Sales Forecasting for Grocery Store, Favorita

Marketing Analytics - Fall'22

Group M

Anudeep Akkana (aa92799)

Muskan Agarwal (ma64547)

Tanvi Dalal (TRD878)

Sreekar Lanka (SL54387)

Prathmesh Savale (ps33296)

AGENDA

- Project Background
- Data & Metrics
- Exploratory Data Analysis
- Time Series Forecasting
- Conclusion

Project Background and Objectives

Project Background

Inventory Problems

- Lack of popular items
- Lost revenue
- Extra product waste

Forecast

- Brick and mortar stores for sales to determine how much inventory to buy

Result

- More accurate forecasting can decrease food waste and increase customer satisfaction

Objectives

Given the daily sales across stores and product families, forecast the future daily sales at store-product level

Perform exploratory data analysis and time series analysis of Favorita stores data

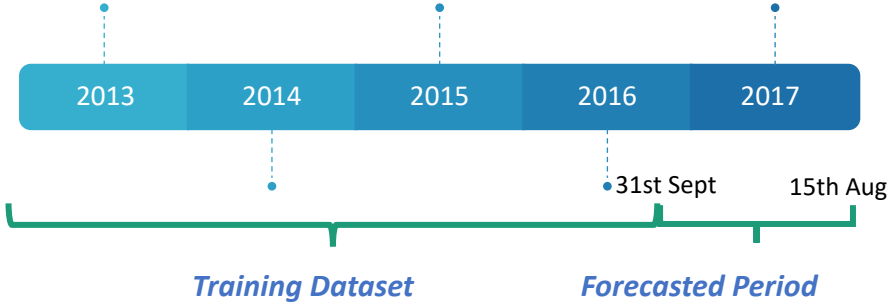
Identify the factors that impact sales

Develop time series model to forecast sales

Data and Metrics

Description

Time series product sales data for Favorita stores



Datasets and Variables:

- **Holiday Events:** Date, type, locale, location, description, transferred
- **Oil:** Date, price
- **Stores:** store_nbr, city, state, type, cluster,
- **Promotions**

Model Evaluation Metric (MAPE)

