**Where to put this document when it’s done**

When your team has finished compiling this Handover Document and all of the associated resources, please follow these instructions:

1. In the root directory\* of your squad’s Microsoft Teams ‘Files’, create a folder called ‘Handover and Showcase’ (if an equivalent folder already exists, just rename the folder).
2. In the ‘Handover and Showcase’ folder, create a new folder for the current trimester (e.g., ‘T3 2020’).
3. Put this document and all of the associated files and resources in this new folder.
4. If your team has old handover files from previous trimesters, please reorganise them to be within this same filing system under the same naming conventions. We’re trying to create uniformity among handover resources!

\*If your squad shares your Files with another squad, navigate to the location that your team would consider to be the root directory for your squad.

A screenshot of a computer

Description automatically generated with medium confidence

Please delete this page before submission.

Project Handover Document

**Deakin 2 Intelligence Consulting**

Melbourne City

*Trimester 1, 2021*

Table of Contents

[1. Project Information 5](#_Toc73607091)

[1.1 Client/Product Owner 5](#_Toc73607092)

[1.2 Academic Mentor/Supervisor 5](#_Toc73607093)

[1.3 Project Team 5](#_Toc73607094)

[2. Project Overview 5](#_Toc73607095)

[3. User Manual 7](#_Toc73607096)

[3.1. Pedestrian count in Melbourne city 7](#_Toc73607097)

[3.2. Energy consumption 9](#_Toc73607098)

[3.3. Pedestrian count for specific day 11](#_Toc73607099)

[3.4. Forecast for Independent variables 12](#_Toc73607100)

[3.5. Prediction of Pedestrian Count at granular level 14](#_Toc73607101)

[3.6. Features of Highcharts 15](#_Toc73607102)

[3.7. Video for user navigation 15](#_Toc73607103)

[4. Completed Deliverables 16](#_Toc73607104)

[4.1 Overall Deliverables 16](#_Toc73607105)

[4.2 Sub-Team A Deliverables 16](#_Toc73607106)

[4.3 Sub-Team B Deliverables 17](#_Toc73607107)

[4.4 Sub-Team C Deliverables 19](#_Toc73607108)

[5. Roadmap 21](#_Toc73607109)

[6. Open Issues 22](#_Toc73607110)

[7. Lessons Learned 22](#_Toc73607111)

[8. Product Development Life Cycle 23](#_Toc73607112)

[8.1 New Tasks 23](#_Toc73607113)

[8.2 Definition of Done 23](#_Toc73607114)

[8.3 Task Review 23](#_Toc73607115)

[8.4 Testing 24](#_Toc73607116)

[8.5 Branching Strategy 24](#_Toc73607117)

[9. Product Architecture 25](#_Toc73607118)

[9.1 UML Diagram 25](#_Toc73607119)

[9.2 Tech Stack 26](#_Toc73607120)

[10. Source Code 26](#_Toc73607121)

[11. Login Credentials 27](#_Toc73607122)

[12. Appendices 27](#_Toc73607123)

# Project Information

## 1.1 Client/Product Owner

**Mr. Will Mcintosh**

Architecture and Data capability Manager

[will.mcintosh@melbourne.vic.gov.au](mailto:will.mcintosh@melbourne.vic.gov.au)

## 1.2 Academic Mentor/Supervisor

**Mr. Thanh Thi Nguygen**

Senior Lecturer, Information Technology (AI)

[thanhthi@d2l.mail.deakin.edu.au](mailto:thanhthi@d2l.mail.deakin.edu.au)

## 1.3 Project Team

**Deakin 2 Intelligence Consulting**

Melbourne City

|  |  |  |
| --- | --- | --- |
| Student ID | Student Name | Role |
| 218590861 | Ayodeji Ladeinde | Senior student |
| 219506525 | Aparna Chintala | Senior student |
| 219398648 | Akhila Manchi | Senior student |
| 220020117 | Nikita Wadekar | Senior student |
| 220186117 | Rohan Man Amatya | Senior student |
| 220365232 | Sivaram Krishnan | Senior student |
| 220608536 | Jason Tsitsopoulos | Junior student |
| 219096246 | Miriam Zhu | Junior student |
| 215101976 | Neet Patel | Junior student |
| 219384532 | Oscar Wu | Junior student |

# Project Overview

According to a report released by Climate Council, clean and healthy environment has a significant impact on the community's habitability. Due to the bushfires and smoke haze in 2020 summer, 80 percent of Australians were affected directly or indirectly. During the catastrophic events, the City of Melbourne council had a comprehensive Climate Change Adaptation Strategy that was issued in 2017, which guided the council to make appropriate decisions. The City of Melbourne council has planted 3000 trees in 2021 to reduce the carbon content in the atmosphere, invested $40 million in stormwater harvesting, working on projects for park expansion, accelerated the recycling hubs and invested $17.1 million to install 2244 solar panels in Melbourne.

The COVID-19 pandemic in 2020-2021 has severely impacted business activities and the pedestrian traffic in the CBD. As per the Australian Bureau of Statistics, 78% of businesses reported no change or increase in revenues in May 2021. 9% of businesses have hired employees in May 2021 and the employment rate is severely affected due to work from home policy of the companies in CBD. Due to the pandemic, 43% of businesses reported that their cash on hand could cover certain business operations for three months. Only 20% of businesses have stopped accessing the support measures provides by the Australian government. The council needs to develop strategies to increase revenue, attract a greater number of people to the city and execute the recovery plan with consistency.

This project aims to assist the City of Melbourne council's decisions to encourage Melbournians to adopt measures that would facilitate reduction of carbon content within the city, organize events within the city to improve the financial status of the businesses affected due to the COVID-19 pandemic, manage the energy consumption levels and invest in renewable energy. The council is committed to investing in infrastructure and contribute towards the next zero emissions. In 2021, Australia’s Technology Investment Roadmap is expected to invest $18 billions of Government investment over the next 10 years and for low emissions technologies drive at least $70 billions of total new investment in Australia by 2030.

The project showcases visualisation of prediction of total pedestrians' count for the next three weeks in the CBD using automated ETL forecasting pipeline and granular level data for 5 sensor locations. Using the independent variables, user can predict the real-time total pedestrians' count in CBD. Independent variables include the specific day using weather forecast features (obtained from reliable sources), whether government has imposed restrictions and status of the public holiday. The project showcases visualization of energy consumption till August 2021 in the Melbourne city. Different time series models were built and the algorithm with the least error was adopted for the predictions. Different regression models were built to predict specific day pedestrian count and the algorithm with the least error was adopted.

Following are the deliverables of the project:

* **Automated ETL forecasting pipeline:** The process of extracting the data, cleaning data, transforming data, executing the time-series model and updating the dashboard on the website is automated using cronjobs. The website is hosted on the EC2 instance and on execution of the job on the server the ETL process is initiated.
* **Forecast pedestrian count:** Time-series model to predict the pedestrian count of the next three weeks using automated ETL forecasting pipeline.
* **Forecast pedestrian count using specific input:** Improved the regression model implemented in Trimester 3, 2020 to predict future pedestrian count depending on some input parameters. This will help the council to make appropriate decisions and achieve net-zero emission by 2040.
* **Forecast granular level pedestrian count:** Forecasted the pedestrian count for top 5 sensor locations using time-series model to obtain the granular level data. The 5 sensor locations are Bourke Street mall south, Victoria point, Collins Place North, Flinders Street and Southern cross station.
* **Forecast energy consumption:** Time series model to predict the energy consumption for March to August 2021 and showcased the forecasted values using a dashboard.
* **Exploratory Data Analysis:** Time series model to predict the independent variables such as weather parameters and recommendation retail price. Performed Exploratory Data Analysis on the independent variables. Represented the forecasted values using dashboard and various graphs.

# User Manual

## Pedestrian count in Melbourne city

Homepage:

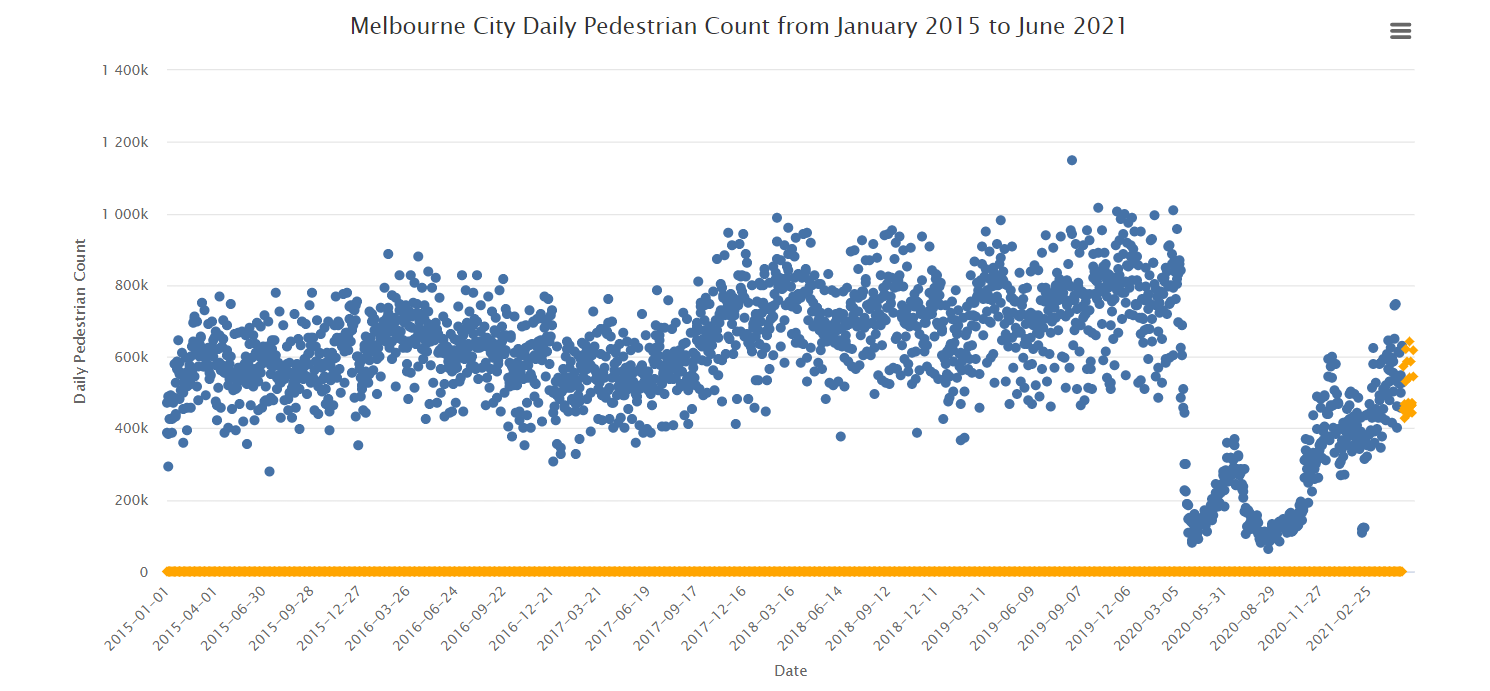


To access the pedestrian forecast for the Melbourne city, hover over ‘Prediction’ and click on ‘Pedestrian Forecast’.

