Predicting Weekly Sales for Walmart
Forecasting and Recommendations

Marketing Analytics Ⅱ Project

Team : Anisha Alluru, Khyathi Balusu, Thirumurugan Vinayagam

Under the guidance of
Prof. Jason Duan

The University of Texas at Austin
Texas McCombs
MS Business Analytics
McCombs School of Business

TABLE OF CONTENTS

1. Background and Problem

The Walmart challenge is from a Kaggle competition submitted by Walmart and was previously used as an interview question for incoming candidates. The dataset of historical sales data was given to candidates to try to predict weekly sales for each department for each store.

The marketing problem we are trying to solve is market demand forecasting using historical data. If the market demand can be accurately predicting it would lead to a decrease in inventory, labor, and marketing costs over Walmart stores.

The data was from a Kaggle competition created by Walmart to predict sales for a given department in a given store. The sales are aggregated from each Walmart since the point of sale system records which department each item is from the overall sales are created each week on Friday. The other data field given in the training dataset is the Holiday which is a formula where, if the date is in a given list of pre-chosen dates, it is marked true as a holiday and false if the date is not a pre-chosen holiday. The pre-chosen holidays were the Super Bowl, Labor Day, Thanksgiving, and Christmas.

The data has been produced by Walmart from February 5th, 2010 to November 1st, 2012 which is roughly two years of data split by weeks. The challenge posed by Walmart is to predict the next week's sales for a given department in a given store.

The research questions we aim to answer are the following.

- Is weekly sales able to be forecasted accurately given the data?
- Can predicting weekly sales lead to an increase in growth for Walmart?
- Is there any correlation between the given variables?
- Are sales during a holiday significantly different than sales during a non-holiday season? If so are there any holidays or periods of time that Walmart should designate as a holiday for forecasting?
- Are weekly sales estimates different for each department or can the department be ignored in predicting future sales? If a department is a necessary attribute for forecasting, can different departments have different marketing activities, such as increasing their portfolio of products in a high sales department? If the department is not important for the

prediction can each store have an overall marketing plan reducing resources and complexity for marketing activities?

- Is the type of store important in the overall prediction of weekly sales?
- Will accurately predicting weekly sales reduce costs leading to a change in marketing activities such as in-store ads?

We aim to answer these research questions by exploratory data analysis and modeling the sales and trying to predict the weekly sales for a store. We believe that models that break data into separate bins and time series forecasting will be successful in predicting weekly sales. Also, for exploratory data analysis, we will use correlation matrices, to see if any numeric variables are highly correlated, and plots to show how sales vary between different stores, departments, holidays, and types of stores.

Forecasting market demand is an important part of decreasing uncertainty in the supply chain which helps to decrease costs, whose savings can be diverted to direct marketing activities. Market demand forecasting helps Walmart reduce inventory and labor costs in the US.

As a large retailer market forecasting is an extremely important part of how Walmart keeps inventory costs to a minimum. By forecasting weekly sales for each department Walmart can reduce or increase its order size of a particular product in a given week or month giving it more control over what products reach consumers at a given time.

Forecasting sales in a given week can help with labor scheduling and cost. Walmart employs around 1.5 million people in the US. Since many workers in the industry are part-time or depend on hours set around a week to a month ahead some labor costs can be reduced if demand is predicted to be low in a future week or increased if the demand is expected to spike over the average orders.

These savings in inventory and labor costs can help Walmart divert the savings into direct marketing such as advertisements. However, with just the demand forecast Walmart can change its marketing activities such as changing a product portfolio for a high-performing or a low-performing department in all stores or even changing products at a specific department in one particular store.

2. Data Summary and Exploratory Analysis

2.1. Data Overview

The dataset was obtained from a Kaggle competition, meaning it is secondary data. The total dataset consisted of 421,570 rows and 16 columns, including weekly sales. The columns of the dataset are store, department, date, the Friday date of a week, a given week's sales, a binary holiday variable, temperature, consumer price index, fuel price, five markdown variables, unemployment, type of store, and size of the store.

2.2. Exploratory Data Analysis

We started the exploratory analysis by describing the mean and standard deviation of each numeric column of the dataset, as shown in figure 1 in the appendix. We ignore store and dept since they are factors and not numeric attributes and we ignore the five markdown variables since they were not useful in any model. The figure shows that the average weekly sales across all stores and departments are 15,981 dollars per week with a standard deviation of 22,711 dollars per week. The average size of a store in square feet is 136,727 sq. ft. with a standard deviation of 60,981 sq. ft. Since Walmart has many large stores the average size is high, which we will address with a future bar chart. The average temperature is 60 degrees Fahrenheit with a standard deviation of 18 degrees.

Finally, the three measures of the economy are the fuel price, the unemployment rate, and the consumer price index. The average fuel price is 3.36 dollars per gallon with a standard deviation of 46 cents, the average unemployment is 7.96 percent with a standard deviation of 1.86 percent, and the average consumer price index is 171.2 dollars with a standard deviation of 39.2 dollars.

After describing the dataset, we moved on to the correlation between the numeric variables: sales, size, temperature, fuel price, CPI, unemployment, and the five markdown variables, since we did not know if they would be useful at this point. We will continue to ignore the markdown variables since they were not helpful in the future.

2.2.1. Correlations

Weekly sales had a high correlation with the size of the store with a 24 percent correlation. This is accurate since a larger store would sell more products than smaller sized stores. Size, temperature, and fuel price have no significant correlation with any other variable. CPI and

Unemployment have a high negative correlation with 30 percent. This makes sense since if unemployment is high the consumer price index decreases and vice versa.

	weeklySales	Size	Temperature	Fuel_Price	CPI	Unemployment
weeklySales	1.00	0.24	-0.00	-0.00	-0.02	-0.03
Size	0.24	1.00	-0.06	0.00	-0.00	-0.07
Temperature	-0.00	-0.06	1.00	0.14	0.18	0.10
Fuel_Price	-0.00	0.00	0.14	1.00	-0.16	-0.03
CPI	-0.02	-0.00	0.18	-0.16	1.00	-0.30
Unemployment	-0.03	-0.07	0.10	-0.03	-0.30	1.00

Table 1 : Correlations

2.2.2. Weekly Sales by Type of Store

After looking at the numeric columns, we looked into the weekly sales filtered by different attributes. The first plot, figure 2 in Appendix, shows the aggregate sales per store colored by the store type. The store with the highest sales is Store 20 with 300 million in sales while the store with the lowest sales is Store 33 with around 36 million in sales. The stores with the highest amount of sales are Type A stores while the lowest are generally B and C however it still includes some Type A stores.

