Monitoring Migration Health Risk in Relation to Climate Change

WIL Project - Group 11

DS Ninjas

Team:

Ho Yin Lam – s3889140

Natasha Sheree Galea – s3540515

Rajnish Kumar Jha – s3821892

Rakyan Satrya Adhikara – s3935019

Richard Joseph Doherty – s3863706

Wing Hang Chan – s3939713

GitHub: https://github.com/nsgalea/DSNinjas

Atlassian: https://wil-project11.atlassian.net/wiki/spaces/WP/overview

Table of Contents

Executive Summary

Every year hundreds of thousands of people move to Australia from various countries around the world. Migration from their home countries to Australia exposes those individuals to psychological and environmental stresses. A wide variety of websites currently exist where interested parties can find information about all aspects of climate data and quite a few make it possible to review migration data, but it is difficult to find any publicly available data linking both climate and migration data. This case study project focused on creating a prototype website to fill that information gap both for potential migrants to Australia and organisations (government and private). The project targets anyone who might be interested in understanding the trends, impacts and potential risks associated with migration patterns. Multiple machine learning models were developed, compared, and optimized to help identify climate zones and then link those climate zones to past, current, and potential migration trends. The initial Minimum Viable Product (MVP) was developed using temperature data and aimed to inform users about the similarity between various countries and Australian postal regions.

1. Introduction

Project Aim

The initial Project Aim statement was: *To create a single web page similar to map explorer* [1] *that provides insurance companies information about migration health risk, and how that risk will be impacted by climate change, taking into account the person's country of origin and what the mean temperature in that region.*

Figure 1 Map explorer of web site Climate Change in Australia

The project started from the initial Project Aim statement (above Figure 1) and moved forward from there with exploration of options for how best to identify available data and use the data available. Initially the aim was to target insurance agencies, however concerns were raised surrounding the legal and ethical concerns of charging more based on a country of origin. It was decided, instead, to focus on the health aspect and create a website with the aim of assisting migrants in mitigating the health risks associated with migration from one climate to another. The website/app would be titled "Monitoring Migration Health Risk in Relation to Climate Change." From that early exploration and analysis, we developed a roadmap for implementation of the target features and functionality via two phases.

Phase 1: Data Application

Phase one was focused on a global comparison of all countries to Australia at once, and providing this information via a PowerBI app. Countries were clustered based on their temperature and temperature variability, and methods setup for testing the accuracy of the model. Development also began on a PowerBI dashboard using randomly generated mock-data that would allow both the model and the dashboard to be developed simultaneously. The focus was on creating an example application that could demonstrate how the data would be useful to provide information surrounding the impact of immigration on individuals' health due to the climate differentiation.

Phase 2: Information Portal

Phase 2 was focused on creating an information portal that would focus on a specific location outside of Australia. As opposed to the cluster-based model, it would allow the user to select a specific location, and then they would be informed about which regions were the most similar to their home location. In the final version, the aim of the portal would be to suggest similar climates, or warmer/cooler temperatures based on the user's selections, as well as highlighting health risks, and potentially encourage more immigration to regional areas of Australia. It would also show climate predictions, and how the environment may change over the next century. For the initial version, a simple model was created in Jupyter Notebook using Euclidean distance to rate the similarity of each region.

2. Problem Statement

In July 2022, over a million people (1,081,610) arrived in Australia, and 968,490 departed [2]. Visitors often come from vastly different climates to Australia [2], which can have a significant effect on their health, due to the sudden change in climate. While the adverse effects of climate have been studied due to global climate change, few studies have focused on the effects of climate differences for migration. A European study found that on average 20% of people travelling to a developed country experienced illness [3]. However, only 10% of those who become ill visit a doctor, and less than 1% are hospitalised [3]. Our project aims to help improve awareness of climate related health risks and provide information to individuals and businesses considering migration to Australia, as well as help with skill shortages and population declines in regional areas by suggesting countries with similar climates to those regions.

Taking the number of people who arrived in Australia in July 2022, on average 200,000 would have been unwell, 20,000 would have visited a doctor and 2,000 would have required hospitalisation [2] [3]. This is likely due to temperature variability between their origin country and destination. Studies have found that temperature variability and whether a temperature is normal for a time of year, has a greater effect on health outcomes than the effects of the temperature itself [4]. For example, higher temperatures increase health risk in spring and early summer, soon after the transition from winter, more than at the end of summer. Similarly, there is an increase in health risks from extreme cold temperatures in late autumn, then in [4]. By providing more education to those considering temporary migration or immigration to Australia, we hope to help reduce these effects. Furthermore, businesses that area considering opening an Australian branch, may consider destinations and choose a location that is similar to their origin country. This would help reduce the associated health costs to the business, from employees getting sick after travel.

Increased temperatures can cause acute increases in mortality over short periods, however cold accounted for 33 per 100,000 deaths in Australia in 2014, compared to heat which accounted for 2 per 100,000 [5]. While most studies have focused on the acute health effects of climate change, there have been less empirical studies into the long-term health consequences. Sunlight exposure, altitude, humidity, extreme weather, and the individual's access to methods to adapt to the climate all have an influence on the health of an individual [6]. Due to these factors, it is important for various organisations, and the individuals themselves to understand the health risks associated with travel. Furthermore, access to more information about the risks can assist travellers with considering the risks associated with the potential destination.

Ethical Considerations

It is important as a part of this project that the ethical risks be considered. There are, two areas of concern. The first is the concern for the safety and health of the individual visiting Australia. It is important that the project considers at risk groups, such as groups from disadvantaged countries, and does not allow them to be highlighted, or taken advantage of. It is important to ensure that they have access to appropriate health care, and in future versions of the project the goal would be to help identify high risk countries that could be given additional healthcare and support to assist in transitioning into Australia. Furthermore, for those who do not speak English as a first language, or may have other cultural needs, such as religious access, it would be ideal to recognize these needs and identify availability of access for those needs, especially when suggesting regional areas for migration, where not all cultural needs may be easily available [7].

The second ethical concern is the effect on the communities themselves. A large migration to some communities may influence the environment of that community in an unwanted way. If communities are close-knit and they may struggle with sudden change. In these cases, it may be important to ensure that the communities are educated, and that migration into those areas occurs at a rate which does not put stress on community resources and culture.

The initial MVP does not include information related to community issues when making recommendations, however it is a future goal of the project.

Benefits

Migration is an important part of Australian economy. Migration helps offset ageing population, increases GDP growth per capita and increases labor productivity [8]. Furthermore, education for international students was Australia's largest non-resource export, earning $18.8 billion in 2014/15 [8]. However, travel to Australia brings increased health risks due to the migration. Currently students pay $478 per year in health cover when visiting Australia [9], and if the student must visit the doctor frequently due to illness, that can become an increased risk to insurance providers, reducing profits. Furthermore, for immigrants who are granted Medicare, the associated health burden is then taken by the government. For this reason, early intervention, and awareness, may assist in reducing those costs and increase the longevity of immigrants.

3. Methodology

Data Exploration

Before the implementation of the clustering method, the data was cleaned and converted into a usable format. Such as removing missing values, checking for distinct values as well as non-relevant features to avoid biases from the model and provide better results (see Figure 2 for the solution pipeline).

