NTHU Business Analytics Using forecasting

Project Report

Forecasting the Quantity of Tourists Volume in Northeast and Yilan Coast National Scenic Area of to better allocate Human Resource

Project Group 4

102078518 AnYu,Luo

102078505 ChienMin,Kao

102078466 Xi Wang

102077503 Daisin,Li

Advisor: Prof. Galit Shmueli

Executive summary

Our business goal is try to better allocate human resource in Northeast and Yilan Coast National Scenic Area Due to the tourism industry booming in Taiwan the past five years, the domestic and foreign tourists are increasing. we select ocean park, hiking trail, rock climbing area, diving space and camp site. These five locations are directly under different units. Try to organize the staffs between these five locations. Our data comes from Tourism Bureau Executive Information System, and the data is measure the monthly tourist volume in Northeast and Yilan Coast National Scenic Area from 2005 to 2013.

Because there are trend and seasonality in most series, to choose the proper forecast methods, first we chose the seasonal naïve to be our benchmark. Second, we also use the Holt-Winters exponential smoothing to be our model which allows the level, trend, and seasonality patterns to change over time. The third method we used is multiple linear regression, and we used month to be our dummy. We drew the forecast and actual charts and residual charts, and we compared the residual chart of these three methods. And we can see in this table, we also compared the RMSE in these three methods.

There are some limitation in this forecasting project. First, we can't find out the difference of tourists volume between weekday and weekend due to lacking of daily data. Second, because there is a big difference of tourist volume between the peak season and low season in Fulong ocean park, it is difficult to forecast the tourist volume accurately. Third, we can't find out the relation of tourist amount between these five places using the data in hand. Finally, it is difficult to give recommendation about human resource allocation in detail because we don't have real human resource arrangement data.

And our recommendations for the Longdong South Ocean Park and Caoling Historic Trail system and Fulong Ocean Park, can use Holt-Winters. On the other hand, in the Longdongwan Coast Park) and Longmen Camp Site, it can simply use seasonal naive will cause the best result.

Business Goal

Our business goal is try to better allocate human resource in Northeast and Yilan Coast National Scenic Area Due to the tourism industry booming in Taiwan the past five years, the domestic and foreign tourists are increasing. The Northeast and Yilan Coast National Scenic Area faces some challenge situations, including parking space, train carrying capacity, temporary labor, and traffic jam problem. We want to balance the tourist quantity and service quality.

We focus on the 5 specific locations in the Northeast and Yilan Coast National Scenic Area, we select ocean park, hiking trail, rock climbing area, diving space and camp site. These five locations are directly under different units. We create the new vision to the Northeast and Yilan Coast National Scenic Area. Try to organize the staffs between these five locations. The data we collected is monthly data, it is limit to forecast the other topic. However there are difference between weekday and weekend. So we can't arrange the human resource dynamically.

Forecasting Goal

We forecast the monthly volume of tourists in Northeast and Yilan Coast National Scenic Area in 2014 and our forecast horizon is 12, means that we want to use the historical data to predict the volume of tourist of next month. Besides, we set it's a prospective forecast and forward-looking goal.

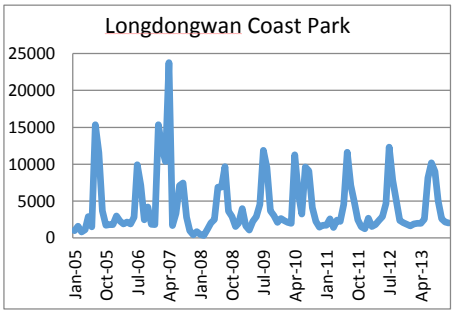
We define the tourists' volume in March 2005 as a unit comparing to other monthly data. We did not get the original labor requirement in these five locations, so we will present the number meaning how many times bigger than the volume in March 2005. We define the forecasting volume in March 2014 as the basic labor requirement. Basic on the trend and seasonal quantity we will organize the temporary labor. Arrange the human resource and do appropriate adjustments. Try to make the most effective human resource using, and reduce the labor costs.