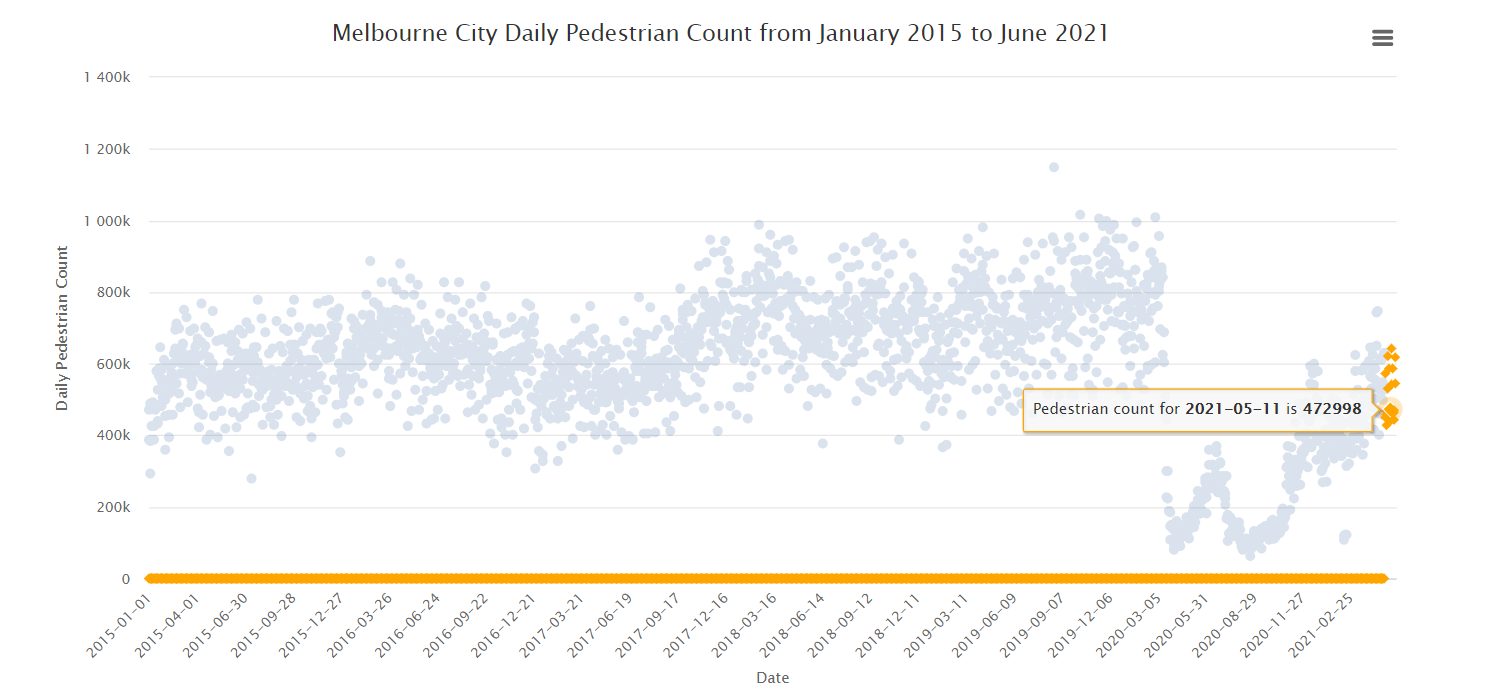


Dashboard for the pedestrian forecast:

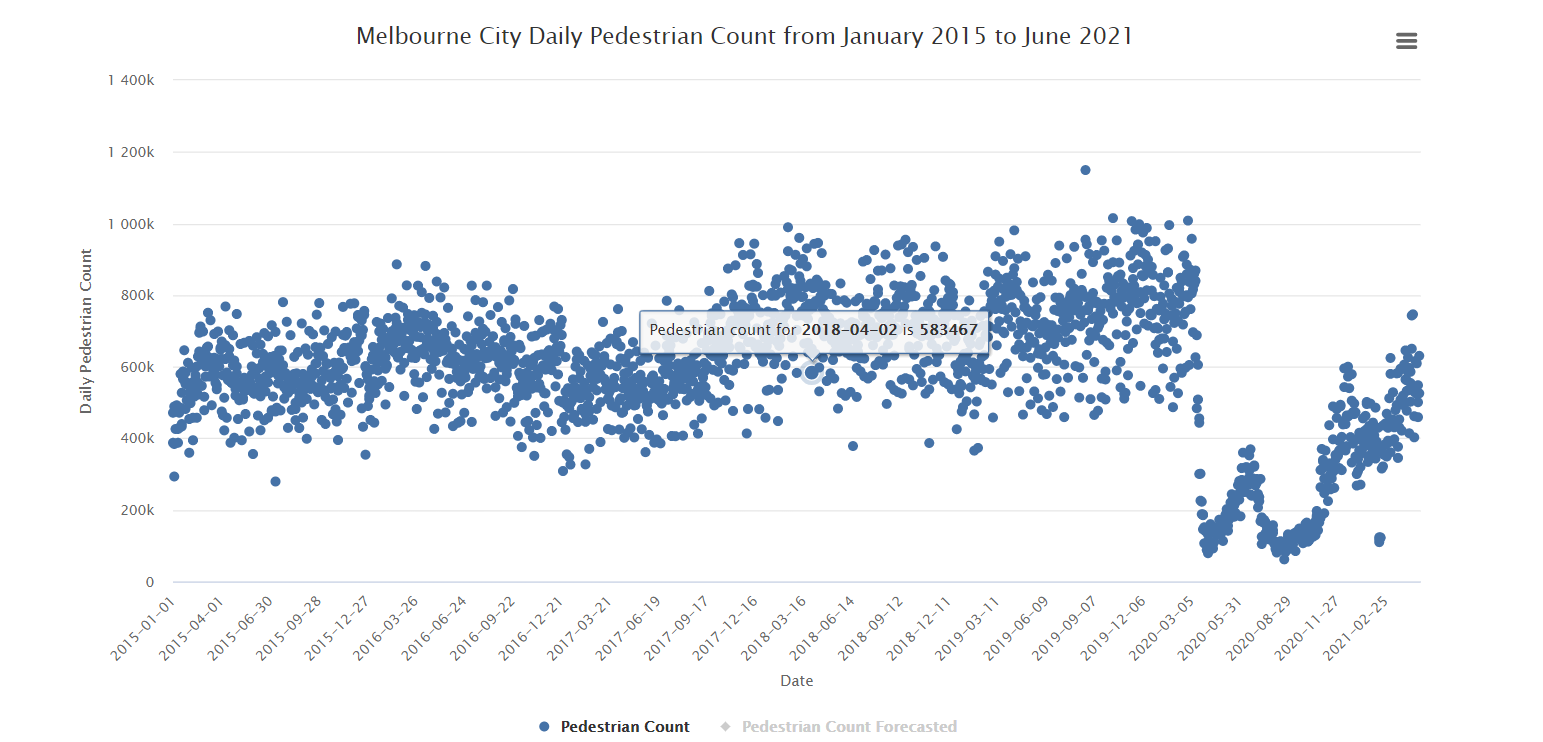
The pedestrian forecast for next three weeks is visible in orange while the actual data from 72 pedestrian sensor location between January 2015 till June 2021 is depicted using blue. The visualization is updated in real time through an automated ETL forecasting pipeline.



To view the pedestrian count, hover over the particular data point.



Using the Highcharts inbuilt features, user can display the visualization for actual data or forecasted values by disabling the other graph.



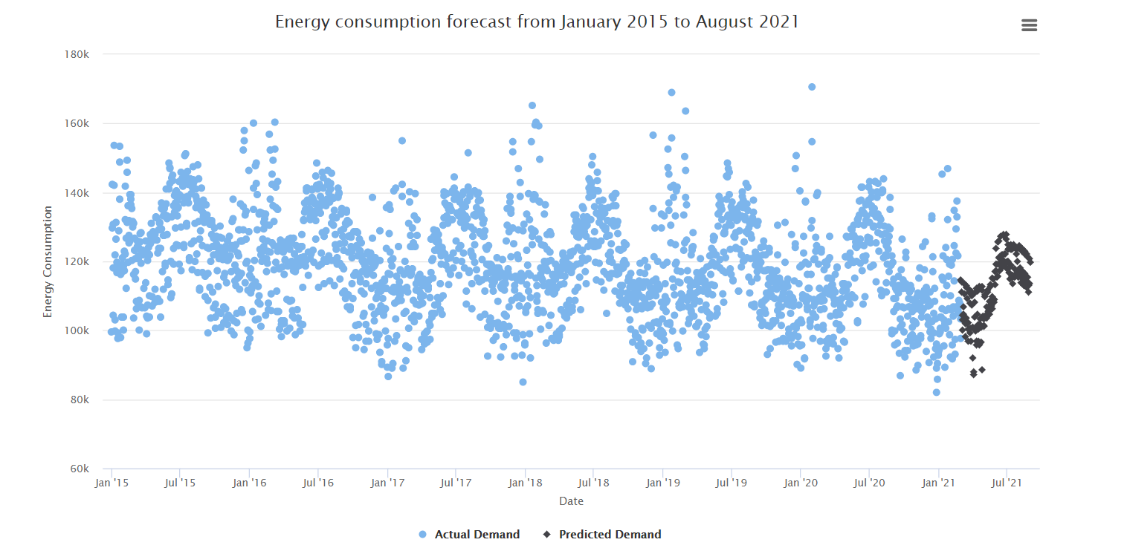
## Energy consumption

To access the energy consumption prediction for the Melbourne city, hover over ‘Prediction’ and click on ‘Energy Forecast’.

