

|  |
| --- |
| ETL TECHNICAL REPORT  2018 |
|  |
| PROJECT 2 – AUGUST 24  Sarah D, Todd G, Ebony P,  Matt F, Shashi P |

# Starbucks to World Happiness

|  |
| --- |
| Below are the sources we chose.  We chose these because the timelines were very close. The Starbucks data included records of each location from February 2017, and the World Happiness data’s research was released for March 20, 2017; the International Day of Happiness.  The data was all formatted as CSV files.  We had to bring in the third Dataset for our Transformation process. |
| **Extract:** For this project we used Kaggle.com to extract our primary Datasets.   1. <https://www.kaggle.com/unsdsn/world-happiness#2017.csv> 2. <https://www.kaggle.com/starbucks/store-locations> 3. <https://www.kaggle.com/koki25ando/country-code> |
| **Transform: Cleaning –** We used pandas to extract and clean these csv files. With many extra variables, we dropped the columns that didn’t pertain to our desired analysis; Things like “Whisker. High” and “Whisker. Low” because we are not making a box plot as one of our later visualizations.  **Formatting –** Using the country code data we found, we joined world data and reformatted our Starbucks Country data from abbreviations to the full country name, so we can later combine them with our World data.  **Load:** Our data was stored in a PostgreSQL database called “SuperUberDuberProject”. Within that are two tables; “happiness” and “store\_directory.”    Final Tables: Conclusion  By looking at these graphs we can see that the highest correlation found is China, but China's happiness ranking is just above the 80th percentile, which is certainly not considered the happiest of over 120 countries. Therefore we learned that Starbucks, sadly, did not actually effect world happiness rankings. |





