Forecasting AsiaYo's One Month Ahead Daily Room Occupancy in Different Cities for Supply Preparation

Sz Wei Wu (Sandy)
Cheng Che Liao (John)
Min Sheng Wu (Akira)
Kai Wei Pai (Kelvin)

Business Analytics Using Forecasting   Group 1

EXECUTIVE SUMMARY

In this project, we collaborate with AsiaYo, an online B&B booking platform company headquartered in Taiwan, to work together on solving their business problem by using forecasting methods. One challenge facing AsiaYo is the revenue lost when they are lack of available rooms on holidays or special peak periods. Considering the enterprise level and resource, we find that it will be more affordable and understandable to focus our solution of this business problem on certain popular areas.

By forecasting the room occupancy of popular cities for next month, the operation team in AsiaYo can better prepare for the upcoming demand and have a reasonable reference for making relevant decisions. In the data processing steps, we first aggregate the daily booking order data into daily room occupancy data. To give a more accurate result, we then use the "cumulative number of orders before 15, 30 and 60 days", "Holidays", and "Sakura Season" as our external data in our model. We tried out different models such as exponential smoothing, ARIMA, linear regression, and neural networks. Eventually, we decide to use linear regression with external data as our main model by reason of better errors and simplicity. For benchmark, we use yearly seasonal naive and monthly seasonal naive to compare the performances with our linear model.

We find that there is a big improvement on the accuracy of our result by adding the external data, especially when using the "cumulative number of orders before 15, 30, and 60 days" data. Another interesting point we find is that our result perform better in Japan comparing to Taiwan due to users' different order behaviors. However, the performance on holidays still needs to be improved. For future work, we will try on modeling only holidays and also bring in other possible related data such as website pageviews or organic search.

PROBLEM DEFINITION

AsiaYo is an online B&B booking platform company headquartered in Taipei, Taiwan. Its business mainly focused in Asian-Pacific area such as Taiwan, Japan, Korea, and Thailand etc. One of the business problem facing AsiaYo is that customers turning to other competing platforms when AsiaYo has no enough available rooms. This problem happens critically on the peak periods.

BUSINESS GOAL

In this project, we aim to help AsiaYo preparing for ***supply deficiency in order to prevent consequential revenue losts***. Our main clients are the COO and the operation team who can adjust different operating strategies, for example, outright purchases. We also believe that our forecast solution can help them on having reasonable data to compare rather than simply using data from “last year”.

FORECASTING GOAL

Based on the business goal, it is necessary to know how many rooms will be taken in the future. *Daily forecast* would be more useful since the accommodation demand often occur at "a-few-day" holidays. We also considered two issues that are important in a implementing point of view. First, we want to retain enough lead time for AsiaYo to make operating decisions. Second, it is easier for them to understand and implement if we focus on popular areas. Overall, we set our forecasting goal as to ***forecast one month ahead daily room occupancy in different cities***.

DATA DESCRIPTION

AisaYo provided the booking data which started from January 1st, 2016 to December 31st and for the check-in data started from January 1st, 2016 to August 31st, 2018. Our data contain multiple series of reservation in 39 cities based on the destination within two countries, Taiwan and Japan. Each reservation includes the information about ***order number***, ***order_time***, ***check_in*** and ***check-out***.

order_date	o_no	country	city	check-in	check-out
2016/01/01	2.01601E+11	Taiwan	Taipei City	2016/01/09	2016/01/10
2016/01/01	2.01601E+11	Taiwan	New Taipei City	2016/01/19	2016/01/20
2016/01/01	2.01601E+11	Taiwan	Taipei City	2016/01/03	2016/01/04
2016/01/01	2.01601E+11	Taiwan	Tainan City	2016/01/11	2016/01/12
2016/01/01	2.01601E+11	Taiwan	Nantou County	2016/02/09	2016/02/10
2016/01/01	2.01601E+11	Taiwan	Yilan County	2016/01/15	2016/01/16
2016/01/01	2.01601E+11	Taiwan	Tainan City	2016/01/24	2016/01/25
2016/01/01	2.01601E+11	Taiwan	Yilan County	2016/02/09	2016/02/10
2016/01/01	2.01601E+11	Taiwan	Tainan City	2016/02/09	2016/02/10

DATA EXPLORATION

Based on our forecasting goal, we are going to forecast the daily occupancy; therefore, we do some preprocess to reach the ideal data as input for the following forecast.