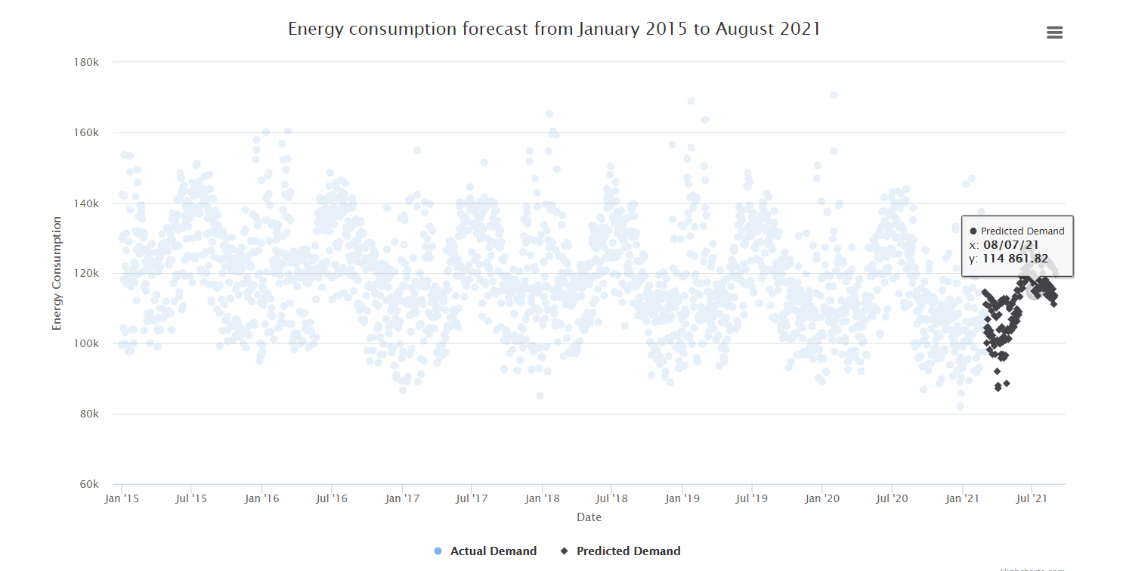


Dashboard for energy consumption prediction:

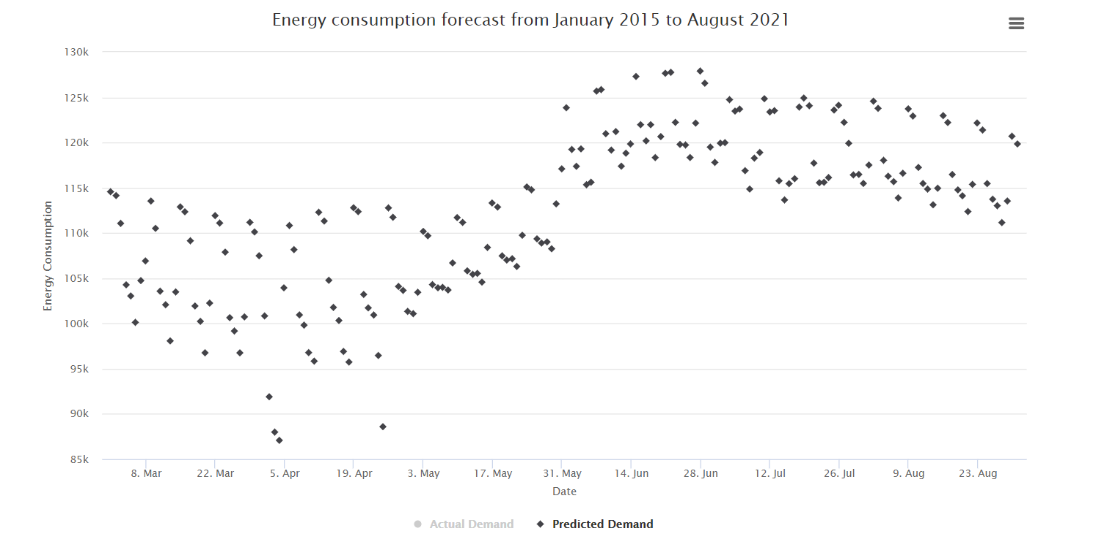
The actual data from AEMO between January 2015 – February 2021 is depicted in blue while the energy consumption forecast for March – August 2021 is visible in black.



To view the energy consumption, hover over the particular data point.



Using the Highcharts inbuilt features, user can display the visualization for actual data or forecasted values by disabling the other graph.

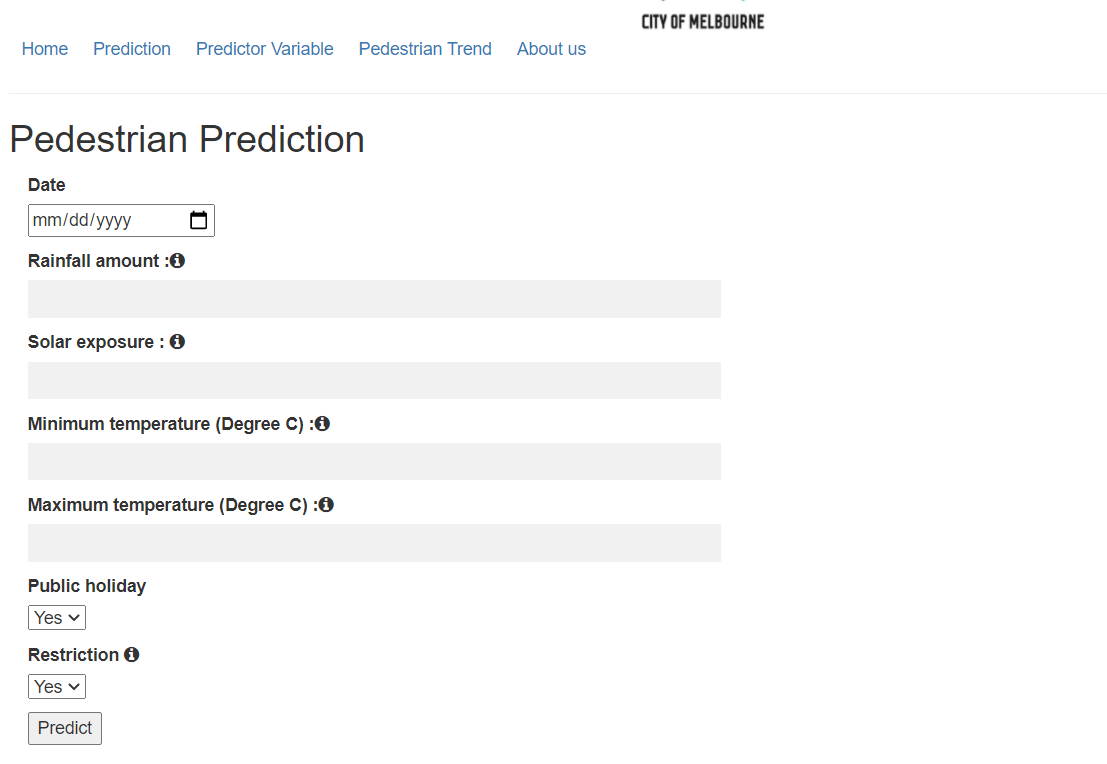


## Pedestrian count for specific day

To access the pedestrian prediction for the Melbourne city, hover over ‘Prediction’ and click on ‘Pedestrian Prediction’.



For the prediction of the pedestrian count for specific day, the Pedestrian Prediction page is accessed. The page contains form element that accepts independent variables as inputs and on submission of the form user can predict the day’s pedestrian count based on the supplied features.



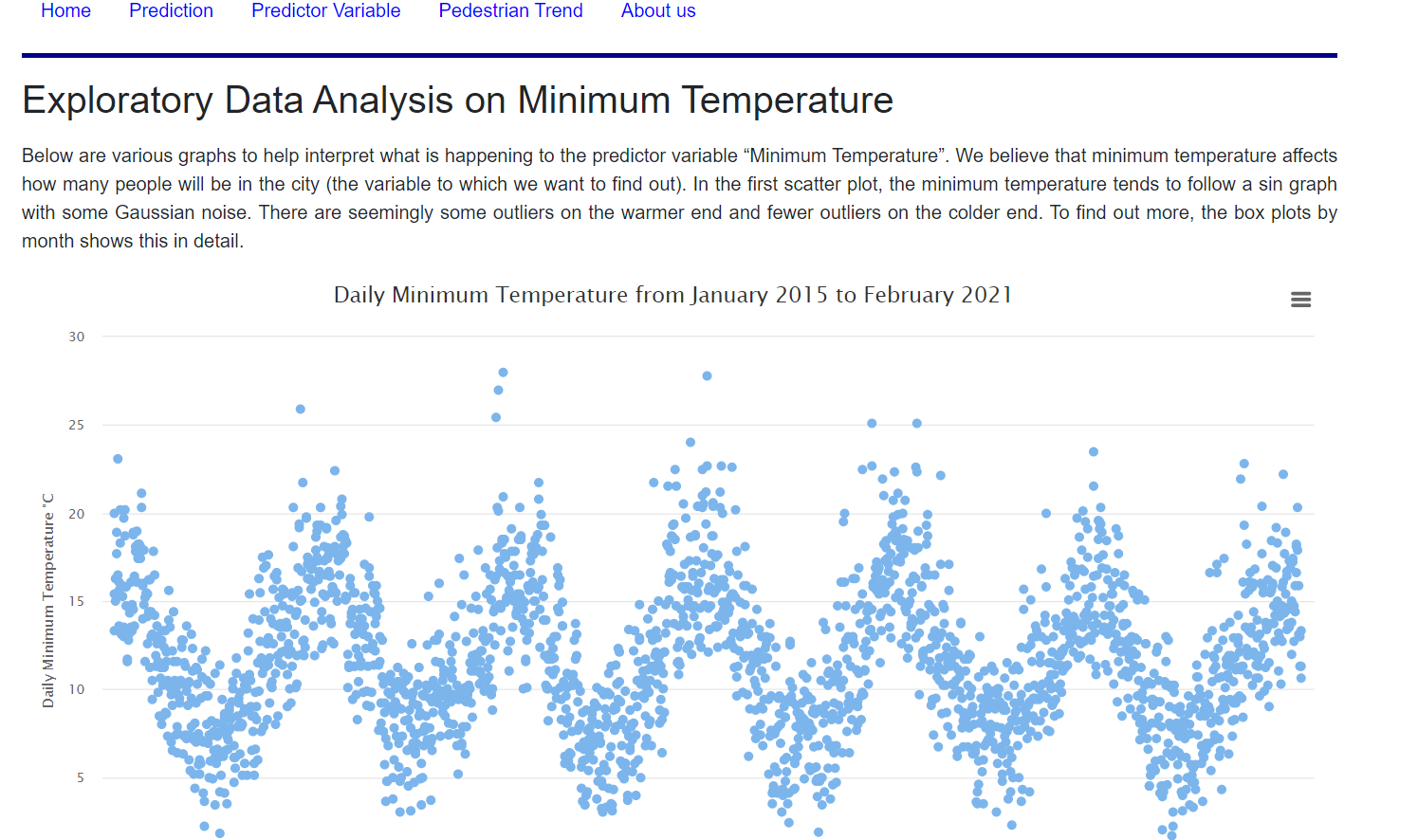
## Forecast for Independent variables

To access the exploratory data analysis of independent variables, hover over ‘Predictor Variable’ and click on any variable for more information.

