Forecasting restaurant sales using data from iChef, weather forecasts, and holiday information

DOES FORECAST WEATHER IMPROVE SALES FORECASTS?

Nicholas Danks, Isaac Martinez, Mahsa Ashouri, Paul Rivera
NATIONAL TSING HUA UNIVERSITY

Executive Summary

One of the largest challenges facing small and medium restaurants is that of competing against big chains with better information infrastructure – an advantage that allows the big chains to plan resource allocation more precisely based on demand and other factors. iChef provides this platform for small restaurants "making enterprise level technologies affordable and understandable for small restaurants". As Taiwan is a subtropical country, often affected by heavy rains and tropical storms, there is a question of if and how weather can affect sales; and how iChef can help its clients to mitigate or capitalize on this challenge.

By more accurately forecasting sales (considering externalities), managers are able to better manage their resources to avoid losses and maximize revenue. By integrating data from iChef Taiwan and World Weather Online – and adding other external information such as holidays - into one database, we were able to propose a model that includes this information on weather and holidays and forecasts daily sales. We benchmark the target model against other simpler models like seasonal naïve, exponential smoothing and another linear regression without external information to compare its performance and consider if the cost of deploying a more complex predictive model is worthwhile for the business.

We find that at the level of daily sales per restaurant, there is little additional predictive accuracy gained by the inclusion of holiday and weather data. However, when we consider a single three hour meal time (such as 5 – 8pm dinner time), we find that the inclusion of weather and holiday data can provide us with higher predictive accuracy and improved forecasts over simpler methods. Since the collection and maintenance of data such as weather forecasts and annual holidays is extremely convenient and easy to conduct, we advise that this data be collected and considered for the prediction of daily sales per restaurant.

Problem description

Business goal:

A common challenge facing every business manager is that of the correct allocation of resources - whether it be personnel or inventory - that is needed to meet actual demand. This is especially important for restaurants, where a mismatch between resource availability can lead to economic losses caused either by excessive stockpiling – this means assets and expenses not generating revenue - or by under allocation - which can lead to service delivery failure, loss of customers and hence reduced revenue. Managers can learn empirically and affect customer patterns by the products they offer, but there are variables that are beyond the control of a manager. Such variables could have a dramatic effect on customer behavior, one such important factor being weather.

Analytics goal:

Different restaurant have different peak hours, where most of the revenue is generated, depending on the type of food they serve. Daily data aggregation is not very helpful since it mixes peak hours with low hours. More details can be observed if the data is aggregated three-hourly because it better shows meal time patterns. Based on the business goal, it is needed to know in which of these three-hourly segments is weather useful for forecasting revenue. Many predictive models were developed and compared.

Data description

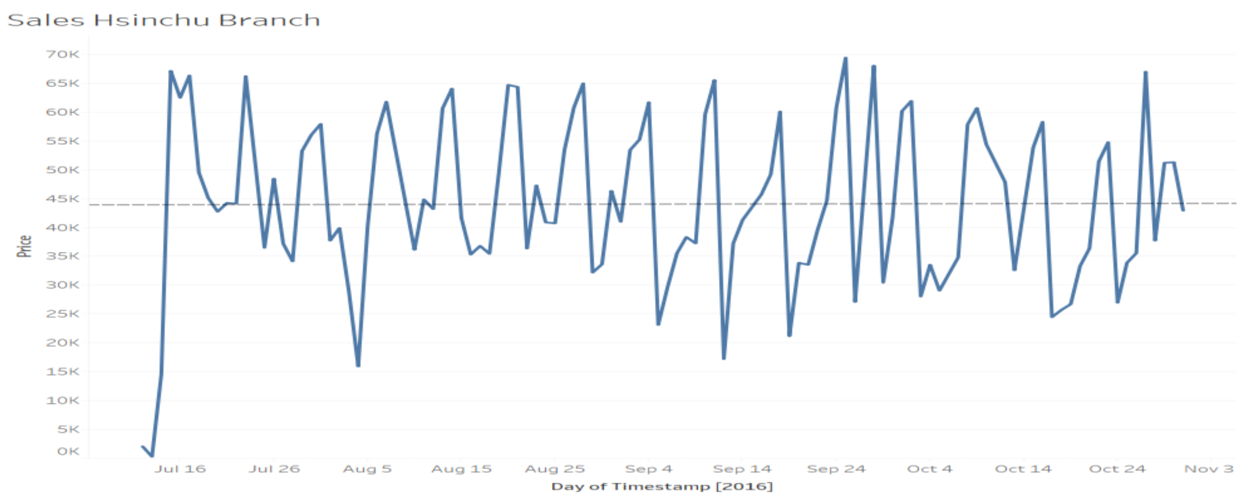
We obtained restaurant data from iChef, which included fields such as time stamp, price and people – just to mention some of the most relevant. This data had entries from mid-July to mid-October and it recorded every sale in the POS for six different restaurants. Weather forecast data was obtained from World Weather Online API, which was given three-hourly and included plenty of weather indicators but the relevant one for this model was precipitation quantity (liters per square meter) and classification (sunny, cloudy, light rain, Heavy rain, etc.)

