Francisco Martinez

Google Sample Data Exploration

Exploring the different sales and visits in Different Channel Groupings

The Google Sample Data is a public data set that contains marketing data on the traffic and transactions that occurred on the Google sales website. The goal of this project is to analyze the different traffic sources, to see which are more stable, produce the most traffic, and have the most sales in the 8/01/16-8/01/17 interval. I will focus mostly on the channel groupings with the most Revenue but the data for the other channels will also be included. The yearly total Revenue, Transactions, Visits, Average Order Values, and Conversion Rates are displayed in the table below.

Google Sample Totals 8/1/2016 - 8/01/2017

Channel	Ave Order Value	Conversion Rate	Revenue	Transactions	Visits
(Other)	9,990,000	0.008333333	9,990,000	1	120
Affiliates	66,375,556	0.000548680	597,380,000	9	16,403
Direct	195,962,393	0.015514662	434,840,550,000	2,219	143,026
Display	515,378,026	0.024273395	78,337,460,000	152	6,262
Organic Search	91,142,281	0.009385131	326,380,510,000	3,581	381,561
Paid Search	90,937,161	0.018913370	43,558,900,000	479	25,326
Referral	117,522,986	0.052872050	651,429,910,000	5,543	104,838
Social	37,530,840	0.000579346	4,916,540,000	131	226,117

Ave Order Value, Conversion Rate, Revenue, Transactions and Visits broken down by Channel.

Referral, Direct and Organic search are the channels that produce the most Revenue. However Display has the highest Average Order Value with Direct succeeding it, then Referral and Organic Search afterwards. So in general, Display has less Revenue than the top three highest earning channels, but makes more revenue from each transaction. Customers from the display channel spend the most per transaction. The customers most likely to buy are from Referral and Display. Social and Organic Search have the most visits, but less of a conversion rate and Revenue compared to Referral and Direct. Social has more views, but more fickle customers. Organic Search has the highest amount of views, but only the fifth highest conversion rate. This hints that the amount of visitors is not the most important variable for predicting the amount of transactions. A linear regression test can be done to confirm the lack of correlation between the two variables.

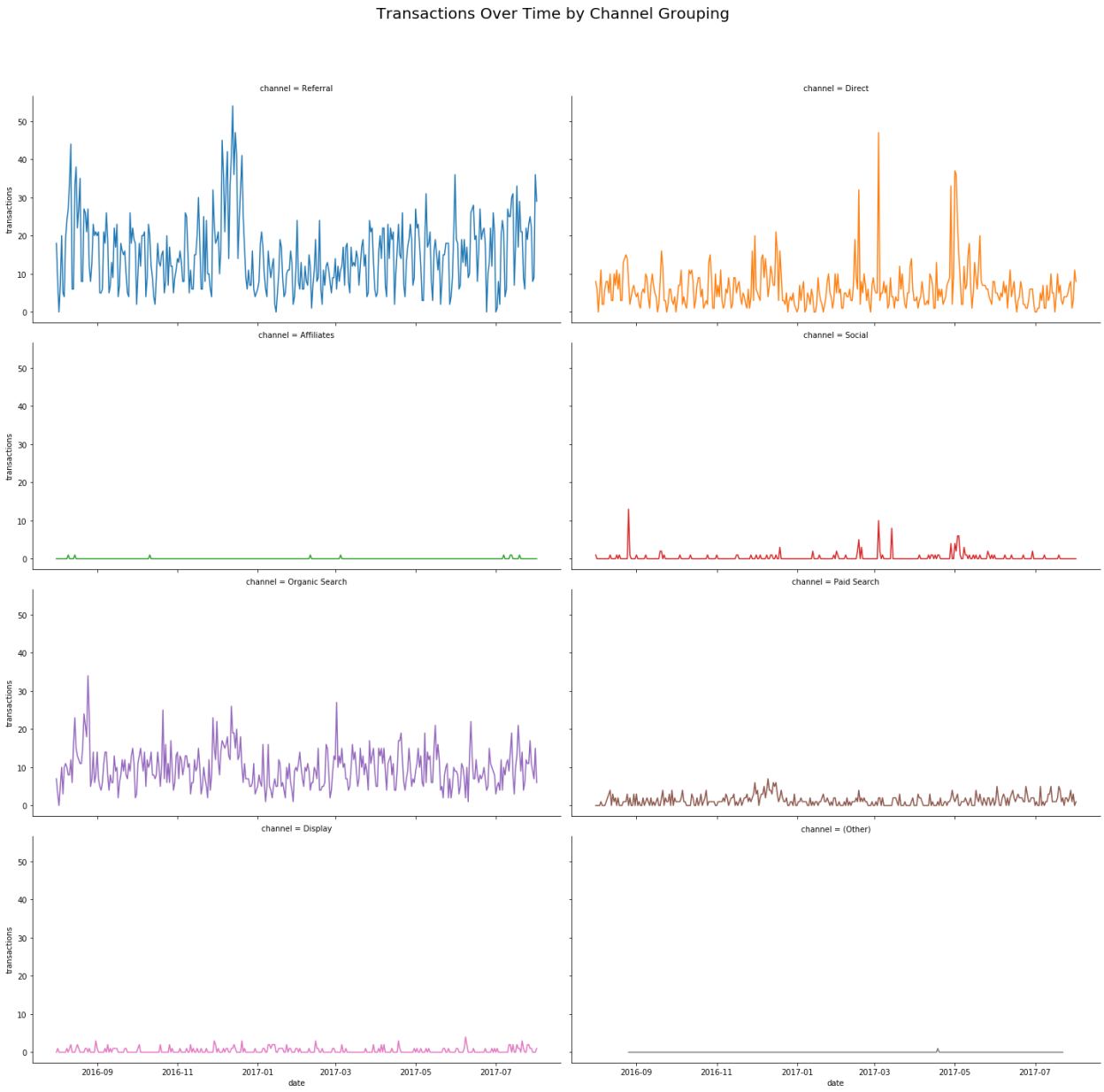
The next analysis will consist of exploring the data throughout the year to see the consistency, peaks and shifts of the channels for each variable. The following line graphs display the revenue over time for each channel grouping.

