National College of Ireland

The Prediction and Optimization of Smart Energy Usage through Machine Learning Recommendations

MSc Research Project
Data Analytics

Mark McGrane
Student ID: x19140606

School of Computing
National College of Ireland

Supervisor: Dr. Catherine Mulwa

National College of Ireland
Project Submission Sheet
School of Computing

National College of Ireland

Student Name:	Mark McGrane
Student ID:	x19140606
Programme:	Data Analytics
Year:	2021
Module:	MSc Research Project
Supervisor:	Dr. Catherine Mulwa
Submission Due Date:	16/08/2021
Project Title:	The Prediction and Optimization of Smart Energy Usage through Machine Learning Recommendations
Word Count:	XXX
Page Count:	31

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:	
Date:	21st September 2021

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):	

The Prediction and Optimization of Smart Energy Usage through Machine Learning Recommendations

Mark McGrane
x19140606

1 Hardware Requirements and Technologies Used

Before implementing this project, please ensure that your local machine has the minimum memory requirements and operating specifications as detailed in Table 1. It may be necessary to zoom into 125% or higher to see the definition of some of the screenshots. If the system requirements and software specified within this section are already installed on the local machine, it is possible to move to Section 7

Table 1: Device and Operating Specifications

Device Processor	AMD Ryzen 7 3700U with Radeon Vega Mobile Gfx 2.30 GHz
Installed RAM	16.0 GB (13.9 GB usable)
System Type	64-bit operating system, x64-based processor
Windows Edition	Windows 10 Home
Windows Version	20H2
Windows Operating System Build	19042.1165
Windows Experience	Windows Feature Experience Pack 120.2212.3530.0

The practical element of the report will use various technologies and programming languages. There will be the need to download and configure these on a local system to display output. The following will need to be implemented.

- R
- R Studio
- PostgreSQL
- Anaconda and Jupyter Notebook
- Tensorflow and Keras

2 R

The language R is used for data cleaning, analysis and statistical generation. Downloading and installation is demonstrated in the following two chapters.

2.1 Downloading R from www

Figures 1-4 demonstrate how to download R from the internet. Open any web browser and navigate to www.r-project.org. Once there, click on the "to download R" hyperlink, which should take you to the web address cran.r-project.org/mirrors.html. Please see Figure 1 and Figure 2. The remainder of the steps are based on the assumption that the download and installation are to a windows machine described above in Section 1.

The hyperlink "Download R for Windows" should be clicked and then saved in the normal folder used for downloads on the local device. Please see figures Figure 3 and Figure 4.

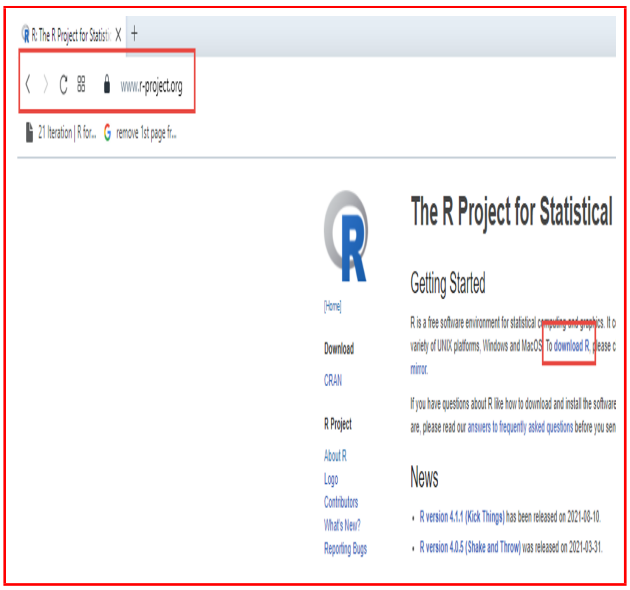

Figure 1

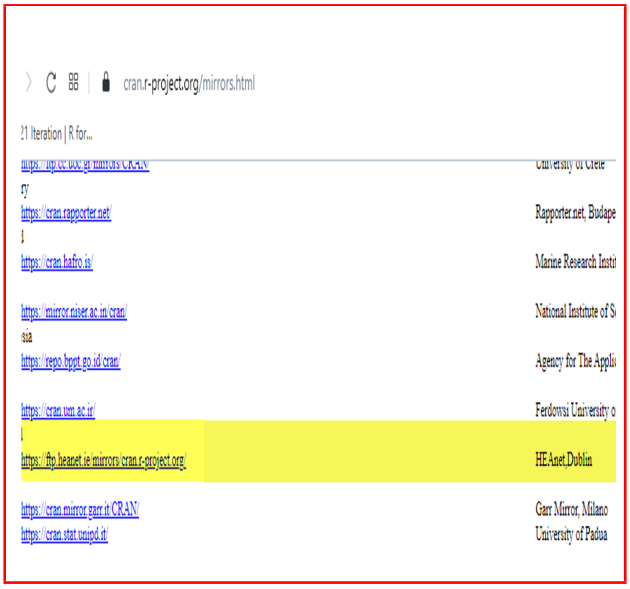

Figure 2

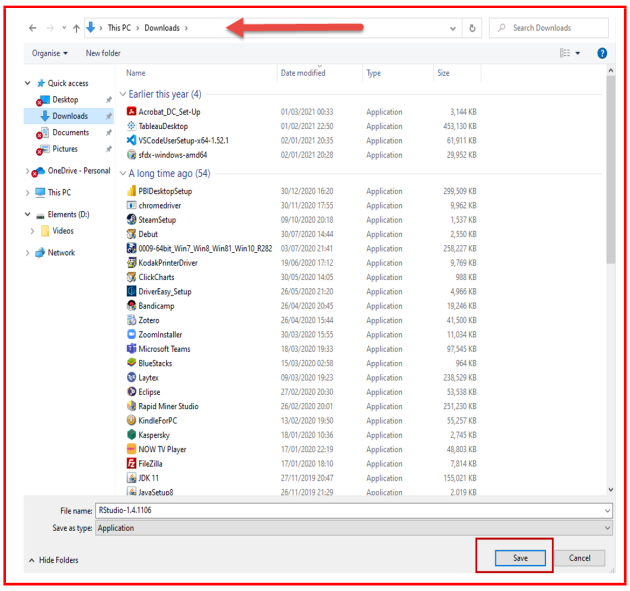

Figure 3

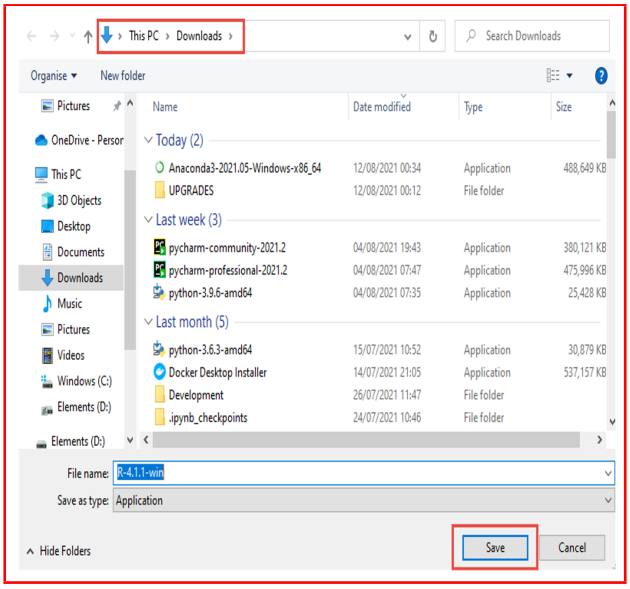

Figure 4

2.2 Installing R locally

Once downloaded entirely, double click on the application to begin the installation process, as shown in Figure 5. Figure 6 shows how to select an alternative file path to install the source files if that is required. Unless specifically mentioned, the default path is used to install all software applications within this project.

