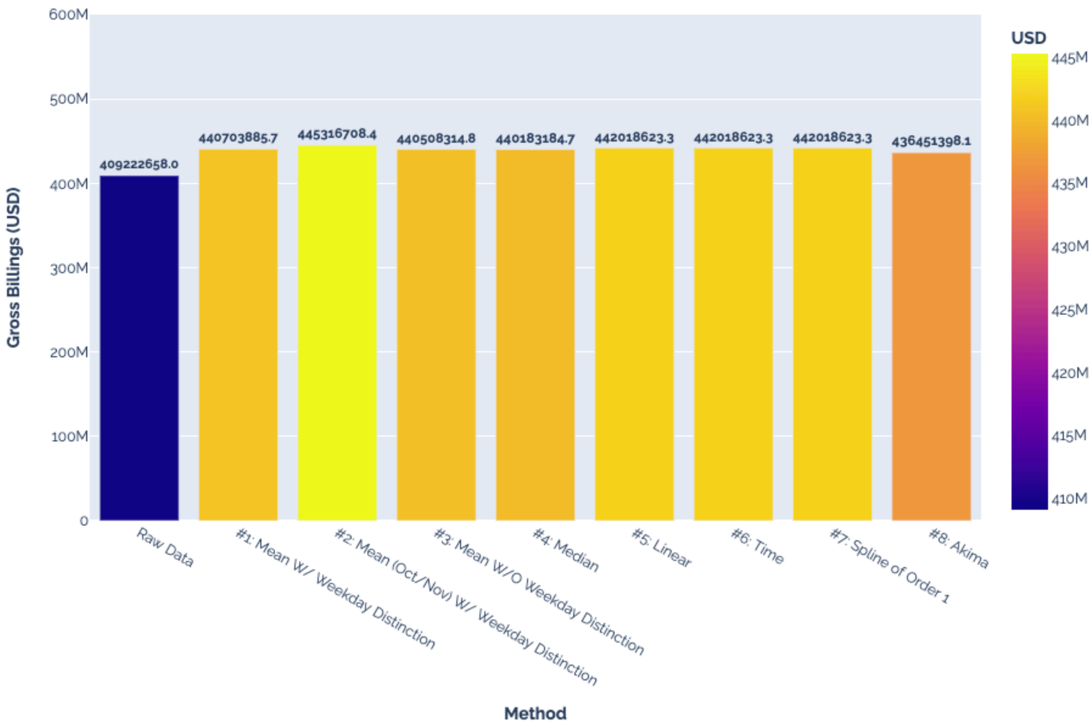
Final Estimations for Groupon's Q4 2013 North America Billings

Groupon: Q4 2013 North America Billings Estimation by Segment ($ million)			
Local	$	440,703,885.679	(440.7M)
Goods	$	282,245,671.041	(282.2M)
Travel	$	70,552,062.125	(70.6M)
Total	$	793,501,618.845	(793.5M)

Eight Estimations for Groupon's Q4 2013 North America Local Billings

To estimate Q4 2013 North America Local Gross Billings, I experimented with eight methods as shown in the graph below. Later in this report, I will elaborate on how my final estimation (Method #1) factored in adjusted data, seasonality, and a weekly pattern.



