

MACHINE LEARNING

FINAL PROJECT MEMO

## STORE SALES

## Team Name

## Snehithulu

## Members

## Mothukuri Ruthvik Sai

## Lalam Trilochan Naidu

## Gera Praneeth

**Abstract:**

Forecasting is important in many fields, including meteorology, economic growth, and demographic patterns. Accurate product demand forecasting is critical in the retail sector, particularly for brick-and-mortar food shops, for effective inventory management. The problem is finding the correct balance, since overstocking wastes perishable items while understocking results in lost income and unsatisfied consumers. Machine learning appears to be a potential method for improving forecasting accuracy in retail.

Finally, incorporating machine learning into forecasting practices has the potential to revolutionise the retail industry by allowing businesses to navigate the complexities of inventory management with precision and agility, ultimately improving customer experience and financial performance.

**Problem Statement:**

The forecasting methods used in retail, particularly for brick-and-mortar grocery stores, are currently subjective, lack sufficient data support, and are not easily automated. This poses challenges for retailers, as inaccurate forecasts can lead to issues such as overstocking perishable goods or running out of popular items, resulting in lost revenue and customer dissatisfaction. The complexity of the problem is heightened by factors such as the addition of new store locations, diverse product offerings, changing seasonal preferences, and unpredictable product marketing strategies.

No Worries!!

Join us in this mystical pursuit, as we enchant data into predictive spells, harnessing machine learning to transform sales forecasting into an art of wizardry! my problem statement

**Positive Impact:**

Predicting sales for a store can have several positive impacts on various aspects of its operations and overall performance. Here are some potential impacts:

* Inventory Optimization
* Cost Reduction
* Customer Satisfaction
* Marketing and Promotions
* Staff Management, Data-Driven Decision-Making and Many More…

It empowers the store to make informed decisions across various aspects of its operations, ultimately contributing to its success in the market.

**Approach:**

Step 1: Loading Data

Step 2: Exploratory Data Analysis

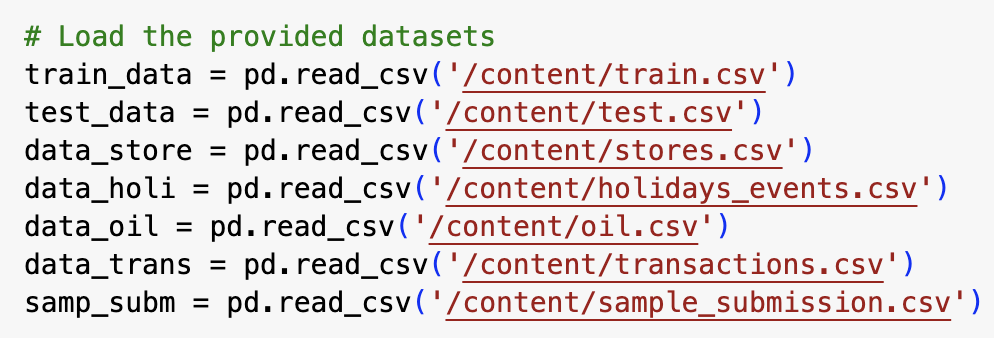
Step 3: Model Building & Deployment

Step 4: Conclusion

**Extracting Data:**

Favorita stores are located in Ecuador.

* The training data includes dates, store and product information, whether that item was being promoted, as well as the sales numbers.

****

**Exploratory Data Analysis:**

Short description of **Key Fields**

* date: Date of entry
* store\_nbr: Store Number
* sales: the turnover for any given day (target variable).
* family: Product Family
* store\_type: Type of Store
* cluster: Store Cluster
* city: City
* state: State the store is located in
* onpromotion: indicates whether a store is running a promo on that day.
* holiday\_type & locale : If it was a holiday and type of holiday
* transactions: Number of transactions
* sales: sales volume

**A screenshot of a computer code

Description automatically generated**

* I did Basic Exploratory Data Analysis to uncover historical sales patterns, including trends and seasonality. By visualizing sales data over time, it becomes possible to identify recurring patterns associated with specific seasons, holidays, or other external factors that influence sales.

