

## **Introduction:**

It is known that clothing stores like Kohls, Target and Macys are all of equally good quality. But still we can see many people preferring one upon others. So, the question here is, are the number of people who prefer these stores approximately same or different. If different, how much is the variation?

**Audience:** Any new immigrant to the US would want to shop in big departmental stores like Macys, Target and Kohls especially in the holiday season. Excluding the regular customers for these stores, other people would also be curious to know which store gives good quality products. The highly liked store by the people will always be the first choice for the new immigrants. This information is indicated by the positive tweets for these stores. Secondly, the marketing department of these shops would also want to know this rate to figure out if they are doing better than their competition. This use case might serve as a tool to get feedback from the people.

**Data Collection:** The data is collected through **Streaming API** in python. To get enough data, I had to run the codes separately (at least 2-3 times), for each store and for 2-3 days. I had 3 separate programs doing the same cleaning function and determining the polarity of each tweets for 3 stores. (One program for each store).

**Data Analysis:** Samples of same size (280) were collected for each store. Firstly, **Proportion Tests** were conducted with all the 3 stores (One for each combination of the stores. Altogether, thrice). The results for all **the 3 combinations were alike**. We can say that **Kohls has almost 42%** of the positive comments in our sample, **Macys has 35%** of the positive comments in our sample and **JCPenney has 41%** of the positive comments in our sample. From all the proportions test we have conducted, we can conclude it is simply possible that these stores **might have almost equal customer base**.

Secondly, from 1 sample proportion test on the negative comments for each of the store, we can say that most of the comments received from the people are not negative from which we can conclude that it is reliable to buy merchandize from these stores and the choices are not that bad.

Then, by conducting the chi square test on all the 3 categories (Positive, Negative and Neutral) of comments about each store, we can conclude that comments of all the 3 categories are equally likely.

Finally, from all the 3 tests conducted, we can say that

1. These 3 stores have almost equal customer base.
2. It is reliable to buy from these stores.
3. They have quite a decent choice of clothes.
4. Kohls and JCPenney have comparatively been recommended more.

The plot for the Positive comments (from the sample) is given below.

**Kohls,Macys and JCPenny Postive Analysis**

