**Starbucks Data ETL Project**

*Team Members: Medina & Desiree*

The purpose of this project is to utilize ETL process to help review and analyze information from various sources. ETL stands for Extract, Transform and Load. Therefore, ETL was used to confirm if Starbucks are located in primarily in higher income zip codes.

We obtained the data from two separate csv files to help us with this project. The csv files were pulled from Kaggle and Data.world websites:

Kaggle: https://www.kaggle.com/starbucks/store-locations

Data.world: https://data.world/jonloyens/irs-income-by-zip-code

**ETL**

*Extract*:

Panda library in python was used to read the data from each csv file and review the data that is included in each csv file.

*Transform*:

Once the csv data was extracted and reviewed, we transformed and cleaned the data:

* Removed all data that was not from the U.S.
* Removed rows that contained no values (NaN)
* Converted the Postal Codes in the same format (5-digits)
* Modified the columns
* Joining tables created from each csv file

*Load:*

The files were stored as csv files, so that it be used for additional analysis.

**CHALLENGES**

During this project we did encounter some challenges. The transformation phase consisted of the challenges because:

* It was difficult to get the run the correct python codes for modifying the Postal Codes:
* The Postal Codes were not formatted equally, ranging from 4 to 9 digits
* The leading “0” was dropped from some Postal Codes values