Sales of Books Forecasting

Jiahong Lin[1],

[1] Nanjing University of Science and Technology, China

Introduction

Time series sales forecasting refers to the process of predicting future sales revenue for a certain period of time. This forecasting is typically based on historical sales data and uses statistical and machine learning algorithms to predict future sales trends. This project is about the time series sales forecast of different commodities in different stores in different countries. This project requires to forecast the sales of six commodities in two stores in four countries in 2021 through the data from 2017 to 2020.

Time Series Aggregating For the prediction of time series with complex factors, data analysis is required to combine the time series under different factors to simplify the prediction target.

Feature extraction The feature set of the training set is formed by extracting the existing features of the time points on the timeline of the past data, and then the future time points are predicted using the time point features. *outlier interpretation* or *object explanation*.

Time Series Aggregating

- The purpose is to eliminate the complex factors caused by different countries, different stores and different products. Simplify the prediction target through data analysis.
- The analysis found that there is a fixed pattern in the proportion of sales of different products in different countries and different stores.

The trend of sales proportion in different countries

The trend of sales proportion in different stores

The trend of sales proportion for different products

Feature Extraction

Through the previous data analysis, it is found that the sales volume of stores and countries has a fixed proportion, while the sales volume of different products has a periodicity. Therefore, it is only necessary to predict

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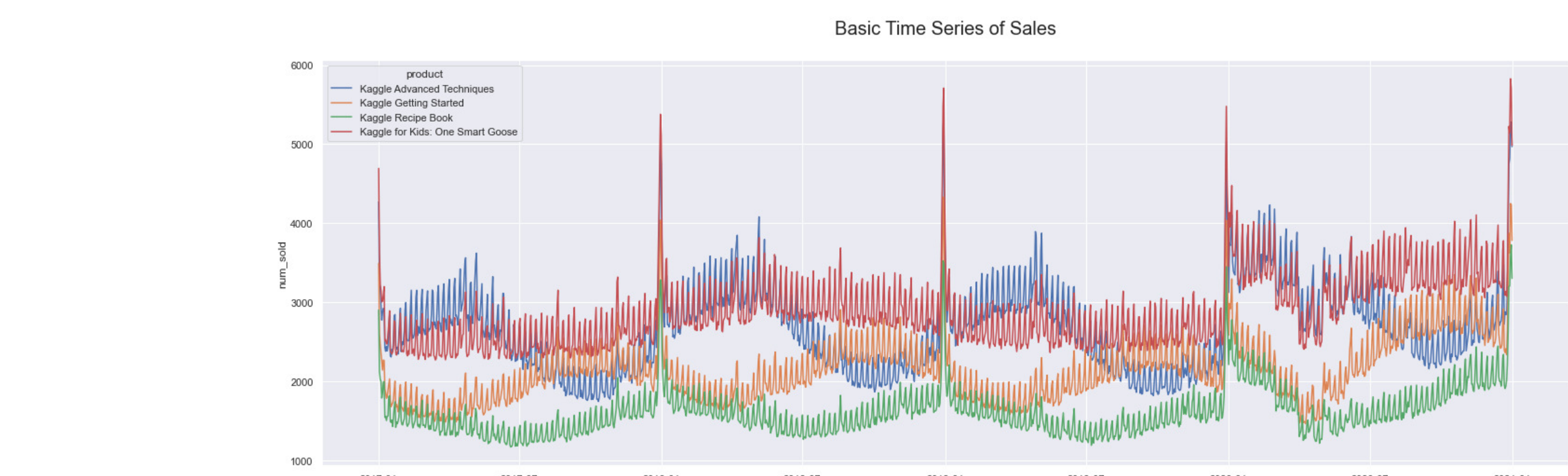
Aggregated Time Series

Time Feature Extraction Seasonal patterns in sales were discovered through observation. 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.

weekday feature

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.

