INTRODUCTION TO DATA MANAGEMENT PROJECT REPORT

(Project Semester August-December 2021)

RETAIL DATA ANALYTICS

Submitted by-

Saurabh Gupta

Registration No- 11914681

Section- KM005

Course Code-INT 217

Under the Guidance of –

Komal Arora

Discipline of CSE/IT

Lovely School of Computer Science & Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that Saurabh Gupta bearing Registration number 11914681 has completed INT 217project titled, **"Retail Data Analytics"** 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.

Signature and Name of the Supervisor
Designation of the Supervisor
School of Computer Science & Engineering
Lovely Professional University
Phagwara, Punjab.

DECLARATION

I, Saurabh Gupta, student of B. Tech. 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: 18 December 2021 Signature: 

Registration No.11914681 Name of the student: Saurabh Gupta

Acknowledgement

I would like to express my gratitude towards my University as well as my course instructor Ms. Komal Arora for providing me the golden opportunity to do this wonderful project on Data Analysis which also helped me in doing a lot of homework and learning. As a result, I came to know about so many new things. So, I am really thank full to them.

Moreover I would like to thank my friends who helped me a lot whenever I got stuck in some problem related to my course and also this project. I am really thankfull to have such a good support of them as they always have my back whenever I need.

Also, I would like to mention the support system and consideration of my parents who have always been there in my life to make me choose right thing and oppose the wrong. Without them I could never had learned and became a person who I am now.

I have taken efforts in this project. However, it would not have been possible without the kind support and help of my instructor, friends and university. I would like to extend my sincere thanks to all of them.

Contents Of the Report

INTRODUCTION

This data set provides with historical sales data for 45 stores located in different regions - each store contains a number of departments. The company also runs several promotional markdown events throughout the year. These markdowns precede prominent holidays, the four largest of which are the Super Bowl, Labor Day, Thanksgiving, and Christmas. The weeks including these holidays are weighted five times higher in the evaluation than non-holiday weeks.Within the Excel Sheet, there are 3 Tabs – Stores, Features and Sales.

Historical sales data, which covers to 2010-02-05 to 2012-11-01.

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 analysing, 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 of the Analysis

- Predict the sales for each store for the following year
- Store wise sales during holiday week
- Most profitable type of store
- Model the effects of markdowns on holiday weeks
- Top 10 store on the basis of consumer price index(CPI)

Source of dataset

Source of the Dataset: https://www.kaggle.com/manjeetsingh/retaildataset

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 analyses the Retail data analytics 2010-02-05 to 2012-11-01 dataset.

Sample of Dataset:

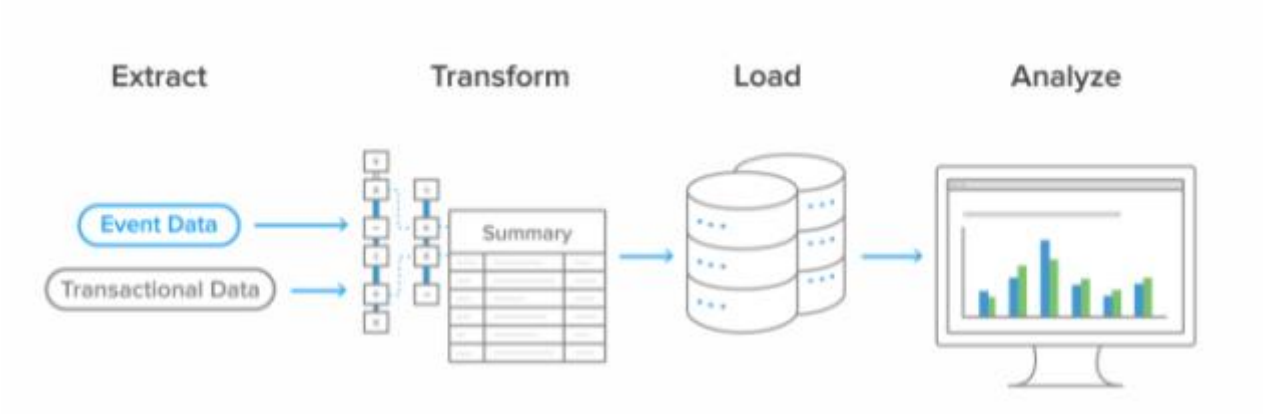
M4

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	IsHoliday	
2	1	5/2/2010	42.31	2.572	NA	NA	NA	NA	NA	211.0963582	8.106	FALSE	
3	1	12/2/2010	38.51	2.548	NA	NA	NA	NA	NA	211.2421698	8.106	TRUE	
4	1	19/02/2010	39.93	2.514	NA	NA	NA	NA	NA	211.2891429	8.106	FALSE	
5	1	26/02/2010	46.63	2.561	NA	NA	NA	NA	NA	211.3196429	8.106	FALSE	
6	1	5/3/2010	46.5	2.625	NA	NA	NA	NA	NA	211.3501429	8.106	FALSE	
7	1	12/3/2010	57.79	2.667	NA	NA	NA	NA	NA	211.3806429	8.106	FALSE	
8	1	19/03/2010	54.58	2.72	NA	NA	NA	NA	NA	211.215635	8.106	FALSE	
9	1	26/03/2010	51.45	2.732	NA	NA	NA	NA	NA	211.0180424	8.106	FALSE	
10	1	2/4/2010	62.27	2.719	NA	NA	NA	NA	NA	210.8204499	7.808	FALSE	
11	1	9/4/2010	65.86	2.77	NA	NA	NA	NA	NA	210.6228574	7.808	FALSE	
12	1	16/04/2010	66.32	2.808	NA	NA	NA	NA	NA	210.4887	7.808	FALSE	
13	1	23/04/2010	64.84	2.795	NA	NA	NA	NA	NA	210.4391228	7.808	FALSE	
14	1	30/04/2010	67.41	2.78	NA	NA	NA	NA	NA	210.3895456	7.808	FALSE	
15	1	7/5/2010	72.55	2.835	NA	NA	NA	NA	NA	210.3399684	7.808	FALSE	
16	1	14/05/2010	74.78	2.854	NA	NA	NA	NA	NA	210.3374261	7.808	FALSE	
17	1	21/05/2010	76.44	2.826	NA	NA	NA	NA	NA	210.6170934	7.808	FALSE	
18	1	28/05/2010	80.44	2.759	NA	NA	NA	NA	NA	210.8967606	7.808	FALSE	
19	1	4/6/2010	80.69	2.705	NA	NA	NA	NA	NA	211.1764278	7.808	FALSE	

ETL Process

ETL, which stands for extract, Transform and load, is a records integration system that combines information from multiple records resources right into a single, consistent information store that is loaded right into a Data warehouse or different target system.

As the databases grew in popularity inside the Seventies, ETL was brought as a procedure for integrating and loading records for computation and evaluation, ultimately turning into the number one method to technique data for data warehousing tasks.



To maintain its value as a tool for decision-makers, Data warehouse system needs to change with business changes. ETL is a recurring activity (daily, weekly, monthly) of a Data warehouse system and needs to be agile, automated and well documented.

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 target database.
- 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.

