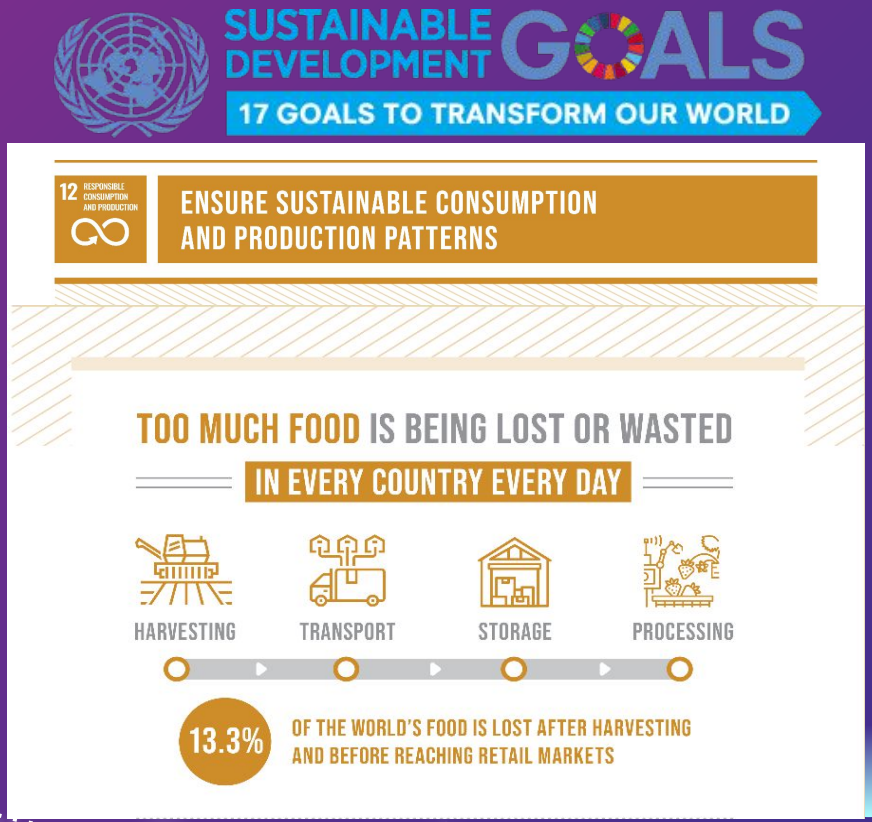
Mini Project

By: Ni Jun Hong Glenn & See Qin Yuan

01

Motivation

Motivation


SUSTAINABLE DEVELOPMENT GOALS
17 GOALS TO TRANSFORM OUR WORLD
12 RESPONSIBLE CONSUMPTION AND PRODUCTION
ENSURE SUSTAINABLE CONSUMPTION AND PRODUCTION PATTERNS
TOO MUCH FOOD IS BEING LOST OR WASTED
IN EVERY COUNTRY EVERY DAY
HARVESTING
TRANSPORT
STORAGE
PROCESSING
13.3%
OF THE WORLD'S FOOD IS LOST AFTER HARVESTING AND BEFORE REACHING RETAIL MARKETS

Problem Definition

02

Exploratory Data Analysis

Data Preparation

Before

	id	date	store_nbr	family	sales	onpromotion
0	0	2013-01-01	1	AUTOMOTIVE	0.0	0
1	1	2013-01-01	1	BABY CARE	0.0	0
2	2	2013-01-01	1	BEAUTY	0.0	0
3	3	2013-01-01	1	BEVERAGES	0.0	0
4	4	2013-01-01	1	BOOKS	0.0	0

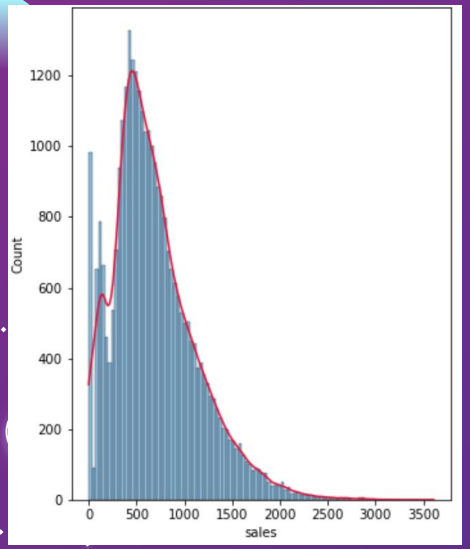
	date	type	transferred
0	2012-03-02	Holiday	False
1	2012-04-01	Holiday	False
2	2012-04-12	Holiday	False
3	2012-04-14	Holiday	False
4	2012-04-21	Holiday	False

