Project Handover Document

Deakin 2 Intelligence Consulting
Melbourne City

Trimester 1, 2021

Table of Contents

1. Project Information

1.1 Client/Product Owner

Mr. Will Mcintosh
Architecture and Data capability Manager
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

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

2. 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% of Australians were affected. 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 3 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 a recovery plan with consistency.

This project aims to assist the City of Melbourne council's decisions to encourage Melbournians to adopt measures that facilitate reduction of carbon content within the city, organize events within the city to improve the financial status of the businesses affected due to COVID-19, 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 billion of Government investment over the next 10 years and for low emissions technologies drive at least $70 billion of total new investment in Australia by 2030.

The project showcases visualisation of prediction of total pedestrians' count for the next 3 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.

3. User Manual

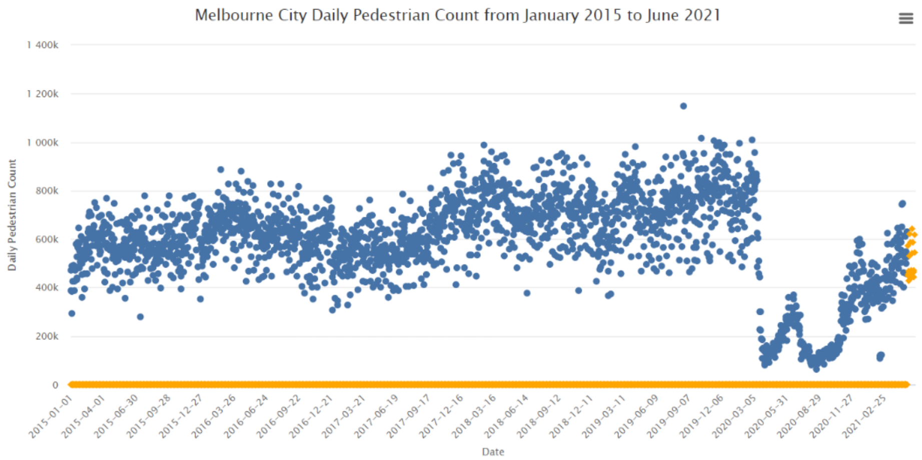
3.1. Pedestrian count in Melbourne city

Homepage:

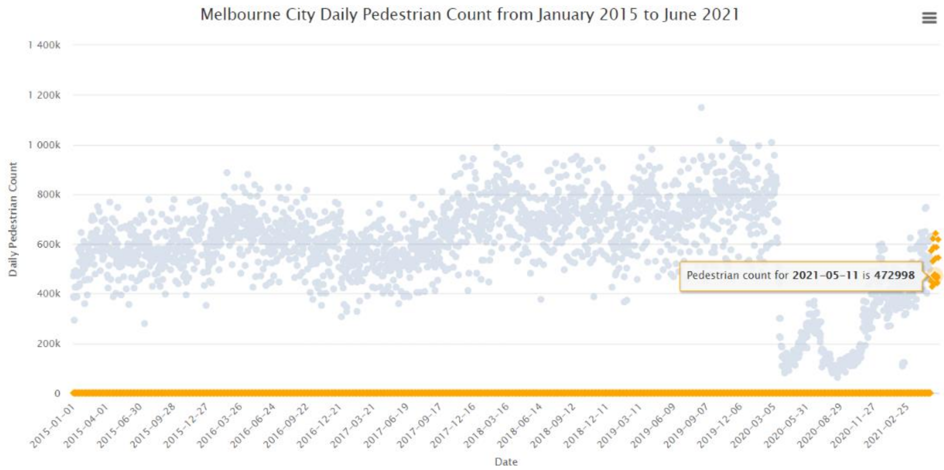
To access the pedestrian forecast for the Melbourne city, hover over 'Prediction' and click on 'Pedestrian Forecast'.

Dashboard for the pedestrian forecast:

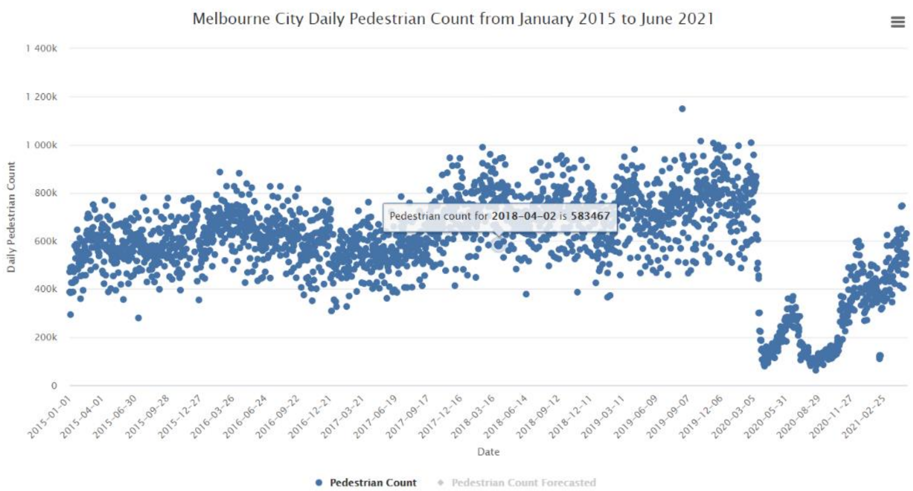
The pedestrian forecast for next three weeks is visible in orange while the actual data from 72 pedestrian sensor location between January 2015 till June 2021 is depicted using blue. The visualization is updated in real time through an automated ETL forecasting pipeline.



