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)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
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.	□

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

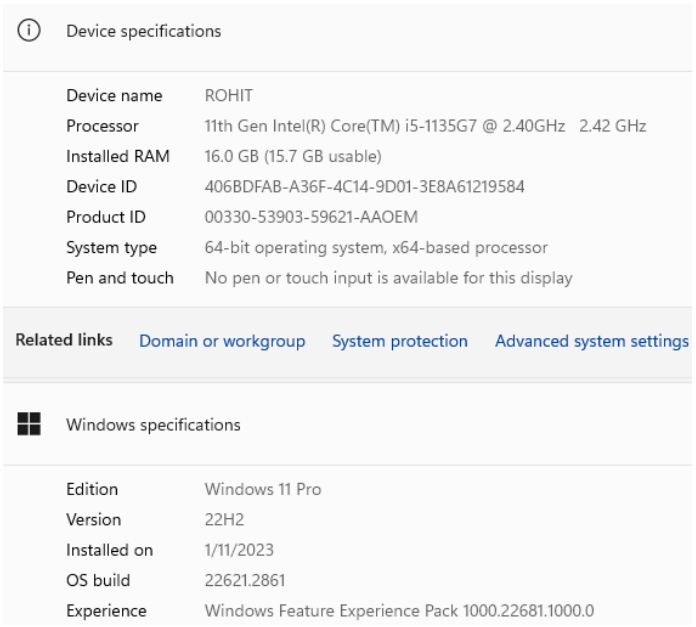
Rohit Puranik
X22165967

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

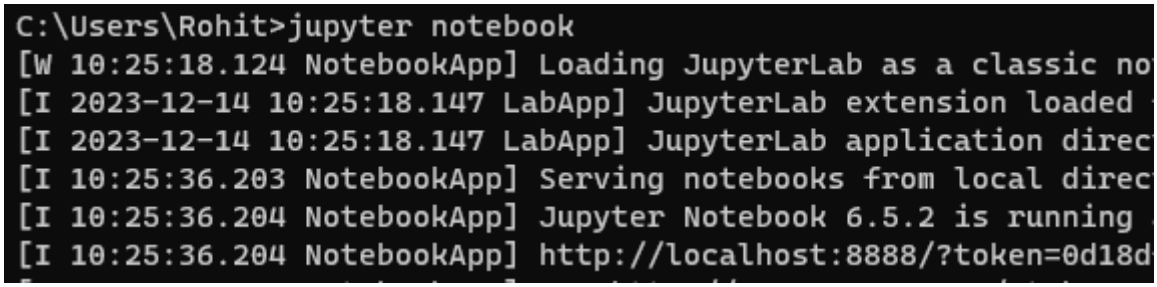
The project was developed on a 64 bit Windows operating system with 16 GB of RAM. Figure 1 displays the system specifications. It is not mandatory to have high end specifications for this project; a processor lower, than an i7 would also suffice.



