

Configuration Manual

MSc Research Project
MSc in Data Analytics

Rohit Puranik
Student ID: x22165967

School of Computing
National College of Ireland

Supervisor: Prof. Arjun Chikkankod

National College of Ireland
MSc Project Submission Sheet
School of Computing



Student Name: Rohit Puranik.....

Student ID: X22165967.....

Programme : Data Analytics..... **Year:** 2023-24

Module: MSc Research Project

Lecturer:
Submission Due Date: 14th December 2023
.....

Project Title: Configuration Manual
Word Count: 490..... **Page Count:** 6.....

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Rohit Puranik

Date: 14th December 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	<input type="checkbox"/>
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	<input type="checkbox"/>
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	<input type="checkbox"/>

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	

Penalty Applied (if applicable):	
----------------------------------	--

Configuration Manual

Rohit Puranik
X22165967

1 Introduction

This document contains all the information, for implementing the project titled "Examining Income Disparity between Urban and Rural Counties in Ireland and Predicting Income Disparity Among Irelands Counties." This guide focuses on the stages of the code starting from data collection to evaluating the model building phase.

2 Hardware Requirement

The project was developed on a 64 bit Windows operating system with 16 GB of RAM. Figure 1 displays the system specifications. It is not mandatory to have high end specifications for this project; a processor lower, than an i7 would also suffice.

Device specifications	
Device name	ROHIT
Processor	11th Gen Intel(R) Core(TM) i5-1135G7 @ 2.40GHz 2.42 GHz
Installed RAM	16.0 GB (15.7 GB usable)
Device ID	406BDFAB-A36F-4C14-9D01-3E8A61219584
Product ID	00330-53903-59621-AAOEM
System type	64-bit operating system, x64-based processor
Pen and touch	No pen or touch input is available for this display
Related links Domain or workgroup System protection Advanced system settings	
Windows specifications	
Edition	Windows 11 Pro
Version	22H2
Installed on	1/11/2023
OS build	22621.2861
Experience	Windows Feature Experience Pack 1000.22681.1000.0

