Valiance Assignment

Submitted by:
Bhaskar Arun

Statement of purpose

We are required to analyse the data provided by the client and provide insights on the impact of **Ambassadors, Professional Athletes** and **Wholesale stores** on the sales in each **DMA(Zip Code).**

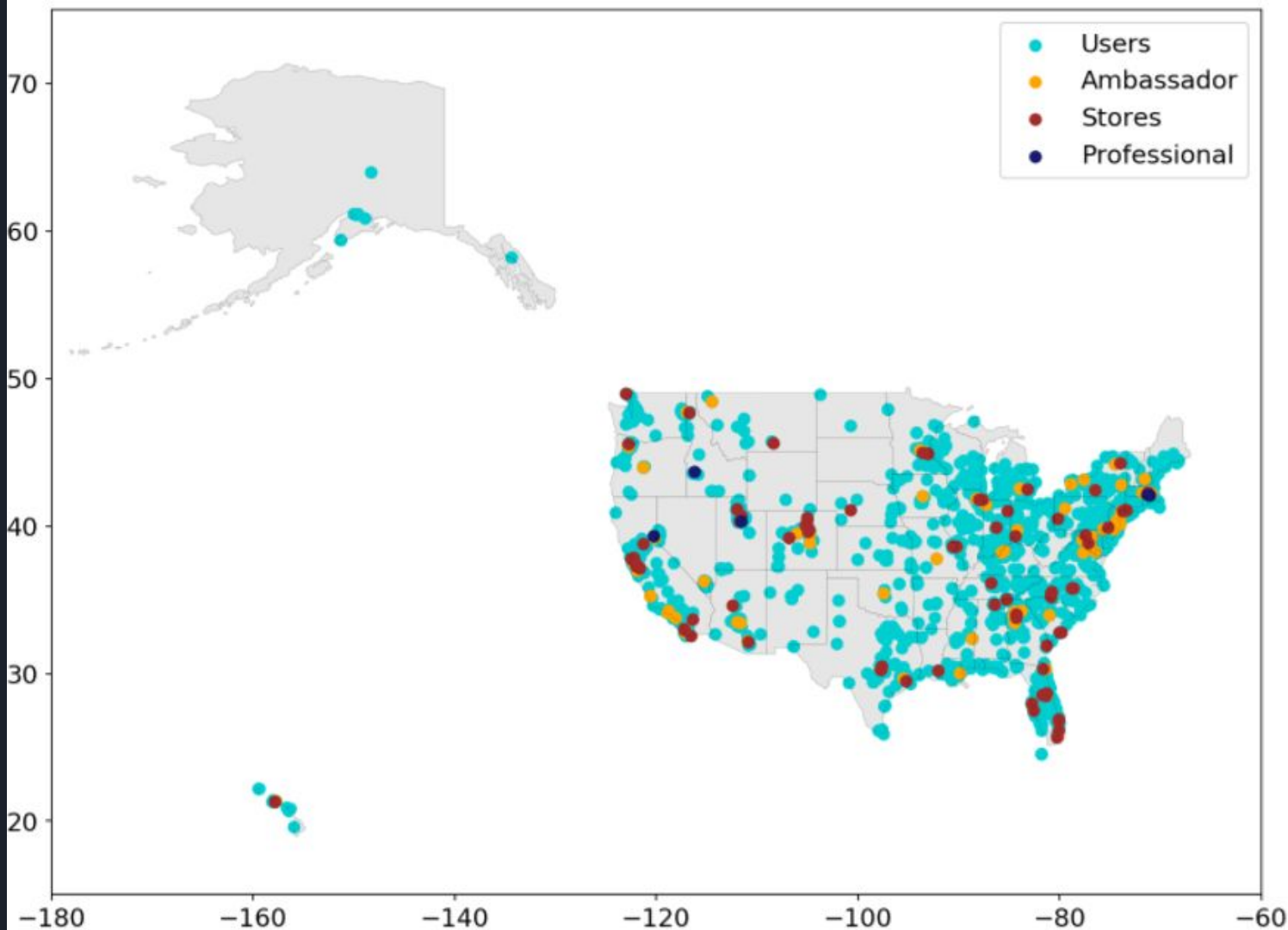
1. Getting the data ready to explore.