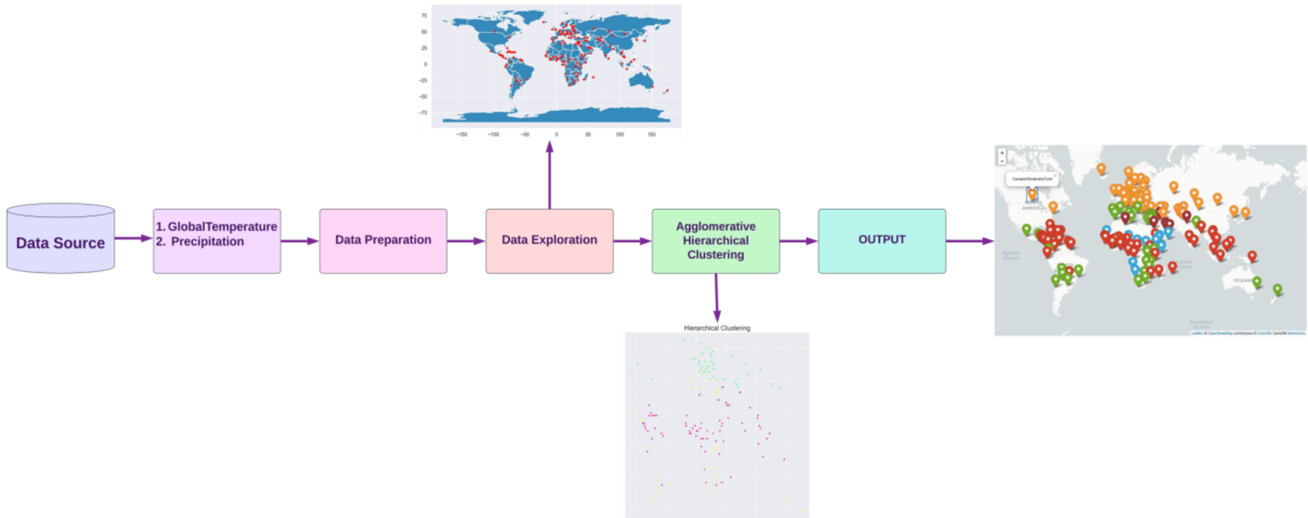


Figure 2 Data Pipeline

Data was located using open-source platforms, to solve the business problem of health issues during migration to Australia. Climate data gathered with a focus on factors related to temperature, humidity, rainfall, altitude, UV, etc. with the aim being to use that data in determining which countries/geographic areas are most similar in terms of climate to postcodes of Australia.

A high level of research and literature review was also completed by all team members, to build an understanding of the needs and possible solutions. After reviewing the available data, the decision was made to limit focus to temperature data. This provides only a partial view of climate but would be appropriate to provide informative insights via the MVP with the hope that it could be expanded to include other climate factors in future phases.

Data Preparation

Several Jupyter Notebook files were created early to assist the team in being able to work on the project without running into bottlenecks. Initially, for example, Risk data was generated randomly and placed into model-data/Risk{CountryName}.csv files, ranked 1 for low risk and 5 for high risk. These files were able to be used from early on, to generate the PowerBI, while development on the final model was completed. In the late stages of the project, these files were updated with accurate data.

Most files were uploaded directly to GitHub, however the map files were quite large. The GitHub README was updated with information on how to download the files manually. Kaggle was used for global temperature data, however it is intended to use a more reliable source in the future.

Data used for the MVP

Filename	Description	Fields
data/acorn_sat_v2.2.0_daily_tmin.tar.gz data/acorn_sat_v2.2.0_daily_tmax.tar.gz http://www.bom.gov.au/metadata/catalogue/19115/ANZCW0503900725	Climate Change Data (Australia) – Daily minimum and maximum temperature for stations around Australia	Date Temperature, air – maximum Temperature, air – minimum StationId (filename)
GlobalLandTemperaturesByCity.csv https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data?resource=download	Climate Change: Earth Surface Temperature Data – Kaggle dataset sourced from http://berkeleyearth.org/	Date, AverageTemperature, AverageTemperatureUncertainty, City, Country, Latitude, Longitude
Station locations http://www.bom.gov.au/climate/how/newproducts/images/IDCJDC04_stations.txt	Station Locations, used to gather the Lat and Lon of the stations, so they could be mapped to postcodes	Bureau of Meteorology Station Number, Station Name, Latitude, Longitude
https://ggis.un-igrac.org/layers/igrac:other_climate_2007_koppen_geiger	Shape file with climate zones (Koppen Geiger), shape file used to render world map, climate zones and used to validate climate models	Climate, Geometry,
data/POA_2021_AUST_GDA94_SHP/POA_2021_AUST_GDA94.shp https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026/access-and-downloads/digital-boundary-files	Postal Area Shape Files (Australia) Shape files used to render Australian map with postcodes	Postcode (POA_CODE21), Geometry
data/privatehealth-01-sep-2022/Hospital Open 01-Sep-2022.xml https://data.gov.au/data/dataset/private-health-insurance/resource/92963b8b-6fb5-43d1-9068-7a363ab17a4f	Insurance Data (data.gov.au) not used in MVP, but intention was initially to use it to attempt to put a value on the health effects.	

Table 1 Data source table

Australian Climate Data (AustralianClimateData.ipynb)

The Australian Climate Data, which had been sourced from the Bureau of Meteorology, was split into one file per station, further split into Minimum Temperature files and Maximum Temperature files. To make the data easier to work with, and to allow it to be rendered on the Australian Postcode Shape files, it was converted from separate files into a single data table. These data tables were also combined with another file that had the latitude and longitude of each station.

This was done by first finding the 3 geographically nearest stations to each postcode. Then the min and max temperature files for each of the 3 nearest stations per postcode were loaded, and their mean minimum and mean maximum temperature were taken. However, since this data also had to be used with the global temperature data, the minimum and maximum temperature data was then converted to AverageTemperature, AverageDailyUncertainty (average variance for each day during a season) and AverageUncertainty (variance across the whole season) for each season. This would allow the post code data to be used with the global data.

The final output included the following columns, which was saved as "model-data/AustralianPostcodeTemperatures.csv"

```
['POA_CODE21',                      'AverageDailyUncertainty_Autumn',
'AverageDailyUncertainty_Spring',   'AverageDailyUncertainty_Summer',
'AverageDailyUncertainty_Winter',        'AverageTemperature_Autumn',
'AverageTemperature_Spring',             'AverageTemperature_Summer',
'AverageTemperature_Winter',             'AverageUncertainty_Autumn',
'AverageUncertainty_Spring',             'AverageUncertainty_Summer',
'AverageUncertainty_Winter', 'Max_Autumn', 'Max_Spring', 'Max_Summer',
'Max_Winter', 'Min_Autumn', 'Min_Spring', 'Min_Summer', 'Min_Winter']
```

Global Climate Data (GlobalTemperatureData.ipynb)

Global climate data was cleaned up, and split into seasons based on the hemisphere. The global data was by date, similar to the Australian Climate data, so the data was converted from daily data to seasonal data for each City. The final output (model-data/GlobalTemperaturesWide.csv), was similar to the Australian Climate Data, so that it could be used to generate a cluster model and predict Australian Postcode clusters.

```
['City', 'Country', 'Lat', 'Lon',, 'AverageDailyUncertainty_Autumn',
'AverageDailyUncertainty_Spring',   'AverageDailyUncertainty_Summer',
'AverageDailyUncertainty_Winter',        'AverageTemperature_Autumn',
'AverageTemperature_Spring',             'AverageTemperature_Summer',
'AverageTemperature_Winter',             'AverageUncertainty_Autumn',
'AverageUncertainty_Spring',             'AverageUncertainty_Summer',
'AverageUncertainty_Winter', 'Max_Autumn', 'Max_Spring', 'Max_Summer',
'Max_Winter', 'Min_Autumn', 'Min_Spring', 'Min_Summer', 'Min_Winter']
```