iChef Raw Data:

invoice_uuid	item_name	item_uuid	people	type	dining	price	timestamp
000026EA-B2E6-41C8-9A99-6C5	羅馬-松露野菇	48a0422a-9bf0-4fe2-ba6e-768e6c6e239c	1	combo	takeout	380	2016/6/15 09:27:30
000026EA-B2E6-41C8-9A99-6C5	拿-經典辣味燻	3090bcb2-16e9-4971-8087-b3f92ee6343d	1	combo	takeout	220	2016/6/15 09:27:30
000026EA-B2E6-41C8-9A99-6C5	拿-經典瑪格	5ebb6673-0086-4b8a-a052-cdb3424ee3c3	1	combo	takeout	180	2016/6/15 09:27:30
000026EA-B2E6-41C8-9A99-6C5	電話	6785d383-5a26-4133-ae11-1e9c1bd4462b	1	item	takeout	0	2016/6/15 09:27:30
000026EA-B2E6-41C8-9A99-6C5	羅馬-義式果香	9ba372f0-038f-4b3b-9afa-833abe6bdfe0	1	combo	takeout	360	2016/6/15 09:27:30
000026EA-B2E6-41C8-9A99-6C5	拿-堤諾先生	57525b6c-4086-4bc1-afa4-aa3def92eab4	1	combo	takeout	200	2016/6/15 09:27:30
000026EA-B2E6-41C8-9A99-6C5	羅馬-西西里蒜	76381357-b595-4836-a35a-013da575d847	1	combo	takeout	400	2016/6/15 09:27:30
000037B7-5F4E-4196-867B-6D0	酥炸酒醋蘑菇	f703cbf4-5c90-4958-b6e1-a1d8aa97723f	2	combo	indoor	99	2016/5/13 13:44:21
000037B7-5F4E-4196-867B-6D0	羅馬-堤諾先生	88a36403-6c9b-4beb-b1b9-7ea40463061e	2	combo	indoor	320	2016/5/13 13:44:21
000037B7-5F4E-4196-867B-6D0	(I)水蜜桃冰茶	2e886631-5ac0-413a-9d8a-06bf136a0f69	2	combo	indoor	70	2016/5/13 13:44:21

Word Weather Online Raw Data:

date	astronomy/0/sunrise	astronomy/0/sunset	astronomy	astronomy	maxtempC	maxtempF	mintempC	mintempF	uvIndex	hourly/0/t	hourly/0/t
3/1/2016	6:19 AM	5:58 PM	No moonr	10:50 AM	21	69	16	60	0	0	13
3/2/2016	6:18 AM	5:59 PM	12:18 AM	11:35 AM	24	75	19	65	0	0	16
3/3/2016	6:17 AM	5:59 PM	1:10 AM	12:22 PM	22	72	20	68	0	0	20
3/4/2016	6:16 AM	6:00 PM	2:02 AM	1:15 PM	25	77	20	68	0	0	19
3/5/2016	6:15 AM	6:00 PM	2:53 AM	2:10 PM	25	76	21	70	0	0	21
3/6/2016	6:14 AM	6:01 PM	3:44 AM	3:10 PM	24	75	21	70	0	0	20
3/7/2016	6:13 AM	6:01 PM	4:33 AM	4:12 PM	25	77	22	71	0	0	20
3/8/2016	6:12 AM	6:02 PM	5:21 AM	5:15 PM	25	78	22	72	0	0	21
3/9/2016	6:11 AM	6:02 PM	6:07 AM	6:19 PM	18	64	17	62	0	0	21
3/10/2016	6:10 AM	6:03 PM	6:53 AM	7:25 PM	17	62	13	56	0	0	15



Brief data preparation details

The iChef data was aggregated per invoice, and then aggregated into three hourly periods that match daily meal times (Breakfast, lunch, dinner and late night). Then the data from World Weather Online (WWO) was matched with these three hourly periods. We also merged with the dataset of National holidays (including typhoon days) for another external variable.