Python Libraries and Modules Used

Pandas, NumPy, SciPy, Calendar, and Plotly for graphing and visualization

Raw Data Overview

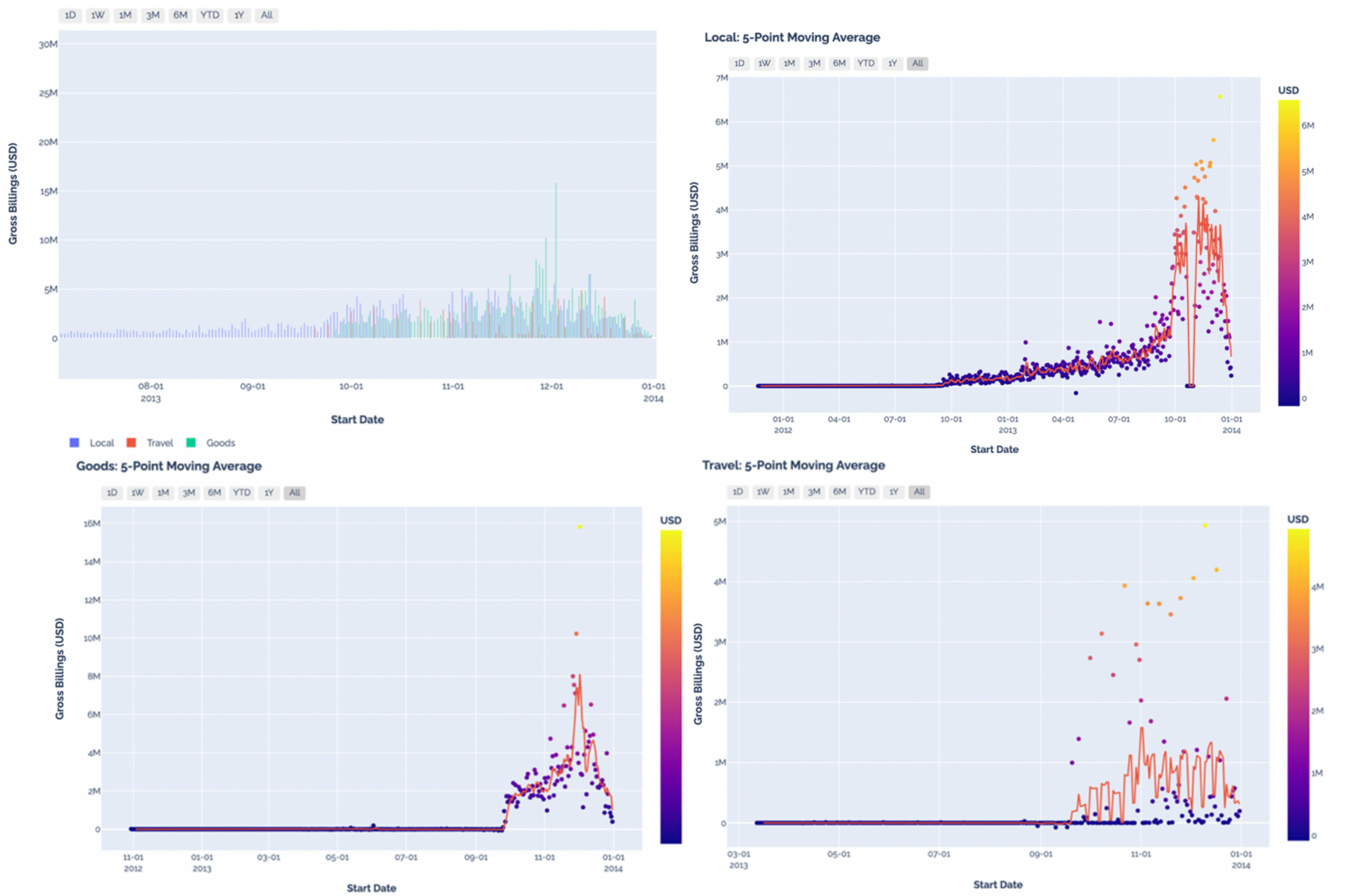
Segment	Inventory Type	Deal ID	Units Sold	Billings	Start Date	Deal URL
Goods	First - Party	14623	14623	14623	14623	14623
	Third - Party	611	611	611	611	611
Local	Third - Party	120576	120576	120576	120576	120576
Travel	Third - Party	2724	2724	2724	2724	2724

Segment	Inventory Type	Units Sold	Billings
Goods	First - Party	9595441.070	261094302.324
	Third - Party	824305.234	21151368.717
Local	Third - Party	13924480.251	409222657.982
Travel	Third - Party	378910.200	70552062.124

Segment	Units Sold	Billings
Goods	10419746.304	282245671.041
Local	13924480.251	409222657.982
Travel	378910.200	70552062.124

I imported the Q4 2013 Raw Data and created a Pandas dataframe called `grpn_df`. To check the quality of my imported data, I checked for: the size of data (138,534 total rows – matches Excel data), empty cells (none), and duplicates of Deal ID (none – each row has a unique Deal ID). Of the 138,534 active deals, there were 26,774 deals with 0 Units Sold and 0 Billings, which meant these deals did not generate Gross Billings in Q4. I did not remove these deals from the data because they were still active in Q4. About 78.6% (94,823 deals) of active Local deals, 96.4% (14,680 deals) of active Goods deals, and 82.9% (2,257 deals) of active Travel deals generated Gross Billings in Q4. I broke down the raw data based on Segment and Inventory Type: Local performed the strongest among the Segments with the largest number of deals (shown above left) and highest Gross Billings followed by Goods and Travel (shown above middle). Only Goods had both First-Party and Third-Party deals, which is likely because Groupon sells products directly or as agent on behalf of the merchant. If the raw data does not need any adjustments, I can use `grpn_df.groupby([‘Segment’]).sum()` to estimate Q4 2013 North America Gross Billings by Segment (shown above right).

Data Resampling and At-First-Glance Patterns



I created individual dataframes for each Segment with “Start Date” set as the index. While I am using time-series data where the data is indexed in time order, it is important to note that the “Start Date” is the date the deal launched – not necessarily the date the billings occurred.

I resampled the data to compute daily Gross Billings by Segment. For example, daily Local Gross Billings of 233,797.393 on Start Date of 12/31/13 meant deals that launched on 12/31/13 generated 233,797.393 of the Total Q4 Local Gross Billings.