Correlation Matrix

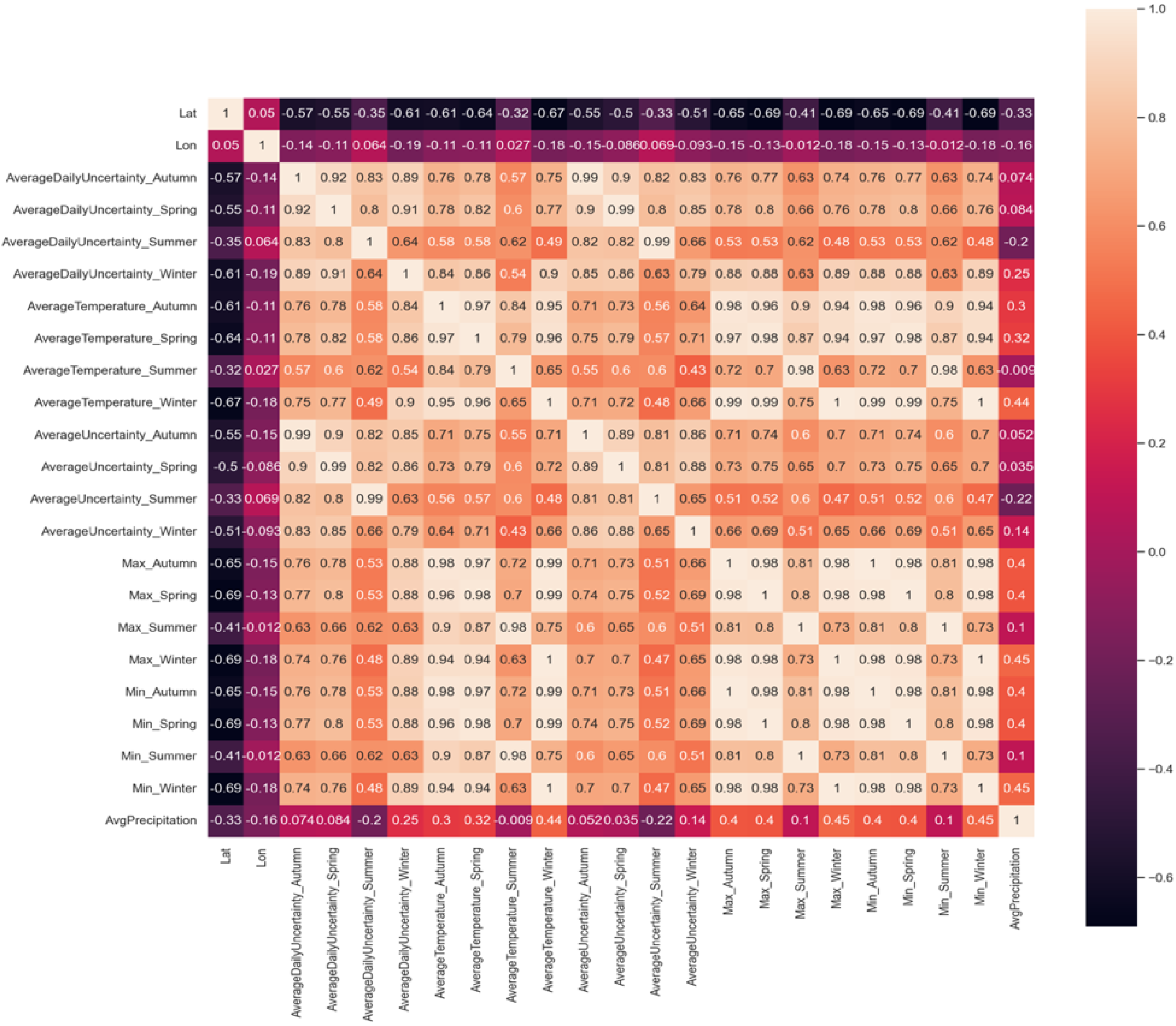


Figure 3 Correlation Plot for Temperature Data

The correlation plot was used to understand which variables are related to each other and the strength of this relationship. A correlation plot typically contains several numerical variables, with each variable represented by a column. The rows represent the relationship between each pair of variables that can help interpret the result statistically and easy to find the relationships between two variables to define a positive relationship and a negative relationship. Henceforth I have taken this process to analyze the features which can help to avoid any biases when it comes to the model interpretation.

Deployment of clustering technique

Clustering is one of the most well-known techniques when it comes to calculating the distance between two data points and finding more similarity-based ones, such that hidden patterns in the data can be identified and provide succinctly insights. Several clustering methods were tested, including Agglomerative Clustering method which was tested with precipitation data. To create an accurate model, first required data analysis of detectable dimensions, associated with metrics or other spatial embeddings [10]. This allows the model to be visualized. In

Figure 4, One example of visible dimensions (Longitude and Latitude) was used to analyze the potential clusters.

By systematically matching variables, clusters, or variables and clusters, Agglomerative cluster analysis creates a distinct set of nested categories or clusters. All clusters and unflustered variables are tested in all possible pairs at each stage, starting with the correlation matrix. The pair that produces the highest average intercorrelation inside the trial cluster is chosen as the new cluster. This technique progresses sequentially from tighter, less inclusive clusters through larger, more inclusive clusters and is continued until all variables are clustered in a single group.

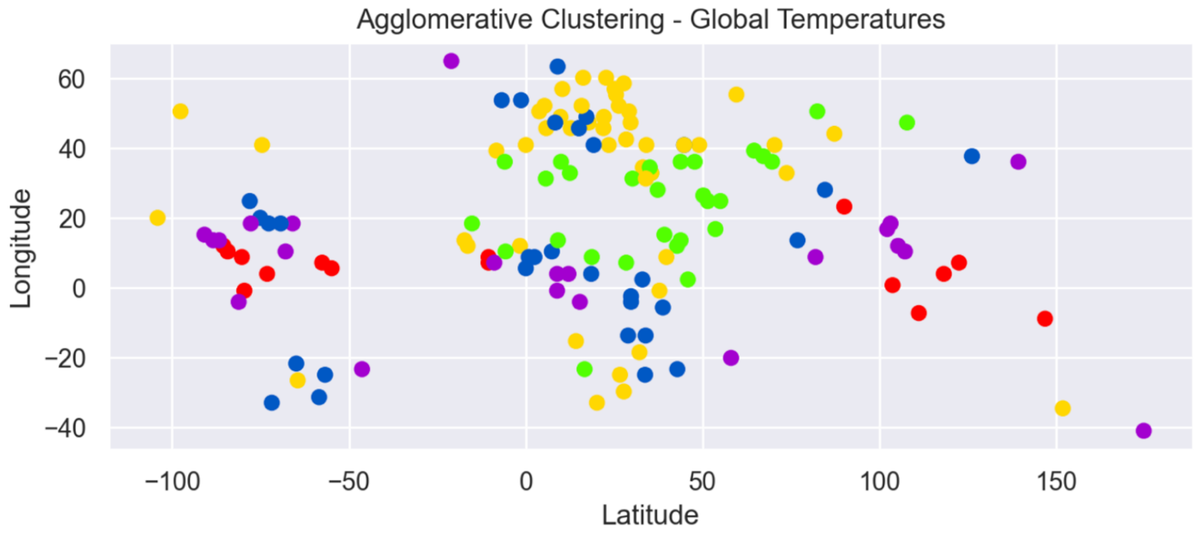


Figure 4 Visualising clusters using Longitude and Latitude

The data are grouped into groups in a tree structure in an agglomerative hierarchical clustering method. Every data point is first treated as a separate cluster in a hierarchical clustering process. The following steps are then repeatedly carried out by it: Find the two clusters that are possibly the nearest to one another and merge the two most comparable clusters. These procedures must be repeated until all the clusters are combined. The goal of hierarchical clustering is to create a hierarchy of nested clusters. An algorithm called hierarchical clustering, commonly referred to as hierarchical cluster analysis, divides objects into clusters based on how similar they are; thus, we have calculated the data points by using Euclidean distance. And to group data, we need a way to measure the elements and their distances relative to each other to decide which elements belong to a group.

Method of Euclidean distance

$$d(p, q) \quad = \sum_{i=1}^{n} (q_i - p_i)2$$

Equation 1 Euclidean distance equation

P, q = two points in Euclidean n-space

q_i , p_i = Euclidean vector, starting from the origin of the space (initial point)

n = n-space

As mentioned above, each cluster has been assigned according to its cluster id with country-wise and temperature so that the closest data point would be able to compute and allocate accordingly.

