**Analytics Intern Test**

**Ran Tokman**

**Importing the Data**

I used Python to import the data from the csv file in the URL into a Pandas data table.

**A screenshot of a cell phone

Description automatically generated**

**Missing Data**

**A screenshot of a cell phone

Description automatically generated**Using a Python function, it was immediately evident that tremendous amounts of the raw data is null. The yellow color is a visualization of missing data. It would be helpful in future analysis to require the brands to fill the complete information form before they can create a trial account.

**Exploratory Data Analysis**

Using the Tableau tool, I created visualizations to better understand the data given. You can see them below.

**Product Category**

As we can see in the chart below, the Apparel & Jewelry Brands had the highest rate of not upgrading to a paid subscription, by far. Based on this raw data alone, we can assume that this category of brands is the least interested in what Opensponsorship has to offer. That being said, this data does not reflect those brands that created an account and did upgrade to a paid subscription.

If we had a percentage of total accounts, it would be a better indicator than the sum of brands that didn’t upgrade.

A screenshot of a cell phone

Description automatically generated

**Geographical Location**

Based on the raw data, most brands that did not upgrade to a paid subscription are domestic companies. Similar to the “Product Category” chart, it is difficult to make an assumption without have the data with the total brands that created an account. Yes, drastically more domestic brands didn’t upgrade, but what is that sum tell us relative to the total?

A screenshot of a cell phone

Description automatically generated

**Onboarding**

This chart tells us that most brands overwhelmingly did not reach the last step of the onboarding process. It begs the question, did they take the time to actually use the account?

A screenshot of a cell phone

Description automatically generated

**Active Usage**

To try and answer that, I created a chart that illustrates the number of accounts created on a single day vs. the number of accounts that were last active on that day. They are trends are almost identical, meaning that almost as many accounts were deactivated as activated on any single day.

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generatedThe chart below further emboldens the assumption. I created a chart that calculated the actual number of active days in the accounts the brands created. As can be seen here, most brands overwhelming stopped using the account on the same they created them(0 active days). What can we do to encourage them to use the trial run for an extended period of time?