Mean absolute percentage error MAPE

$$M = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

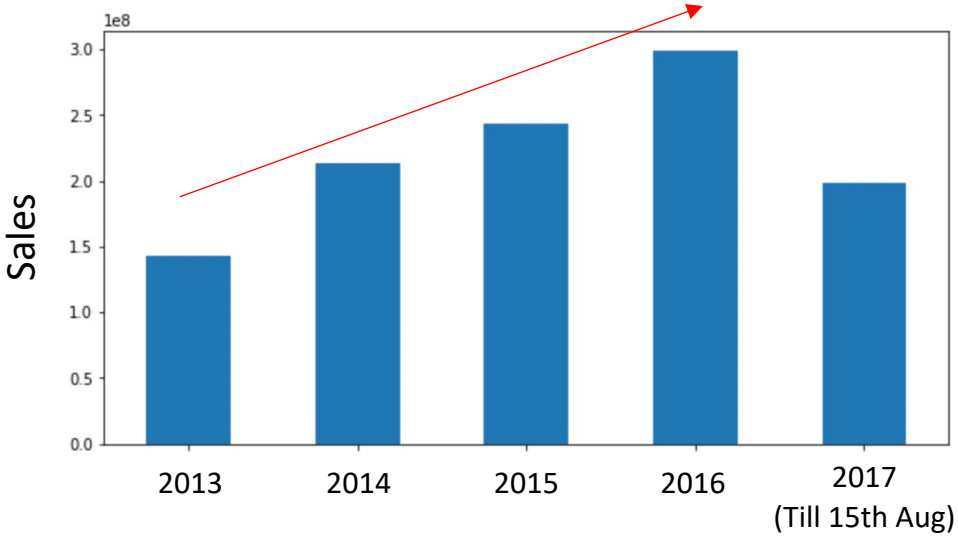
M = mean absolute percentage error

n = number of times the summation iteration happens

A_t = actual value

F_t = forecast value

EDA – Overall Sales are increasing every year for Favorita



Top 4 Stores by Sales

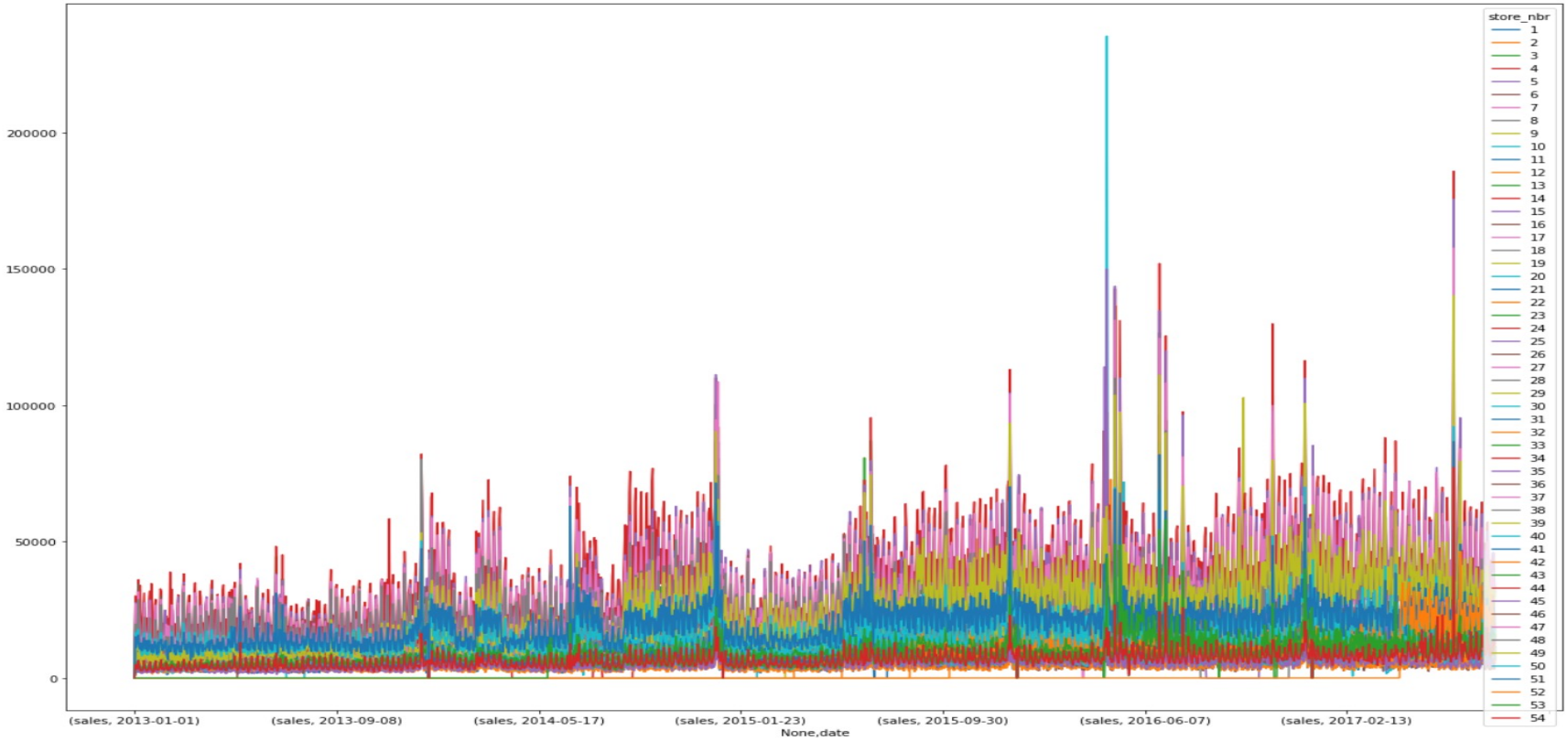
Store No.	Sales(K)	City	State
44	63.3	Quito	Pichincha
45	55.6	Quito	Pichincha
47	52.0	Quito	Pichincha
3	51.5	Quito	Pichincha

Last 5 Stores by Sales

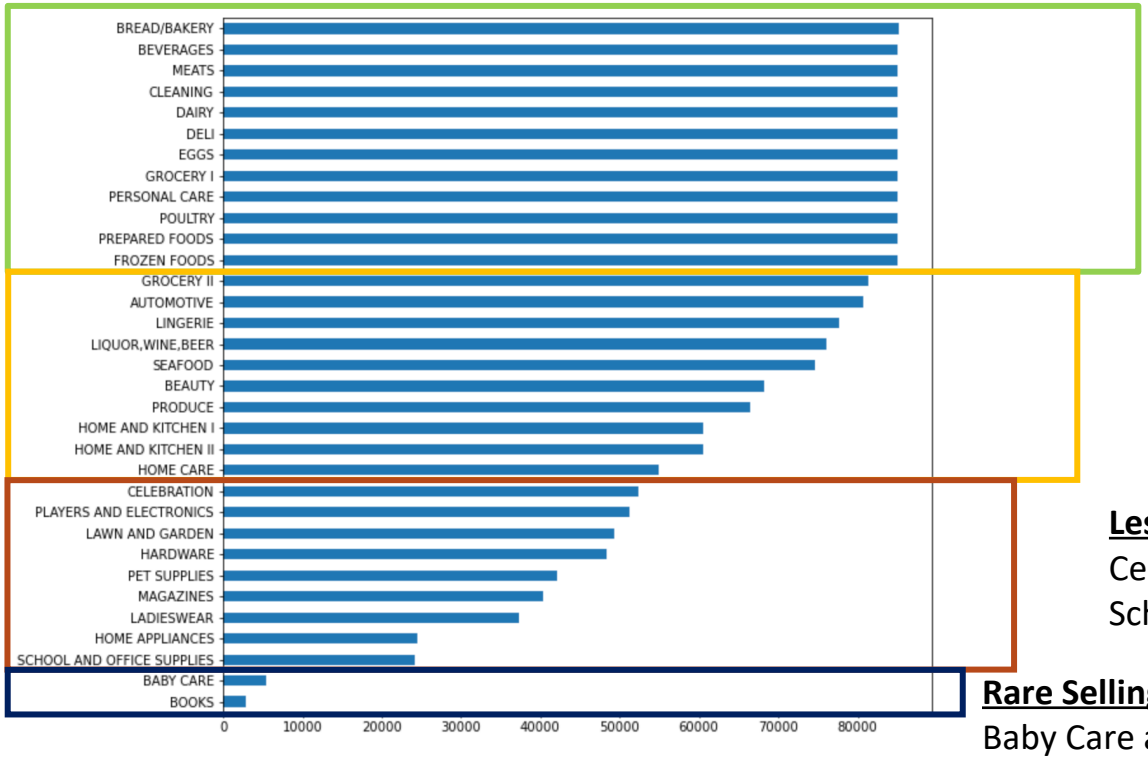
Store No.	Sales(K)	City	State
35	7.8	Playas	Guayas
30	7.5	Guayaquil	Guayas
32	6.1	Guayaquil	Guayas
22	4.2	Puyo	Pastaza
52	2.7	Manta	Manabi