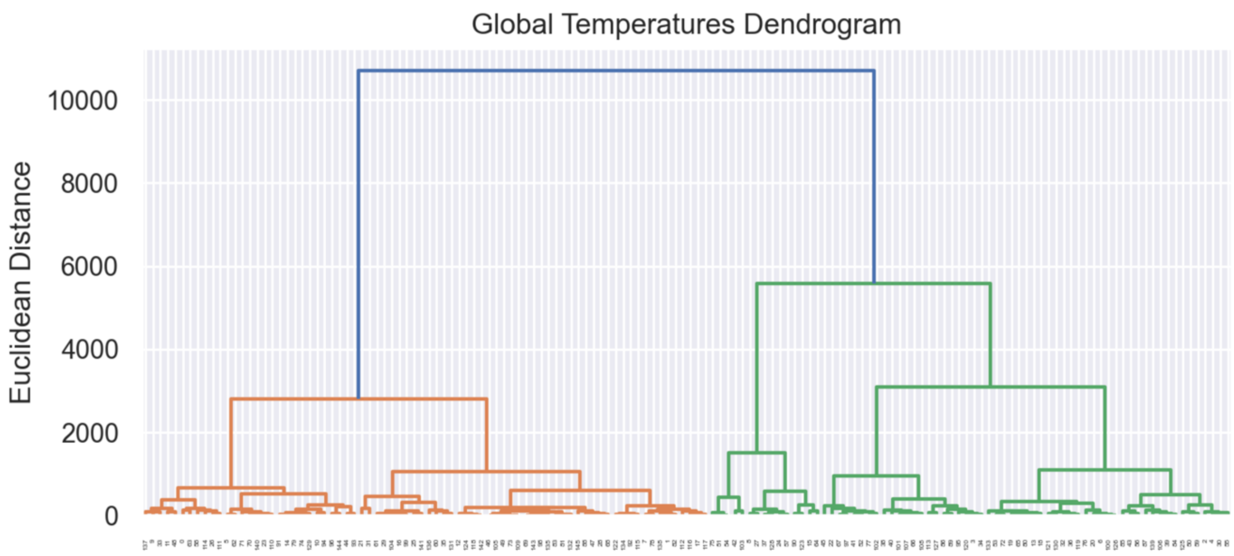


Figure 5 Global Temperatures Dendrogram

In the above (Figure 5) The portion in the dendrogram in which the rectangle has the maximum height, we can choose as it represents the maximum Euclidean distance between an optimal number of clusters. Therefore, we can cut the upper heightened rectangles and can find the optimal number of clusters which is 4 as observed in Global temperature data. Due to the complexity of the data, one of the drawbacks of a dendrogram for agglomerative hierarchical clustering is to find the correct clusters suitable for the data [11]. Hence, we have chosen 5 clusters to represent the temperature conditions e.g. – Very hot, warm, pleasant, cold, and very cold by measuring their distance through the closest data points.

Figure 6 Categorisation of clusters

As we have performed "Agglomerative Hierarchical Clustering" which is a type of unsupervised ml technique for the modelling, based on the global temperature dataset with spring, autumn, winter, precipitation, and summer in particular countries, where each distance score defines between every pair of points. The distance metric is highly customizable capturing any notion of dissimilarity such as Euclidean distance [12]. Likewise, from the above global temp data, each flag represents a climate temperature according to its colour and label. Based on weather conditions which is (one of the reasons for migrants), organizations can predict the probability of the migrants to make insurance business plans.

Model Evaluation

To compare the models, it was necessary to develop a method of comparing the models. During the research phase, the Koppen-Geiger climate model was used as a means of evaluating the model. The Koppen-Geiger model is based on “threshold values and seasonality of monthly air temperature and precipitation” and is split into five main classes and 30 sub-types [13].

To evaluate the model, the clusters were compared to Koppen-Geiger per major climate zone. It was expected that if the model was accurate, the climate zones would group in the same way as known clusters of the Koppen-Geiger classification system (Arid, Tropical, Temperate, Polar, Cold). However, since the initial MVP only used Temperature and not precipitation, it was not expected to accurately match Koppen-Geiger at this point. For example, it was found that the model did poorly in differentiating Arid and Tropical zones, as the major difference between these regions is related to rainfall.

Using this method of evaluation, it was possible to get an Accuracy and Gini-Purity measure for each zone. Accuracy favoured the most common cluster in a region, for example if the Arid region had 85% of cluster 1, it would be given an accuracy score of “85%”. Gini on the other hand, would give a 0 if the accuracy was 100%, however Gini was also affected by the distribution of all the elements, not just the most common value. Overall, they were very similar in how they rated the models, but it was decided to use Accuracy as the final metric for choosing a model.

In the end four methods were tested, *Birch, K-Means, Bisecting K-Means and Agglomerative Clustering*, using 5, 6 and 7 clusters. Birch with 5 clusters was found to be the best model using this method of evaluation, and only temperature data.

Accuracy	5 Clusters	6 Clusters	7 Clusters
K-Means	52%	49%	46%
Birch	63%	60%	52%
Bisecting K-Means	57%	55%	46%
Agglomerative	51%	47%	46%

Table 2 Comparison of average model cluster accuracy

Model Development

A Birch clustering model was used to cluster our Global Temperature Data with 5 Clusters. The five clusters were based on research into climate zones [14] and known regions were used to validate data but we also noted that it may be worth having more clusters in the future based on further optimization and testing of the model, possibly up to 30 climate groups based on the full climate set of Koppen-Geiger climate groups.

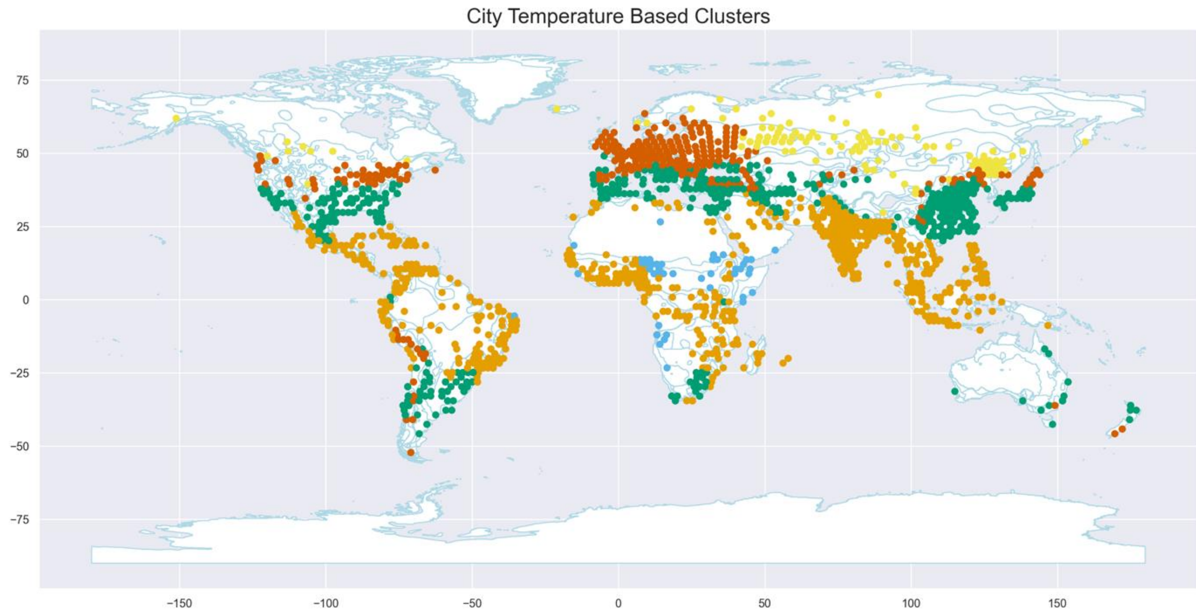


Figure 7 Birch model with 5 clusters, rendered on map

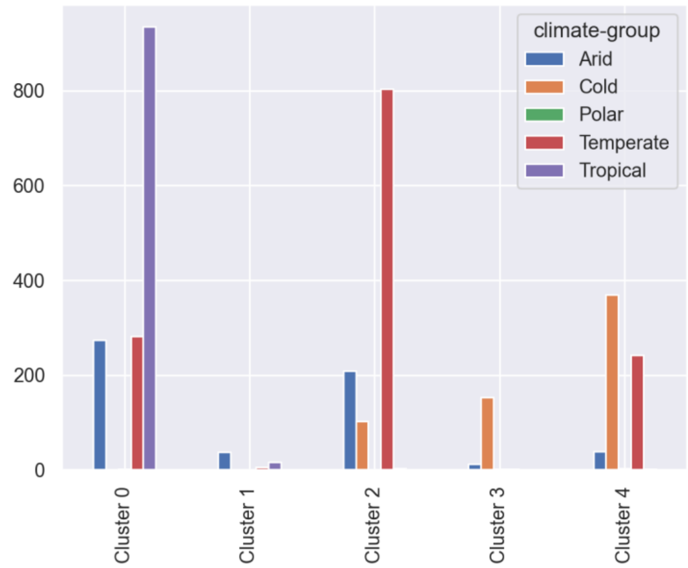


Figure 8 Distribution of Clusters using Birch with 5 Clusters

As can be seen in Figure 8, whole the Tropical and Temperate zones were able to be clearly identified, the Polar and Cold zones ended up mostly being put together in the same cluster. However, the reason for this is likely due to a lack of data samples. In addition, the Arid zone which is also known as the “Dry” climate zone, was unable to be clearly separated. The Arid zone ended up split relatively evenly between Cluster 0 (Tropical) and Cluster 2 (Temperate). Thus Arid (Cluster 1) was the least accurately modelled area, followed closely by Polar.

