**Assignment Write-Up – Fortella.AI**

**Title:** Data Analysis on Baby Names

**Overview:** It is a data set of baby names since 1880 derived from Social Security applications. It lists the first name and gender.

**Introduction and Purpose:** The dataset contains names of individuals born every year from 1880 until 2018. The purpose of doing this analysis is to predict how many individuals could be born over the next few years, is there a rise or fall in unisex names (male / female sharing the same first name), how many children could be possibly born with the same name in the next year.

**Analysis and Approach:** Following are few observation I made before trying to solve two small use-cases using basic Machine Learning algorithms.

1. Individuals (Male/Females) born year over year in the past 140 years.
2. Analysis showing most occurring Male/Female names.
3. Analysis showing the trend of how many individuals were born decade over decade.
4. Most common ‘Unisex Names’ found in the dataset over past 140 years.
5. Analysis shows the trend of how percentage of Unisex names are decreasing over the years.
6. Predict the number of children to be born in the year 2019.

This analysis shows a forecast of how many individuals could be born in the next year. Using Simple Linear Regression, this analysis was implemented using Independent Variable ‘Year’ and Target Variable ‘Names Occurences’.

1. Predict at least 50 names with the number of occurrence who will share the same name.

Created a dataframe called ‘recentNames’ which contains the data for the last 5 years (2014 onwards) from the parent dataset ‘babyDataset’. Using one for loop I iterated over the data and the 2nd for loop I prepared a dataset of a single name and number of occurrence of that name in the past 4 years. After following the above approach, using Random Forest algorithm I predicted the number of individuals who will born with the same names in the year 2019.

**Machine Learning Algorithms used:**

1. Linear Regression
2. Random Forest

**Conclusion and Future Work:**

Attached python projects will definitely give you an idea of how many individuals were born and what names were given to them over a course of 140 years. The analysis will also give a forecast of how many individuals will be born in the next year and what names could be possibly given to them.

Few projects that could be done with the same dataset are as follows.

1. Predict from name, if male or female.
2. Predict for future if there would be more males or females.