

Ved Nigam (vx190021)
Mitchell Craven (mac190011)
Saketh Dontikurti (sxd190083)

Walmart Retail Sales Analysis – Project Proposal

Team name : Team Great Value

Data Source:

<https://www.kaggle.com/datasets/rutuspatel/walmart-dataset-retail>

We found this dataset on Kaggle and it contains data from 45 Walmart locations in the United States. The goal of analysis on this dataset would be to see if factors such as Holidays, Temperature, Fuel Prices, CPI (Consumer Price Index), and Unemployment in the country affect the weekly sales of the Walmart store. The dataset has 8 variables and 6,435 observations. We will be trying to fit the data on a linear regression model using R.

Variables:

- Response Variables:
 - Weekly Sales: Sales for the given store in dollars
 - CPI: Average cost of a grocery shopper in the United States in dollars
- Predictor Variables:
 - Unemployment: rate in the United States
 - Temperature: in Fahrenheit on the day of the specific Walmart store
 - Fuel Price: price in dollars of fuel for the consumer
 - Holidays: factor variable to see if the week is a special holiday in the United States; i.e Thanksgiving, Christmas, Easter, etc.)

Data Analysis plan: Our plan is to mainly look at the Weekly Sales and CPI as response variables and use the rest as predictors to fit a multi linear regression model. Another possible approach is to build separate linear regression models for each predictor to see which attribute most closely has a relationship with the Weekly Sales. We also might use Holidays as an response variable, and fit it with an multilogistic regression model to see if we can use the data to predict whether a week is classified as a holiday week or not. In that case, we would use weekly sales and CPI as predictor variables.

Ved Nigam (vxn190021)
Mitchell Craven (mac190011)
Saketh Dontikurti (sxd190083)

Responsibilities: We would like to evenly distribute the work among our three team members. We will try to collectively work on the data cleaning and data analysis aspects of the project. Then since we have different response variables we could split up the work by each assigning a different response variable and coming together for the report and presentation. Hopefully we can meet every other week and constantly communicate during class so we are all on the same page.