At first glance, Local was the only Segment for which Q4 Gross Billings accumulated Gross Billings from deals that launched as early as 2012 and consistently throughout 2013. This may indicate that Groupon retained many prolonged partnerships with local merchants that saw success and extended their deals. For example, consumers satisfied with Groupon deals from local restaurants and nail salons may have become regulars at those shops, repeatedly and frequently purchased the Groupon deal, and even referred others to the business – a three-way partnership that would have increased Groupon’s Gross Billings by building customer loyalty and retention. This may explain why even though 21.4% of active Local deals did not generate Gross Billings in Q4, they may have remained active in anticipation of future purchases.

Goods and Travel had data points heavily centered around Q4 Start Dates. This could have been due to the nature of the Goods and Travel deals that remain active for limited periods, lose popularity over time, or experience decrease in demand especially as these Segments faced competition from established power players such as Google, Amazon, and Expedia that not only have larger subscriber bases but also directly interfere with Groupon's market acceptance by offering similar deals. It could have, however, also pointed to the effectiveness of the personalized targeting technology Groupon used (source: Groupon's 2012 annual report) for their Local deals. Groupon could consider implementing this technology in the other Segments as a marketing strategy to boost Billings.

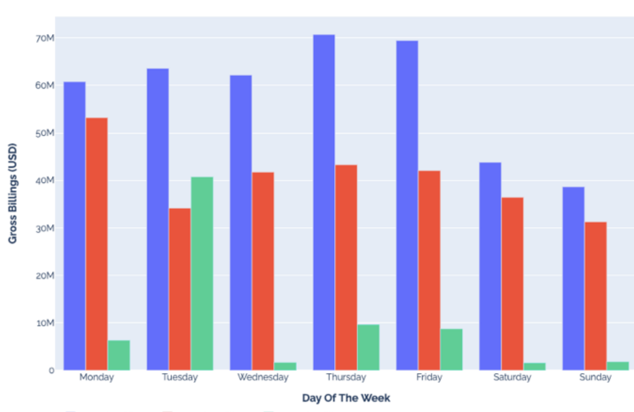
All three Segments had some of their highest daily Gross Billings for November and early to mid-December Start Dates. This reflected the "seasonal buying pattern where demand declines during customary summer vacation periods and increases during the fourth quarter holiday season" (source: Groupon's 2012 annual report). I will further discuss seasonality throughout the report. Businesses often ramp up themed emails, search engine ads, and social media promotions during the holiday season. Since they are one of Groupon's traffic sources, they may have indirectly increased Groupon's online exposure to bolster Gross Billings. Merchants may have offered more deals to Groupon and vice versa with Groupon pushing merchants to release more deals, resulting in not just increased demand but also increased supply. Furthermore, Groupon's management reported that "they [were] very pleased with the early returns from the site design launched on November 1" (source: Morgan Stanley report) and that "mobile download accelerated to almost 10M in Q3 from 7M in Q1/Q2" (source: Deutsche Bank report). The successful website revamp and surge in active users, especially as Groupon transitioned from an email-driven to a mobile-forward approach, could have contributed to the upward trend in daily Gross Billings of November and onward Start Dates.

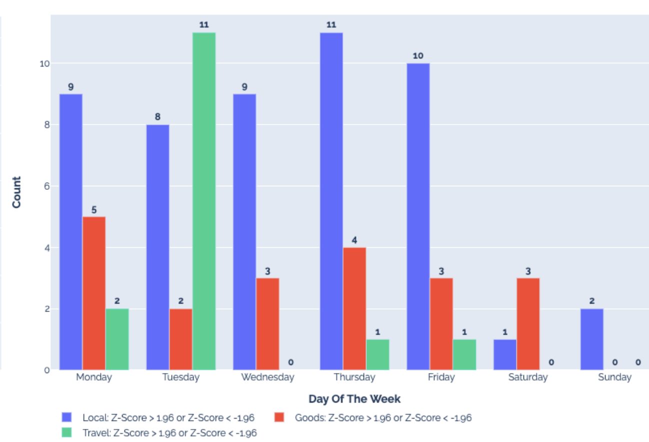
Statistical Analysis (Distribution, Z-Score & Correlation), Seasonality, and Weekly Patterns

I used SciPy functions and my own functions to statistically analyze the data.

Among various normality tests, I used the D'Agostino-Pearson test – specifically over the Shapiro-Wilk test – because the dataset contained non-unique values (such as 0). When tested for normal distribution, Local was the only Segment with data that seemed to come from a normal distribution but only across Q4 Start Dates.