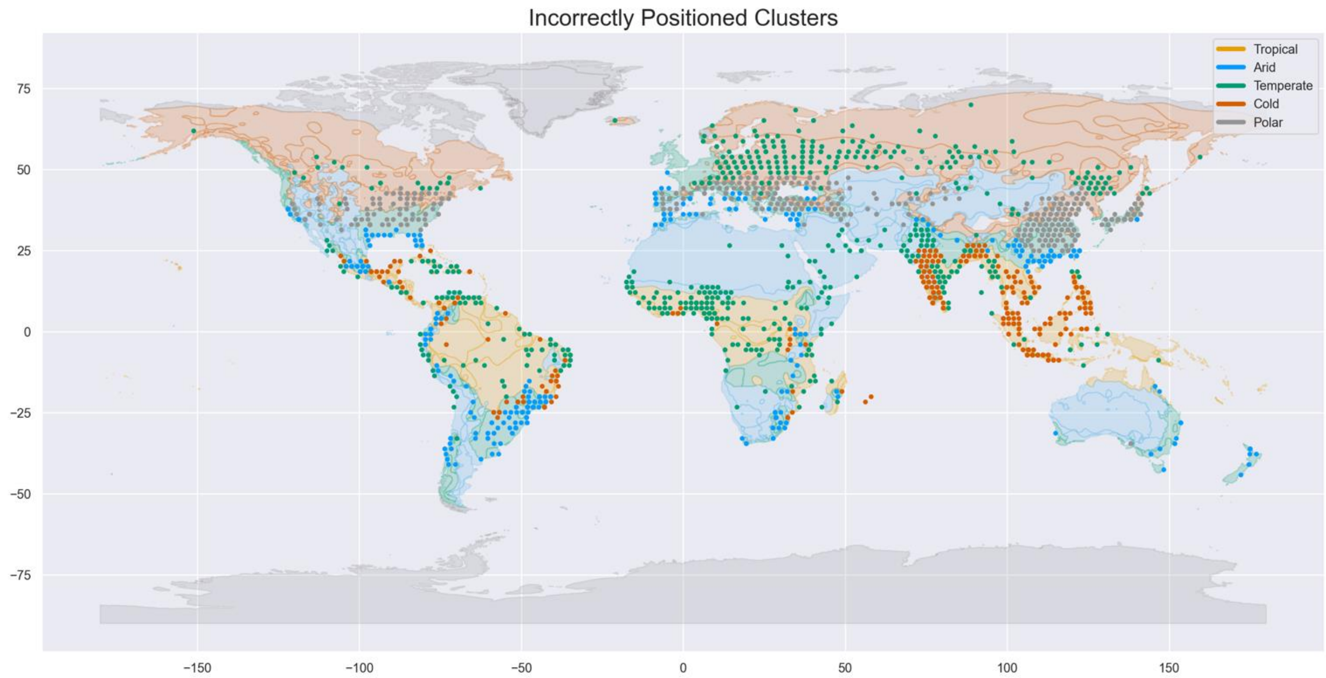


Figure 9 Incorrectly Positioned Clusters

In Figure 9, the out of position points were rendered over the map to emphasise the incorrect locations. In this case each region was coloured by the Koppen-Geiger region, then the points overlaid that did not match with the most common value for that region.

Visualizing the clusters

After implementation, the resulting clusters seemed to make sense visually when rendered on the map. That was a good sign, but further analysis was required to compare clusters to known climate regions. For each shape in the shape file, a determination of which of the cities it contained was made, along with determination of their most frequent cluster.
Purity and distribution of clusters within the zone were also used to validate the clusters.

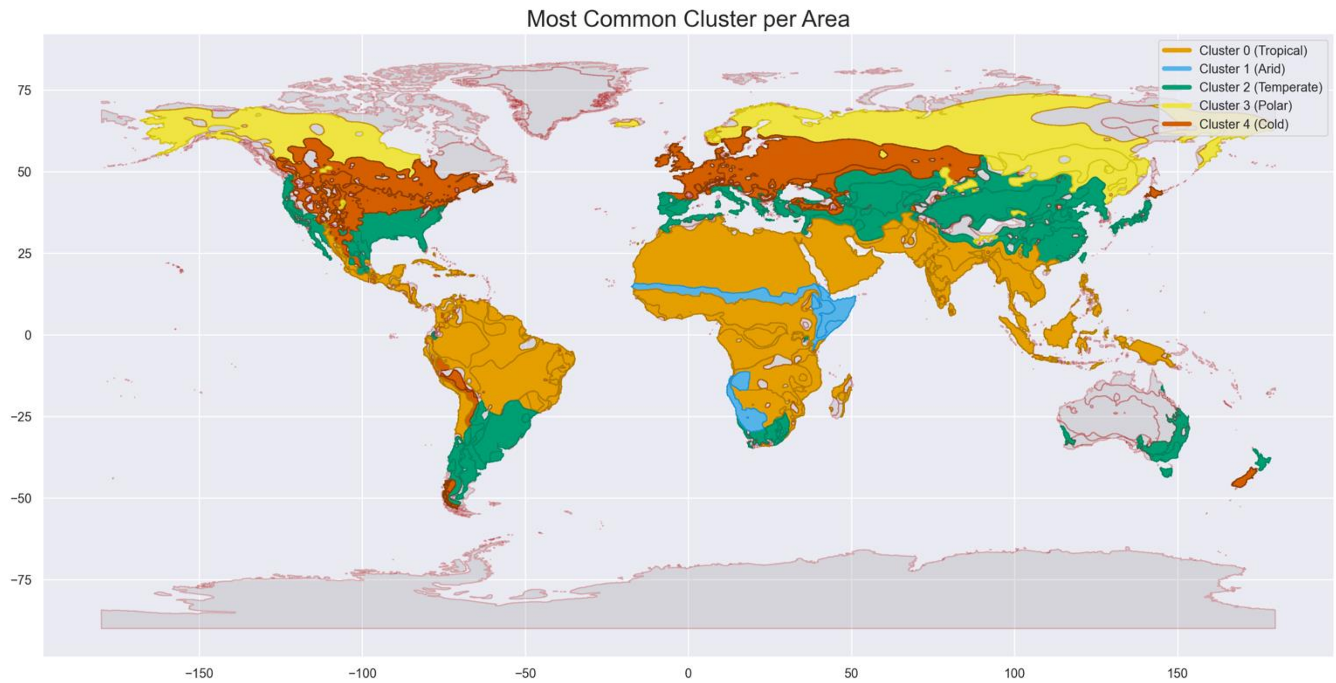


Figure 10 Coloured by Most Common Cluster

Classification of Australian postcodes based on clusters and recommend areas

Predication of the clusters based on Australian Postcode Climate Data could now take place using Global Climate Regions to predict Australian Postal Regions. Then, we were able to make recommendations based on climate. Stations were mapped to postcodes using 3 nearest stations. Then, seasons were extracted and given the same fields as above. Uncertainty didn't exist in the dataset and had to be calculated.