Next, the overall sales were split by the type of store in figure 2. The total amount of sales is 6.7 billion in sales of which Type A consists of the most, at 4.3 billion, Type B is next at 2 billion, and Type C is 0.4 billion.

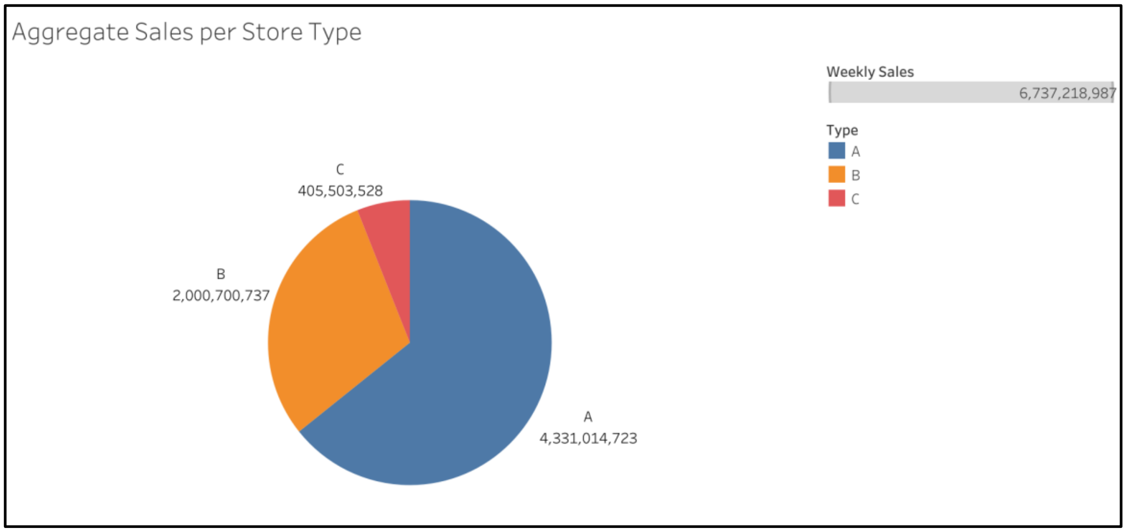

Figure 2: Aggregates Sales - Store Type

Using these two figures we came to the conclusion that Type A stores are the largest Walmart stores, such as supercenters, Type B are normal Walmart stores, and Type C stores are smaller stores or the neighborhood grocery stores. To test this hypothesis we created a plot of the size of the stores colored by the type of stores. The plot, figure 3 in the appendix, shows that the prior hypothesis is correct however there are two Type A and two Type B stores in the smaller end of store size.

After looking at sales filtered by stores, we looked at sales filtered by department. As shown in figure 4 in the appendix, the department with the highest sales is Department 92, with a close second of Department 95, with over 480 million in sales each while the department with the lowest sales has only 14 sales. This shows that not all departments are equal in terms of sales meaning that the department should be an important component in a given model.

2.2.3. Holiday Sales

The final variable we analyzed is the binary holiday variable. First, we plotted overall sales when it is a holiday versus when it is not a holiday. As expected the overall sales are much higher when it is not a holiday, especially since there are only four holidays coded in the data. However, the four holidays do contribute 7.5 percent of the overall sales.

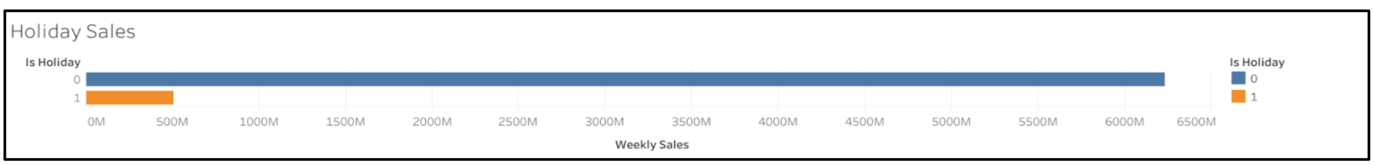

Figure 3 : Holiday Sales

2.2.4 Weekly Sales Trend

The final plot we created plots the weekly sales in a line graph split by the type of store and the sales split by the holiday column. For Type A and Type B stores, there is a large peak around December and January which is reasonable since it is the largest retail season. There is also a large peak in the holiday sales, which represents Black Friday and the Christmas season for both Type A and Type B stores. There is also no trend for Type A and Type B stores. For the Type C stores, the overall sales do not have any significant spikes but it seems to have an upward shift or trend in 2012. The strangest thing about Type C stores is that the sales actually decrease during the holiday season, which might be plausible since Type C stores are smaller stores meaning people shop at the large stores which takes away sales from the Type C stores.