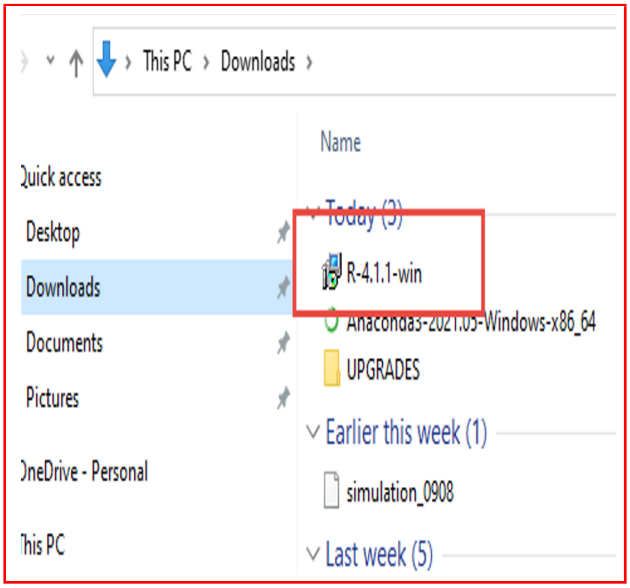

Figure 5

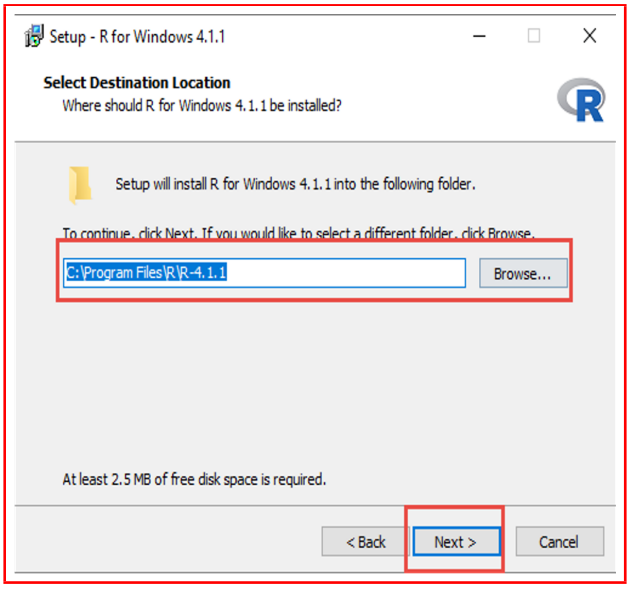

Figure 6

Figures Figure 7 ,Figure 8 ,Figure 9 , and Figure 10 demonstrate the last steps of the implantation. If the four options are not checked as demonstrated in Figure 7, manually select them before continuing. The accepted defaults were used as shown in Figure 8. Whilst a customized startup is a viable option; it is not advised as it deviates from the steps taken in the original implantation. Figure 9 and Figure 10 can be customized as desired. Even though a desktop shortcut and quick launch shortcut were originally chosen, selecting these should not impact results.

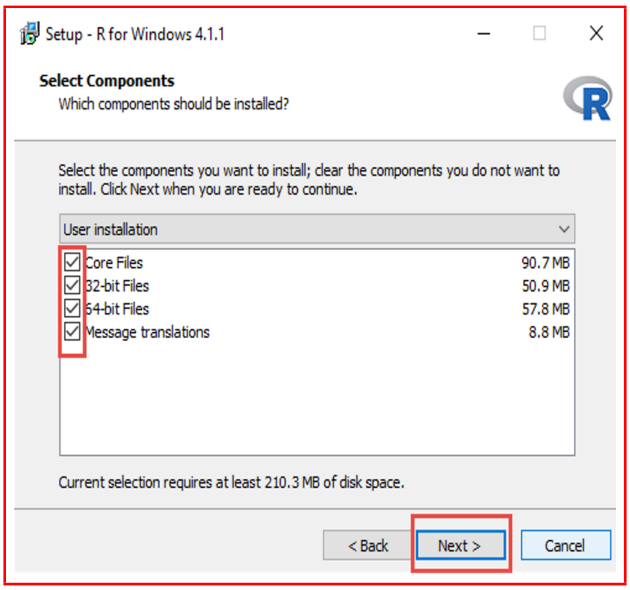

Figure 7

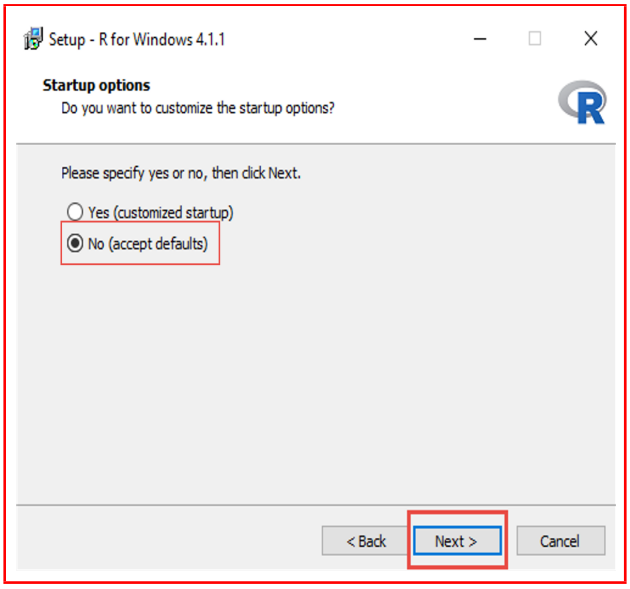

Figure 8

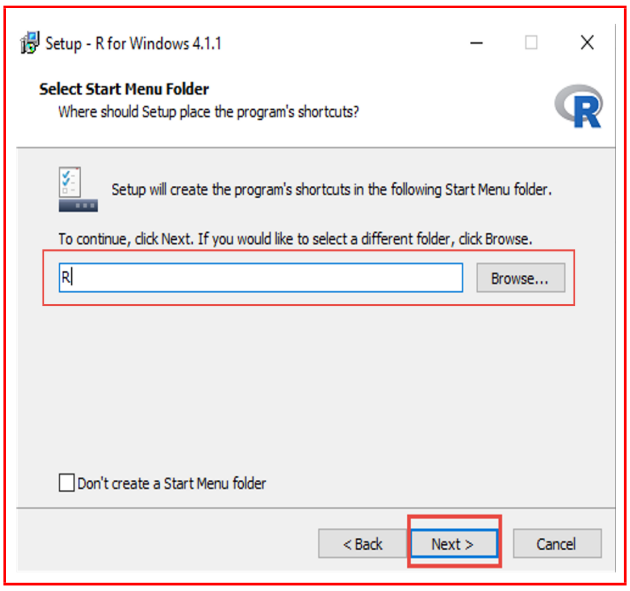

Figure 9

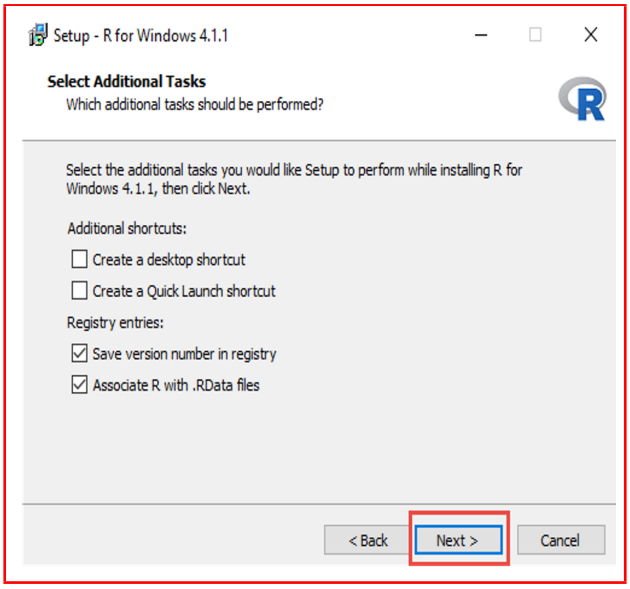
Setup - R for Windows 4.1.1
Select Additional Tasks
Which additional tasks should be performed?
Select the additional tasks you would like Setup to perform while installing R for Windows 4.1.1, then click Next.
Additional shortcuts:
Create a desktop shortcut
Create a Quick Launch shortcut
Registry entries:
Save version number in registry
Associate R with .RData files
< Back
Next >
Cancel

Figure 10

3 R Studio

3.1 Downloading R Studio from www

R Studio is an open-source Integrated Development Environment (IDE) that provides a front end client to develop and run R code. R Studio can be downloaded from www.rstudio.com/produts/rstudio/downloads. Several paid subscription products can be used, but the free version highlighted in Figure 11 was more than capable of the project requirements. There are several potential downloads for the system in question, as demonstrated in figure Figure 12. Please download the most applicable to you. The file should be saved in the standard folder used for downloads on the local device, as shown in Figure 13.

