	795490682	26 July 2016	2225	152.58	97.44	339490.5	216804	
Australia and Oceania	New Zealand	Fruits	Online	H	08 September 2014	142278373	04 October 2014	2187	9.33	6.92	20404.71	15134.04	
Europe	Moldova	Personal Care	Online	L	07 May 2016	740147912	10 May 2016	5070	81.73	56.67	414371.1	287316.9	
Europe	France	Cosmetics	Online	H	22 May 2017	898523128	05 June 2017	1815	437.2	263.33	793518	477943.95	
Australia and Oceania	Kiribati	Fruits	Online	M	13 October 2014	347140347	10 November 2014	5398	9.33	6.92	50363.34	37354.16	

3.5 Data Cleaning:

Initial:

Ship Date
6/27/2010
9/15/2012
05-08-2014
07-05-2014
02-06-2013
2/21/2015
4/27/2011
7/27/2012
8/25/2015
5/30/2014
07-12-2011
8/19/2014
03-01-2017
2/13/2017
2/23/2014
06-03-2012
12/18/2016
4/18/2015

Final:



	Sales Channel	Order Priority	Order date	Order ID	ship date	Units Sold	Unit Price
1	line	H	28-05-2010	669165933	27-06-2010	9925	
2	line	C	22-08-2012	963881480	15-09-2012	2804	
3	line	L	02-05-2014	341417157	08-05-2014	1779	
4	line	C	20-06-2014	514321792	05-07-2014	8102	
5	line	L	01-02-2013	115456712	06-02-2013	5062	
6	line	C	04-02-2015	547995746	21-02-2015	2974	
7	line	M	23-04-2011	135425221	27-04-2011	4187	
8	line	H	17-07-2012	871543967	27-07-2012	8082	
9	line	M	14-07-2015	770463311	25-08-2015	6070	
10	line	H	18-04-2014	616607081	30-05-2014	6593	
11	line	H	24-06-2011	814711606	12-07-2011	124	
12	line	H	02-08-2014	939825713	19-08-2014	4168	
13	line	L	13-01-2017	187310731	01-03-2017	8263	
14	line	H	08-02-2017	522840487	13-02-2017	8974	
15	line	C	19-02-2014	832401311	23-02-2014	4901	
16	line	M	23-04-2012	972292029	03-06-2012	1673	
17	line	M	19-11-2016	419123971	18-12-2016	6952	
18	line	C	01-04-2015	519820964	18-04-2015	5430	
19	line	L	30-12-2010	441619336	20-01-2011	3830	
20	line	L	31-07-2012	322067916	11-09-2012	5908	
21	line	L	14-05-2014	819028031	28-06-2014	7450	
22	line	H	31-07-2015	860673511	03-09-2015	1273	
23	line	L	30-06-2016	795490682	26-07-2016	2225	
24	line	H	08-09-2014	142278373	04-10-2014	2187	
25	line	L	07-05-2016	740147912	10-05-2016	5070	
26	line	H	22-05-2017	898523128	05-06-2017	1815	
27	line	M	13-10-2014	347140347	10-11-2014	5398	
28	line	L	07-05-2010	686048400	10-05-2010	5822	

3.6 Data Insights through Visualizations using Power BI