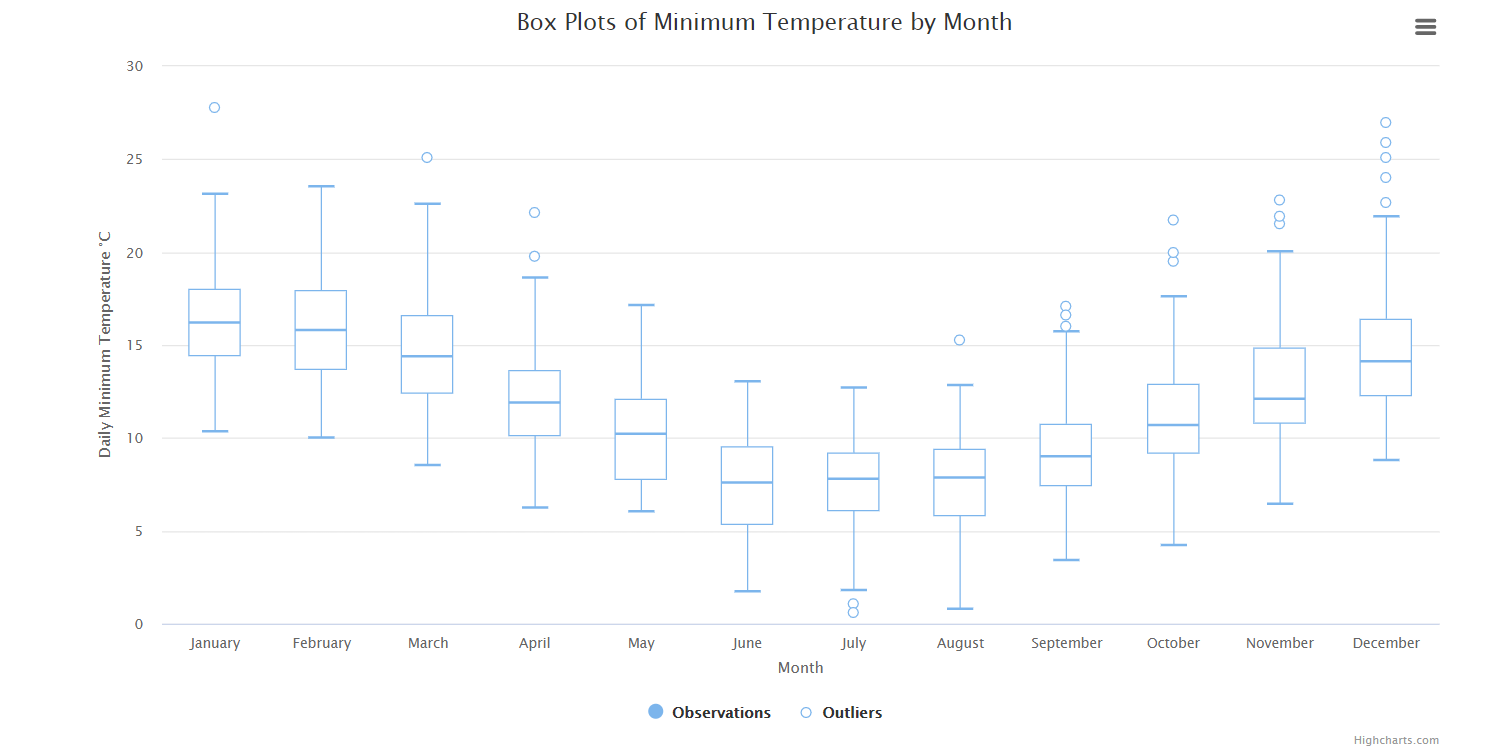


Exploratory Data Analysis for ‘Minimum Temperature’:

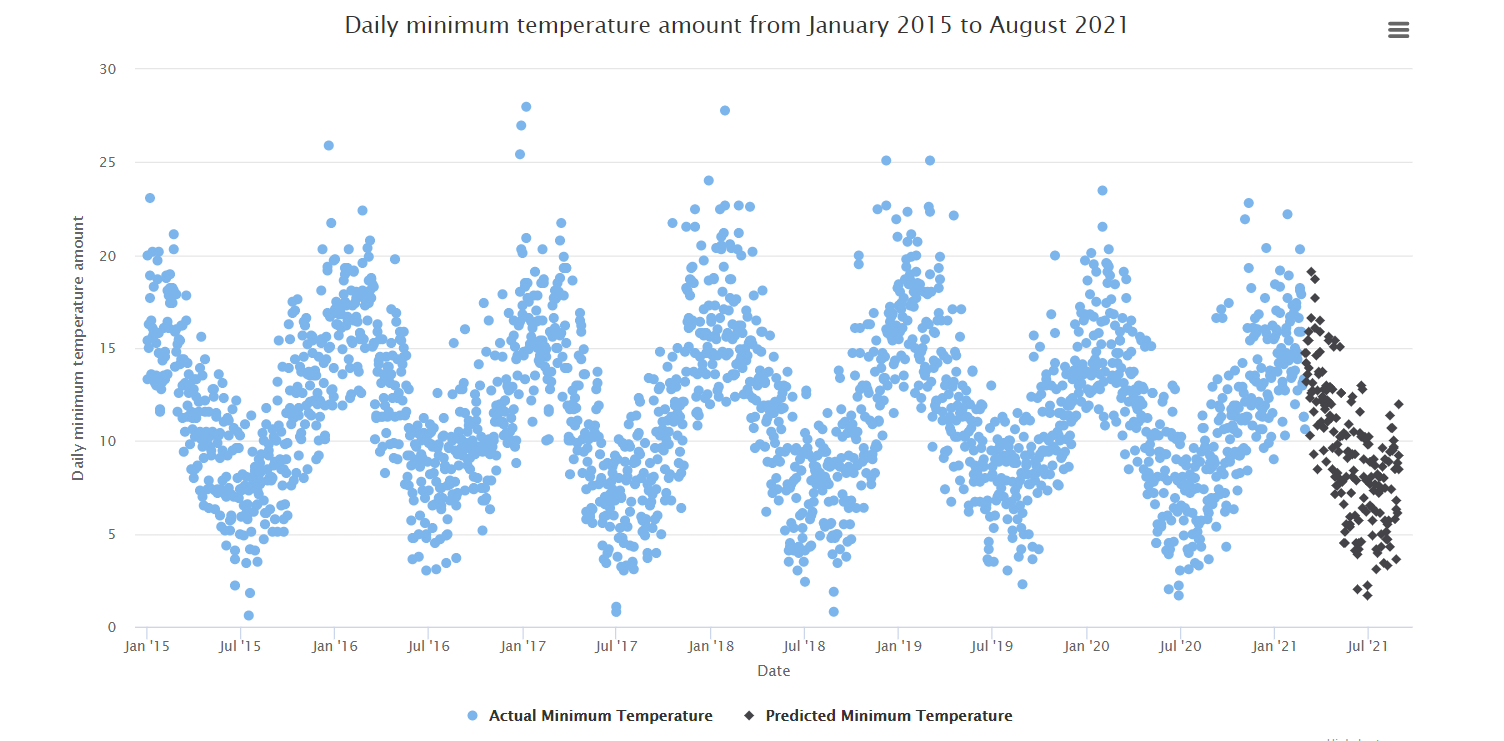
Actual data collected from Bureau of Meteorology for January 2015 to February 2021 is depicted in blue.



Distribution of data using box plot:

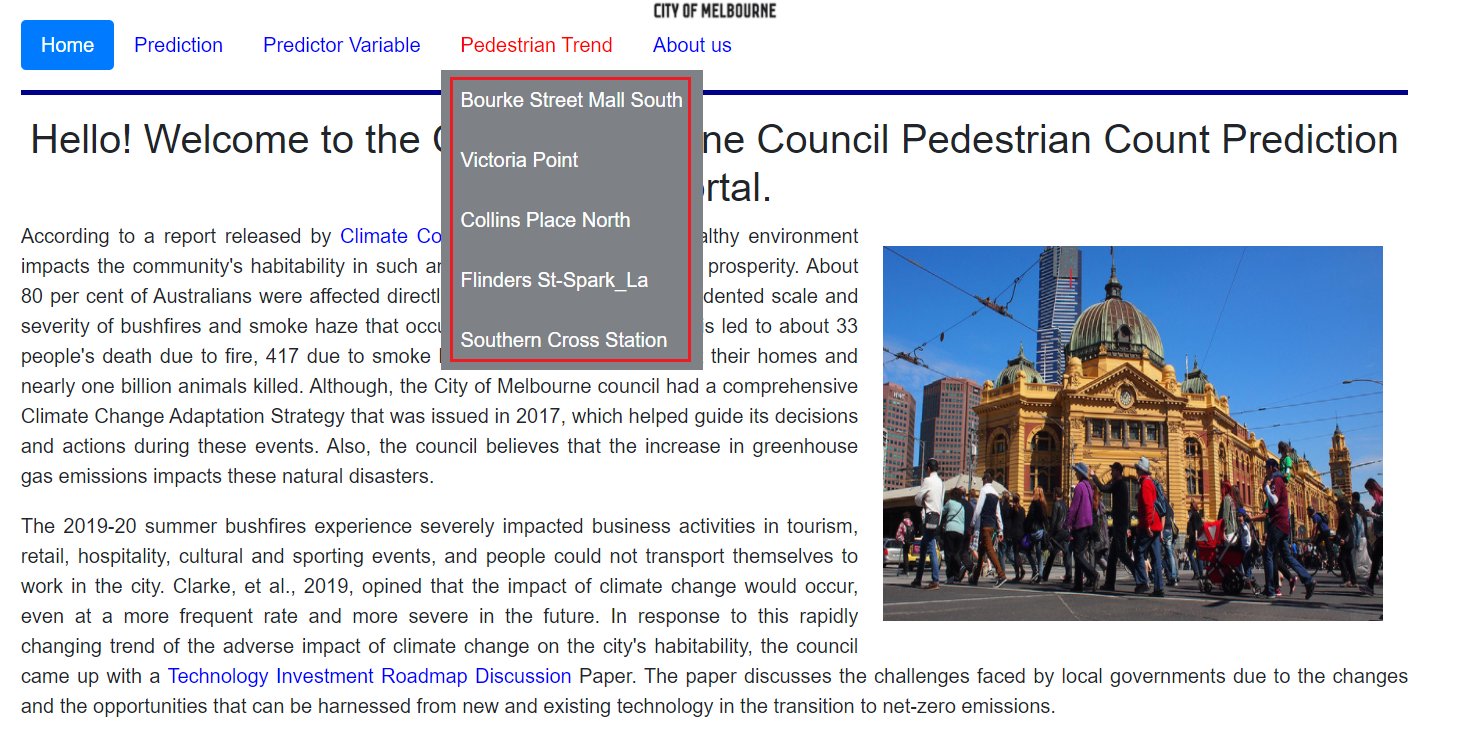


Prediction for Minimum Temperature is depicted is in black.



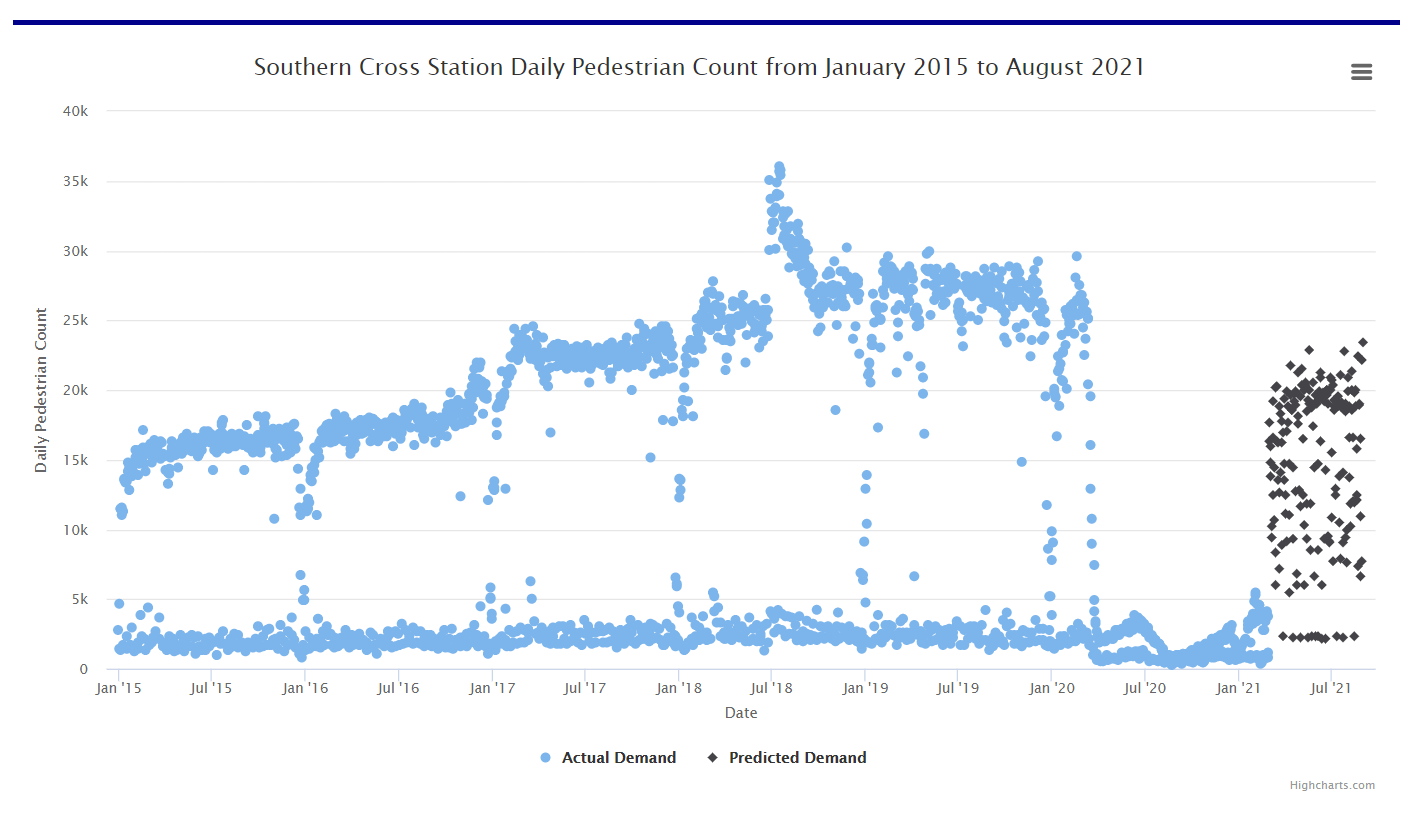
## Prediction of Pedestrian Count at granular level

To access the granular level prediction, hover over ‘Pedestrian Trend’ and click on location for more information.



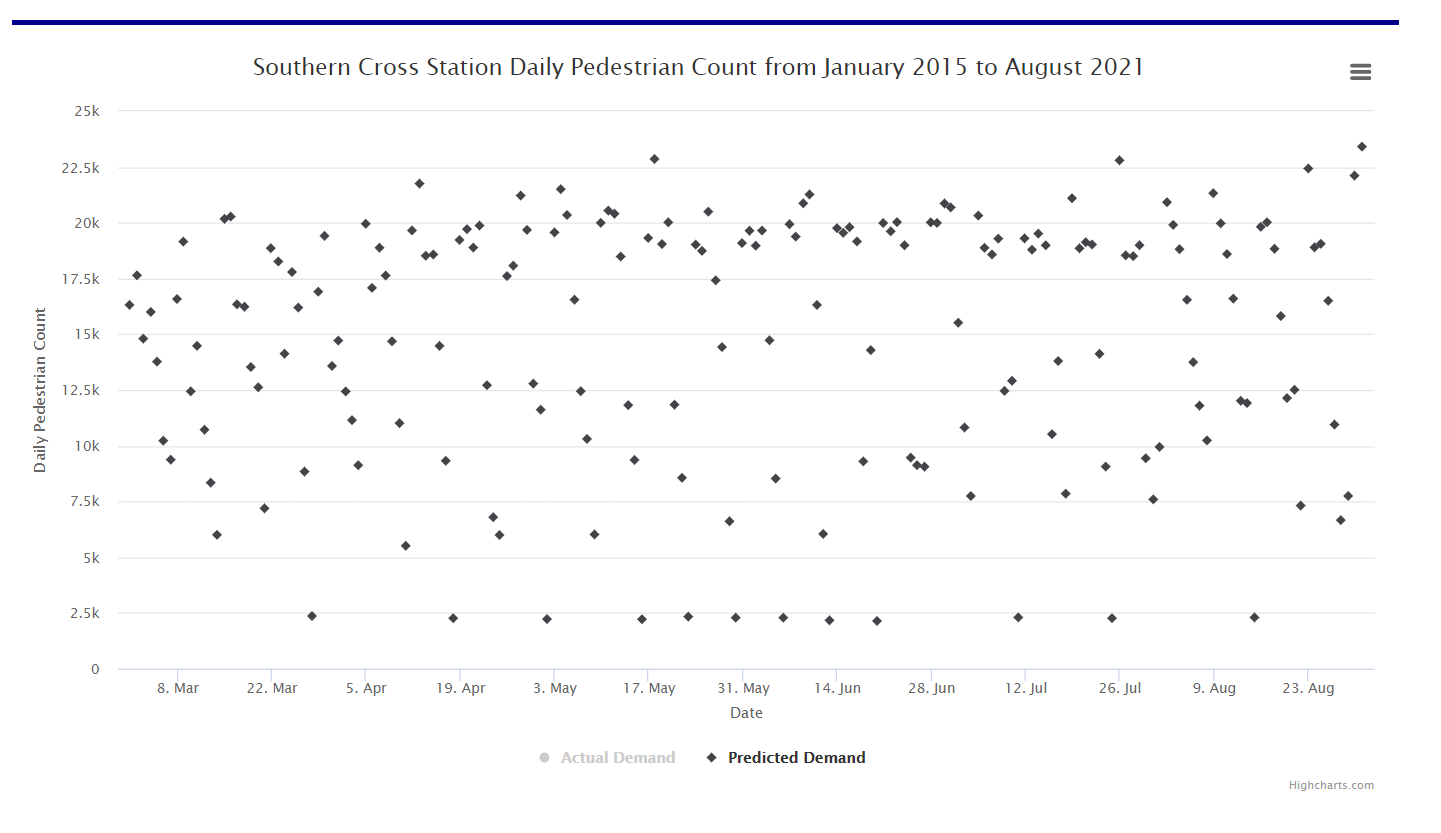
Pedestrian count prediction for Southern Cross station:

The actual data is collected from the sensor located in Southern Cross station and is depicted in blue. The prediction of the pedestrian count is depicted in black.