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

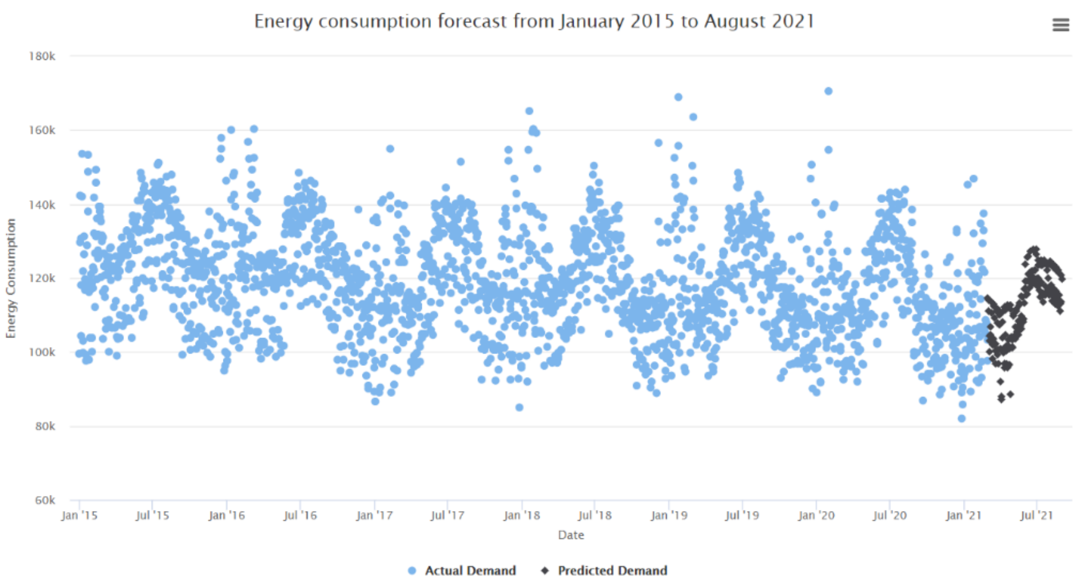


3.2. Energy consumption

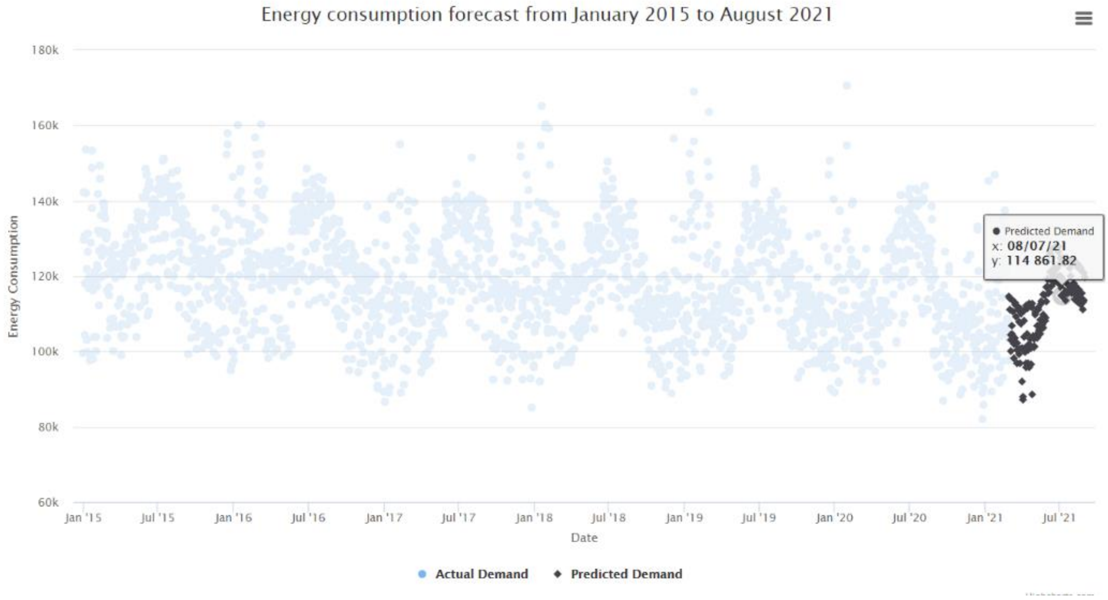
To access the energy consumption prediction for the Melbourne city, hover over ‘Prediction’ and click on ‘Energy Forecast’.

Dashboard for energy consumption prediction:

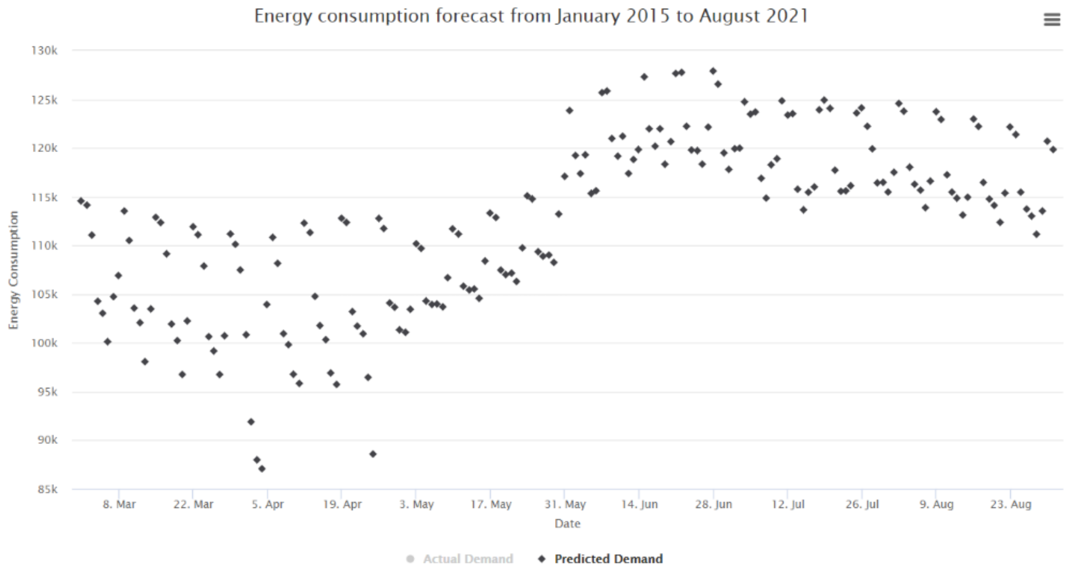
The actual data from AEMO between January 2015 – February 2021 is depicted in blue while the energy consumption forecast for March – August 2021 is visible in black.



