YipitData - Groupon Analysis

04/05/2019

This report is an analysis of alternative data for Groupon. The report is divided into two parts: first is a sell-buy recommendation based on estimates of Groupon's billings for Quarter four of 2013, and the second is a solution for removing Groupon's duplicated deals in Belgium for March 2015.

Part 1

First, we checked for anomalies in each deal by observing whether either the units sold or the billings were zero. No such anomalies were found. Then, we analyzed deals where the billings or the units sold were a negative value. Even though this might initially seem like an anomaly, it was not because negative billings or units sold define a refund. We started the next anomaly test by splitting the data set into three segments as local, goods and travel deals. Then, we aggregated the data by days. Later, we checked if there were new deals every day between October 1st, 2013 and December 31st, 2013 (dates encompassed by Q4 2013). In the goods segment, every day in the period of interest had new deals, from this we determined that there were no anomalies in the goods segment of the data. In the travel segment, we found that there were 18 days in the period of interest without new deals. In the local segment, we found 11 days without new deals. The difference between days without deals in the travel and the goods segments was that while the travel days were sporadic, local days were consecutive between October 20th, 2013 and October 30th, 2013. Therefore, we decided not to treat the travel days as an error but to treat the local days as missing data which would be a systematic error. Hence, we needed to estimate the billings and the units sold for these 11 days.

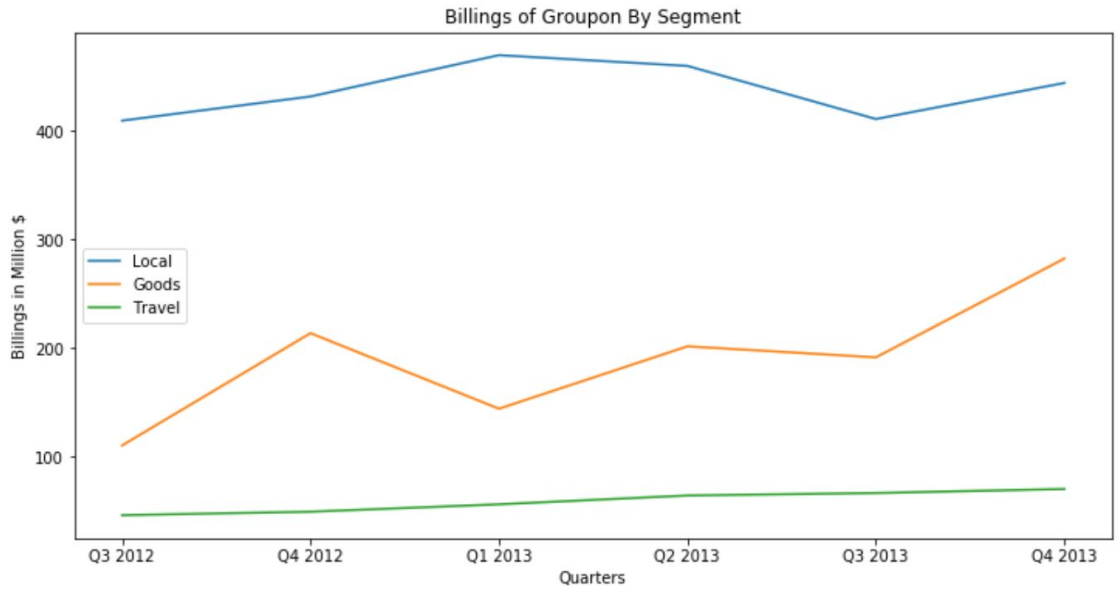
In estimating the billings and units sold in the local segment for 11 days without new deals, first, we took the mean of the billings and units sold and filled in the missing days with these values. Although this is an easy way to roughly estimate the missing data in those days, it is an inaccurate estimation. This is because deals that started before the dates of interest could have had less or more sales or could have been a factor of the seasonality of the deals. So we decided to build a linear regression model to estimate the data of the missing days. After building the linear regression model, we got an estimation with an average accuracy of only 50%. This percentage was not convincing, which made us build other models. Ada Boost, Random Forrest, Decision Tree models with 85% training and 15% testing data gave us an accuracy of 78%, 76%, and 73% respectively. We determined that Ada Boost was the most useful testing method and we estimated the billing first. We tuned the model's number of estimators and learning rate features by using Grid Search. This yielded an 85% accuracy. After finding this model that was the most accurate estimation for billings, we used Ada Boost and Grid Search to estimate the units sold which gave us an accuracy of 70%.