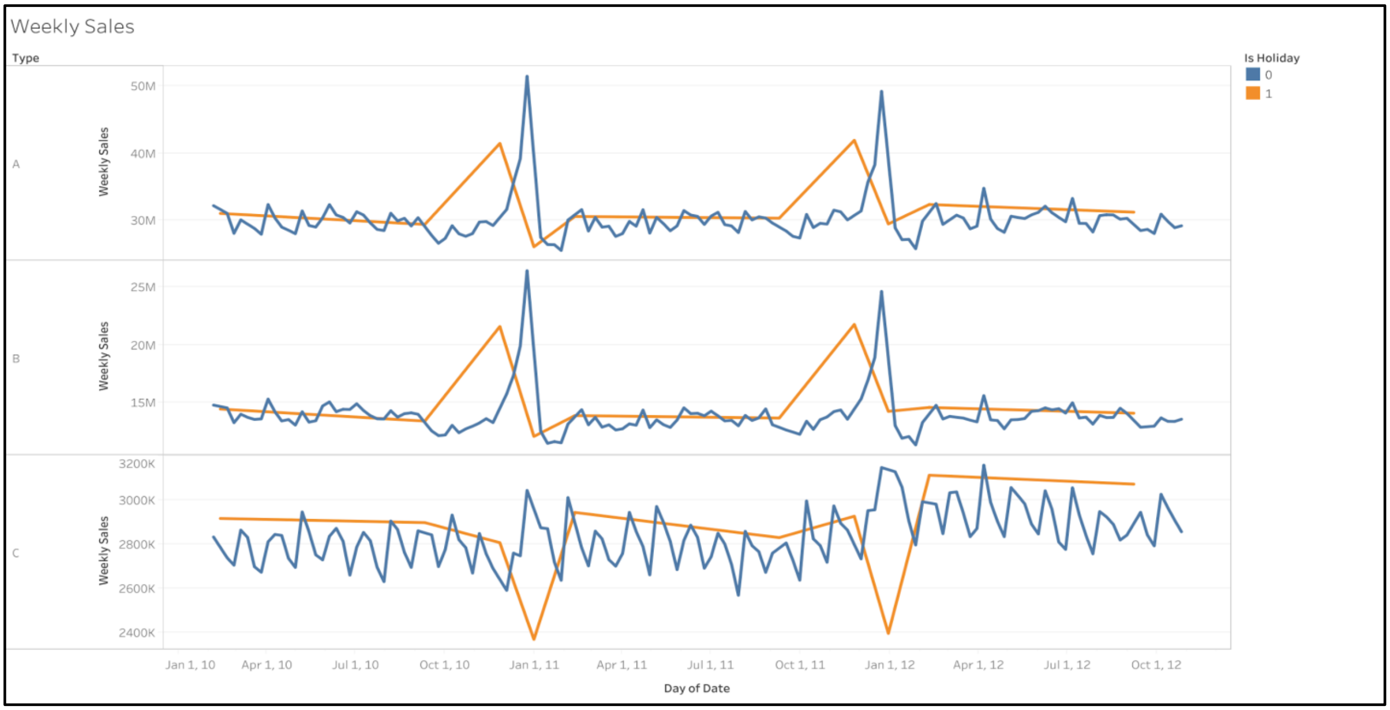

Figure 4: Weekly Sales OVerview

3. Data analyses, Key findings, Conclusions

3.1 Choice of Models

The goal being to predict weekly sales, we treated this as a regression problem and chose models accordingly. We built the following models for the below reasons:

- **Linear Regression** - It's a simple model with an interpretable equation and the coefficients give us the direction of influence on the dependent variable.
- **Random Forest** - This tree model gives more accurate results as it is an ensemble of many individual models. The random selection of features to build each model makes the trees built less correlated and so this could improve the results. Also, it gives the important features by using information gain.

- **Decision Tree** - This tree model is high on explainability. The step of caution is that the predictions are sometimes coarse as one prediction is made for an entire subdivision of the feature space and is also susceptible to overfit.
- **XGB** - XGB is an advanced version of Gradient Boosting. It is highly effective in handling weighted data. Gradient Boosting would use up higher computation power for larger datasets and Extreme Gradient Boosting (XGB) comes into picture when lower CPU power needs to be used.
- **Long Short-Term Memory Model** - LSTM is used on data that deals with time series elements, especially with larger datasets. It is highly effective in using optimum computation power, while rendering statistically accurate results

3.2 Choice of Metric

The metric considered to evaluate the performance of the model was RMSE (root mean score error)l. It's defined as,

$$RMSE = \sqrt{(f - o)^2}$$

where, f is forecasted values or expected values and o is observed values. It is essentially the standard deviation of the residuals. It is a numerical measure of how spread out are these residuals.. This score is high when our predicted value is not close to actual values and vice versa. Lower score shows a more accurate model, excluding cases that deem over fitting.

3.3. Models

3.3.1. Linear Regression Model

The first model that was run was a standard linear regression model. This model is basically a linear approach to modelling the relationship between a dependent variable and one or more independent variables which hold some explanatory power. In this case, we have more than one explanatory variable. Hence, it is a multiple linear regression.

As expected, this model could not fit the highly complex data at hand. The linear regression model gave a **test root mean square error of $21,745.77 which was the highest** among all the models that were run.

3.3.2. Random Forest Model

Random Forest is an ensemble tree-based algorithm which creates multiple decision trees from randomly selected subsets of data from the training set. These decision trees are ultimately aggregated on the basis of votes from the different trees to finalize the best tree.

After tuning the random forest model using GridSearchCV, number of trees as 10 and minimum sample split of 4 gave us a **root mean square error of $4662.29** on the test set.

A visualization of the fit for each individual test record can be seen below:

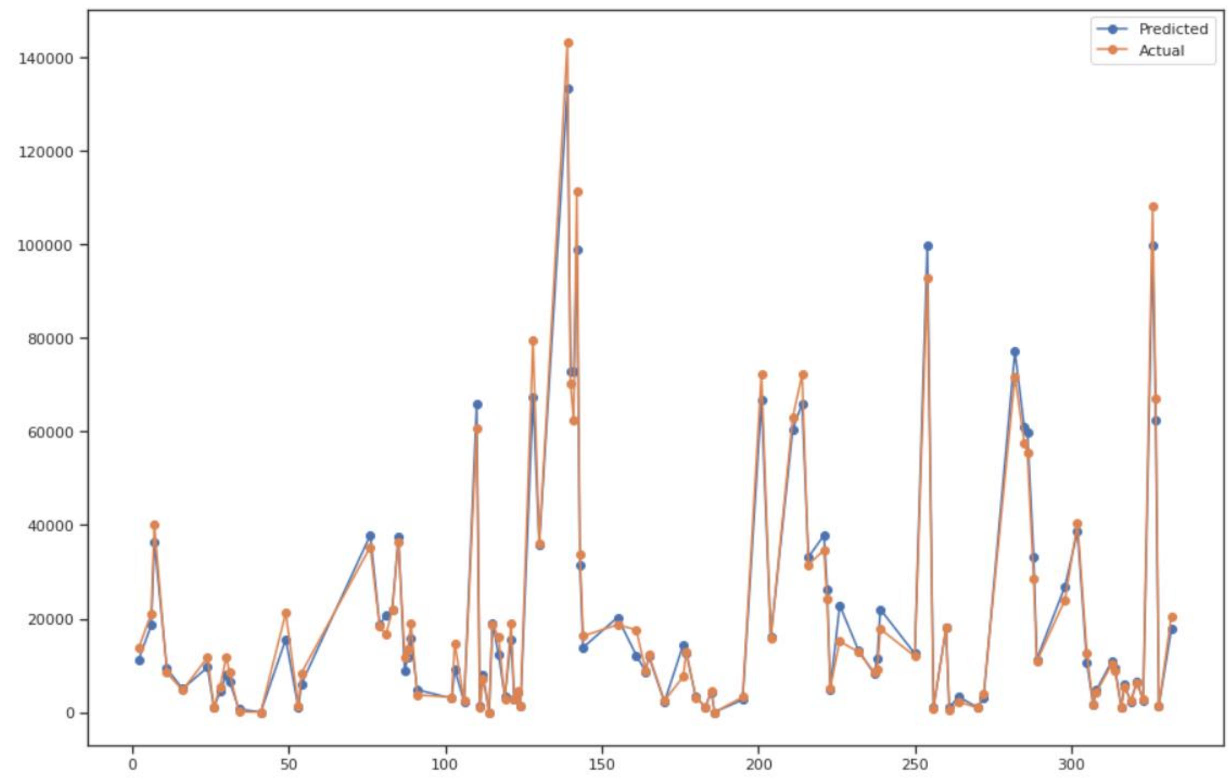

Figure 5: Random Forest

3.3.3. Decision Tree Model

We then used the Decision Tree algorithm to regress weekly sales values.. Decision Tree uses recursive partitioning to split the feature space based on decision thresholds and identify important features using information gain from the feature. Both categorical and numerical data can be handled by Decision Tree and for this data a tree with zero depth gave us a **root mean square error of $5558.45**

A visualization of the fit for each individual test record can be seen below:

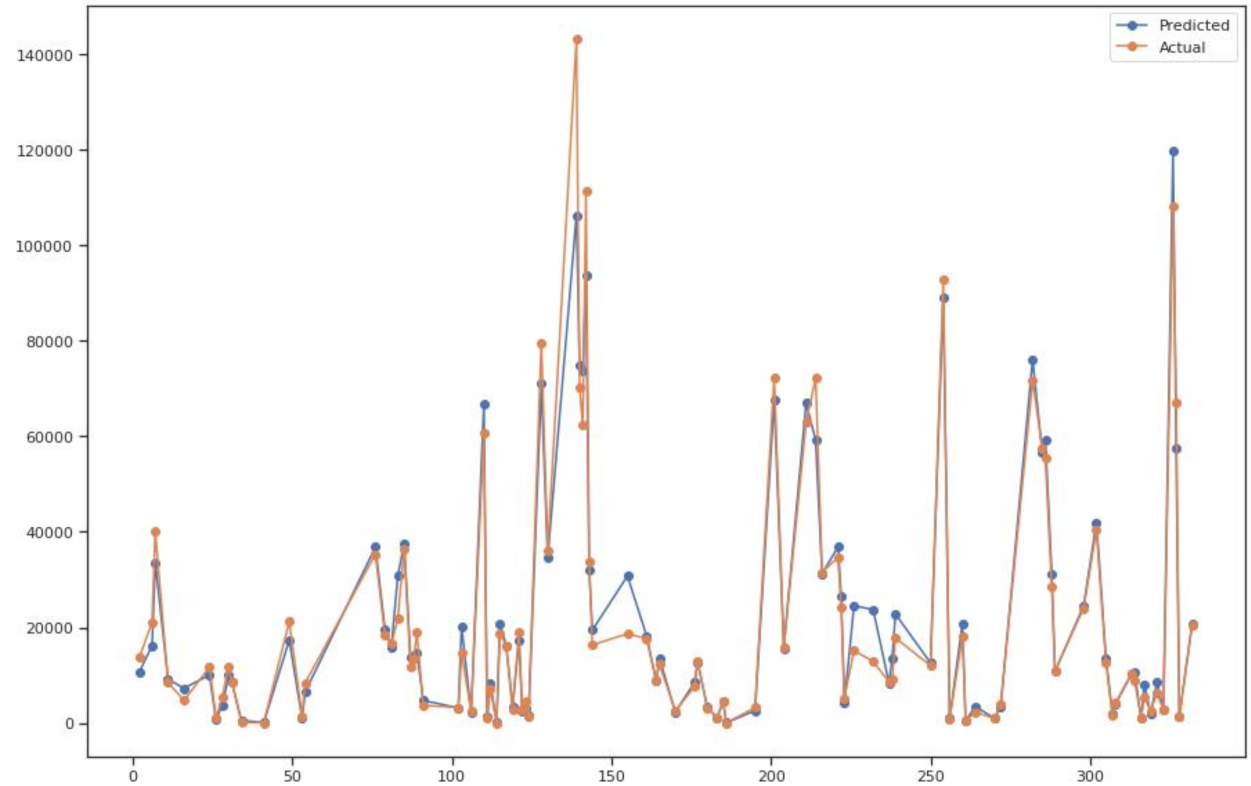

Figure 6: Decision Tree

3.3.4. XGB Model

GB learns and computes new residuals at every step based on predictions made on previous steps . These will be used as leaves for the next tree. This process goes on until iterations and estimators match. The final residual is the mean of all the residuals at every step. For this model, we have given 100 estimators which mean 100 iterations take place. Gradient Boosting uses depth greater than 1 and for our model we have depth as 6. Extreme Gradient Boosting was performed with **a root mean square error of $5321.36**, higher than random forest.