**Further Analysis:**

* Equador's economy is dependent on the crude oil price. Let's examine first the relationship between crude oil and grocery sales and transactions.

**A graph showing a line graph

Description automatically generated with medium confidence**

­­­­­­­­­­­A graph showing a wave of blue and orange lines

Description automatically generated

A graph of a sound wave

Description automatically generated with medium confidence

The graphical representation of the number of sales over time compared to the date indicates that there is no discernible correlation between sales, oil prices, and transactions. This observation suggests that there is no specific relationship between the economic status of the country and daily grocery consumption.

* Moreover, Let's check the items that are most sold and the promotion to see which items influence the most for the total sales

A pie chart with numbers and a white background

Description automatically generatedA pie chart with numbers and a list of items

Description automatically generated with medium confidence

The **Top 5** most sold are Grocery, beverages, cleaning, dairy, and produce. Grocery + beverages account for more than 50% of total sales.

After conducting the analysis, I aim to delve deeper into understanding the impact of sales during a specific month.

A graph of blue and white bars

Description automatically generated

Sales analysis in different time frames

* On a daily basis, Saturday and Sunday show the highest sales.
* On a monthly basis, December sales are particularly strong.
* On a yearly basis, It is growing at a steady pace.

A graph of blue rectangular bars

Description automatically generated with medium confidence

* The average holiday sales are equivalent to Saturday and Sunday sales.

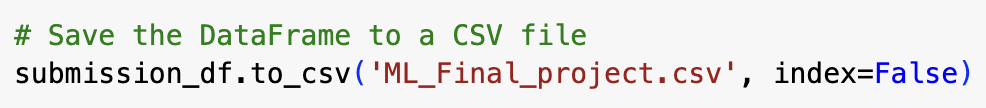
**Model Building & Deployment:**

Ultimately, in the process of model building, I considered three algorithmic approaches: Linear Regression, Gradient Boosting Regression, and Random Forest. Following validation and the computation of the root mean square least error (rmsle), it was determined that Gradient Boosting Regression emerged as the most suitable choice.

A screenshot of a computer program

Description automatically generated

* As stated earlier, the root mean square least error (rmsle) for Gradient Boosting is 2.06, demonstrating its superiority over other models. Additionally, it is recognized that Gradient Boosting can capture intricate relationships and non-linear patterns in the data, making it adept at fitting complex, non-linear functions.. This approach can be found in the source code go through it for reference.



**Conclusion:**

In summary, the analysis found no significant correlation between sales, oil prices, and transactions, suggesting no specific link between the country's economic status and daily grocery consumption. The top 5 items, including Grocery, beverages, cleaning, dairy, and produce, contribute over 50% of total sales.

Further examination revealed high daily sales on Saturdays and Sundays, strong monthly sales in December, and steady annual sales growth. Interestingly, holiday sales match weekend averages.

The model-building phase favored Gradient Boosting with a superior root mean square least error (rmsle) of 2.06, showcasing its ability to capture intricate relationships and non-linear patterns.

In conclusion, the sales prediction project for Favorita stores in Ecuador is complete, with results stored in the "ML\_project" CSV file. The detailed findings can be explored in the source code for reference.

**References:**

[1] [KASHISH RASTOGI. 📝Store Sales Analysis⏳ Time Serie](https://www.kaggle.com/kashishrastogi/store-sales-forecasting)

[2] [HOWOO JANG. First kaggle notebook. Following TS tutorial](https://www.kaggle.com/howoojang/first-kaggle-notebook-following-ts-tutorial)

[3] [EKREM BAYAR. Store Sales TS Forecasting - A Comprehensive Guide](https://www.kaggle.com/ekrembayar/store-sales-ts-forecasting-a-comprehensive-guide)

[4] [Ryan Holbrook. Time Series](https://www.kaggle.com/learn/time-series)