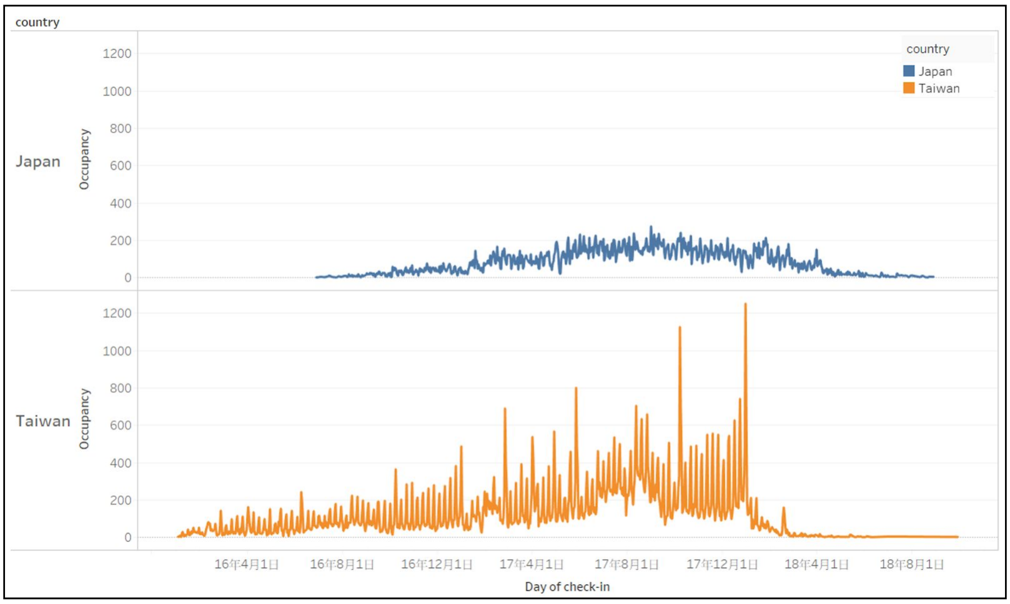


Figure.1 Time plot of occupancy in Japan and Taiwan

The visualization (Appendix.1) reveals how different the order behavior vary from countries. Visitors to Japan tend to book in the early stage and those to Taiwan prefer to order at the near time. By observing this phenomena, we would like to make good use of the order data as one of the predictor to forecast the upcoming occupancy and try different forecast horizon to the performance as well.

1. Derived Variables -
    a. Weekend
    b. Occupancy: Subtract ***Check-out*** by ***Check-in.***
    c. Cumulative number of orders before ***15***, ***30***,***45*** and ***60*** days:
    Subtract **order_data** by ***Check-in*** and cumulate the orders.
2. External Data -
    a. Holiday: For both country and the information come from the website.
    b. Sakura: Exceptional for Japan.

Finally we choose several top cities in each country, five in Taiwan and four in Japan, for the forecasting and select data only in 2017 because there were some policy issue caused the fluctuation in 2016 and uncovered reason for low occupancy in 2018.

city	Date	occupancy	weekend	long_holiday	cum_15	cum_30	cum_45	cum_60
Taipei City	2017/1/1	56	1	1	22	18	8	4
Taipei City	2017/1/2	23	0	1	8	2	2	1
Taipei City	2017/1/3	26	0	0	13	3	2	0
Taipei City	2017/1/4	28	0	0	10	2	1	0
Taipei City	2017/1/5	28	0	0	11	1	1	0
Taipei City	2017/1/6	44	0	0	17	3	2	1
Taipei City	2017/1/7	69	1	0	24	7	3	3
Taipei City	2017/1/8	32	1	0	12	2	1	1
Taipei City	2017/1/9	27	0	0	8	2	1	1

METHODS

We have tried simple methods like exponential smoothing, and more complicated methods such as ARIMA and Neural Networks. Eventually, we found that ***Linear Regression*** performs the best with simplicity. However, the results are highly underestimated; therefore, we seek for the external information for better result.

The final method we use is ***Linear Regression*** with ***external data***. The external data includes ***cumulative number of orders before n days*** and ***holidays***. For Taiwan, we only use Taiwanese holiday; for Japan, we use Japanese holiday and the "Sakura season". For different forecast horizons, we use different *n* of the "cumulative number of orders before n days" data depending on the horizons.

EVALUATION

1. Error increase as forecast horizon rise

Figure 2 shows the actual (blue), benchmark (black), and forecast time series for daily room occupancy in Taipei. Overall, all forecast methods' performance are better than seasonal naïve. Moreover, we can observe that as forecast horizon increase, error (Appendix 2) tend to rise, which means when we want to forecast long-term ahead, the uncertainty will also increase at the same time.