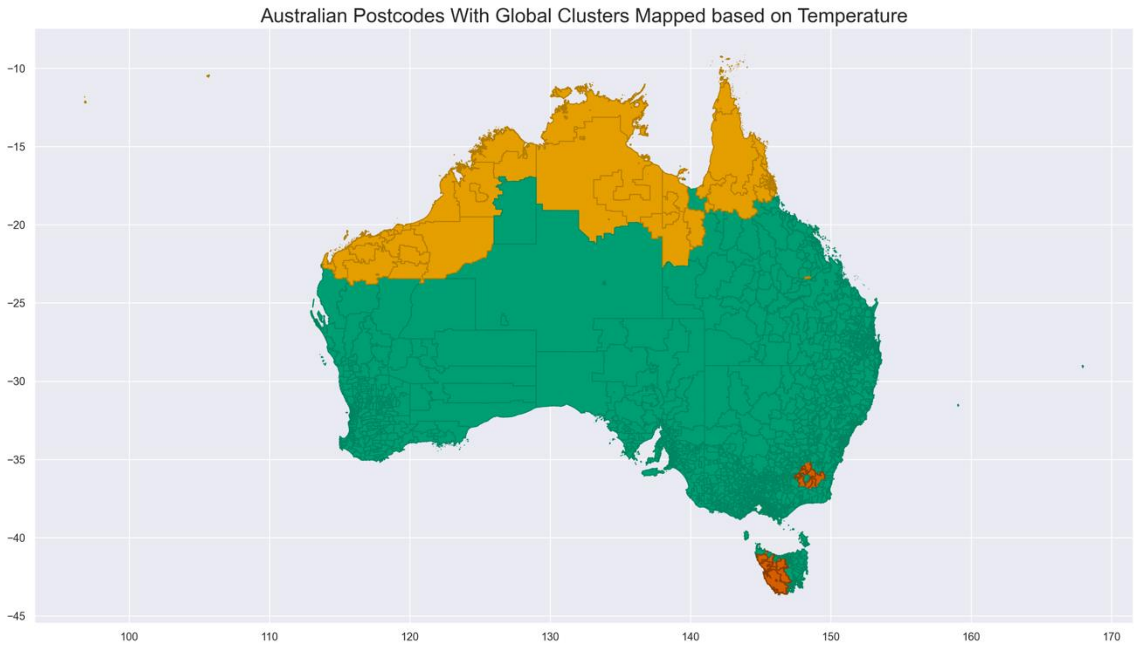


Figure 11 Global Clusters used to Predict Australian Postcode Cluster

Recommended Regions (based on similarity to home city/location)

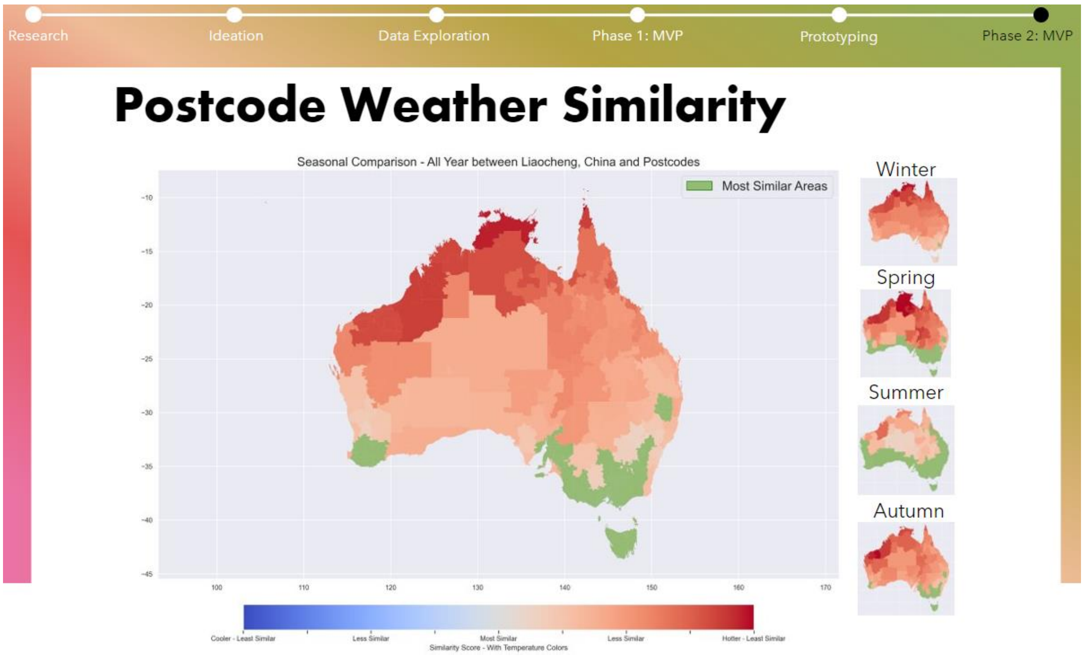


Figure 12 Regional similarity between postcodes and Liaocheng, China

The second phase involved developing a model that predicted the most similar and least similar regions based on a selected location. As can be seen in Figure 11, the Euclidean distance of all of Australia was calculated and then the most similar regions highlighted in green. From there, as regions were further (Euclidean distance) from the most similar region, they were mapped to a gradient of red (warmer than the home location) and blue (cooler than the home location).

Some regions such as India, had a relatively equal distance across all postcodes, and thus large amounts of Australia were highlighted (see Figure 13), and this was using the distance relative only to the selected city, not all cities. This may mislead the user into thinking their country's weather is closer to Australia's climate than it really is. For these reasons, this was only a very early phase model, which still required a lot of work to clearly portrait the information. Future development of the model would involve improving the clarity of the model, and adding additional information to help the user understand the risks (such as information about which temperatures are more likely to cause health concerns).

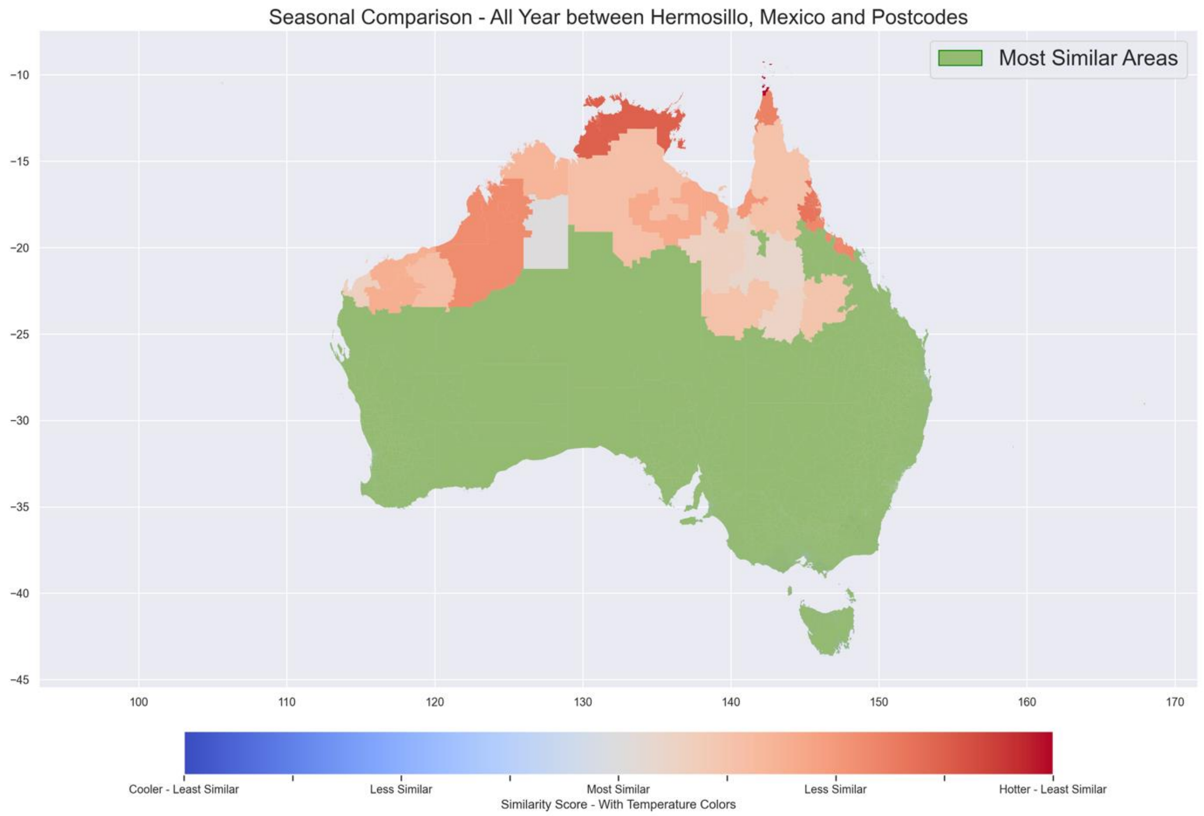


Figure 13 Hermosillo, Mexico had a large number of areas with similar Euclidean Distance

4. Findings

In this section, we will discuss the findings and in doing so, best to visualize the result after modelling the data considering not everyone has the right level of knowledge to translate the data by themselves. We used **PowerBI Desktop** to visualize the data considering it's user-friendly, it's very cost-effective compared to other Data Visualization tools, and despite all the convenience, it's a powerful tool to use. After we visualize the data, we gathered some interesting points that can be learned from.

Average Temperature Per Season Per Country

For this part, we chose six different countries, including Australia, based on most common immigrants' countries of origin and countries that have extreme climates. Those countries that have met the criteria are *China, India, New Zealand, United Kingdom, United States*. We analyse the temperature of all six countries and rank them based on the average value per season.