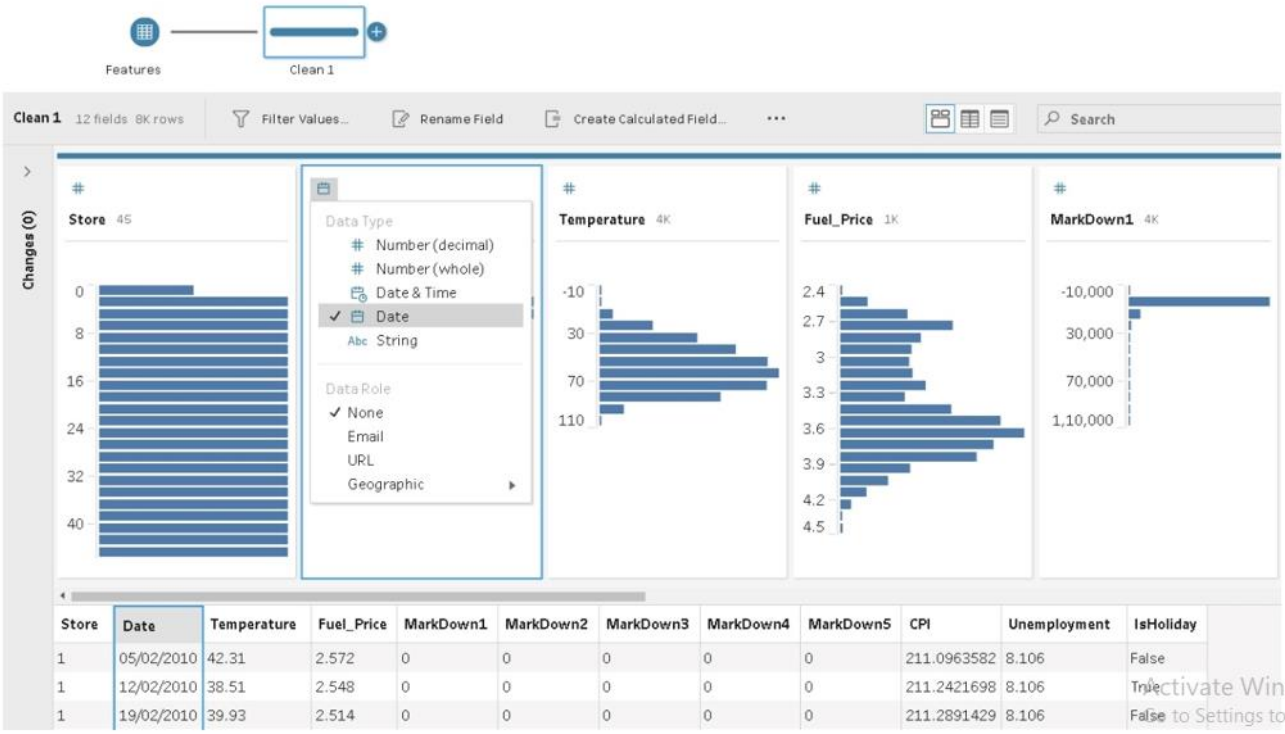
1. E – Extraction

2. T – Transformation

3. L – Loading

ETL Process Used in Project

Data type changed of date column.



Analysis on dataset

INTRODUCTION

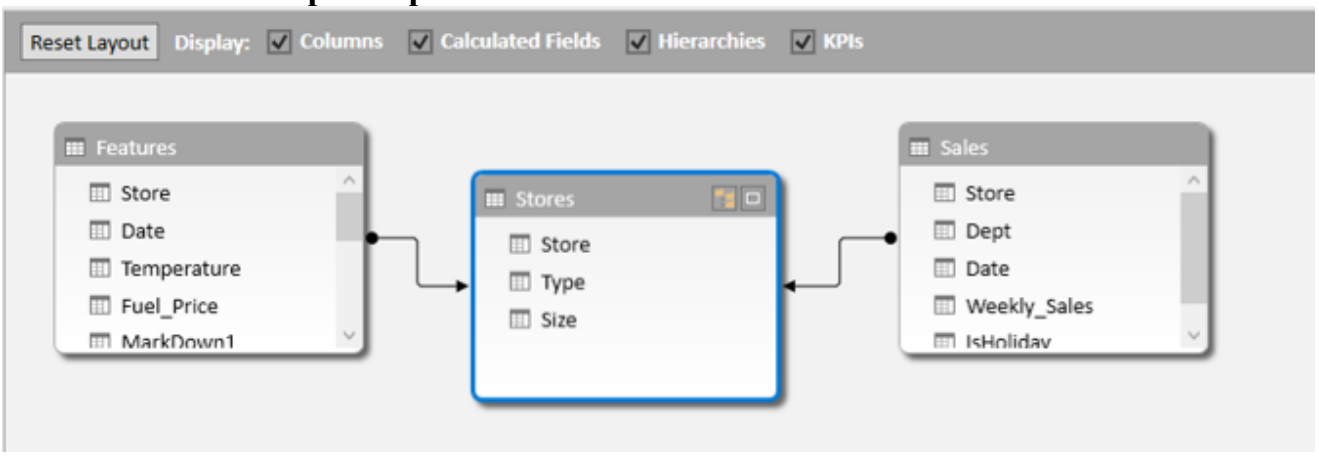
This dataset consists of 3 different worksheets named as Features, Sales and Stores. This dataset revolves around weekly sales of 45 stores which are categorized as A, B and C type.

