

Names of Applicants of Social Security Issued at birth in USA

Project prepared for CEBD 1160
By Wilson Lee



Project Objective

- Find out the most used name in USA history
- The distribution of the name among the different state
- See the change in popularity of a given name



Data Source

The data for this project is provided by the USA Social Security Administration.

The database consists of the number of Social Security Card applications at birth for a given name.

Website Link:

<https://www.ssa.gov/OACT/babynames/limits.html>



Provided Data Format

The provided data is a set of text files, 1 file per state, and is comma delimited.

The application count is aggregated by

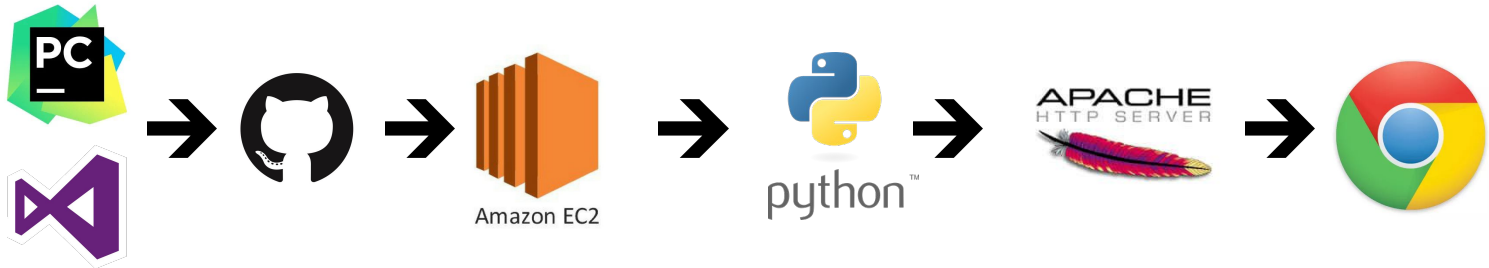
- State of application
- Gender of the applicant
- Year of birth of the applicant
- Name of the applicant

Example Entry:

AK,F,1910,Mary,14

Project Development Process

- Python script and web page developed locally
- Pushed to github repository
- Pull to Amazon's AWS EC2 machine
- Run Python script on EC2 to clean and aggregate data
- Host web page on EC2 through Apache2
- Debug web page through Chrome





Why AWS EC2 + Web Server

Available options:

- S3 + Static web hosting
 - Con: Limited to 2,000 put requests for free tier. Generated database had more than 31,000+ file
- Award Space + web hosting
 - Con: I tried to upload all fills and was blocked for uploading after 10,000 files

Solution:

- Upload the source code and original data set to EC2 and generate the 31,000+ files on EC2.
- Create a web server similar to Airflow



Project Output

The following webpage is the deliverable of this project:

<http://ec2-18-222-252-207.us-east-2.compute.amazonaws.com/c/bd1160/>

This web page consist of the following

- A chart with the the total occurance of a selected name in USA over the time frame of the data
- The number of occurance of the name in a given year per year
- The “Top Ten” occurance of male, female and unisex names



GitHub Repository

The GitHub Repository for this project is found in the following link:

https://github.com/wlee0515/US_Name_Per_State_Project



Questions?