Display was stated to have customers that spend the most per transaction, however most of the spending occurred in the interval between the middle of 03/2017 and the middle of 05/2017. Interestingly, no other channel has this spike of revenue produced in the same interval. Possible causes could be because of a specific product or targeted ads. The Direct Channel has multiple high peaks or Revenue spread out throughout the year compared to Display. One is in the beginning of 9/2016, another around 12/2016, the max on 02/2017, and a series of peaks around 7/2017. Direct at all other times is consistently under .5 in revenue. Organic Search and Referral are more consistent in revenue with less drastic peaks compared to the previous channels. This shows that the revenue is more constant throughout the for these two variables and a more reliable source of revenue compared to the previous two channels. However there are noticeably time intervals where there are consistently higher peaks of revenue.

A further step to include in this analysis of revenue is to investigate the cause of the drastic peaks of the Display and the Direct channels in order to understand how to recreate the peaks or even create a more consistent stream of peaks throughout the year. A possible avenue for investigation is to look at the websites with the display ads that were linked to the

web store between the middle of 03/2017 and the middle of 05/2017. Another investigation might be to look at the products sold during the peaks of the Direct Channel.

The next analysis is on the amount of transactions over time. The following graphs show this according to channel groupings.



The display channel has a consistent number of transactions throughout the year including the 03/2017-05/2017 interval when the revenue spiked. This shows that the number of transactions and average order value remained consistent throughout the year except for when the average order value raised considerably in the interval of 03/2017 -05/2017. Referral transactions fluctuate from high to low but seem to be consistently high. it seems to rise when the Referral revenue rises, hinting at a low average order value. This would mean that Referral makes most

of its revenue off of a large amount of transactions of low value. Organic search also seems to have a constantly high amount of transactions hinting at it serving most of its revenue from a high amount transactions of low average order value. The Direct channel transactions do not seem to follow the revenue peaks in the beginning, but does seem to follow the revenue peaks at around 03/2017, suggesting a low average order value before 03/2017 and a higher value afterwards. Organic search transactions seem pretty consistent.

To confirm the previous statements, the average order value throughout the year will be examined . The following graph examines the average order value by channel groupings.

The display channel did have a high average order value from 03-2017 to 05-2017. This confirms that most of the Revenue was made in the stated interval and that the revenue was mostly from a small amount of transactions of high value. Organic Search does have a low average order value as predicted before with small peaks at the end of the timeline. The direct

channel has high average order values on the first three months with a low order value afterwards. This validates my previous statement that the first peaks in revenue for the direct channel was caused by a pretty low amount of transactions of higher value, with the later peaks in revenue caused by a large amount of transactions of low average order value. This is interesting because in general, most high peaks of revenue for each channel is either caused by a high frequency of transactions of low average value such as referral, or a low frequency of transactions with high average order value such as display. Understanding why these peaks happened for Direct allows different strategies of generating revenue, either trying to elicit more transactions and low order values or less transactions with high order values. Lastly, the average order value of Organic Search and referral shows that there was a constant stream of revenue throughout the year caused by many transactions with a lower average order value.