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



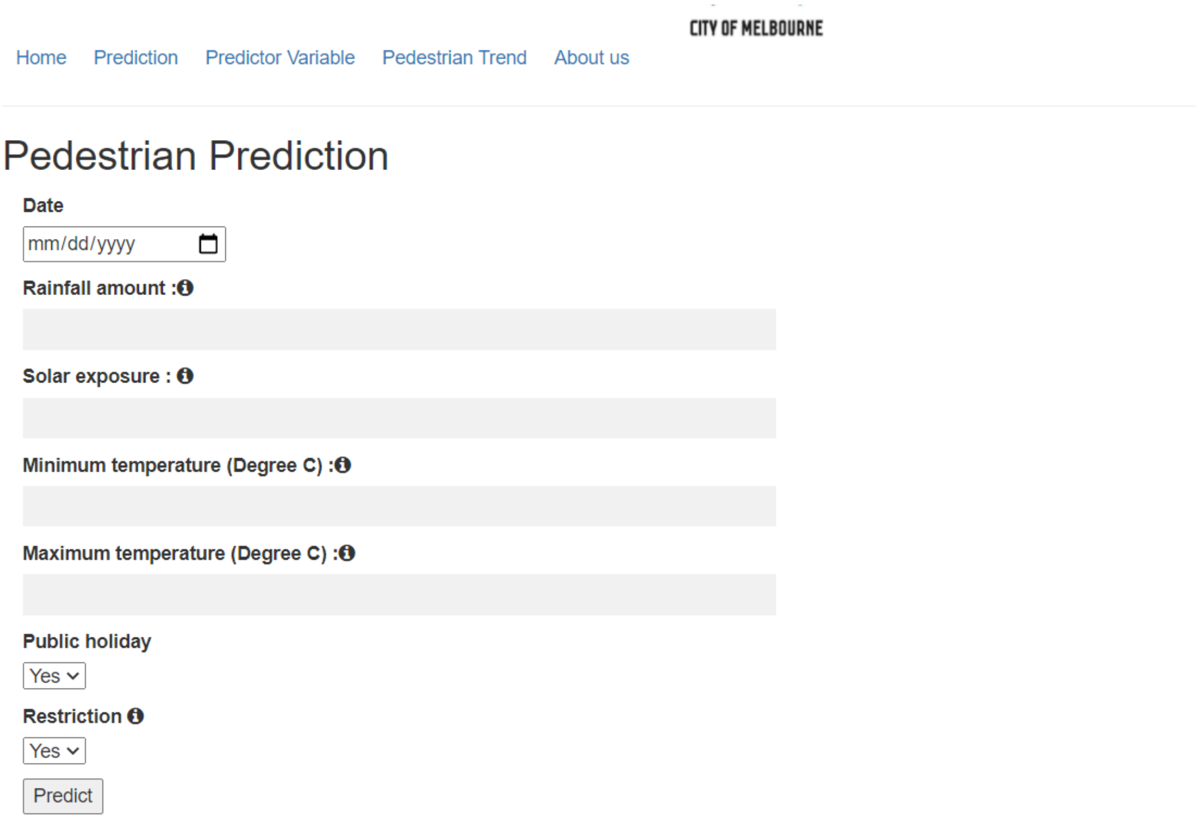
Using the Highcharts inbuilt features, user can display the visualization for actual data or forecasted values by disabling the other graph.



3.3. 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.

