**Analyst Evaluation -** Neil Schwalb

Introduction

With the impending launch of the new North American product, it is necessary to analyze the US and Canadian markets in an effort to determine where the product should be launched first. The end goal of this analysis is to determine which North American country's VIP users should be the first to receive the new product. To complete this exercise, the following thought process and questions shall be answered.

1. Identify the top 10 stores, globally, by successful sales, prior to 2014 for currently active users.
2. Compare and contrast the sales distributions between the US and Canadian stores.
3. Make an initial hypothesis on whether Canada's or the US's VIPs shall be used as the target group for the new product.
4. Look at an additional factor that could affect the country for the product launch and re-evaluate the initial hypothesis.

Results

Step 1.

The top 10 stores, as seen below, range from 91 sales to 53 sales from 2010 to 2014.

Successfully Processed Orders by Store (Top 10)
Prior to 2014
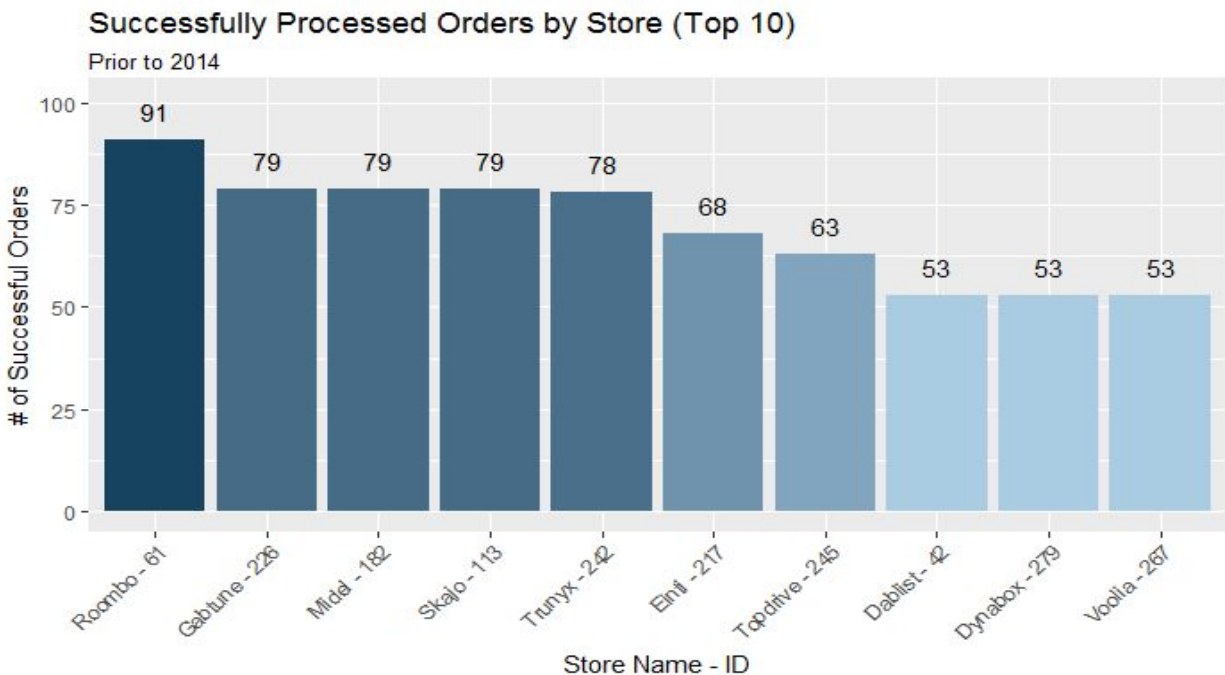
Step 2.

Overall the distribution of sales in the US has a positive skew (majority of sales are on the lower end), with 85% of the 154 stores yielding less than $12,000 apiece and the other 15% yielding anywhere from $35,000 $147,000 each. A little more that 40% of the US stores produced

revenue between $3,00 and $7,600, with median sales just under $5,000. The bottom 85% of the sales have a fairly uniform distribution, although the bottom half of the $150 to $11,000 range contains more stores than the upper half. Oppositely, the top 15% of stores, ranging from $155,000 to $616,000 in individual sales, have a slightly higher density of stores in in its upper half than the its bottom. The top 15% of stores account for more than 76% of the total revenue in the US over the past 7 years.