EDA - Time Series for Stores

All stores have similar time series shapes



Different Product Types



Everyday Selling
Bread/Bakery, Dairy, Meats etc. which are daily need products

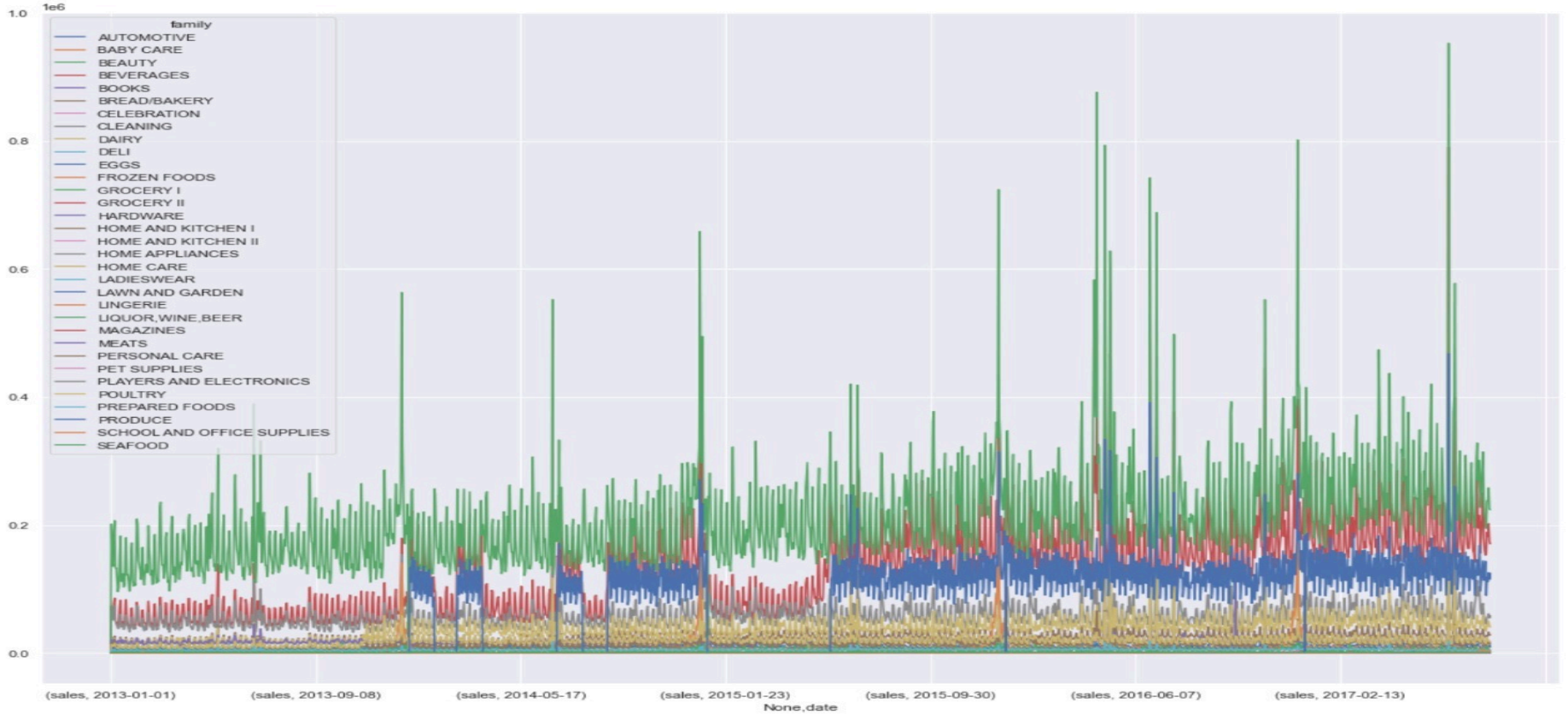
Required but less frequent
Liquor, Kitchen , Home care

