**Infosys Limited**

<Population projection of Government data>

**Requirements Specification Document**

<27-Jan-2022>

**REVISION LIST**

| Ver.Rev | Date | Authors | Description |
| --- | --- | --- | --- |
| 1.0 | 27-Jan-2022 | Renu Sri Gandham | Base version |

***TABLE OF CONTENTS***

1. PROJECT / PRODUCT OVERVIEW 3

2.DATSET DESCRIPTION…………………………………………………………………………………….....3

3. USER STORIES 3

3. HARDWARE AND SOFTWARE REQUIREMENTS 4

a. Deployment Environment Requirements 4

b. Development Environment Requirements 4

# PROJECT / PRODUCT OVERVIEW

As part of a country’s population estimate data, the respective government has collated the statistics for previous years and want to make projection for subsequent year i.e 2020.

Population projection is important since it helps people i.e government, researchers, make decisions about the future. Here are some possible ways that people use the result of population projection. 1. Estimate the basic need for human, such as demand for food, water, power, transportations. The company has given the data and has advised the team to build a model which can predict the data for subsequent year.

1. **Dataset Description:**

The population data has details for males and females for different age brackets [ Age 0 to 90+ years] is captured for 39 consecutive years (1981 – 2019) as per the process. The dataset representing the process measurement points is described as follows:

* File names

mid-year-pop-est-20-time-series-3\_1981.csv

mid-year-pop-est-20-time-series-3\_1982.csv to ……

mid-year-pop-est-20-time-series-3\_2019.csv

* Data Set Characteristics: Multivariate
* Area: Computer
* # of Records: 30 records from each file
* Number of Attributes:95

**Problem statement:**

Build a machine learning time series model which would help forecast/ project the population estimate for subsequent year i.e 2020 based on the current and historic stats information.

# USER STORIES

***Read more about how to write User Stories, examples, explanation etc.*** [***here***](https://lex.infosysapps.com/toc/lex_29125677084960174000/about) ***and*** [***here***](https://apc01.safelinks.protection.outlook.com/?url=http%3A%2F%2Frgalen.com%2Fagile-training-news%2F2013%2F11%2F10%2Ftechnical-user-stories-what-when-and-how&data=02%7C01%7Ckomal_papdeja%40infosys.com%7C62533c0c7f954f19a02208d7592536a1%7C63ce7d592f3e42cda8ccbe764cff5eb6%7C1%7C0%7C637075888684136902&sdata=%2BK7pe46UJomLPfE1EAh%2B%2FXPfIR4qLB1AHKTLnmGTEnQ%3D&reserved=0)

Suppose you are in that technical team, and you are asked to do the following task:

|  |  |  |  |
| --- | --- | --- | --- |
| **USER STORY ID** | # | **PRIORITY** | Should have completed below |
|  | **1** | **HIGH** | Using panda’s library in Python, extract the year information from each dataset and make it as a column in the dataset.  Arrange the data from multiple files into one single by melting and casting data sets |
|  | **2** | **HIGH** | Perform EDA to analyze past behavior of the population data  Make a note of obvious temporal structures, like   * 1. trend   2. seasonality   3. irregularity   4. Cyclic pattern   5. anomalies like missing data * Check for the stationarity in the data for existence of constant mean and constant variance. * Perform AIC and BIC tests for stationarity |
|  | **3** | **HIGH** | Build Machine learning model- 1:   * Plot ACF and PACF plot to help better estimate (p,d,q) for ARIMA model * Train an ARIMA (Auto Regressive + Integration + Moving Average) model and evaluate it. * Acceptance criteria for ML model accuracy is 70% or more. |
| **Acceptance Criteria** | | | |
| * *The quality Output from EDA suitable for model building is the first stage acceptance criteria.* * *ML model accuracy is the second stage acceptance criteria.* * *Report or Visual Dashboard is the third stage acceptance criteria.* | | | |
|
|
|
| **Any Other Information** | | | |
| NA | | | |
|
|

# HARDWARE AND SOFTWARE REQUIREMENTS

## Deployment Environment Requirements

**Hardware requirements:**

* Disk Space: 16 GB or more
* Processor: 1.4 GHz or more 32 bit/64 bit
* Memory: 4 GB or more

**Software requirements: Python anaconda**

Download and install the software from <https://www.anaconda.com/products/individual>

**Power BI:** Download and install the software from url

<https://app.powerbi.com/SignupRedirect?ru=https%3A%2F%2Fapp.powerbi.com%2Fhome>

## Development Environment Requirements

Use Ijupyter or Spyder installed as part of anaconda download.