Quarterly Profit Projection

A Time-series Analysis With FBprophet

11/4/2020

Background

MW.com is a fictional publicly traded e-commerce company that sells men's clothing across three websites: *suits.mw.com*, *shoes.mw.com* and *shirts.mw.com*. Assume that by collecting data from *mw.com*'s three websites, we've been able to come up with very accurate estimates for what Go.com will report for 1Q15 revenue. The data set contains the daily revenue data that collected from *mw.com*'s three websites for 1Q15 and the company reported data for the last five quarters for both revenue and profit.

Goal

Based on the 1Q15 revenue data and the historical data to come up with an estimate for what *mw.com* will report as their 1q15 profit.

Tools

The majority part of the analysis is done in python with pandas, numpy and fbprophet

Initial Analysis

From the 'Historical Data' tab, we can calculate the **Profit Margin** (profit by revenue) of the past five quarters as below:

Profit Margin	Q4 2013	Q1 2014	Q2 2014	Q3 2014	Q4 2014
suits.mw.com	10.000%	10.000%	10.000%	10.000%	10.000%
shoes.mw.com	0.500%	0.500%	0.500%	0.500%	0.500%
shirts.mw.com	2.000%	2.000%	2.000%	2.000%	2.000%
Total	2.004%	1.845%	1.818%	1.864%	1.634%

Table 1: Historical Profit Margin

Although, the **Total Profit Margin** varies among different quarters, the **Individual Profit Margin** (10% for *suits.mw.com*, 0.5% for *shirts.mw.com*, and 2% for *shoes.mw.com*) remains constant. Thus, the model of a good estimate of the profit could be:

$$\hat{Profit} = 10\% * \hat{R_{suits}} + 0.5\% * \hat{R_{shirts}} + 2\% * \hat{R_{shoes}} \quad (1)$$

$\hat{R_{cars}}$ represents the estimated total revenue of *suits.mw.com* in 1Q2015. Same for $\hat{R_{shoes}}$ and $\hat{R_{shirts}}$. Import the **Sales Data** into Jupiter notebook. Split the data set into three subset dataframe `suits`, `shoes` and `shirts`. Make the `Date` as the index and `Revenue` as the only column in each dataframe. Each dataframe only contains the daily revenue data for the corresponding website. Table 1 is the summary of the daily revenue. The long distances between the **min** and **max**, high **std** is common for *mw.com*. The irregularity is the missing revenue data in `shoes`. The missing data can also be seen in figure 1.

	suits	shirts	shoes
count	90.00	90.00	90.00
mean	186194.72	361441.59	52626.09
std	158793.62	351150.91	65608.55
min	3790.80	6999.62	0.00
25%	75800.40	118879.35	9312.00
50%	111863.40	214161.58	31199.52
75%	285843.60	605310.78	53204.16
max	571988.40	1395072.32	239376.48

Table 2: The summary of the daily revenue

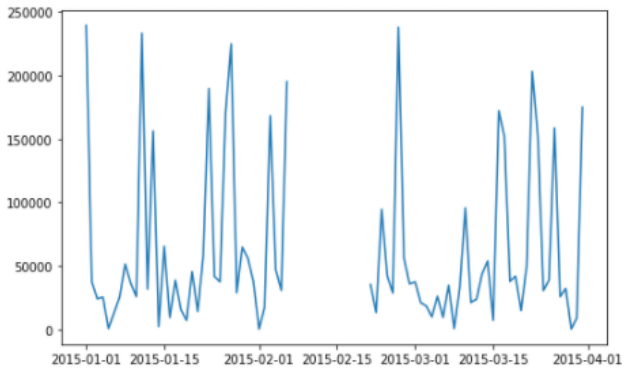

Figure 1: Daily Revenue of shoes.mw.com in 1Q2015

I ran the following command:

```
pd.date_range(start = '1/1/2015', end = '3/31/2015').difference(suits.index)
```

to examine the value of `Date`. In `suits` and `shirts`, the value of `Date` is distinct. In `shoes`, I find the missing revenue value on the following dates from '2015-02-07' to '2015-02-20'.

Predict Revenue

Without any irregualrity found in `suits` and `shirts`, it is reasonable to use the sum of the daily revenue in 1Q2015 as the estimated total revenue

for *suits.mw.com* and *shoes.mw.com*. The remaining problem is to estimate the revenue during the period of 2015-02-07 to 2015-02-20 from *shirts.mw.com*. Figure 1, the daily revenue of *shoes.mw.com* could be treated as time-series data. Train the data set with `fbprophet` and predict on the missing dates, the predicted revenue during the missing period and its range is as below:

	ds	yhat	yhat_lower	yhat_upper
0	2015-02-07	37311.902044	-44201.601974	122405.112780
1	2015-02-08	52258.709559	-22095.681487	138220.195713
2	2015-02-09	64589.951678	-14213.299162	143706.393815
3	2015-02-10	102027.043724	22904.649464	176705.679671
4	2015-02-11	25417.996280	-55716.866984	106097.404794
5	2015-02-12	90642.861609	16753.737413	169618.572469
6	2015-02-13	61158.218270	-23360.203937	141707.442990
7	2015-02-14	37173.488718	-41043.626289	113626.207892
8	2015-02-15	52120.296233	-28147.477974	131077.065365
9	2015-02-16	64451.538352	-16582.986107	138868.173701
10	2015-02-17	101888.630399	26012.857013	188411.665541
11	2015-02-18	25279.582955	-54613.251860	107934.468438
12	2015-02-19	90504.448284	8242.825837	165088.630196
13	2015-02-20	61019.804944	-14264.476716	142968.057967