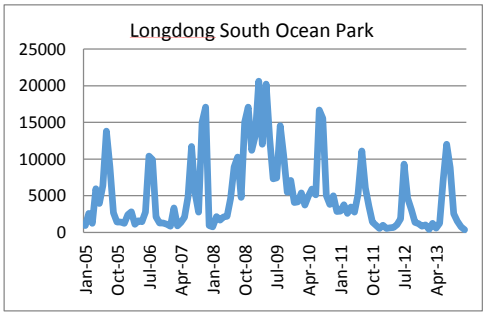
Data

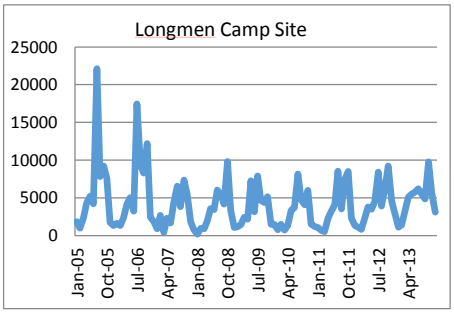
Our data comes from Tourism Bureau Executive Information System, and the data is measure the monthly tourist volume in Northeast and Yilan Coast National Scenic Area from 2005 to 2013. Figure1 is the raw data of this area, and for better visualization, we took log of the data of those five places.

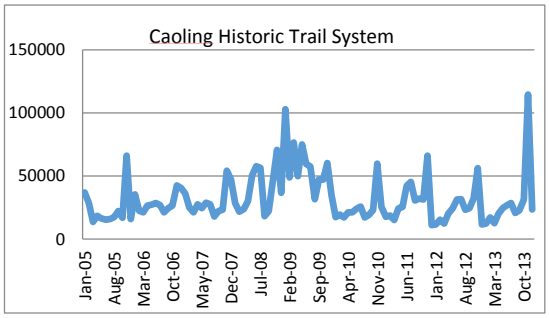
We found out that there're missing values in Jan-05 to May-05 and Dec-06 of the

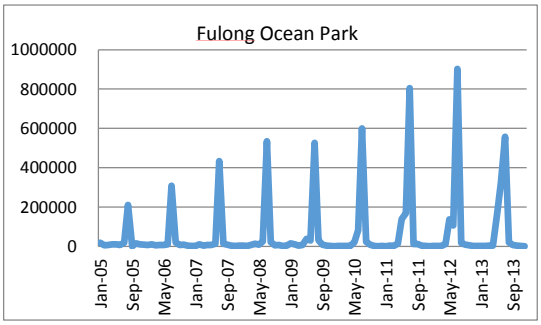
Longdongwan Coast Park series, so we used the value in Jan-04 to May-04 and Dec-05 to impute the missing data. Besides, we took the average of the last month data and the next month to impute several missing ones with a clear upward trend. Then, we drew the charts of these five series to see their characteristic and try to know if there are trend or seasonality.











From these five charts, we found that there are both trend and seasonality in the series of Longdongwan Coast Park, Longdong South Ocean Park, Longmen Camp Site and Fulong Ocean Park, but no trend or seasonality in the Caoling Historic Trail System.

Forecasting period details

Training data: JAN 2005 to DEC 2012 (108 months)

Validation data: JAN 2013 to DEC 2013 (12 months)

Forecast Horizon: JAN 2014 to DEC 2014 (12 months)

Models considered

1. Seasonal Naïve forecast,
2. Linear Regression (t, month as dummy)
3. Holt Winter method without trend

Because there are trend and seasonality in most series, to choose the proper forecast methods, first we chose the seasonal naïve to be our benchmark. Second, we also use the Holt-Winters exponential smoothing to be our model which allows the level, trend, and seasonality patterns to change over time. The third method we used is multiple linear regression, and we used month to be our dummy.

For the Fulong Ocean Park, there is a big difference of tourists amount between peak season and low season every year due to the Ho-hai-yan Rock Festival. Therefore, we separating the series into the peak periods and the low periods and model the two series separately.

Performance evaluation

We drew the forecast and actual charts and residual charts, and we compared the residual chart of these three methods(From figure2a to figure2e). And we can see in table 1, we also compared the RMSE in these three methods. We found in the Longdongwan Coast Park and Longmen Camp Site, use seasonal naïve has the smallest value, in the Longdong South Ocean Park, Caoling Historic Trail System (including Yuanwangkeng Riverside Park), and Fulong Ocean Park, use holt-winters has the smallest value.

