**Project Title:**

Retail Promotional Discount Effectiveness

**Team Members:**

Scott Stevener and Erica Unterreiner

**Project Description:**

The company 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 objective of this dataset is to calculate the markdown and analyze how the profits are affected to determine the best pricing strategies for the products sold. Understanding this information helps stores manage their business better and maintain profitability.

**Data Cleanup & Analysis:**

**Steps to recreate the ETL Process:**

1. **Extract store, sales, and features data from three csv files.**

We obtained historical sales data for 45 stores located in different regions for a single company. The first data set includes discounts applied as well as extraneous information such as the price of fuel, the consumer price index and whether there was a holiday during that week. The second data set includes aggregated weekly sales. The third data set is about the stores.

1. **Transform: what data cleaning or transformation was required.**

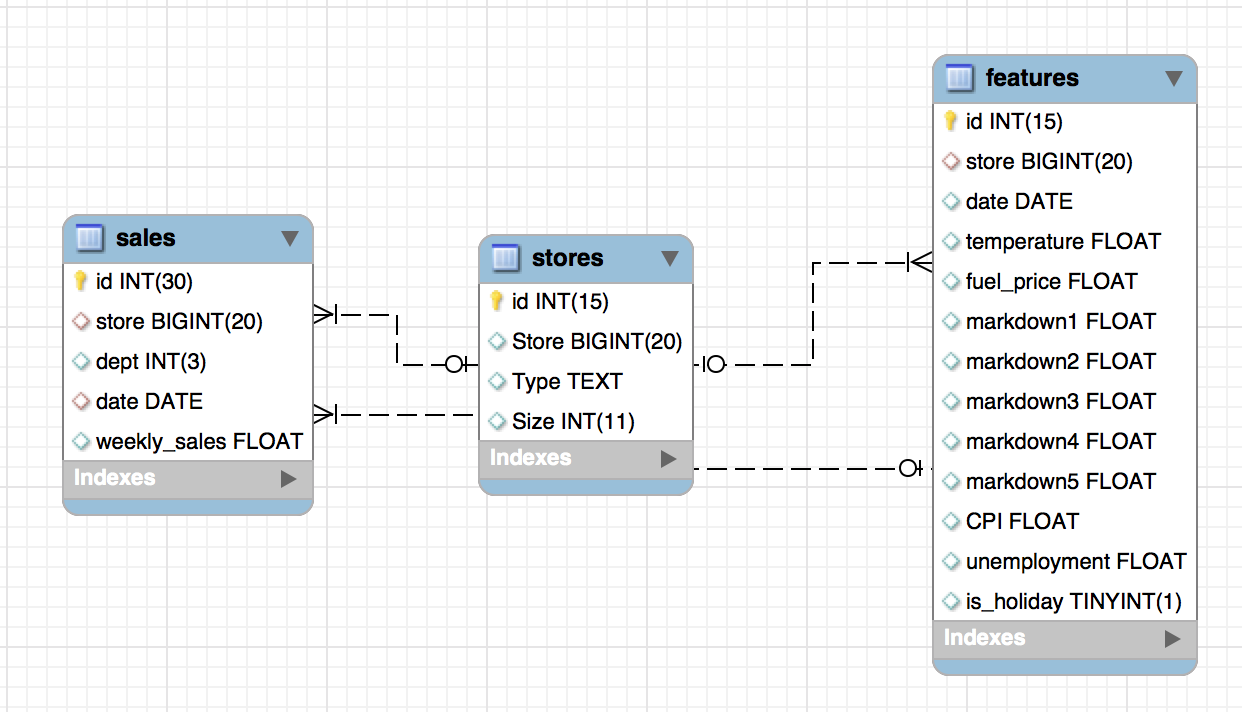
Once imported the data needed to be reviewed for cleanup. Two of the tables include a date field. The dates were not consistently formatted; therefore, a first step was to convert both dates to the same format. Additionally, for weeks when discounting promotions were not in play, the markdown fields contained nulls; therefore, these were filled with zero values.

1. **Load: the final database, tables/collections, and why this was chosen.**

The cleansed data was loaded into a relational database named retail\_db.

