# 

# 

# **FINAL PROJECT**

**Linear Regression to Predict Weekly Sales**

**By**

# **Tien Ho Bradley Haren Jacob Haussler Muhammed Abubaker Syed**

# 

# **STAT 4020: Regression Analysis**

# **Presented to Shuchismita Sarkar**

Major: Supply Chain Management

Data set source: <https://www.kaggle.com/rutuspatel/walmart-dataset-retail>

We want to predict weekly sales based on temperature, fuel price, and unemployment rate. Since one of our team member’s major and specializations is supply chain management, we thought it would be great practice to apply the model to a real use case.

Furthermore, this kind of modeling helps us understand how we can predict sales behaviour or any other variable related to business with respect to variables that are crucial for its prediction.

The dataset has many columns but the columns that are of interest to us are Weekly\_Sales, Temperature, Fuel\_Price, and Unemployment. We are going to train the model based on the average temperature of the week, average fuel price of the week, and the unemployment rate of that week. The response variable is Weekly\_Sales and the regressors are Temperature, Fuel\_Price, and Unemployment. The dataset was available on Kaggle and the description of the dataset is “Historical sales data for 45 Walmart stores located in different regions are available. There are certain events and holidays which impact sales on each day. The business is facing a challenge due to unforeseen demands and runs out of stock sometimes, due to inappropriate machine learning algorithms. Walmart would like to predict the sales and demand accurately”. However, we will use a particular subset of the dataset to predict weekly sales. The rest of the data that is not the response variable or the regressors are discarded since they do not present any real meaning to the model.