While analyzing the sums of visits in the first table, I concluded that the amount of visits seems to not affect the amount of transactions. To study this, using the graphs of visits over time and the conversion rate over time can give us a more in depth analysis. The following graph is on the visits over time.

For the majority of the graphs the amount of visits seem consistent over the year. Referral's visits seemed to follow the amount of transactions. Organic Search seems to have a slight upward trend with Direct having two thin peaks and a couple of shorter thicker peaks. Display has a low amount of views that decreases at around 03-2017 which is when the average order value increases. Lastly social starts the year with 500 to 1000 views, increases drastically over 2000 visits and then drops. Specifically the amount of views for Social do not seem to correlate with a rise of transactions or revenue for social. To see whether these visits affect the amount of transactions the following graph shows the conversion rate throughout the year.

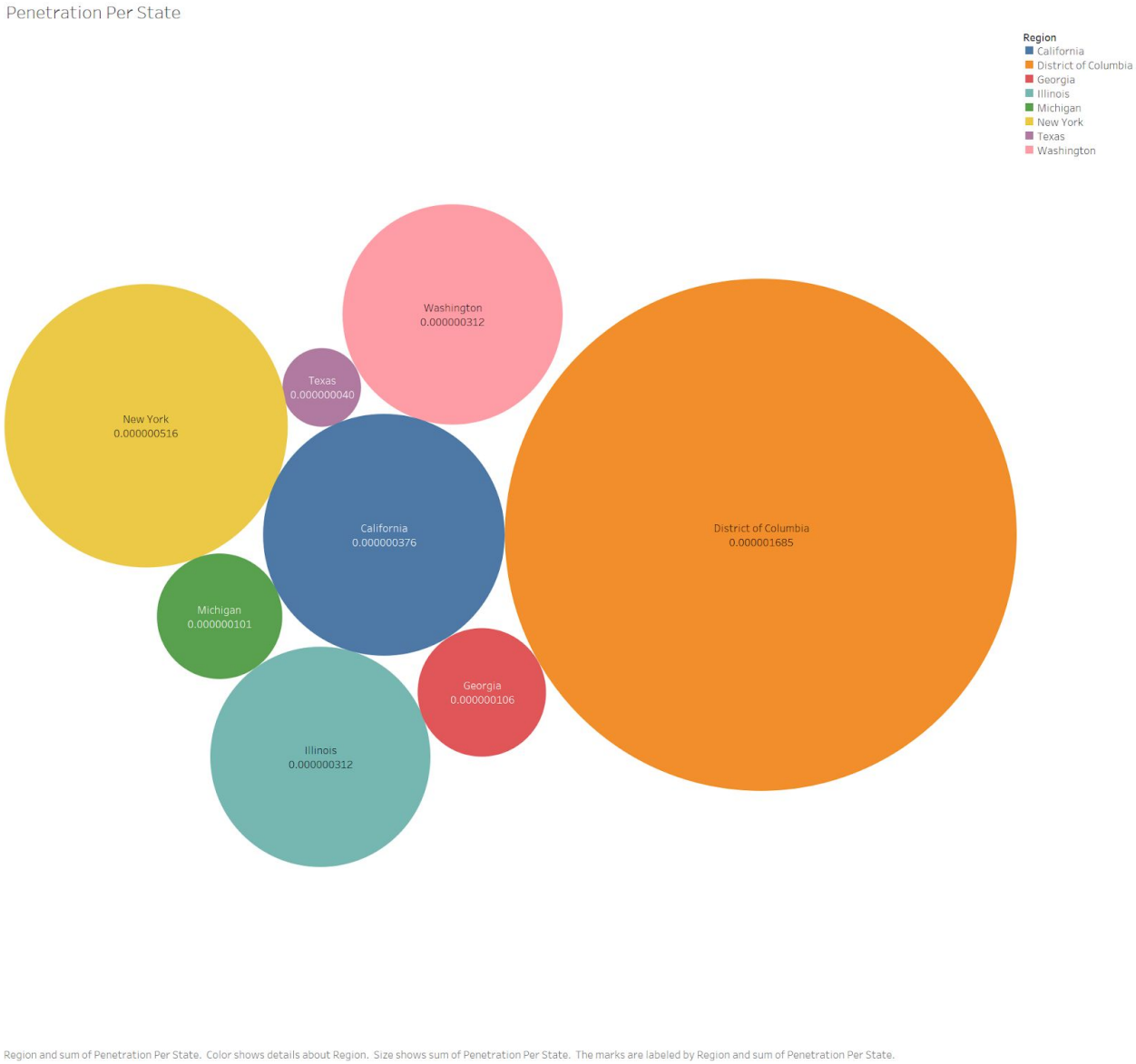
The amount of visits for social at the beginning of the year did not convert to transactions and the conversion rate increased later on because there was a decrease of visits, demonstrating that the visitors from social are among the least likely to buy and least loyal. Organic Search has a pretty consistent low conversion rate. Even though Organic Search has a large amount of transactions, it has such a low conversion rate since it has such a large pool of visitors, many of them who do not buy anything. This means that Organic search visitors are not that loyal, but since the pool of people is so large, there are still many transactions with the small percentage of people who buy. The Referral conversion rate is pretty consistent between .5% and 1% throughout the year, making it one of the highest and stable conversion rates. This suggests that investing in ways to generate views from Referral could be very lucrative since the viewers for Referral are consistent customers. Direct is pretty consistent except for a large peak at

