Retail Store Sales Analysis

Case - Walmart

MBAN_Pin-Chia (Cynthia) Chou_pvc5424

CONTENT

1. Introduction

- What is Sales Analysis
- Key Performance Indicator (KPI)
- Infosphere of Work Domain

Introduction of Work Domain

My work domain focuses on ***Sales Analysis*** for retail store. Sales analysis is to retrieve data to evaluate the sales performance against company's goal. This analysis can provide insights about not only which product holds the top or under performance, but also sales forecasting. As for who will be targeted audiences of the project, it could be staff from sales, marketing team, supply chain or leader team.

In terms of key performance indicator (KPI), **sales revenue** and **monthly sales growth** will be applied to the project. Furthermore, sales revenue can be divided into annually one and quarterly one.

Infosphere of Work Domain

Dataset

Weekly sales from 45 Walmart stores located in different regions

Data Source

Kaggle

Data Volume

421,570 records

Data Period

2010/2/5 ~ 2012/10/26
(143 weeks in total)

2. Extrapolation

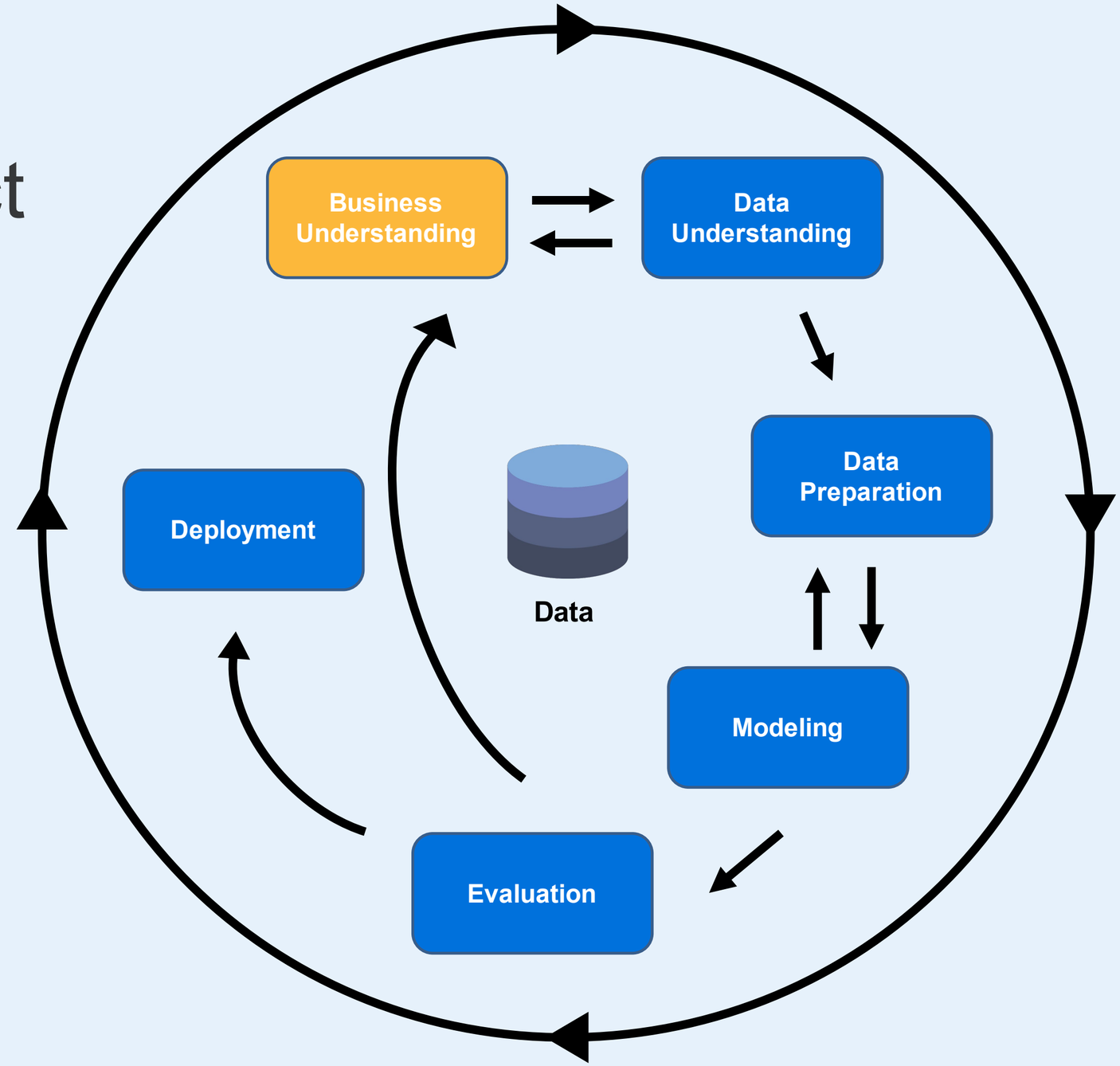
- CRoss Industry Standard Process
- Extract Transform Load (ETL) process