I created a `z_score()` function to detect outliers, which I defined as any data point with a Z-score that is greater than 1.96 or less than -1.96 to test using a confidence interval of 95% and identity data points that are in the top or bottom 2.5% of the data. The function compares the Z-scores of each daily Gross Billings to the mean of the group and returns data points that are significantly greater or less than the mean (at least around 2 standard deviations greater or below the mean). Although I used the term "outlier" to highlight these data points, they were not excluded from any calculations because they were derived directly from real data and therefore are valid. My main goal was to use these unusual peaks or lows to spot patterns. Even though the Goods and Travel data did not seem to depart from a normal distribution, I still used Z-scores for them because as per the Central Limit Theorem, the distribution of the sample mean is most likely not very distinguishable from a normal distribution since both data have finite variance.





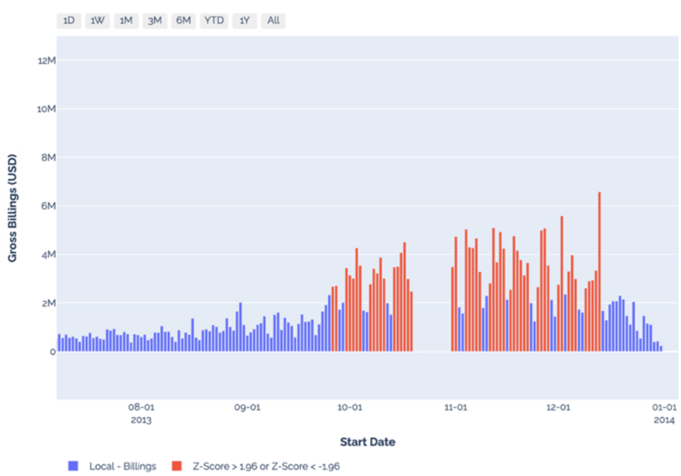
I noticed weekly patterns from the graphs above: the left graph shows the daily Gross Billings for each Segment grouped by the day of the week of the Start Date (for example, 10/01/13 was a Tuesday) and the right graph shows the outlier Z-scores grouped by the day of the week of the Start Date.

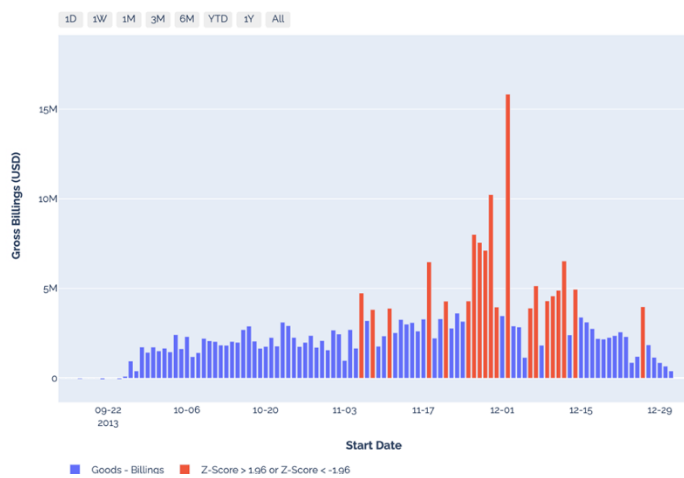
For Local, there seemed to be a weekly pattern where daily Gross Billings of weekday Start Dates (Monday to Friday) were greater than those of weekend Start Dates (Saturday and Sunday) from September to December. Since Local includes deals that are services (for example: food and drink vouchers, discounted art exhibit tickets) that require consumers initiating action to redeem them, consumers may have been more likely to purchase a Groupon deal during the weekdays so that they can redeem it on weekends. Such consumer behavior pattern may have encouraged Groupon to release more Local deals during the weekdays.

For Goods, there seemed to be a weekly pattern where daily Gross Billings of Monday Start Dates from late September to November were higher than the rest. Since Monday is the first day of the work week, there may have been a spike in the email open rate from consumers and therefore, increased traffic across the Groupon platforms. Groupon may have released more Goods deals not only in anticipation of this surge but also as a strategy for customer retention to beat competitors and their competing offers. However, daily Gross Billings of December Start Dates disrupted this pattern possibly due to the overall increased supply and demand for Goods deals leading up to Christmas.

For Travel, there seemed to be a weekly pattern where daily Gross Billings of Tuesday Start Dates were significantly greater than the rest and those of Wednesday, Saturday, and Sunday Start Dates were the lowest, from late September to mid-December. This may indicate that there was low supply and demand of Travel deals on weekends and that the Travel industry at large may have been operating around Tuesday, affecting the way Groupon released their Travel deals.

The Z-score analysis mirrored these weekly patterns. For Local and Goods, most of the outliers were of weekday Start Dates with peaks on Thursday for Local and on Monday for Goods. For Travel, no outliers were of weekend Start Dates and the vast majority of the outliers were of Tuesday Start Dates. Although limited, the similarity in the weekly patterns of daily Gross Billings between Local and Goods may explain why only the correlation between Local and Goods was statistically significant with a p-value < 0.05 when I calculated the Pearson correlation coefficients between the Q4 daily Gross Billings of the three Segments.





