**Understanding-Millennials**

To better understand this generation, we ran an analysis comparing avocado sales vs. Real estate data sets to see if Millennial are truly spending all their money on avocado toast.

[](https://raw.githubusercontent.com/mmsquier/Understanding-Millennials/master/Millennial-meme.PNG)

**Data Gathering**

To conduct the study a dataset created by Haas Avocado Board was obtained using Kaggle and a Homeownership dataset from the United States Census.

* <https://www.census.gov/housing/hvs/data/ann17ind.html>
* <https://www.kaggle.com/neuromusic/avocado-prices>

**Data Transformation**

After downloading the CSV files and looking at the data we noticed that the datasets did not share the same regional labeling. We decided to go with the US Census region categories for it seemed a formal classification. The Avocado dataset City-State classification was altered to match the US Census dataset.

The next step was to create Jupyter notebook and upload the datasets. In the notebook we cleaned up the data utilizing:

* group\_by ()
* sum()
* rename()
* reset\_index()
* melt()

**Loading Data**

After cleaning up the data, it was imported to Postgress SQL software using the steps below:

Two tables were created the "understanding\_millenials" database utilizing CREATE TABLE. Afterwards, the Jupyter notebook was connected to SQL and created an engine. After creating the connection we confirmed that the tables where in the database using engine.table\_names().

The dataframes, "new\_avocado\_df" and "homeowner", from the jupyter notebook were loaded into the "understanding\_millenials" database utilizing the .to\_sql(). With .read\_sql\_query('select\* ' ', con=engine) to pull and read from SQL the database using pandas.