A visualization of the fit for each individual test record can be seen below:

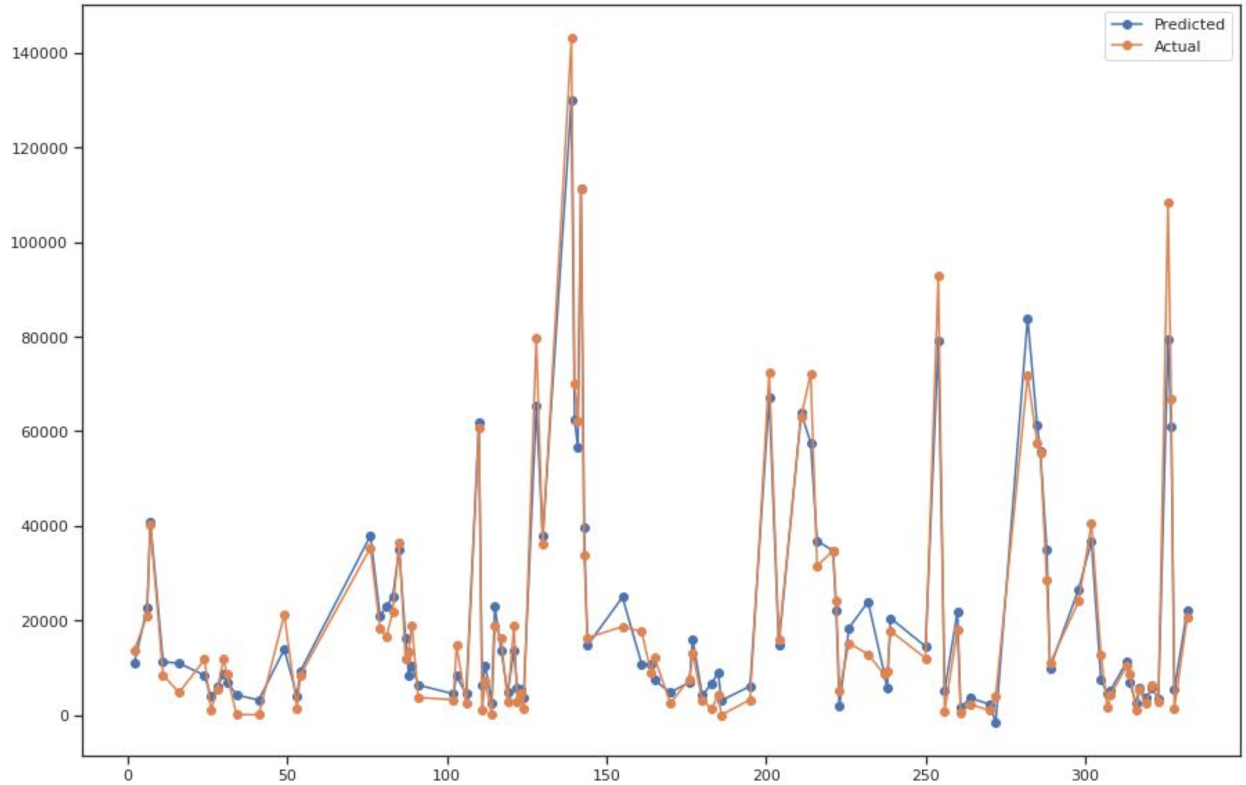

Figure 7: XGB Model

3.3.5. Long Short-Term Memory Model

We decided to create an LSTM since the data is a time series and to take advantage of the small computing time for a neural network. For the LSTM model, the dataset had to be reshaped into each row corresponding to a given date and each column corresponding to a department in each store. To convert the columns into departments per store, we started by creating a unique id for each department in each store, meaning department 1 in store 1, department 1 in store 2, and so on. This process output 3331 unique departments per store. After creating the department per store, which will be called location, we merged the new dataset with the old dataset to create a location column in the large dataset. Using the pivot table function we set date as the index for the pivot table, the location id as each column, and the weekly sales as the values in the table.

This overall pivot table was the dataset for the long short term memory model. The dataset was converted to a series of arrays with each date being an array. The array dataset was split into training and testing, with 70 percent of the data used for training. The model would take three arrays from the training set and predict the next array, which is defined in the n_steps. The three arrays would be split by the split_sequences function which takes a given input and splits it into x and y variables for the model. In this model, the split_seqences function creates the trainX, trainy, testX, and testY that is used for fitting and prediction for the LSTM model.

Next, the Model is called as a Sequential model from the Keras library. The Keras library is an overlay of the Tensorflow library in Python which is often used in machine learning. The first layer added to the model is a Bidirectional LSTM layer with 100 neurons, the default for the LSTM layer, and the Rectified Linear Unit Activation function. The first layer excepts an input vector of size (3,3331). The output of this layer is hidden and passed to the Dense layer returns an output array of the same size. The final part of the model is compiling the model. For the model, we decided to use the adam optimizer since the gradient-based optimization is generally efficient yet still works well in practice. The loss is mean squared error since our comparison between all the models is the root mean squared error (RMSE).

Now that the model has complied, we fit the model to trainX and trainy, running for 500 epochs, which took around five minutes in our case. After fitting the model, we predicted with both the training and testing sets to create a training RMSE and a testing RMSE. It should be noted that since neural networks learn from the data they are given the training and testing RMSEs are different each time you run the model regardless of the data. After running and using the model to predict a few times, **the lowest test RMSE, which is the metric used to compare all of the models, was $4432**, which was the lowest error of all of the models we ran. This means that the LSTM can predict weekly sales within $4432 of the actual value.

3.4 Model Comparison

Comparing the scores on the validation set for the various models, we see that Random Forest and LSTM have the lower RMSE. LSTM has lower RMSE than that of Random Forest. So, we select the LSTM model to base our recommendations on.