How to start the project

CRoss Industry Standard Process for Data Mining (CRISP-DM)

Before doing data analysis, designing a complete plan on the project is a great start.

Firstly, it's important to determine the business objectives. For instance, the biggest goal of the project is to identify sales revenue across time and the impact of different metrics on the sales, so stakeholders can make better decision on supply chain or marketing strategies.

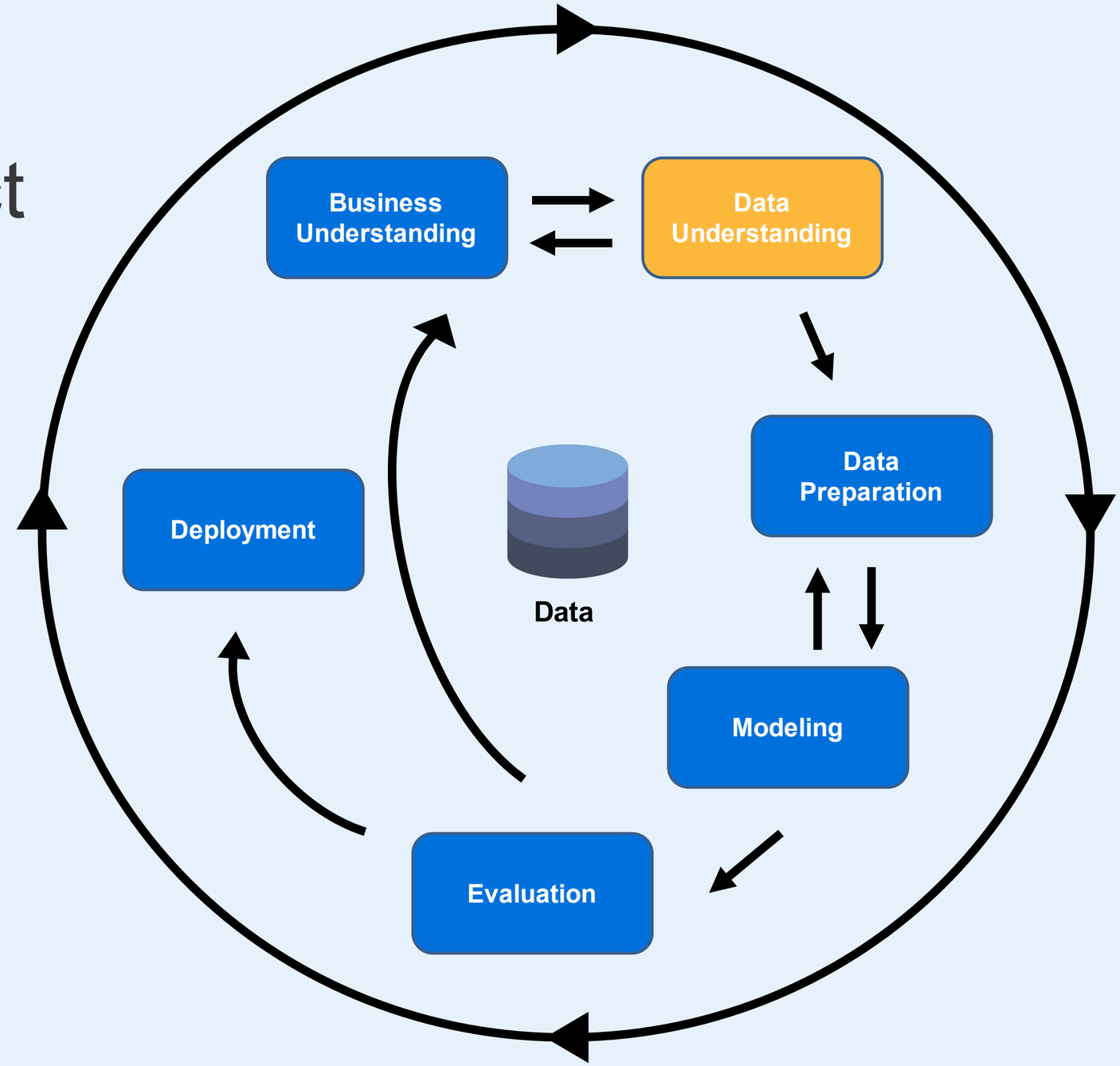


How to start the project

CRoss Industry Standard Process for Data Mining (CRISP-DM)

After defining the project goal, the second step is to collect and explore data that can help accomplish the project objective.

As for how to start data understanding phase, I'll follow *Extract Transform Load (ETL) process* in the next page.