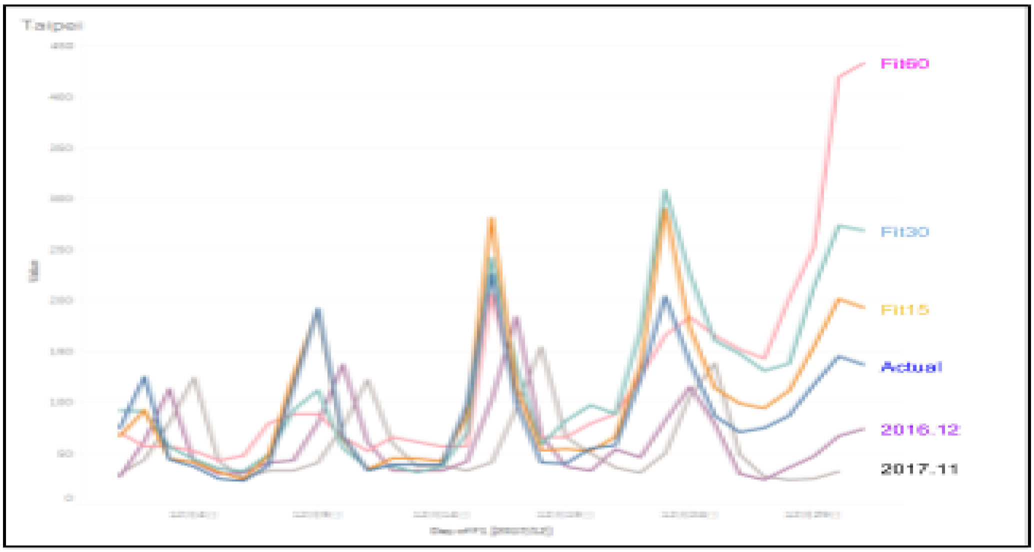


Figure.2 Time plot of forecasted occupancy value in Taipei (validation period)

2. Error in holiday is higher than that of in non-holiday

We found that the error of holiday is larger than that of non-holiday (Appendix 3). As we can see in Figure 3, there are some large overestimate errors on Holidays that we need to be aware of. We believe that people tend to make a reservation early on holiday might led to overestimate, and this is the reason why error is higher in holiday.

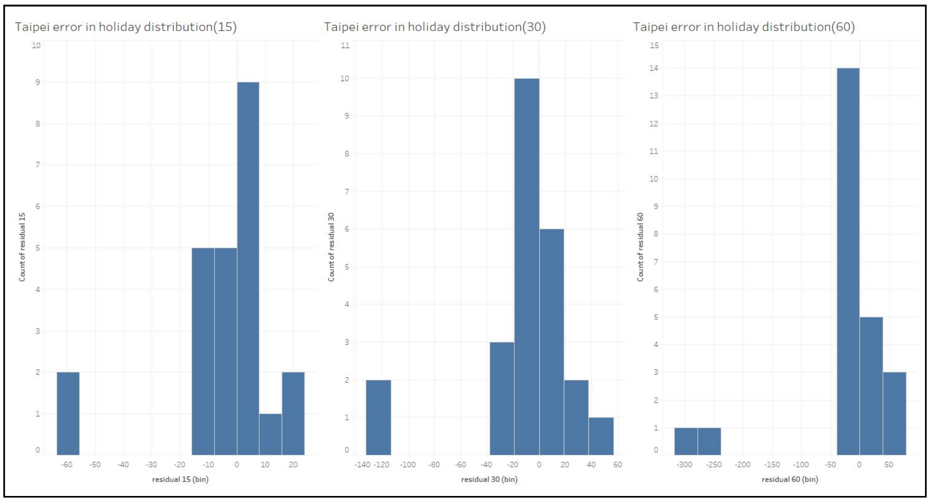


Figure.3 Validation errors distribution chart of Holidays

FUTURE FORECAST

For future forecast, due to the feedback from both the Asiayo data team and class, we will add the data collected from the website into analysis.

The forecast of 2018 is shown below.

	Taipei	Kaoshiung	Tainan	Taichung	Yilan	Fukuoka	Tokyo	Kyoto	Osaka
2018/1/1	71	28	37	46	26	10	46	6	87
2018/1/2	75	15	23	51	8	7	51	2	78
2018/1/3	72	16	19	40	9	4	40	2	58
2018/1/4	58	16	23	55	8	5	55	2	57
2018/1/5	53	18	28	55	17	7	55	3	58
2018/1/6	71	19	37	88	20	5	88	4	53
2018/1/7	55	16	19	50	9	5	50	5	52
2018/1/8	48	11	23	50	9	2	50	6	50
2018/1/9	58	15	23	40	25	2	40	7	47
2018/1/10	41	15	23	45	13	2	45	5	56

