Sales of Books Forecasting

Lin Jiahong

Nanjing University of Science and Technology

2023-01-28

Overview

Problem Definition

Sales of Books Forecast

Data Analysis

Feature Extraction

Model Train

Conclusion

Problem Definition

Sales of Books Forecast

Defn Sales of Books Forecast aims to predict the sales of books in 2021 through the book sales data from 2017 to 2020.

- Data covers different countries and different stores.
- There are cyclical and seasonal changes in book sales.

Data	row_num	date	country	store	product
train	70128	1461	6	2	4
test	17520	365	6	2	4

Data Analysis

Overall data

- Country - Belgium,France,Germany,Italy,Poland,Spain
- Product - [Kaggle Advanced Techniques],[Kaggle Getting Started],[Kaggle Recipe Book],[Kaggle for Kids: One Smart Goose]
- Stores - KaggleMart,KaggleRama
- Time line

Data	Earliest date	Latest date
train	2017 – 01 – 01	2020 – 12 – 31
test	2021 – 01 – 01	2021 – 12 – 31

Monthly sales statistics

- the patterns in sales of all countries and stores are identical.the magnitudes of sales are different

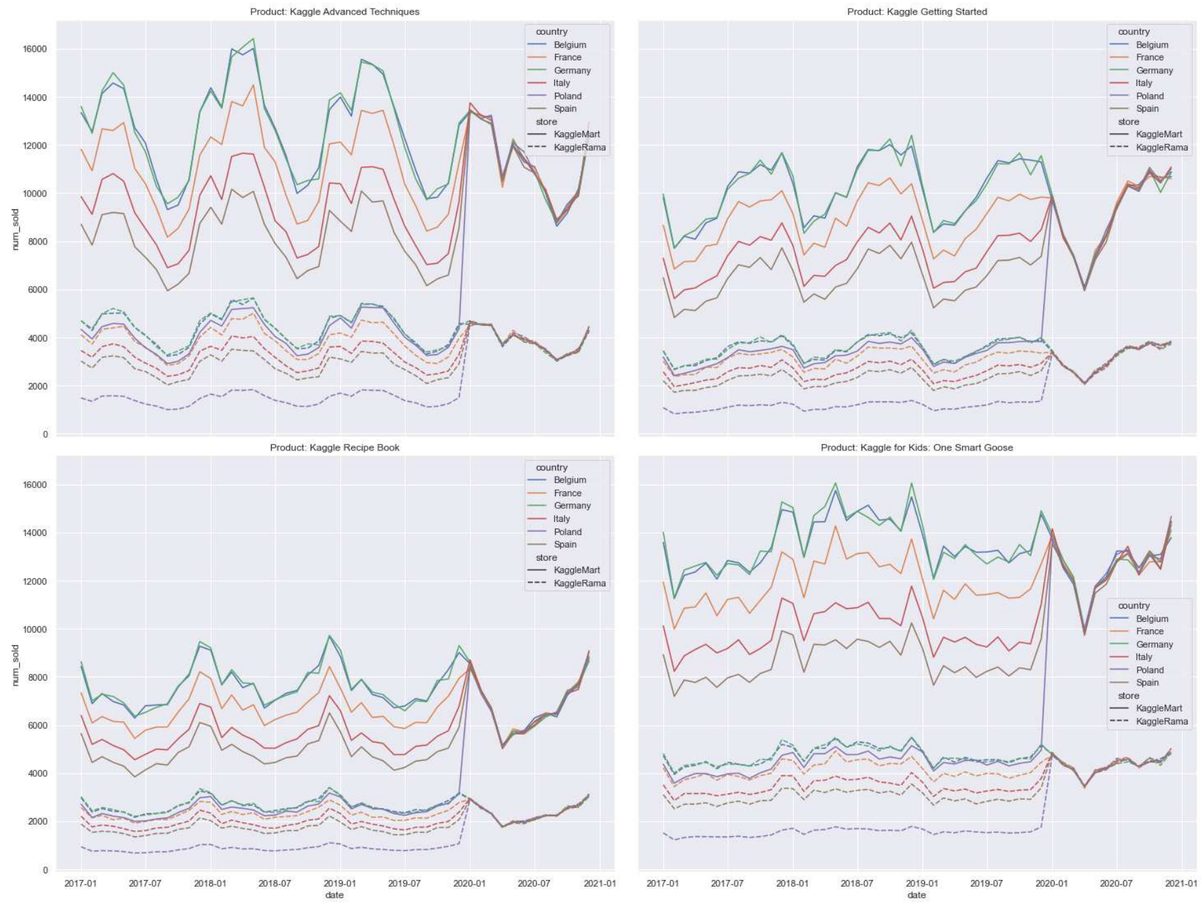

Figure 1: Monthly sales

Aggregating Time Series(Store)

- Store-KaggleMart appears to consistantly have 74.25% of the total number of sales

Store	ratio
KaggleMart	0.742515
KaggleRama	0.257485

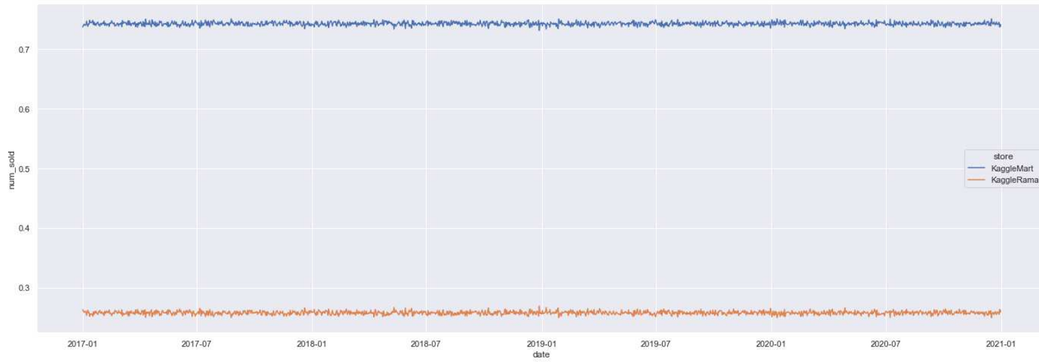

Figure 2: Stores ratio

Aggregating Time Series(Store)

- To compare the trend of the two stores, multiply the sales data of the two stores by a constant.