How to do data understanding- *Extract Transform Load (ETL) process*

As mentioned before, in the second phases of CRISP-DM, we need to extract data from databases, then transform data format to the one easier to analyze. Here is detailed phases of ETL process.

Extraction

In the real world, analysts usually collect sales data from existing database, spreadsheet, CRM software or sales applications.
In this project, my data gathering source is from Kaggle, an online platform for programming learners.

Transformation

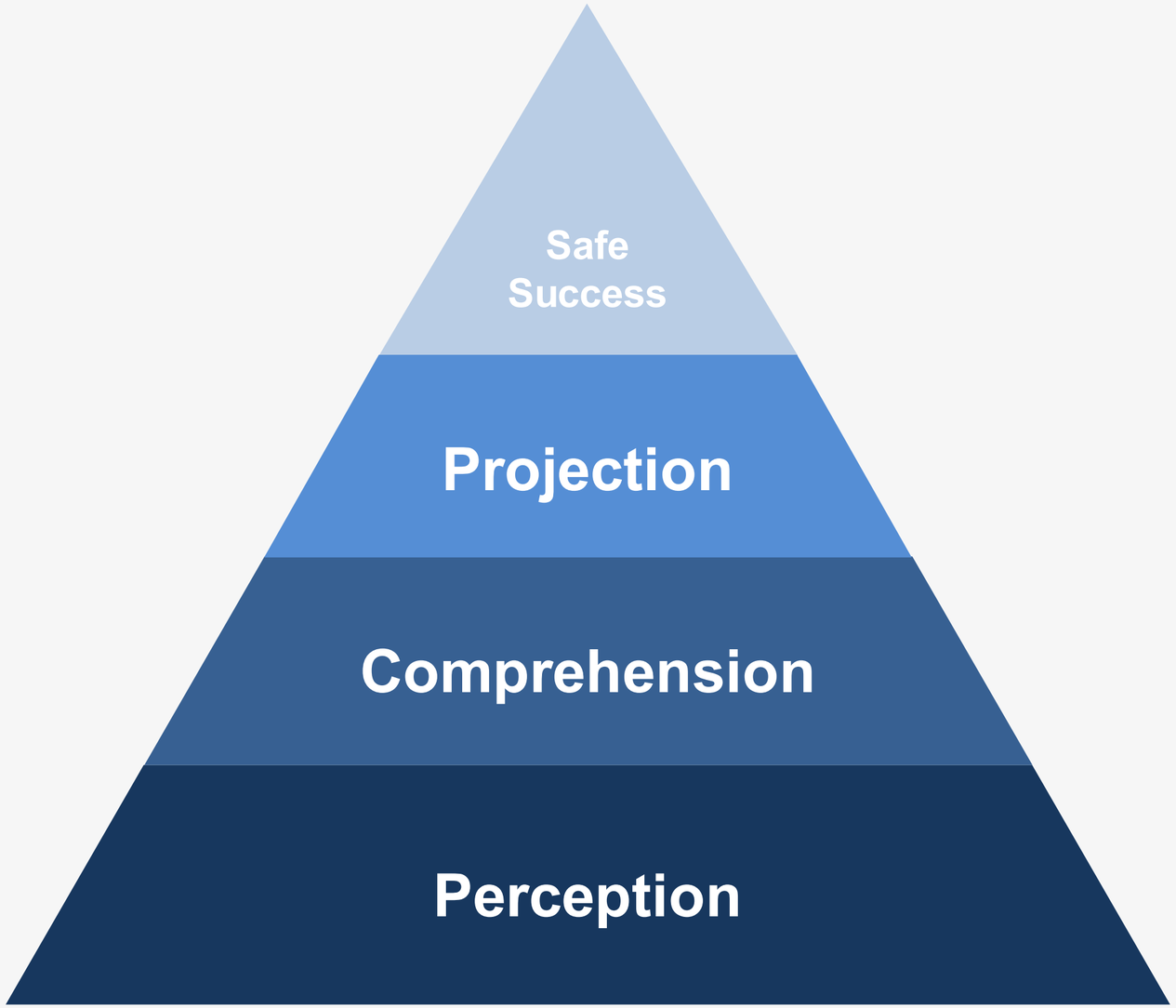
Generally, cleaning, sorting, deduplication, standardization and verification is common sub process. In the project, I will use cleaning and sorting to do data transformation.

