

Capstone Project 2

Rossmann Store Sales Prediction

By

Md Shahrukh Khan
G Manikanta , Shadab Ali Khan

Agenda

1. →About the company
 2. →Objective
 3. →Problem description
 4. →Data set
 5. →Data reading and Cleaning
- EDA
 - Regression used
 - Results
 - Limitations
 - Conclusion

About the company

- ==>Germany's second-largest drug store chain, handling in over 3,000
- ==>Drug stores in 7 European countries.
- ==>Started in 1972 by Dirk Rossmann

Objective

Help Rossmann doing EDA analysis and create a regression model for sales prediction

Problem description

1. Stores sales are influenced by different factors.
2. Our present task is to predict Rossmann stores daily sales using store and other factors

Data Set

Historical sales data for 1,115 Rossmann stores

Files: "Rossmann Stores Data.csv" - historical data including "Sales
• store.csv" - supplemental information about the stores

Data fields : Id, store ,sales ,customers ,open , stateholiday ,schoolholiday
, Storetype, assortment, competition distance ,promo,promo2

Task is to predict : "sales"

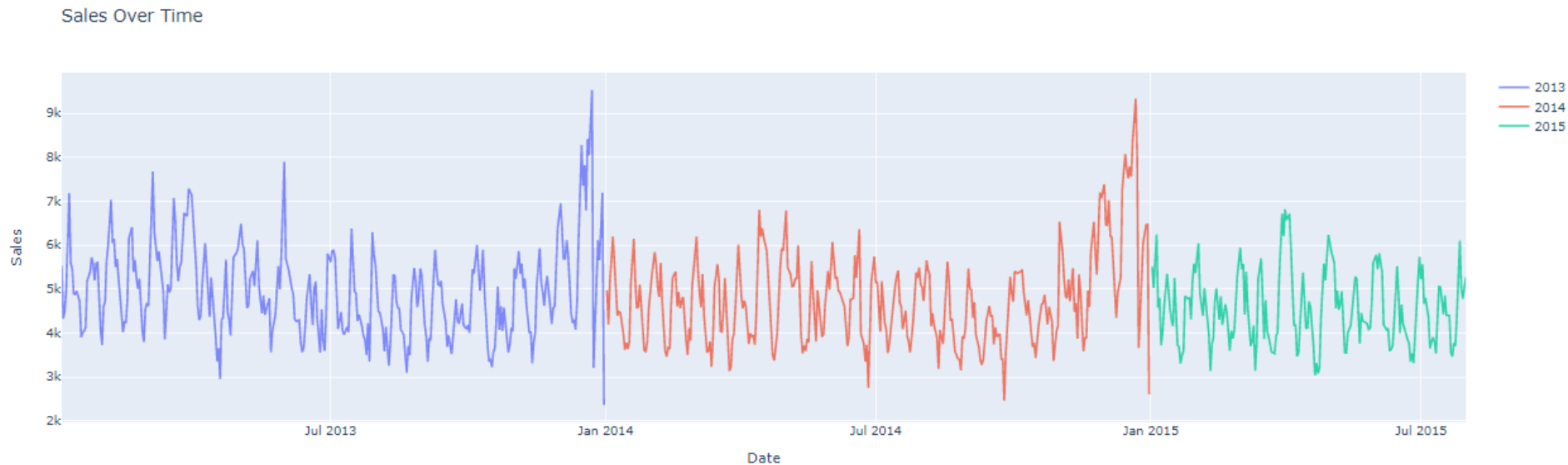
Data reading & cleaning

Rows that have zero sales are removed from training file as they can not be used in evaluation