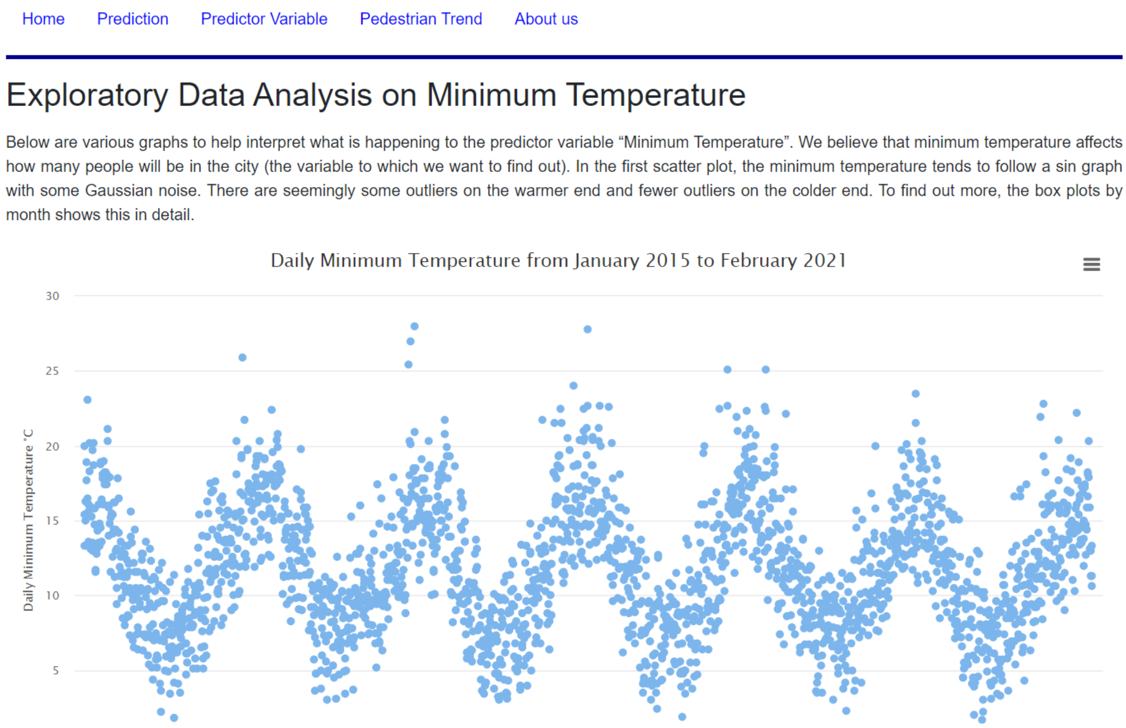


3.4. Forecast for Independent variables

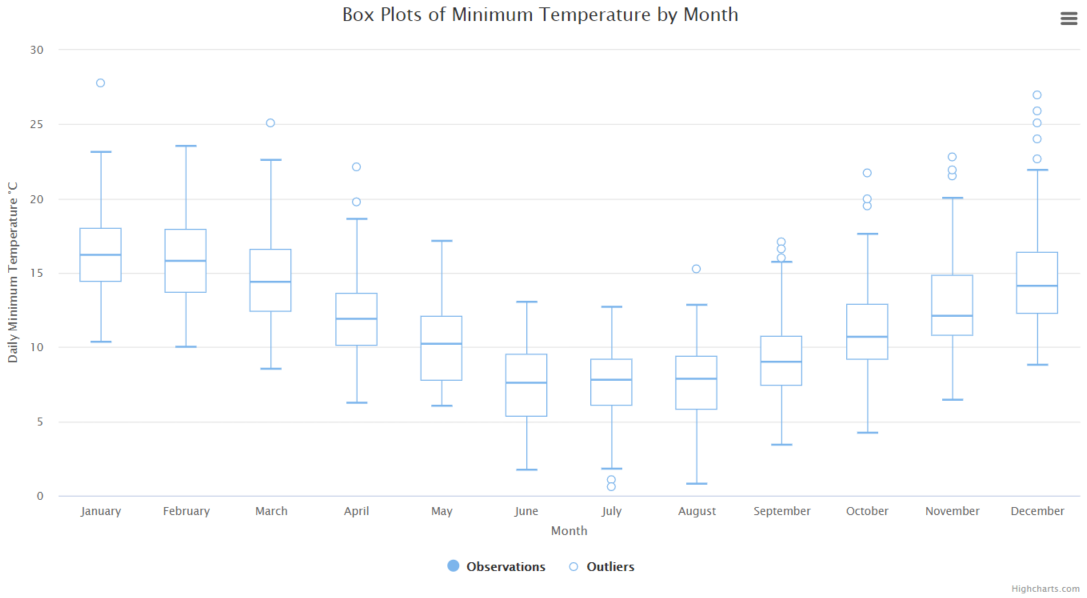
To access the exploratory data analysis of independent variables, hover over ‘Predictor Variable’ and click on any variable for more information.

Exploratory Data Analysis for ‘Minimum Temperature’:

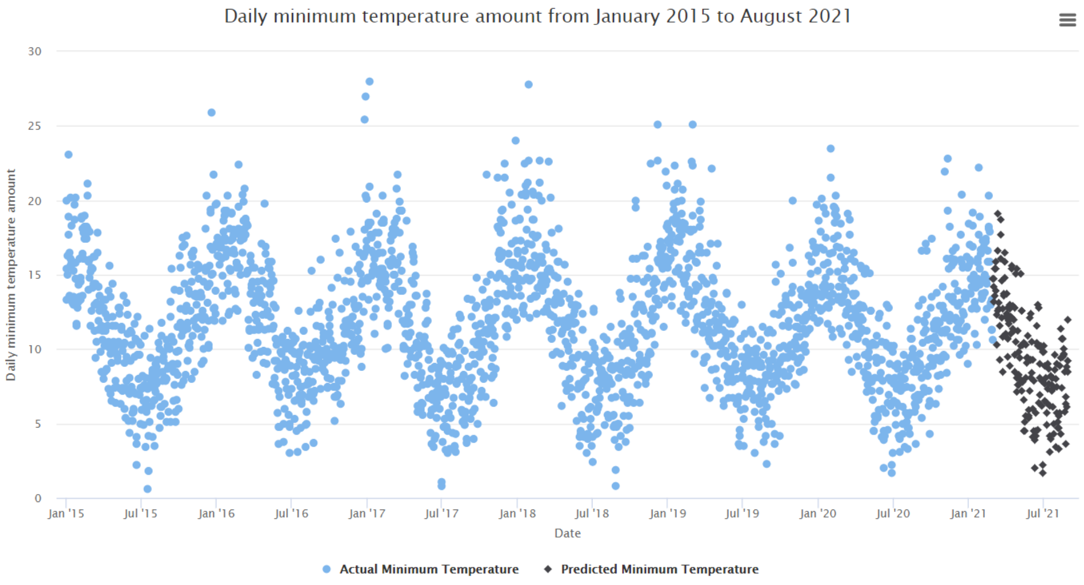
Actual data collected from Bureau of Meteorology for January 2015 to February 2021 is depicted in blue.