Figure(1) System Specification

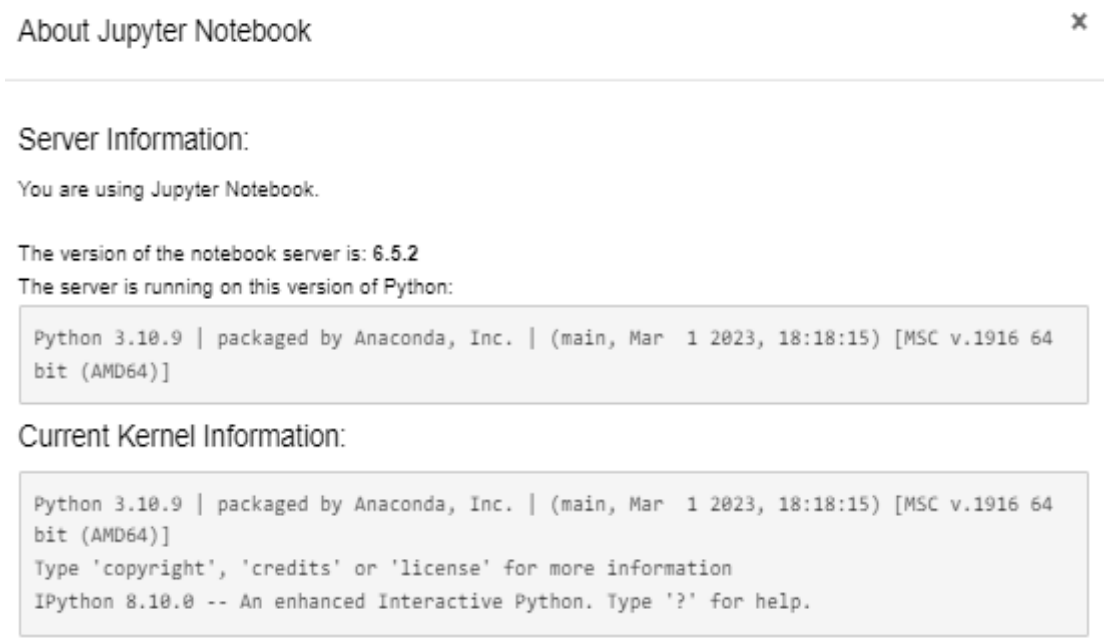
3 Software Requirement

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



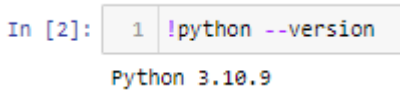
Figure(2) Command promt

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



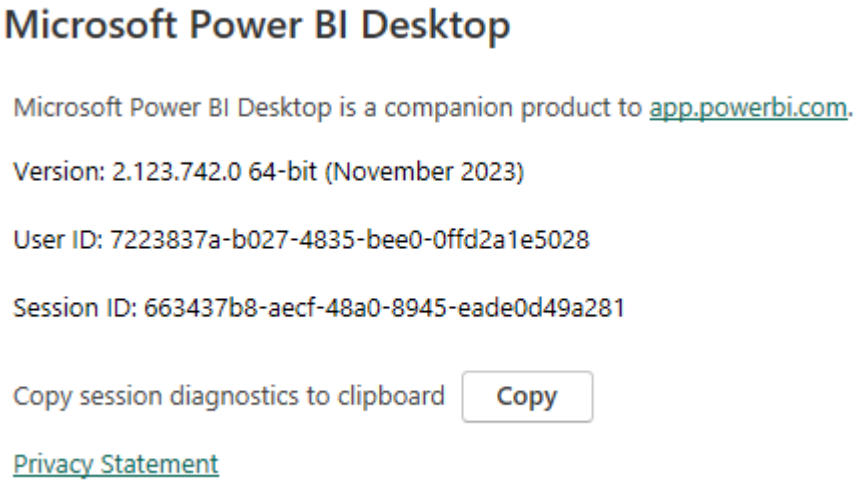
Figure(3) Jupyter Notebook

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



Figure(4) Version of Python used

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



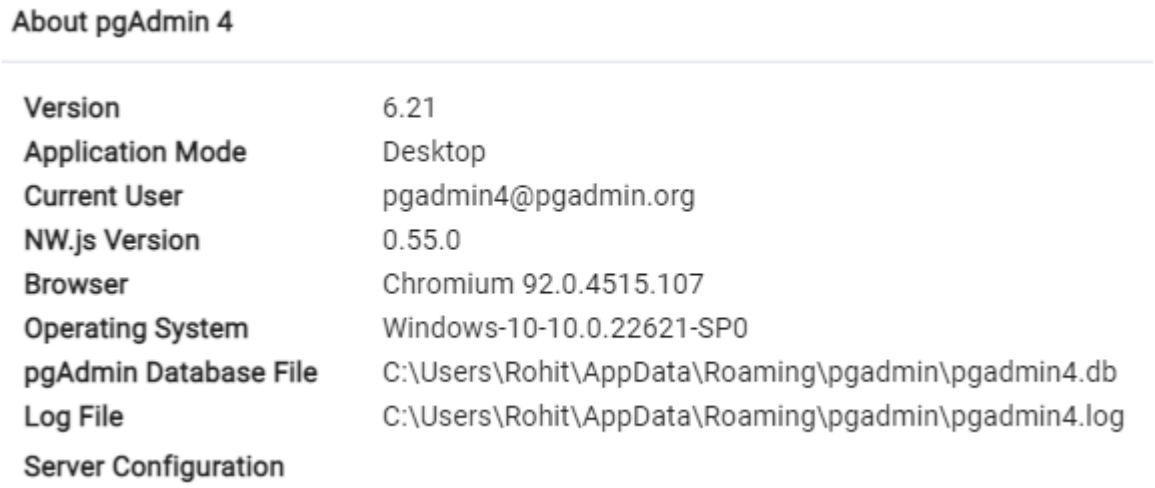
Figure(5) PowerBI software version

Following python libraries are used for data processing & data transformation

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import scipy.stats as stats
from pandas.plotting import scatter_matrix
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, roc_auc_score, roc_curve
from sklearn import linear_model
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.metrics import r2_score
from sklearn.feature_selection import SequentialFeatureSelector
from sklearn import preprocessing