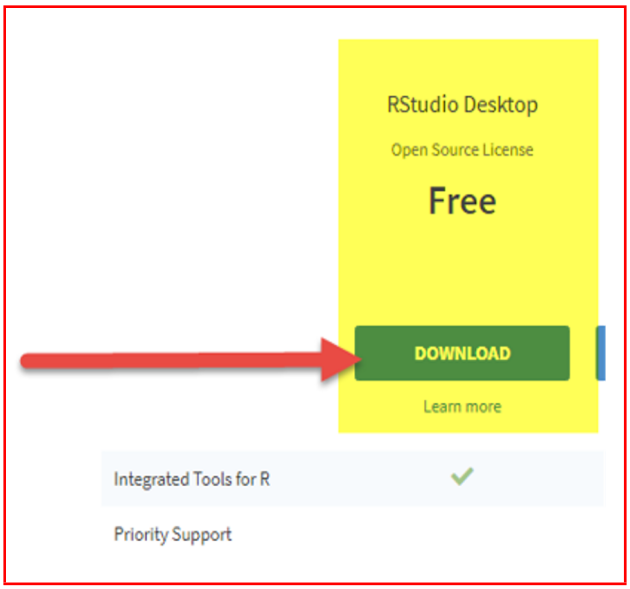

Figure 11

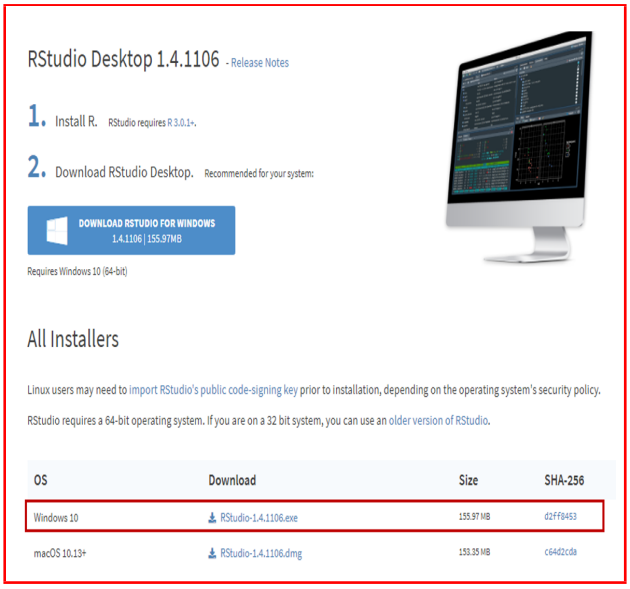

Figure 12

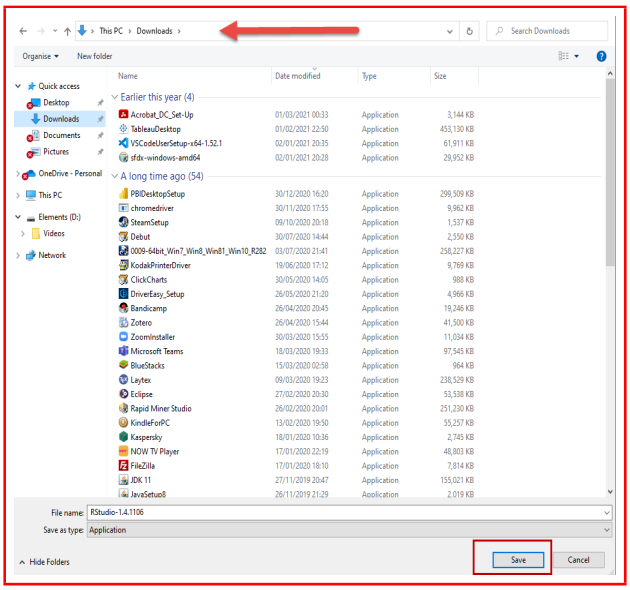

Figure 13

3.2 Installing R Studio locally

Once downloaded entirely, double click on the application to begin the installation process, as shown in Figure 14. Figure 15 and Figure 16 shows how to select an alternative file path to install the source files or rename the start menu option if required. Once again, it is noted that the default options are used for this installation. The remainder of the wizard completes the installation. Nothing was changed from the default from this juncture onwards.

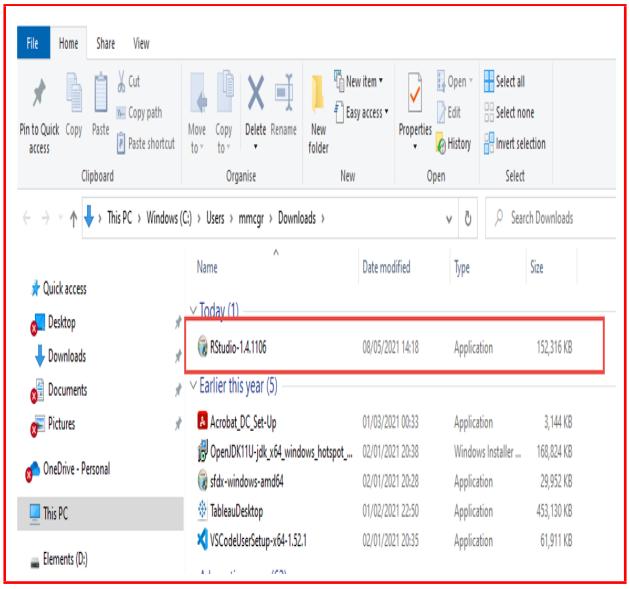

Figure 14

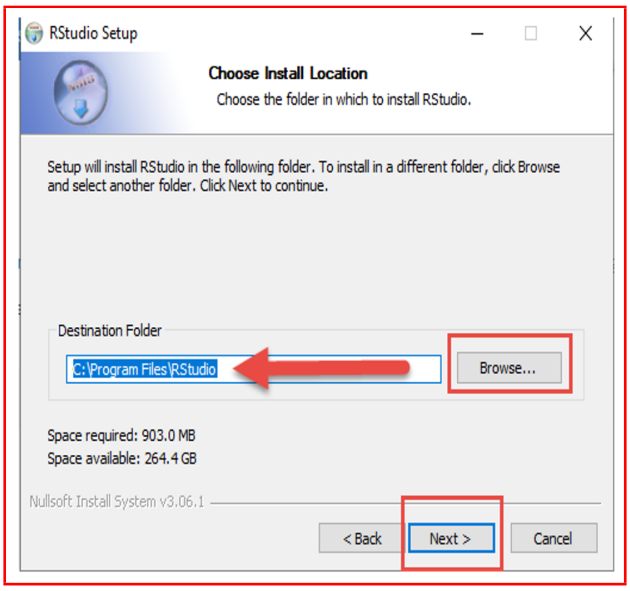

Figure 15

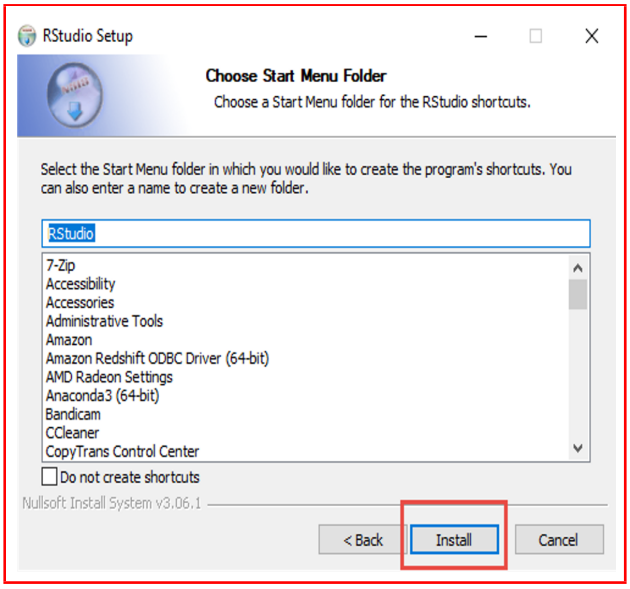

Figure 16

The following file can now be run from within the R Studio application.

- Phase1-DataCleaning

4 PostgreSQL Database

PostgreSQL is a reliable, efficient and robust relational database model which is also open source.

4.1 Downloading PostgreSQL Database from www

Open any web browser and navigate to www.postgresql.org/download. Once there, select the correct operating system, in this case, Windows, as seen in Figure 17. Next, the browser should redirect to www.postgresql.org/download/windows where the "download the installer" link must be selected as seen in Figure 18. Next, select the most recent version to download from the server, as shown in Figure 19. Finally, the file should be saved in the standard folder used for downloads on the local device, as shown in Figure 20.

