Color coding!

RED- confusion, Ask her.

Blue- internet stuff, I am still working on.

Green- Internet stuff or doubt that she clarified.

Train Table

The metadata of the train table is given below.

**Store**

-The store numbers

**Dept**

-The department numbers

**Date**

-The week

**Weekly\_Sales**

-Sales for the given department in the given store,

Feature Table

The metadata of the feature table is given below.

Meta Data of the Data

**Store**

-Store number

**Date**

-Week

**Temperature**

-Average temperature in the region

**Fuel\_Price**

-Cost of fuel in the region

**MarkDown1**

-Anonymized data related to promotional markdowns that Walmart is running.

[*as per author of data set: Weekly sales is the revenue generated by selling stuff in that week. So, the value here represents the amount that Walmart got through sales.*

*Markdowns are usually discount offers kind of thing which are given during special holiday events to boost sales. So, the value here represents the amount of discount given. Markdowns would refer to the total dollar amount of markdowns. There are different types of markdowns (for example, to* ***match a competitor's pricing*** *or for* ***damaged product****), so that's why there are 5 different markdown types. I hope it help.]*

*--as per this logic, the negative value in the Markdown filed might be the overcharged to the customer. (ask her)*

**MarkDown2**

-Anonymized data related to promotional markdowns that Walmart is running.

**MarkDown3**

-Anonymized data related to promotional markdowns that Walmart is running.

**MarkDown4**

-Anonymized data related to promotional markdowns that Walmart is running.

**MarkDown5**

-Anonymized data related to promotional markdowns that Walmart is running.

**CPI**

-The consumer price index

[The Consumer Price Index (CPI) is a measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services.]

*[CPI is used to understand changes in the cost of living over time. When the CPI rises, it indicates that the average price level of goods and services has increased, meaning that inflation has occurred. Conversely, if the CPI falls, it suggests that the cost of living has decreased, indicating deflation.]*

**Unemployment**

-I did not find any description about this field, need to ask her?

-what does number in unemployment mean and what does NA values in unemployment means?

**IsHoliday**

- Whether the week is a special holiday week

Stores Table

The metadata of the stores table is given below.

**Store**

-Stores numbered from 1 to 45

**Type**

-Store type has been provided, there are 3 types — A, B and C.

**Size**

-Stores size has provided

Train and test Table

The metadata of the and and test table is given below.

**Store**

**-**The store numbers

**Dept**

-The department numbers

**Date**

-The week

**Weekly\_Sales**

-Sales for the given department in the given store,

**IsHoliday**

-Whether the week is a special holiday week

Description:

Train dataset is the data set in which we need to work but it does not have a lot of information on it. Each store are categorized into three different types A, B, and C. All the information about the store, its type and size are in the store table so *store* field in the store table is the primary key i.e no values are duplicated in the *store* column in the store table. . We do also have the *store* column in the feature table also which is the foreign key and hence can be duplicated or have null values (but luckily, we don’t have any null values). So, we can connect the train, feature and store table using *store* column, this can be done using R, python, SQL or PowerBI

Now we will join these two tables with our training table.

**Abstract**

This study focuses on the weekly sales of Walmart’s different stores and different departments. The aim is to predict the weekly sales for the last quarter of the 2012 by looking at the sales data from the Jan 2010 to Oct 2012. Walmart’s weekly sale is based on various factors such as average temperature in the region, cost of fuel in the region, store, store’s size, type of the store, department number, MarkDown events i.e kind of promotional events that Walmart runs to increase sales. and finally, whether the week is a special holiday week or not. The dataset includes 421571 rows of data in our training set. To get full access of the data we need to connect the data table feature and store with our train data set. Store table has information about the store where *store* column is the primary key.

**Exploratory Data Analysis (EDA)**

Exploratory Data Analysis (EDA) is a critical step in the data analysis process. It involves examining and visualizing the dataset to summarize its main characteristics, often with the help of statistical graphics and other data visualization methods.

**General Observation from EDA:**

* The biggest stores, Type 'A', have the highest maximum and average sizes.
* The smallest stores, Type 'C', have the lowest maximum and average sizes.
* Type 'B' stores fall in the middle in terms of size.
* Minimum store size is similar across all types.
* Sales increase significantly during the holiday season (November to December) every year.
* On average, each store makes $15,981 in sales.
* Average sales for 2012 appear lower than 2010, mainly because data for the last quarter of 2012 is missing. However, comparing the first three quarters, 2012 has higher average sales than 2010 and 2011.
* Despite fewer holidays compared to non-holidays, average sales during both periods are similar, indicating high holiday sales for Walmart.
* Larger stores (Type A) generally have higher sales compared to smaller ones (Type B and C).
* Most Type C stores are located in regions with higher average temperatures.
* Fuel prices seem consistent across all store types but show an increasing trend over the year.
* 51.11% of stores are Type A, 38.78% are Type B, and the rest are Type C.
* Holiday season doesn't notably affect sales at Type C stores, suggesting they might be in rural areas with fewer people.
* Department 92 is the highest-selling department at Walmart.
* Store 20 is the top-selling store at Walmart.
* Departments 88 to 95 and 38, 40 have notably high sales, indicating they might sell everyday items like food and groceries.
* Some stores with high sales may be located in urban areas.
* And more…

**Final Prediction**

We did try to predict the future sale for the last quarter of 2012 with 95% confidence interval. The the figure the orange shaded region show the confidence interval and Our forecast predicts the highest sales on Nov 23 at $50,965,404.847 with 95% confidence. And I did google it and found that it was black Friday on Nov 12 in 2012, so kind of make sense, why prediction was high on that day!

A screenshot of a computer

Description automatically generated

**Detailed Analysis with Visualization:**

**Here is the dynamic dashboard analyzing the sales data of the Walmart.**