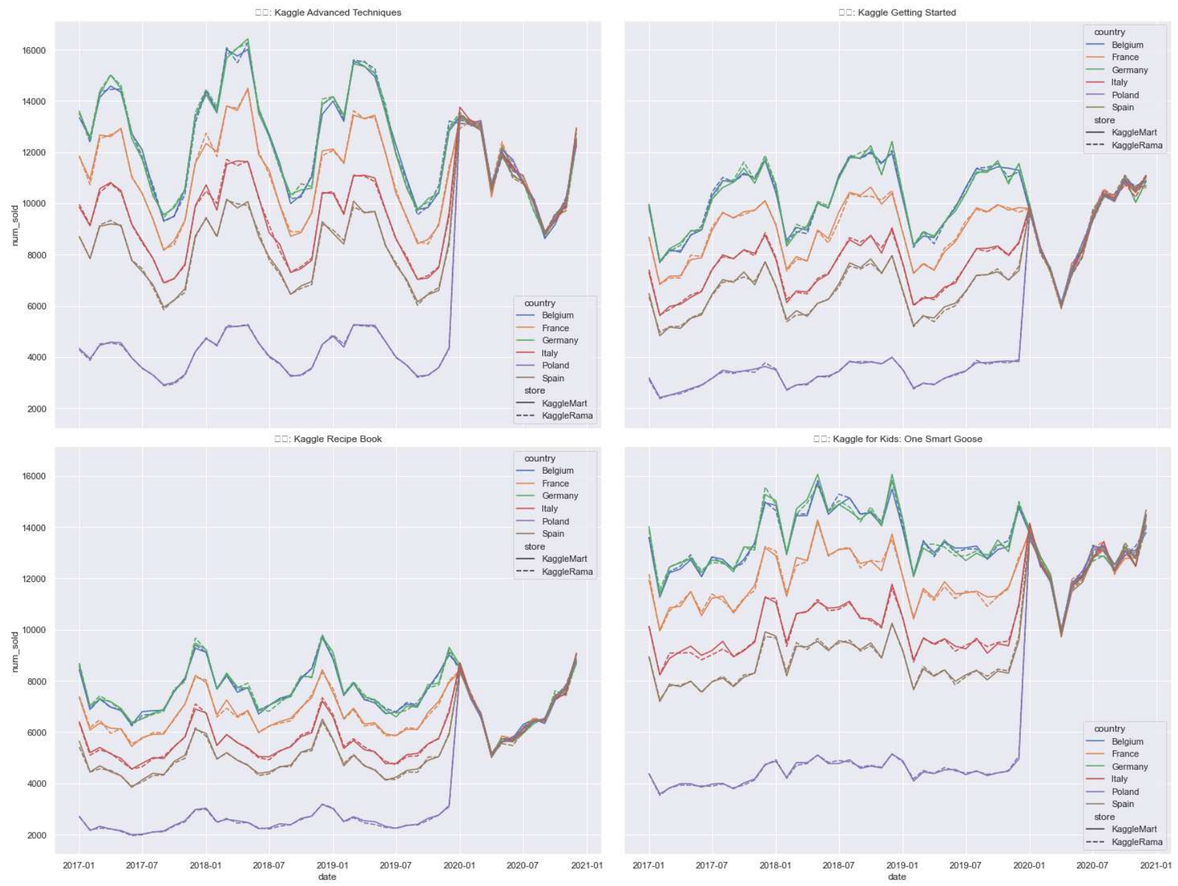

Figure 3: Stores ratio trend

Aggregating Time Series(Country)

- Country-The ratio of total sales in different countries also fluctuates little.

Country	ratio
Belgium	0.218930
France	0.191360
Germany	0.219586
Italy	0.159383
Poland	0.071348
Spain	0.139393

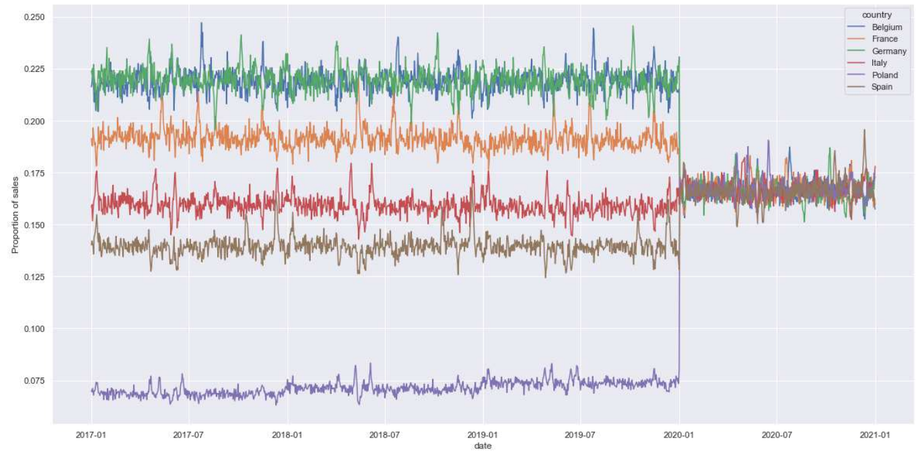

Figure 4: Countries ratio

Aggregating Time Series(Country)

- Multiply all countries by a constant so they are comparable with Belgium.