Loading

After data transformation, analyst will load newly transformed data into a new destination. Data can be loaded all at once (full load) or at scheduled intervals (incremental load). In the project, I will load this dataset from separate csv files into Tableau.

3. Summary

- Situational Awareness
- Summary of Dashboard



To summarize the process of how user can utilize the dashboard, below is three levels of situational awareness from Endsley's model.

1. Perception of new information

In the dashboard, the purpose of mark cards is to highlight overview of sales for the users. In addition, users can utilize the date slicers to select the specific time period. Lastly, there are four visualization to reveal details about sales.

2. Comprehension of the information

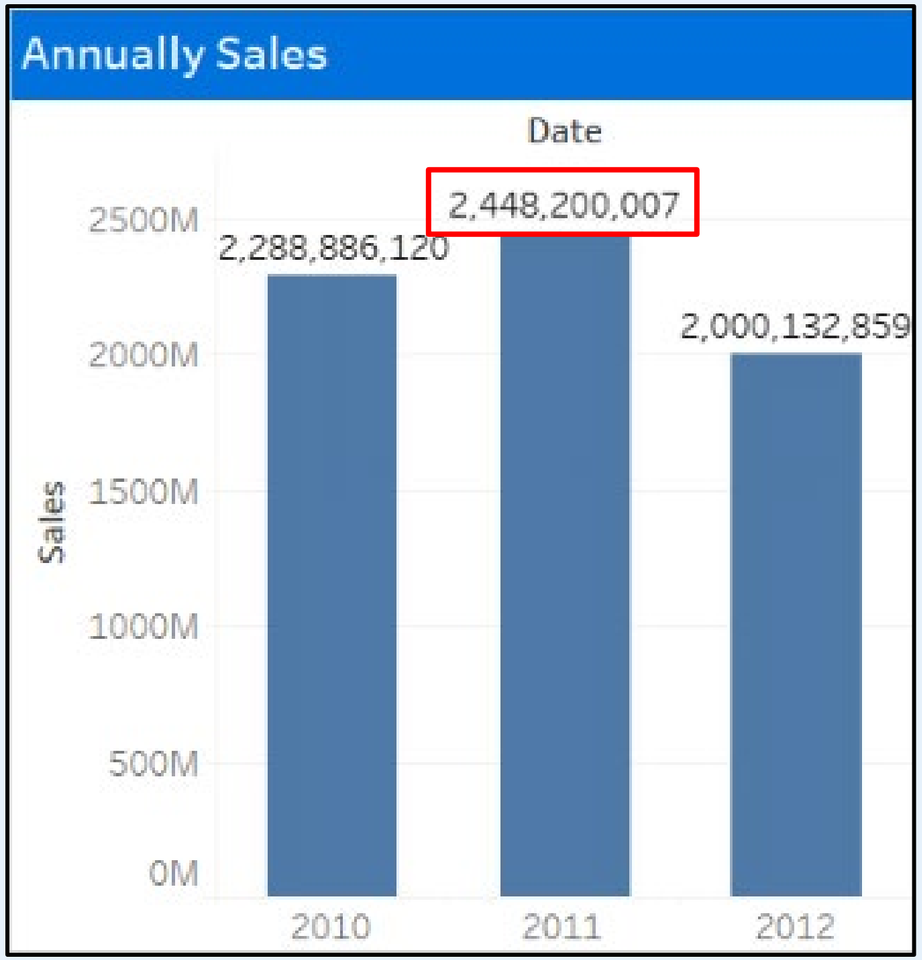
Based on the changes in sales, users can determine which month holds better performance on sales, and reassign staff or adjust the number of inventory. Furthermore, the matrix table provides more in-depth analysis on how consumer price index and unemployment rate affect sales across regions.