The graphs above show the daily Gross Billings vs. Start Date for each Segment with the red bars representing outliers. I used these graphs to further investigate seasonality and weekly patterns. Most of the outliers were daily Gross Billings of Q4 Start Dates, which may indicate that there was increased supply from merchants and demand from consumers for Groupon deals approaching and during the holiday season including Halloween (10/31), Thanksgiving (11/28), Black Friday (11/29), Cyber Monday (12/02), and Christmas (12/25).

Local had some of its highest daily Gross Billings on Start Dates leading up to Cyber Monday from Thanksgiving (11/26 – 12/02) with a peak on 12/13. Goods also had its highest daily Gross Billings on Start Dates leading up to Thanksgiving with a peak on Cyber Monday (12/02) and a second peak on Black Friday (11/29). It was interesting to note the high daily Gross Billings for 12/26 Start Date since 12/26 is past Christmas and there was a sharp decline in daily Gross Billings from 12/23 to 12/24 and 12/25 Start Dates. A contributing factor for this could have been the increased supply and demand of deals that launched on Canada's Boxing Day (12/26). Travel had consistently high daily Gross Billings for every Tuesday Start Dates from October to mid-December with a peak on 12/10.

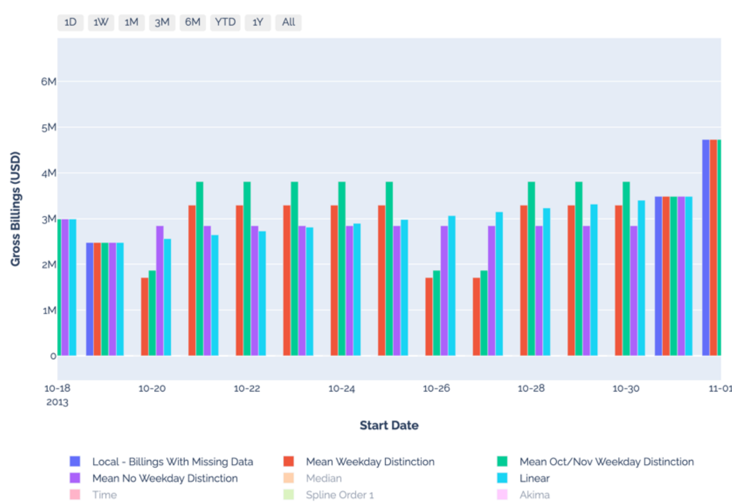
Some of the lowest daily Gross Billings for all three Segments were also of Start Dates that were holidays namely Christmas Eve, Christmas, and New Year's Eve. This may indicate that either Groupon or Groupon's merchants did not launch as many new deals on these dates and that there was low traffic across the Groupon platforms. Overall, the seasonal timeframe was in line with the period of seasonal strength for the industrials sector.

Data Adjustments Using Means, Median & Interpolation

The raw data was missing Local deals that launched between 10/20 and 10/30 inclusive, which meant it was missing Q4 Local Gross Billings that were generated from deals that launched on these dates, if any. I adjusted the data by estimating the daily Gross Billings of the missing Start Dates. I experimented with eight methods to fill in the missing daily Gross Billings data in my dataframe (shown below). I completed all calculations in Python.

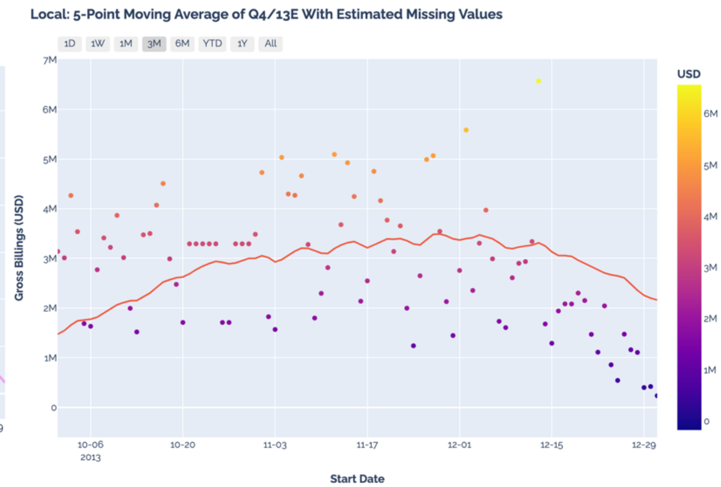
For my first and primary method, I filled in the missing values using a weekday and a weekend mean. To account for the weekly pattern where the daily Gross Billings were higher for Start Dates that were weekdays versus weekends, I distinguished weekdays from weekends and used a separate weekday mean and a weekend mean. I computed both means using daily Gross Billings of Q4 Start Dates to factor in the weekly pattern and seasonality. The period of 10/20 to 10/30 inclusive is a total of 11 days: 8 weekdays and 3 weekends. I estimated the daily Gross Billings of weekday Start Dates (10/21, 10/22, 10/23, 10/24, 10/25, 10/28, 10/29, 10/30) to be the weekday mean (3,293.963.737) and the daily Gross Billings of weekend Start Dates (10/20, 10/26, 10/27) to be the weekend mean (1,709,839.267) (shown in red from the right graph below). After filling in the missing values with the means, I used `.sum()` to add up the daily values and estimate the Total Q4 2013 North America Local Gross Billings: **440,703,885.679**.