we also took into account features related to the day of the week and important dates within the year.
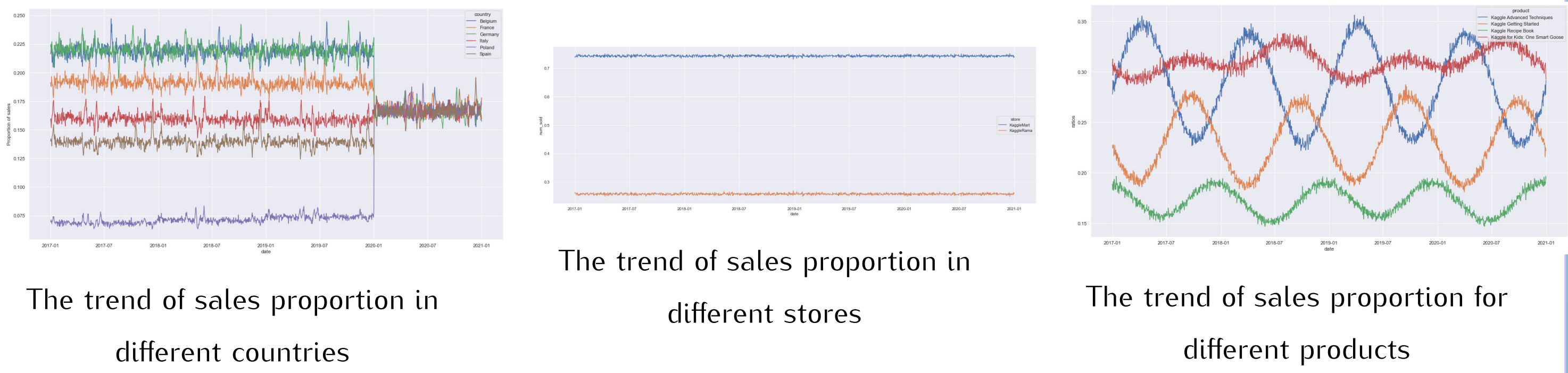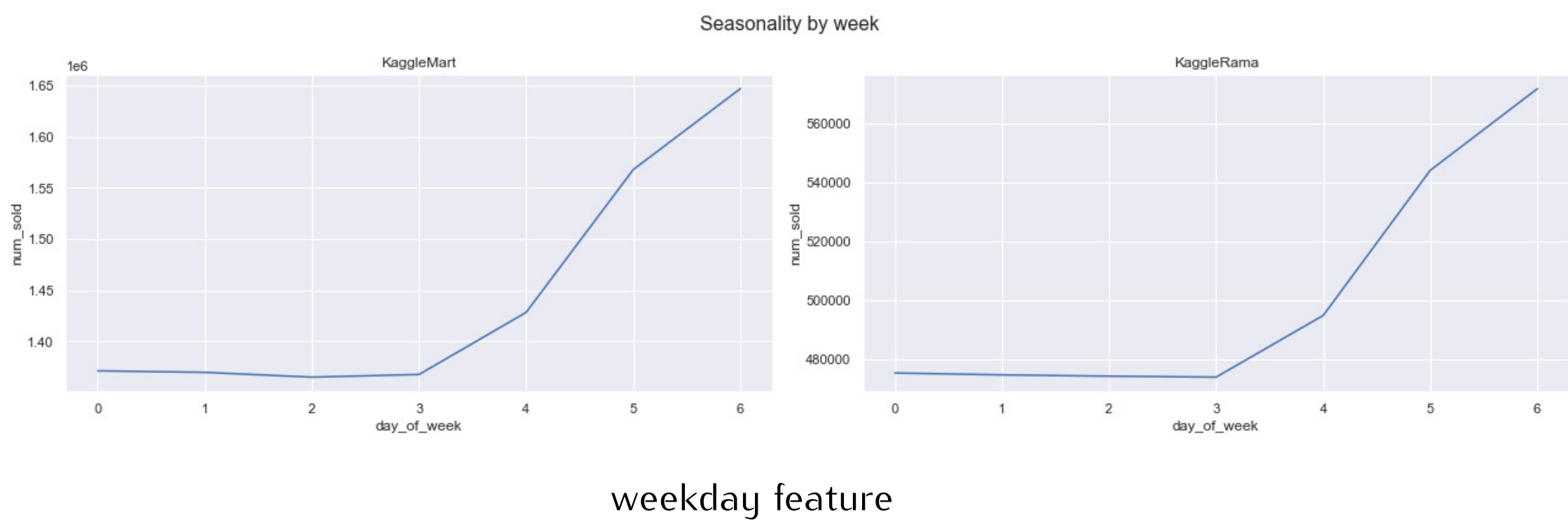
Total Sales Forecasting

