**Sales Prediction and Forecast Volatility according to the Dataset**

**Name:**

**Student ID:**

**The course of Study:**

**Current Situations:**

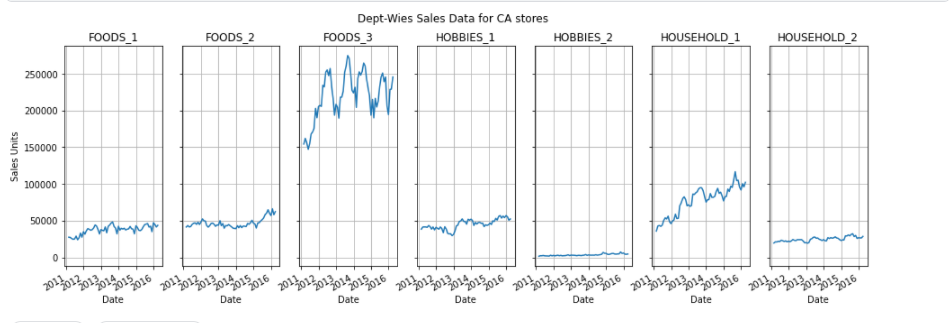
Prediction and forecasting has became a very important parts of the organizations since the market competition became too much harsh and new companies are arriving everyday in the same domain of market these domains can IT, BFSI, FMCG and management consultancy. Most of the companies assigns the statisticians for the forecasting their growth as per the current performances of the employees and the daily sales. Nowadays many companies and organizations are trying to grow their sales of the product it doesn’t matter whether it is management firm or an FMCG their major focus is on the market growth of their product sale and to make their strong consumer base. The job of the statisticians is to create an estimate the average growth of the company according to the current performance nowadays companies hire the data scientists and analysts for this job. In the earlier days analysis and forecasting used to require a lot of calculations and mathematical methods to predict the accurate result which contains a lot of mistakes and sometimes it used to take time to estimate the result.

Nowadays the libraries of python like numpy, tensorflow, pandas, Matplotlib, scikitlearn etc are used to plot and predict the perfect results, the tools like tableu, SPSS etc are the plus in the prediction of the data and providing perfect results. Now all the companies need is to hire the perfect analyst who has got good hands at coding and tools like tableu, SPSS, PowerBi etc.

Our aim is to check the volatility of the predicted results as per the products sales and the users’ interests. For this task we searched the dataset, the dataset which is taken for the particular process is taken from kaggle repository, the particular dataset had many discrepancies for that we prepared the dataset by removing the null and redundant values from the dataset to make it ready to use for the programming and prediction. After the preparation of the dataset we require the particular platform for running the programming and for that we used google colab which is just like the jupyter notebook but better than that as it lessens the CPU runtime and lagging of the system which is drawback of the jupyter notebook. After the selection of coding platform we import the libraries of python these libraries are pandas, tensorflow, matplotlib.pyplot and scikit-learn. These libraries are used to predict the result and forecast according to the interests of users.

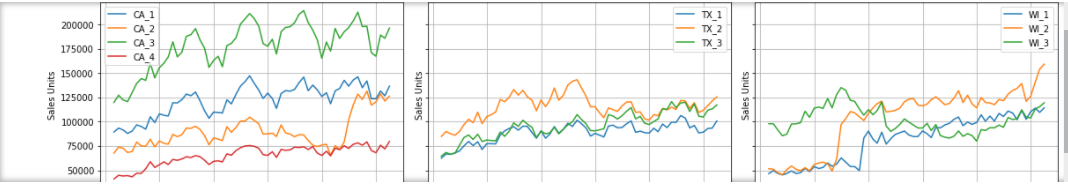
The sales prediction is categorized according to the interests of the users which is “Food”, “Hobbies” and “Household”. The dataset contained the sales data of three major states of US these states are Texas, California and Winchester. We observed the sales as per the categories show that most of the sales of the product is based on foods and cooking materials, which shows that most of the sale is based on the FMCG companies and these companies are taking the major profits from their users as well as from their vendors.

The following graph shows the sales according to the department or categories of the products.