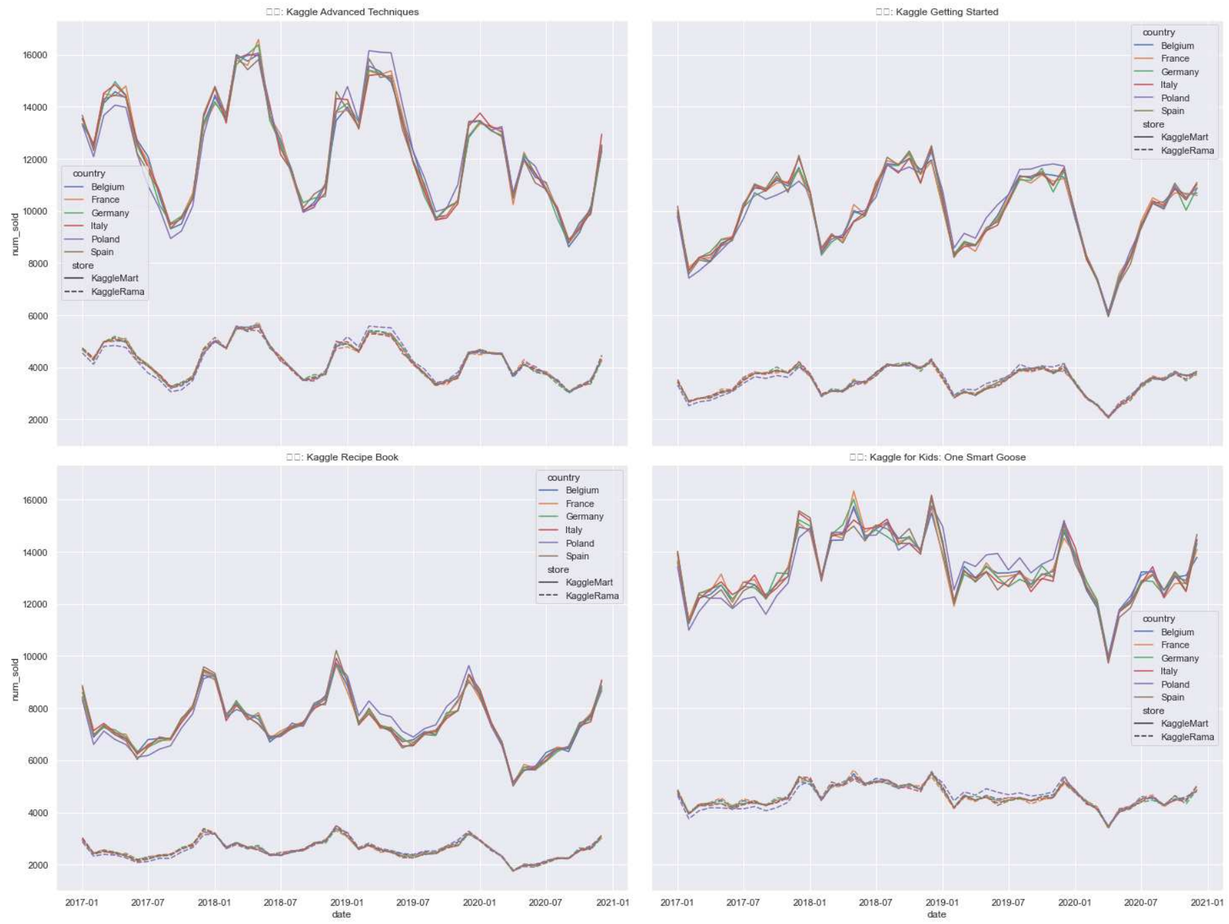

Figure 5: Countries ratio trend

Aggregating Time Series(Country and Store)

- In the plots make all time series inline with the Belgium KaggleMart store by multiplying by a constant.

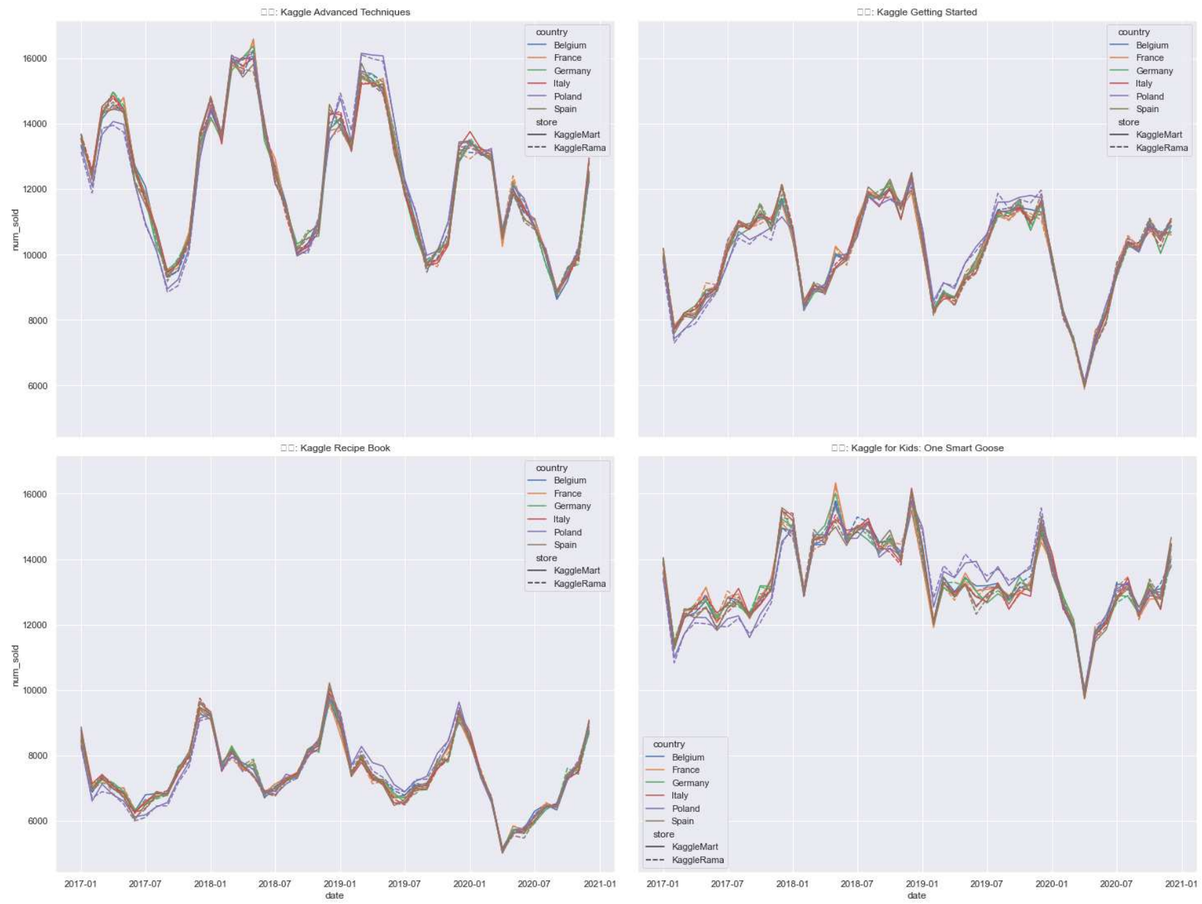

Figure 6: Countries and Store trend

Aggregating Time Series(Product)