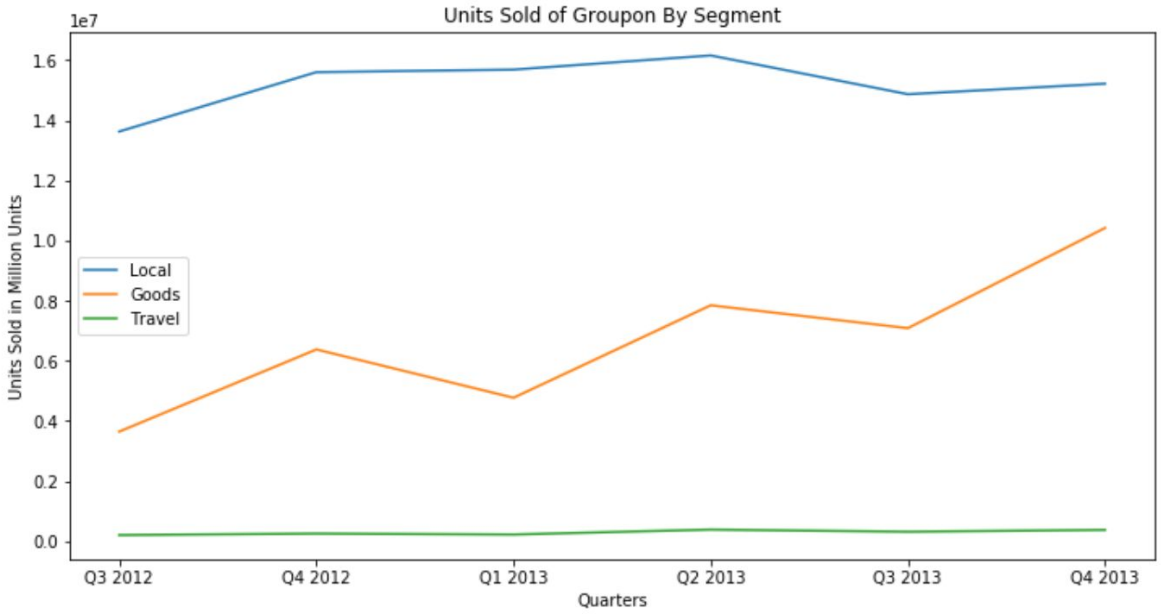
Assuming that all the billings and units sold from the deals in our data set happened during Quater four 2013, we calculated our estimation by segments by summing up all the values for billings and units sold separately. We estimated the number of new deals by counting the deals that had a start date between October 1st, 2013 and December 31st, 2013. Then we compared our estimates of billings to other

