**Summary**

The data used in this Math Senior Project (Math 4395) was acquired from Zappos.com, an online clothing and shoe store. The data was originally an Excel spreadsheet that contains 21,061 customer transactions from January 2013 - December 2013. Some of the fields captured in this data set includes the date of each transaction, the sites visited under the Zappos general website umbrella, the number of visits that occurred in one session (one sitting), the technology device used, such as iPhone, Android, Windows, BlackBerry, etc., the amount the customer spent on each session and each visit, and much more. There is no personal customer information present in this data. Any information that can be used to trace back to a customer is disregarded. However, there is an identifier field named *new\_customer* that differentiates a completely new customer (recently registered or opened an account) from a returning customer (already has an account) and from a guest/visitor (has no account and did not open one at the occurrence of the transaction). Below is a detailed description of all fifteen columns in the dataset from Zappos.com.

***day***

The *day* column represents the date and time of the transaction. The dates are in the “MM/DD/YYYY 12:00:00 AM” format and all transactions were for the year 2013. This field has no blanks:

* 2366 transactions were on January 2013;
* 2137 were on February 2013;
* 2327 transactions occurred on June 2013;
* 2035 were on July 2013;
* 2462 were on August 2013;
* 2347 transactions were captured on September 2013;
* 2464 were on October 2013;
* 2389 were on November 2013; and
* 2534 transactions happened on December 2013

As we can see, only nine months’ worth of customer transaction were collected in this dataset. There are no data for the months of March, April and May. From the above, December 2013 has the largest number of customer transactions followed by October.

***site***

The *site* column is the company’s site visited by users and - in the scope of this dataset - they include Acme, Botly, Pinnacle, Sortly, Tabular, and Widgetry. There are no blanks in this field:

* 7392 users visited the Acme site;
* 804 visited Botly;
* 5725 visited Pinnacle;
* 5532 users visited the Sortly site;
* 804 visited Tabular; and
* 804 visited the Widgetry site

Majority of the users visited the Acme, Pinnacle and Sortly site - in that order.

***new\_customer***

This column identifies if a user is a new customer, returning customer or neither. New customers are assigned the data value 1, returning customers are assigned the value 0, while neither are null (blanks):

* 7066 users are returning customers (0);
* 5736 users are new customers (1); and
* 8259 users are neither (null)

Majority of the users were either new or returning, which is a good thing when we delve deeper into finding correlation between several fields.

***platform***

This column contains the type of device (Android, iPhone, iOS, Windows, etc.) the users use when navigating through the Zappos website to search or buy a product:

* 3172 users were on Android devices;
* 1589 users used a Blackberry device;
* 1349 users were on ChromeOS;
* 3435 users were on iOS devices;
* 459 users were on iPad devices;
* 468 users were on iPhone devices;
* 2036 users used Linux devices;
* 333 users used Macintosh devices;
* 2054 users were on MacOSX devices;
* 327 devices used were Other;
* 74 users were on SymbianOS;
* 1641 devices used were Unknown;
* 2399 devices used were Windows;
* 1315 devices used were WindowsPhone;
* 410 of the devices were blanks

***visits***

The *visits* are integers that represent the number of distinct website visits. Distinct visits mean the different pages or sites a user looks through in Zappos.com. For instance, a user could be viewing a product or several products under the Sortly site but another product(s) under the Widgetry site. These would be considered distinct visits. The number of visits in this data goes from 0 to 136057.

***distinct\_sessions***

The number of distinct session (integer) is how many times a user surfs through the website without leaving desk or in one sitting. One distinct session may have multiple sites. The data values range from 0 to 107104.

***orders***

This is the number of website orders (integer) made by each user (customer transactions). The values range from 0 to 4916.

***gross\_sales***

This field captures the total gross sales for the website orders. A user may not have placed order or bought anything, or they may have placed an order and did not buy it in that session: 9576 fields are blanks.

***bounces***

This is an integer column that refers to the number of visits that only viewed one page. The *bounces* in this data set ranges from 0 to 52598.

***add\_to\_cart***

This numerical column is the number of visits that added a product to cart. Ideally, all visits should end with a purchase or an item in the cart at least. The range of this data values are 0 to 3966.

***product\_page\_views***

This numerical data values are the number of product pages views per session and per visit. The values range from 0 to 146866.

***search\_page\_views***

This is the number of search pages viewed per visit per session.

***conversion\_rate***

This is a calculated measure =’orders’/’visits’

***bounce\_rate***

This is a calculated measure = ‘bounces’/’visits’

***add\_to\_cart\_rate***

This is a calculated measure =’add\_to\_cart’/’visits’