- The change trend of the sales volume of the four books is cyclical.

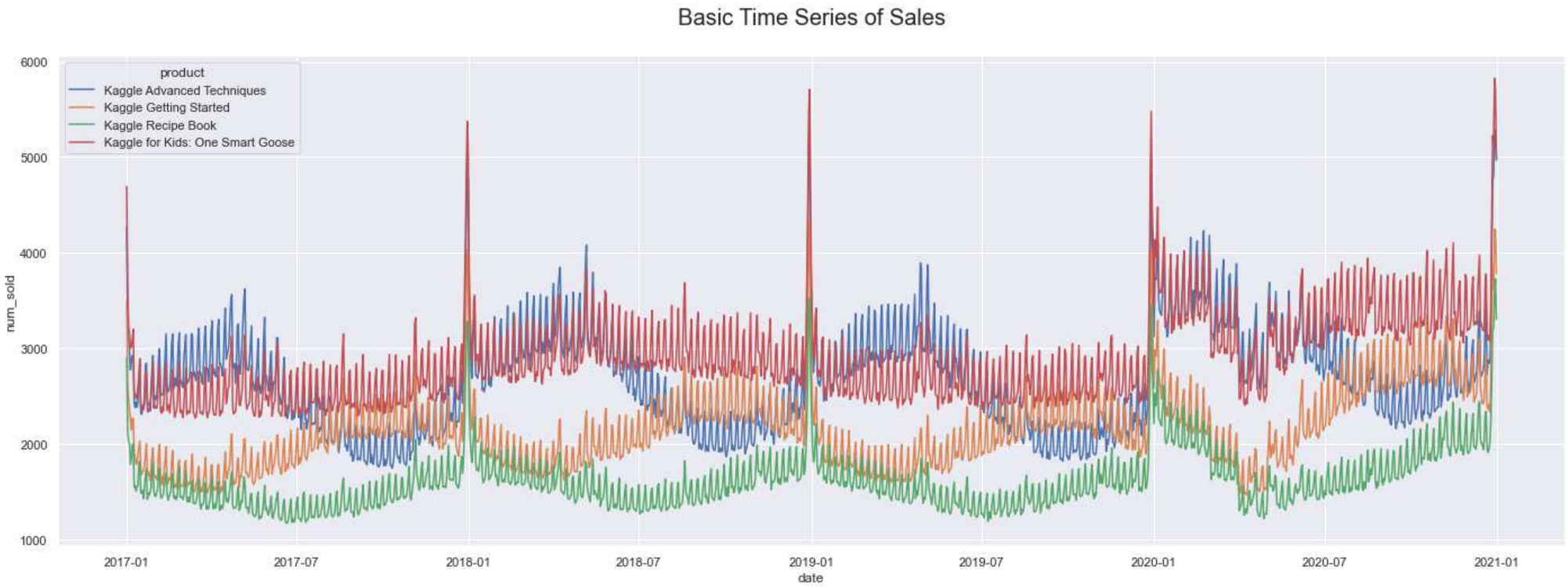

Figure 7: Sales of Product

Aggregating Time Series(Product)

- The change trend of the sales proportion of the four books has rules.