Under forecasting will result in decline in service quality. Considering the real work content, we will arrange the daily staff schedule to maintain the service quality. Over forecasting will result in rising cost, so we are more concerned about over forecasting. We found that in every series, there are over forecasting in some months, and Caoling Historic Trail System has the highest degree of over forecasting, which should be enhanced the forecasting accuracy specifically.

Recommendations

- In the Longdong South Ocean Park, Caoling Historic Trail system, and Fulong Ocean Park, use **Holt-Winters** will cause the best result.
- In the Longdongwan Coast Park and Longmen Camp Site, can simply use **seasonal naive** will cause the best result.
- Forecasts can be rolled forward on a monthly basis to include fresh data.
- We can compare to the basic unit- number of tourists in March 2005(the lowest number in our series). Based on the results of the previous forecasting, setting up the permanent and temporary labor requirements. And the forecast numbers are used to set up the labor in the location (decide how many full time employee and part time employee need to hire in low-season and high-season).

Limitation

Due to lacking of daily data, we can't find out the difference of tourists volume between weekday and weekend. Because there is a big difference of tourist volume between the peak season and low season in Fulong ocean park, it is difficult to forecast the tourist volume accurately. Using the data we have, we can't find out the relation of tourist amount between these five places. Because we don't have the real human resource arrangement data of these five places, we can't give recommendation about human resource allocation in detail.