Start Date	Raw Data	#1: Mean W/ Weekday Distinction	#2: Mean (Oct/Nov) W/ Weekday Distinction	#3: Mean W/O Weekday Distinction	#4: Median	#5: Linear	#6: Time	#7: Spline of Order 1	#8: Akima
2013-10-18	2991174.718	2991174.718	2991174.718	2991174.718	2991174.718	2991174.718	2991174.718	2991174.718	2991174.718
2013-10-19	2477089.304	2477089.304	2477089.304	2477089.304	2477089.304	2477089.304	2477089.304	2477089.304	2477089.304
2013-10-20	0.000	1709839.267	1869443.992	2844150.616	2814593.337	2561149.652	2561149.652	2561149.652	2280300.758
2013-10-21	0.000	3293963.737	3810714.804	2844150.616	2814593.337	2645210.000	2645210.000	2645210.000	2158839.228
2013-10-22	0.000	3293963.737	3810714.804	2844150.616	2814593.337	2729270.348	2729270.348	2729270.348	2105425.642
2013-10-23	0.000	3293963.737	3810714.804	2844150.616	2814593.337	2813330.696	2813330.696	2813330.696	2112780.924
2013-10-24	0.000	3293963.737	3810714.804	2844150.616	2814593.337	2897391.044	2897391.044	2897391.044	2173625.999
2013-10-25	0.000	3293963.737	3810714.804	2844150.616	2814593.337	2981451.392	2981451.392	2981451.392	2280681.793
2013-10-26	0.000	1709839.267	1869443.992	2844150.616	2814593.337	3065511.740	3065511.740	3065511.740	2426669.232
2013-10-27	0.000	1709839.267	1869443.992	2844150.616	2814593.337	3149572.088	3149572.088	3149572.088	2604309.240
2013-10-28	0.000	3293963.737	3810714.804	2844150.616	2814593.337	3233632.436	3233632.436	3233632.436	2806322.744
2013-10-29	0.000	3293963.737	3810714.804	2844150.616	2814593.337	3317692.784	3317692.784	3317692.784	3025430.668
2013-10-30	0.000	3293963.737	3810714.804	2844150.616	2814593.337	3401753.132	3401753.132	3401753.132	3254353.939
2013-10-31	3485813.480	3485813.480	3485813.480	3485813.480	3485813.480	3485813.480	3485813.480	3485813.480	3485813.480
2013-11-01	4730799.690	4730799.690	4730799.690	4730799.690	4730799.690	4730799.690	4730799.690	4730799.690	4730799.690



For my second method, I followed the same substitution approach but only used daily Gross Billings of Start Dates in October to November to compute the means because I noticed that the December data points diverged slightly from the weekday-high/weekend-low daily Gross Billings pattern after 12/13. As expected, this increased the weekday mean from 3,293.963.737 to 3,810,714.804, the weekend mean from 1,709,839.267 to 1,869,443.992, and the Q4 Local Gross Billings estimation from **440,703,885.679** to **445,316,708.390** (shown in green from right graph above). I did not use this method for my final estimation because only the daily Gross Billings of two Start Dates in December were detected as outliers so I cannot assume most of the December data to be unusual.

The other methods of substitution were using a single mean without a weekday/weekend distinction, median, and interpolated values from Linear Interpolation, Time Interpolation, Spline Interpolation of Order 1, and Akima Spline Interpolation. I did not use these methods for my final estimation because they did not reflect the weekly pattern (shown in the left graph below). All eight estimations are reported in the graph on Page 1.



To summarize, I estimated Q4 2013 North America Local Gross Billings by filling in missing values for daily Gross Billings using a weekday mean and a weekend mean, and using .sum() to add up the total: **440,703,885.679**. The right graph above shows that the means seem to moderately fit the 5-point moving average trendline.