3. Projection of impact on the sales

In the final step, we can use these historical sales data to do sales prediction. In the project, I haven’t applied any regression model to do prediction yet.

4. Visualization

- Annually Sales
- Quarterly Sales
- Monthly Sales Growth
- CPI & UNRATE



Visualization 1 – Annually Sales

Data Process

Using year and sum of weekly sales to create the column chart. This column chart provides the overall insights about the trend of sales for stakeholders. In addition to the height of columns, the data labels is helpful to reveal annually sales.

Data Insights

The chart indicates the highest total sales about 2.5 billion in 2011, and lowest sales in 2012. It's necessary to break down sales into quarterly or monthly to identify what factor causes the highest or lowest sales.

Visualization 2 – Quarterly Sales

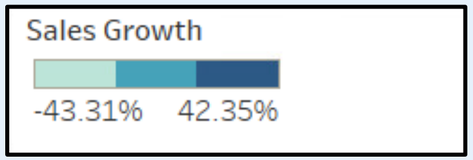
Data Process

Using quarterly date and sum of weekly sales to create the column chart. This column chart reveals the details about the quarterly sales each year.

Data Insights

Generally, there are two important holidays in Q4, Black Friday and Christmas, so Q4 usually accounts for the highest sales. In the chart, it shows the increase in the sales from Q1 to Q4. However, in 2012, the sales of Q4 is the lowest one.

Monthly Sales Growth		
2010	February	
	March	-4.42%
	April	27.21%
	May	-19.32%
	June	2.96%
	July	20.98%
	August	-19.32%
	September	-5.53%
	October	22.50%
	November	-6.59%
	December	42.35%
2011	January	-43.31%
	February	13.82%
	March	-3.74%
	April	26.30%
	May	-19.81%
	June	4.47%
	July	21.15%
	August	-17.97%
	September	17.10%
	October	-17.02%
	November	14.68%
	December	37.07%
2012	January	-41.37%
	February	13.72%
	March	20.54%
	April	-18.40%
	May	-0.08%
	June	27.46%
	July	-22.07%