Appendix

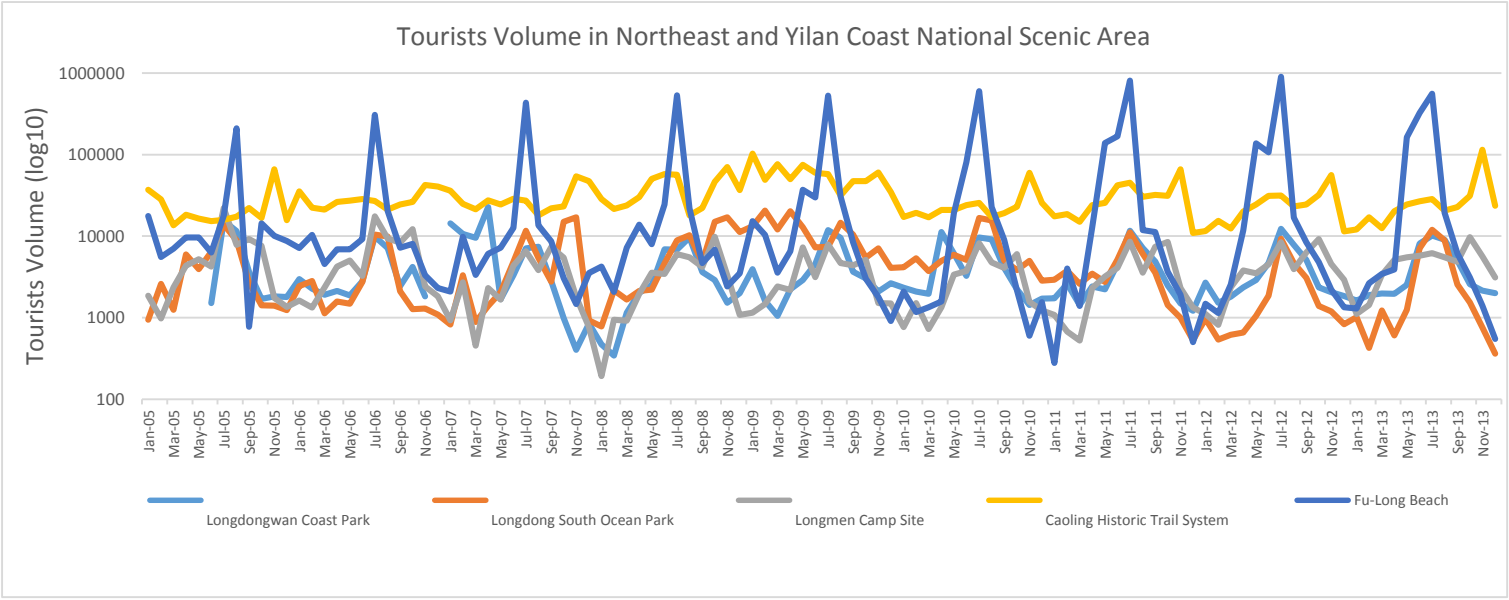


Figure 1 raw data of this area

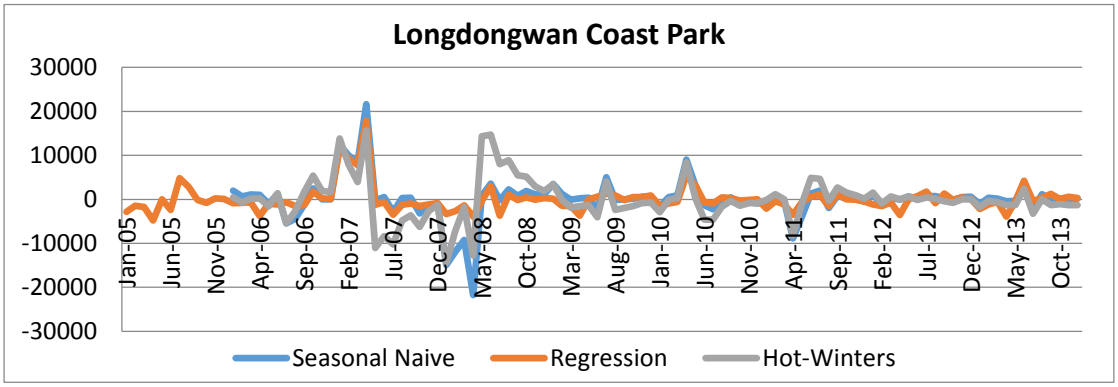


Figure2a The residual in three different methods for Longdongwan Coast Park

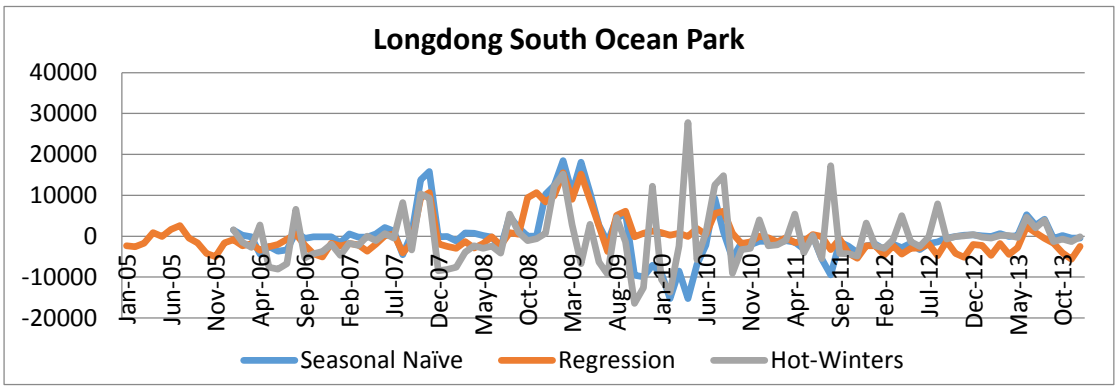


Figure2b The residual in three different methods for Longdong South Ocean Park

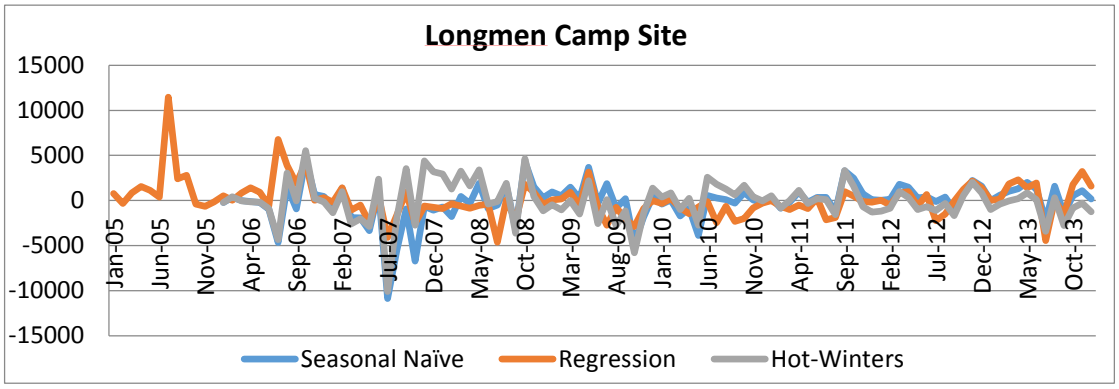


Figure2c The residual in three different methods for Longmen Camp Site

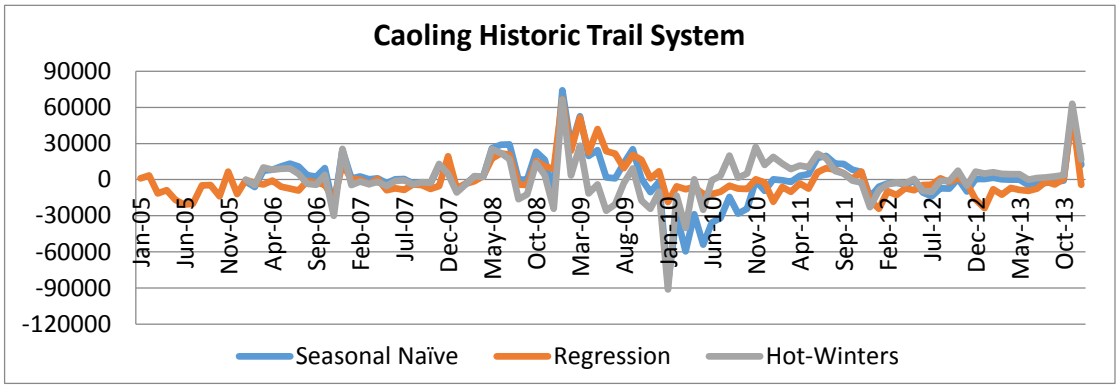


Figure2d The residual in three different methods for Caoling Historic Trail System