Less frequent and Seasonal
Celebration supplies, Hardware, School & Office supplies

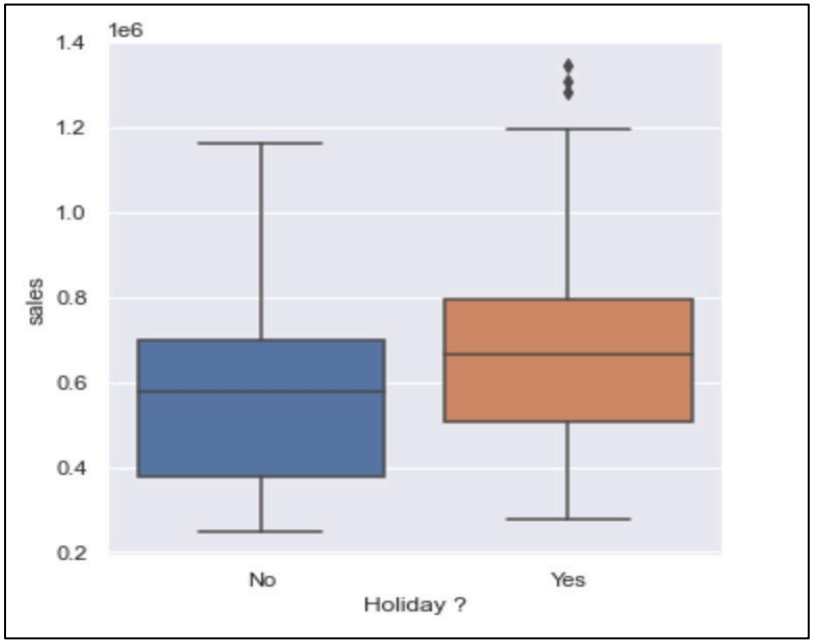
Rare Selling
Baby Care and Books

EDA - Time Series for product families

Different product families show variety of time series shapes



Impact of external factors on sales - Holidays



OLS Regression Results

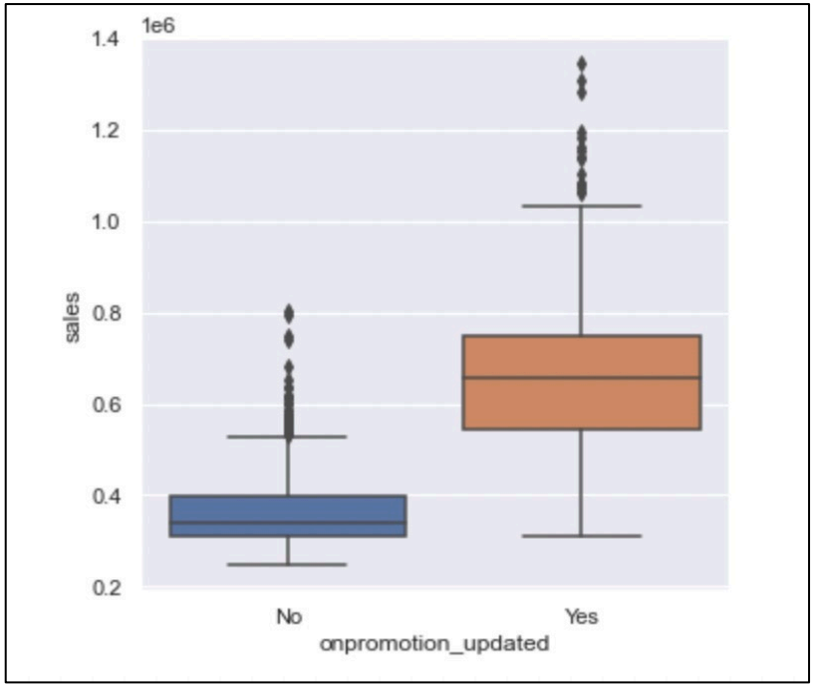
Dep. Variable:	y	**R-squared (uncentered):**	0.187
Model:	OLS	**Adj. R-squared (uncentered):**	0.186
Method:	Least Squares	**F-statistic:**	269.9
Date:	Mon, 07 Nov 2022	**Prob (F-statistic):**	9.06e-55
Time:	22:17:26	**Log-Likelihood:**	-17236.
No. Observations:	1178	**AIC:**	3.447e+04
Df Residuals:	1177	**BIC:**	3.448e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
x1	6.684e+05	4.07e+04	16.430	0.000	5.89e+05	7.48e+05