I estimated Q4 2013 North America Goods Gross Billings: **282,245,671.04132** using:

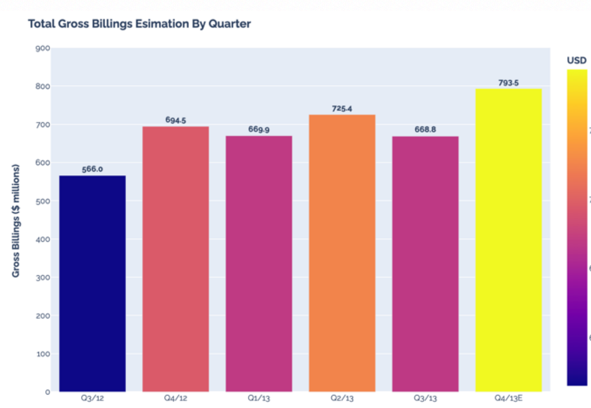
```
grpn_df.loc[grpn_df['Segment'] == "Goods", 'Billings'].sum().
```

I estimated Q4 2013 North America Travel Gross Billings: **70,552,062.1245** using:

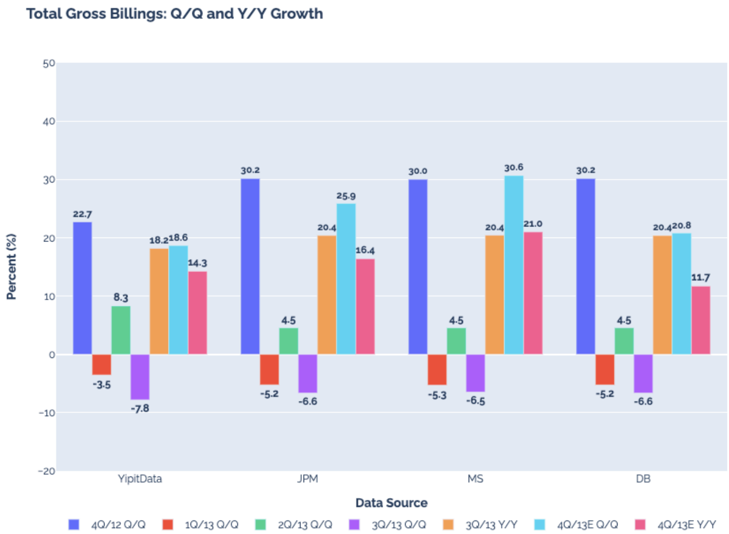
```
grpn_df.loc[grpn_df['Segment'] == "Travel", 'Billings'].sum().
```

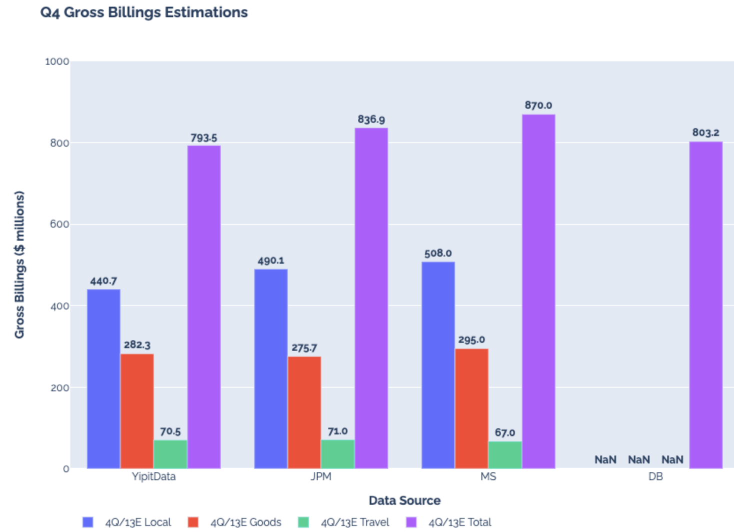
Summary Table of Final Estimations including Q/Q Growth & Y/Y Growth