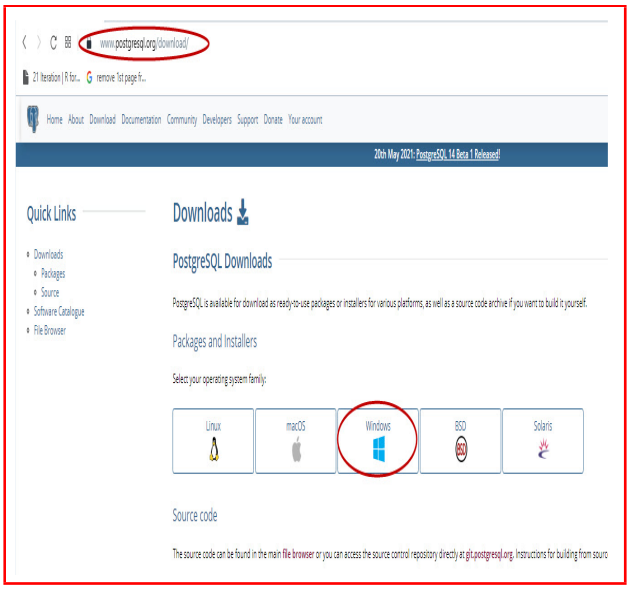

Figure 17

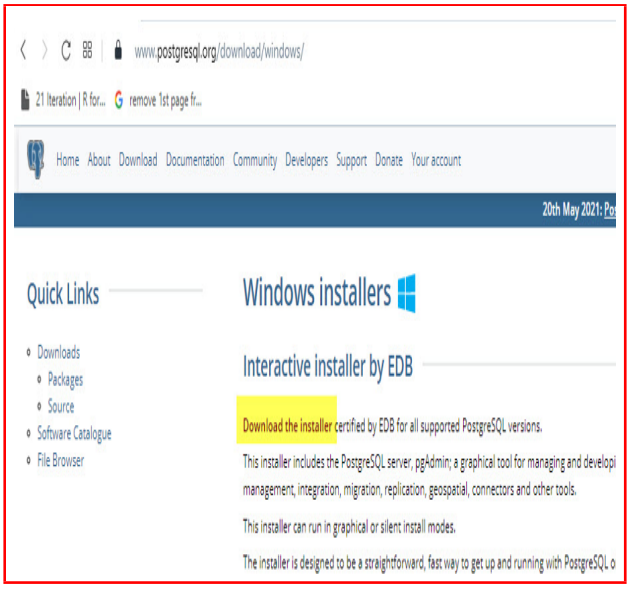

Figure 18

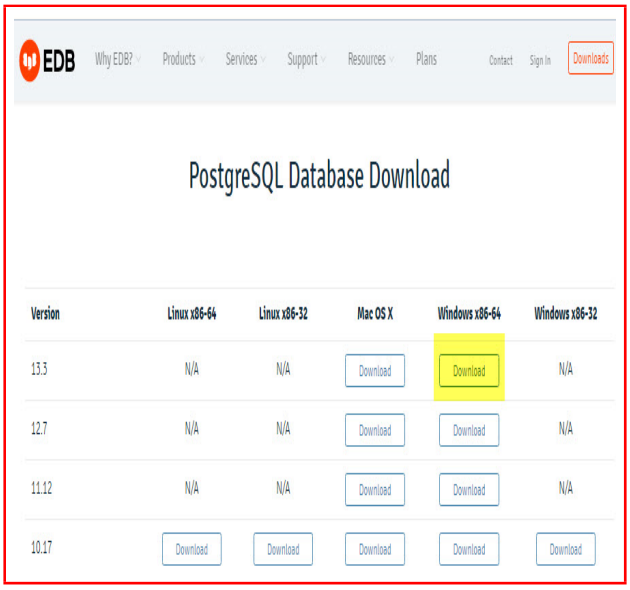

Figure 19

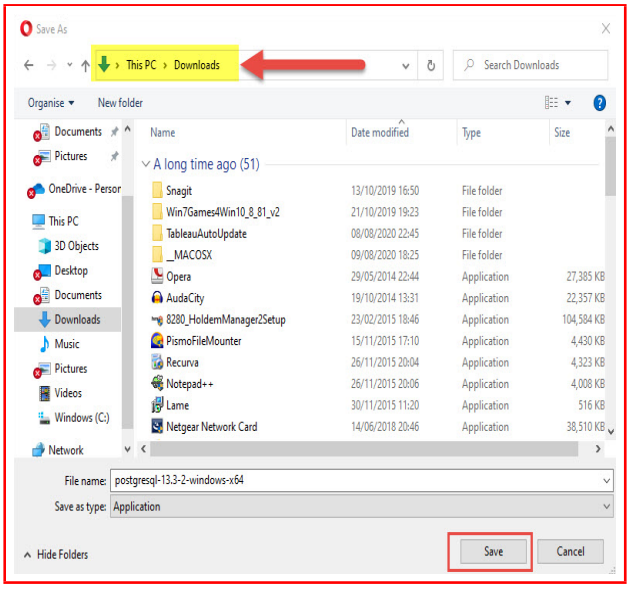

Figure 20

4.2 Installing PostgreSQL Database locally

Once downloaded entirely, double click on the application to begin the installation process, as shown in Figure 21. Once again, the default path is used for this installation, as demonstrated in Figure 22. All four are selected when selecting components to install as part of the installation, as shown in Figure 23.

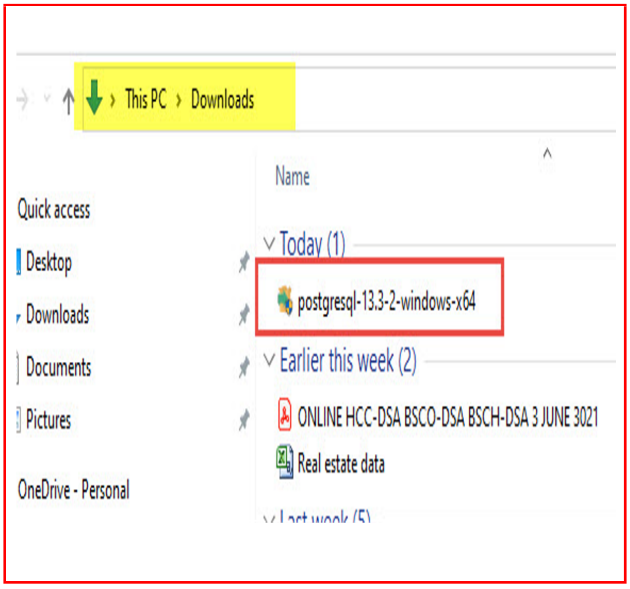

Figure 21

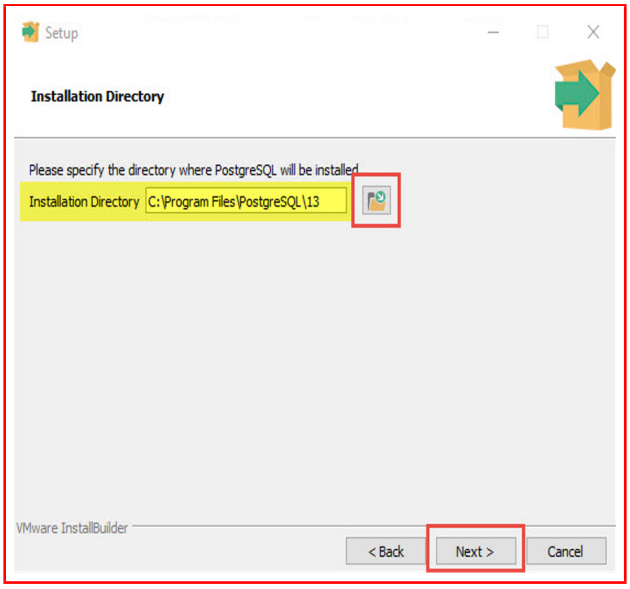

Figure 22

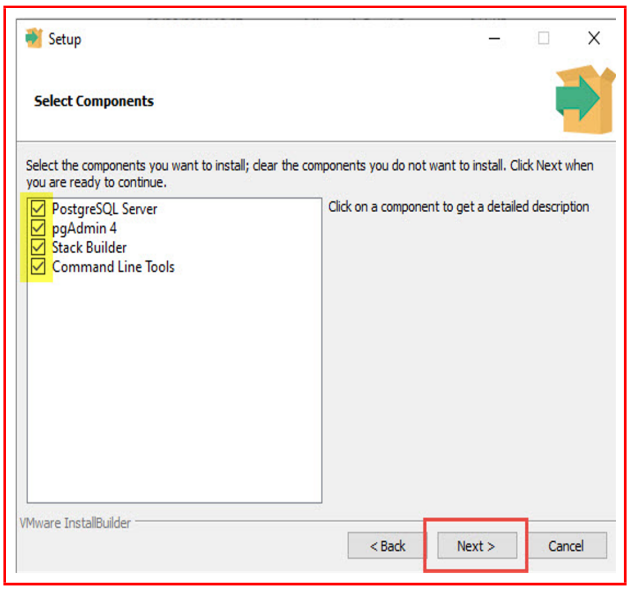

Figure 23

4.3 Configuring PostgreSQL Database locally

Unlike other software installations within the project, the configuration is needed once PostgreSQL is downloaded and installed locally. First, the password needs to be set, as shown in Figure 24. Please note the sensitivity of the case. There is no password reset option; it is advisable to make a separate password record once it is created or keep it to DataAnalytics as demonstrated. Next, the server will "listen" for communications on a dedicated port number. Port 5432 is advised as the port to declare, as is shown in Figure 25. Finally, the relevant locale details are set, as shown in Figure 26. For this project, it is was decided not to launch stack builder upon completion of the configuration.

Therefore, please ensure this box is not ticked, as shown in Figure 27, before clicking the "finish" button.