Visualization 3 – Monthly Sales Growth

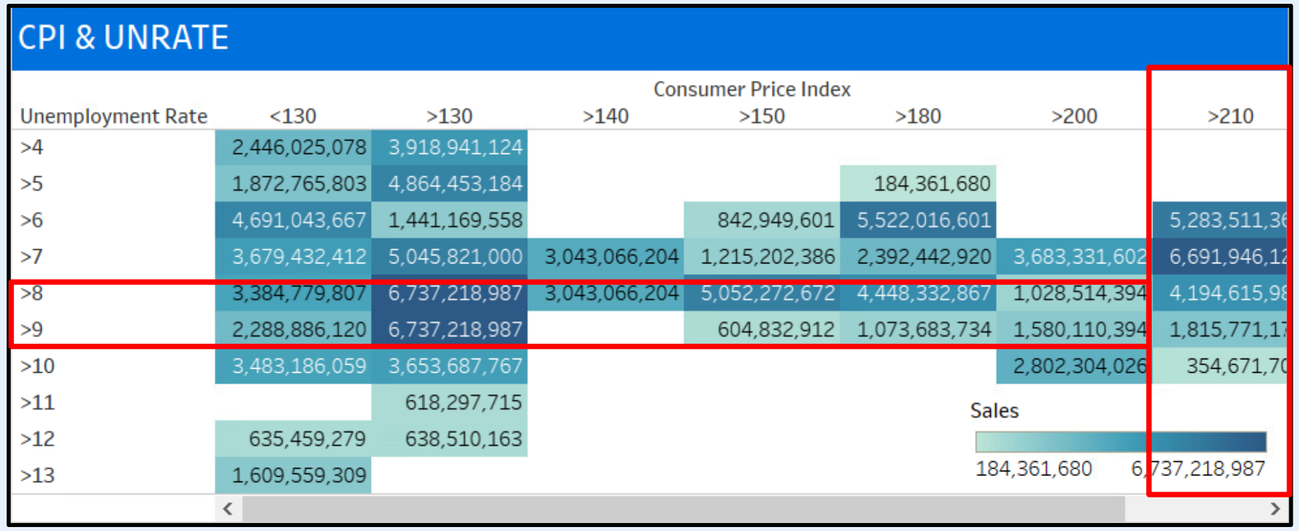
Data Process

Using month date and difference in sales to create a highlight table, and 3 stepped colors show level of sales growth.

Data Insights

The table demonstrates the highest growth is 42.35% in December 2010 and the lowest one is -43.31% in the following month. It could be assumed that the higher sales happened in December because of holiday season, but in January, there wasn't much holiday discount and customers obviously spent less in Walmart than they did in December.