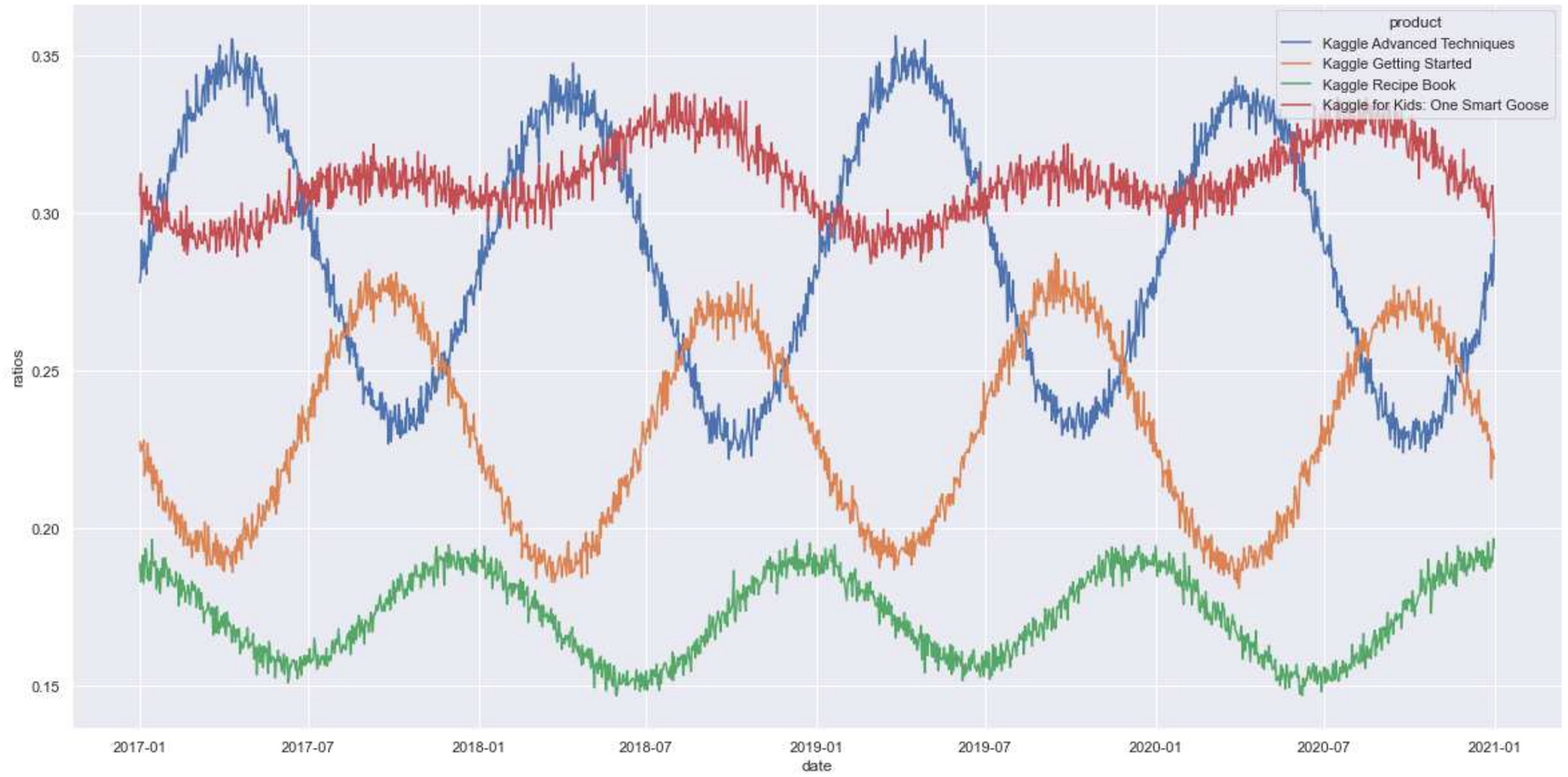

Figure 8: Product ratio trend

Aggregated Time Series

- aggregate the sales timeline to consider how to forecast the overall sales volume.

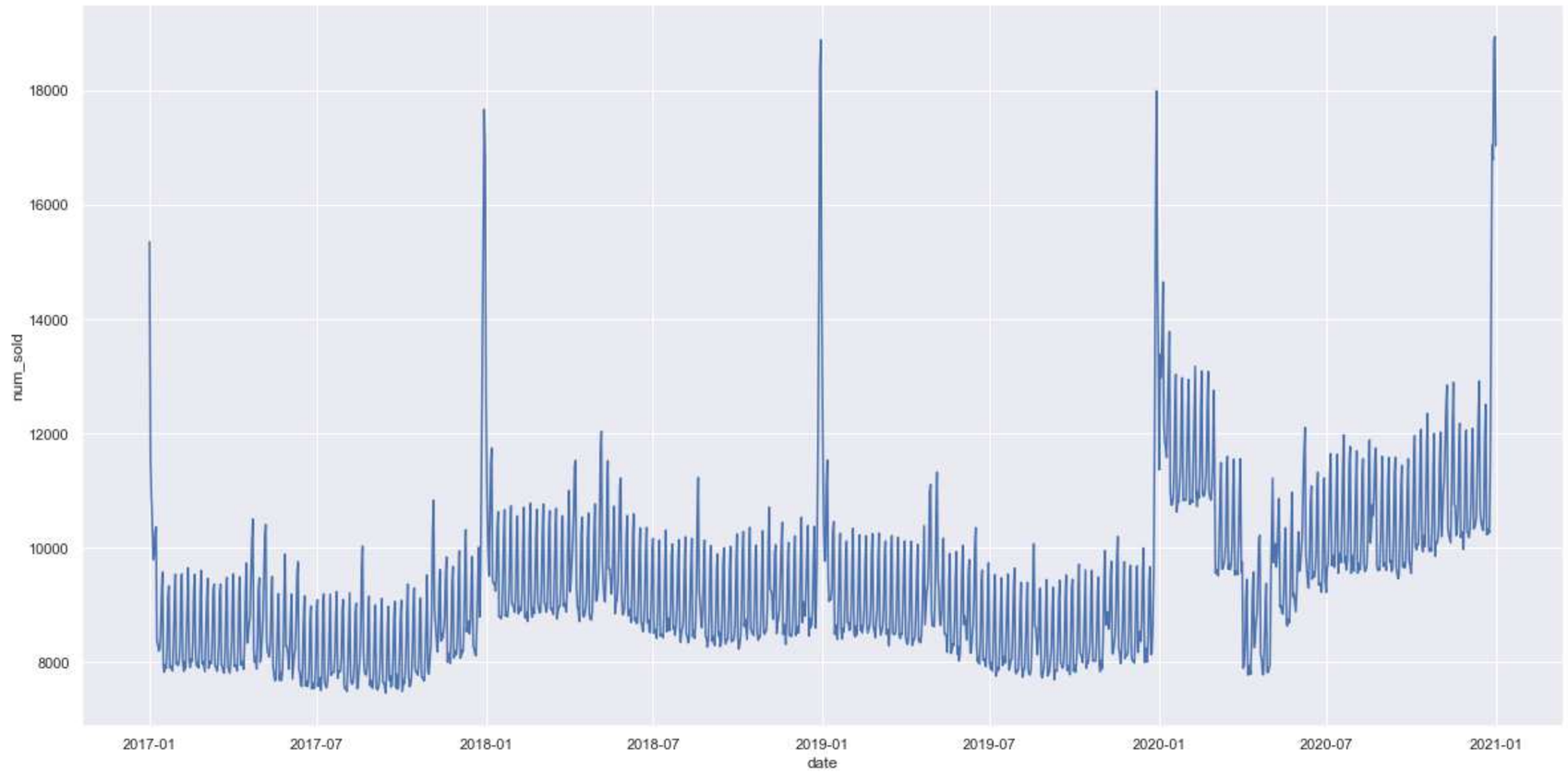

Figure 9: Aggregated time series

Feature Extraction

Forecast target

- the total sales of each day.
- the sales volume of different products in each day of the year.