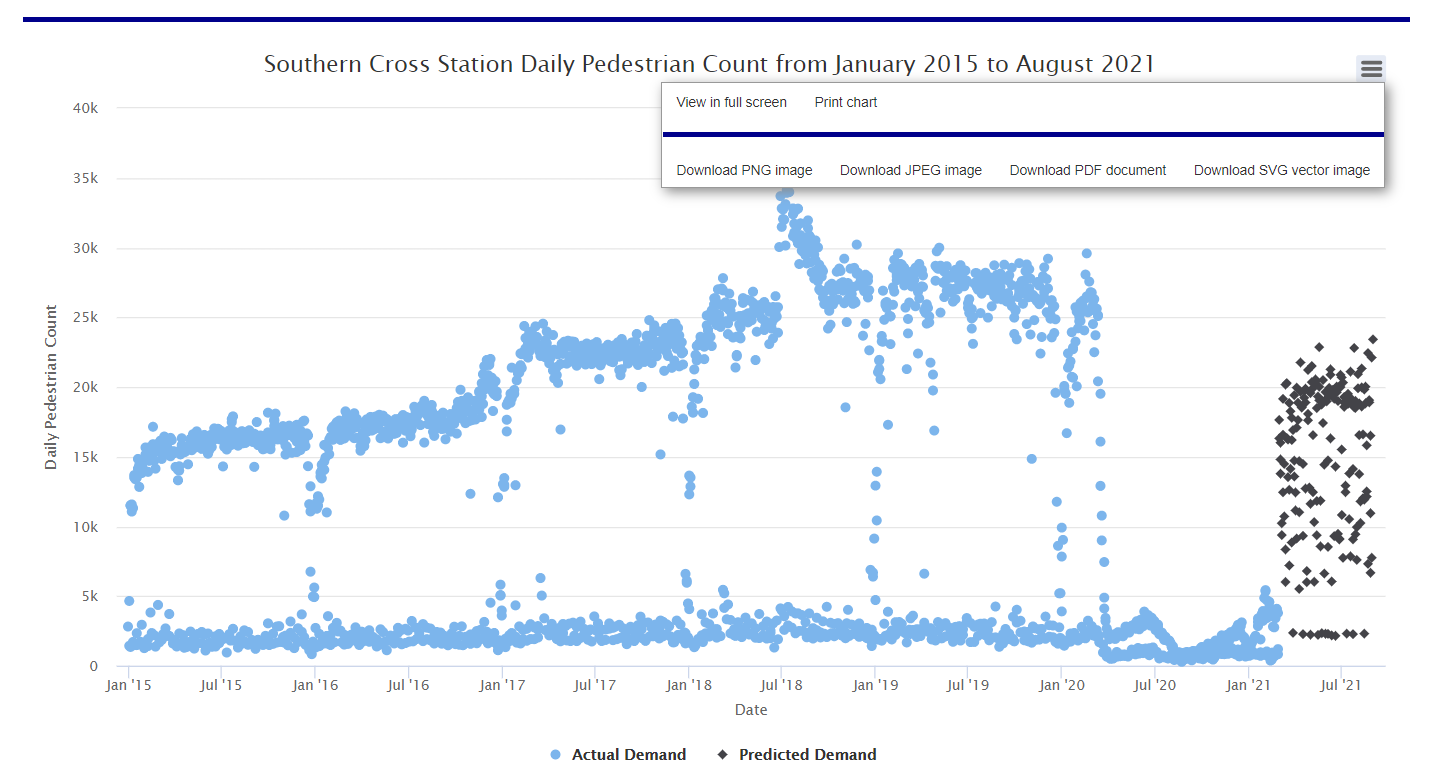


## Features of Highcharts

User can disable a scatter plot by clicking on the name of the graph:



Using Highcharts user can view the graph in full screen, print the visualization, download the visualization as image and pdf file. Click on the menu icon to access the features.



## Video for user navigation

Please click on the following link to view the entire product and the user navigation through the website:

[User Manual](https://deakin365.sharepoint.com/sites/Data2IntelligenceConsulting/Shared%20Documents/D2I%20(Melbourne%20City)/T1%202021/Handover%20Document%20and%20Video/User%20Manual.mp4)

# Completed Deliverables

Provide a list of product features and/or deliverables, including a brief description, that have been completed this trimester. Please relate these deliverables to their corresponding Trello cards if this is possible.

Only include features and/or deliverables that are fully complete – incomplete work is to be listed in section 4. Roadmap.

Make sure to explicitly highlight which features and/or deliverables where completed this Trimester and which Squad member(s) were primarily responsible for their completion.

Also, please indicate where each of the completed deliverables can be found (E.g., MS Teams, Bitbucket repository) and make sure to include a URL link to the resource.

## 4.1 Overall Deliverables

## 4.2 Sub-Team A Deliverables

This sub team was responsible for delivering two statistical model for Melbourne city pedestrian count prediction capability services. A regression model and a time series model were provided for integration into two features of the website.

The dataset for training both types of models were obtained using the same approach. Firstly using this [notebook](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/sub_team_a/dataset/download/dataset_download.ipynb), the pedestrian count for the city of Melbourne for various sensor locations were downloaded using API endpoints from the council. The hourly pedestrian counts obtained were then aggregated to daily counts. Now, to train statistical models, we need to create features for these regression outputs. Weather data from Bureau of Meteorology were used as features. Moreover, the covid restriction and the holiday variables were also integrated. Both these features were binary in nature. For the initial stages, Aparna Chintala was responsible for manually obtaining the data from various sources and then merging them.

Because the regression model in the prior trimester didn’t yield good enough results, Rohan Amatya was responsible for identifying ways of improving the data pre-processing process, modelling process, evaluating various approaches in terms of test performance and automating the whole ETL (Extract Transform Load) forecasting pipeline.

Firstly, in terms of pre-processing, all the data were normalized(standardized). This ensured that the models were easier to train and resulted in faster convergence. Since time-series modelling was also added this trimester for dedicated automated pedestrian forecasting on a periodic basis, the pre-processing for this modelling process involved additional steps. Feature engineering was carried out as implemented in this [notebook](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/sub_team_a/time_series/neural_networks/time_series_forecasting.ipynb) to ensure that the data was transformed in a format suitable for time series modelling.

Now, multiple types of statistical regressors were trained and evaluated for both types of problems. For regression problem which was carried out by Miriam Zhu and Akhila Manchi, [ticket](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/sub_team_a/time_series/neural_networks/time_series_forecasting.ipynb) shows that the RandomForest Regressor yielded the smallest test error. The scaler(pre-processing) and the trained model obtained from this evaluation was integrated into the [website resource](http://13.250.31.141:8080/Pedestrian_prediction) to obtain the on-demand pedestrian count predictions capability.

Similarly, for the time-series forecasting multiple models were trained and evaluated. VAR(Vector Autoregression) , RNN(Recurrent Neural Network), GRU (Gated Recurrent Unit), LSTM(Long Short Term Memory), Stacked LSTM were evaluated by Rohan Amatya. [EDA](https://trello.com/c/oY8IQJXj/109-st-a-perform-eda-stationarity-check-on-the-dataset-for-time-series-forecasting-of-pedestrian-count) (Exploratory Data Analysis) was also carried out to understand the data, perform imputations, aggregate information and perform stationarity checks. This [ticket](https://trello.com/c/ie1FfIRG/110-sta-implement-time-series-forecasting-using-sequential-neural-network-model-like-rnnlstm-for-pedestrian-count) and notebooks [1](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/sub_team_a/time_series/neural_networks/time_series_forecasting.ipynb), [2](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/sub_team_a/time_series/neural_networks/multistep_time_series_forecasting.ipynb) can be viewed for reference. Here, windowing technique was used for feature engineering. Since the dataset was 2 weeks late and we were tasked with providing prediction of at least 1 week from the current date, we needed to predict for 3 weeks in reality. This prediction of 3 weeks has been brought down from the considerable timeframe of 6 months as done in the prior trimester based on advice from Rohan where he mentioned that predictions would be more uncertain as the time difference increased. Finally, the single-step multivariate time series forecasting was converted to multi-step to incorporate the mentioned changes. This [web resource](http://13.250.31.141:8080/Pedestrian_forecast) shows the customer-facing exposure of the feature implemented.

For the integration of both the features into the website, Ayodeji Ladeinde aided the respective sub-team members.

Now, most projects would consider data collection and prediction as a one-time process. However, to add value to the client we automated the whole [ETL prediction pipeline](https://trello.com/c/b6u8tSRM/137-st-a-automated-prediction-pipeline), which would periodically refresh the [dashboard](http://13.250.31.141:8080/Pedestrian_forecast). This was envisioned, planned and implemented by Rohan. Here, a complete end-to-end cloud solution was used to automate the pipeline in order to make the process robust and at the same time reduce cost as much as possible. The data collection and the model training are an expensive process. For manual training, google colab was used which provided computation resources and GPUs. However, when using dedicated servers of similar configurations, these would yield high cost. So, in order to automate the pipeline and reduce the server cost, the automatic ETL pipeline was implemented using AWS services.

A [lambda function](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/sub_team_a/etl_automation/lambda/server_start.py) is used to start the resource-intensive server in the stipulated time in a periodically in order to perform the automation. The lambda function is triggered by a cloud watch event rule for the cron job. The entire ETL implementation is provided [here](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/sub_team_a/etl_automation/lambda/server_start.py), which is to be deployed over the server. Please note that in order to use these resources, some knowledge regarding AWS cloud services such as lambda, EC2, cloudwatch, log groups, IAM is required.

## 4.3 Sub-Team B Deliverables

The objective goals of Sub-Team B is: visualization of 5 sensor locations across Melbourne. Granular prediction pedestrian count for 5 major sensor locations within the COM. This would enhance the council’s ability to make appropriate decisions on measures to promote walking such locations. The top 5 sensor location are: Flinders street, Collins place north, Bourke Mall, Southern Cross, Victoria Point.Visualization of trend of the pedestrian count for 5 major sensor locations from the start period of when the sensor was mounted to February 2021 and prediction till August 2021.

In iteration 0, Deji, Jason and Oscar are formed sub team b group and come up with project proposal.

In iteration 1:

1. Deji Extract and analyze pedestrian count data from City of Melbourne. (Reviewed by Jason and Oscar). [Trello](https://trello.com/c/cocriQsU)
2. For weather features extraction. Oscar imported and cleaned rainfall amount; Deji did Solar Exposure; Jason did min and max temperature, respectively. [Trello](https://trello.com/c/Zy09oMZ7).
3. Produce some descriptive statistics to help you decide how to predict the independent weather features into the future. Each member in sub team b did exploration analysis for weather features. Like boxplot, histogram, pie chart, bar chart. [Trello](https://trello.com/c/WyZualdH)
4. Forecasted five sensor location

Firstly, sub-team B predicted weather features, including Rainfall, Min and Max temperature, solar exposure across city of Melbourne from Mar 2021 to August 2021. Plus, sub-team B made a prediction of pedestrian counts in the TOP 5 sensor locations, from Mar 2021 to August 2021 by these feature attributes. Secondly, exporting output to files, which enables sub-team B to use them for visualization. Finally, python flask presents these figures into [website page](http://13.250.31.141:8080/Predictor_Variable).