Raw data uploaded in power pivot-

[Store]

Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment
1	2/5/2010 ...	42.31	2.572	0	0	0	0	0	211.096...	
1	2/12/201...	38.51	2.548	0	0	0	0	0	211.242...	
1	2/19/201...	39.93	2.514	0	0	0	0	0	211.289...	
1	2/26/201...	46.63	2.561	0	0	0	0	0	211.319...	
1	3/5/2010 ...	46.5	2.625	0	0	0	0	0	211.350...	
1	3/12/201...	57.79	2.667	0	0	0	0	0	211.380...	
1	3/19/201...	54.58	2.72	0	0	0	0	0	211.215...	
1	3/26/201...	51.45	2.732	0	0	0	0	0	211.018...	
1	4/2/2010 ...	62.27	2.719	0	0	0	0	0	210.820...	
1	4/9/2010 ...	65.86	2.77	0	0	0	0	0	210.622...	
1	4/16/201...	66.32	2.808	0	0	0	0	0	210.4887	
1	4/23/201...	64.84	2.795	0	0	0	0	0	210.439...	
1	4/30/201...	67.41	2.78	0	0	0	0	0	210.389...	
1	5/7/2010 ...	72.55	2.835	0	0	0	0	0	210.339...	
1	5/14/201...	74.78	2.854	0	0	0	0	0	210.337...	
1	5/21/201...	76.44	2.826	0	0	0	0	0	210.617...	

Connection made in power pivot-

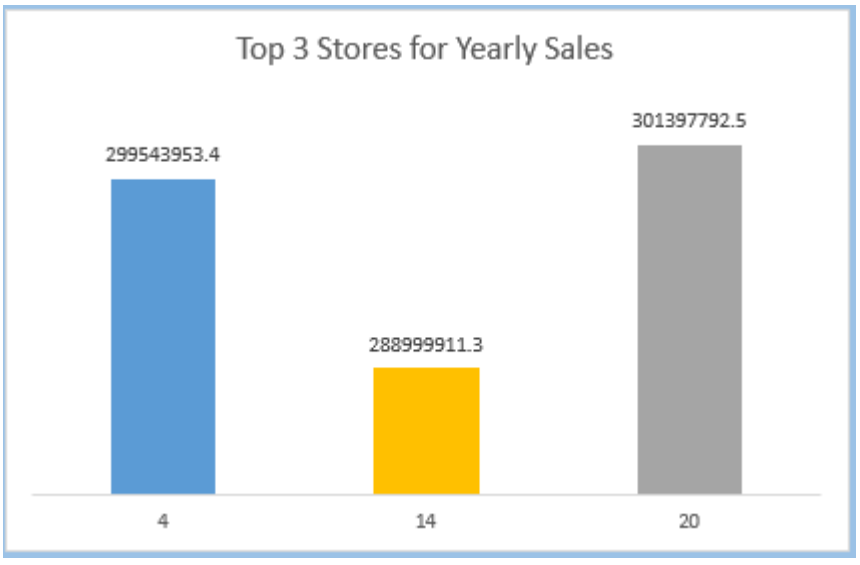


The Cleaned Retail Dataset:

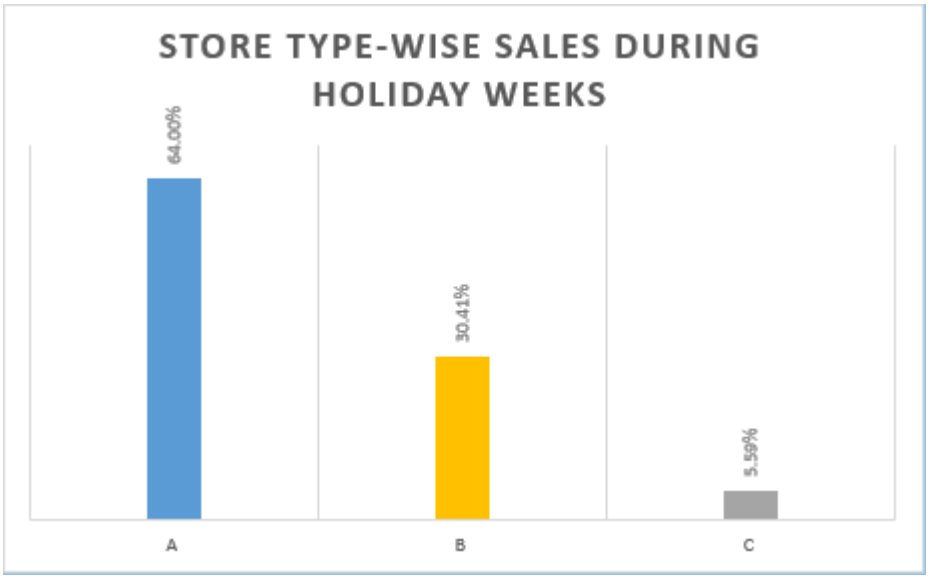
Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	IsHoliday
1	2/5/2010	42.31	2.572	0	0	0	0	0	211.0963582	8.106	FALSE
1	2/12/2010	38.51	2.548	0	0	0	0	0	211.2421698	8.106	TRUE
1	2/19/2010	39.93	2.514	0	0	0	0	0	211.2891429	8.106	FALSE
1	2/26/2010	46.63	2.561	0	0	0	0	0	211.3196429	8.106	FALSE
1	3/5/2010	46.5	2.625	0	0	0	0	0	211.3501429	8.106	FALSE
1	3/12/2010	57.79	2.667	0	0	0	0	0	211.3806429	8.106	FALSE
1	3/19/2010	54.58	2.72	0	0	0	0	0	211.215635	8.106	FALSE
1	3/26/2010	51.45	2.732	0	0	0	0	0	211.0180424	8.106	FALSE
1	4/2/2010	62.27	2.719	0	0	0	0	0	210.8204499	7.808	FALSE
1	4/9/2010	65.86	2.77	0	0	0	0	0	210.6228574	7.808	FALSE
1	4/16/2010	66.32	2.808	0	0	0	0	0	210.4887	7.808	FALSE
1	4/23/2010	64.84	2.795	0	0	0	0	0	210.4391228	7.808	FALSE
1	4/30/2010	67.41	2.78	0	0	0	0	0	210.3895456	7.808	FALSE
1	5/7/2010	72.55	2.835	0	0	0	0	0	210.3399684	7.808	FALSE
1	5/14/2010	74.78	2.854	0	0	0	0	0	210.3374261	7.808	FALSE
1	5/21/2010	76.44	2.826	0	0	0	0	0	210.6170934	7.808	FALSE
1	5/28/2010	80.44	2.759	0	0	0	0	0	210.8967606	7.808	FALSE
1	6/4/2010	80.69	2.705	0	0	0	0	0	211.1764278	7.808	FALSE
1	6/11/2010	80.43	2.668	0	0	0	0	0	211.4560951	7.808	FALSE
1	6/18/2010	84.11	2.637	0	0	0	0	0	211.4537719	7.808	FALSE
1	6/25/2010	84.34	2.653	0	0	0	0	0	211.3386526	7.808	FALSE

