Candidate: Aram Moghaddassi
Submission Date: October 19, 2018

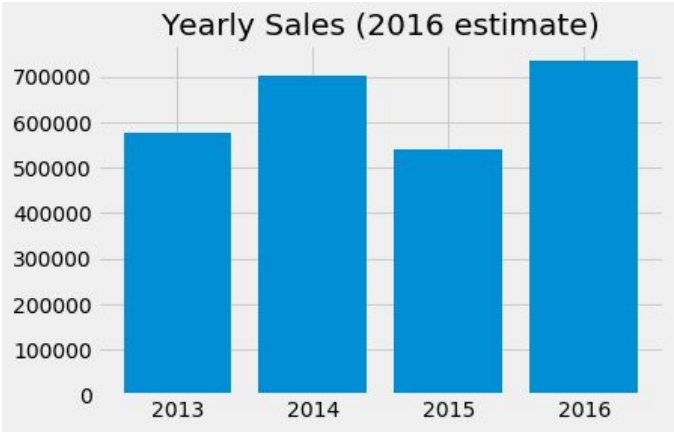
Executive Summary

Sales spiked in 2014 due to a string of purchases made by one user, but otherwise have remained roughly constant over the last three years of full data. Additionally, 2016 seems to be on track to be the best performing year thus far, as per the data we have. Revenue streams are diversified across many regions and customers, but only a few of these segments account for the majority of sales (as is to be expected).

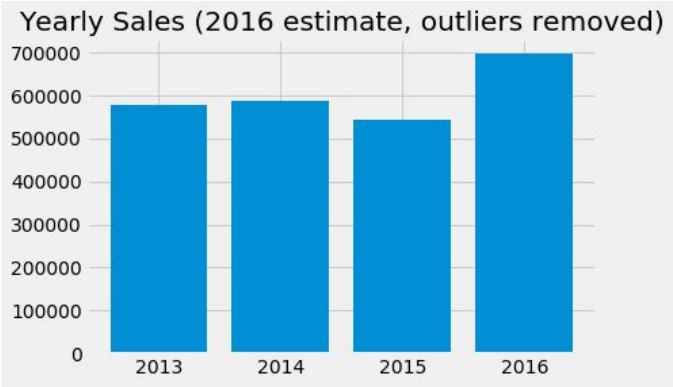
The business has solid sources of recurring revenue, and continues to attract new customers year to year. I would recommend an investment given that the business is able to articulate a sound plan of growth for the future; specifically, the business should have a plan to generate strong Q4 sales in 2016 and capitalize on one-time purchase customers.

Sales Trajectory

Total sales were about ~550,000 in 2013 and 2015, with 2014 being the highest sales year at ~700,000 total sales. There is only data up to April for 2016, so I used the average proportion of sales from January to April from 2013-2015 to generate an estimate for 2016 sales. This is shown in the figure below:

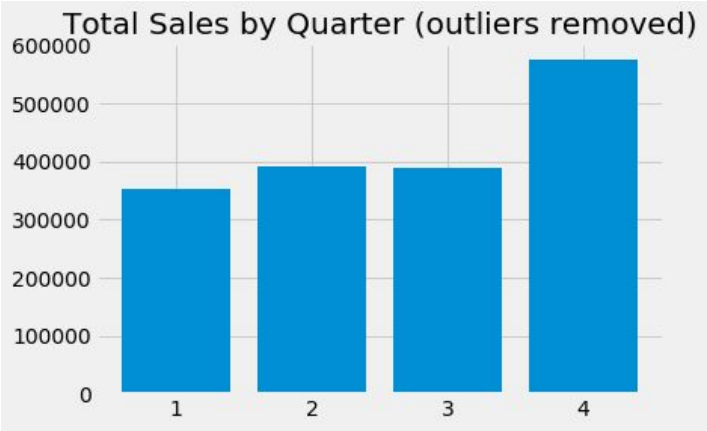


As mentioned in the summary, the high sales of 2014 can be attributed to an outlier user who made ~110,000 in purchases over 10 days in December, 2014. As this user didn't make any other purchases following this, I also considered the sales plots and estimates that didn't include this user's sales. This is shown in the figure below:



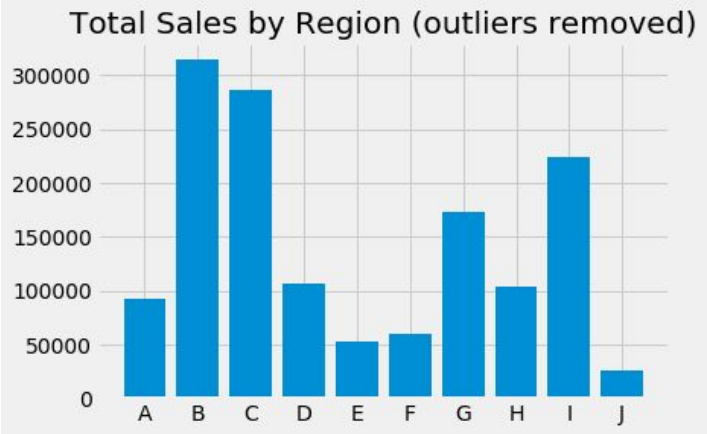
This plot shows much more consistent sales from 2013-2015, and puts 2016 on track to be the best performing year thus far.

Seasonal differences should also be noted: Sales are fairly constant over quarters 1-3, but there is an uptick in sales in the fourth quarter in all three years where data is available. This data is shown in the figure below:



Customer Behavior and Regional Differences

Though the business operates in 10 regions, four of them produce the majority of sales:



Breaking down sales by customers also leads to a similar conclusion. Purchases by customers seem to follow a Pareto-like distribution; the top 20% of customers account for ~ 72% of total sales. This is to be expected as business sales are often modeled well by this distribution. On a year by year basis, customer sales were fairly consistent with the total distribution (after accounting for the outlier user), implying that there are many one-time purchase customers, but those that account for many sales are coming back year to year.

Data Analysis Methodology

This section will give an overview of my approach to the dataset as to give context to the findings described above. All analysis was completed in Python (matplotlib, pandas, numpy) in the Jupyter Notebook environment (source code file is also included in the report).

The dataset offered multiple criteria with which I could evaluate the quality of sales; I segmented on time, region, and customer. To consider time, I first plotted sales by year, and then by quarter. These plots are shown in the sales trajectory section. 2014's unusually high Q4 sales seemed like an outlier, and motivated the finer grain analysis.

To consider region, I plotted total sales by region, and then sales by year by region for only the top total sales regions. The Q4 2014 outlier was now reflected in the unusually high sales bar for region C in 2014; the other regional bars seemed fairly consistent over time. I then broke down sales by customer, first finding customers with the highest lifetime purchases. The distribution of customer sales followed a Pareto-like distribution, which makes sense given the dataset's context. Once again I broke down this plot into yearly sales distribution plots, finding the 2014 was again the only inconsistent plot in relation to the overall distribution.

I then isolated a single user (ID# 864) who accounted for the high sales performance in Q4 of 2014. The user was then removed from the dataset, and sales trajectory and estimate plots were again drawn, but now without the outlier user included.