After

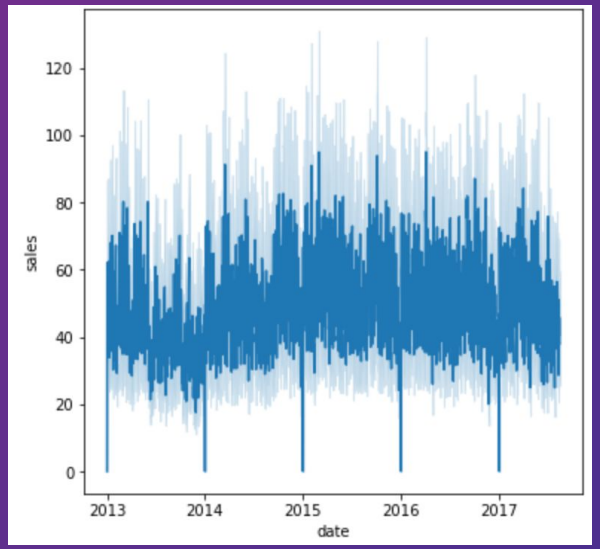
	id	date	store_nbr	family	sales	onpromotion	holiday
5	5.0	2013-01-01	1.0	BREAD/BAKERY	0.0	0.0	1.0
8	8.0	2013-01-01	1.0	DAIRY	0.0	0.0	1.0
9	9.0	2013-01-01	1.0	DELI	0.0	0.0	1.0
10	10.0	2013-01-01	1.0	EGGS	0.0	0.0	1.0
11	11.0	2013-01-01	1.0	FROZEN FOODS	0.0	0.0	1.0