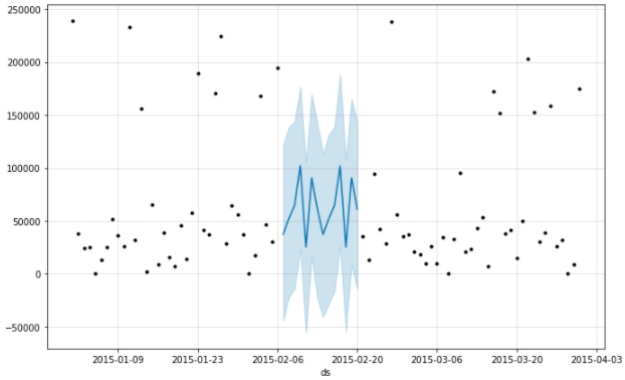

Figure 2: Daily Revenue of boats.go.com in 1Q2015 with prediction from 2015-02-07 to 2015-02-20

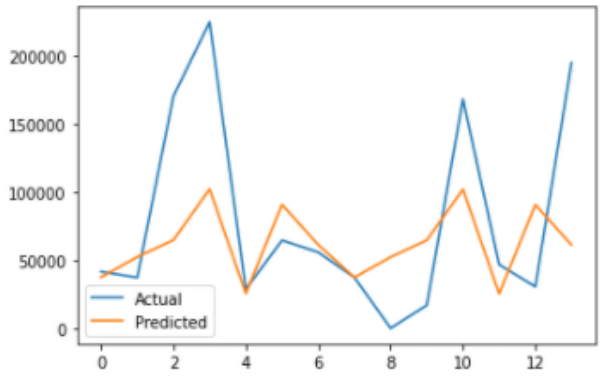

Figure 3: Prediction vs actual revenue in the 14 days prior to 2015-02-07

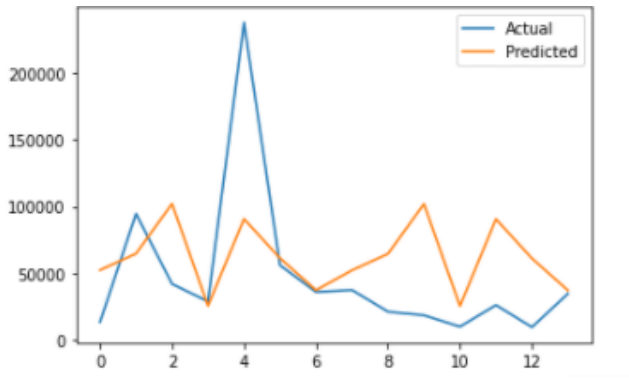

Figure 4: Prediction vs actual revenue in the 14 days after to 2015-02-20

To examine the accuracy of the `fbprophet` model, test the model on the 14 days before 2015-02-07 and 14 days after 2015-02-20. The **mean absolute error** is 47,363.687 and 40,152.591 respectively. And the comparison of the actual and predict revenue line is showed in Figure 3 and 4. Both plots show that the `fbprophet` model fits the data well. Especially in the 14 days prior to 2015-02-07. `fbprophet` model predicts the trends accurately. Thus, I predict revenue of *shoes.mw.com* for 1Q2015 to be [5,602,191.68 - 47,363.69, 5,602,191.68 + 40,152.59]. And the final estimation for the 1Q2015 profit of *shoes.mw.com* is between **$ 111,096.56** to **$ 112,864.89**.

Conclusions

With the predicted revenue of *shoes.mw.com*, recorded revenue of *suits.mw.com* and recorded revenue of *shirts.mw.com*, and the **Profit Margin** stays constant. The estimate profit of *mw.com* in 1Q 2015 is **$ 1,949,498** to **$ 1,951,248**. The breakdown of the projected profit is as figure 6.

Profit	Q4 2013	Floor	Cap
suits.mw.com	$1,675,752.48	$1,675,752.48	$1,675,752.48
shirts.mw.com	$162,648.71	$162,648.71	$162,648.71
shoes.mw.com	$112,043.83	$111,096.58	$12,846.89
Total	2.004%	$1,949,497.75	$1,951,248.08

Figure 5: go.com 1Q2015 Profit projection breakdown

More to consider

The revenue in 1Q2015 from *suits.mw.com* increased about 200% comparing to 4Q2014. It would be worthy to investigate the cause of such increase and further evaluate if the **Profit Margin** of *shoes.mw.com* would remain at 10% in a significant larger scope. The analysis of predicting the missing daily revenue lack of considerations in potential outliers and influential data points. Obtain information on whether there was special promotion on certain dates would be significant to further improve the accuracy of the prediction.