We use Ridge regression from the "linear_model" module to correlate the relationship between sales and time features, and predict on the test set.

weekday feature

Specific Sales Forecast

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.
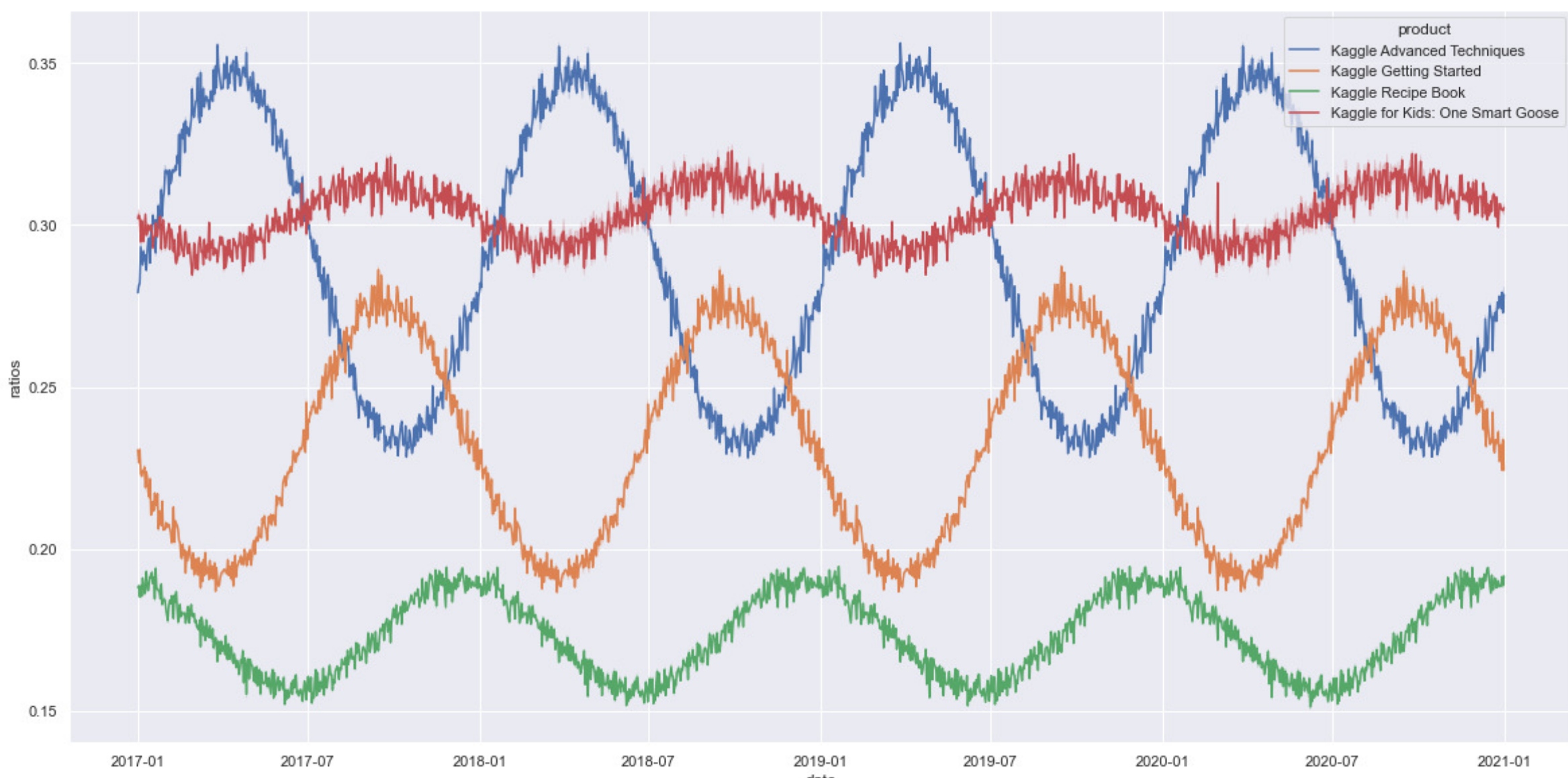
Product Ratio Forecast

Country and Store Ratio Forecast • we assume that the proportion of sales in various countries in 2021 is the same as in 2020. In 2020, the proportion of sales in various countries is the same, each accounting for 1/6.