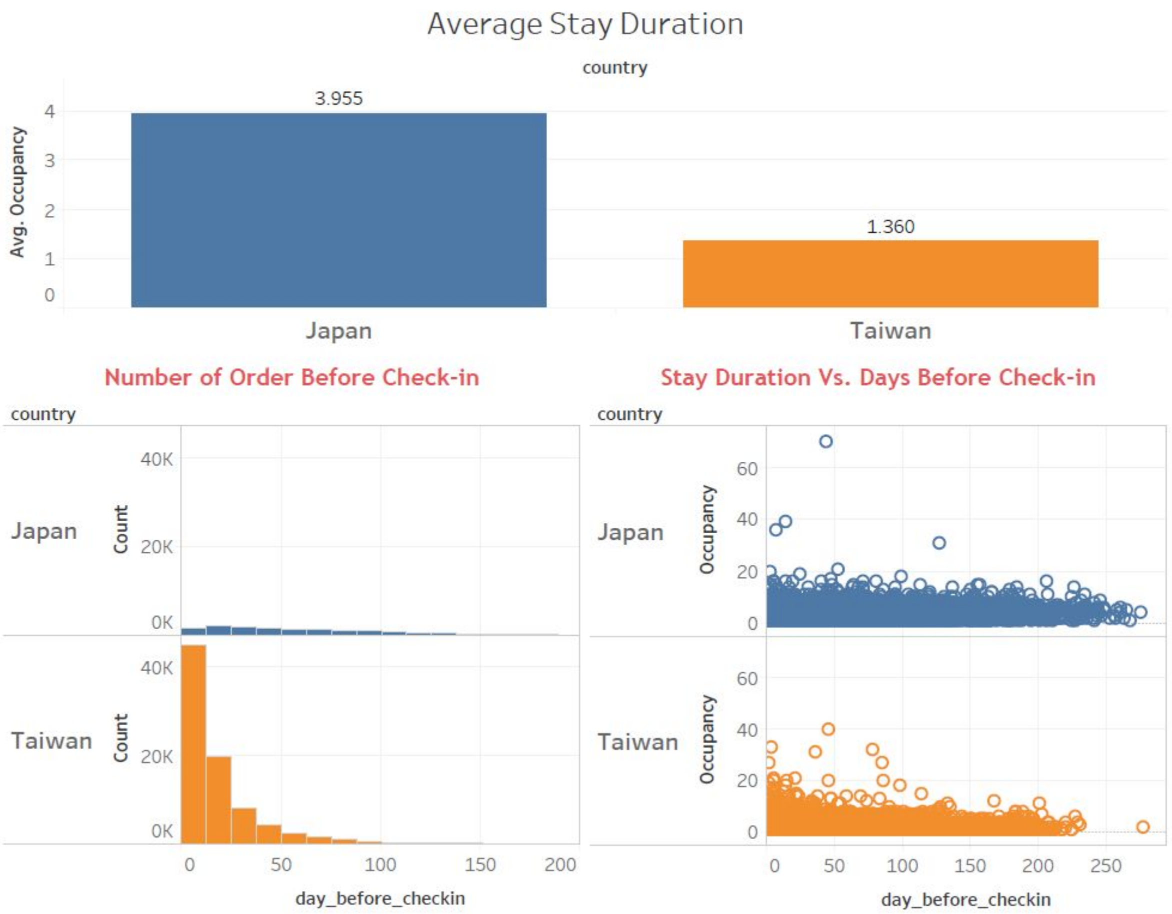
CONCLUSION

Based on our model for forecasting ***1 month ahead daily room occupancy*** can mostly capture the pattern in various region, especially in Japan due to different booking behaviors. However, the model forecast on *holidays are not as good as on not-holiday*. This is a really important issue for company to prepare enough rooms for customer. We will try to deal with this problem by building a new model for holiday to reduce noise come from ordinary day.

To sum up, every member in our team thought that we learned a lot from this course. People from various background can help us in different perspectives by using certain domain knowledge. For instance, business ones can help to define business problems and pick useful attributes. On the other hand, engineer can apply different techniques to find a best way for achieving business goal. It’s this course that teach us a really important lesson called teamwork.

APPENDIX

Appendix.1 Combination of Visualization



Appendix.2 RMSE Table of different cities in validation period (Taiwan)

	Taipei	Yilian	Tainan	Taichung	Kaohsiung
RMSE 15	27.98	12.21	9.66	25.04	7.14
RMSE 30	55.97	25.87	29.95	83.13	10.50
RMSE 60	88.15	22.19	44.12	45.37	13.77

Appendix.3 RMSE Table of Holidays and Not-Holidays in validation period

	Holiday	Not-holiday
RMSE 15	18.43	12.96
RMSE 30	41.78	22.01
RMSE 60	86.84	29.14

Appendix.4 Validation errors distribution chart of Non-Holidays