Demand Analysis and Forecasting - Rothman Retail

Can Yang, Haode Sun, Kangtong Guo,
Xinyu Jia, Xujia Zhang

Demand Analysis and Forecasting-Rossmann

1 Background

2 Data Description

3 Approach and Methods

4 Data Analysis

5 Forecast Analysis

6 Future Work &Limitation

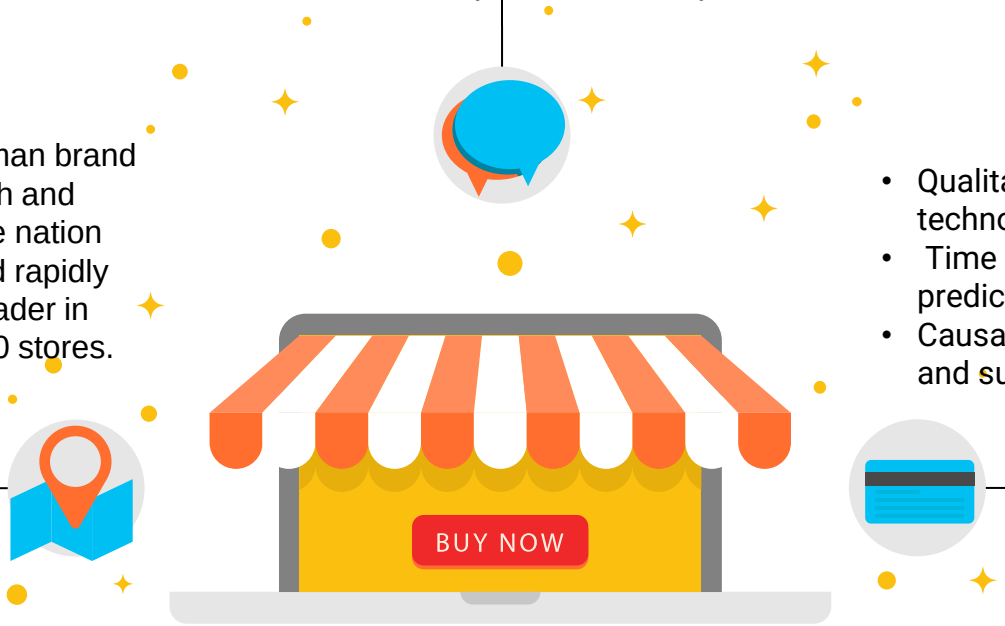
Background

Opportunity

- Leverage customer and store data to support decision making
- Discuss how to better improve business performance

Rossmann

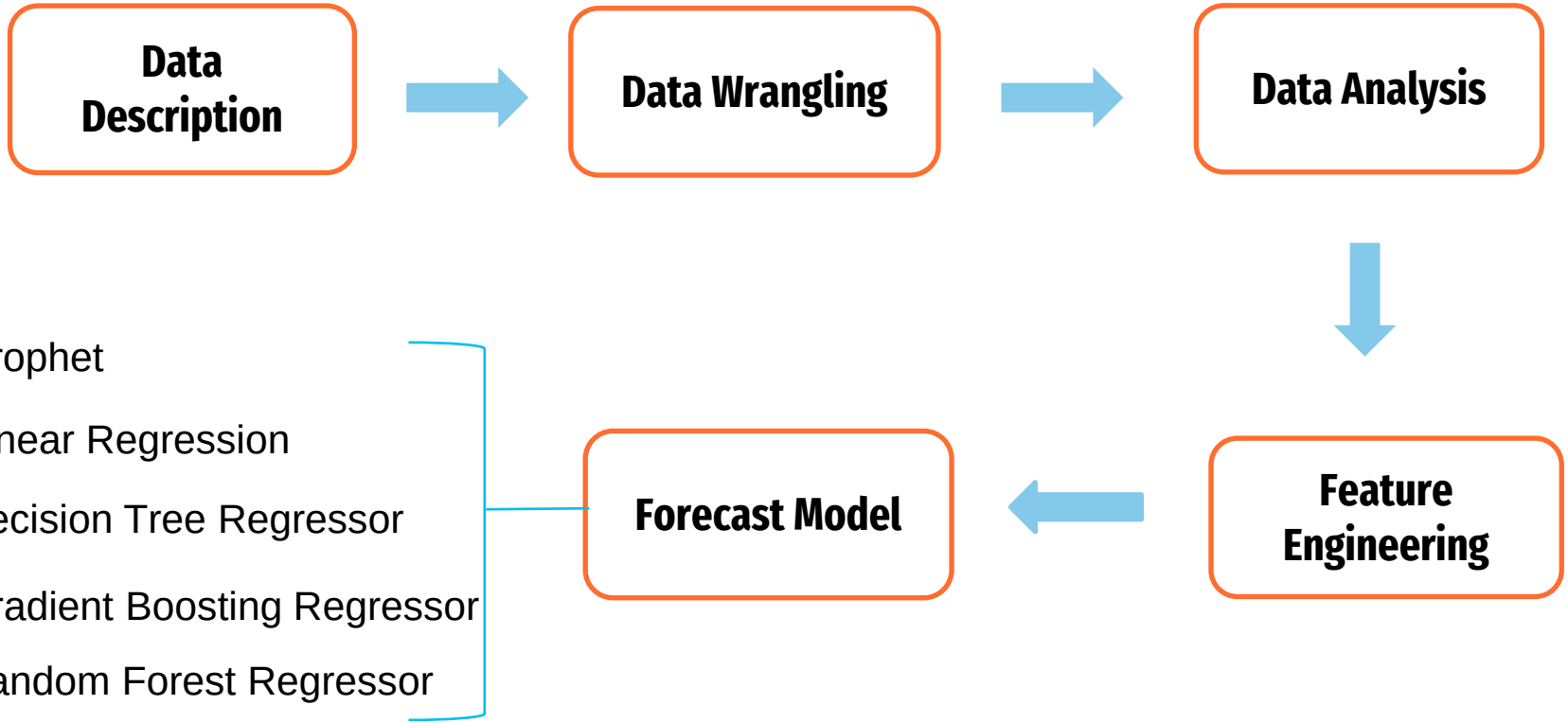
- Rossmann is the first German brand to introduce discount health and beauty retail concept to the nation
- Rossmann had progressed rapidly and became the market leader in northern Germany with 100 stores.



