**Project Title:**

Promotional Discount Effectiveness

**Team Members:**

Scott Stevener and Erica Unterreiner

**Project Description:**

The objective of this dataset is to understand how profits may be influenced by market conditions and to determine the best pricing strategies (i.e. markdown levels) for various market conditions. Understanding this information helps stores better manage their business and maintain/increase profitability.

**Data Cleanup & Analysis:**

The company runs several promotional markdown events throughout the year. These markdowns precede prominent holidays; the four largest are the Super Bowl, Labor Day, Thanksgiving, and Christmas. Weeks including holidays are weighted five times higher in the evaluation than non-holiday weeks.

**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 market conditions information, such as price of fuel, consumer price index (CPI), and whether there was a holiday during that week. The second data set includes aggregated weekly sales. The third data set contains information about each store.

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

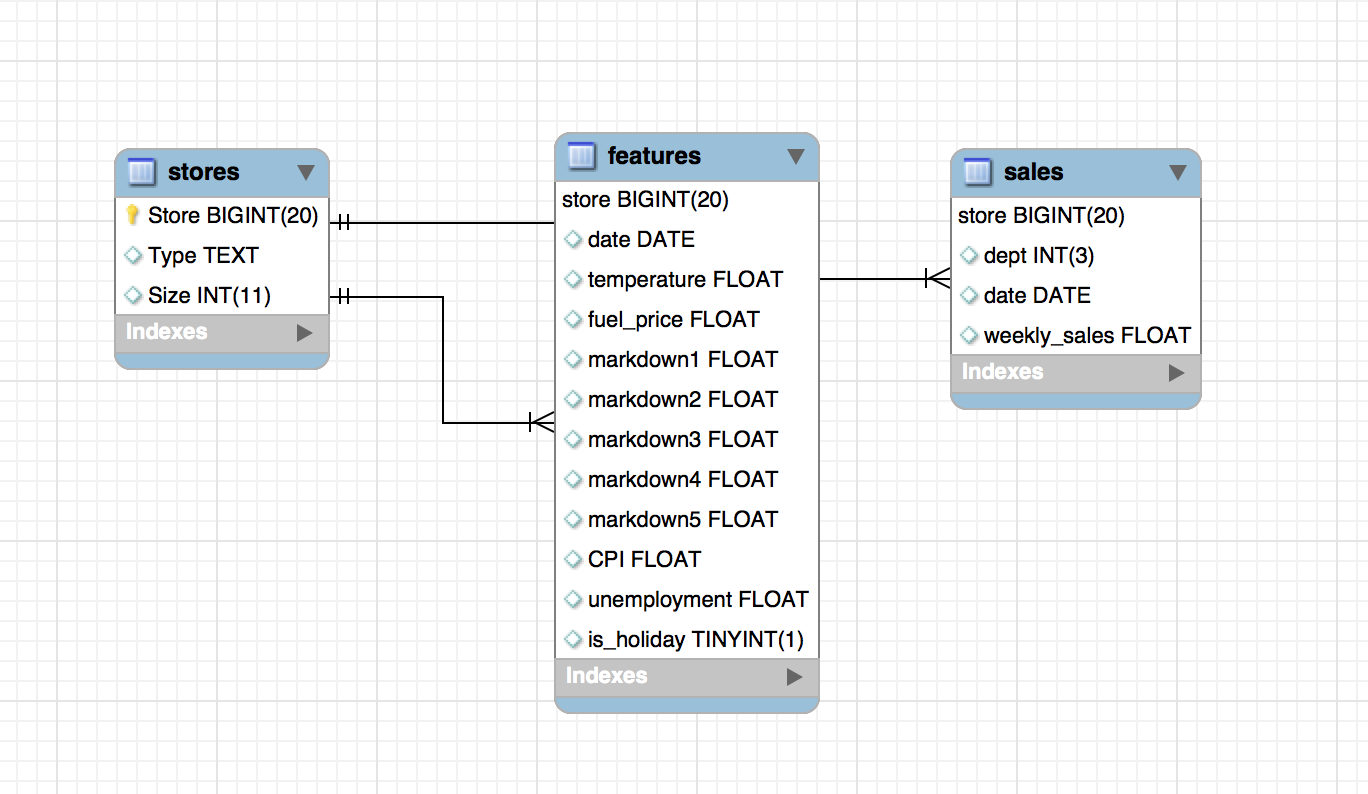
Once imported, the data was reviewed for cleanup. Two tables include a date field, however, these fields are not consistently formatted. Date fields were converted to a matching format. Additionally, in weeks when discounting promotions were not applicable, the markdown field contained null values. The null values were converted to zeros.

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, departments, and market conditions 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 markdowns and analyzing influence of market conditions to determine the best pricing strategies.  Five different markdowns means that the price of a product has been reduced five (5) times.  Calculating markdowns – Using an item originally priced at $150, a 10% markdown is $15.00 ($150 x .10), resulting in a sale price of $135 ($150 minus $15.00). A 20% markdown is $30 ($150 x .20), with a sale price of $120 ($150 minus $30). A 30% markdown results in a sale price of $105; 40% is $90; and 50% is $75. |

|  |  |  |  |
| --- | --- | --- | --- |
| Columns of Table: STORES | | | |
| Physical Name | Data Type | Null Option | Comment |
| Store | BIGINT(20) | Not Null | The identifier for each store |
| Type | TEXT | Not Null | Store typing consists of categorizing stores by certain factors (e.g. size, sales, etc.) Valid values include A, B, C |
| Size | INT(11) | Not Null | Square footage of each store |

|  |  |  |  |
| --- | --- | --- | --- |
| Columns of Table: FEATURES | | | |
| Physical Name | Data Type | Null Option | Comment |
| Store | BIGINT(20) | Not Null | The identifier for each store |
| Date | DATE | Not Null | 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) | Not Null | Indicator of whether or not the week contains a holiday. |

|  |  |  |  |
| --- | --- | --- | --- |
| Columns of Table: SALES | | | |
| Physical Name | Data Type | Null Option | Comment |
| Store | BIGINT(20) | Not Null | The identifier for each store |
| Dept | INT(3) | Not Null | The identifier of the store department. |
| Date | DATE | Not Null | Represents the week |
| Weekly\_Sales | FLOAT | Not Null | Sales for a given department in a given store. |

|  |  |  |  |
| --- | --- | --- | --- |
| Columns of Table: RETAIL\_COMBINED | | | |
| Physical Name | Data Type | Null Option | Comment |
| Store | BIGINT(20) | Not Null | The identifier for each store. |
| Type | TEXT | Not Null | Store typing consists of categorizing stores by certain factors (e.g. size, sales, etc.) Valid values include A, B, C. |
| Size | INT(11) | Not Null | Square footage of each store |
| Dept | INT(3) | Not Null | The identifier of the store department. |
| Date | DATE | Not Null | Represents the week |
| Weekly\_Sales | FLOAT | Not Null | Sales for a given department in a given store. |
| 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. 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) | Not Null | Indicator of whether or not the week contains a holiday. |