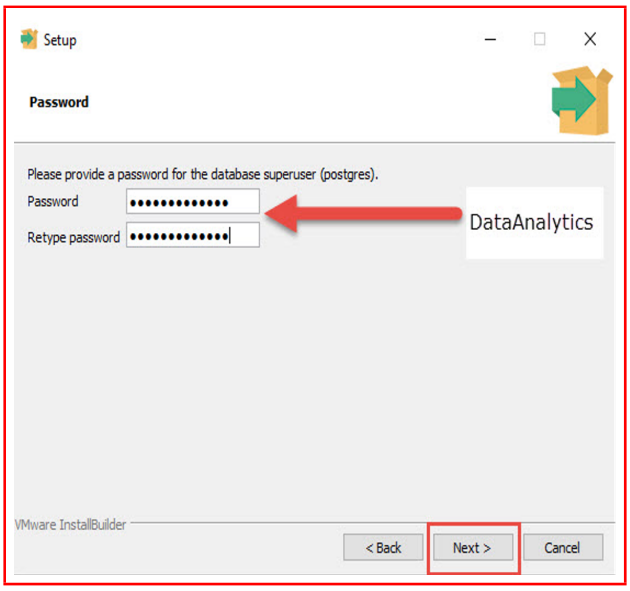

Figure 24

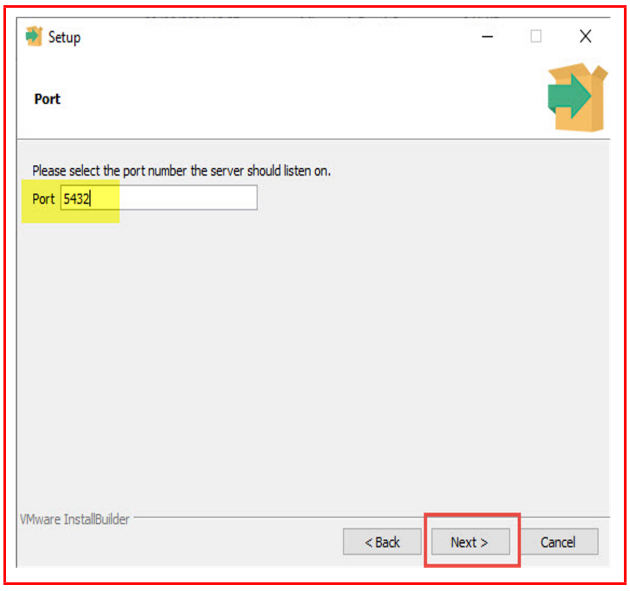

Figure 25

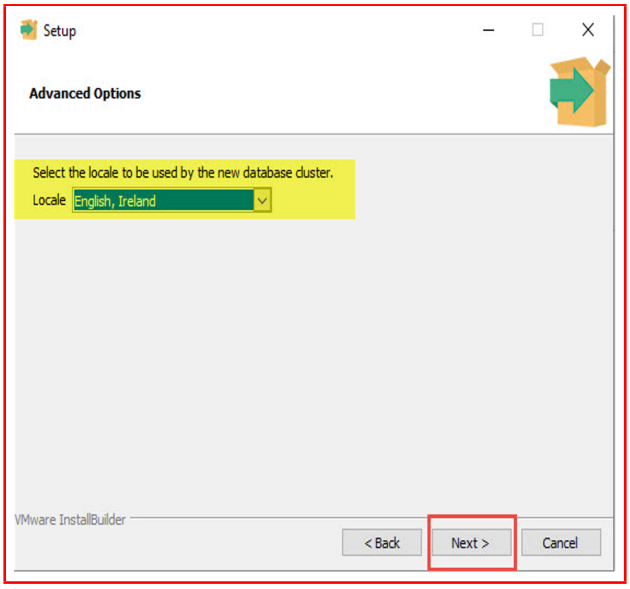

Figure 26

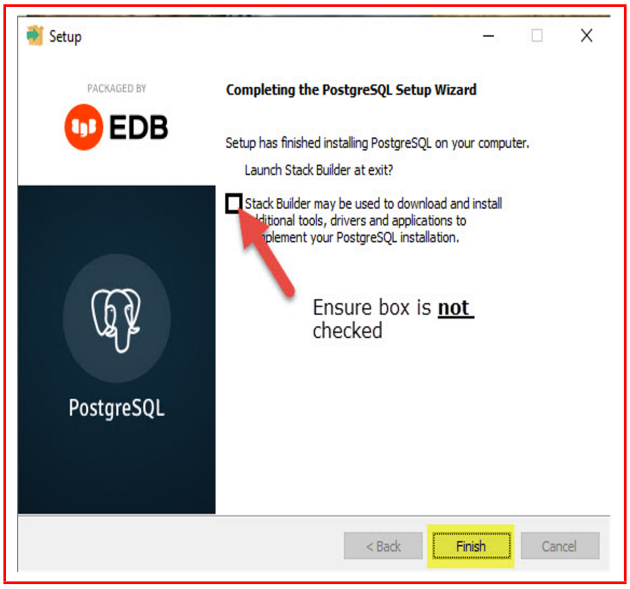

Figure 27

4.4 Recap of the customised configuration

Please ensure the following configurations have been implemented precisely as outlined in the previous subsection 4.3 to avoid connectivity issues later within the project.

- **Password:** DataAnalytics (Case Sensitive)
- **Port:** 5432
- **Locale:** English, Ireland
- **Launch Stack Builder At Exit?:** Box Unchecked

5 Anaconda and Jupyter Notebook

The distribution and development of Python applications occur within Anaconda, which like R Studio, is an open-source environment that can be downloaded online.

Anaconda is downloaded from www.anaconda.com as shown in Figure 28. Also, similar to R Studio, there are many potential downloads (both free and subscription-based) for the system in question. Please download the most applicable to you. However, the individual open source edition highlighted is sufficient for the needs of this project. Once selected, the browser should lead you to www.anaconda.com/products/individual where the download can begin as shown in Figure 29 and Figure 30. The file should be saved in the standard folder used for downloads on the local device, as shown in Figure 31.