Then two data files were merged i.e. test data is merged with store data

Exploratory data analysis

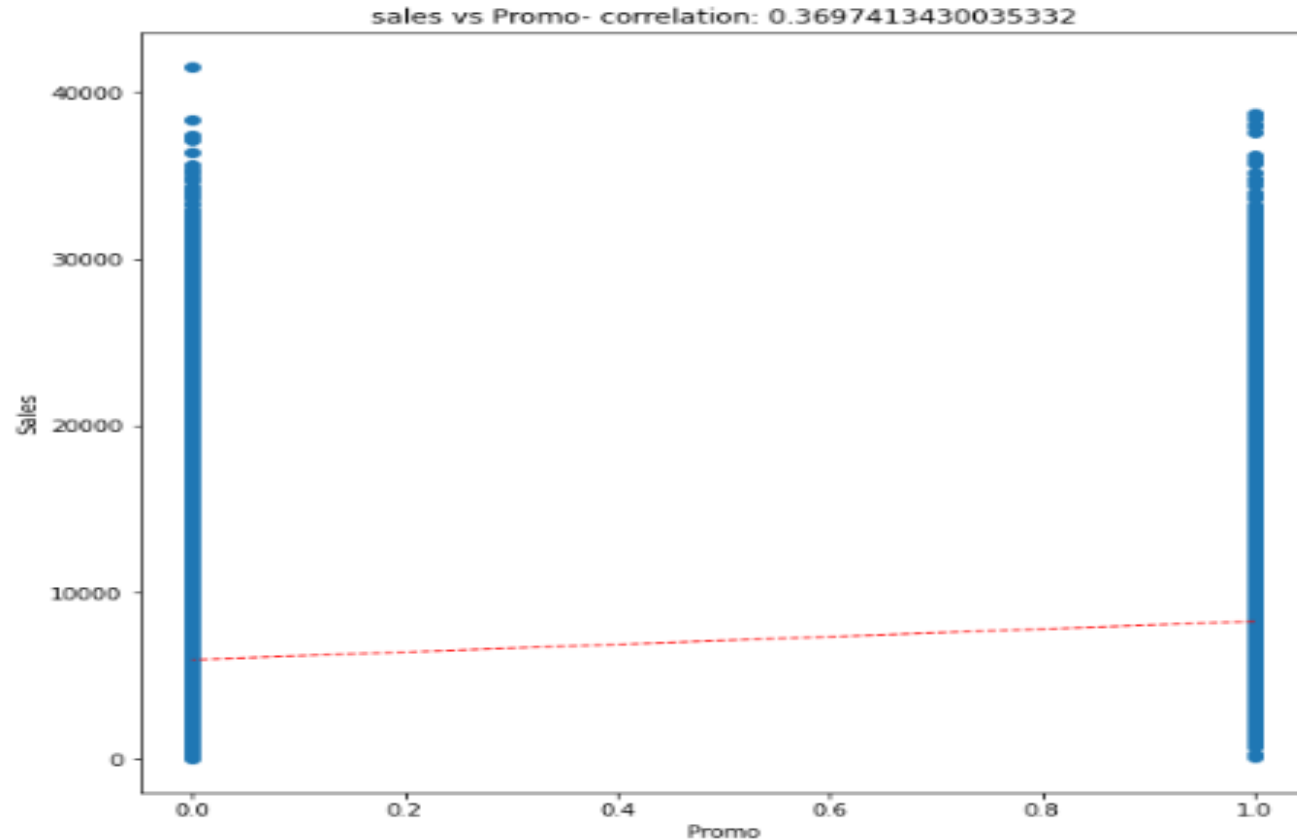
Sales Over time



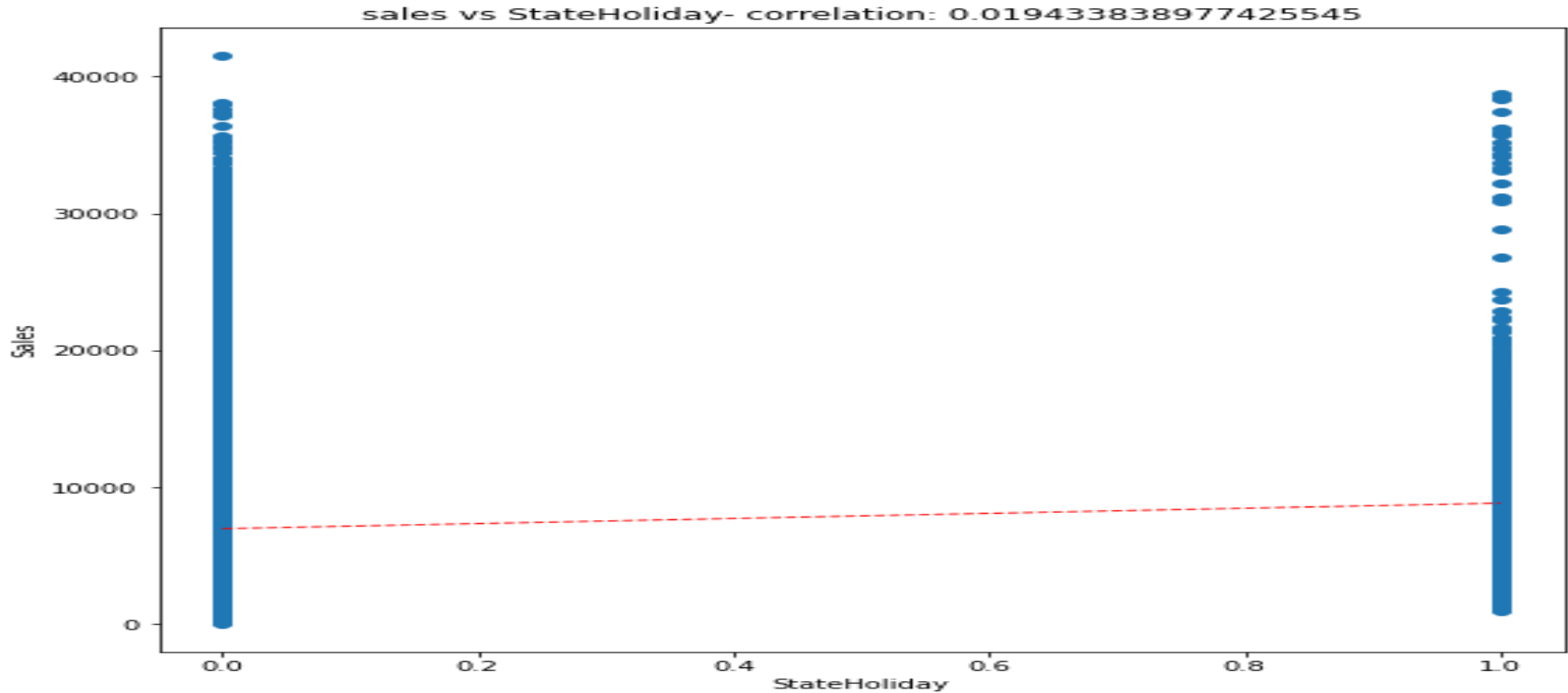
Correlation



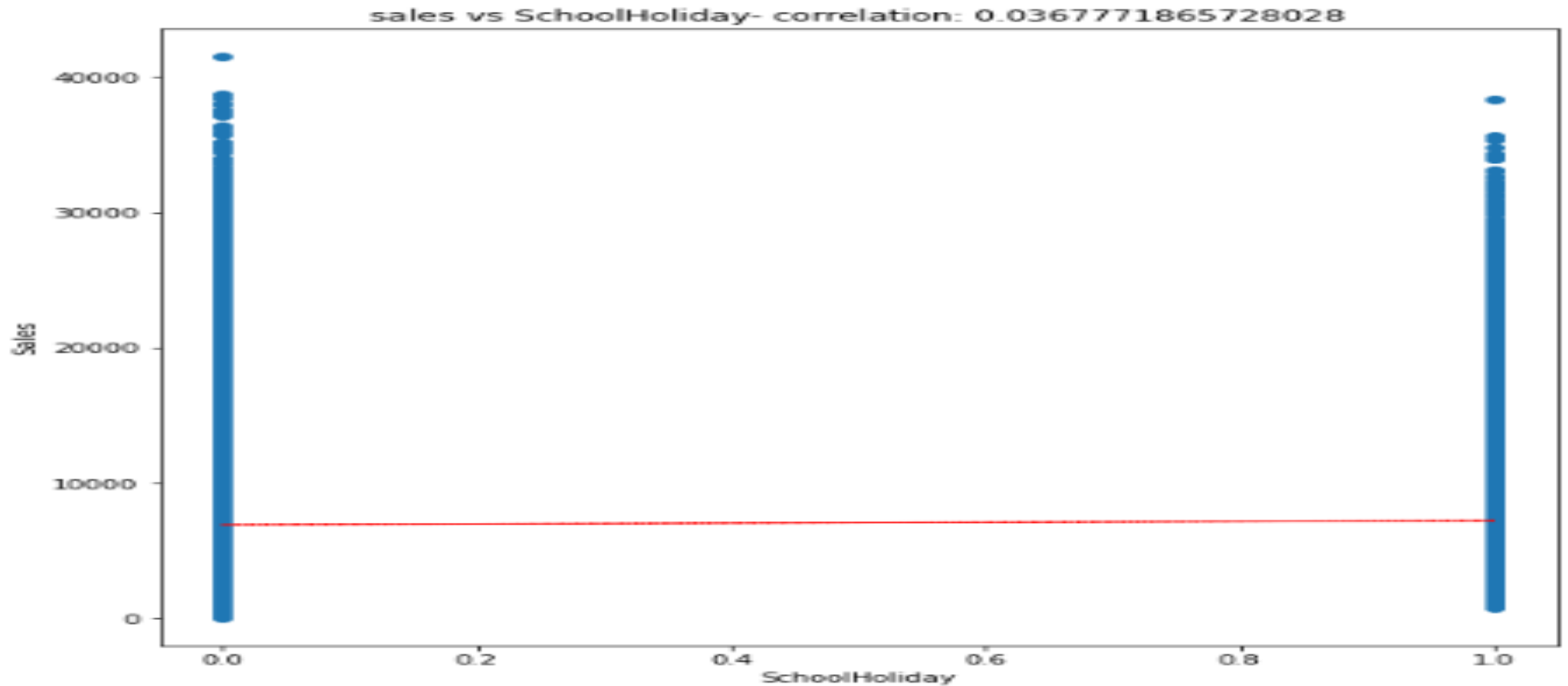