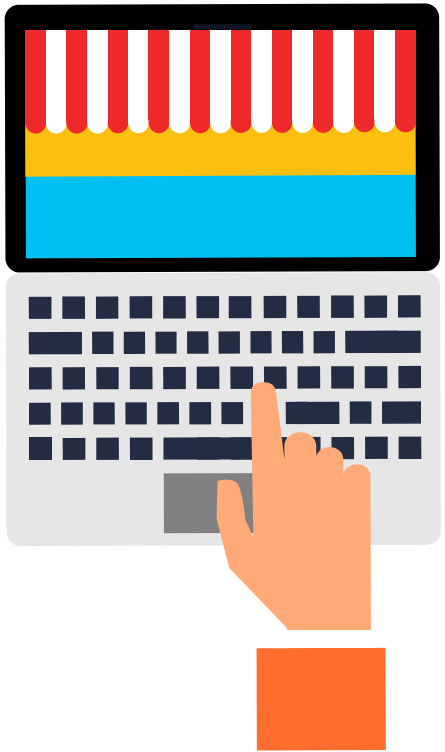
Common Types of Forecasting

- Qualitative techniques: used in new technology areas
- Time series analysis: suitable for predict with clear and stable trends.
- Causal models: with historical data and sufficient analysis

Approach and Methods



Description of Data



- Train
- Test
- Store
- state_names
- store_states

Train

	Store	DayOfWeek	Date	Sales	Customers	Open	Promo	StateHoliday	SchoolHoliday
0	1	5	2015-07-31	5263	555	1	1	0	1
1	2	5	2015-07-31	6064	625	1	1	0	1
2	3	5	2015-07-31	8314	821	1	1	0	1
3	4	5	2015-07-31	13995	1498	1	1	0	1
4	5	5	2015-07-31	4822	559	1	1	0	1

Figure 1

Test

	Id	Store	DayOfWeek	Date	Open	Promo	StateHoliday	SchoolHoliday
0	1	1	4	2015-09-17	1.0	1	0	0
1	2	3	4	2015-09-17	1.0	1	0	0
2	3	7	4	2015-09-17	1.0	1	0	0
3	4	8	4	2015-09-17	1.0	1	0	0
4	5	9	4	2015-09-17	1.0	1	0	0

Figure 2

Store_states

	Store	State
0	1	HE
1	2	TH
2	3	NW
3	4	BE
4	5	SN

Figure 3

Store_names

	StateName	State
0	BadenWuerttemberg	BW
1	Bayern	BY
2	Berlin	BE
3	Brandenburg	BB
4	Bremen	HB