Distribution of data using box plot:

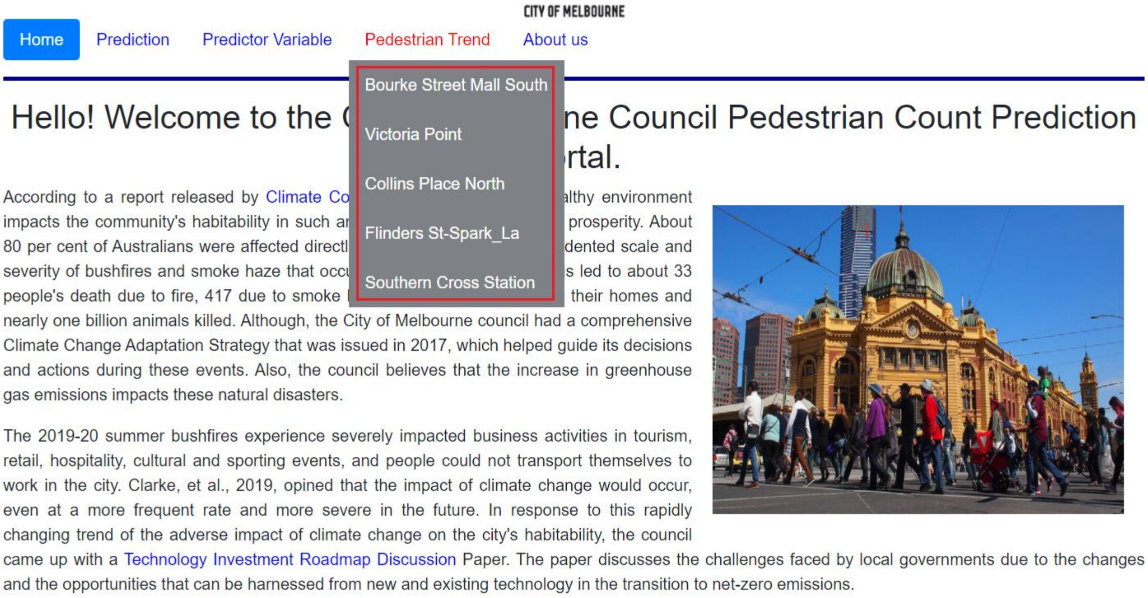


Prediction for Minimum Temperature is depicted is in black.



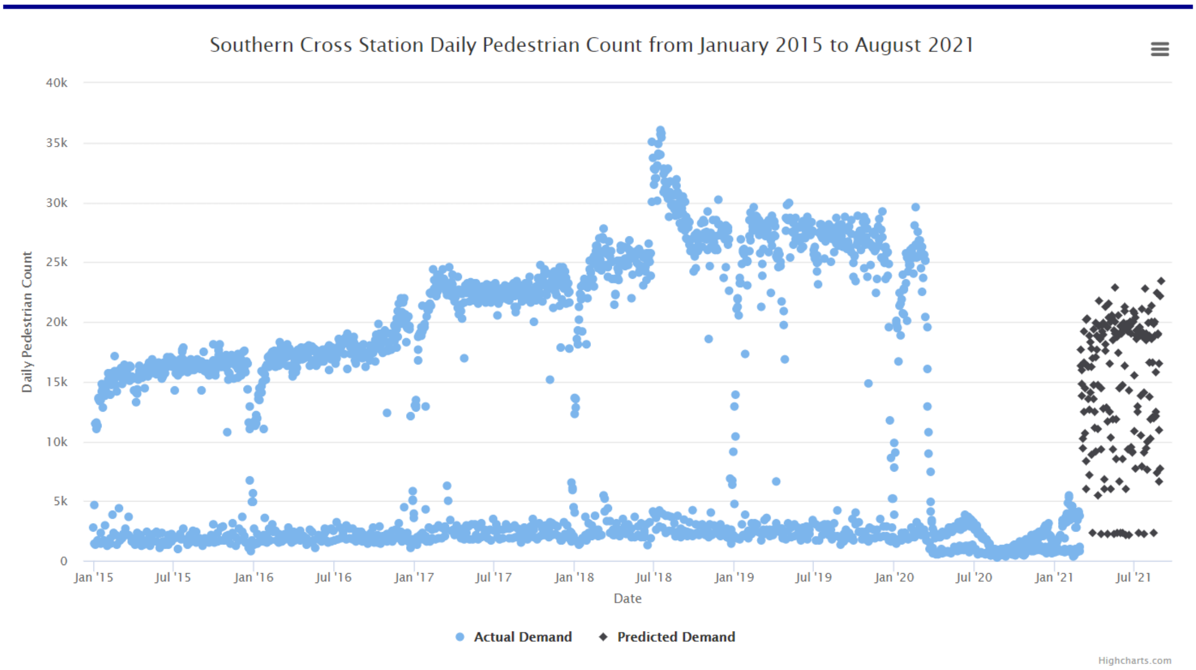
3.5. Prediction of Pedestrian Count at granular level

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



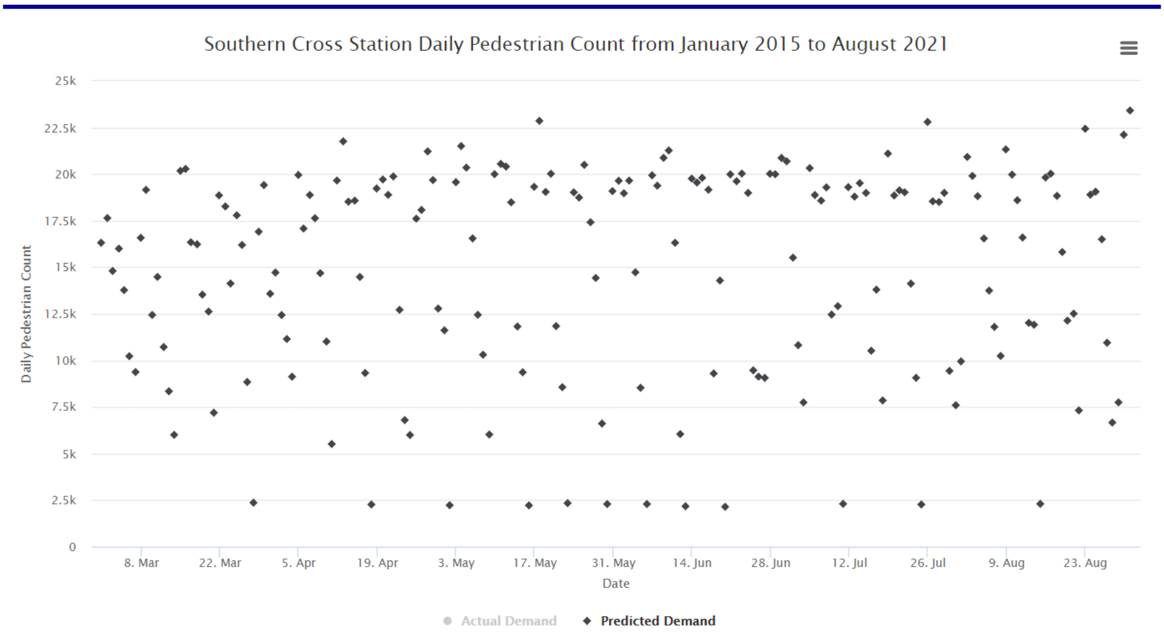
Pedestrian count prediction for Southern Cross station:

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.