Model	MAE ($)	MSE ($)	RMSE ($)
Linear Regression	14556	475M	21745.77
Random Forest	1779.29	22M	4662.29

Decision Tree	2146.55	35M	5558.45
XGB	3065.33	28M	5321.36
LSTM	**2072.94**	**18M**	**4432.66**

Table 2: Result scores from models

A visualization of the metrics for each model can be seen below:

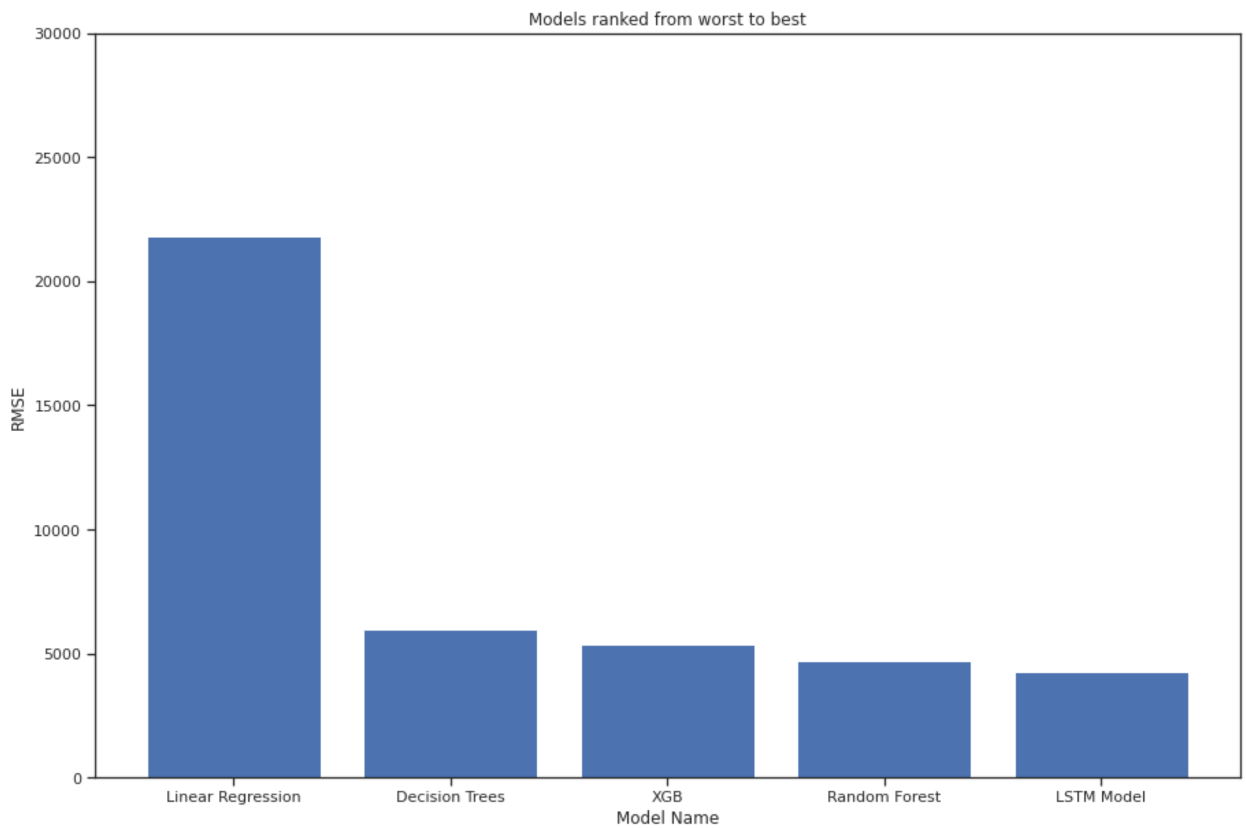

Figure 8: XGB Model

4. Marketing strategy recommendations, limitations and future research directions

4.1. Recommendations

We recommend Walmart to deploy the LSTM model to predict future sales every week. The weekly sales are able to be forecasting accurately within $4432.66 of the actual value of next week's sales. If Walmart is able to deploy the model over all of its stores it would lead to inventory and labor savings for Walmart compared to using no market demand forecasting methods.

There are no correlations that Walmart can influence which would lead to market growth or a change in sales. However, we recommend Walmart looks into holiday sales at Type C stores since those stores, unlike all other stores, have lower sales during the holiday season. We also recommend that Walmart adds Black Friday and New Years to their list of holidays. The sales during the current holidays that are a part of the data are significantly different than the sales during the non-holiday season.

As per the department attribute, departments cannot be ignored in predicting future sales. Therefore, since department is a necessary attribute for forecasting, different departments must have different forecasts for future sales. This means that each department should have unique marketing activities, such as increasing their portfolio of products in a high sales department or decreasing its portfolio for departments with low sales. Since the department sales forecast is different for each store it means that each unique department in every store should have a different marketing plan and each store cannot have an overall marketing plan. Although this means that resources and complexity for marketing activities cannot be reduced by aggregating marketing activities per store, our forecasting method does mean that savings can be realized through changing departments independent from the store number.

Although a limitation of the LSTM model is that the output cannot directly be interpreted we can see that the type of store is important in the overall prediction of weekly sales. Through exploratory data analysis we saw that Type A and Type B stores are large stores with a spike in sales during each holiday season and during November and January, which we recommended Walmart to add to the holiday list as Black Friday and New Years respectively. Type C stores are distinctively different from the other stores since there are no large spikes in the data and a large decline in sales during the holiday season. Our analysis recommends that the type of store should be included in the model for weekly sales, which it is in the LSTM model.

Our final question was if accurately predicting weekly sales reduces costs leading to a change in marketing activities such as in-store ads. Since our recommendation was that Walmart deploy the LSTM model, our question is now if our model would reduce costs for Walmart. We can say that costs such as inventory and labor costs can be reduced. Also since our model forecasts market demand for each unique department and store, Walmart can further reduce costs by changing the size of departments or the departments included in each store.

Through our recommendation of the deployment of the LSTM model we have solved our marketing problem of predicting market demand forecasting using historical data and decreasing inventory, labor, and marketing costs over Walmart stores.

