Data Process & Report

# Visa Passport Data

## Python

#### Steps

1. Import dependencies – pandas, sqlalchemy, numpy, datetime, config
2. Import raw files from Kaggle, “us\_perm\_visas.csv”
3. Remove unnecessary columns
4. Combine similar columns
5. Calculate adjusted Yearly Wage
6. Remove duplicated columns
7. Set Case Number as index
8. Export copy of reduced version of csv
9. Make connection to PGAdmin to read in reduced data into a table
10. Additional cleaning of cells

## PGAdmin

#### Steps

1. Create database ETL\_project
2. Create table of variables

# Census Data

## Python

#### Steps

1. Import dependencies – pandas, numpy, config files, sqlalchemy, config
2. Import raw file from Census, “ASE\_2016\_00CSCBO01.csv”
3. Remove unnecessary columns
4. File had two rows of headers, removed top headers and made second row of header primary headers
5. Printed name of remaining columns
6. Removed spaces and reduced verbosity of column names to fit in SQL data
7. Separated files into integer columns and character columns
8. Made copies of character columns into categorical columns
9. Printed out column names with SQL syntax, indicating as numeric or character
10. Exported copy of file as CSV

## PGAdmin

#### Steps

1. Created table with file exported from Python
2. Created crosswalk files between categorical data and corresponding category numbers
3. Created reduced table that dropped text versions of categorical data
4. Exported csv file with numeric versions of categorical data
5. Exported csv files of all crosswalk files