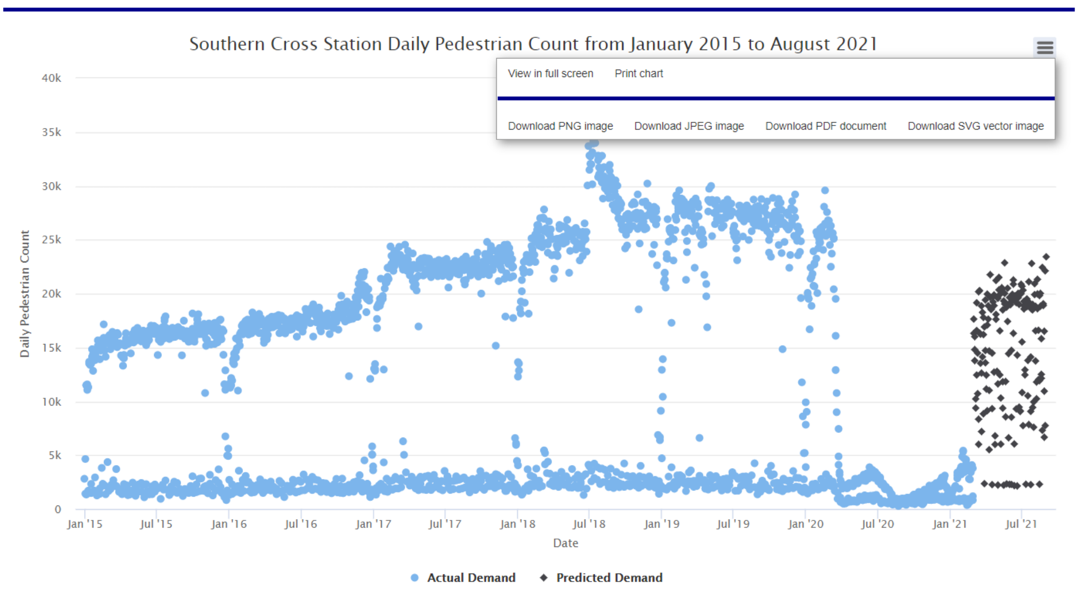


3.6. Features of Highcharts

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.



3.7. Video for user navigation

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

[User Manual]

4. Completed Deliverables

4.1 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, 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-19 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 desired 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 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 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 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) and Stacked LSTM were evaluated by Rohan Amatya. EDA (Exploratory Data Analysis) was also carried out to understand the data, perform imputations, aggregate information, and perform stationarity checks. This ticket and notebooks 1, 2 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 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, which would periodically refresh the dashboard. 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 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, 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.

4.2 Sub-Team B Deliverables

The objective goals of Sub-Team B were visualising the 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 formed Sub-team B and come up with the project proposal.

Iteration 1:

- Deji extracted and analysed the pedestrian count data from City of Melbourne, which was then reviewed by Jason and Oscar. Trello
- For weather features extraction, Oscar imported and cleaned rainfall amount; Deji did Solar Exposure; Jason did min and max temperature, respectively. Trello
- Descriptive statistics were produced to help decide how to predict independent weather features. Each member in sub-team b did exploration analysis for weather features like boxplots, histograms, pie charts and bar chart. Trello
- Forecasted five sensor locations with weather feature variables. Trello

Iteration 2:

- Visualised predictions for sensor count and Independent variables using Highcharts. Trello
- Make the pedestrian prediction webpage more interactive (e.g. make sliders). Trello

Firstly, sub-team B predicted weather features, including Rainfall, Min and Max temperature, solar exposure across city of Melbourne from March 2021 to August 2021. Plus, sub-team B made a prediction of pedestrian counts in the TOP 5 sensor locations, from March 2021 to August 2021 by these feature attributes. Secondly, exporting output to files, enabled sub-team B to use them for visualization. Finally, by using Python Flask these figures were able to be presented onto a website page.

Data Acquisition:
All the weather dataset can be found in BOM. Pedestrians counts based on Top 5 sensor location can be found here.

Data Manipulation and EDA:
Sub team B decided to use datetime index as attribute to clean the dataset. Code can be found here. Furthermore, sub team B also reported the findings into word document here. That is, the summary statistics of weather feature in terms of day of week, month, and year.

Machine learning:
After wrangling and tidying the 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.

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.

Deliverable Outcome (Sub Team B):

- Researched variables and used reliable datasets for the project.
- Scraped and cleaned independent variables.
- Trained and forecasted 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.
- Visualised prediction sensor counts using Highcharts.
- Applied different time series models, like Arima Walking forward and LSTM. The variation explanation of each feature increased by average 5% for each sensor location.
- Compared to Tableau, Highcharts visualisation is faster to load on the website page.

4.3 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 were 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
2. 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 and Trello
3. Performing Exploratory Data Analysis on the dataset
   - Neet Patel was responsible for this deliverable
   - This can be found in Bitbucket and Trello
4. Determining stationarity of the dataset
   - Neet Patel and Nikita Wadekar were responsible for this deliverable
   - This can be found in Bitbucket and Trello
5. 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 and Trello
6. 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 and Trello
7. 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 and Trello
8. Incorporate the dashboard in the webpage
   - Sivaram Krishnan was responsible for this deliverable
   - This can be found in Bitbucket and Trello

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 – August2021. 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 R^2 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).

