**Project on Data Science with Python**

**Objective**–The United States Social Security Administra4on (SSA) has made available data on the frequency of baby names from 1880 through 2016, this data set is often used in illustrating data analytics. The project requires analyzing the data considering different methods.

**The description of the dataset is mentioned below**

* The data can be obtained at:

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

The format of data is as given below.

**Name, Male/Female, Number of babies**

Mary,F,7065

Anna,F,2604

Emma,F,2003

Elizabeth,F,1939

**Problem statement:**

* Visualize the proportion of babies given for a particular name
* Determine the naming trend
* Determine the most popular names in each year

**Lab Environment**: We will require to install python 3 along with the all related libraries like numpy , pandas , matplotlib etc.

**Domain** – This problem statement belongs to Regional domain.

**Hint for doing Project Work**

The analysis of given data can be done using following points as hint.

* Check the Data
* Load Data – Library “Pandas” read CSV can be used with loop, so all data will be in single data frame.
* Grouping the data based on sex – Data Frame Manipulation
* Aggregate the data at the year and sex level – SQL like data frame manipulation
* Prepare the data for plot – Matplotlib with Bar and box plots
* Plot the total births by sex and year - – Matplotlib with Bar and box plots
* Reorganize the data
* Analyzing Naming Trends