Figure 4

Data Analysis-data processing

1.Find and remove outliers

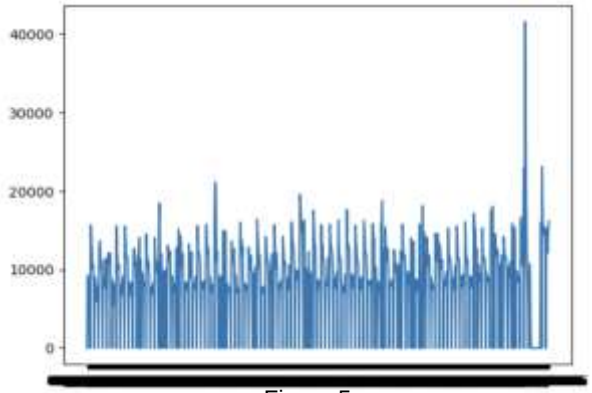


Figure 5

2.Fill in missing values and do logarithmic changes

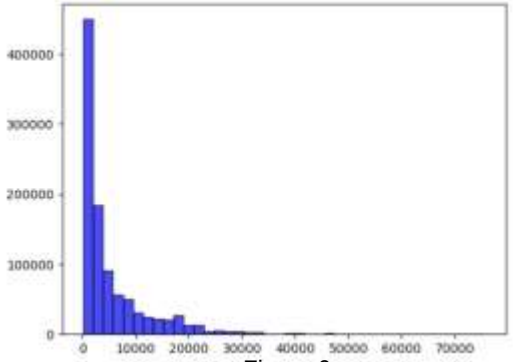


Figure 6

after transformation

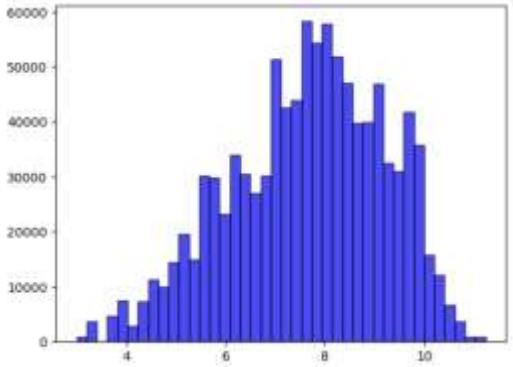


Figure 7

3.Delete last month two columns

4.Add variable Promo2SinceWeeks

5.Delete Promo2SinceWeek' ,'Promo2SinceYear

6.Discard samples with stores closed and zero turnover

Data Analysis-Data Overview

Element heatmap

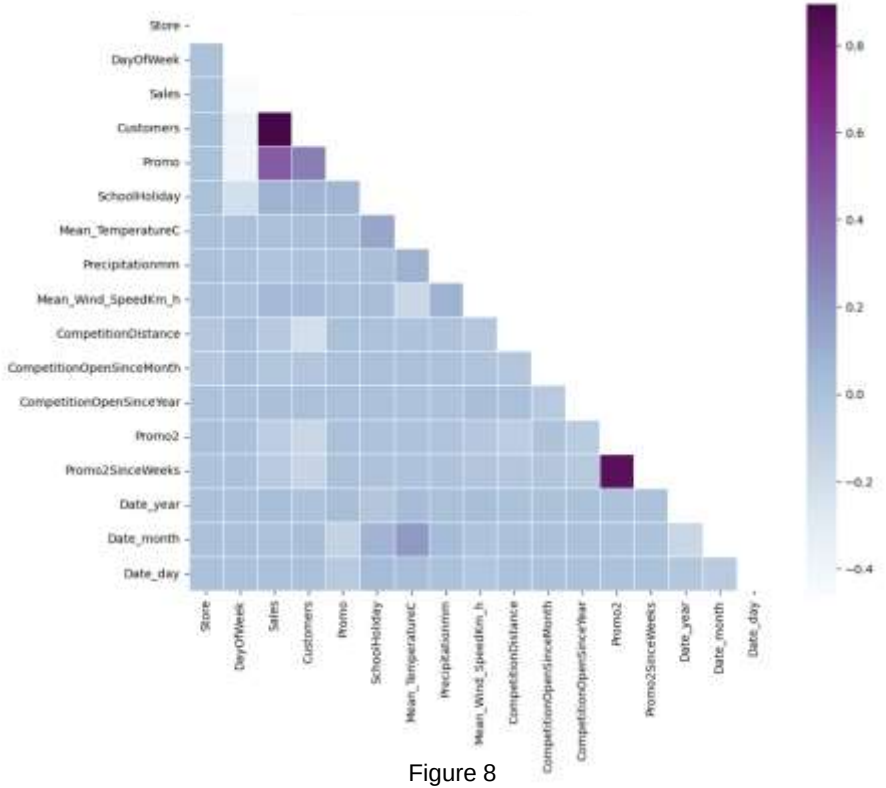