We selected 2 main restaurants, one in Taipei and one in Hsinchu. The time periods displayed double seasonality (both Daily and Weekly). The weather external variable was collected by API call from the WWO API and then transformed into binary variables, calculated using the precipitation data from WWO. Outliers were calculated from extreme values on the data. All this process was automated in a Python programming language script in preparation for receiving more data from iChef, since we had two batches of data and aim to maintain high automation of the forecasting process.

Forecasting solution:

In the initial data exploration phase, many forecasting models were considered and applied on this data, such as seasonal naïve forecasts, exponential smoothing, linear regressions with and without external variables and neural networks in order to compare their predictive performance. It was found that there were certain days that couldn't be adequately forecast by any of the models. It was decided to call these days "outliers". Since the data obtained only covered three months, it wasn't possible to observe yearly patterns, however, it was possible for us to observe monthly and weekly seasonality and trend. It is clear that although much of the signal was captured by the models, there remains a lot of noise and potential for domain experts to identify further signals.

Finally, three final models were tried and compared, these are

1. Exponential Smoothing
2. Linear Regression Model 1 with Trend, Seasonality and the Outlier variables

3. Linear Regression Model 2 with Trend, Seasonality, Outliers, Public Holiday, Rainy and Typhoon variables

The exponential smoothing model will serve as a benchmark for 3 reasons. First, it is cheap to compute and highly automatable. Secondly, it does not consider external variables and thirdly it has a relatively high level of predictive accuracy. For these reasons, it serves as a good benchmark and comparison for our Regression models. The second model, Regression model without weather data, was created to serve as a comparative for the inclusion of further variables but still excludes weather information. If the Regression including weather performs no better than the regular regression, we can conclude that weather forecast data has no value in the forecasting of the restaurant sales.

The reason for the third model is that Trend and Season can explain normal calendar days, but public holidays are expected irregular days with a different demand and “Rainy” and "Typhoon" are the “Unpredictable” variable. By including these additional variables and comparing performance to the comparative Regression and the Exponential Smoothing model, we can establish whether the weather data is useful for forecasting restaurant sales. As was previously mentioned, outliers are the days for which there was very high or very low sales with no apparent reason. These are probably caused by the owner’s decision to close the restaurant or hold a private event.

The three different models were run on the 4 different three-hourly time series (breakfast, lunch, dinner and late night) for each of the 2 restaurants selected, and also on a daily aggregated data. The Findings for the restaurant in Hsinchu were are as follows:

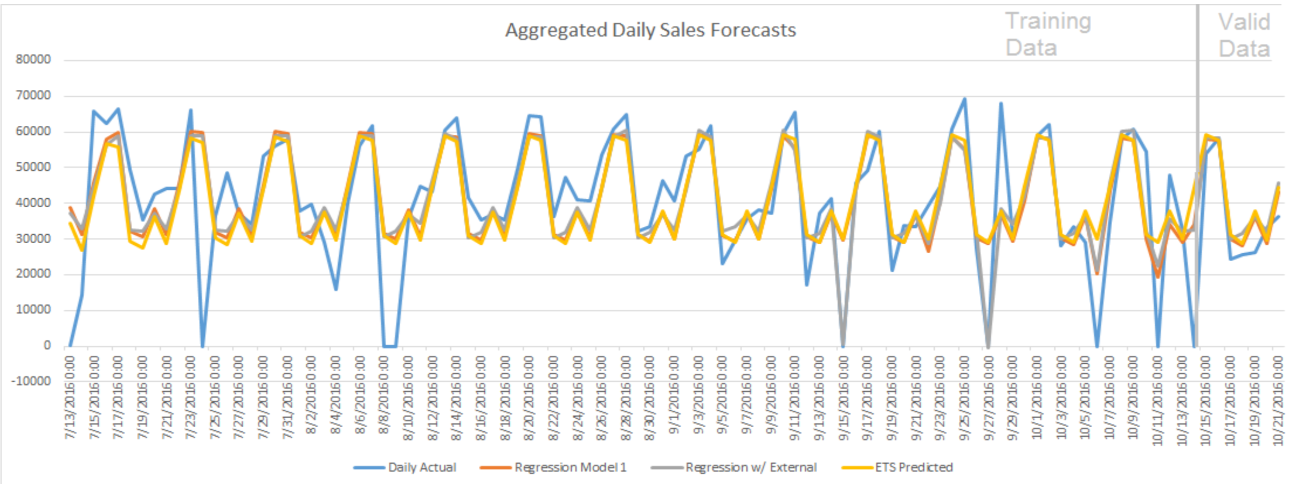
- On the daily aggregated data, the best performing model is the ETS as shown on the following table:

	Regression Model 1 **No Weather Data**	**Regression Model 2** **Include Weather**	**ETS Model (Benchmark)**
RMSE	**6,551.33**	**6,975.45**	**6,455.36**
MAPE	**18.32%**	**18.94%**	**18.09%**

- On the three hourly data for the dinner time period (5pm – 8pm), Regression Model 2 (Including Weather) performs the best.

	Regression Model 1 **No Weather Data**	**Regression Model 2** **Include Weather**	**ETS Model (Benchmark)**
RMSE	**1,675.92**	**1,315.72**	**1,773.50**
MAPE	**16.26%**	**18.52%**	**17.56%**

Finally, we present the model predictions for the daily aggregated sales for the Hsinchu restaurant for all 5 models in order to illustrate the predictive performance:



Conclusions

On the whole, we found that forecast weather can have an impact on forecast sales per restaurant. However, counter-intuitively, for some time slots the weather contributes positively to the forecast while in others it contributes negatively. This is an interesting finding and might be due to some phenomena interpretable by domain knowledge. Referring to Tables 1 and 2 of the appendix, we see that the Rainy variable has a positive co-efficient of NTD 1,942.33. Additionally, when forecasts are aggregated to the daily level, we see little or no predictive power improvement over the simpler benchmark ETS model.

Advantages

Since restaurants get most of their revenue from certain peak hours, management can maximize this revenue by applying the linear regression model that includes external information, for better forecast sales given certain conditions, including weather. At the same time, weather is a good indicator because it's easy to access and maintain.

Limitations

Aggregate Scale dilutes the performance improvement obtained by the model including external information and at the same time shows more extreme error values than the regression models.

Operational Recommendations

This research shows that there is value in including weather information in sales forecasts for restaurants. We propose that iChef collect weather forecast weather information on their services and use this information for predictive purposes at the meal time level in order to give restaurant managers an improved tool to take better planning decision. However, for daily sales the ETS model has the best predictive performance at the lowest computational and operational cost.

Appendix

Table 1 Second Model: Regression without weather information – Hsinchu 5pm

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	13778.75	1840.34	7.487	6.03e-11	***
trend	-10.81	22.43	-0.482	0.63096	
season2	-5374.40	2169.89	-2.477	0.01524	*
season3	1116.06	2170.24	0.514	0.60841	
season4	7476.51	2211.21	3.381	0.00109	**
season5	7914.85	2220.01	3.565	0.00060	***
season6	-2897.03	2220.69	-1.305	0.19556	
season7	-4851.44	2211.79	-2.193	0.03101	*
train.outlier.ts	-2284.09	2740.13	-0.834	0.40686	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Model parameters for the Linear regression model without weather information

Table 2 Third Model: Regression with weather information – Hsinchu 5pm

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	13023.820	1841.434	7.073	4.64e-10	***
trend	-11.851	23.238	-0.510	0.611434	
season2	-4236.080	2180.308	-1.943	0.055462	.
season3	1257.744	2153.445	0.584	0.560782	
season4	7564.563	2190.020	3.454	0.000876	***
season5	8419.189	2212.790	3.805	0.000272	***
season6	-2840.355	2178.437	-1.304	0.195933	
season7	-3835.738	2206.494	-1.738	0.085896	.
train.rainy.ts	1942.334	1355.800	1.433	0.155770	
train.outlier.ts	-2706.793	2714.691	-0.997	0.321654	
train.holiday.ts	6.131	2203.445	0.003	0.997787	
train.typhoon.ts	-10088.829	4319.223	-2.336	0.021945	*

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Model parameters for the Linear regression model with weather information