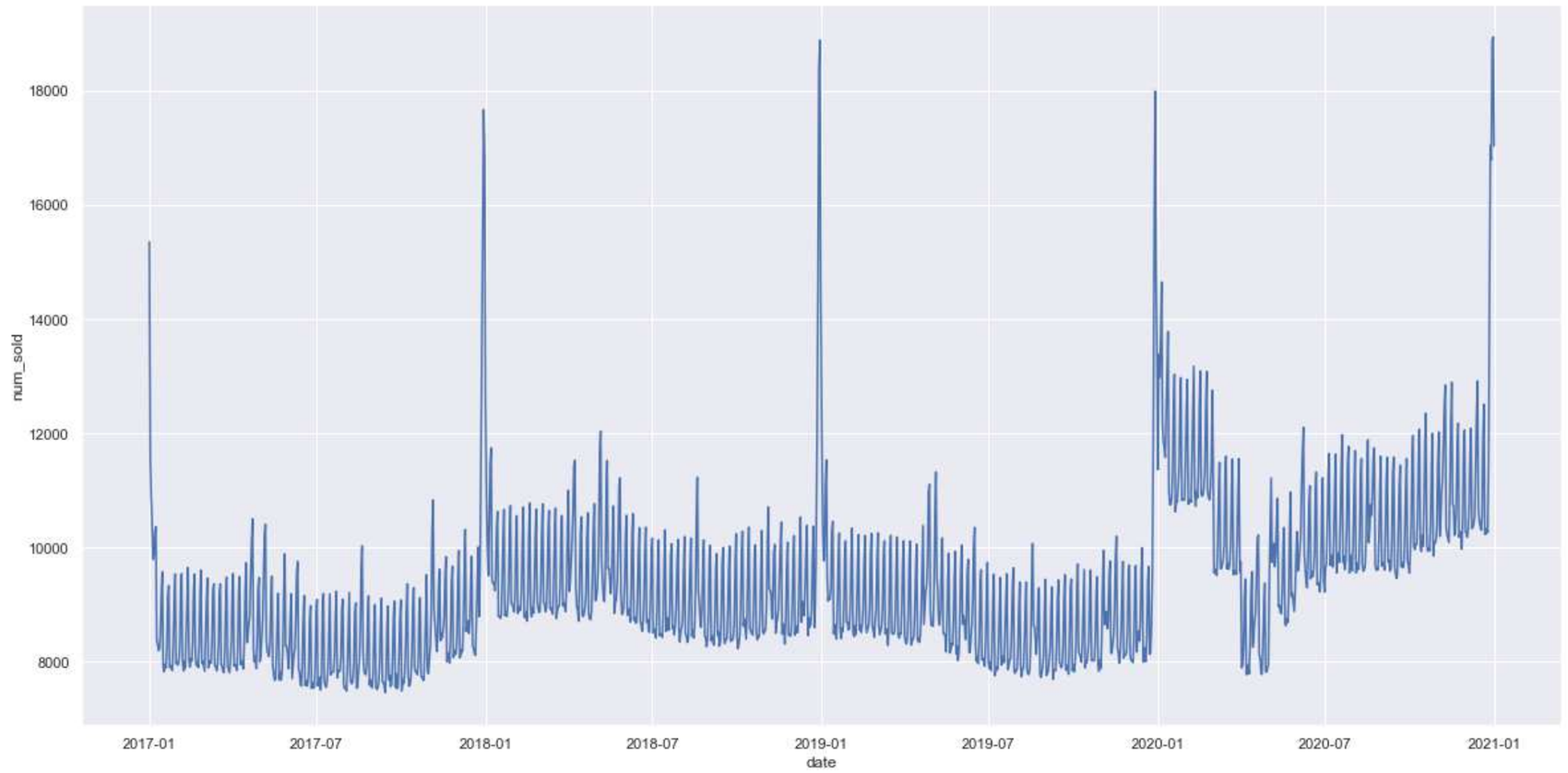

Figure 10: Aggregated time series

Time Feature Extraction

- Seasonal patterns in sales were discovered through ob- servation. We extracted features such as the month of the year, the day of the week, and the day of the year from the time series.

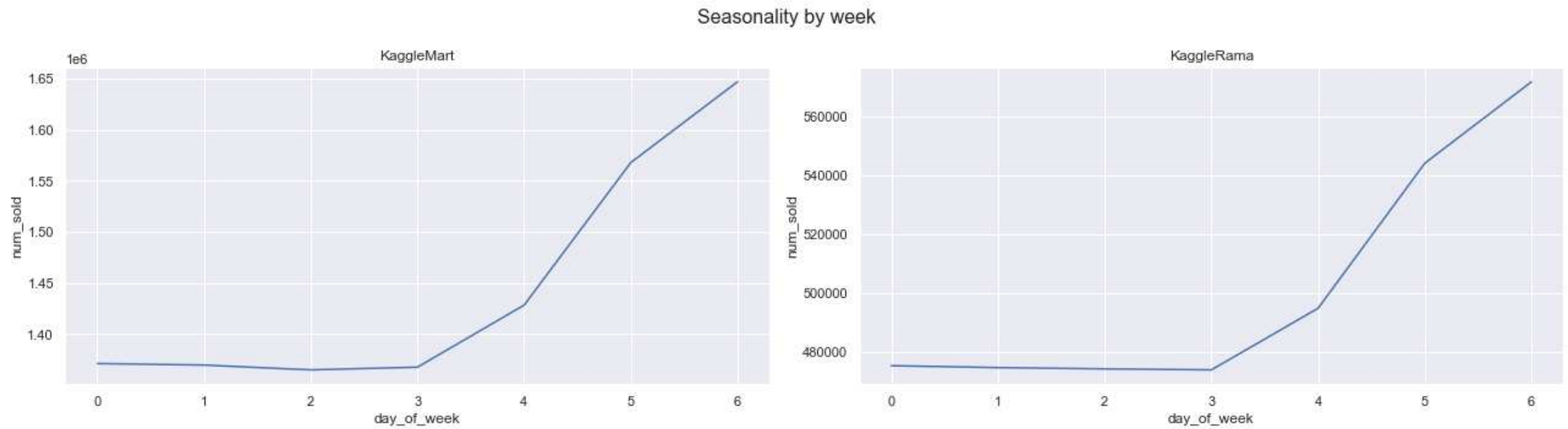

Figure 11: weekday feature

Time Feature Extraction

- We converted the monthly data into numerical values that are suitable for input into a Fourier transform. This allows us to perform a Fourier transform on the monthly data and obtain the coefficients of the various frequency components.
- Also took into account features related to important dates within the year.
- The features of the training set include 23 feature items.