Figure 8

Changes in sales and number of customers per year

Sales

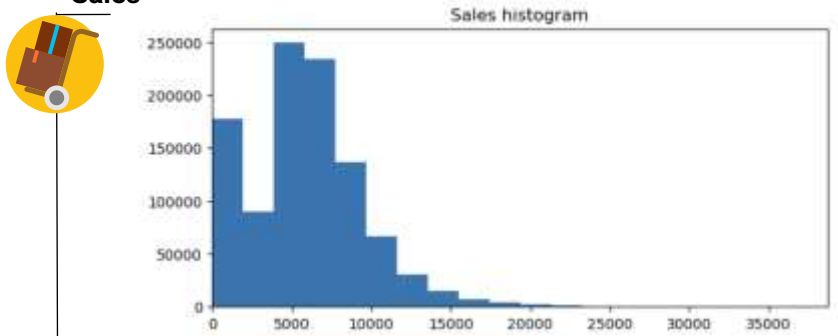


Figure 9

Customers

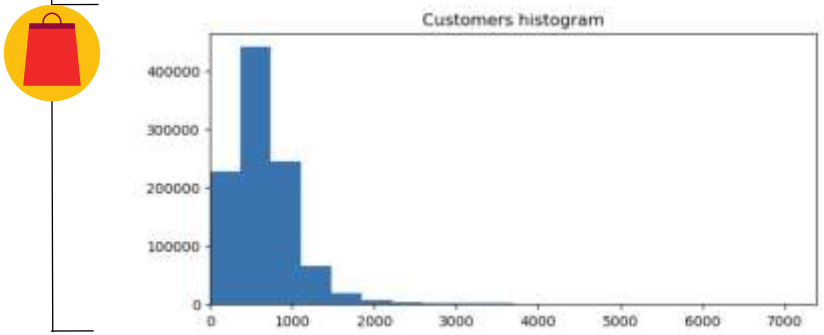


Figure 10

Data Analysis - Data Visualization

Effects of Promotions on Different Types of Stores

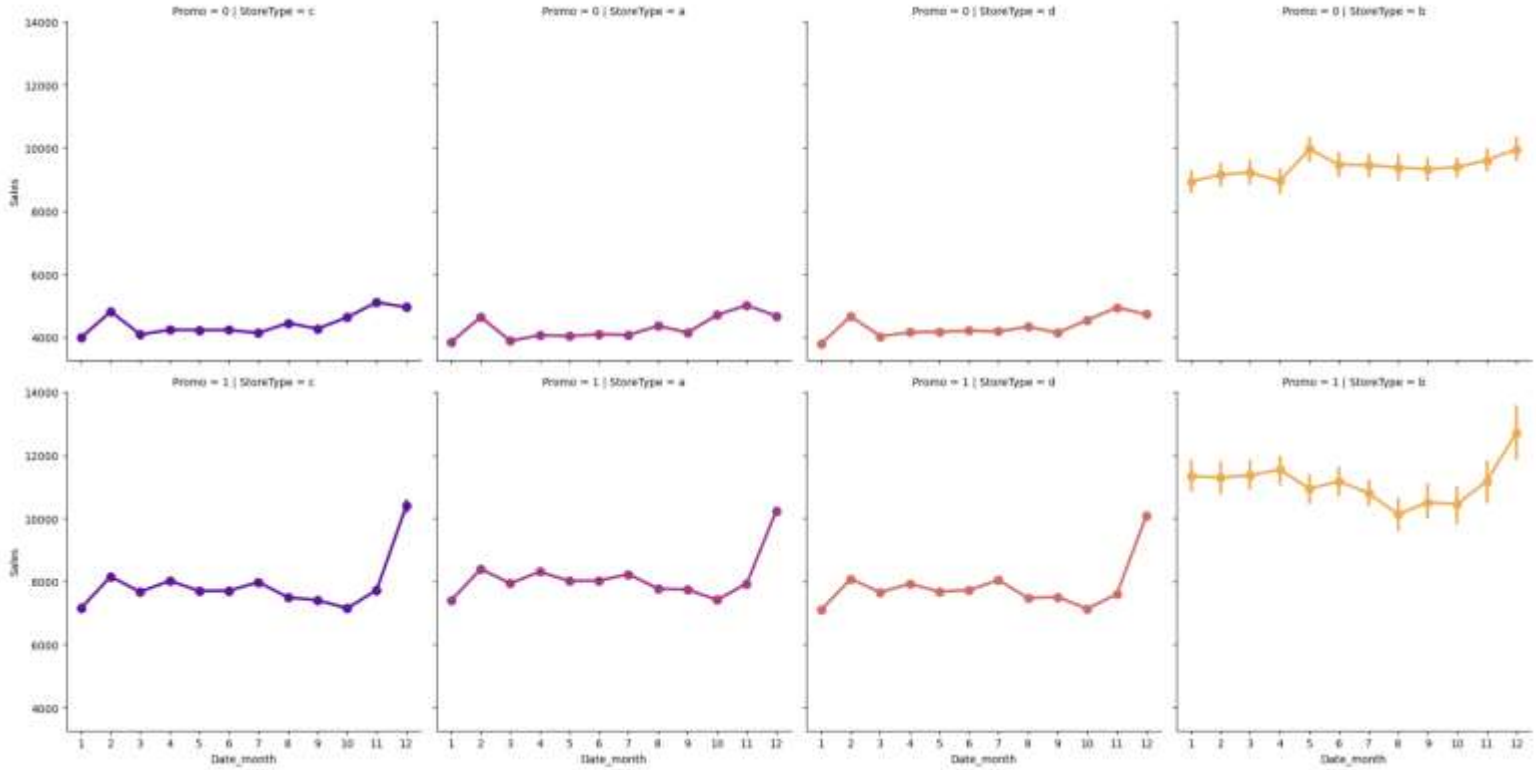


Figure 11

Data Analysis - Data Visualization

Sales volume of different types of stores on working days and non-working days

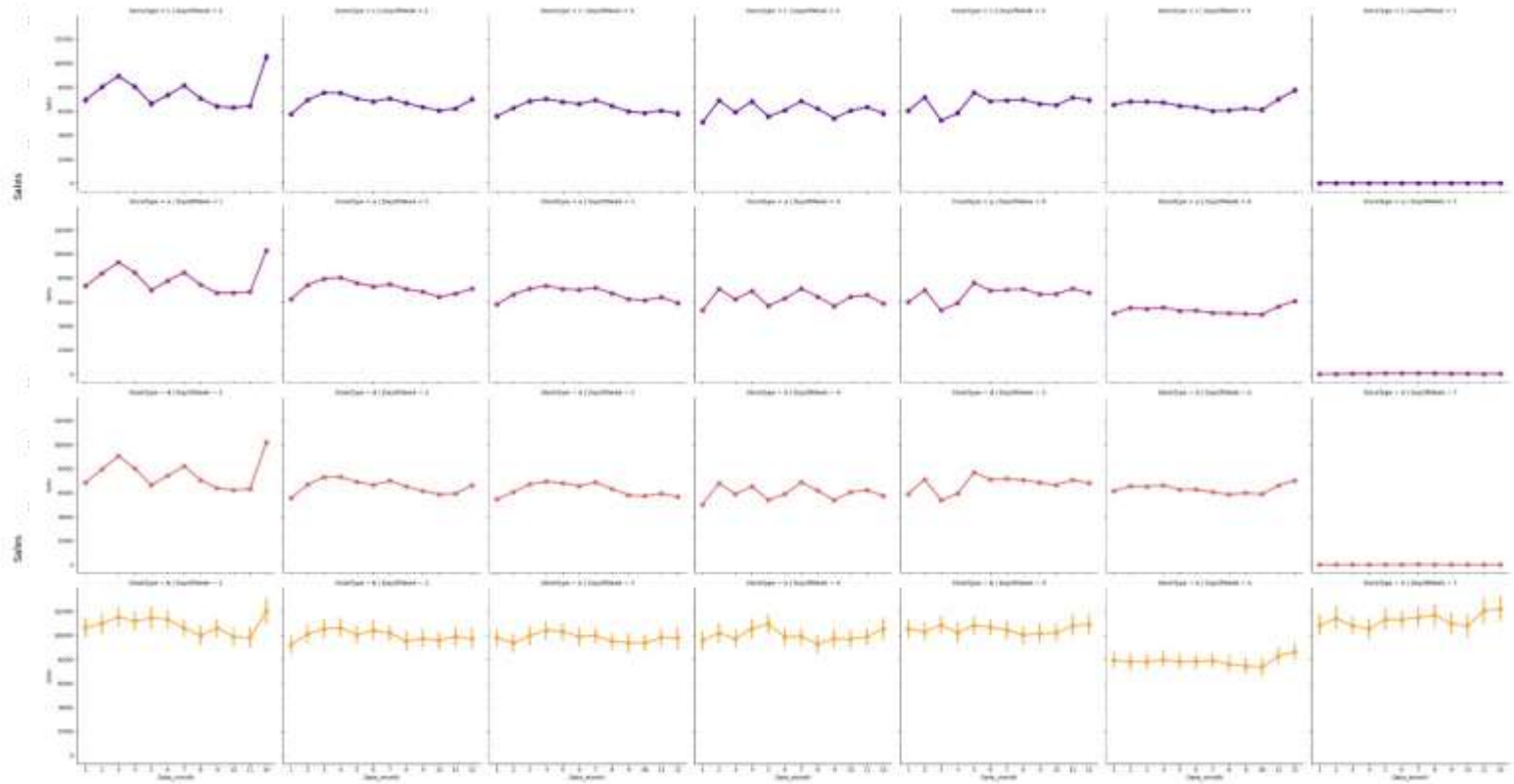


Figure 12

Data Analysis - Data Visualization

