

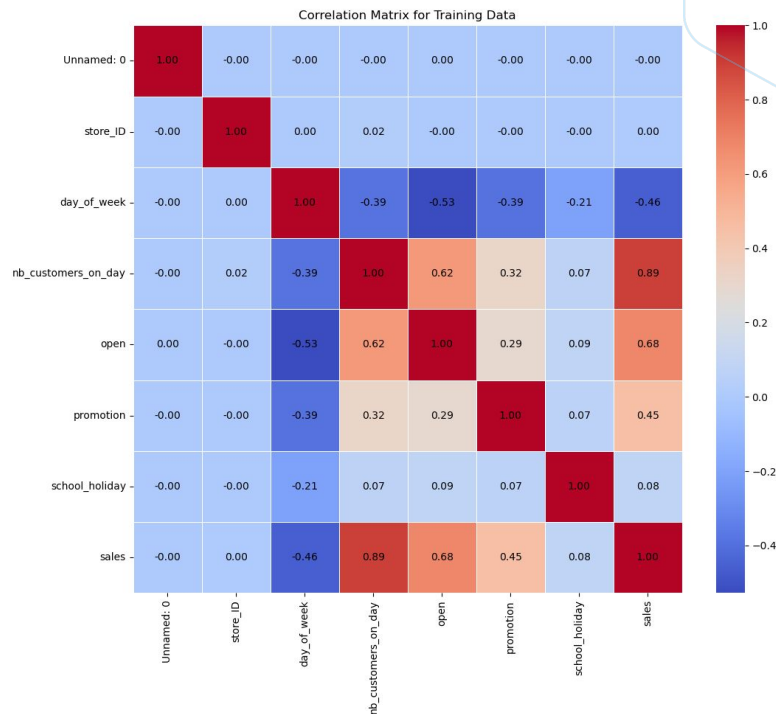
Iron Kaggle

Data Science and Machine Learning

Exploratory Data Analysis & Data Preparation

For both datasets:

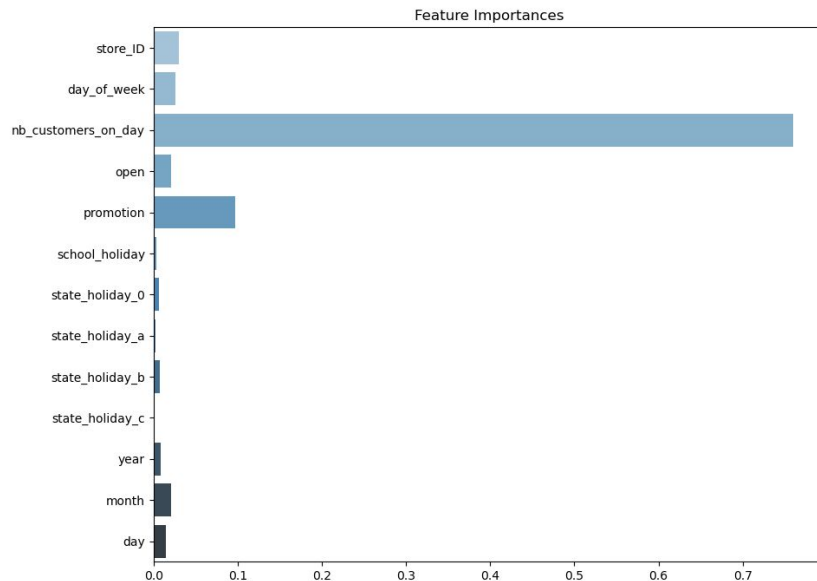
- **Converted “date” into year/month/day and dropped “date” info**
- **Converted “state_holiday” to dummy variables**
- **Dropped index and unnamed columns**
- **Correlation between variables**
 - "nb_customers_on_day" and "sales"



XGBoost Regressor code

- **High performance and accuracy predicting sales**
 - Has the ability to handle complex relationships and interactions within structured data
- **Scalability**
 - Capable of handling large datasets quickly
- **Feature importance**
 - Help understanding which factors have the most influence in sales predictions

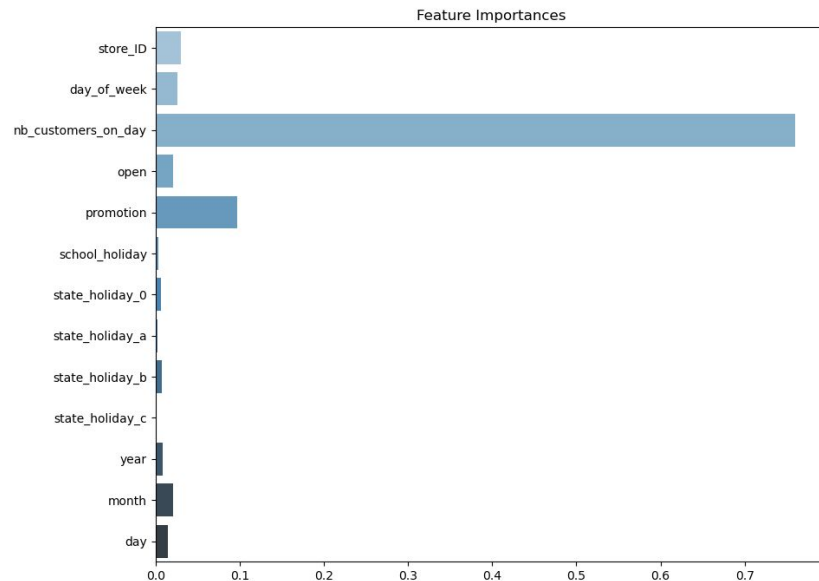
XGBoost



XGBoost Regressor code

- **High performance and accuracy predicting sales**
 - Has the ability to handle complex relationships and interactions within structured data
- **Scalability**
 - Capable of handling large datasets quickly
- **Feature importance**
 - Help understanding which factors have the most influence in sales predictions
- **Expected R^2 score: 0.933**

XGBoost



Notes for future

- **Try different models**
- **Further tuning of model hyperparameters**



Thank you!

Data Science and Machine Learning