US Sales Distribution by Store (2010 - Present)
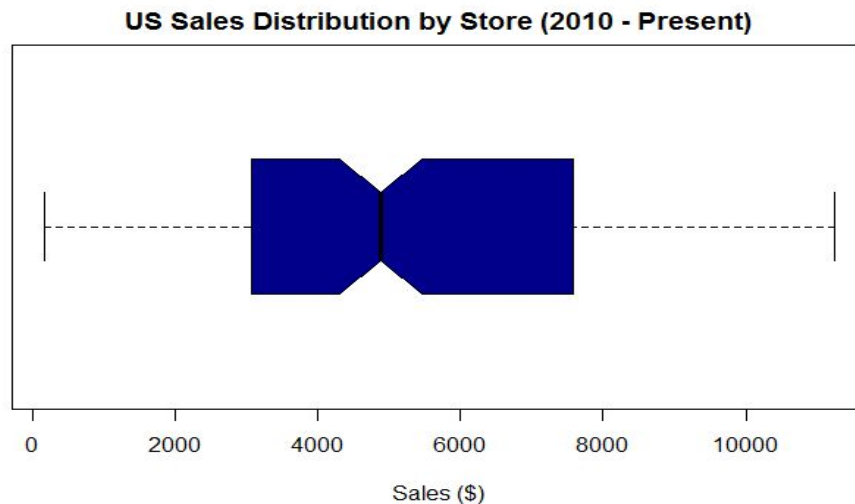
Sales ($)

Similarly, the Canadian market has a greater distribution of sales per store on the lower end of the spectrum, with the bottom 85% having a fairly normal distribution from $37 to $10,700 per store. The sales across that range, much like in the US, sit primarily on the bottom half. However, the top 15% of Canadian store sales (7 stores) don't follow a uniform distribution, with 6 of the 7 stores falling between $60,000 and $78,000 and the final store yielding almost $120,000 over the 6.5 year span.

**Canadian Sales Distribution by Store (2010 - Present)**
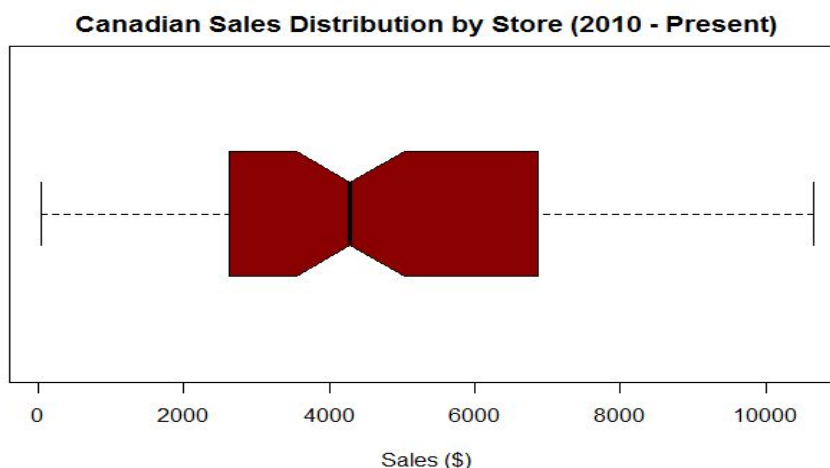
Sales ($)

Step 3.
Using the top ~10% of stores by sales over the past 7 years as the VIP users, the US would have 18 user and Canada would have 6. In that time, on average, each Canadian user has yielded approximately $81,000 in sales, whereas US users have yielded approximately $93,000

in sales each, a 13% increase to their Canadian counterparts. The US has also produced more total sales, $1.7 million compared to Canada's $480,000 over the past 6.5 years. However, Canadian users have produced approximately 1 more sale per store per year than their US counterparts (15.4 to 14.5, respectively). With all of that said, it is recommended that the US be prioritized for the new product due to the greater number of VIP users (18 to 6) and overall greater financial footprint on the continent. The 24 VIP users would allow for more feedback on the product and the nearly $260,000 in US annual sales would provide a better test of the product's efficacy.

Step 4.
Another metric that would be recommended to evaluate is the growth of the store's number of sales annually by country. This would provide an indicator of whether or not the stores are growing and how quickly they are growing.