5. 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 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 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 3 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.

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 0 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.

6. Open Issues

Pedestrians counts extract from City of Melbourne from Jan 2015 to Mar 2021. However, there are missing values, which lead to length of dataset range is not same. Suggestion: Wisely choose data range to fit this problem.

The historical data and the forecasted data for the pedestrian count needs to be integrated properly. Currently, all the forecasted values for the historical data are set to 0 which is why Highcharts display yellow data instances at the bottom.

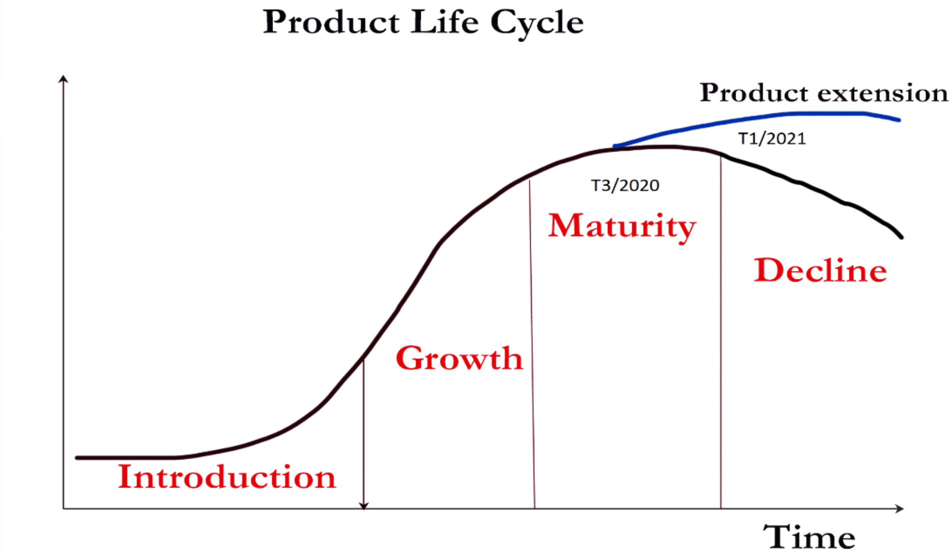
Due to Covid-19, there is a short memory trend in time series model, which lead to low value in variable explanation (R^2). LSTM performed better compared to KNN, RF, SVR.

7. Lessons Learned

In this trimester, we had two products integrated to form one product. But due to lack of communication with the senior members of last trimester we were not able to extract the product. The squad worked on the second product and enhanced it. For future squads, proper communication and coordination is necessary to transition the information. The senior team members should share the server resources and one of the junior team members should be given the responsibility to understand the server resources. This will avoid the loss of code.

Keeping an internal channel as a means of communication within the team, as not all discussions in teams should be showcased to non-team members irrespective of our objective to earn marks. For future squads, ensure every team member is given equal opportunity to contribute irrespective of their level of technical know-how or how well they could express or not express themselves as well as looking at submissions objectively devoid of any negative nuance or exerting undue ego on other members.

8. Product Development Life Cycle



For this trimester, the team started off the product development phase by conducting further research on the client's long-term goals to determined more requirements that could be implemented to extend the product earlier delivered more progress upon the achievements of T3/2021 rather than allowing the project decline or close.

We devised ways to deliver more features product to meet the client's need by conceptualising more tools that used and data repositories that can serve as a reliable source of intelligence for our client's decisions towards achieving its long-term goals, analysing the impact of the new features on the client's decision-making ability and analysing the impact of the overall goal.

The work methods we devised is to brainstormed ideas and collaborated with our client on how the project should progress. We drafted an initial understanding of the project's critical tasks, with priority given to identifying more relevant and reliable sources of data for advance features. Subsequently, the team was divided into sub-teams like the identified deliverables in ensuring each of the deliverables received the much-needed concentration and give each member the ability to contribute towards the attainment of the deliverable in an objective manner. Each sub team led by a senior member were responsible for a feature of the final product.

We looked at the numerous options for delivering the project's end product that would match the defined goals while considering the dataset availability and schedule constraints. Our focus was on adding value, preserving the quality of the final release, and managing the workload via agile technologies like a Trello board and a Bitbucket git repository, all while iterating on stakeholder’s feedback until each product was perfected on time.

The habits the team developed intuitively over the course of the Trimester was to ensure we meet with the client regularly and utilised feedbacks to ensure that every component of the project met his expectations, allowing the team to become more efficient and productive.

8.1 New Tasks

Following the successful determination of the Trimester's set of deliverables, the next squad meeting focused on discussing the lessons learned from the previous sprint. The squad brainstorms and members present their developed and well-detailed user stories during subsequent meeting, which is followed by a planning meeting where the squad brainstorms and members present their developed and well-detailed user stories. Senior members go over the user stories in further detail and break them down into new tasks, which are then placed to the Trello board's Product backlogs list and allocated to interested members.

8.2 Definition of Done

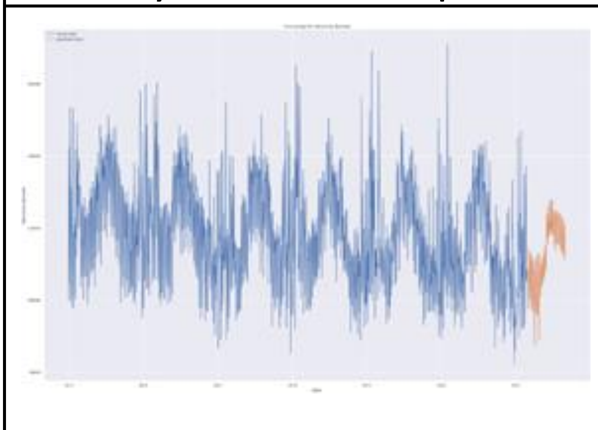
Completion is the last phase in the task lifecycle. After a thorough evaluation, the reviewee would either provide suggestions for improvements or return the card to the in-progress list. In other circumstances, the work is transferred from the Review/Testing list to the Completed list if the reviewee is pleased and has guaranteed that the Task satisfies a sufficient standard. It is assumed that after the task has been tagged as complete, the Bitbucket repository will be updated with the source code/dashboard used in the task, along with an appropriate commit note.