Holidays have a positive impact on sales

The sales go up by 668K at an overall level if it is a holiday given all else constant

Impact of external factors on sales - Promotions



OLS Regression Results

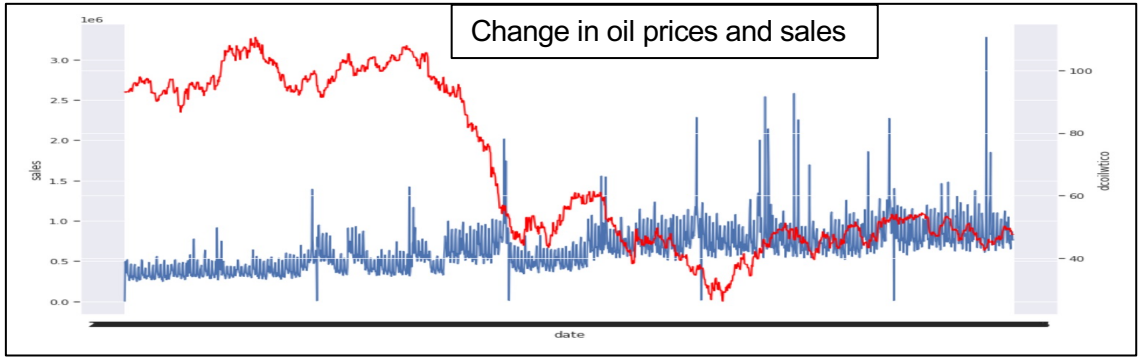
Dep. Variable:	y	**R-squared (uncentered):**	0.834
Model:	OLS	**Adj. R-squared (uncentered):**	0.834
Method:	Least Squares	**F-statistic:**	5847.
Date:	Mon, 07 Nov 2022	**Prob (F-statistic):**	0.00
Time:	21:53:04	**Log-Likelihood:**	-16089.
No. Observations:	1163	**AIC:**	3.218e+04
Df Residuals:	1162	**BIC:**	3.218e+04
Df Model:	1		
Covariance Type:	nonrobust		

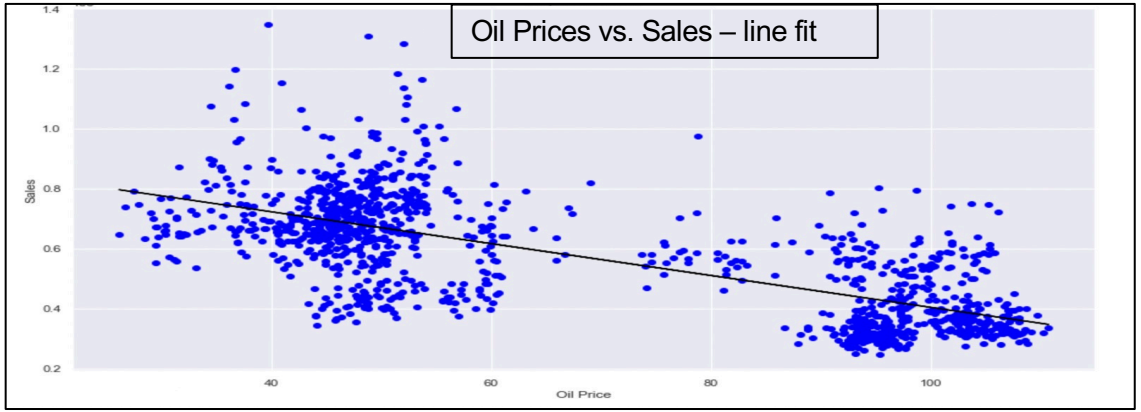
	coef	std err	t	P>|t|	[0.025	0.975]
x1	6.467e+05	8457.384	76.468	0.000	6.3e+05	6.63e+05

Promotions have a positive impact on sales

The sales go up by 647K at an overall level if it is a promotion day given all else constant