5.1 Downloading Anaconda from www

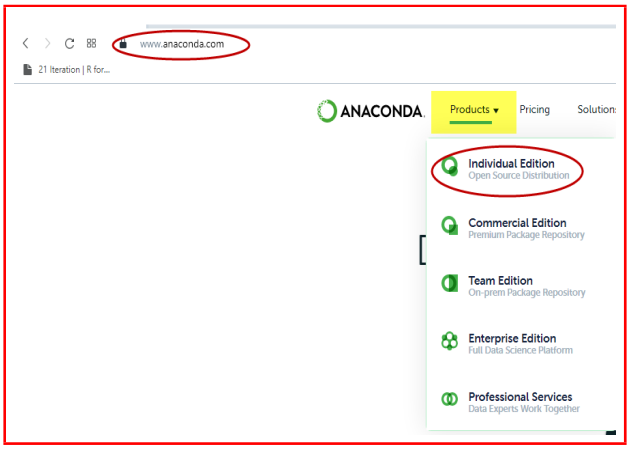

Figure 28

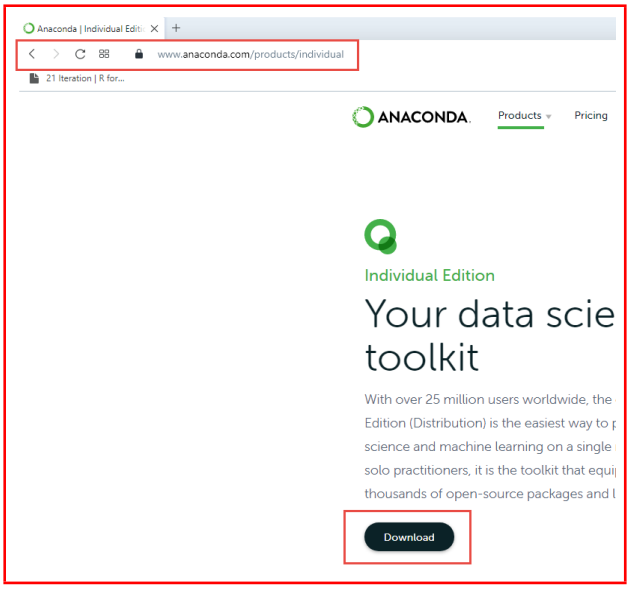

Figure 29

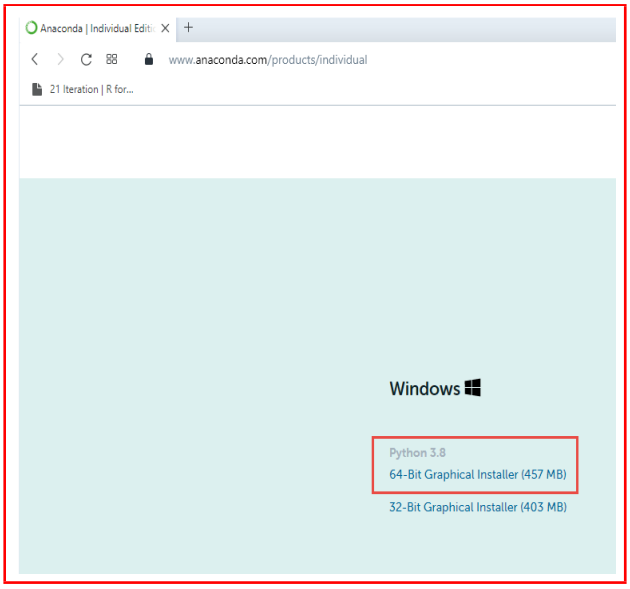

Figure 30

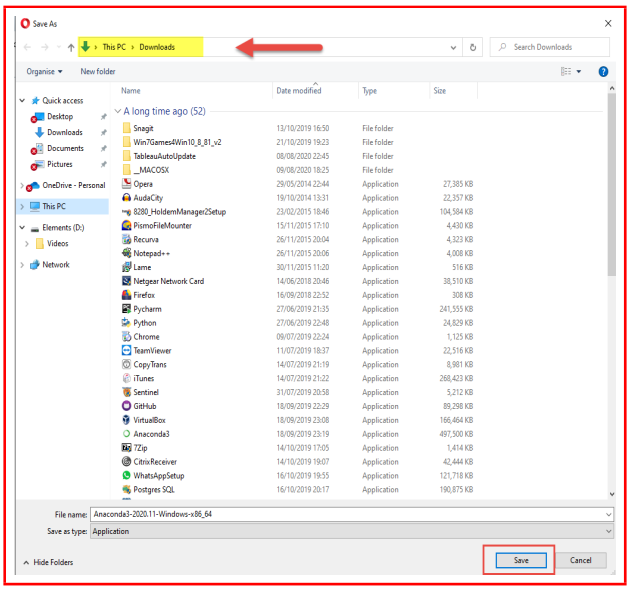

Figure 31

5.2 Installing Anaconda locally

Once downloaded entirely, double click on the application to begin the installation process, as shown in Figure 32. The system on which you are installing Anaconda may have multiple users. It is the individual's choice to download for all users on the local device or just themselves, as shown in Figure 33. The primary installation of Anaconda for the project was just installed for the current user as recommended within the application. Figure 34 shows how to select an alternative file path to install the source files if that is required. Once again, it is noted that the default path is used for this installation.

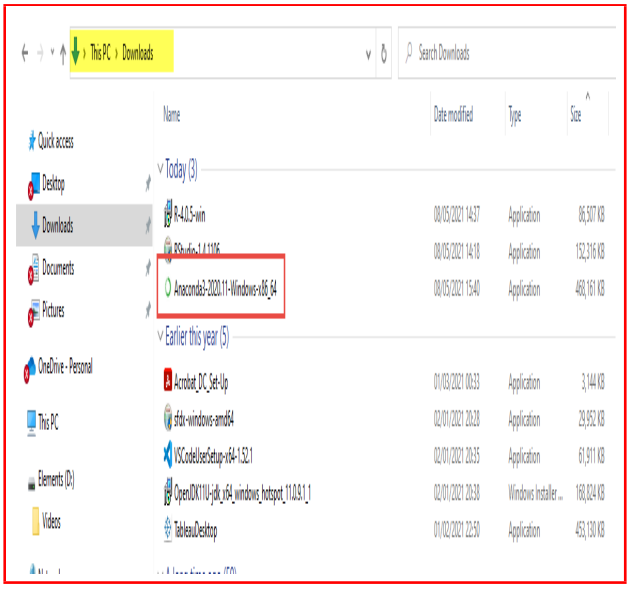

Figure 32

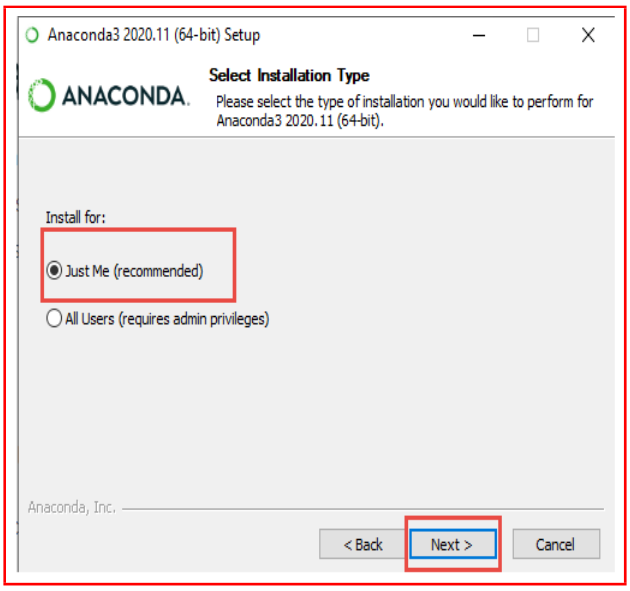

Figure 33

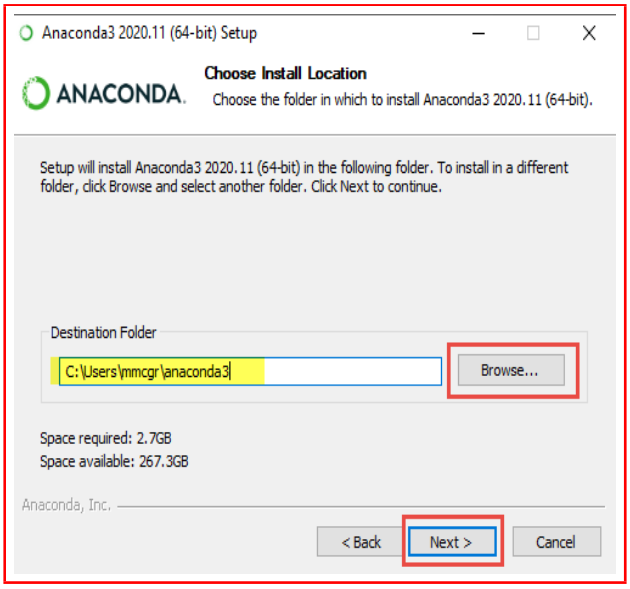

Figure 34

Not seen with R or R Studio, an advanced installation section is presented in Figure 35 with the facility to "Add Anaconda3 to my PATH environment variable" and "Register Anaconda3 as my default Python 3.8". It is anticipated that **diverging from the selection shown within the figure of option one unticked and option two ticked could impact the set-up within Anaconda.** The remainder of the wizard completes the installation. Nothing was changed from the default from this juncture onwards.

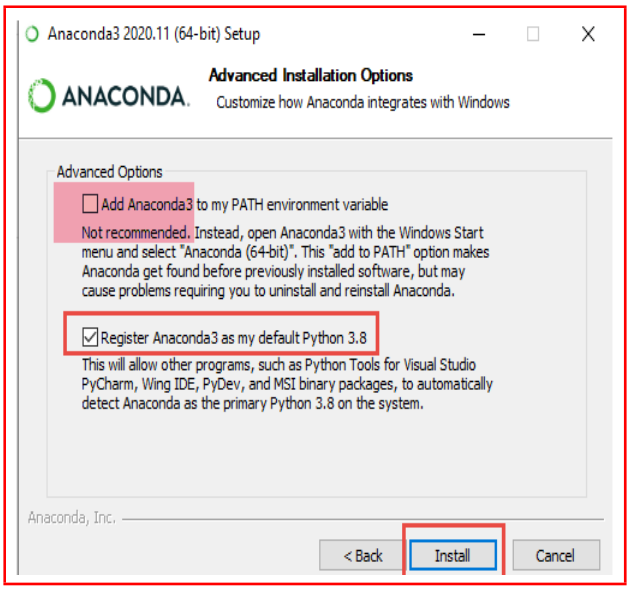

Figure 35

5.3 Installing Jupyter Notebook

Once installed, Anaconda can be opened through the start menu or possibly a desktop shortcut if one was created as part of the set-up. Figure 36 outlines the final step necessary

to install Jupyter Notebook within the Anaconda environment. The wizard can be run with automatic default values and completes in a few short minutes.

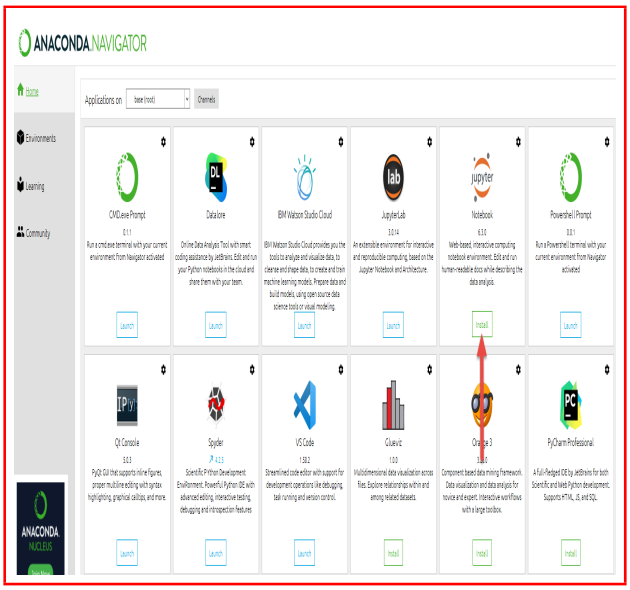

Figure 36

The following files can now be run from within the Jupyter Notebook application.