feature	value
month_sin	0.5
month_cos	0.866025
year	2020
important_dates_1	0/1
day_of_week_1	0/1

Model Train

Total Sales Forecasting

- Use Ridge regression from the "linear_model" module to correlate the relationship betweensales and time features , and predict on the test set.
- Train the model with time features as X and sales as y.
- The sales of the test set is predicted according to the time features of the test set.

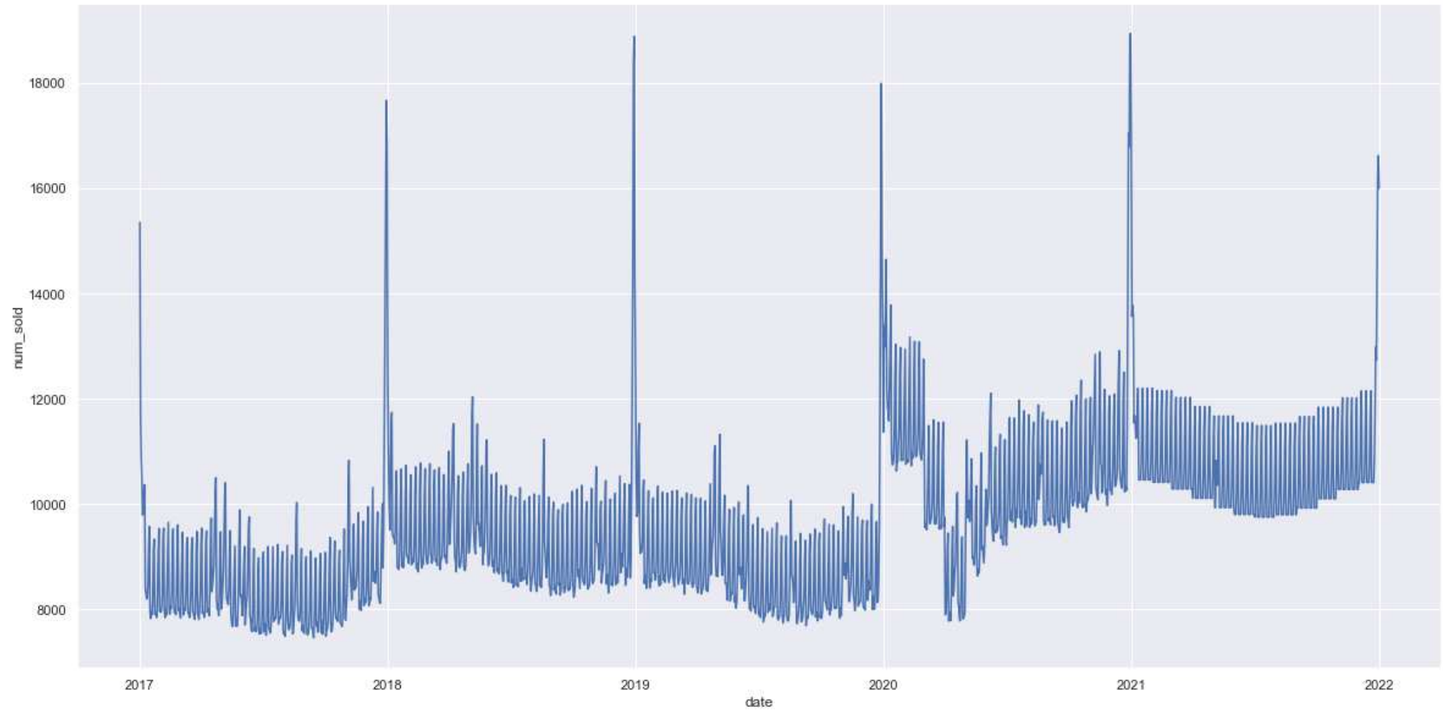

Figure 12: total sales forecasting

Product Ratio Forecast

We found that the proportion of sales for a product has a cyclical variation with a period of two years. Therefore, we assign the daily proportion of sales for each product in 2019 to 2021.