Data Visualisation

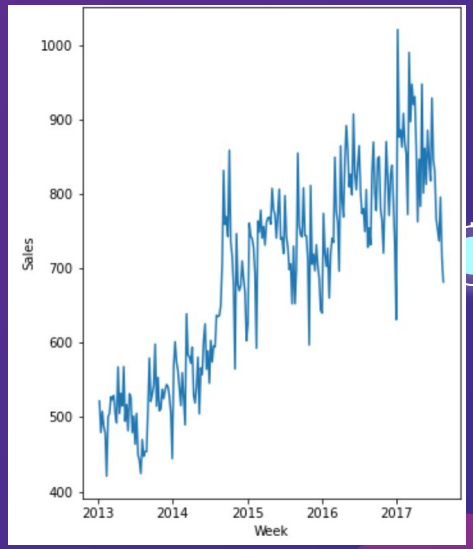
Distribution



Line Plot: Daily

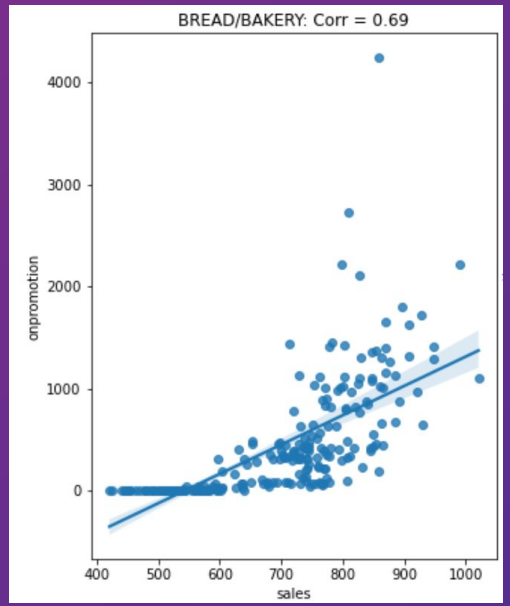


Line Plot: Weekly

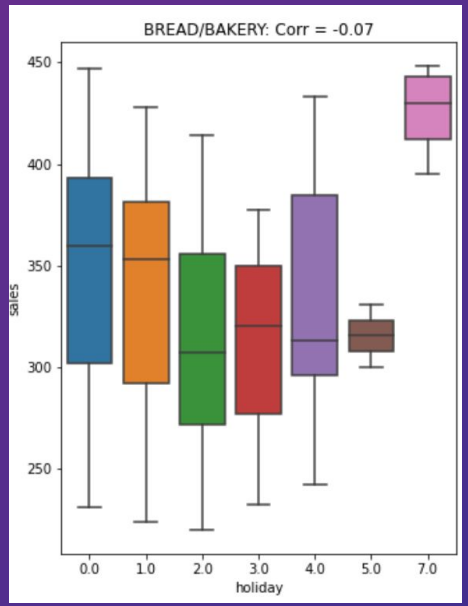


Correlation Analysis

Scatter Plot



Box Plot



03 Machine Learning

ARIMA Model

ARIMA

AutoRegression (AR)

Uses past values to predict future values.

Differencing (I)

Value to make past value stationary

Moving Average (MA)

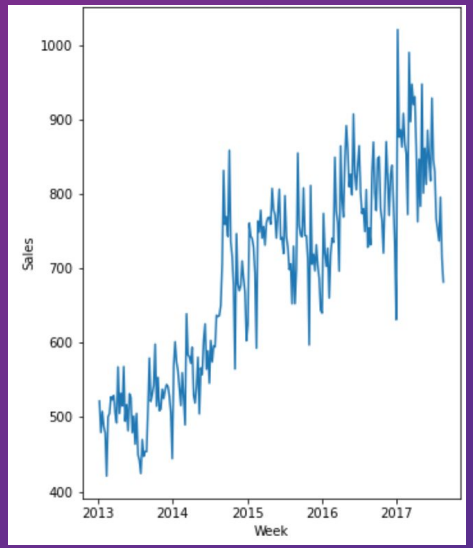
Uses errors of past values to predict future values

AR MA

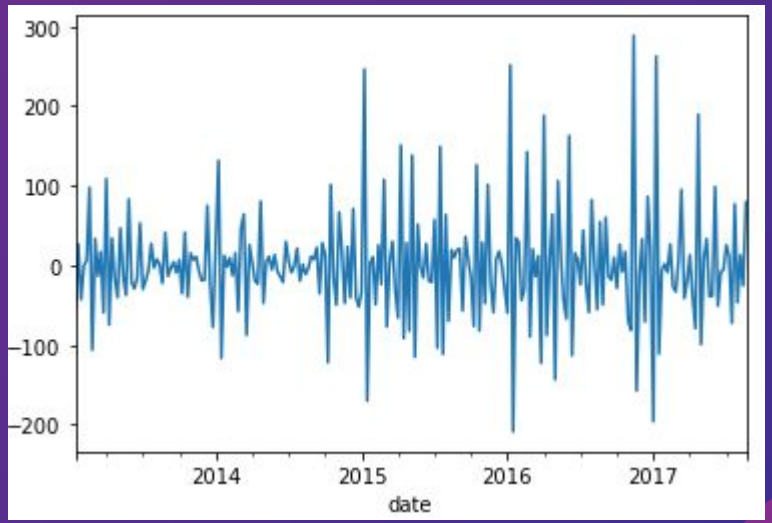
$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + .. + \phi_q \epsilon_{t-q}$$