	3Q/2012	4Q/2012	Q/Q Growth	1Q/2013	Q/Q Growth	2Q/2013	Q/Q Growth	3Q/2013	Q/Q Growth	Y/Y Growth	4Q/2013E	Q/Q Growth	Y/Y Growth	2013E
Billings ($ million)														
Local														
YipitData	408.94	431.10	*5.42%*	469.10	*8.81%*	459.20	*-2.11%*	410.41	*-10.62%*	*0.36%*	**440.70**	*7.38%*	*2.23%*	1,779.42
JPM	355.74	430.26	*20.95%*	450.14	*4.62%*	450.46	*0.07%*	402.80	*-10.58%*	*13.23%*	490.14	*21.68%*	*13.92%*	1,793.54
MS	356.00	430.00	*20.79%*	450.00	*4.65%*	450.00	*0.00%*	403.00	*-10.44%*	*13.20%*	508.00	*26.05%*	*18.14%*	1,811.00
DB	N/A	N/A	*N/A*	N/A	*N/A*	N/A	*N/A*	403.00	*N/A*	*N/A*	N/A	*N/A*	*N/A*	N/A
Goods														
YipitData	110.54	213.70	*93.32%*	144.30	*-32.48%*	201.60	*39.71%*	191.50	*-5.01%*	*73.24%*	**282.25**	*47.39%*	*32.08%*	819.65
JPM	152.12	240.85	*58.33%*	165.36	*-31.34%*	196.88	*19.06%*	194.57	*-1.17%*	*27.91%*	275.72	*41.71%*	*14.48%*	832.52
MS	152.00	241.00	*58.55%*	165.00	*-31.54%*	197.00	*19.39%*	195.00	*-1.02%*	*28.29%*	295.00	*51.28%*	*22.41%*	852.00
DB	N/A	N/A	*N/A*	N/A	*N/A*	N/A	*N/A*	195.00	*N/A*	*N/A*	N/A	*N/A*	*N/A*	N/A
Travel														
YipitData	46.51	49.70	*6.86%*	56.50	*13.68%*	64.60	*14.34%*	66.88	*3.53%*	*43.80%*	**70.55**	*5.49%*	*41.96%*	258.53
JPM	44.51	47.85	*7.51%*	65.82	*37.55%*	64.86	*-1.45%*	67.64	*4.28%*	*51.96%*	71.02	*5.00%*	*48.42%*	269.34
MS	45.00	48.00	*6.67%*	66.00	*37.50%*	65.00	*-1.52%*	68.00	*4.62%*	*51.11%*	67.00	*-1.47%*	*39.58%*	266.00
DB	N/A	N/A	*N/A*	N/A	*N/A*	N/A	*N/A*	68.00	*N/A*	*N/A*	N/A	*N/A*	*N/A*	N/A
Total														
YipitData	565.99	694.50	*22.71%*	669.90	*-3.54%*	725.40	*8.28%*	668.79	*-7.80%*	*18.16%*	**793.50**	*18.65%*	*14.26%*	2,857.59
JPM	552.37	718.95	*30.16%*	681.32	*-5.23%*	712.21	*4.53%*	665.00	*-6.63%*	*20.39%*	836.88	*25.85%*	*16.40%*	2,895.40
MS	553.00	719.00	*30.02%*	681.00	*-5.29%*	712.00	*4.55%*	666.00	*-6.46%*	*20.43%*	870.00	*30.63%*	*21.00%*	2,929.00
DB	552.40	719.00	*30.16%*	681.30	*-5.24%*	712.20	*4.54%*	665.00	*-6.63%*	*20.38%*	803.20	*20.78%*	*11.71%*	2,861.70



Buy Or Sell Recommendation and Ideas for Further Analysis





I recommend buying the Groupon stock. The left graph above shows that Groupon had a steady Y/Y Growth for Q3/Q4 2013 and a generally upward trending quarterly growth despite small dips. While my Q4 Gross Billings estimations were lower than those by J.P. Morgan (JPM), Morgan Stanley (MS), and Deutsche Bank (DB), they all anticipate the Q/Q Growth and Y/Y Growth for Q4 to be positive.

I also factored in information from the Equity Research reports to my recommendation. According to JPM and DB, Groupon's Q3 revenue was negatively impacted by seasonality, disruptive changes to Gmail format, and deal fatigue. They still each maintained a Neutral and Buy rating, citing good progress in Groupon's transition from a "push" to a "pull" approach and its competitiveness in the leisure, recreation, and foodservice markets. Given that seasonality is a known risk and Groupon is expected to introduce a new email product in 2014 (source: DB report), I believe Groupon is fixing many of the addressed problems and moving forward in the right direction. MS rated Groupon "Overweight" highlighting a positive momentum in North America, a high number of active users, and an expanding mobile platform. I can see these strengths playing a role in my analysis.

As Groupon continues to build a stronger marketplace with more, better, and diverse merchants and invest in both the newer (mobile-forward) and the traditional (email-driven) approaches, I expect an upward growth trajectory in both Gross Billings and Gross Profit. Therefore, I recommend buying the Groupon stock as a long-term investment even though the stock price may be volatile in the short-term as the company undergoes changes.

If I had additional data, I think I would be able to give an even more data-oriented recommendation and overall analysis. Data I am interested in seeing include:

1) Historical data on web traffic, mobile downloads, and email open rates to observe course of transition
2) A/B Testing results to evaluate the November 2013 site redesign and the 2014 new email layout
3) Engagement overview (total visits, visit duration, pages per visit, bounce rate, conversion rate, exit rate)
4) Traffic by geography in the North America region to detect geographical patterns correlated with seasonality
5) Traffic sources (direct, referrals, search engines, social media, emails, display advertisements, other ad channels) especially since most of the Groupon deals' inventory type was third-party
6) Data from 1) to 5) on competitors for competitor analysis
7) Nasdaq Composite data and historical stock data for Groupon, which I can use to run machine learning algorithms and predict future stock trends (I have done a project on this before!)

*I would be more than happy to share my full Python code upon request.