As it can be seen in the 'Successful Sales per Annum' graph, while both countries have a generally upward trend from 2011 to 2015, Canada appears to drop off slightly during 2016. Also, looking at the 'VIP Successful Sales per Annum' graph, it can be seen that US VIPs have increased their number of sales from 2011 to 2015, whereas Canadian VIPs' sales have remained fairly constant during that time. The Canadian VIPs have suffered a similar drop-off from 2015 to 2016 as their general users. Having such a large dip in total sales during that year would make the Canadian VIP users a better candidate for the new product, as any increase in sales due to the product would be much more easily recognized. Taking this into consideration, the Canadian VIP users may serve as a better group for the new North American product release.
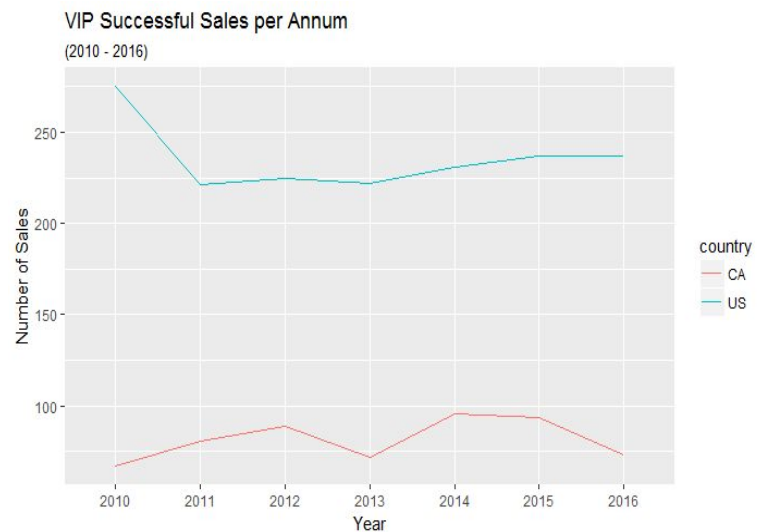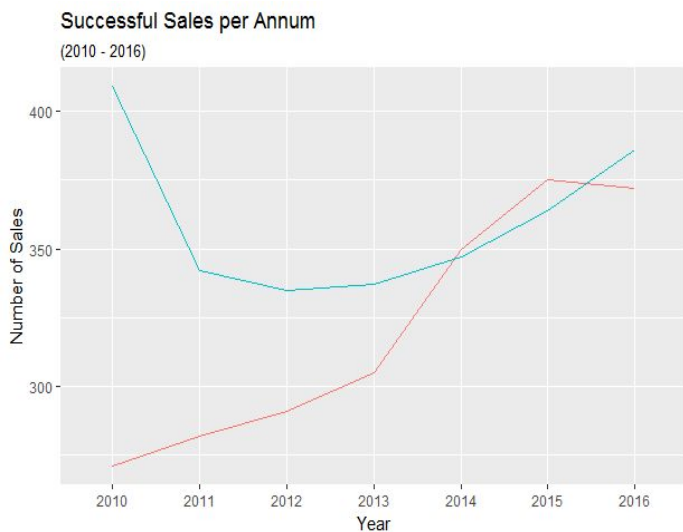
Conclusion
In conclusion, the US is the larger market with the more stores and overall total sales. However, based on the current downward trend of the Canadian VIP users' sales, it is recommended that they be prioritized for the new product launch in North America. They are more inclined to try

something new as their volume of sales is declining and results of the product launch would be more visible.

Methods

Pulling and Wrangling Data

To pull the correct data, MySQL Workbench was used to analyze the available tables content and size before pulling them into R for further analysis. All of the tables were of a manageable size and were all deemed necessary due to the required analysis. The tables were pulled into R using the RMySQL package and loaded as data frames.

Assumptions

Based on the data provided, the following assumptions were made as part of the analysis:
● Successful sales are considered to be orders that were not cancelled.
● Currently active users are considered to be users that do not have an 'account_paused' timestamp.
● The VIP users represent the top 10% of North America, by sales ($) since 2010.

Data Validation

The data was initially validated by visually looking at trends and analyzing some summaries of the tables such as how many paused accounts are there currently, how many orders have been cancelled, and whether or not there is a 1:1 mapping between users and stores.

Data Analysis

To determine the top 10 stores by sales, the users were filtered for active users and then the stores were filtered based on the previous result. To get all past successful orders, all_orders was filtered for all non-cancelled orders as well as dates processed_dates before 2014. The numbers of orders were then evaluated based on the count of orders joined with the the active_stores information.

The US and Canadian sales distributions were based off of the sum of all successful order totals by store for the two countries. A box and whiskers plot and it's statistics were then used to show the distribution without outliers. As the outliers for both the US and CA roughly represented the top 15% of each country, they were evaluated separately using box and whiskers plots and their related statistics. The means and medians were compared for each stage of box plots to help determined skew as well.

Simplified versions of the data points from the sales distribution were used to evaluate which country should be prioritized for the new product launch. The VIPs were selected through filtering the top 10% of North American users and then the sums of the order totals and numbers of orders were manipulated to find averages over years and stores.

The numbers of sales by year were taken through summation of order counts by year and country. They were subsequently plotted and through visual inspection, trends were assessed.