Visualization 4 – CPI & UNRATE



Data Process

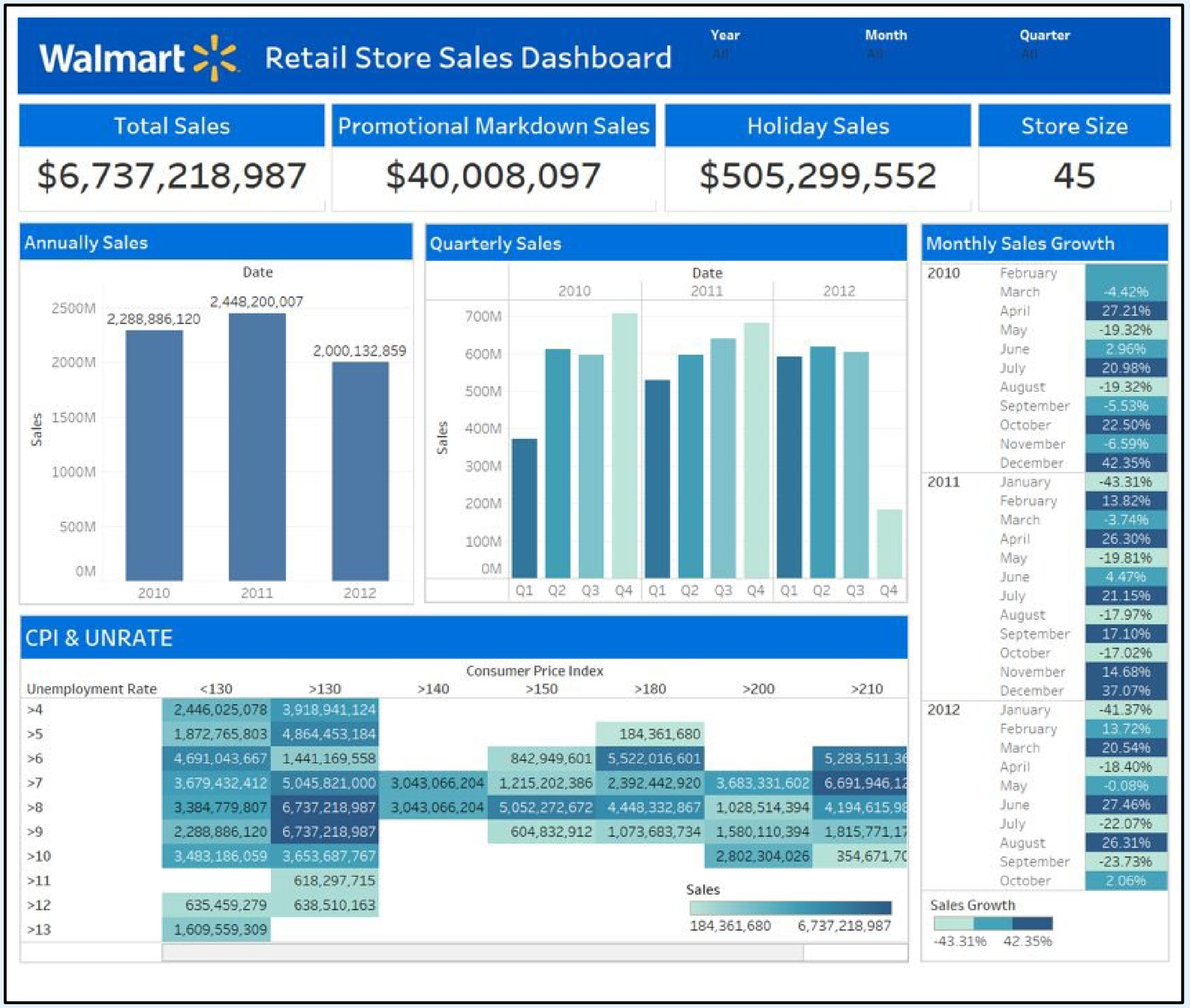
Comparing sum of weekly sales under various unemployment rate and consumer price index to create a matrix table, and 5 stepped colors reveal level of total sales.

Data Insights

In the matrix table, it makes sense that stores with lower CPI generates more sales because the price of groceries is relatively cheap and customers' purchase willingness may be higher. On the other hand, some stores with higher UNRATE also create more than 1 million sales. It could be possible that the price of groceries in Walmart is affordable for customer faced with unemployment issue.

5. Dashboard

- Mark Cards
- Visualizations



Dashboard Demo

In addition to four visualizations, this dashboard includes mark cards on the top, highlighting the overall sales under different situations.

From the visualizations, it's easily to determine annually and quarterly sales. Moreover, given monthly sales growth rate, stakeholders will be aware of human or inventory shortage in holiday seasons and reassign resources in advance. Lastly, CPI and UNRATE are important factor to indirectly affect sales. Therefore, the goal of matrix table is to sheds light on if stores in the region where CPI or UNRATE is higher holds relatively high sales than others.

Reference

- "How to Perform a Sales Analysis (Step-by-Step with Methods & Metrics)". *The Close Sales Blog.* Jan 28, 2020.
- "CRISP-DM". *Data Science Process Alliance.*
- *Walmart Sales Prediction*. Kaggle
- "What Is Extract, Transform, Load? Definition, Process and Tools". Talend - A Leader in Data Integration & Data Integrity.
- "Sales Data Analysis | How to Analyze Sales Data". *Pipedrive*.
- "Consumer Price Index (CPI)". *Investopedia.*
- *Consumer Price Index, 1913-. Federal Reserve Bank of Minneapolis.*
- "Unemployment Rate". *FRED, Federal Reserve Bank of St. Louis*. Jan. 1, 1984.
- *"Got Situational Awareness?". PeakMind.* Nov. 30, 2008.

Thank you