**APPENDIX:**

**ENTITY RELATIONSHIP DIAGRAM**

****

**TABLE AND COLUMN METADATA FOR END USER ANALYSTS**

|  |  |
| --- | --- |
| Tables included in Entity Relationship Diagram | |
| Physical Name | Description |
| STORES | Anonymized information about the 45 stores, indicating the type and size of store. Stores has been divided into 3 types A, B, C, based on certain factors (size, sales, etc). |
| FEATURES | Contains additional data related to the store, department, and regional activity for the given dates. |
| SALES | Historical sales data, which covers the time period of January 2012 through November 1, 2012. |
| RETAILED\_COMBINED | Target table with merge of stores, features and sales from January 2012 through November 2012. This table is to be used for calculating the markdown and analyzing how the profits are affected for determining the best pricing strategy for your product.  How to calculate markdown- just imagine an item is priced at $150 at a retail store. 15 percent - Markdown is $22.50 ($150 x .15) and the sale price is $127 ($150 minus $22.50) 20 percent - Markdown is $30 ($150 x .20) and the sale price is $120 ($150 minus $30). Five different markdowns means that price of that specific product has been reduced five (5) times. For example, $50 is the actual price of a product, if there is a 10% markdown, the price becomes $45 (markdown 1), then we may lower the price again and so on. |

|  |  |  |
| --- | --- | --- |
| Columns of table: STORES | | |
| Physical Name | Data Type | Comment |
| ID | INT(15) | The unique identifier of a record on this table |
| Store | BIGINT(20) | The identifier for each store |
| Type | TEXT | Store typing consists of categorizing stores by certain factors (e.g. size, sales, etc.) Valid values include A, B, C |
| Size | INT(11) | Square footage of each store |

|  |  |  |
| --- | --- | --- |
| Columns of table: FEATURES | | |
| Physical Name | Data Type | Comment |
| ID | INT(15) | The unique identifier of a record on this table |
| Store | BIGINT(20) | The identifier for each store |
| Date | DATE | Represents the start of the week |
| Temperature | FLOAT | Represents the average temperature in the region |
| Fuel\_Price | FLOAT | Represents the cost of fuel in the region |
| MarkDown1 | FLOAT | Represents anonymized data related to promotional markdowns. MarkDown data is only available after Nov 2011, and is not available for all stores all the time. Any missing value is marked as zero. |
| MarkDown2 | FLOAT | Represents anonymized data related to promotional markdowns. Any missing value is marked as zero. |
| Markdown3 | FLOAT | Represents anonymized data related to promotional markdowns. Any missing value is marked as zero. |
| Markdown4 | FLOAT | Represents anonymized data related to promotional markdowns. Any missing value is marked as zero. |
| Markdown5 | FLOAT | Represents anonymized data related to promotional markdowns. Any missing value is marked as zero. |
| CPI | FLOAT | Represents the Consumer Price Index. |
| Unemployment | FLOAT | Represents the unemployment rate. |
| IsHoliday | TINYINT(1) | Indicator of whether or not the week contains a holiday. |

|  |  |  |
| --- | --- | --- |
| Columns of Table: SALES | | |
| Physical Name | Data Type | Comment |
| ID | INT(30) | The unique identifier of a record on this table |
| Store | BIGINT(20) | The identifier for each store |
| Dept | INT(3) | The identifier of the store department. |
| Date | DATE | Represents the week |
| Weekly\_Sales | FLOAT | Sales for a given department in a given store. |

|  |  |  |
| --- | --- | --- |
| Columns of Table: RETAIL\_COMBINED | | |
| Physical Name | Data Type | Comment |
| Store |  | The identifier for each store. |
| Type |  | Store typing consists of categorizing stores by certain factors (e.g. size, sales, etc.) Valid values include A, B, C. |
| Size |  | Square footage of each store |
| Dept |  | The identifier of the store department. |
| Date |  | Represents the week |
| Weekly\_Sales |  | Sales for a given department in a given store. |
| Temperature |  | Represents the average temperature in the region |
| Fuel\_Price |  | Represents the cost of fuel in the region |
| MarkDown1 |  | Represents anonymized data related to promotional markdowns. Any missing value is marked as zero. |
| MarkDown2 |  | Represents anonymized data related to promotional markdowns. Any missing value is marked as zero. |
| Markdown3 |  | Represents anonymized data related to promotional markdowns. Any missing value is marked as zero. |
| Markdown4 |  | Represents anonymized data related to promotional markdowns. Any missing value is marked as zero. |
| Markdown5 |  | Represents anonymized data related to promotional markdowns. Any missing value is marked as zero. |
| CPI |  | Represents the Consumer Price Index. |
| Unemployment |  | Represents the unemployment rate. |
| IsHoliday |  | Indicator of whether or not the week contains a holiday. |