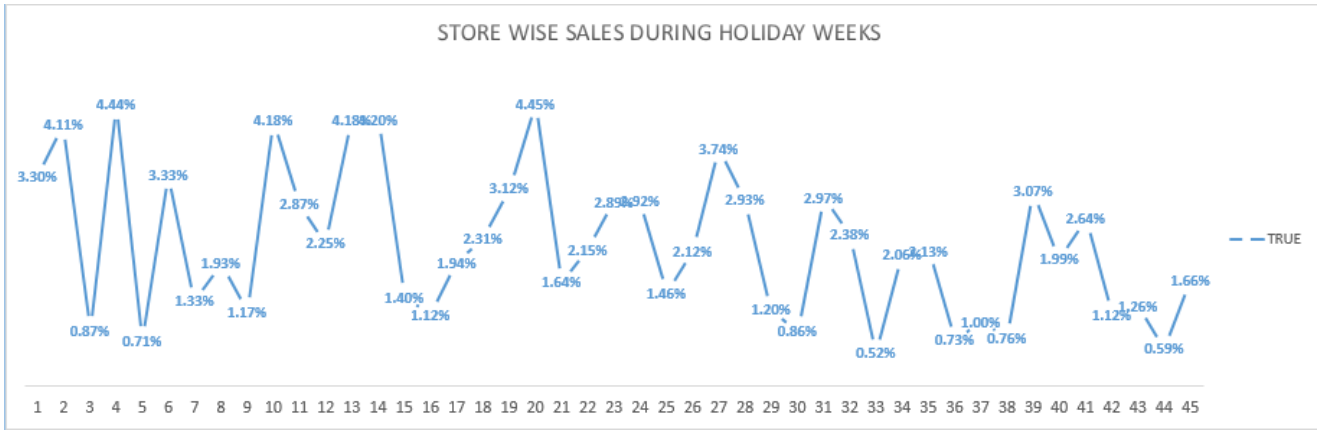
OBJECTIVES-

Predict the sales for each store for the following year:

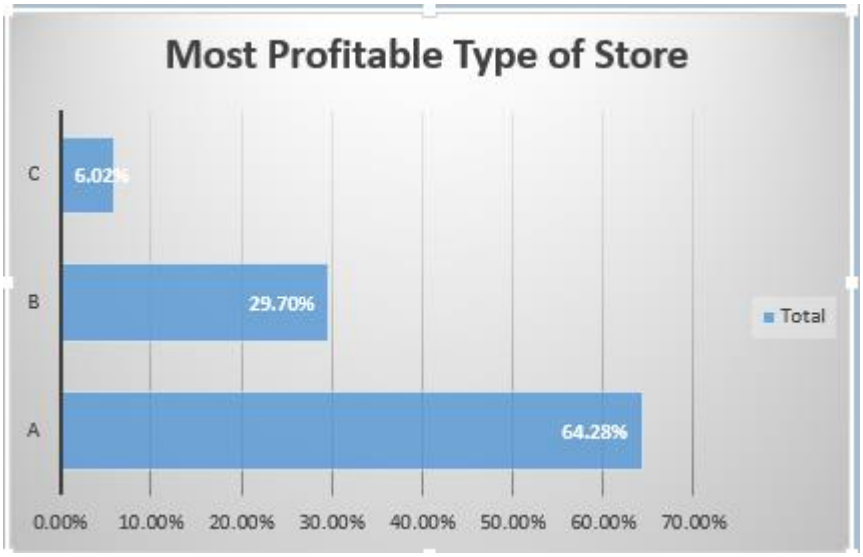


Store wise sales during holiday week:

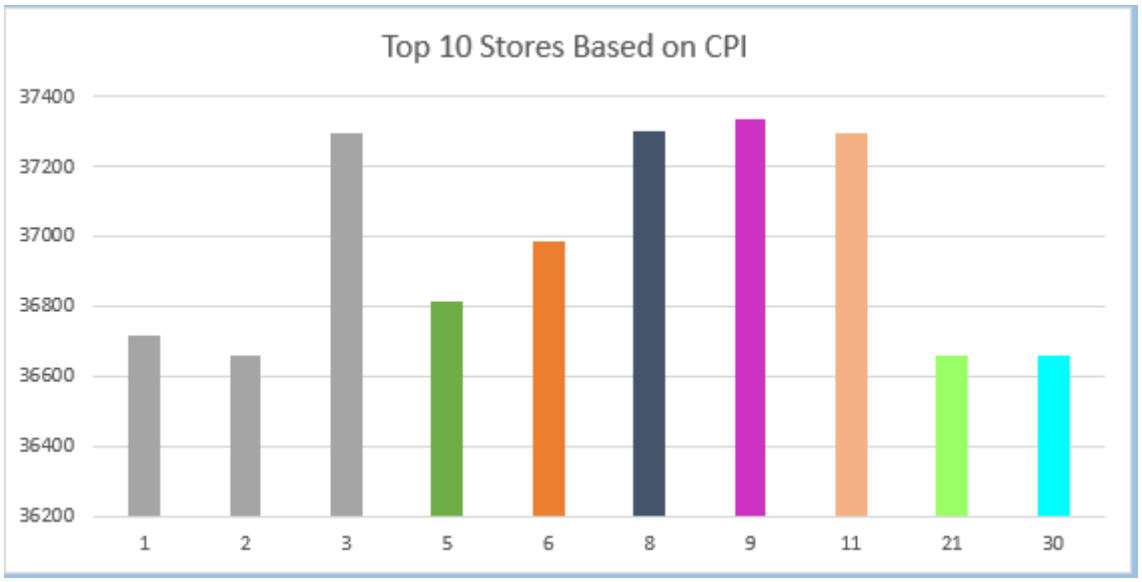




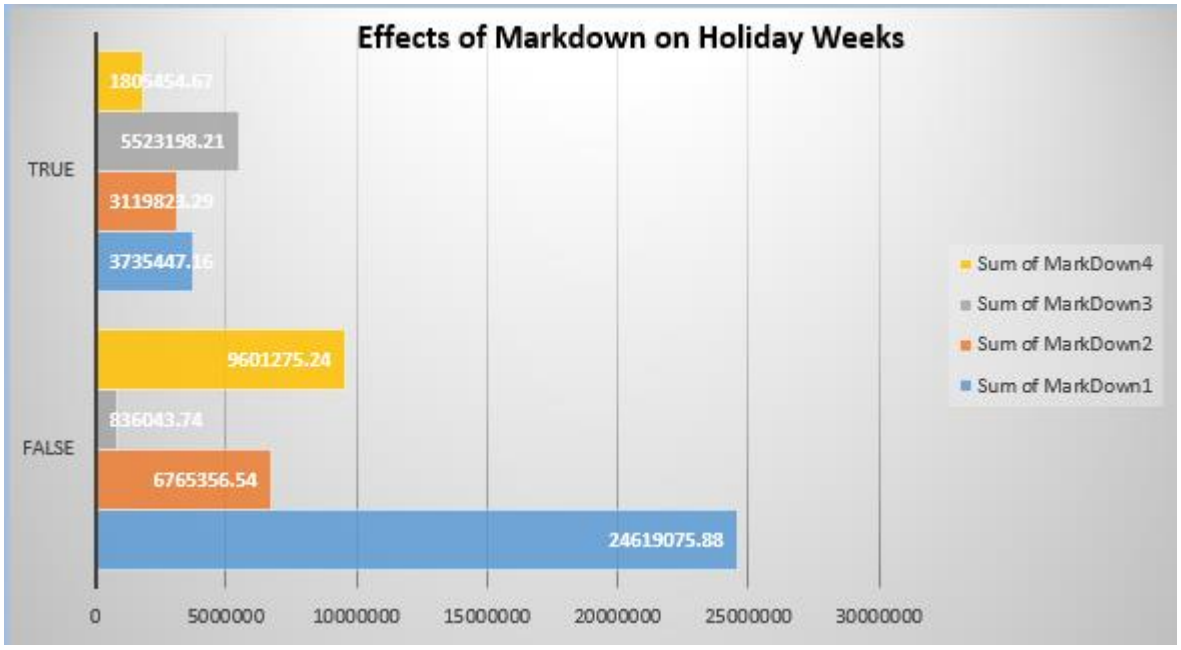
Most profitable type of store:



Top 10 store on the basis of consumer price index (CPI):

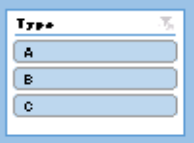


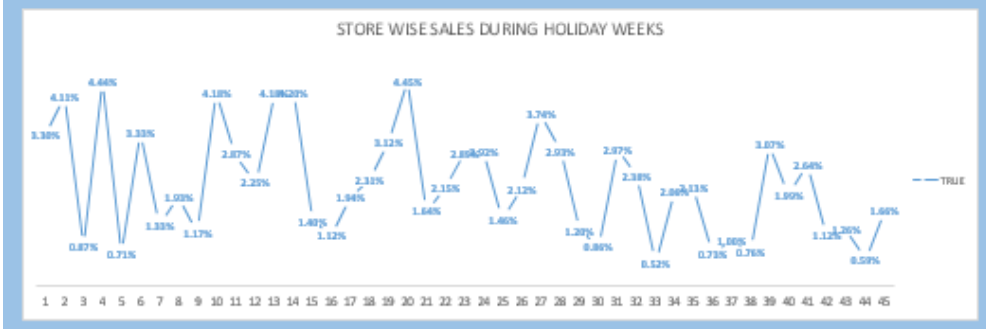
Model the effects of markdowns on holiday weeks:

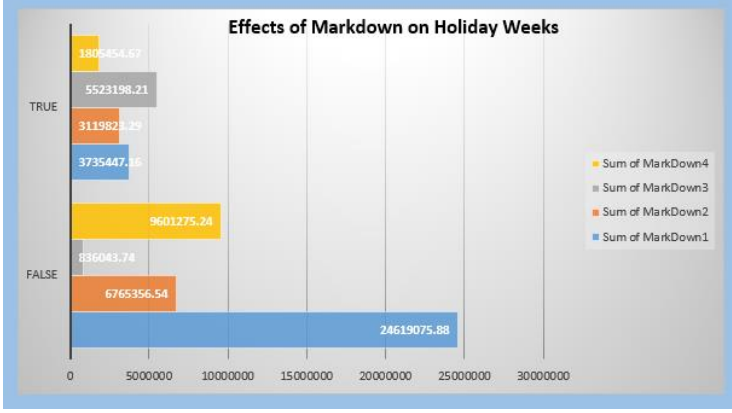


Overall Analysis

ANNUAL TRENDS







List of Analysis with Results

- Predict the sales for each store for the following year:
  The top store was store no. 20
- Store wise sales during holiday week:
  Type A store was more profitable during holiday weeks
- Most profitable type of store:
  Type A store was most profitable
- Top 10 store on the basis of consumer price index (CPI):
  Store 9 was the top store on the basis of CPI.
- Model the effects of markdowns on holiday weeks:
  Markdown 1 had the most effect on the holiday weeks

Bibliography

- https://www.kaggle.com/manjeetsingh/retaildataset
- https://www.tableau.com/academic/students
- https://www.geeksforgeeks.org
- www.youtube.com