Group Project #1.5

(Nelly, Jaber, Hal)

Final Report

**Extract**

1) We selected two different data sources

* 1. From the website Kaggle.com we selected a json file of 2014 311 Call Center data
     1. <https://www.kaggle.com/cityofLA/los-angeles-311-call-center-tracking-data>
  2. From the website data.gov we selected a csv file of 2010 census data
     1. <https://catalog.data.gov/dataset/2010-census-populations-by-zip-code>

**Transform**

1) For the Call Center data, we performed the following cleaning/transformation:

a. We had to change the format of the values in the zip code column top string format from a float format.

b. We eliminated all NaN values from the from the Call Center data

c. We removed any columns we felt unnecessary for our data analysis

d. We renamed columns so that they would be easier to convert into a database

e. We truncated the call center data to 500 rows to make it easier to convert to a database

2) For the Census data, we performed the following cleaning/transformation:

a. We removed any columns we felt were unnecessary for our data analysis

b. We renamed columns because the columns headers for the json data came through as numbers and we wanted to be consistent with how our data looked.

**Load**

1) We decided to load the final database into MySQL because we found relationships between our two data sets and we felt a relational database would best suit our data sets.

2) We will be joining our data sets on zip code since that data is available in both data sets.