Here we can clearly see that the food department are in big profits and after that most of the sales are done in the household materials of the store and least money has been spent on the hobbies by the users.

After that the second analysis we have done is on the states, to see which state is spending money on food and which is spending more on hobbies and household materials. The following graph shows the statewise sales according to the dataset.

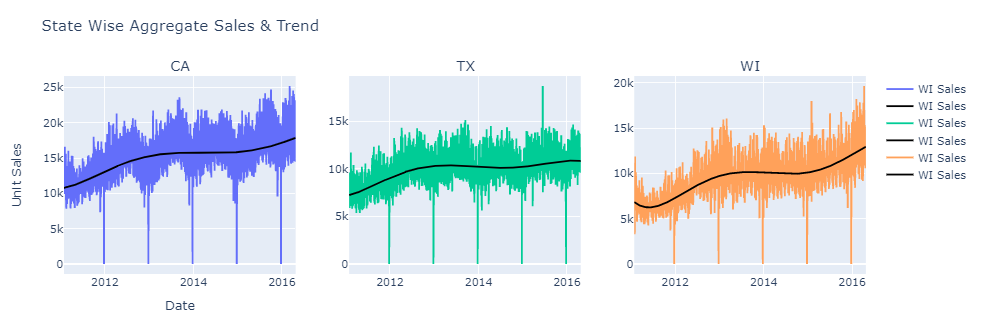


Here we can see that again food sales won the race while household is on the second place, one more thing which we can see that sales level in the Texas(Tx) is equally placed for all the three categories while on the other two we can see that food sales are touching the peaks and others are too low than that.

After this we did the aggregate the sales calculations for that we trained our model for the aggregate calculations and to show the stats of the aggregate calculations according to the aggregate sales in the different states the following code shows the training of the model.



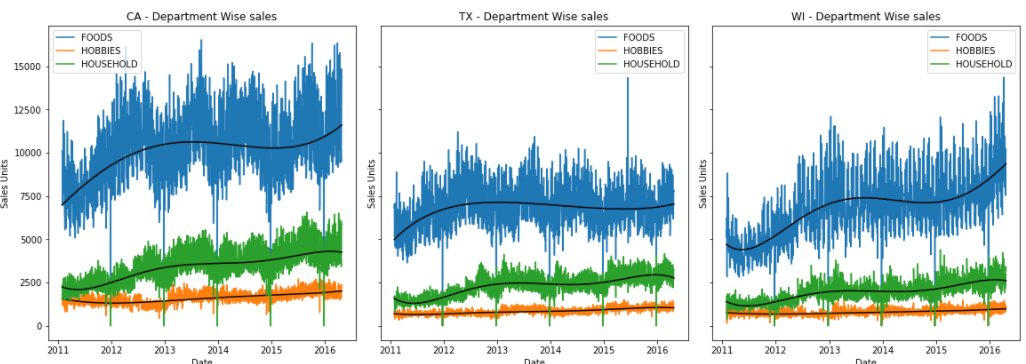
The aggregate sales graph is shown below which tells us the aggregate sales happening in different states ie aggregate statewise sales,



After that we calculated aggregated sales as per the categories happened in different states the following code shows the training of model to show the aggregate statewise sales according to the categories.



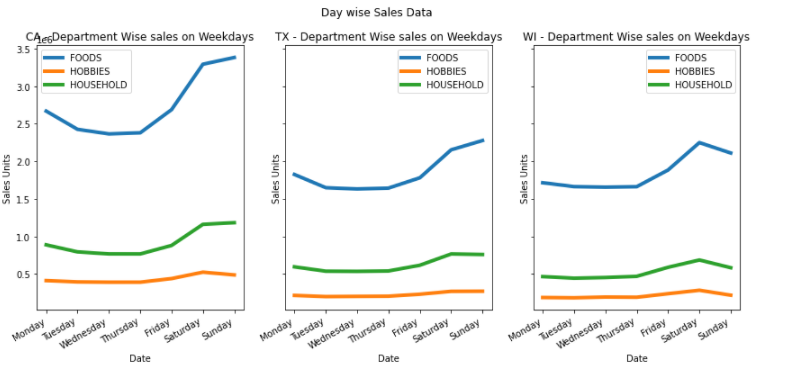
The calculated results below show the aggregate prediction results of sales as per the categories.