8.3 Task Review

Work reviews and testing are required steps in the process before the task can be finished. An allocated reviewee, often a senior member, attentively reviews the work for this assignment in this stage. The reviewer verifies that best practises have been exhibited, such as providing pertinent comments and using suitable indentations. The reviewer would also confirm that the code produces the required result.

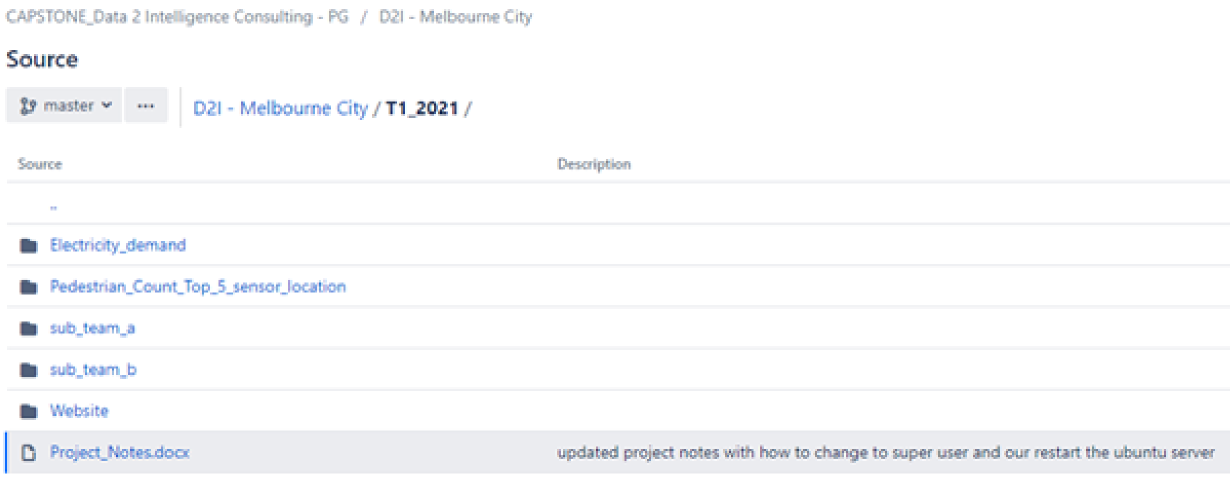
8.4 Testing

Product testing was completed by ensuring that forecasted graphs in Python were 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 planned to do. As Highcharts is interactive, 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.

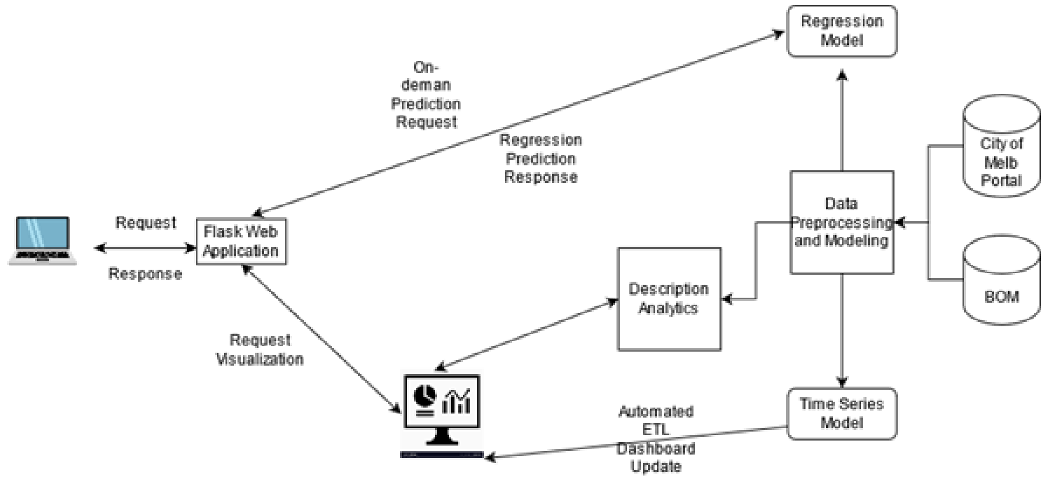


Within each folder, there were multiple folders such as 'Dataset', 'Python codes', 'Website', etc. The squad had decided to commit files to the relevant sub-team folder, to ensure that they Bitbucket is laid out neatly, instead of creating a messy and unorganised branch. Additionally, using an organised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

9. Product Architecture

9.1 UML Diagram



9.2 Tech Stack

The whole of the project can be divided into two folds:

1. The customer facing web application
2. 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.

10. 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

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

Pedestrian

- Browse CAPSTONE_Data 2 Intelligence Consulting - PG /D2IC-PG/repos/d2i---melbourne-city/browse/T1_2021/sub_team_a

11. 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 the tools used in this trimester are private. Read document carefully.

12. Appendices

Link to the team's Pitch video:

- https://deakin365.sharepoint.com/sites/Data2IntelligenceConsulting/Shared%20Documents/D2I%20(Melbourne%20City)/T1%202021/Handover%20Artefacts/D2I%20-Melbourne%20City-%20Pitch%20Video.mov

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/

Link to the live website

- http://13.250.31141.8080/