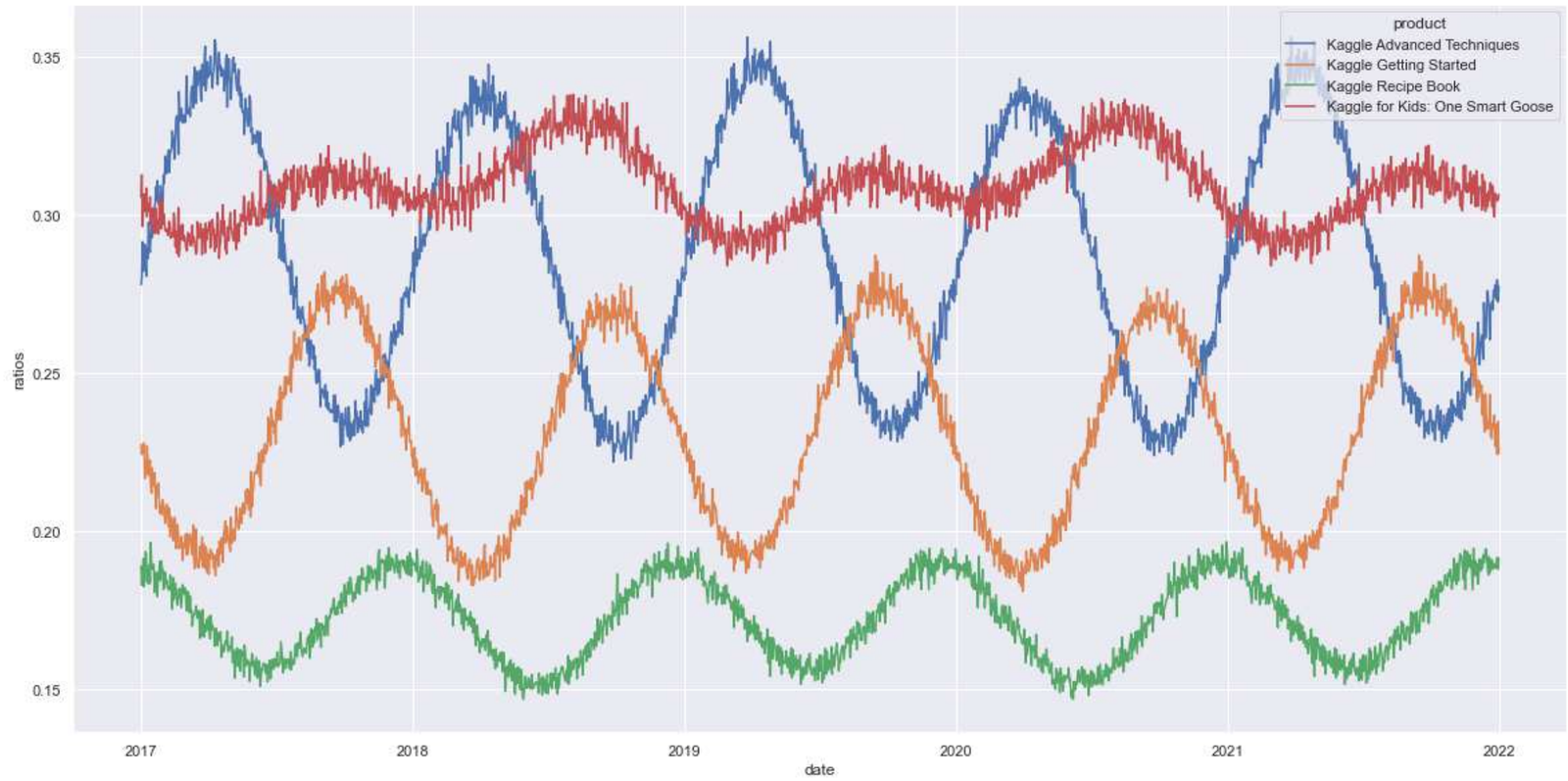

Figure 13: Product Ratio Forecast

Country and Store Ratio Forecast

- We assume that the proportion of sales in countries in 2021 is the same as in 2020. In 2020, the proportion of sales in countries accounts for 1/6,
- The proportion of different stores remains fixed, KaggleMart accounts for 75%, and KaggleRama accounts for 25%

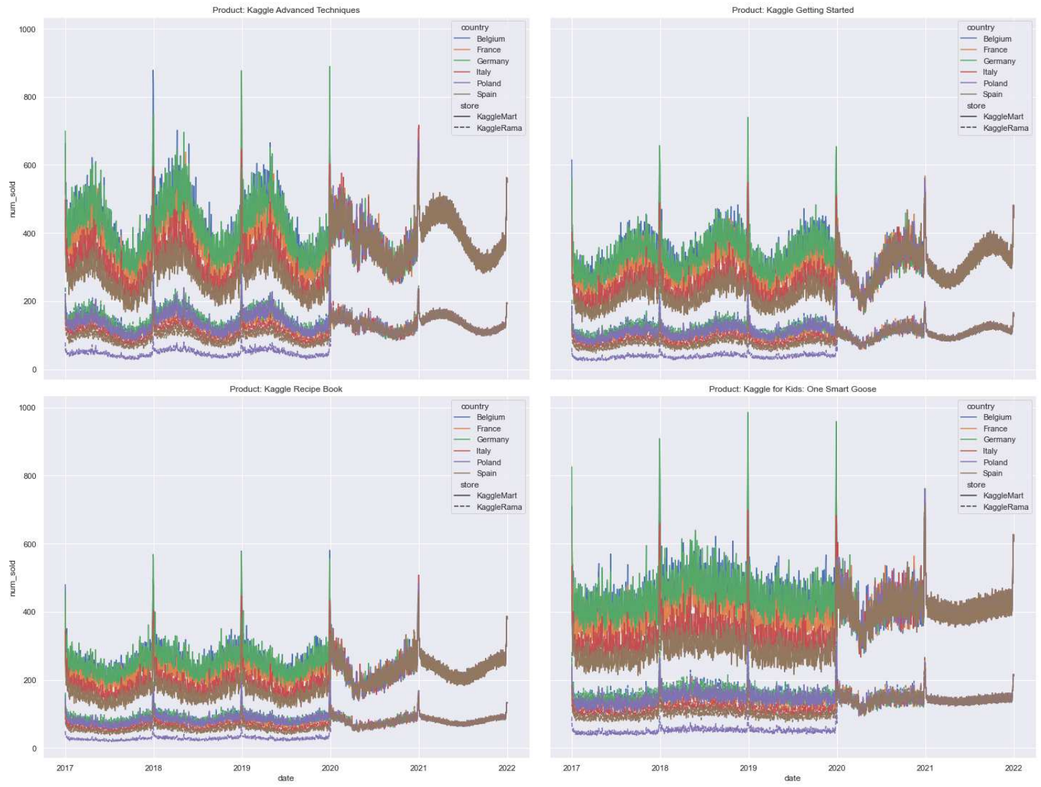

Figure 14: Final Forecasting

Conclusion

Conclusion

- This is a time series forecasting problem that includes complex elements.
- Simplifying the effects of complex factors through analyzing patterns discovered through single factor analysis.
- Use linear regression method to predict the relationship between sales volume and time characteristics.

Questions?

Contact Information

Jiahong Lin

School of Economics and Management

Nanjing University of Science and Technology, China

KKSMI18@163.COM