Impact of external factors on sales - Oil





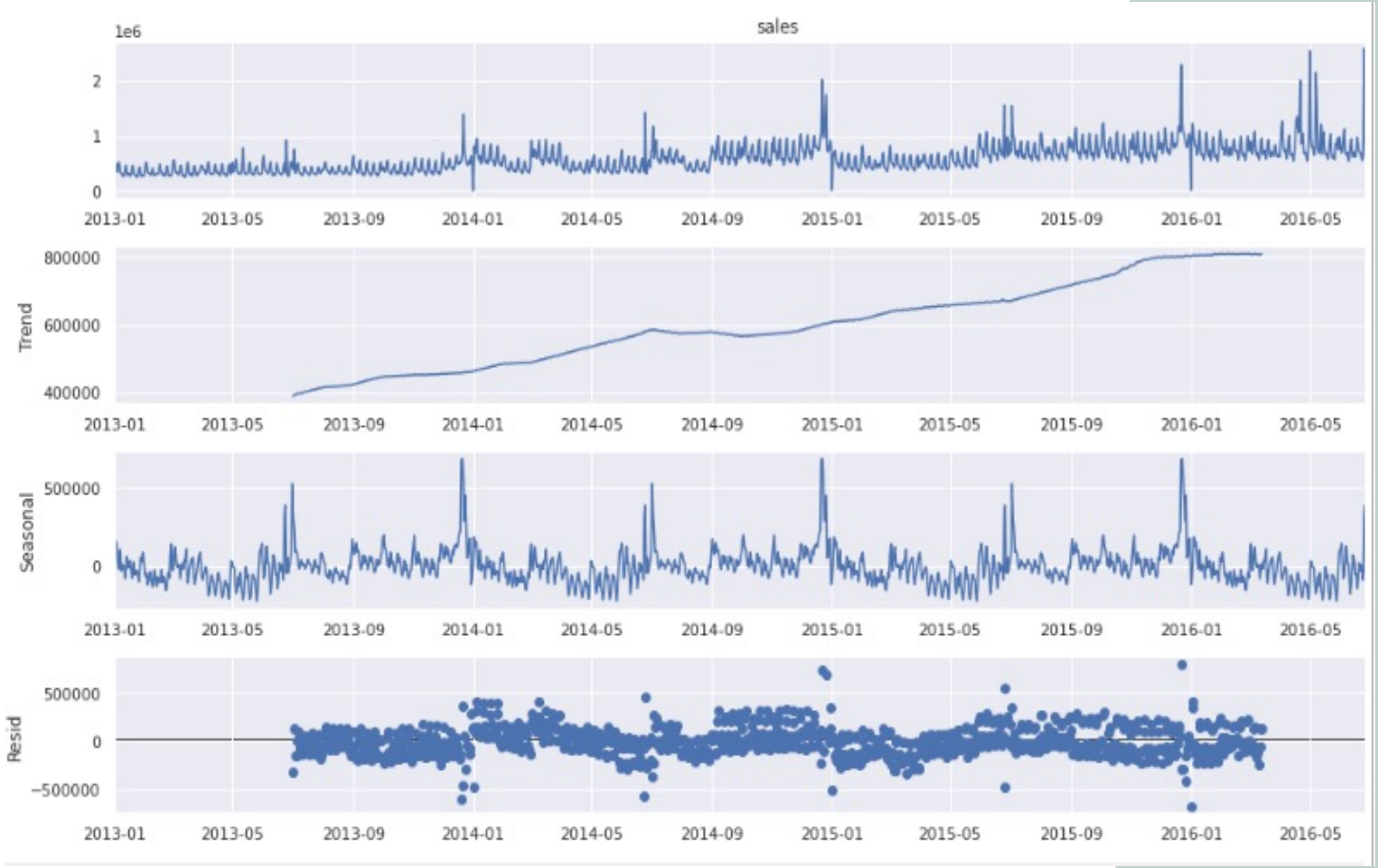
OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.310
Model:	OLS	**Adj. R-squared:**	0.309
Method:	Least Squares	**F-statistic:**	754.9
Date:	Tue, 08 Nov 2022	**Prob (F-statistic):**	1.36e-137
Time:	11:22:22	**Log-Likelihood:**	-23177.
No. Observations:	1684	**AIC:**	4.636e+04
Df Residuals:	1682	**BIC:**	4.637e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.057e+06	1.58e+04	66.894	0.000	1.03e+06	1.09e+06
x1	-5984.1028	217.801	-27.475	0.000	-6411.291	-5556.914