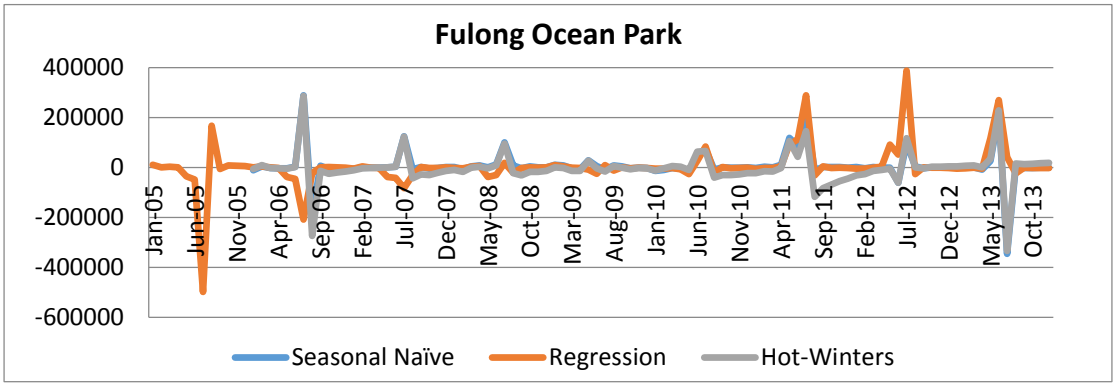


Figure2e The residual in three different methods for Fulong Ocean Park

Table 1 The RMSE for three different method in five series

	Longdongwan Coast Park	Longdong South Ocean Park	Longmen Camp Site	Caoling Historic Trail system	Fulong Ocean Park
Seasonal Naïve	1285.862	2111.999	1284.579	17365.791	118222.566
Linear regression	1871.812	3189.572	2097.878	18653.67	80349.4(high/low season)
Holt-Winters	1578.34 (Additive)	1954.886 (Additive)	1407.918 (Additive)	15749.95 (No-Trend)	11765.62 (Multiple)