Figure 14 Mapping of Average Temperature Per Country

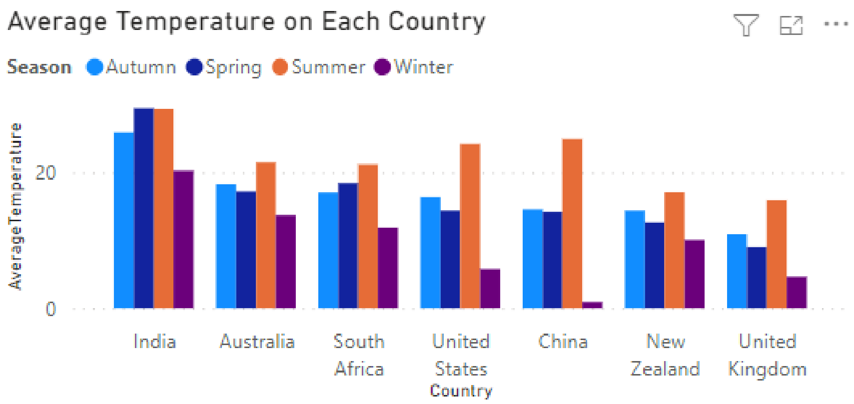


Figure 15 Bar Graph for Average Temperature Per Season Per Country

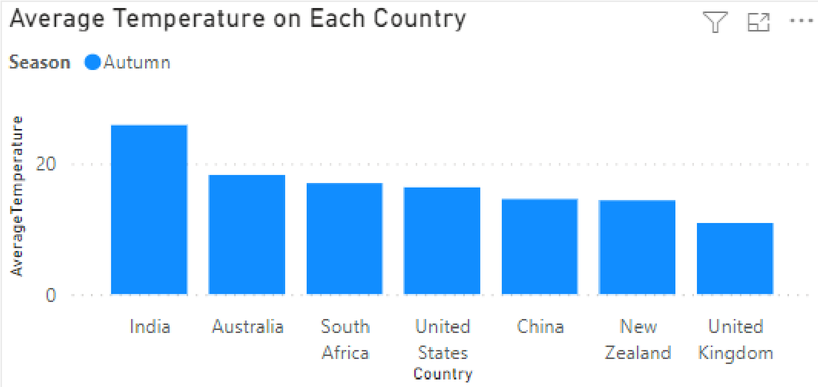


Figure 16 Bar Graph for Average Temperature Per Country on Autumn Season

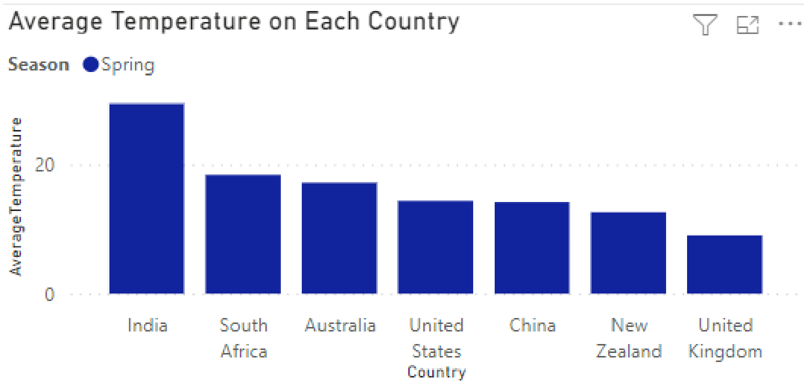


Figure 17 Bar Graph for Average Temperature Per Country on Spring Season

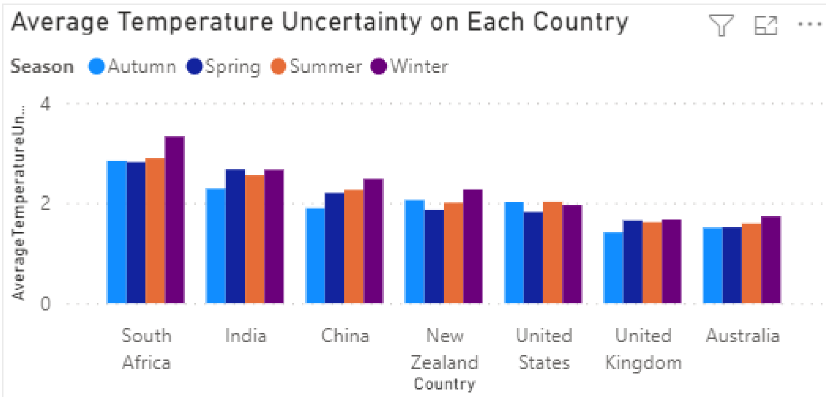


Figure 18 Bar Graph for Average Temperature Uncertainty Per Season Per Country

Based on the graphs above, it is shown that:

1. *India* has the highest average temperature compared to other countries across all seasons.
2. *The United Kingdom* has the lowest average temperature in **autumn**, **spring,** and **summer** seasons.
3. *China* has the lowest average temperature during the **winter** season.
4. *Australia* and *South Africa* are similar from one to another in every season.
5. Average temperature in **autumn** and **spring** is almost the exact same in rank per country.
6. Average temperature uncertainty is similar from one season to another in rank per country.

Note that temperature uncertainty shows how each country's temperature fluctuates. Higher temperature uncertainty means the selected country's temperature fluctuates more, the result of a less stable climate.

Timeline Graph for Temperature in Australia

We also decided to create a timeline graph for temperature in Australia. However, in this case, we can only do one region at a time, and we decide to go to *Kalumburu*, *Western Australia*.

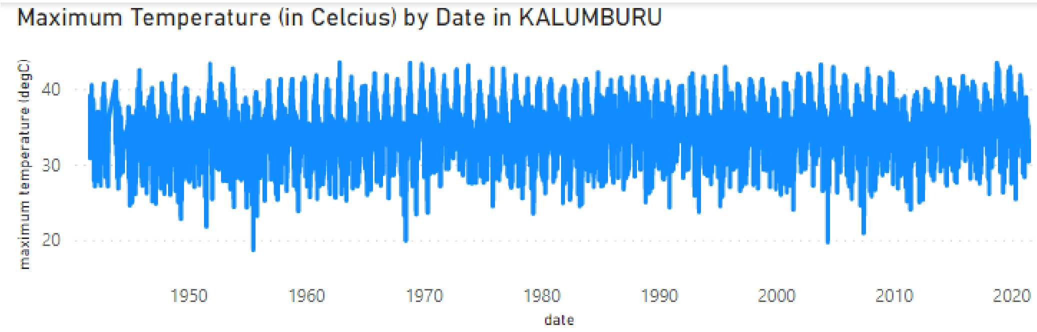


Figure 19 Line Graph for Maximum Temperature (in Celsius) by Date in Kalumburu

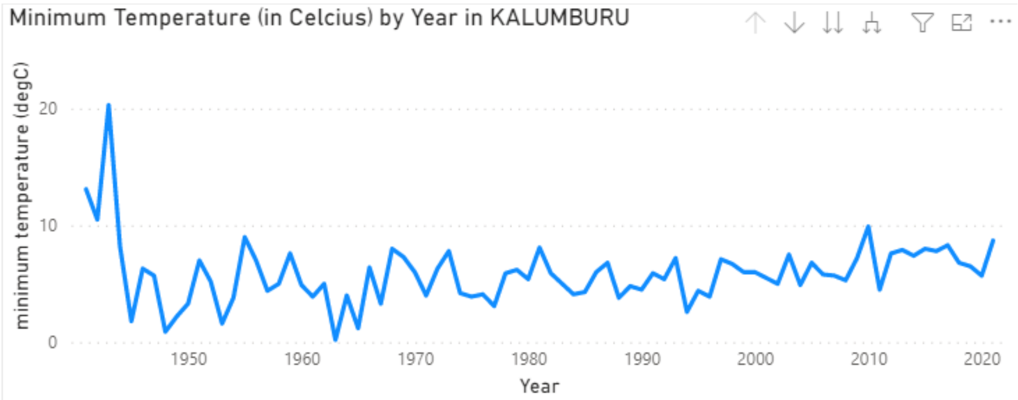


Figure 20 Line Graph for Minimum Temperature (in Celsius) by Year in Kalumburu

Based on the graphs above, we can see that in *Kalumburu, Western Australia*, the highest temperature is above 45°C in the 1960s and the lowest temperature is around 0°C in the 1940s.