Figure(1) System Specification

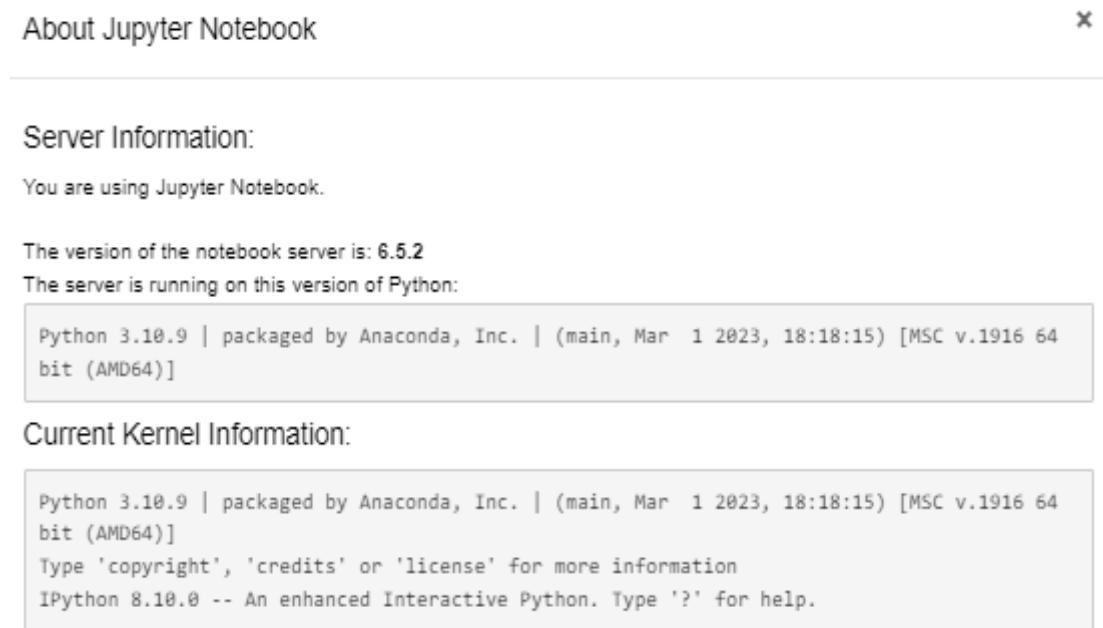
3 Software Requirement

Jupyter notebook is used as an integrated development environment, where in it needs to be restarted by triggering the command jupyter notebook in the command prompt

```
C:\Users\Rohit>jupyter notebook
[W 10:25:18.124 NotebookApp] Loading JupyterLab as a classic no
[I 2023-12-14 10:25:18.147 LabApp] JupyterLab extension loaded
[I 2023-12-14 10:25:18.147 LabApp] JupyterLab application direc
[I 10:25:36.203 NotebookApp] Serving notebooks from local direc
[I 10:25:36.204 NotebookApp] Jupyter Notebook 6.5.2 is running
[I 10:25:36.204 NotebookApp] http://localhost:8888/?token=0d18d
```

Figure(2) Command prompt

The research was conducted on the the below verrsion of Jupyter Notebook as shown in Figure(3)



Figure(3) Jupyter Notebook

The following version of the Python shown in Figure(4)

```
In [2]: 1 !python --version
Python 3.10.9
```

Figure(4) Version of Python used

For visualization purpose PowerBI reporting tool is used which is as per the below version

Microsoft Power BI Desktop

Microsoft Power BI Desktop is a companion product to app.powerbi.com.

Version: 2.123.742.0 64-bit (November 2023)

User ID: 7223837a-b027-4835-bee0-0ffd2a1e5028

Session ID: 663437b8-aecf-48a0-8945-eade0d49a281

Copy session diagnostics to clipboard

Copy

[Privacy Statement](#)

Figure(5) PowerBI software version

Following python libraries are used for data processing & data transformation

```
1 import numpy as np
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 import seaborn as sns
5 import scipy.stats as stats
6 from pandas.plotting import scatter_matrix
7 from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, roc_auc_score, roc_curve
8 from sklearn import linear_model
9 from sklearn.model_selection import train_test_split
10 from sklearn import metrics
11 from sklearn.metrics import r2_score
12 from sklearn.feature_selection import SequentialFeatureSelector
13 from sklearn import preprocessing
14 from sklearn.model_selection import cross_validate
15 from sklearn.metrics import explained_variance_score
16 from sklearn.metrics import mean_absolute_error
17 from sklearn.preprocessing import StandardScaler
18 from sklearn.linear_model import LogisticRegression
19 from sklearn.neighbors import KNeighborsClassifier
20 from sklearn.tree import DecisionTreeClassifier
21 from sklearn.ensemble import RandomForestClassifier
22 from sklearn.svm import SVC
23 from sklearn.ensemble import GradientBoostingClassifier
24 from sklearn.naive_bayes import GaussianNB
25 from sklearn.utils import resample
26 import psycpg2
27 import psycpg2
28 import matplotlib.pyplot as plt
29 import mplotcursors
30 from scipy.stats import ttest_rel
```

Figure(6) Python Libraries

Relational database requirement: PostgreSQL is essentially used to store data of the output