- Phase2A-PipInstalls
- Phase2B-DataVisualisation
- Phase2C-DataSculpting
- Phase2D-TransferToPostgreSQL

6 Tensorflow and Keras

6.1 Creation of new Anaconda Environment

The installation of Tensorflow and Keras for the Deep Learning model development will require additional configuration within Anaconda. As shown in Figure 37, this process starts with creating a new environment upon which these additional packages can be installed. As shown in Figure 38, it may be necessary to update your Anaconda Navigator if a more up to date version is available. This is always the preferred option to ensure you are running the most recent packages available within the application.

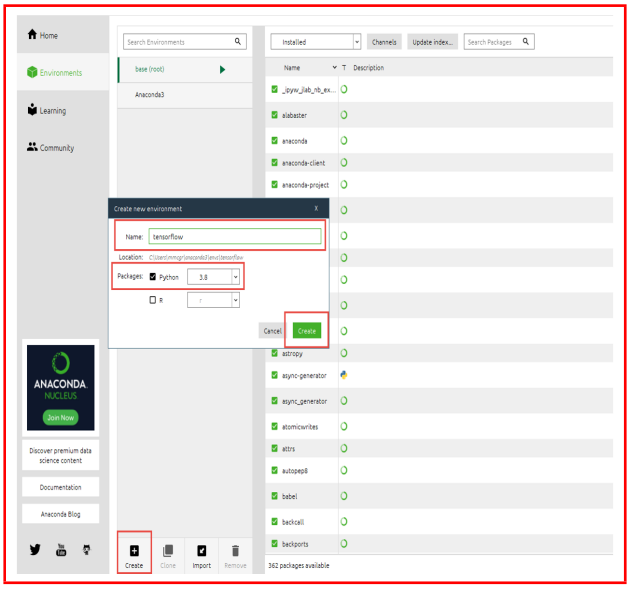

Figure 37

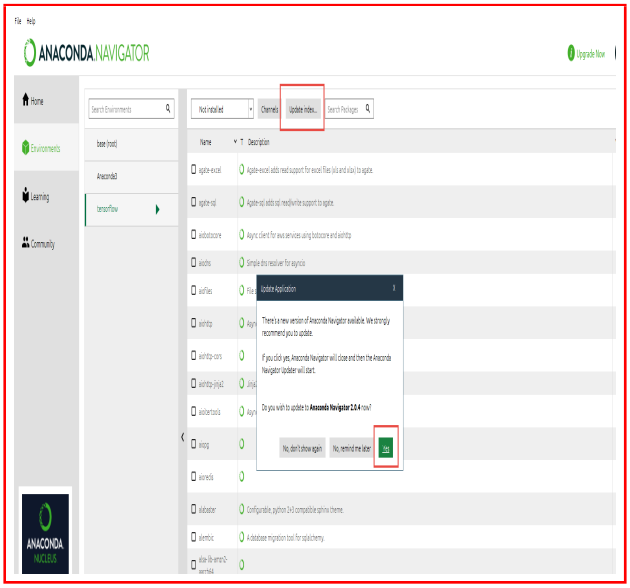

Figure 38

6.2 Installation of Tensorflow and Additional Packages

Once the new environment is created, it will be necessary to specifically install the packages used for the development of the Deep Learning Model. Figure 39 demonstrates how the Tensorflow package is installed within the new environment. For the purposes of the code run in the last file of the project, it will be necessary to repeat this final step to install the following packages.

- Matplotlib
- Pandas
- Seaborn
- Keras

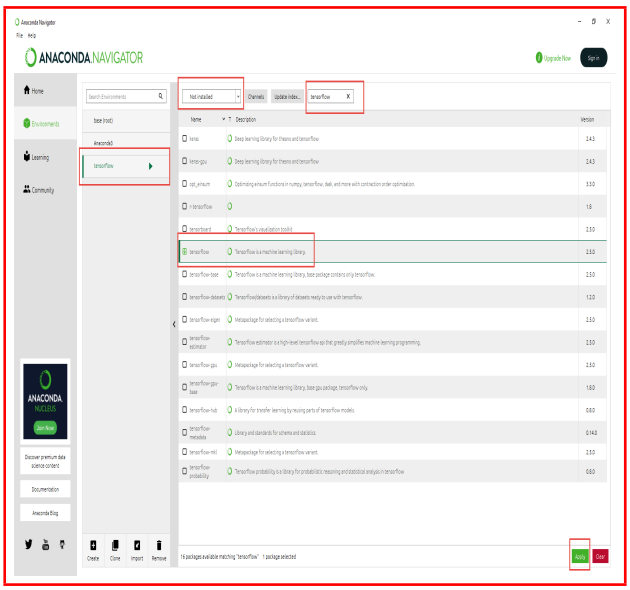

Figure 39

It is possible to verify these packages have been installed by checking the "Installed" section of the packages within the new environment as shown in Figure 40

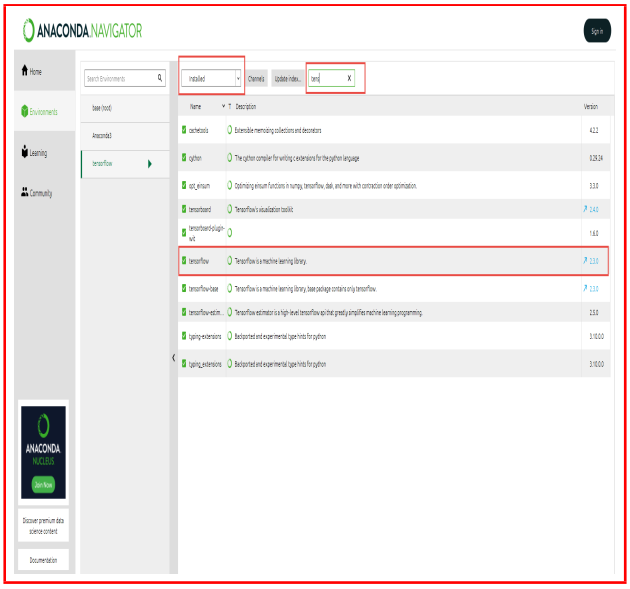

Figure 40

The final steps are shown in Figure 41 and Figure 42. This involves launching and installing Jupyter Notebook as done previously in section 5.3. N.B. Jupyter must be installed and launched within the new environment which was just created.

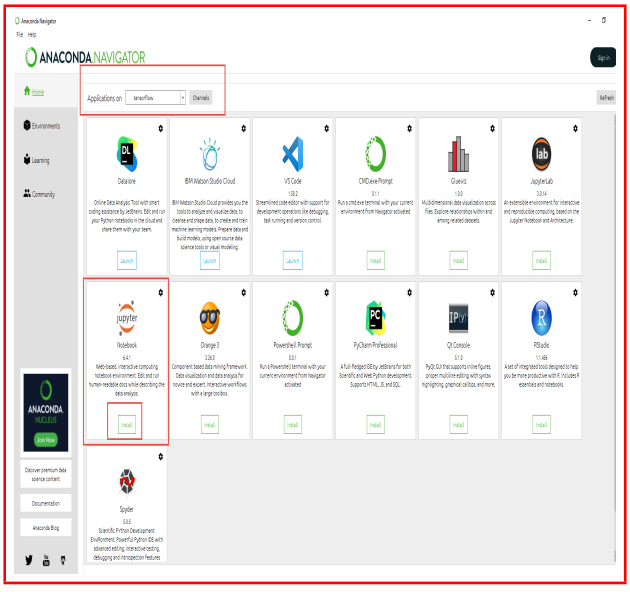

Figure 41

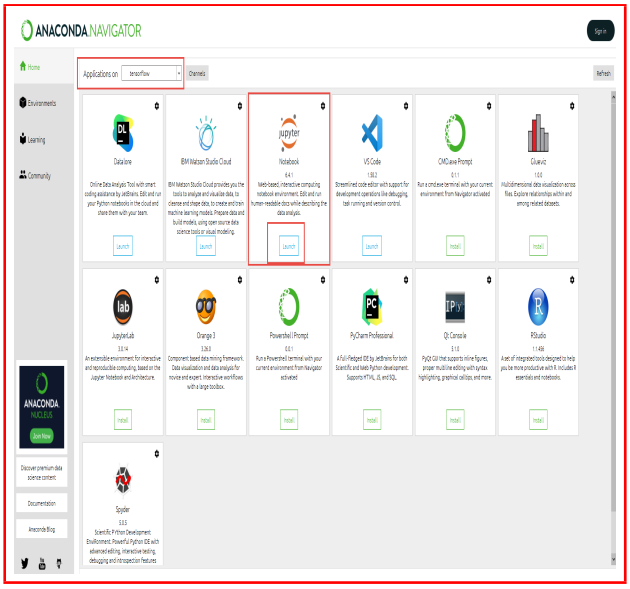
ANACONDA NAVIGATOR
Sign in
Home
Environments
Learning
Community
Applications on
tensorflow
Channels
Refresh
Datalore
IBM Watson Studio Cloud
Notebook
VS Code
CMD.exe Prompt
Glueviz
JupyterLab
Orange 3
Powershell Prompt
PyCharm Professional
Qt Console
RStudio
Spyder
ANACONDA
NUCLEUS
Join Now
Discover premium data science content
Documentation
Anaconda Blog

Figure 42

7 Final Configuration Details

7.1 Order Of Configuration

It is of vital importance that the installation of software is done in accordance to sections 2, 3, 4 , 5 , 6 and in the order specified. This will enable the R and Python scripts to behave with the consistency expected and experienced during the implementation and testing phase. Table 2 gives a breakdown of the milestones needed for each of the files to be executed.