4.2. Limitations

- Though the models are quite accurate, they do not exactly predict the weekly sales differently for different stores. Each store has a different average value for sales which isn't captured well enough in these models.

- One of the limitations of the LSTM model is that there is no way to directly interpret the output without transforming it.

- As the dataset used here is relatively small, the loss difference is not extraordinary . However, in large datasets which may run into gigabytes, this trick of simple averaging may reduce the loss to a great extent.

4.3. Future Research

- In order to harness the nature of the data in a better way, it would be wise to run other time-series models as well. One of the most popular time-series models is the ARIMA model which would be able to predict the sales based on past values.

- Along with this, another possibility would be to run a mixed effect model which will be able to differentiate between the different stores and predict the weekly sales differently for each store. This would be particularly useful as each store has a distinct demand which drastically affects its sales.

- In order to make the modelling process more efficient, some modifications can also be done to the data in the future. This includes, splitting the date feature into three columns - days, month, weeks.

- Additionally, the dataset includes special occasions such as Christmas, pre-Christmas, Black Friday, Labour day, etc. People obviously tend to shop more on these days than usual. Adding these as a new feature to data will also improve accuracy to a large extent.

- There is also a missing value gap between training data and test data with 2 features i.e. CPI and Unemployment. If this gap is reduced, the performance can ultimately be improved.

5. Appendices

5.1. Exploratory Data Analysis : More visualizations

5.1.1 Statistics

	mean	std
Store	22.200546	12.785297
Dept	44.260317	30.492054
weeklySales	15981.258123	22711.183519
Size	136727.915739	60980.583328
Temperature	60.090059	18.447931
Fuel_Price	3.361027	0.458515
MarkDown1	7246.420196	8291.221345
MarkDown2	3334.628621	9475.357325
MarkDown3	1439.421384	9623.078290
MarkDown4	3383.168256	6292.384031
MarkDown5	4628.975079	5962.887455
CPI	171.201947	39.159276
Unemployment	7.960289	1.863296