Stationarity

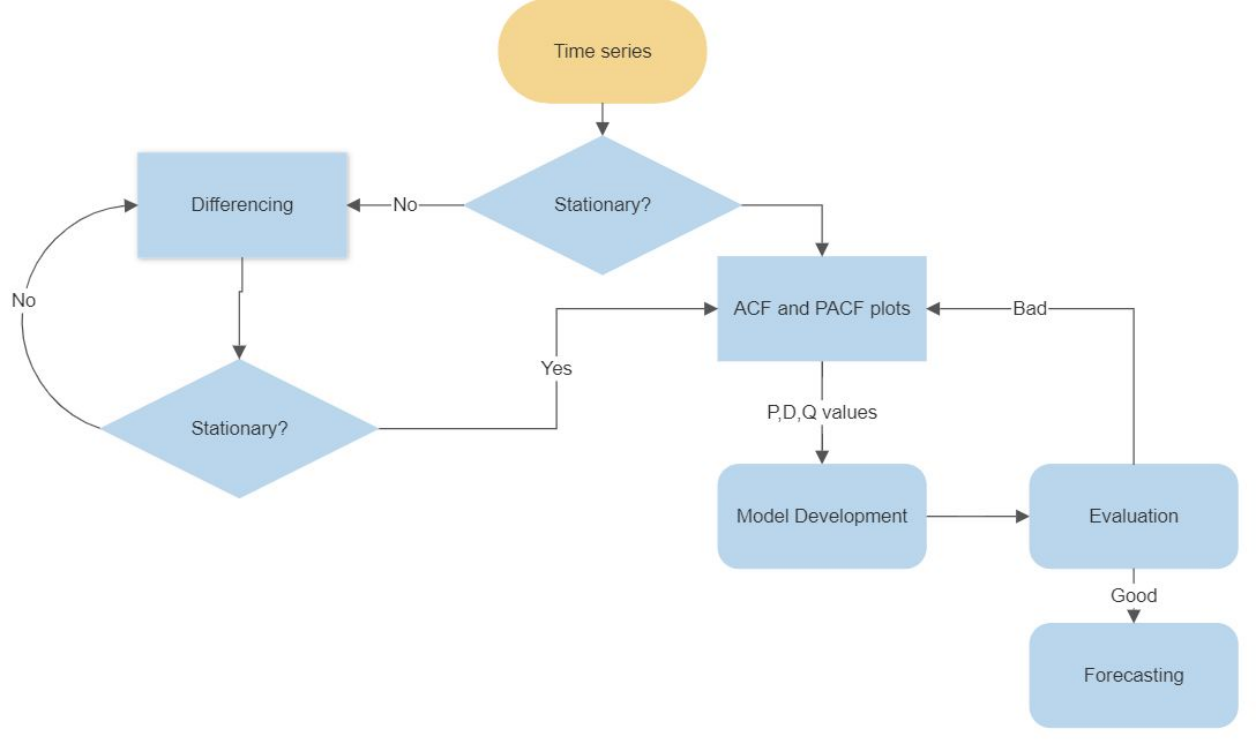
Before differencing



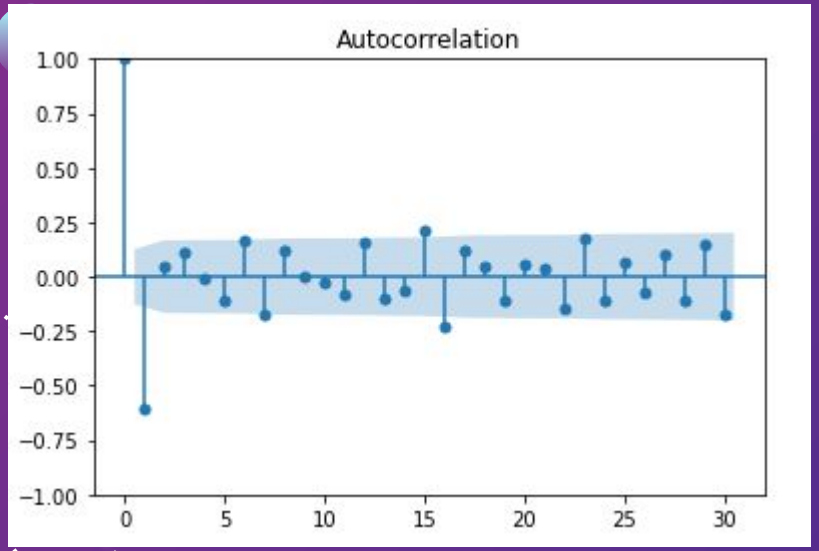
After differencing

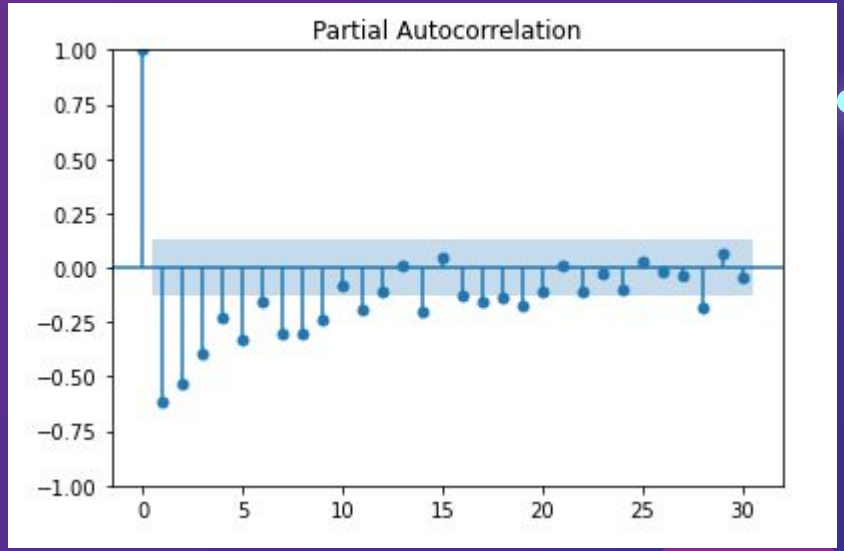


Model Training Process

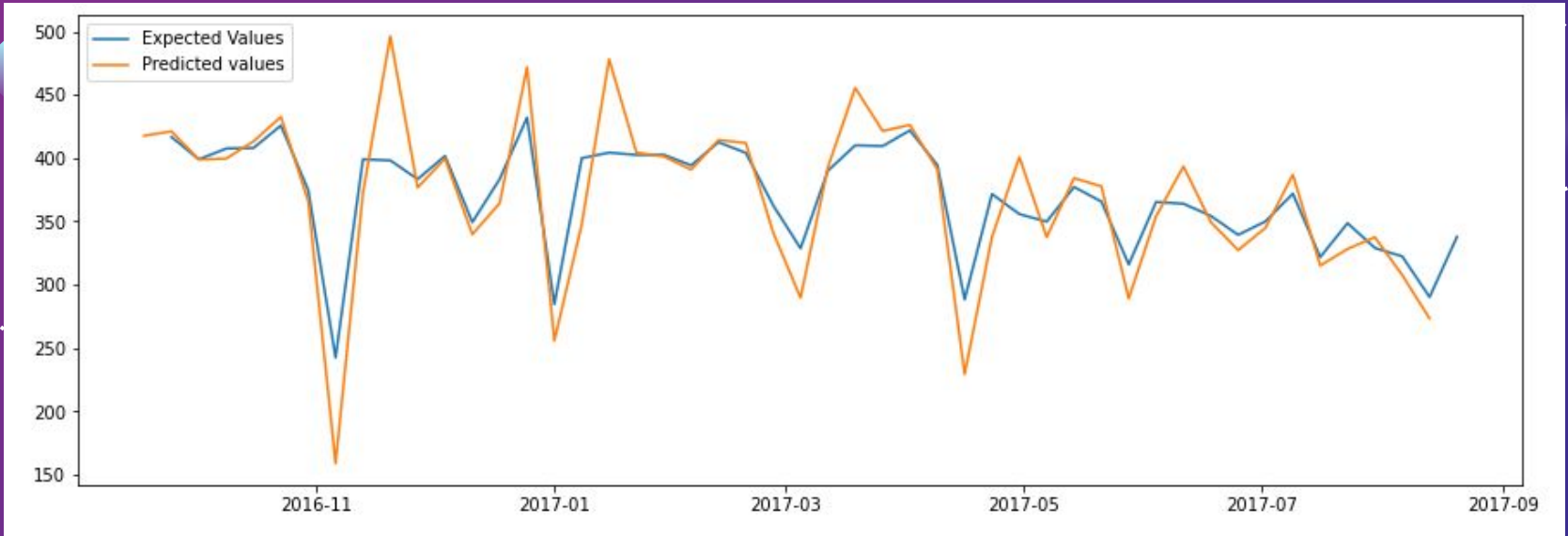


ACF and PACF test





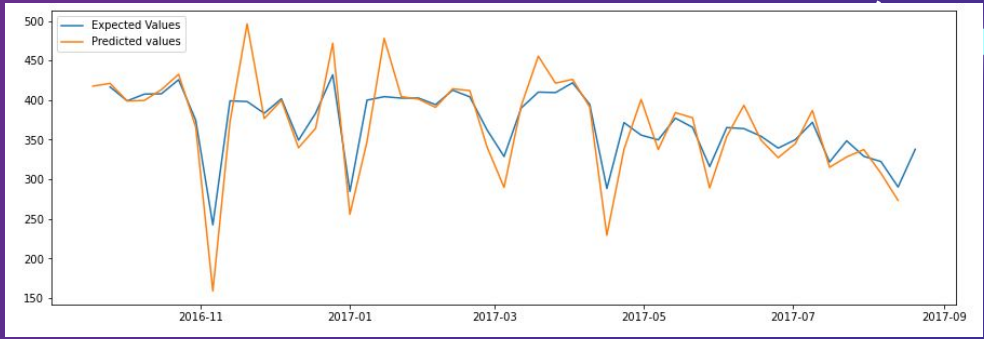
Results


Expected Values
Predicted values
500
450
400
350
300
250
200
150
2016-11
2017-01
2017-03
2017-05
2017-07
2017-09

Improvements

Observations