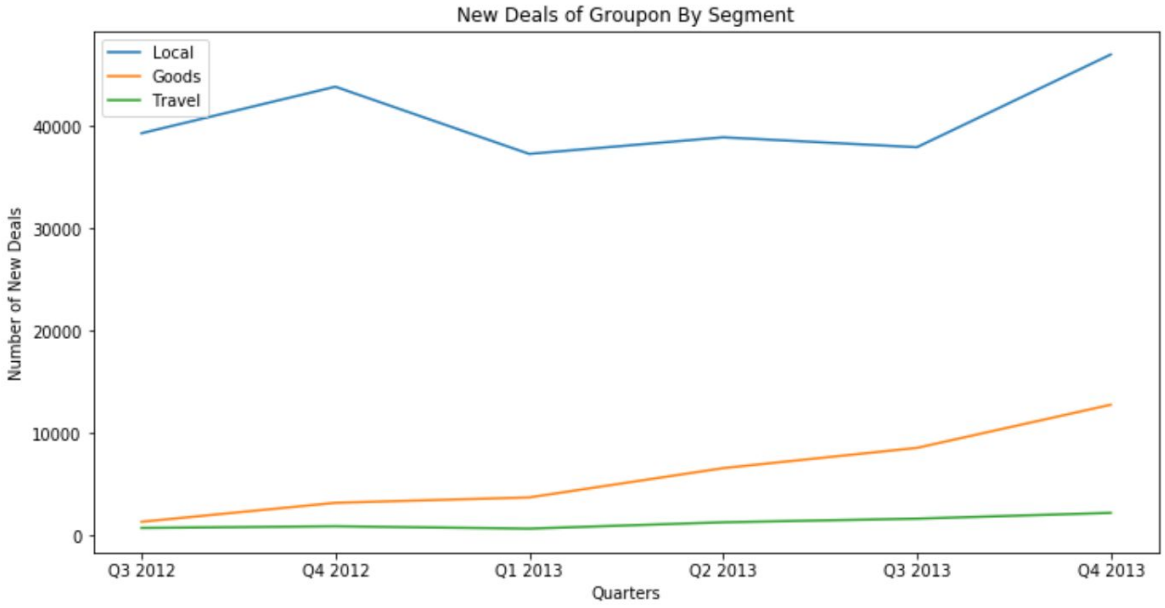
financial firms such as JP Morgan, Morgan Stanley, and Deutsche Bank. These values can be seen in the graph below. In the local segment, our estimate at 444 million dollars was less than the average of the other firms at 474 million dollars. However, our estimation was still within the margin of error. For the goods segment, our estimation at 282 million dollars was over the average estimation of the other firms at 261 million dollars, and yet still within the margin of error. Our estimation for the travel segment at 71 million dollars was just right compared to the average estimation of the other firms at 70 million dollars.

	Local	Goods	Travel
YipitData Est.	**444**	**282**	**71**
JP Morgan Est.	490	276	71
Morgan Stanley Est.	508	295	67
Deutsche Bank Est.	423	211	71
Average Est.	474	261	70

Given all the estimates of the billings, units sold, and new deals of Groupon for Q4 2013, as seen in the graphs below, we can say that Groupon is on the rise in every segment compared to the previous quarter. Thus, our recommendation was to buy Groupon's stock.







In order to see the technical analysis yielding this report, an HTML file that can be opened in any web browser can be found with the submission.

Part 2

This part detected Groupon's duplicated deals in March 2015 in Belgium. The methodology used was based on splitting the deal URL and using the deals name and sub-name. For example:

- deal URL: http://www.groupon.be/deals/antwerpen_fr/abricot/38364572
- deal name: antwerpen_fr
- deal sub-name: abricot
- deal number: 38364572

The dataset was sorted by the deal URL, then the deal URL was split into three columns as; the deal's name, sub-name, and number. The deal was considered to be duplicated if the deal's name, sub-name, billing amount, and units sold were the same unless it was the first one between the duplicated deals. The first deal (that had duplicates) and deals without duplicates were considered valid deals. The duplicated deals were found via Pandas 'duplicated()' function that yielded a boolean output of true for duplicates and false for valid deals. That output was then translated to a valid/duplicate column called deals_status. We sanity checked our work by comparing the total number of deals before with the sum of the total number of valid deals and the total number of duplicate deals. After that, we summed the total billings and units sold for the deals that were valid only. This sum represents the total billings and units sold for Groupon in March 2015 in Belgium. The total billings were **EUR 6,027,920** and the total units sold were 4,472,507 and the total number of valid deals was 45,923.

An excel file is attached with the submission that includes all deals of Groupon in Belgium in March 2015 with the status of each deal (as valid or duplicated).