- ➔ **Cleaning the Zip Codes and states.**
- ➔ **Getting the Latitude and Longitude Coordinates for each Zip Code.**
  http://federalgovernmentzipcodes.us/
- ➔ **Using a GeoJson to create a map using MatplotLib**

  Used the geoJson from an old project of my own available on github

  https://raw.githubusercontent.com/skepticalbayes/DataScience/master/Intro/data/us-states.json

The geographical locations for all the users, Ambassadors, Professionals and the wholesale stores.



2.Now we analyze the sales data.

- ➔ Separating out the average user from the Ambassadors and Professional athletes as we have to analyze their impact on sales in each DMA so we keep their purchases and spendings aside.
- ➔ Aggregating the purchases and spending in each DMA. We calculate the correlation between spendings and purchases to be able to rely on either metric.
- ➔ We create a dataframe where we have number of Ambassadors within 25 miles of a DMA, number of Ambassadors in a DMA, number of Professionals in a DMA and the number of stores.

First we analyze the effect of an ambassador within 25 miles of a DMA

- We make 2 groups one with no ambassadors within 25 miles and one with one or more ambassadors within 25 miles as denoted this as Nearby Ambassadors in the dataframe.
- Now we calculate the T-test for the means of the two groups. If the p-value is smaller than the threshold, e.g. 1%, 5% or 10%, then we reject the null hypothesis of equal averages.
- We get p-value of 0.00015 confirming the alternate hypothesis.
- The DMAs with an Ambassador within 25 miles drives up the average spending from 250$ to 303$.

Now we analyze the effect of multiple ambassadors within a DMA instead of one.

- ➔ We make 2 groups one with one ambassador and one with more than one ambassadors as denoted by Ambassadors in the dataframe.
- ➔ Now we calculate the T-test for the means of the two groups.
- ➔ We get p-value of 0.58 rejecting the alternate hypothesis.
- ➔ Owing to very few data points(2) and high variance for multiple Ambassadors we can't trust the high average spending in DMAs with multiple Ambassadors.
- ➔ Which lands us to conclude that having multiple ambassadors in a DMA rather than one does not ensure improved sales.

Now we analyze the effect of Professionals within a DMA compared to an Ambassador.

- We make 2 groups one with Professionals one with Ambassadors.
- Now we calculate the T-test for the means of the two groups.
- We get p-value of 0.0012 confirming the alternate hypothesis.
- Both DMAs with a Professional athlete have recorded just one purchase each compared to the average of 3.55 purchases in the DMAs with Ambassadors.
- Even though just 2 DMAs have Professional Athletes we can conclude the impact of Professional athletes to be counter productive.

Now we analyze the effect of Wholesale Stores in a DMA.

- ➔ We make 2 groups one with Stores one without.
- ➔ Now we calculate the T-test for the means of the two groups.
- ➔ We get p-value of 0.0037 confirming the alternate hypothesis.
- ➔ DMAs with a Store have recorded an average of 495$ in spendings compared to 270$ in DMAs without.
- ➔ We can conclude that the stores are definitely driving up the sales in their respective DMAs.

DMAs with Stores have an average of 495$ in spendings and 4.29 purchases

THAT'S MORE THAN THE IMPACT OF AMBASSADORS AND PROFESSIONALS COMBINED

Window of opportunity.

Major Talking points

Professional athletes proving out to be counter productive for sales.

Wholesale distributers are a must with a more uniform distribution Nationwide

What to look into

Clustering of Ambassadors not improving sales

What to look out for

Ambassadors being more uniformly distributed preferably within 25 miles of a DMA drives up sales

What the data is telling us

We should try getting our stores to remote locations like the north-west where sales are active

We should distribute our resources more uniformly instead of clustering them.

We should look into why our sponsored Professional athletes are proving out to be counter productive

Thank you!