Impact of school holidays on sales of different types of stores

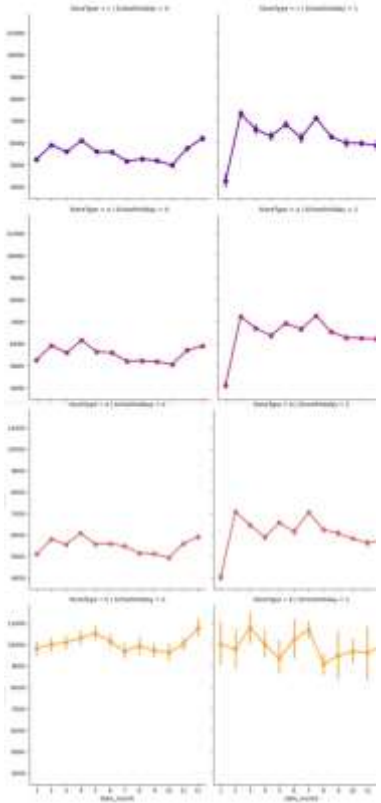


Figure 13

Data Analysis - Data Visualization

Impact of stateholidays on sales of different types of stores

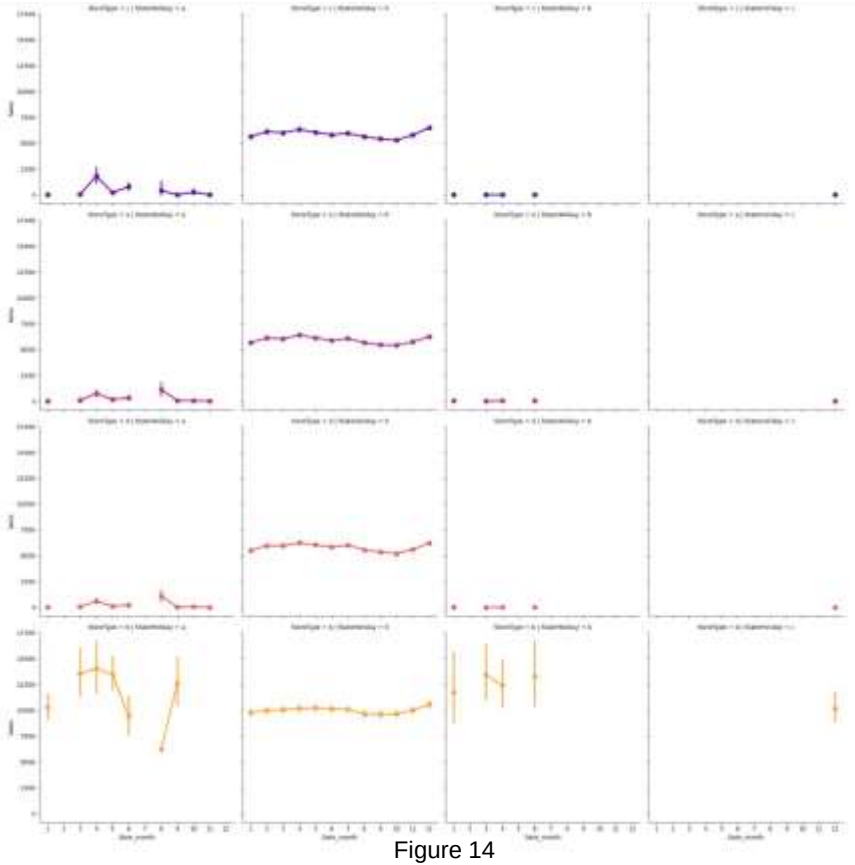


Figure 14

Data Analysis-spatial analysis of sales volume

Spatial Distribution of Sales in Different Types of Stores

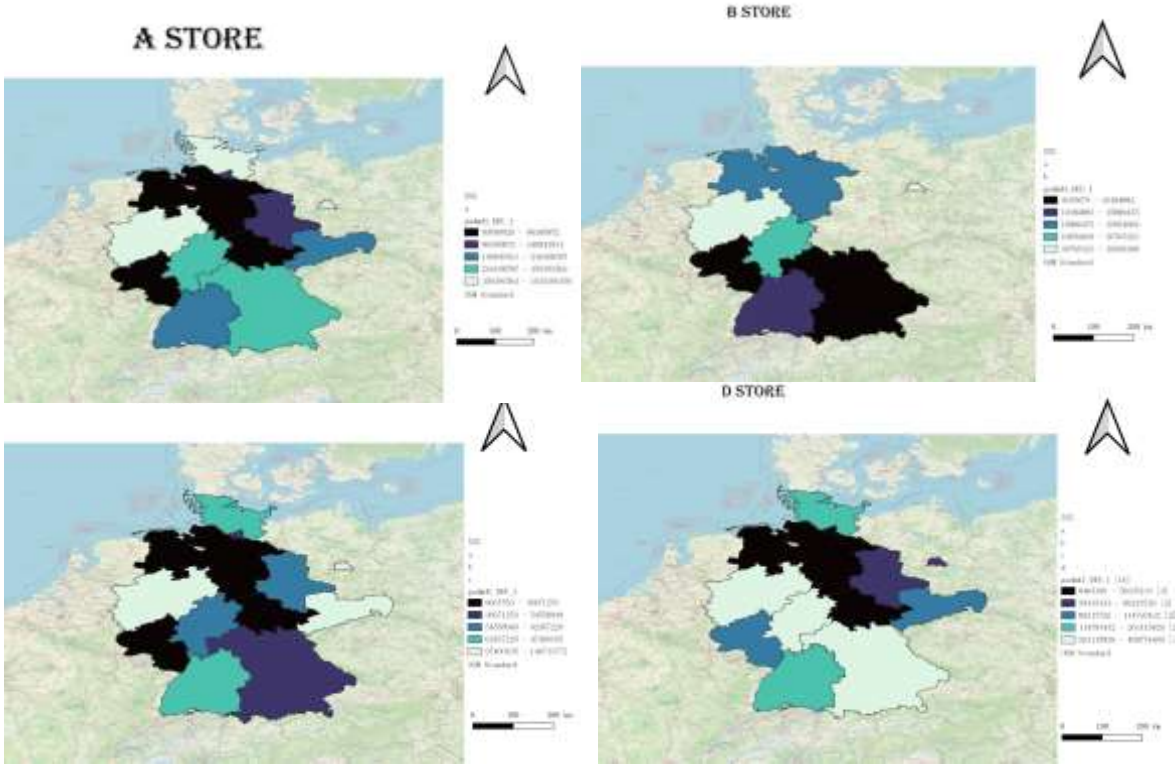


Figure 15

Uneven distribution

Stores a, c, and d are not distributed in 4 districts, and store b is not distributed in 8 districts

sales rankings