Figure 1

5.1.2. Bar Plot for Weekly Sales on Type of store

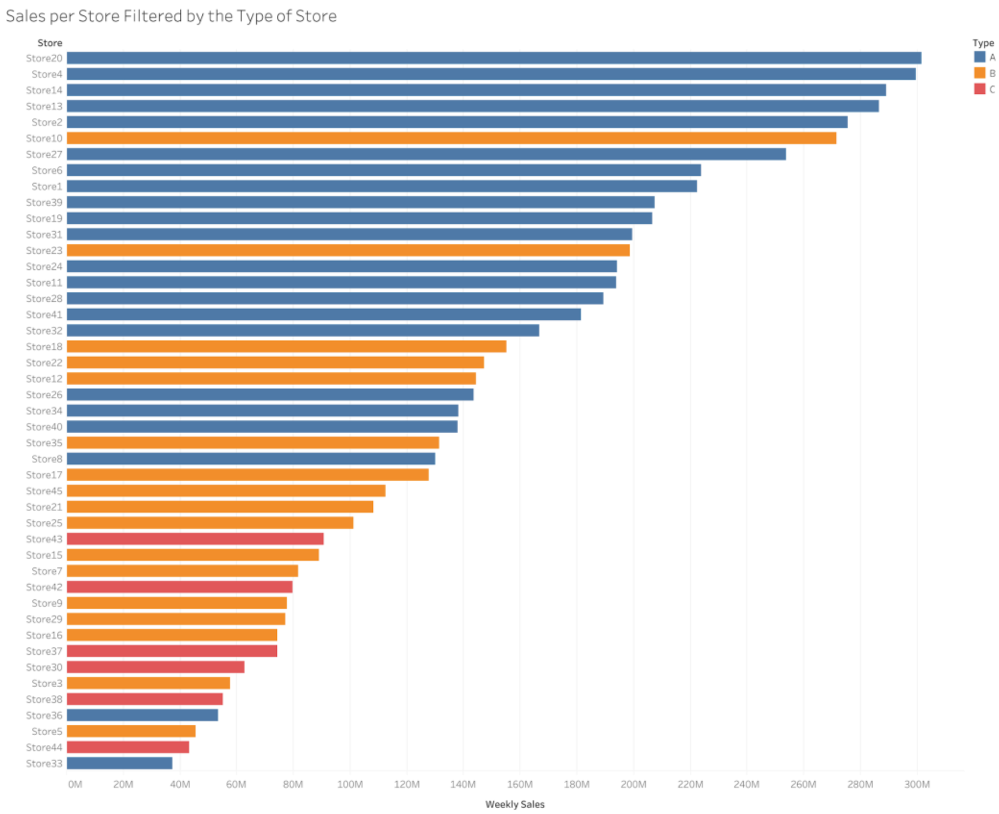

Figure 2

5.1.3. Bar Plot for Weekly Sales on Size of Store

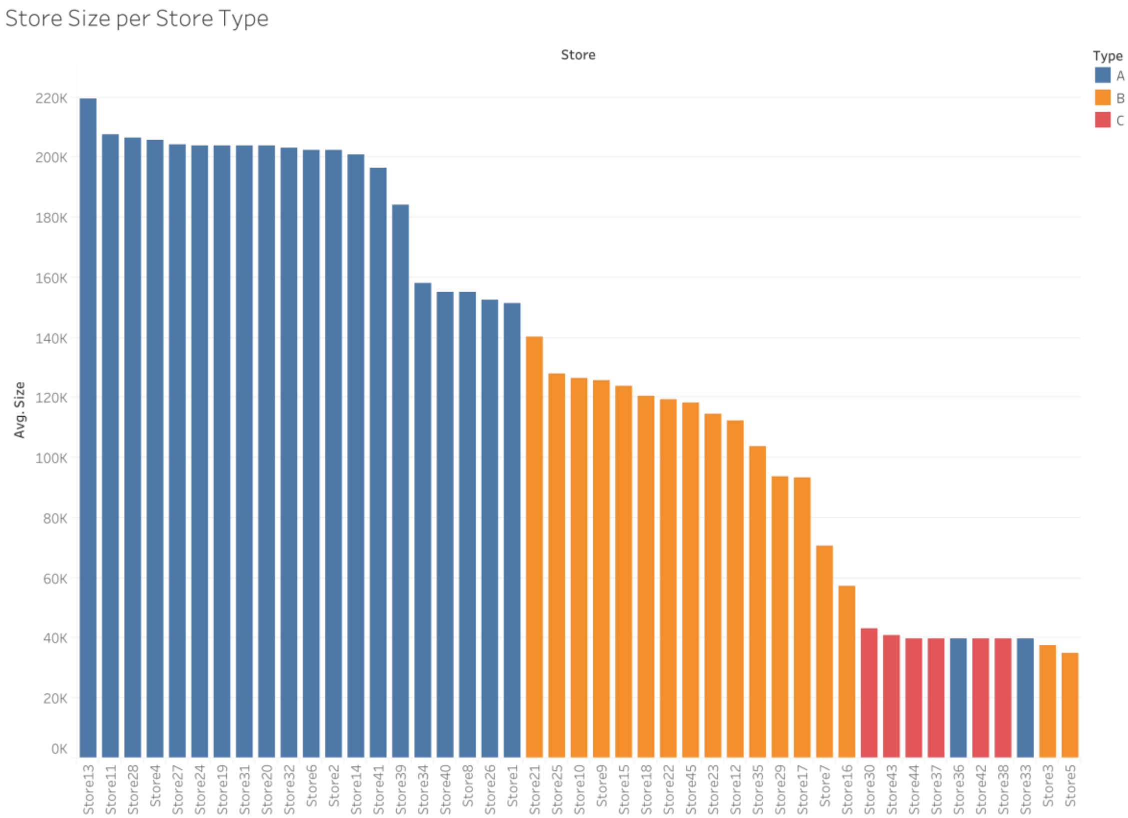

Figure 3

5.1.4. Department Sales

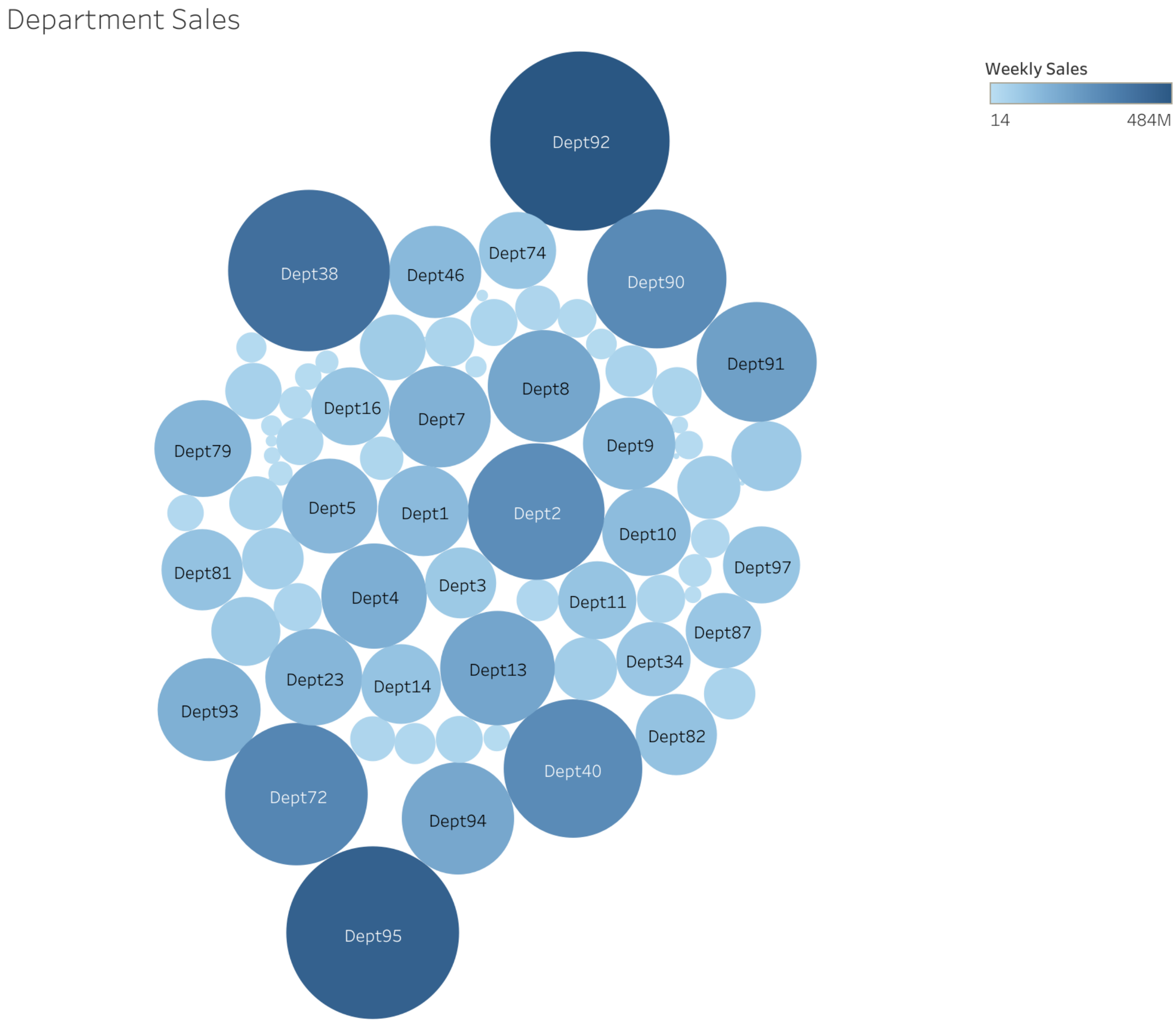

Figure 4

6. References

1. Source: https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting/
2. https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-series-forecasting/
3. https://www.statisticshowto.com/probability-and-statistics/regression-analysis/rmse-root-mean-square-error/
4. https://xgboost.readthedocs.io/en/latest/python/python_api.html#module-xgboost.sklearn