**Data Acquisition**: All the weather dataset can be found in [BOM](http://www.bom.gov.au/jsp/ncc/cdio/weatherData/av?p_nccObsCode=136&p_display_type=dailyDataFile&p_startYear=&p_c=&p_stn_num=086338). Pedestrians counts based on Top 5 sensor location can be found [here](http://www.pedestrian.melbourne.vic.gov.au/#date=02-06-2021&time=13).

**Data Manipulation and EDA**: Sub team B decided to use datetime index as attribute to tidying dataset. Code can be found [here](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/sub_team_b/iteration_1). Furthermore, sub team B also reported the findings into word document [here](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/sub_team_b/iteration_1_report). That is, the summary statistics of weather feature in terms of day of week, month and year.

**Machine learning:** After wrangling and tidying dataset, sub team B compared five models for each sensor location and picked the best model for predicting pedestrian counts, to get predicting value with high performance. Code can be found [here](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/sub_team_b/Pedestrian_Count_Top_5_sensor_location).

**Visualization:**  Interactive features, like sliders for independent features, are used to capture user input and pedestrians counts by HighCharts. Moreover, sub team B embedded the product into the website. Flask code can be found [here](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/Website/app.py).

**Deliverable Outcome (Sub Team B):**

* Researched variables and using reliable dataset for the project.
* Scrape and clean independent variables
* Train and forecast independent variables
* Summary statistics of weather features
* Modified the menu bar on the home page to improve user interaction.
* Added more features to the home page and improved product deliverables
* Visaulized prediction sensor counts using Highchart
* Applied different time series models, like Arima Walking forward, LSTM. The variation of feature increased by average 5% for each sensor location.
* Compare to Tableau, Highchart visualization is faster in loading the website page.

## 4.4 Sub-Team C Deliverables

The primary goal of Sub-Team C involved predicting the energy consumption for the City of Melbourne Council. This would enhance the council’s ability to make appropriate decisions as to the capacity of the renewable energy sources that can support the volume of energy being consumed within the metropolis. The trends of energy consumption was forecasted from March 2021 to August 2021. All the project deliverables from 1 – 8, have been completed during this trimester, i.e. T1 2021. Sub-Team C was responsible for completing the following project deliverables:

1. Obtaining the dataset from Kaggle

* Nikita Wadekar was responsible for this deliverable
* This can be found in [Bitbucket](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/Electricity_demand/Dataset/electricity_demand.csv)

1. Update the dataset with the appropriate parameters till February 2021

* Neet Patel, Sivaram Krishnan and Nikita Wadekar were responsible for this deliverable
* This can be found in [Bitbucket](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/Electricity_demand/Dataset/electricitydemand_data.csv) and [Trello](https://trello.com/c/cLKbgNnq/104-st-c-updating-data-set-from-7-october-2020-to-28-february-2021)

1. Performing Exploratory Data Analysis on the dataset

* Neet Patel was responsible for this deliverable
* This can be found in [Bitbucket](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/Electricity_demand/Python%20Codes/Exploratory%20Data%20Analysis%20for%20'Elec%20Demand'%20Dataset.ipynb) and [Trello](https://trello.com/c/GBWEfqtZ/105-st-c-performing-pre-processing-and-exploratory-data-analysis-on-the-electricity-demand-data-set)

1. Determining stationarity of the dataset

* Neet Patel and Nikita Wadekar were responsible for this deliverable
* This can be found in [Bitbucket](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/Electricity_demand/Python%20Codes/Stationarity%20of%20variables%20'demand',%20'RRP'%20and%20'min_temp'.ipynb) and [Trello](https://trello.com/c/QDQFLrxr/129-st-c-checking-for-stationarity-in-the-dataset)

1. Develop and run multiple multivariate-time series models

* Neet Patel, Sivaram Krishnan and Nikita Wadekar were responsible for this deliverable
* This can be found in [Bitbucket](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/Electricity_demand/Python%20Codes/Stationarity%20of%20variables%20'demand',%20'RRP'%20and%20'min_temp'.ipynb) and [Trello](https://trello.com/c/y3s0mrfM/106-st-c-develop-and-run-the-time-series-model-on-the-energy-consumption-data-set)

1. Forecast values from March 2021 to August 2021

* Neet Patel, Sivaram Krishnan and Nikita Wadekar were responsible for this deliverable
* This can be found in [Bitbucket](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/Electricity_demand/Python%20Codes/Stationarity%20of%20variables%20'demand',%20'RRP'%20and%20'min_temp'.ipynb) and [Trello](https://trello.com/c/e4mXOENZ/130-st-c-forecast-the-energy-consumption-from-march-2021-to-august-2021)

1. Develop a dashboard to present the predicted electricity demand

* Neet Patel and Nikita Wadekar were responsible for this deliverable
* This can be found in [Bitbucket](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/Website) and [Trello](https://trello.com/c/sy8fH5Fz/107-st-c-develop-dashboard-using-highcharts-for-forecasted-electricity-demand)

1. Incorporate the dashboard in the webpage

* Sivaram Krishnan was responsible for this deliverable
* This can be found in [Bitbucket](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/Website) and [Trello](https://trello.com/c/6x0yJFsk/108-st-c-design-and-develop-a-new-web-page-electricity-demand-forecast)

All aspects of the project deliverables were carried out successfully by Sub-Team C.

**Iteration 0**

The first task in Iteration 0 involved obtaining the ‘Daily Electricity Price and Demand Data’ dataset from Kaggle (<https://www.kaggle.com/aramacus/electricity-demand-in-victoria-australia>). The Kaggle dataset consisted of 14 attributes and 2,016 observations. An observation was recorded at the end of each day. Subsequently, the next step involved extracting the known data from the Bureau of Meteorology (BOM) and updating the dataset for the attributes. This was completed as the Kaggle dataset was limited from 1 January 2015 – 6 October 2020. Hence, BOM was utilised to update the dataset values from 7 October 2021 to 28 February 2021. After extracting the known data and updating the dataset, important attributes with the dataset were analysed. To manage these tasks, each team-member was assigned specific tasks on the Trello board. It was unanimously decided with the sub-team to leave comments under the relevant card in Trello, detailing the work that was completed.   
After each team-member had successfully carried out their task, the next step involved dropping certain attributes such as ‘demand\_pos\_RRP’ and ‘RRP\_positive’ from consideration as they displayed weak correlation with other variables and were not important factors in being able to accurately predict electricity demand.

**Iteration 1**

At the commencement of Iteration 1, team-members were responsible for updating the ‘RRP’ and ‘Demand’ attribute from October 2020 to February 2021. These values were obtained from the Australian Energy Market Operator (AEMO). The trends with the electricity demand were visualised through multiple plots. Moreover, the dataset was also updated in terms of allocating values for the attributes ‘Public Holiday’ and ‘School Day’ from October 2020 – August 2021.   
Subsequently, the next task involved researching Exploratory Data Analysis (EDA) to obtain a better understanding of the dataset and determine the underlying trends within the dataset. Other work being done simultaneously included obtaining the summation of the electricity demand from verified data sources and predicting the RRP values using multivariate time-series models such as the Facebook Prophet model and VAR model. The trends in the RRP were also visualised by plotting multiple graphs. The final work in Iteration 1 involved determining the stationarity of the dataset through visualisations and by conducting the Augmented Dickey Fuller test. It was noticed that there were extreme outliers in the RRP dataset. As a result, these outliers were removed and replaced with the median RRP value. Discussions were held about researching several multivariate time-series models to predict the energy consumption. It was decided by the team to use values from January 20215 – December 2020 as the ‘Training data’ and January 2021 – February 2021 as the ‘Testing data’. The team specifically focused on tuning the parameters to obtain accurate performance metrics in terms of Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and R2 score.

**Iteration 2**

In Iteration 2, each team-member was allocated multivariate time-series models to research and implement on the dataset. Some of these models included: Prophet model, Vector Auto Regression (VAR), VAR with Moving average, Simple RNN Model, LSTNet Model, Long Short-term Model (LSTM), VAR moving average with Exogenous regressors. Out of these models, some models could not be implemented on the dataset due to their complexity/lack of resources, weak performance metrics and inaccurate predictions. As a result, these models were dropped from consideration. Out of these models, the SimpleRNN, Prophet and LSTM model were used in predicting the energy consumption from March 2021 – August 2021. Each team-member was allocated a model, and they were implemented to forecast the energy consumption. The model which showed the best performance in terms of the forecasted graph and performance metrics was selected to be the model to add to the webpage, and ultimately present to the client. Prior to iteration 2, it was decided that the Sub-team C would utilise ‘Tableau’ for visualisation. However, due to suggestions from Sub-team C members and other squad members, it was then decided that ‘Highcharts’ would be utilised to visualise the forecasts. Subsequently, the code for the forecasted graph was prepared using JavaScript. The forecasted graph was then integrated on the energy prediction webpage (<http://13.250.31.141:8080/Energy_forecast>).

# Roadmap

The primary objective of the tribe Data to Intelligence (D2I)-Melbourne City is to analyse the datasets available from the City of Melbourne (COM) data source and develop algorithm to provide valuable information to the council which will assist in decision making. In the trimester 3 2020, the squad has focused on one of the council's long-term goals, which is, achieving a net-zero emission by 2040. The squad visualized the pedestrian traffic and forecasted the pedestrian count from November 2020-March 2021. The relevant data visualizations assisted the council to make appropriate decisions for achieving net-zero emission by 2040.

In the trimester 1 2021, the squad decided to focus on improving the machine learning model for pedestrian count and adding new feature ‘Energy consumption forecast’. During the client meeting, Mr. Mcintosh requested for granular level pedestrian traffic across the city of Melbourne. The squad selected 5 sensor locations, to visualize the data distribution and to predict the pedestrian count for these locations. The collection of data, transforming data, executing the model and developing the data board was a manual process. However, with automating the ETL prediction pipeline, the team decided to add great value to the project. After implementation of automation, no human intervention would be needed to get the latest forecast. Tableau loads slowly on the website due to which the team decided to explore and use Highcharts. After several deliberations, the squad formed a Work Breakdown Structure in iteration 0 and divided into sub teams.

The deliverables that were planned for implementation were completed and presented by the end of this trimester. The product gives the client an overall historical view of pedestrian data traffic, facilitate future pedestrian count for next three weeks in Melbourne city, improved prediction for specific day using independent variables, granular level information of 5 sensor locations and energy consumption forecast till August 2021. The most valuable deliverable is the automation ETL forecasting pipeline which was successfully implemented in this trimester.