Different types of stores have different sales rankings in different regions

Data Analysis- why we choose store ‘a'?

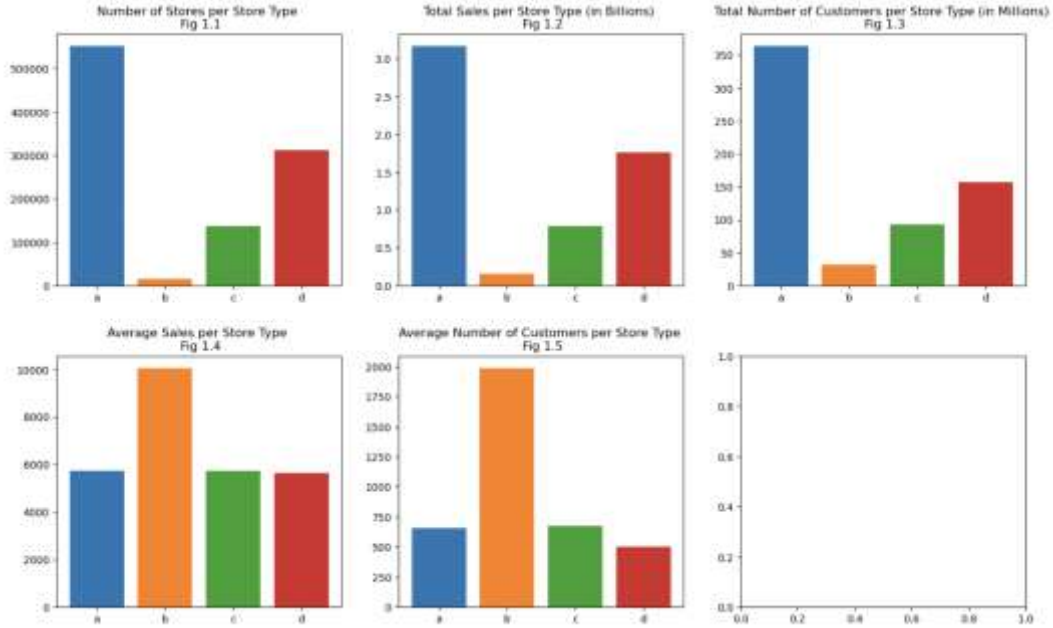
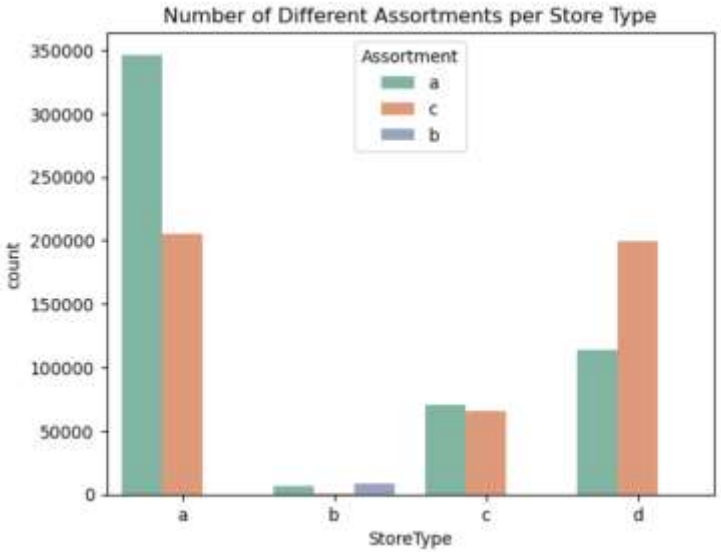


Figure 16



Assortment: a = basic, b = extra, c = extended

Figure 17

- The total number of stores, total sales and customers in store a are far higher than those in other stores
- The average sales and average number of customers of store a are similar to those of other stores
- The number of stores a is the largest, and the basic type stores are far more than other stores