There is a strong negative correlation between sales and oil prices

Time Series Decomposition



Original Time series

General Trend

Seasonality

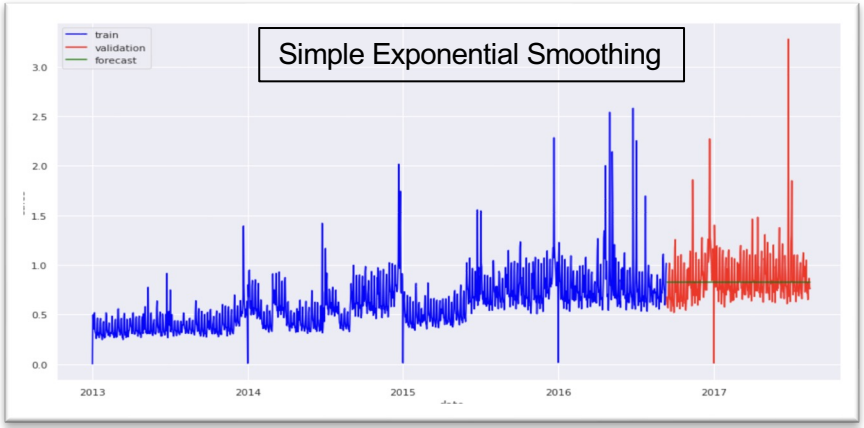
Residuals

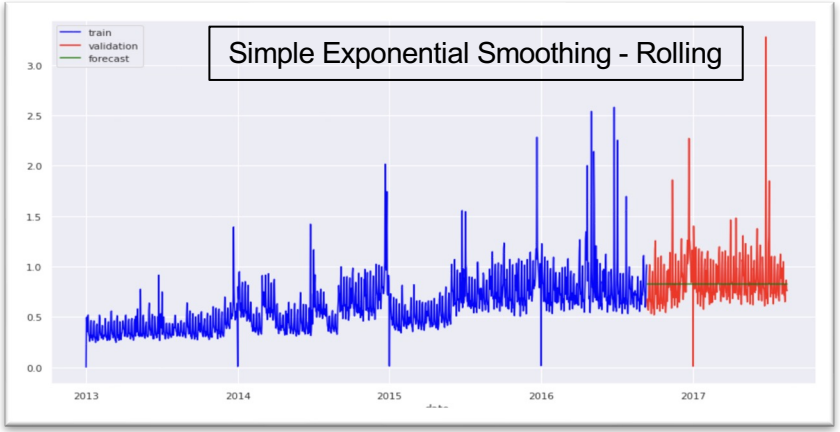
Results for different iterations

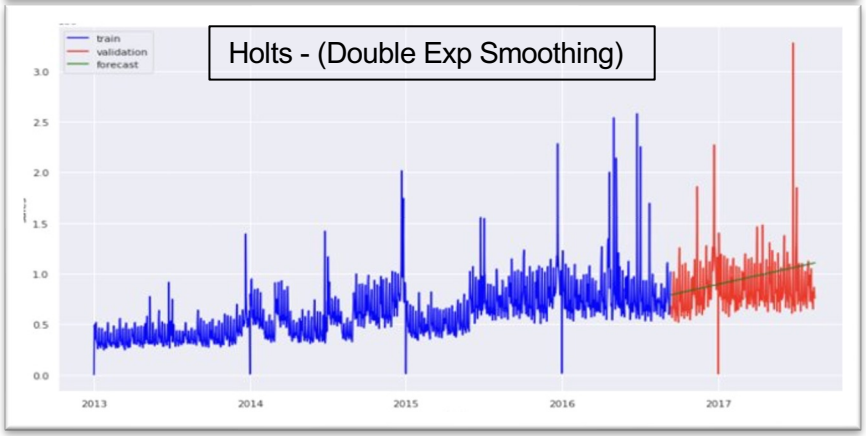
Method	MAPE
Simple Exp smoothing	37%
Rolling exponential smoothing	37%
Holts (double exp smoothing)	46.14%
Holts (double exp smoothing) - rolling	35.17%
Holts Winter (triple exp smoothing)	48.45%
Holts Winter (triple exp smoothing) - rolling	45.99%
Propagate trend and seasonality	35.09%
Propagate trend and seasonality and adjust residue	35.08%

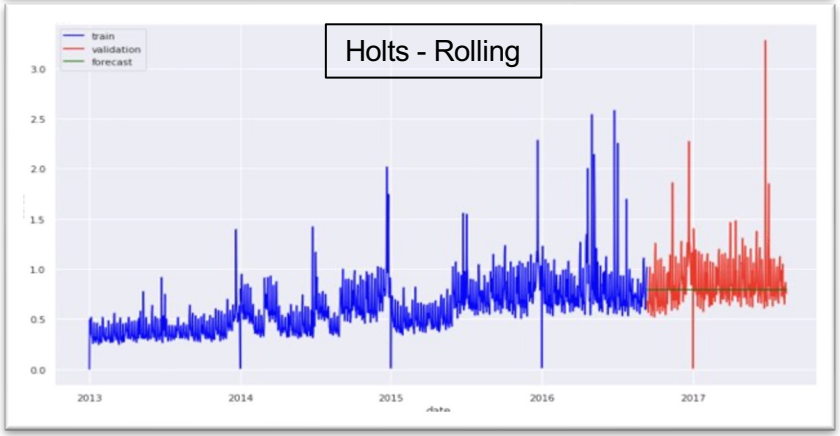
- Rolling point forecasts have better accuracy compared to longer forecasts
- Trend and seasonality propagation with adjusted residue have the best results
- Double exponential smoothing results are comparable to propagating trend and seasonality

Forecast visuals - (Simple Exponential Smoothing and Holts)