Sales vs Promo



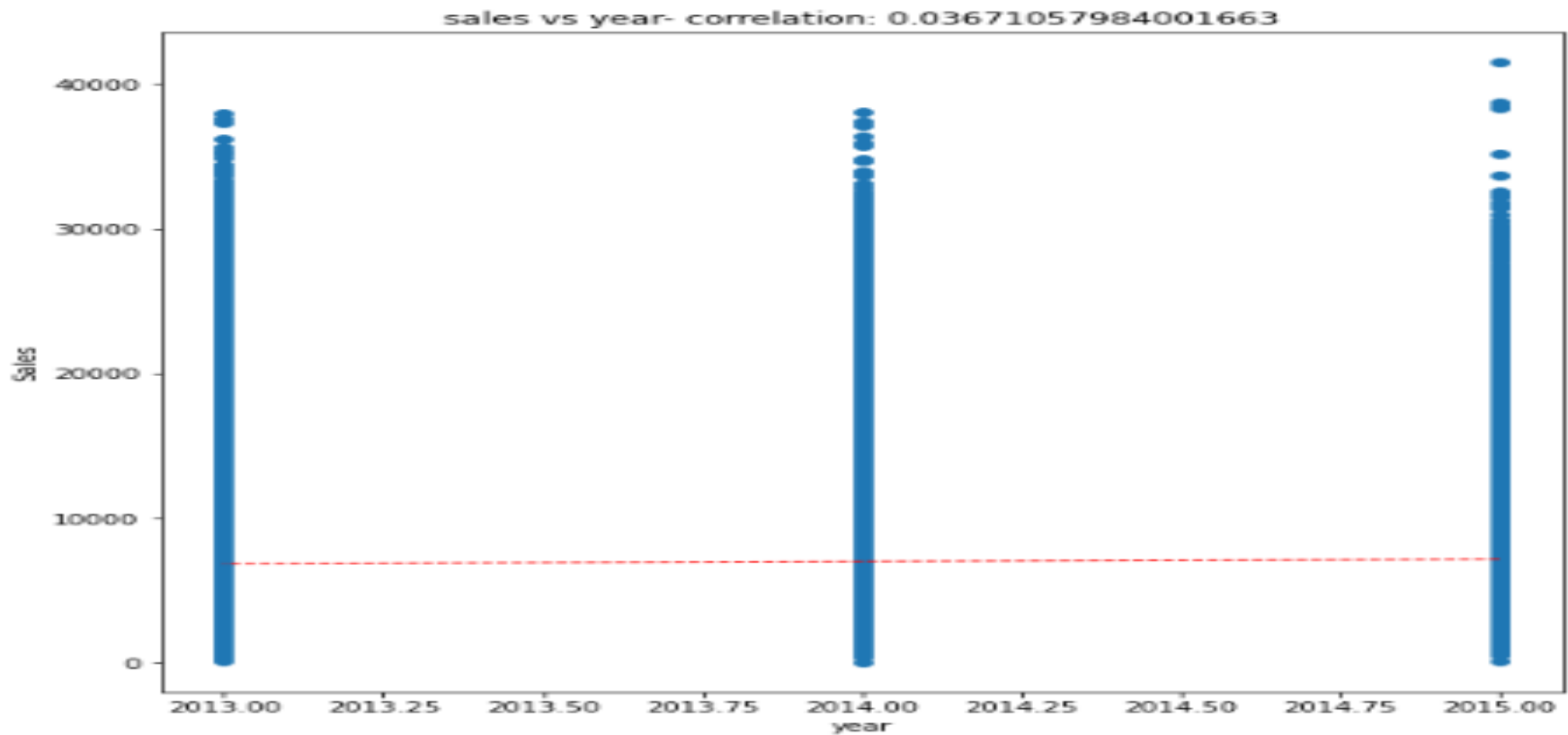
Sales vs Stateholiday



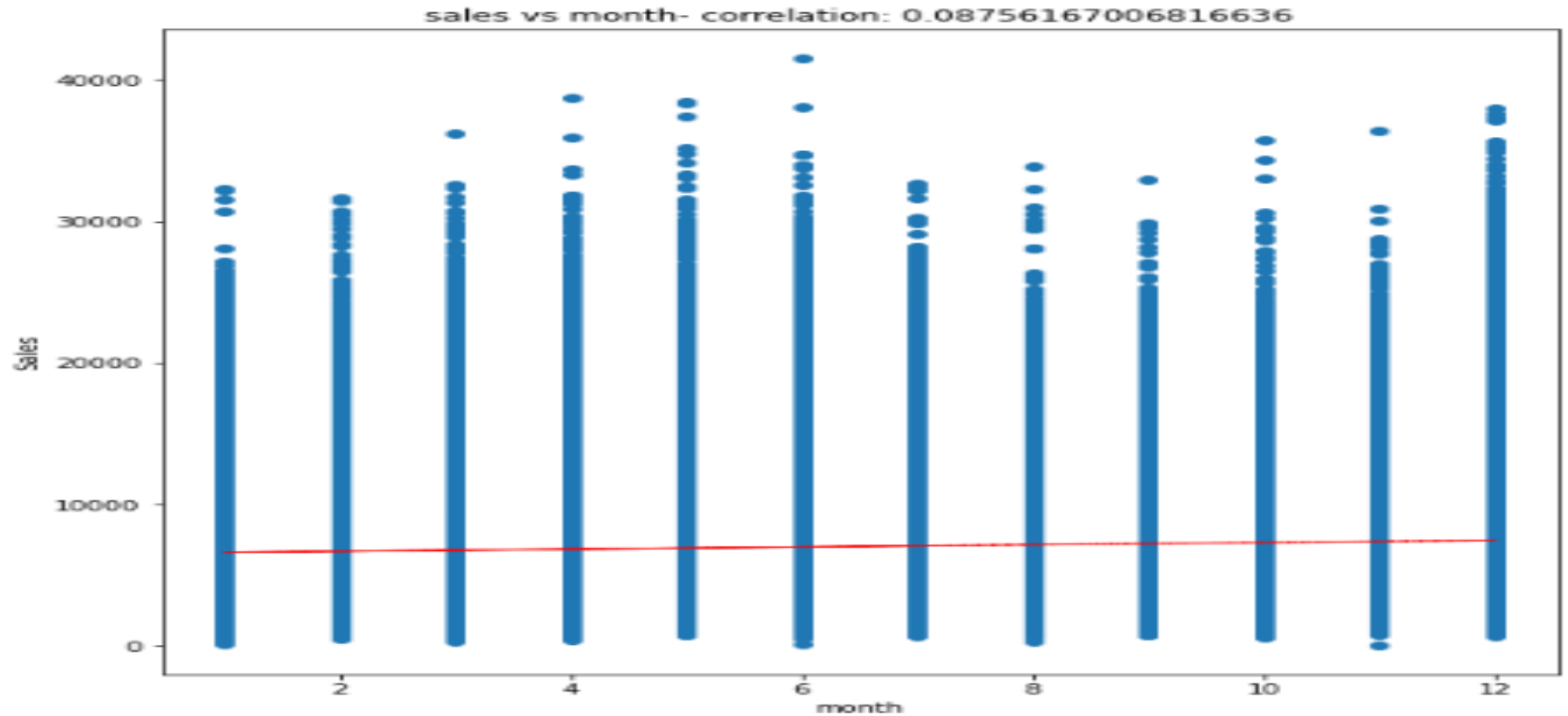
Sales vs School holiday



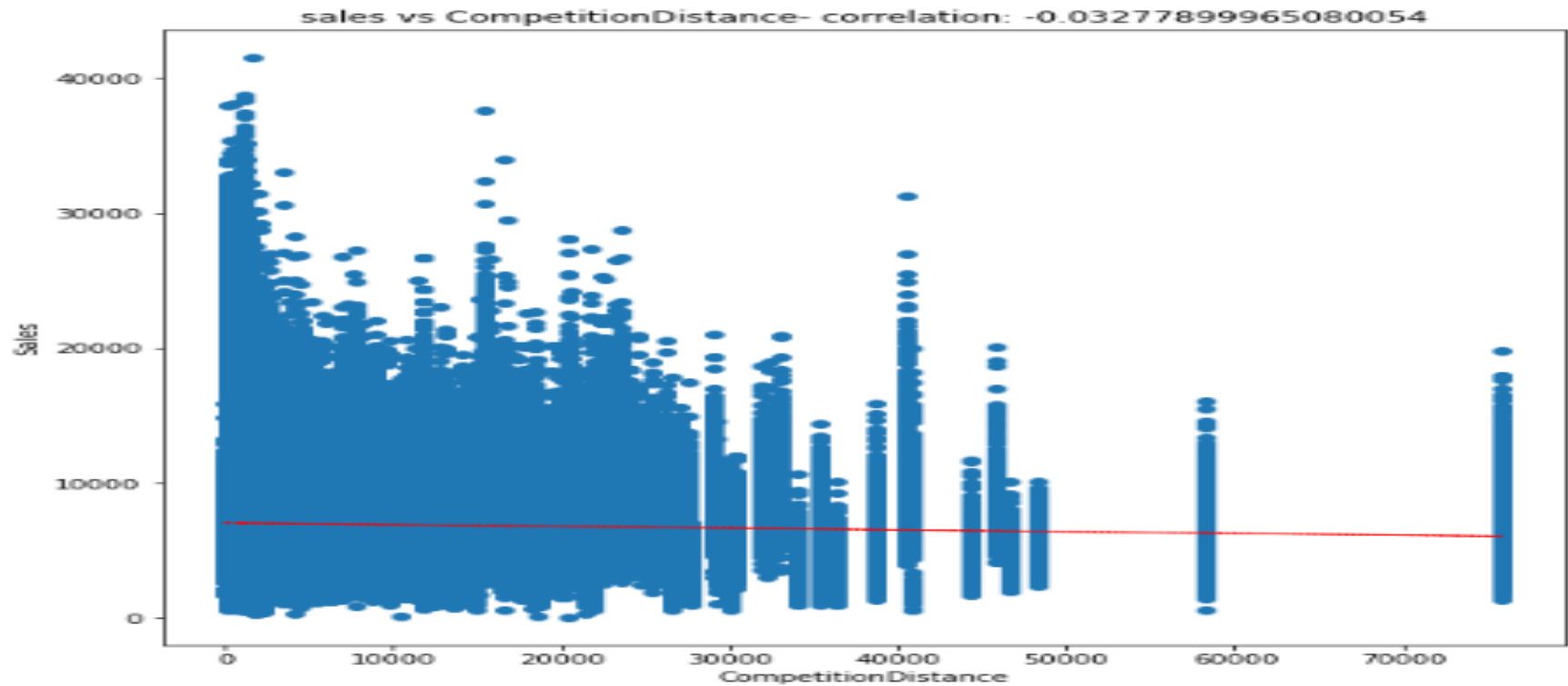
Sales vs Year



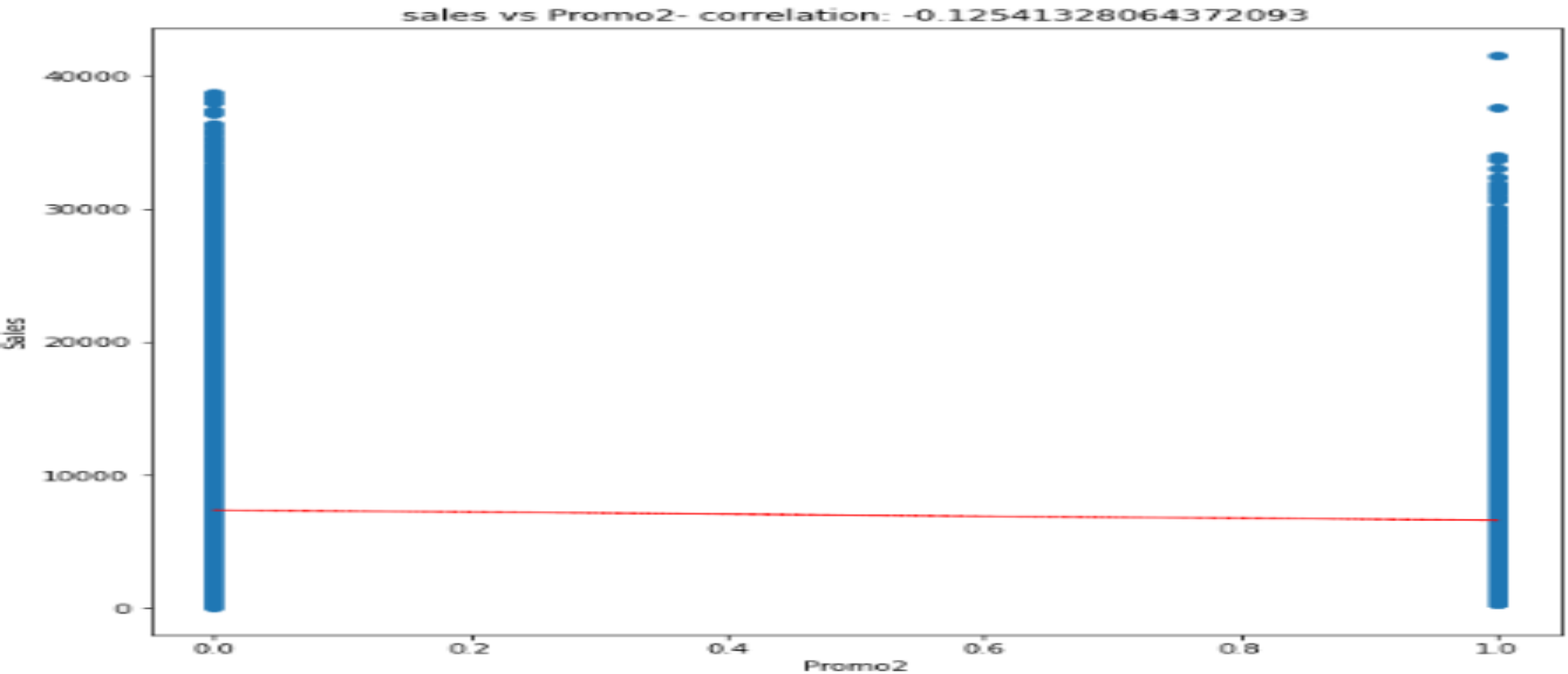
Sales vs Month



Sales vs Competition Distance



Sales vs Promo



Regression used

Linear regression :

Regression Model Score : 0.75

Out of Sample Test Score : 0.74

Training RMSE : 1562.10

Testing RMSE : 1573.52

Training MAPE : 17.02

Testing MAPE : 17.21

Regression used

Lassio

Regression Model Score : 0.75

Out of Sample Test Score : 0.74

Training RMSE : 1562.17

Testing RMSE : 1573.89

Training MAPE : 17.02

Testing MAPE : 17.21

Regression used

Decision tree:

Regression Model Score : 0.8497

Out of Sample Test Score : 0.78

Training RMSE : 1216.92

Testing RMSE : 1417.25

Training MAPE : 12.49

Testing MAPE : 14.75

Regression used

Random Forest:

Regression Model Score : 0.97 ,

Out of Sample Test Score : 0.836

Training RMSE : 476.77

Testing RMSE : 1239.04

Training MAPE : 4.83

Testing MAPE : 12.66

Limitations

- Lack of competitor data
- Lack of information on product type
- Lack of information on store type

Conclusion

Promotions playing a vital role in sales

Variations in sales for stores open on Sunday



thank
you