**Problem:**

There are many seasons that sales are significantly higher or lower than averages. If the company does not know about these seasons, it can lose too much money. Predicting future sales is one of the most crucial plans for a company. Sales forecasting gives an idea to the company for arranging stocks, calculating revenue, and deciding to make a new investment. Another advantage of knowing future sales is that achieving predetermined targets from the beginning of the seasons can have a positive effect on stock prices and investors' perceptions. Also, not reaching the projected target could significantly damage stock prices, conversely. And, it will be a big problem especially for Walmart as a big company.

**Aim:**

My aim in this project is to build a model which predicts sales of the stores. With this model, Walmart authorities can decide their future plans which is very important for arranging stocks, calculating revenue and deciding to make new investment or not.

**Solution:**

With the accurate prediction company can;

* Determine seasonal demands and take action for this
* Protect from money loss because achieving sales targets can have a positive effect on stock prices and investors' perceptions
* Forecast revenue easily and accurately
* Manage inventories
* Do more effective campaigns

**Plan:**

1. Understanding, Cleaning and Exploring Data
2. Preparing Data to Modeling
3. Random Forest Regressor
4. ARIMA/ExponentialSmooting/ARCH Models

**Metric:**

The metric of the competition is weighted mean absolute error (WMAE). Weight of the error changes when it is holiday.

***Understanding, Cleaning and Exploring Data:*** The first challange of this data is that there are too much seasonal effects on sales. Some departments have higher sales in some seasons but on average the best departments are different. To analyze these effects, data divided weeks of the year and also holiday dates categorized.

***Preparing Data to Modeling:*** Boolean and string features encoded and whole columns encoded.

***Random Forest Regressor:*** Feature selection was done according to feature importance and as a best result 1801 error found.

***ARIMA/ExponentialSmooting/ARCH Models:*** Second challange in this data is that it is not stationary. To make data more stationary taking difference,log and shift techniques applied. The least error was found with ExponentialSmooting as 821.

**Findings:**

* Although some departments has higher sales, on average others can be best. It shows us, some departments has effect on sales on some seasons like Thanksgiving.
* It is same for stores, means that some areas has higher seasonal sales.
* Stores has 3 types as A, B and C according to their sizes. Almost half of the stores are bigger than 150000 and categorized as A. According to type, sales of the stores are changing.
* As expected, holiday average sales are higher than normal dates.
* Top 4 sales belongs to Christmas, Thankgiving and Black Friday times. Interestingly, 22th week of the year is the 5th best sales. It is end of May and the time when schools are closed.
* Christmas holiday introduces as the last days of the year. But people generally shop at 51th week. So, when we look at the total sales of holidays, Thankgiving has higher sales between them which was assigned by Walmart. But, when we look at the data we can understand it is not a good idea to assign Christmas sales in data to last days of the year. It must assign 51th week.
* January sales are significantly less than other months. This is the result of November and December high sales. After two high sales month, people prefer to pay less on January.
* CPI, temperature, unemployment rate and fuel price have no pattern on weekly sales.

More detailed finding can be found in notebooks with explorations.

**Future Improvements:**

* Data will be made more stationary with different techniques.
* More detailed feature engineering and feature selection will be done.
* More data can be found to observe holiday effects on sales and different holidays will be added like Easter, Halloween and Come Back to School times.
* Markdown effects on model will be improved according to department sales.
* Different models can be build for special stores or departments.
* Market basket analysis can be done to find higher demand items of departments