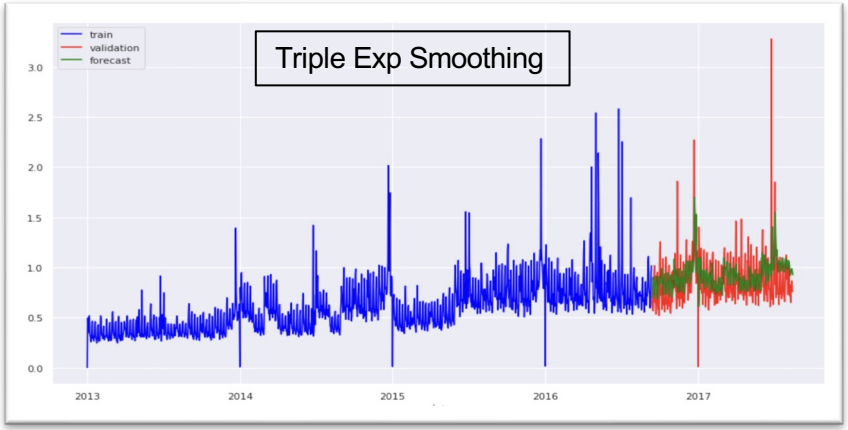


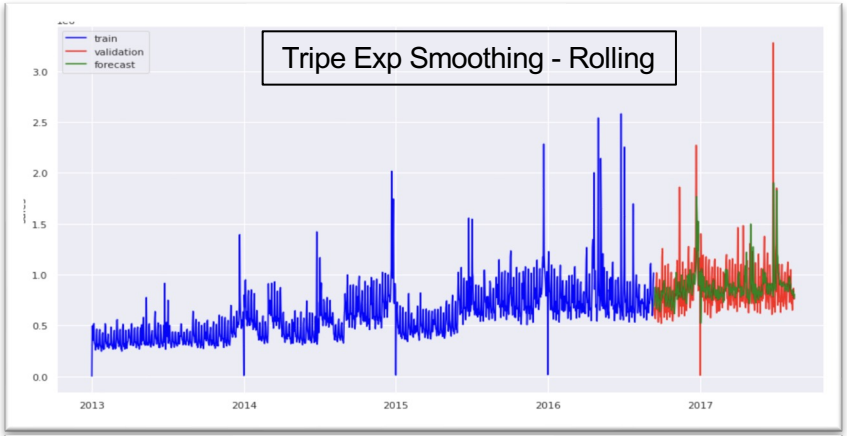


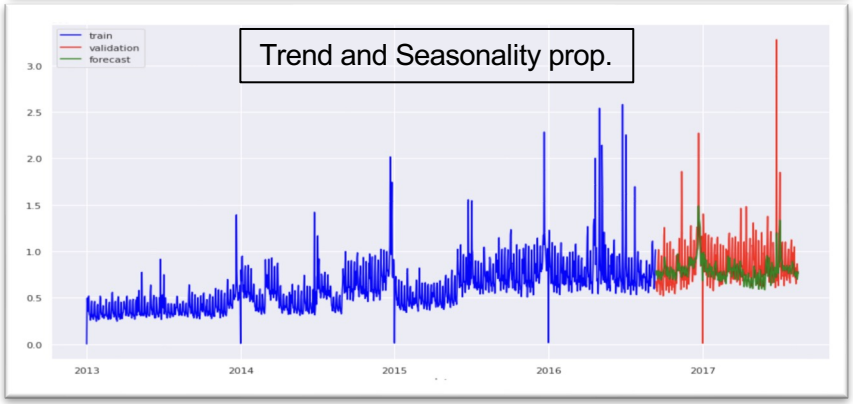


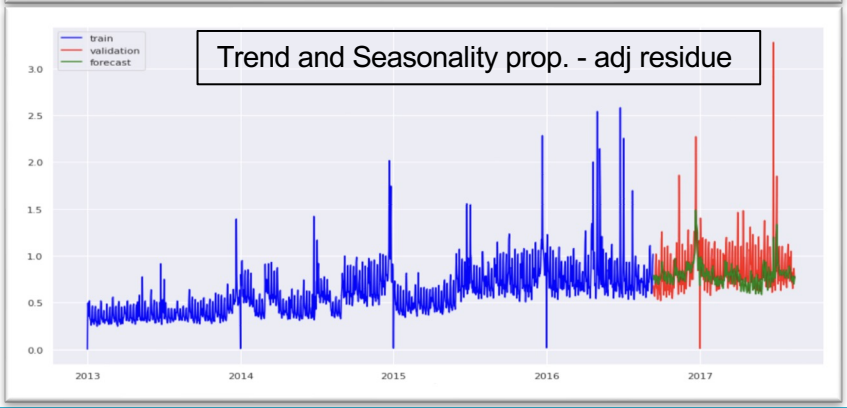


Forecast visuals









Recommendations

- Using this forecast data, Favorita stores can plan inventory for the coming months
- It is also important to track the oil prices as they influence sales
- The surge in sales during holidays and promotions can also be planned since we included the adjustment for these external regressors in the model
- At a store level - Favorita should focus on store 3, 44, 45, 47 and 49

Next Steps

- Going forward... use store and product level forecasts at a daily level with adjustment for external regressors to determine inventory planning at a store level

Thank You

Appendix

Time series shape clustering –