Table 2: When Scripts Can Be Run

Application	**What Can Be Run Once Installed**
R	
R Studio	Phase1-DataCleaning
PostgreSQL	
Anaconda and Jupyter Notebook	Phase2A-PipInstalls Phase2B-DataVisualisation Phase2C-PreProcessing Phase2D-TransferToPostgreSQL Phase3A-DailyEnergyRegressonPredictionModel-RandomForrest Phase3B-DailyEnergyRegressonPredictionModel-DecisionTrees Phase3C-DailyEnergyRegressonPredictionModel-KNN Phase4A-RecommendationSystemDevelopment Phase4B-RecommendationSystemAnalysis
Tensorflow and Keras	Phase5-DeepNeuralNetworkRecommender

7.2 File Directories

2 local file paths will need to be set up on the local system to handle

- The input files are provided within the HUE Dataset. Once set up, please transfer the Residential 1-28.csv, Holidays.csv, WeatherYVR.csv and All-Residential.txt files there.
- The output files generated by the R and Python scripts as part of the project execution.

When implementing the project the directory `C:\Users\mmcgr\OneDrive\Documents\College2019_2020\Semester4\ResearchProject\Development\InputFiles\OriginalHUE_Data` was used for storing the input files (Residential 1-28.csv, Holidays.csv and WeatherYVR.csv)

The directory `C:\Users\mmcgr\OneDrive\Documents\College2019\textunderscore2020\Semester4\ResearchProject\Development\FilesForAnalysis` was used for all files that are created, exported to and subsequently read back in from the project by the R and Python scripts.

7.3 Replacing Text Within Files

It will be necessary to make the following replacements in the files listed below to ensure automation.

- **Phase1-DataCleaning**
  Replace `"C:/Users/mmcgr/OneDrive/Documents/College2019_2020/Semester4/ResearchProject/Development/InputFiles/OriginalHUE_Data"` with the local directory where the input files are stored

  Replace `"C:/Users/mmcgr/OneDrive/Documents/College2019_2020/Semester4/ResearchProject/Development/FilesForAnalysis"` with the local directory where the output files will be created, exported to and subsequently read back in from.

- **Phase2A-PipInstalls**
  No changes needed

- **Phase2B-DataVisualisation**
  Replace `"C:/Users/mmcgr/OneDrive/Documents/College2019_2020/Semester4/ResearchProject/Development/FilesForAnalysis"` with the local directory where the output files will be created, exported to and subsequently read back in from.

  N.B. This will need to be done in 2 places within the file.

- **Phase2C-PreProcessing**
  Replace `"C:/Users/mmcgr/OneDrive/Documents/College2019_2020/Semester4/ResearchProject/Development/FilesForAnalysis"` with the local directory where the output files will be created, exported to and subsequently read back in from.

  N.B. This will need to be done in 2 places within the file.

- **Phase2D-TransferToPostgreSQL**
  Replace `"C:/Users/mmcgr/OneDrive/Documents/College2019_2020/Semester4/ResearchProject/Development/FilesForAnalysis"` with the local directory where the output files will be created, exported to and subsequently read back in from.

- **Phase3A-DailyEnergyRegressonPredictionModel-RandomForrest**
  Replace `"C:/Users/mmcgr/OneDrive/Documents/College2019_2020/Semester4/ResearchProject/Development/FilesForAnalysis"` with the local directory where the output files will be created, exported to and subsequently read back in from.

- **Phase3B-DailyEnergyRegressonPredictionModel-DecisionTrees**
  Replace `"C:/Users/mmcgr/OneDrive/Documents/College2019_2020/Semester4/ResearchProject/Development/FilesForAnalysis"` with the local directory where the output files will be created, exported to and subsequently read back in from.

- **Phase3C-DailyEnergyRegressonPredictionModel-KNN**
  Replace `"C:/Users/mmcgr/OneDrive/Documents/College2019_2020/Semester4/ResearchProject/Development/FilesForAnalysis"` with the local directory where the output files will be created, exported to and subsequently read back in from.

- **Phase4A-RecommendationSystemDevelopment**
  Replace `"C:/Users/mmcgr/OneDrive/Documents/College2019_2020/Semester4/ResearchProject/Development/FilesForAnalysis"` with the local directory where the output files will be created, exported to and subsequently read back in from.

  N.B. This will need to be done in 3 places within the file.

- **Phase4B-RecommendationSystemAnalysis**
  Replace `"C:/Users/mmcgr/OneDrive/Documents/College2019_2020/Semester4/ResearchProject/Development/FilesForAnalysis"` with the local directory where the output files will be created, exported to and subsequently read back in from.

  N.B. This will need to be done in 2 places within the file.

- **Phase5-DeepNeuralNetworkRecommender**
  Replace `"C:/Users/mmcgr/OneDrive/Documents/College2019_2020/Semester4/ResearchProject/Development/FilesForAnalysis"` with the local directory where the output files will be created, exported to and subsequently read back in from.

  N.B. This will need to be done in 2 places within the file.

7.4 File Use

Table 3 gives a summary as to what files are used as inputs and are outputted as a result of each script. All input and output files are provided as part of the ICT Artefact Upload. Still, it is essential to note that the output files may be overridden as the scripts are executed.

Table 3: How CSV Files Are Used Within Scripts

Script	**Input CSV Files**	**Output CSV Files**
Phase1-DataCleaning	Residential-1.csv — Residential-28.csv Weather-YVR.csv	energyUsageHistory WithWeather-V1.csv
Phase2A-PipInstalls	—	—
Phase2B-DataVisualisation	energyUsageHistory WithWeather-V1.csv	energyUsageHistory WithWeather-V2.csv
Phase2C-PreProcessing	energyUsageHistory WithWeather-V2.csv	energyUsageHistory WithWeather-V3.csv
Phase2D-TransferToPostgreSQL	energyUsageHistory WithWeather-V3.csv	
Phase3A-DailyEnergyRegressonPredictionModel-RandomForrest	energyUsageHistory WithWeather-V3.csv	—
Phase3B-DailyEnergyRegressonPredictionModel-DecisionTrees	energyUsageHistory WithWeather-V3.csv	—
Phase3C-DailyEnergyRegressonPredictionModel-KNN	energyUsageHistory WithWeather-V3.csv	—
Phase4A-RecommendationSystemDevelopment	energyUsageHistory WithWeather-V3.csv	recommender LogHistory.csv recommenderSystem TestingResults.csv
Phase4B-RecommendationSystemAnalysis	recommender LogHistory.csv recommenderSystem TestingResults.csv	—
Phase5-DeepNeuralNetworkRecommender	recommenderSystem TestingResults.csv energyUsageHistory WithWeather-V3.csv	—

References

Anaconda,Data science technology for groundbreaking research (n.d.). https://www.anaconda.com/. Accessed: 2021-06-08.

Keras,Simple. Flexible. Powerful. (n.d.). https://keras.io/. Accessed: 2021-07-01.

Matplotlib: Visualization with Python (n.d.). https://matplotlib.org. Accessed: 2021-05-21.

Pandas: fast, powerful, flexible and easy to use (n.d.). https://pandas.pydata.org. Accessed: 2021-05-23.

PostgreSQL:,The World's Most Advanced Open Source Relational Database (n.d.). https://www.postgresql.org/. Accessed: 2021-06-06.

RStudio Connect 1.9.0,R Studio Connect (n.d.). https://www.rstudio.com/. Accessed: 2021-06-06.

Seaborn: Visualization with Python (n.d.). https://seaborn.pydata.org. Accessed: 2021-05-21.

The R Project for Statistical Computing,R (n.d.). https://www.R-project.org/. Accessed: 2021-05-30.