03/2017 and another at 05/2017, which seems to match peaks in transactions around the same dates. On 05/2017 the conversation rate peak also matches an increase of visits, which means that at 05/2017 there was an influx of transactions made from an increasing number of loyal visitors. Understanding what caused an influx of loyal customers in this short amount of period could allow us to raise the conversion rate throughout the year and make generating views more valuable. Display has a low amount of visitors, but the second highest conversion rate.The conversion rate shifts throughout the year with many high peaks, the highest at 04-2017, and others at 09-2016 and 01-2017. SInce there are so few viewers compared to the other big channels, every small shift in transactions and visitors creates a huge shift for the conversion rate. This means that display viewers are few but loyal, and are willing to spend the most amount of money out of all the channels. Increasing the visitors and figuring out what caused the increase of average order value is essential for raising revenue for display.

This analysis created many insights on the nature of each channel and possible ways to increase revenue. Referral is the channel with the highest revenue, most transactions, highest conversion rate, and third highest average order value. The tables showed that the Revenue for Referral is pretty consistent throughout the year and is generated mostly through a high frequency of transactions of pretty low average order values (compared to Display and Direct). Referral has a consistent amount of views and conversion rate throughout the year, making it a good channel that can generate transactions through views. Investing in how to generate views seems to be a good tactic for generating revenue for Referral. Direct was shown to have the second highest revenue and second highest average order value, fourth highest conversion rate and third highest transactions. Direct has high peaks of revenue spread out throughout the year. The first peaks were caused by a lower amount of transactions of high average order value and the later peaks were caused by a higher frequency of transactions of lower average order value. This means that there was a lower conversion rate in the first half and higher in the last half of the year. Understanding the cause of change of the conversion rate and average order value of Direct throughout the year can help make strategies for increasing revenue. Organic Search has the largest amount of visitors which means that even with the low conversion rate, it generates many transactions. Most of the revenue is generated through transactions of lower average order value. Display has the highest average order value, second highest conversion rate, fourth highest revenue, and fifth highest transactions. The revenue is consistently low except with a peak of revenue from about 04-2017 to 05-2017 caused by an increase of average order value. Transactions and visits are consistently low compared to other channels however the conversion rate is consistently high with high peaks at 04-2017. Display a small amount of loyal visitors who spend the most money compared to other channels. Understanding what caused the increase of average order value from 04-2017 to 05-2017 and increasing visitors are possible ways of increasing revenue for this channel. Social is the channel with the second most views but second smallest conversion rate. With all these insights, there is still much analysis that can be done to understand the shifts we saw in the data. However once we understand these changes we can create strategies to increase revenue for each channel

Project 2 Penetration Per State

The last analysis is investigating which region has the highest Penetration Per state. The following bubble graph displays the ratio of the amount of transactions per population of certain states. The chart is missing some states since the rest of the google state transaction data was not public, and the rest of the state transaction data was null.



The region with the highest penetration per state was the district of Columbia, which is not a state, but makes sense since it is a small region. Afterwards it is New York, California, and Washington. These states have some of the largest populations in the U.S., so it is interesting that they would have the largest amount of transactions per population. Possible reasons could be that California and Washington are big tech centers, and all three have large urban populations.