Forecast Analysis—Prophet

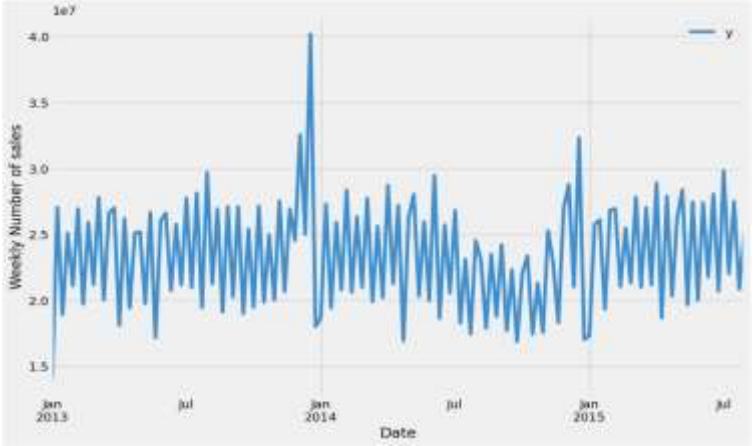


Figure 18

- The plot above represents the visualization data of store A's weekly sales from January 2013 to July 2015.
- Almost every year around December sales reach a peak.

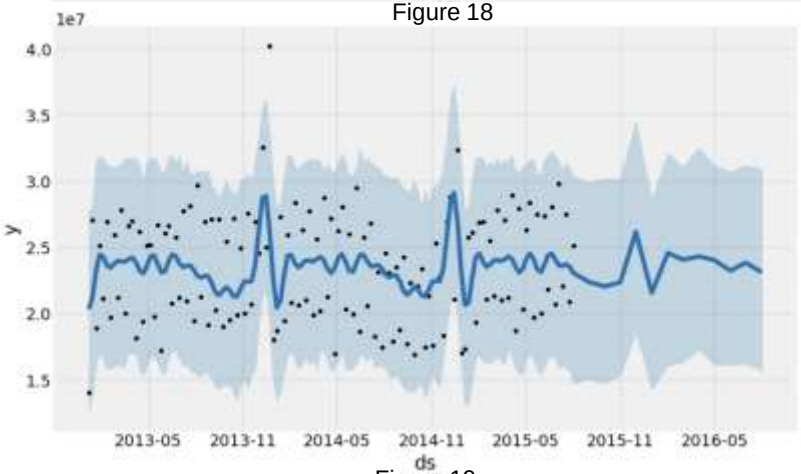


Figure 19

- The plot below represents the prediction of sales with the prophet model.
- A slow downward trend from August 2015 - November 2015, followed by a rise, peaking around December, before falling again and eventually levelling off.

Forecast Analysis—Plotting the forecasted components

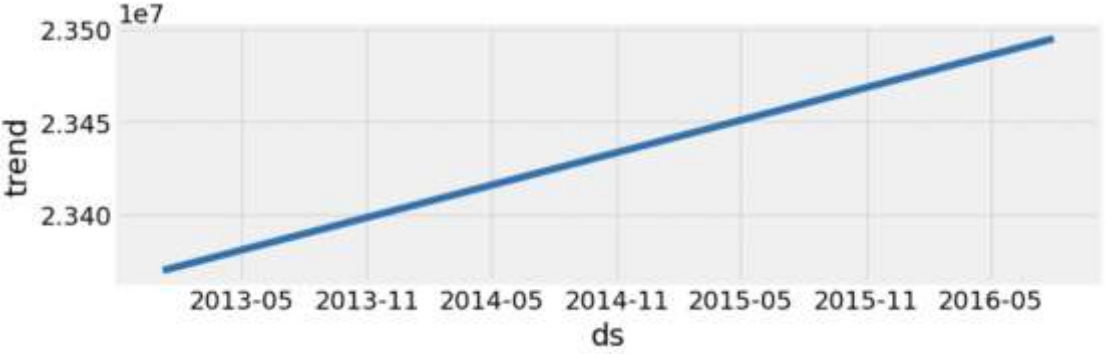


Figure 20

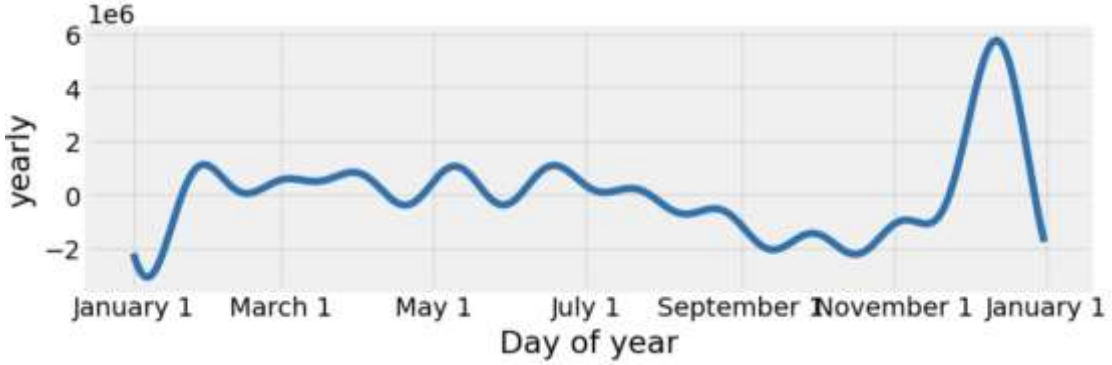


Figure 21

- The first plot shows that the weekly volume of sales has been linearly increasing over time.
- The second plot shows that the largest sales occurs during the months of February and December.