Following are certain deliverables which are planned for the next trimester to enhance the project features:

* Updating the pedestrian count forecast for Melbourne to remove the zero days from the visualization.
* Automating the pedestrian count prediction for the 5 sensor locations.
* Analysing the pollution data (<https://aqicn.org/map/australia/>), weather data (<http://www.bom.gov.au/>) and impact of pollution levels in Victoria (<https://www.abs.gov.au/statistics/health/causes-death/provisional-mortality-statistics/latest-release>) using Bayesian network. Develop machine learning algorithm to make appropriate predictions which will assist the client towards the goal of net zero emission.
* Analysing the revenues of the business and the impact of COVID-19 pandemic in Melbourne CBD (<https://www.abs.gov.au/statistics/economy/business-indicators/business-conditions-and-sentiments/latest-release>) using Bayesian network. Develop machine learning algorithm to make appropriate predictions to assist the council to take measures to increase revenue of the businesses.
* Improving the homepage and user interface of the website.

# Open Issues

List all of the issues and challenges that the team is still facing, and any progress that has been made so far to address them.

The purpose of this section is to flag things that may interfere with the future squad’s ability to work on the project, and to give advice as to how these issues could be fixed in future.

Here are some examples of Open Issues:

* Software compatibility issues that arise when members of the team use different version of software.
* An unclear process for reviewing completed tasks on Trello, leading to a backlog of work that is sitting somewhere between unfinished and finished.
* An essential team member had to leave the squad with no notice, and there is currently a skill void in their place.

# Lessons Learned

List key lessons learned from the project this Trimester and what you recommend future squads should do differently. You must also explain why you believe this to be the case.

In particular, try to think about processes or technology that you would recommend be changed in the future; things that an uniformed squad may mistake for a good idea at first, but later learn to be ineffective.

For example, maybe your team had challenges communicating their progress during panel presentations, but towards the end of the Trimester, you developed an effective method for conveying progress accurately. This would be a great thing to talk about.

# Product Development Life Cycle

This section should explain how your team undertakes work. It is an attempt to codify your processes so that they can be understood and followed by new members.

As a squad, you may not have clearly defined your Product Development Life Cycle, and that’s okay! This is an excellent opportunity to explain the work methods, processes and habits that your team has been developing intuitively over the course of the Trimester.

## 8.1 New Tasks

How are new tasks created?

How does your squad form new ideas about work that needs to be done and turn those ideas into distinct, actionable tasks?

For example, maybe your team meets at the start of each week, reviews your progress in your current sprint, makes a big long list of everything to be done, and then converts that list into a series of cards on Trello. This process would be something you talk about in this section.

## 8.2 Definition of Done

How does the team know when a task is done?

What are criteria for a successfully completed task?

This may seem obvious, but it in a software development project having a definition of done can ensure a certain standard of work that holds all team members accountable. For example, messy, clunky code that “just works” is very different to clean, well-commented code that works AND is easy to understand. Which would you prefer to be your team’s definition of done?

## 8.3 Task Review

Who reviews a task once it’s been marked as done?

How does the team ensure that all work is looked over before it’s contributed to the main repository or working prototype?

If you don’t currently have a system for reviewing tasks, make sure to flag this for next trimester’s team to work on as soon as they begin.

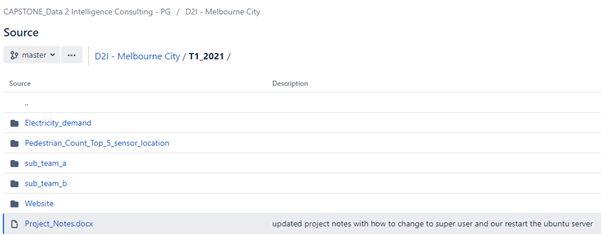
## 8.4 Testing

Product testing was completed by ensuring that forecasted graphs in Python were completely replicated on the webpage. By comparing the shape of the forecasted graphs in Python against the graphs in Highcharts, we could be confident that the product accomplished what it was originally planned to do. As Highcharts is interactive (hovering over a particular observation reveals its y-axis value), we were also able to confirm the predicted values from the webpage against the predicted values obtained from Python. In the table below, the similarity between the 2 graphs and the values for the Energy consumption forecast can be observed.

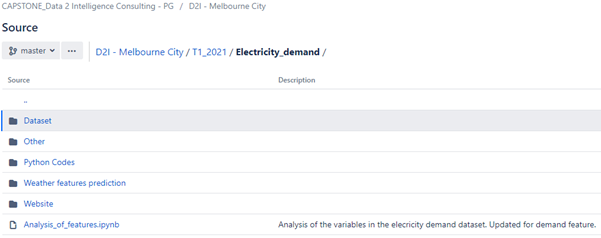
|  |  |
| --- | --- |
| **Python Forecasted Graph** | **Highcharts Forecasted Graph** |
|  |  |

## 8.5 Branching Strategy

Within the D2I – Melbourne City master-branch, there are currently 4 main folders. For this trimester, the squad created a new folder ‘T1\_2021’ to showcase the tasks being completed. Within this folder there are 5 main folders: one for each sub-team, one for the Pedestrian count and one for the website. The bitbucket contains all the evidence of the work completed during Trimester 1 2021, from python file, csv files, docx files, html files and js files.



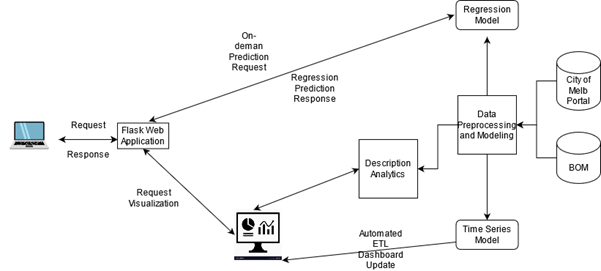
Within each folder, there were multiple folders such as ‘Dataset’, ‘Python codes’, ‘Website’, etc. The squad had decided to commit files to the relevant sub-team folder, to ensure that they Bitbucket is laid out neatly, instead of creating a messy and unorganised branch. Additionally, using an organised and cleaner format, makes it easier to locate certain files.



During the initial weeks of the project, it was recommended that every squad-member commit their work after the successful completion and review of allocated tasks. It was also recommended that relevant comments be made in the Bitbucket, which allows other squad-members to understand the type of work completed.

# Product Architecture

## 9.1 UML Diagram



## 9.2 Tech Stack

The whole of the project can be divided into two folds: (i) the customer facing web application and (ii) the offline data analytics eco-system.

For the website, an AWS ec2 instance (t2.micro) was used for hosting. This server is of free tier and has less computation and storage capacity. AWS was chosen as it is used extensively in industries and gaining experience on this would be beneficial for future career. The web application was implemented using full stack technologies such as flask, HTML, CSS, java script. Flask is a web application framework for python than is easy to understand and helps for rapid prototyping. Interactive dashboard was created using High Charts which processed the flat files created by the downstream analytical process.

Now for the data analytics processes, google colab was used mostly for the exploratory data analysis and manual data modelling (model training and evaluation). Google colab provides us with power full Nvidia Tesla GPU and 12GB of RAM which were highly beneficial to implement and test our workflow. It also provides support for python modules such as scikit-learn, Tensorflow and Keras out-of-the-box which was extensively used for the model(s) training and evaluation. These libraries helped to compile, train and evaluate machine learning and deep learning models. Moreover, the python module pandas was also used extensively to process the flat files.

Now, for the automated ETL forecasting pipeline, since the free-tier ec2 instance having RAM of only 1GB wouldn’t be adequate to perform the training of neural networks, we had to use a more expensive t2.xlarge AWS ec2 instances. This was chosen due to its multi-core and 16GBs of RAM which helped to perform the training process. To reduce cost, we also used AWS lambda function to start the ec2 instance to execute the periodic forecasting ETL pipeline. The lambda function was triggered using AWS cloudwatch event rule which is where we enter the cron job execution timing rules.

# Source Code

**Data Source**

The pollution data:

* <https://aqicn.org/map/australia/>

Weather data:

* <http://www.bom.gov.au/>

The impact of pollution levels in Victoria:

* <https://www.abs.gov.au/statistics/health/causes-death/provisional-mortality-statistics/latest-release>

The impact of COVID-19 pandemic in Melbourne CBD:

* <https://www.abs.gov.au/statistics/economy/business-indicators/business-conditions-and-sentiments/latest-release>

Daily Electricity Price and Demand Data:

* <https://www.kaggle.com/aramacus/electricity-demand-in-victoria-australia>

**Deliverables**

Main website:

* [Browse CAPSTONE\_Data 2 Intelligence Consulting - PG /D2IC-PG/repos/d2i---melbourne-city/browse/T1\_2021/Website/templates](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/Website)

Pedestrian count top 5 sensor location:

* [Browse CAPSTONE\_Data 2 Intelligence Consulting - PG /D2IC-PG/repos/d2i---melbourne-city/browse/T1\_2021/Pedestrian\_Count\_Top\_5\_sensor\_location](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/Pedestrian_Count_Top_5_sensor_location)

Pedestrian

* [Browse CAPSTONE\_Data 2 Intelligence Consulting - PG /D2IC-PG/repos/d2i---melbourne-city/browse/T1\_2021/sub\_team\_a](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/sub_team_a)

# Login Credentials

* Bitbucket and Trello uses Student Login Authentication

Repository Link: <https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse>

* AWS account for ec2 instance.

Note: Most of tools using in this trimester are private. Read document carefully.

# Appendices

Link to the team’s showcase video:

* <https://deakin365.sharepoint.com/:v:/r/sites/Data2IntelligenceConsulting/Shared%20Documents/D2I%20(Melbourne%20City)/T1%202021/Handover%20Document%20and%20Video/User%20Manual.mp4?csf=1&web=1&e=xfKT5l>

Link to our Trello board

* https://trello.com/b/sxBfi5DY/d2i-melbourne-city-trello

Link to Trello roadmap

* https://trello.com/b/unvtjPUX/d2i-melbourne-city-roadmap

Link to BitBucket repository

* [https://bitbucket-students.deakin.edu.au/projects /D2IC-PG/repos/d2i---melbourne-city/browse/T1\_2021](https://bitbucket-students.deakin.edu.au/projects/D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/)/

Link to the live website

* [http://13.250.31141.8080/](http://13.250.31.141:8080/)