- The proportion of different stores remains fixed, KaggleMart accounts for 75%, and KaggleRama accounts for 25%.
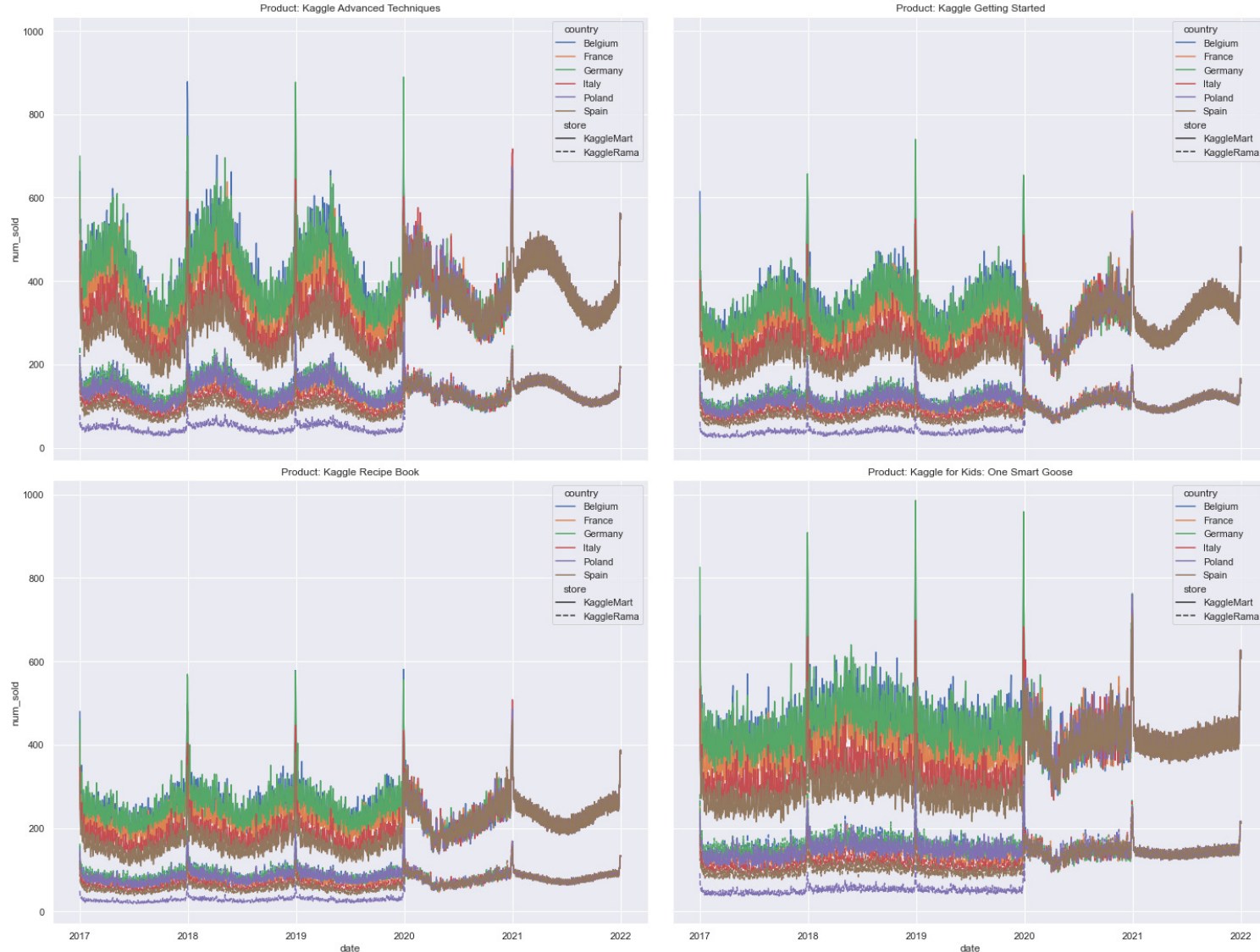
Final Forecasting

Conclusion

Problem Definition This is a time series forecasting problem that includes complex elements.

Strategies Simplifying the effects of complex factors through analyzing patterns discovered through single factor analysis.

Prediction method Use linear regression method to predict the relationship between sales volume and time characteristics.