Our inspiration stemmed from when we were given this data from Goldman Sachs and told to discover something new. As we began to discuss the data, we began to discover that we wanted to explore a topic in regards to ethnicity/race. Our group is 50% Mexican, 50% white and 100% female. We wanted to use our diverse backgrounds to our advantage in this challenge.

Using the data from Goldman Sachs as well as outside data from the US Census Bureau. Our code analyzes the data and seeks to find if there is a relationship between the proportion of Hispanic residents of a specific zip code and the number of taco/burrito options offered in that same zip code.

We found with this data set that there was no strong correlation between the proportion of the Hispanic population and taco/burrito options. We wanted to further filter the data in hopes of finding some kind of correlation by eliminating chain restaurants. We suspected maybe there would be a correlation between the Hispanic ratio and the number of burrito/taco options from smaller, non-chain, and therefore more authentic restaurants. However, due to time constraints and lack of integrated data manipulation knowledge, we were unable to do so.

We built this data through reading in the data sets and cross-examining the data in order to rearrange into needed subsets of data; the subsets were of the form that we could perform linear regression on.

We then plotted the data in a scatter plot alongside the linear model.

One of the most difficult challenges we ran into is being very unaccustomed to the language and data manipulation. This was all of our first times at a Datathon and we are in the learners' category.

Another challenge was that the data set was mostly of the categorical type of data. This made it hard for us to perform analysis on the data without making any biased grouping.

The time constraint was also a large challenge for us as we needed to use a large amount of time to learning the language.

There were also many spaces in the data set in which were blank and therefore did not provide us with a complete set of data. The variables we first were interested in analyzing, such as ingredients, we mostly incomplete. We also do not know how well the data was collected or how truly random it was.

We were proud to have completed a program that works for our first datathon.

We were also proud of making connections within the coding community and learning new coding methodology.

We learned different ways to perform regression on a set of data whether in the linear, multivariable or logarithmic style through the learning seminars.

We also learned how removing outliers could, in turn, affect correlation.

Hopefully, we will compete in another Hackathon in the future as a group.