Store Sales Prediction

Outline

01

Introduction & Project workflow

Goal and workflow of project

02

Data Description

Describe the data

03

Data preprocessing

Analysis and visualization

04

Machine learning Model

Models

Introduction

The goal of this project, I will use machine learning to Store Sales Prediction to provide the stores with essential insights into the upcoming inventory and cash flow for the following months.

Project workflow

1. Import libraries
   - Read Data
2. Data Preprocessing
   - Data visualization
3. Data splitting
   - Build Models
4. Evaluating Model Performance

Data Description

This dataset can be found at Kaggle. This dataset contains over 18000 rows with 8 features and after engineering features 16. The dataset that they provided contains the information of sales with:

- ID: Unique identifier for a row
- Store id: Unique id for each store
- Store type: Type of the store
- Location type: Type of the location where the store is located
- Region Code: Code of the region where the store is located
- Date: Information about the date
- Holiday: If there is a holiday on the given date
- Discount: If the store offers a discount on the given date

Data preprocessing

- Analysis and visualization

Analysis and visualization

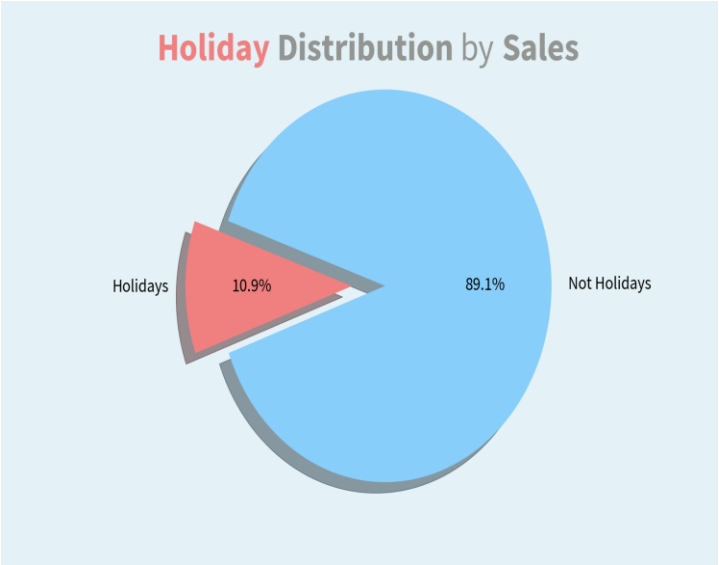
The best stores (by number of sells)

The best store types (by number of sales)

Analysis and visualization

Best months by Total sells

Holidays Distribution by Sales



Machine learning Model

Models

Random Forest

R2 Score: 0.7776877526730435
MAE:5911.999665476047
MSE:75175976.76668999
RMSE:8670.408108427768

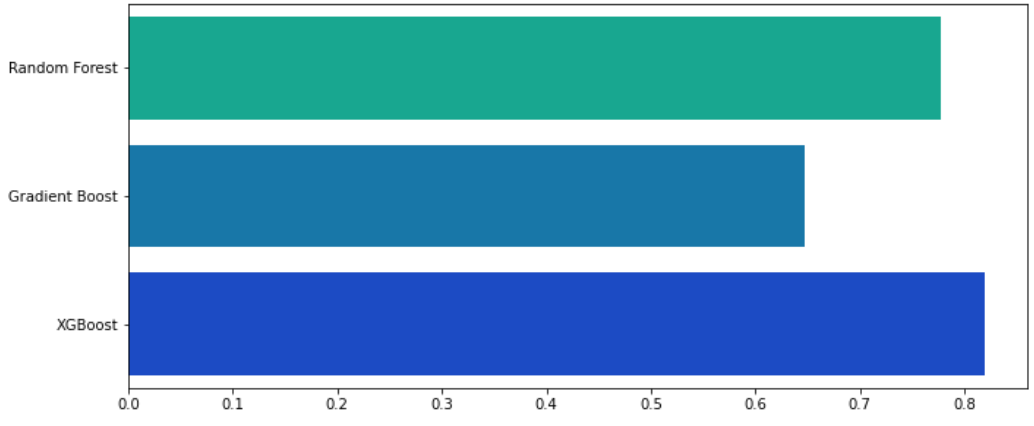
Gradient Boosting RegressorDecisionTreeClass

R2 Score: 0.6469144534012485
MAE:7619.312688639989
MSE:119397609.29465993
RMSE:10926.921309072373

XGBRegressor

R2 Score: 0.8188080066409554
MAE:5452.611182863876
MSE:61270961.21266244
RMSE:7827.576969449898

Scores



Thanks!