



Rossmann Pharmaceutical Store Sales Forecasting

By Harsh Vardhan



Business Problem Statement

Rossmann is a large pharmaceutical retail chain in Europe. They want to **predict sales** of their stores 6 weeks into the future.


Goal is to use **past data + store info + events** to predict future sales.

Accurate prediction helps in **inventory planning, staff management, and targeted promotions.**



Project Objectives

- Understand customer behaviour (Task 1: EDA)
- Build a machine learning model to predict Sales
- Create a deployable web app to predict sales for uploaded CSV
- Submit structured code, charts, report, and app link via GitHub



Dataset Overview

File	Description
train.csv	Store-wise historical sales data
test.csv	Store-wise future data without sales
store.csv	Store information (StoreType, Promo2, Distance, etc.)
sample_submission.csv	Format for final prediction file



Features in the Data

Numerical: Customers, CompetitionDistance, Sales

Categorical: StoreType, Assortment, StateHoliday, Promo

Date: Converted into Year, Month, Day, WeekOfYear,
IsWeekend



Exploratory Data Analysis (EDA)

Charts Used:

- Sales Over Time
- Promo vs Sales
- StoreType vs Sales
- StateHoliday vs Sales
- Feature Correlation Heatmap

Key Insights:

- Promo increases sales
- StoreType A performs best
- Sales dip during state holidays



Data Preprocessing & Feature Engineering

- Removed closed stores (Open=0)
 - Filled missing values in CompetitionDistance, Promo2SinceWeek, etc.
- Converted Date into multiple time-based features
- One-hot encoded categorical variables (StoreType, Assortment, etc.)



Model Training

- Used: RandomForestRegressor
- RMSE on validation: ~1100
- Trained on 80%, tested on 20%
- Used engineered features



Prediction Output

- Prediction done on test.csv
- Format matched with sample_submission.csv
- Output file: rossmann_submission.csv

Final Deliverables

- rossmann_submission.csv ✓
- Jupyter Notebook (Project-6-notebook.pynb) ✓
- Charts as PNGs ✓
- GitHub Repo ✓
- PPT Presentation ✓



Learnings & Takeaways

Hands-on understanding of EDA, ML modeling, and feature engineering

Real-world business problem solving with data

Working with real sales data and interpreting trends



THANK YOU