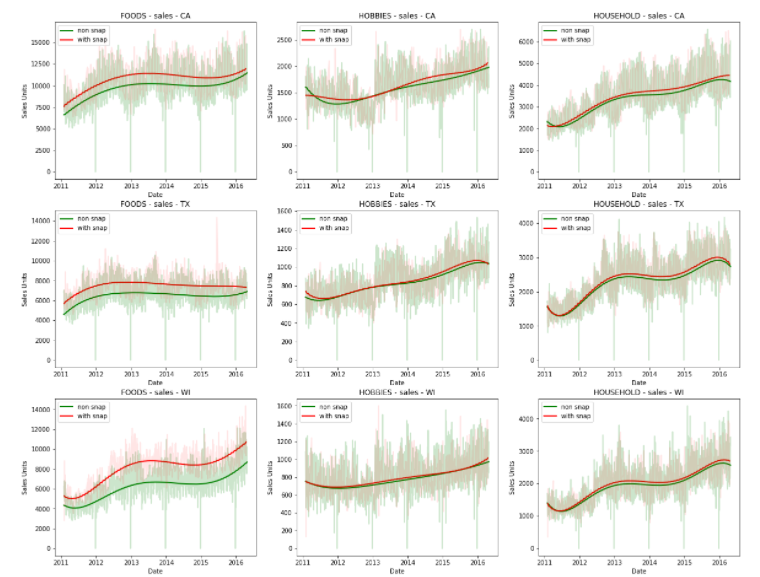
After that we ran our final analysis according to the states to check how much sales happens on the weekdays as well as on weekends in different categories. The following code shows the training of the model.



The above code calculates the sales on weekdays and weekends in different states and in different department; the following graph shows the result of above code.



Thus this is we calculated the sales volatility according to different categories and we can see that food industries are winning every sales thus as a final analysis we did the final analysis food sales according to different states, which showed us the very perspective to view the sales of food products statewise.



The above results can show us how differently the food sales go in just three states in a few last years.

Thus this is how we calculated the sales prediction and volatility of the sales as per the different departments and the interests of users on the basis of the data of three states.

**Problem Areas:**

The main problem we faced while doing the forecasting and calculating is the excessive runtime the bigger dataset means the longer runtime unless you have got a powerful processor, if you don’t have powerful processor, don’t use the system CPU better is to go for online platforms like colab and kaggle notebook. Another which occurred was the appropriate dataset we tried many datasets but some of them are either too big or too small, and others were too dirty in context with data as there were so many redundancy and null values. The final problem was during the calculations of the larger results using the scikit-learn library, many libraries are either removed or updated from the sklearn repository so we have to go for the alternative for that.

**Future Works:**

After the successful implementation of the basic model now our future task is to create a model which can take larger dataset of many categories not only Food or households. Our future target is to create a model which can predict the comparative results by taking the data of almost 5 or 6 industries together and can predict the better output with better accuracy.

**References:**

* Aishwarya, K., 2020. *FOOD DEMAND PREDICTION USING MACHINE LEARNING*. [online] Available at: <https://www.irjet.net/archives/V7/i6/IRJET-V7I6686.pdf> [Accessed 1 June 2021].
* Barbosa, N., 2015. *Demand forecasting for production planning in a food company*. [online] Available at: <https://www.researchgate.net/publication/285219852\_Demand\_forecasting\_for\_production\_planning\_in\_a\_food\_company> [Accessed 1 June 2021].
* Ethical Advocate, 2019. *7 Biggest Ethical Issues Facing The Agricultural Industry | Ethical Advocate*. [online] Available at: <https://www.ethicaladvocate.com/7-biggest-ethical-issues-facing-agricultural-industry/> [Accessed 1 June 2021].
* Silva, J.C., Figueiredo, M.C. and Braga, A.C., 2019, July. Demand forecasting: A case study in the food industry. In *International Conference on Computational Science and Its Applications* (pp. 50-63). Springer, Cham.