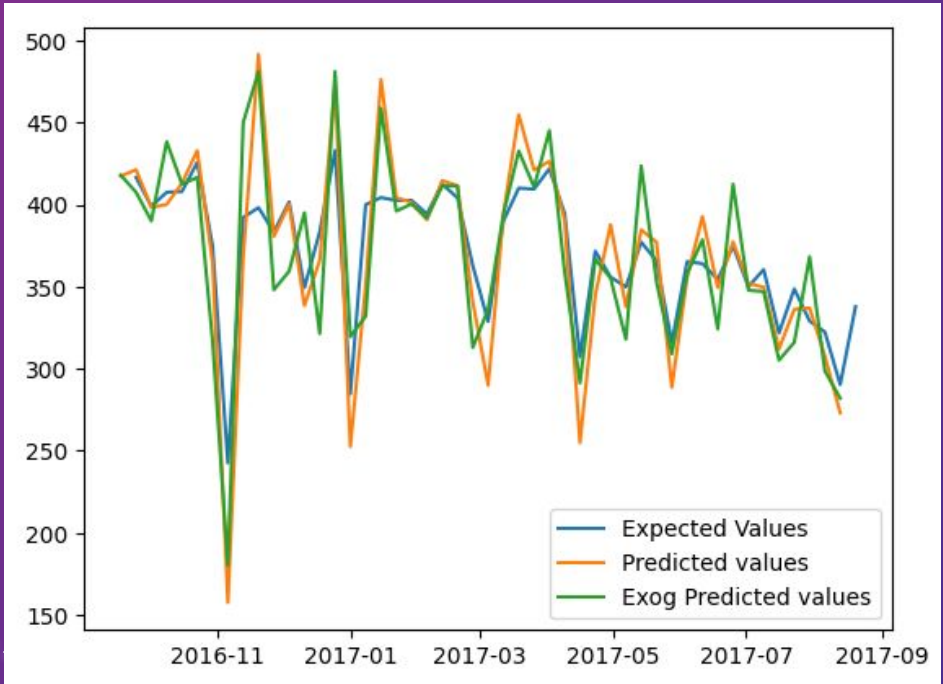
- RMSE: 70.652
- Model performed quite accurately



Can it be improved?

- ARIMA only uses past values to predict future sales
- There might be other variables that can also affect sales
  - Presence of holidays
  - Items on promotion

Results (Addition of exogenous var)



RMSE:

- Without Exog: 70.652
- With Exog Variables: 66.708

04

Conclusion

Improvements

What we achieved

- Train a model that uses past values and other independent variables to predict sales

Contributions to store owners

- Model can be used to predict the future sales of the different food categories and use the values to order the right amount of food for the following weeks.
- This can help to curb down on food wastage due to over-ordering.

Data-driven Insights

- Store owners should take note of holiday periods and how they give promotions on their products as they can affect the demand of these products.

Thank you!