

## Project Description

---

### Business Context

Rossmann operates over 3,000 drug stores in 7 European countries. Currently, Rossmann store managers are tasked with predicting their daily sales for up to six weeks in advance. Store sales are influenced by many factors, including promotions, competition, school and state holidays, seasonality, and locality. With thousands of individual managers predicting sales based on their unique circumstances, the accuracy of results can be quite varied. You are provided with historical sales data for 1,115 Rossmann stores. The task is to forecast the "Sales" column for the test set. Note that some stores in the dataset were temporarily closed for refurbishment.

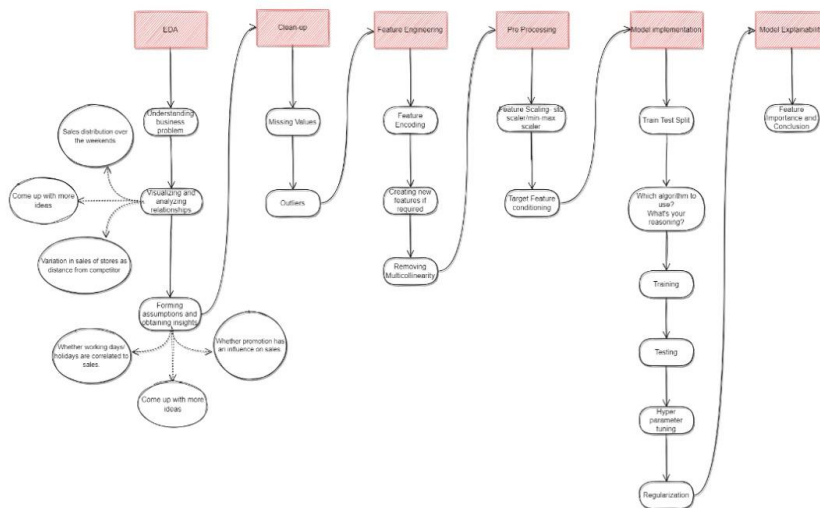
### Data Description:

Fields	Description
Id	Unique entry id
Store	store_id
Sales	Sales made for the day
Customers	Footfall for the day
Open	Open or closed
StateHoliday	State Holiday or not
SchoolHoliday	School Holiday or not
StoreType	Type of stores
Assortment	Type of assortment
CompetitionDistance	Distance from the nearest competition
Promo	Store running promotion or not
Promo2	Store running consecutive promotion or not

## Main Libraries to be Used:

- Pandas for data manipulation, aggregation
- Matplotlib and Seaborn for visualisation and behaviour with respect to the target variable
- NumPy for computationally efficient operations
- Scikit Learn for model training, model optimization, and metrics calculation

## Project Architecture:



## Project Evaluation Criteria

---

- Efficient EDA
- Encoding if necessary.
- Feature selection, new feature creation
- Dealing with multicollinearity if any
- Feature scaling
- Understanding the target feature and its distribution
- Modeling - use at least two algorithms
- Evaluation and improvement of model.
- Feature Importance and Conclusion
- Understanding of how your project is useful to stakeholders?

## Rubrics

Rubrics	Weightage	Max Credits
GitHub Commits	5	100
Summary and Technical Documentation in Collab Notebook	10	100
EDA and Visualization	5	100
Looking for and Handling NaN/ Null/ Missing Values and Outliers	2.5	100
Finding Correlation in Variables (Both Dependent and Independent, Visualizations on Data)	10	100
Pick Appropriate Independent Variables, Test Train Split, Train Model	10	100
Prediction and Calculate Some Evaluation Metrics for Model	10	100
Number of Models Experimented (Atleast 2)	5	100
Hyperparameter Tuning	5	100
Final Summary of Conclusion	2.5	100
Commented Code	5	100
Proper Output Formatting	2.5	100
Modularity of Code	2.5	100
Video Presentation	20	100
Fluency and Grammatical Accuracy in Video	5	100

## [Data store](#)

## [Rossman data store](#)