from sklearn.model_selection import cross_validate
from sklearn.metrics import explained_variance_score
from sklearn.metrics import mean_absolute_error
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.utils import resample
import psycopg2
import psycopg2
import matplotlib.pyplot as plt
import mplcursors
from scipy.stats import ttest_rel
```

Figure(6) Python Libraries

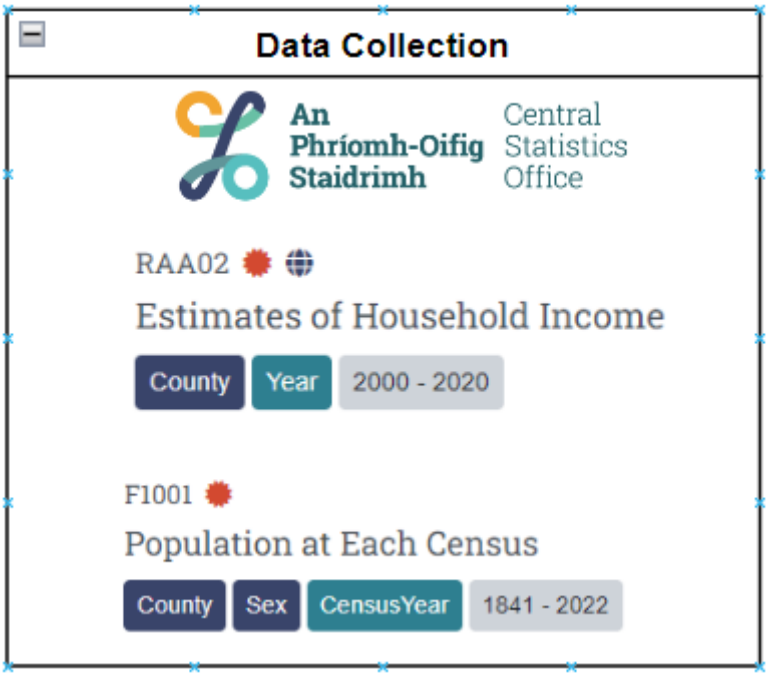
Relational database requiement: PostgreSQL is essentially used to store data of the output



Figure(7) pgAdmin 4

4 Data Requriments

The study requies 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

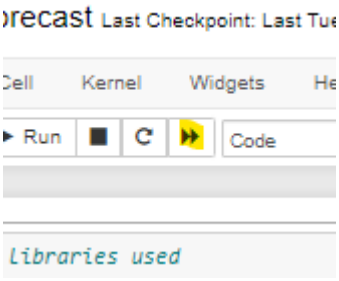
5 Initial Setup

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

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

Figure(9) Data Sets

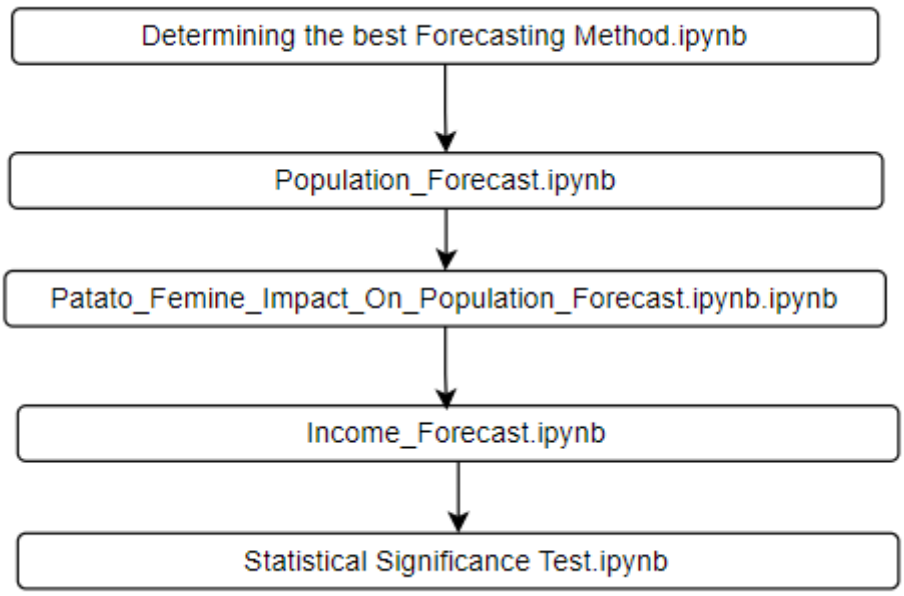
Once all the paths are changed all the jupyter notebooks need to be triggered by execute button highlighted in yellow



Figure(10) Data Sets

6 Execution Order

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



Figure(11) Execution order