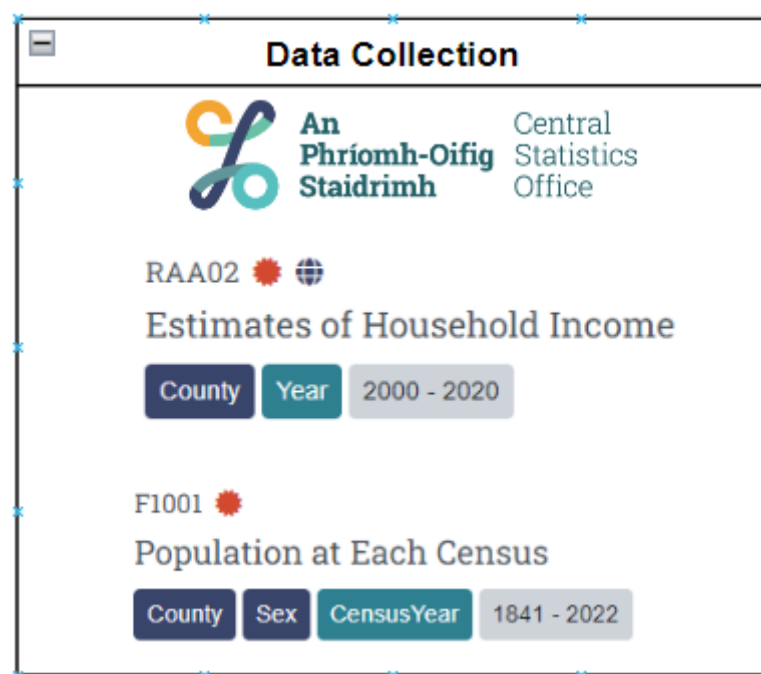
About pgAdmin 4

Version	6.21
Application Mode	Desktop
Current User	pgadmin4@pgadmin.org
NW.js Version	0.55.0
Browser	Chromium 92.0.4515.107
Operating System	Windows-10-10.0.22621-SP0
pgAdmin Database File	C:\Users\Rohit\AppData\Roaming\pgadmin\pgadmin4.db
Log File	C:\Users\Rohit\AppData\Roaming\pgadmin\pgadmin4.log
Server Configuration	

Figure(7) pgAdmin 4

4 Data Requirements

The study requires 2 dataset given and published by Center of statistics Ireland to be downloaded in .xlsx format and kept in a folder



Figure(8) Data Sets

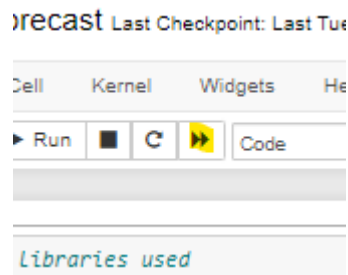
5 Initial Setup

All the file paths as per below code snippet needs to be updated

```
1 # Specify the file path to population dataset
2 DS2_file_path_input = 'C:/Users/Rohit/Downloads/FY001.20231029T171018.xlsx'
3 sheet_name = 'Unpivoted'
4 df_population = pd.read_excel(DS2_file_path_input,sheet_name=sheet_name)
```

Figure(9) Data Sets

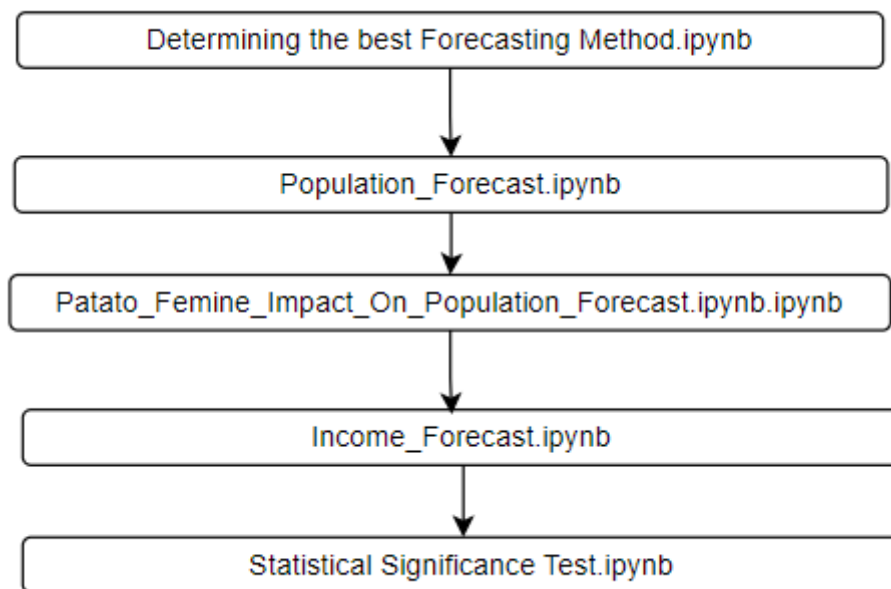
Once all the paths are changed all the jupyter notebooks need to be triggered by execute button highlighted in yellow



Figure(10) Data Sets

6 Execution Order

The following .ipynb files need to be executed in the below order



Figure(11) Execution order