Forecast Modeling—Machine Learning Method (cross_val_score)

model	0	1	2	3	4	mean	std
Lr_model	0.569101	0.569961	0.570956	0.568883	0.569902	0.569760	0.000734
Tree_model	0.611801	0.608037	0.615249	0.610808	0.609766	0.611132	0.002406
Rfr_model	0.689484	0.687019	0.692625	0.687995	0.687995	0.688970	0.001996
Gbdt_model	0.646806	0.649195	0.646593	0.647646	0.647646	0.647739	0.000982

Sheet 1

Forecast Analysis-Conclusion

	MSE	R^2
Lr_model	6.727429e+06	0.570969
Tree_model	5.998547e+06	0.617453
Rfr_model	**4.869550e+06**	**0.689452**
Gbdt_model	**5.499765e+06**	**0.649262**

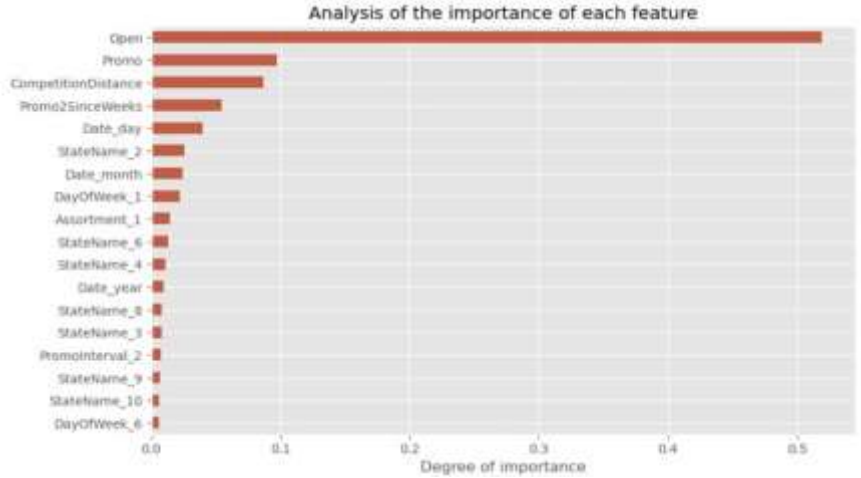
Sheet 2

Forecast Modeling

The top ten columns with positive correlations

```
Sales              1.000000
Open               0.658507
Promo              0.461796
StateHoliday_0     0.251995
DayOfWeek_1        0.224626
DayOfWeek_2        0.152107
Assortment_1       0.111187
StateName_2        0.096848
DayOfWeek_3        0.095561
DayOfWeek_5        0.093995
Name: Sales, dtype: float64
```

after prediction



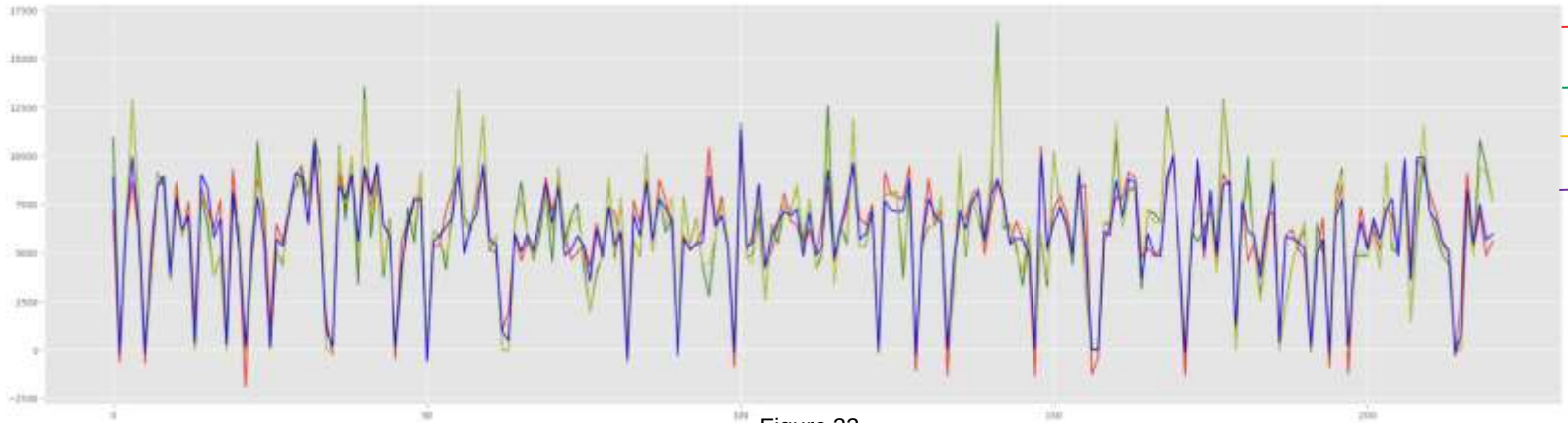


Figure 22

Future work

- Select more predictive models to test in order to get the optimal model.
- Optimising the model by adjusting the parameters

Limitation

- Does not take into account the characteristics of consumers in different regions and cannot make targeted recommendations
- There is no universality in analysing just one typr shop and making predictions.

References

- A.S. Watson Group - World's largest international health & beauty retailer, leader in O+O retail. (2015). *Rossmann*. [online] Available at: https://www.aswatson.com/our-brands/health-beauty/rossmann/.

- Brownlee, J. (2017). *Introduction to time series forecasting with python: how to prepare data and develop models to predict the future*. Machine Learning Mastery.

- Catal, C. , Ece, K. , Arslan, B. & Akbulut, A. (2019). Benchmarking of Regression Algorithms and Time Series Analysis Techniques for Sales Forecasting . Balkan Journal of Electrical and Computer Engineering , 7 (1) , 20-26 . DOI: 10.17694/bajece.494920

- kaggle.com. (n.d.). *Tutorial: Time Series Forecasting with Prophet*. [online] Available at: https://www.kaggle.com/code/prashant111/tutorial-time-series-forecasting-with-prophet.

Thank you!

Can Yang, Haode Sun, Kangtong Guo,
Xinyu Jia, Xujia Zhang