Note that for the timeline graph, the data for maximum temperature have less empty value compared to the data for minimum temperature, which is why the line graph for minimum temperature is yearly, not by date. Thus, makes both graphs not in sync from one to another.

Migration Risk Mapping

For this part, we chose four different countries based on the locations of said countries, which are *Brazil* and *New Zealand*, and common immigrants' country origins, which are *United States* and *India.* We analyse the migration risk by mapping them per state on each country.

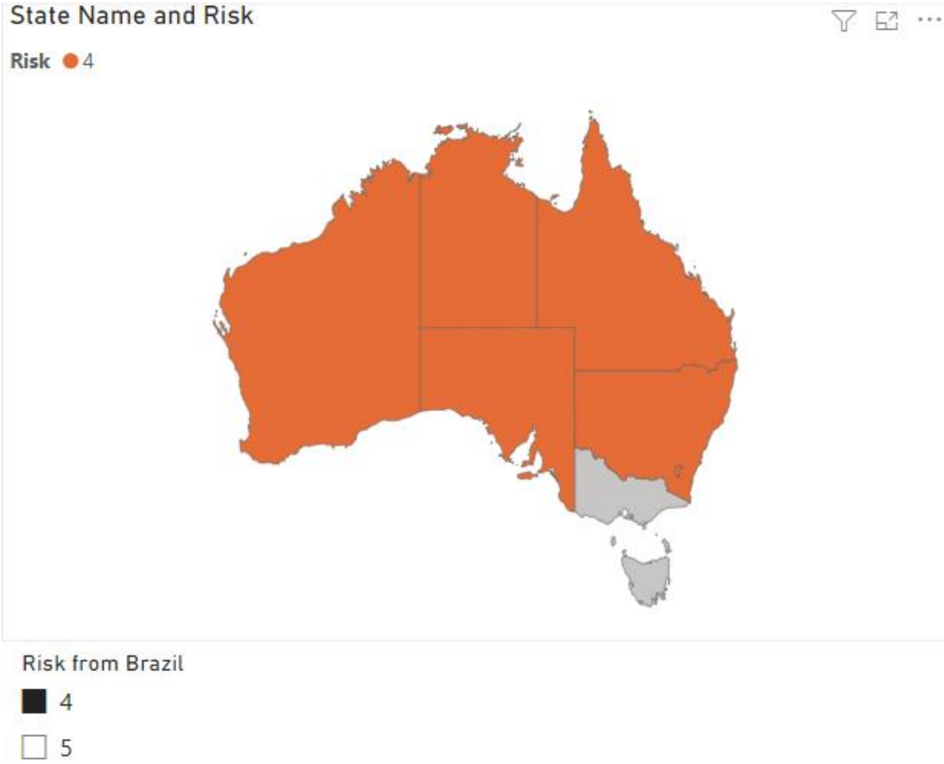


Figure 21 States That Have Risk Level 4 from Brazil

Figure 22 States That Have Risk Level 5 from Brazil

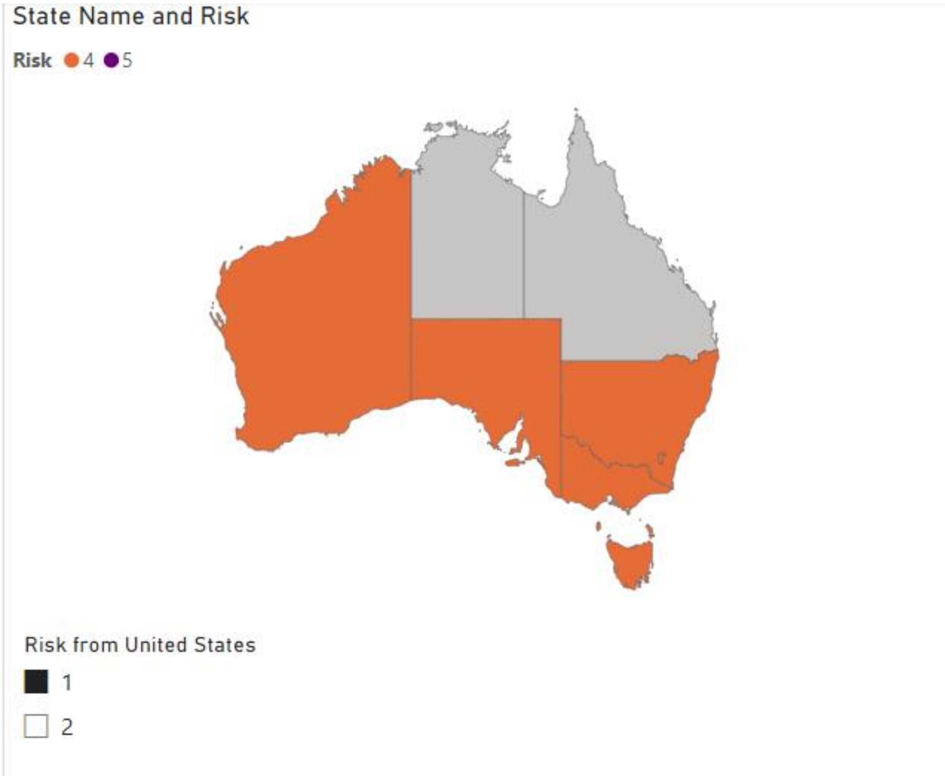


Figure 23 States That Have Risk Level 1 from United States

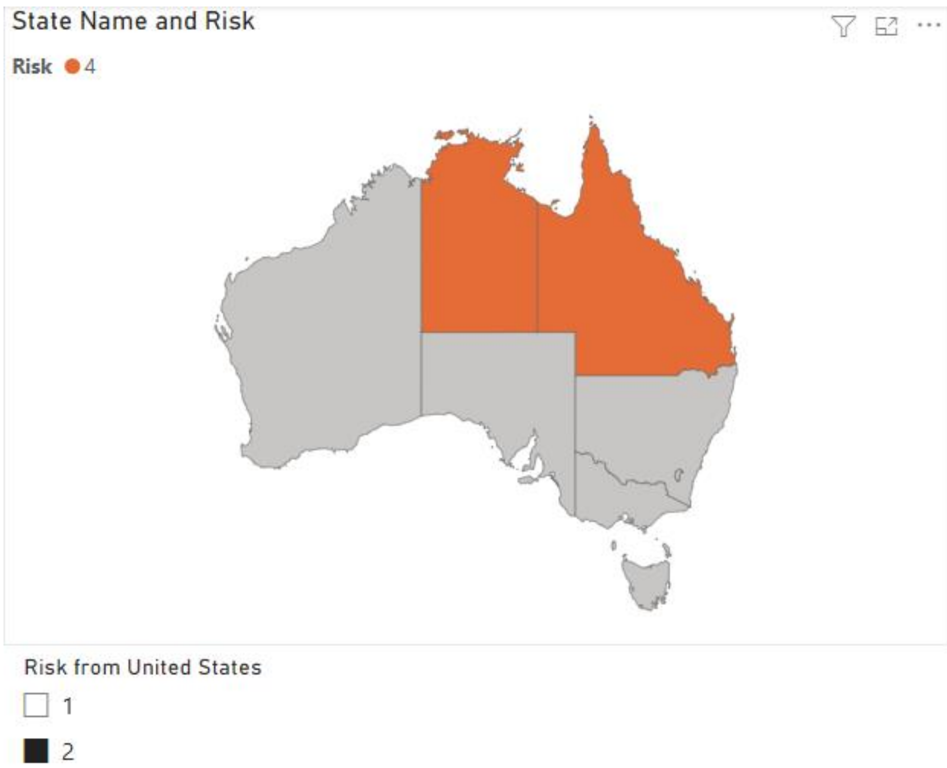


Figure 24 States That Have Risk Level 1 from United States

Figure 25 Risk Levels for Other Countries in Every State of Australia

Based on the graphs above, we can see that:

- Immigrants from *New Zealand* have lowest risk compared to other countries (risk level 1).
- On average, immigrants from *India* have the highest risk across Australia (risk level 5).
- Immigrants from *Brazil* have higher risk in *Victoria* and *Tasmania* compared to other states in Australia.
- Immigrants from *United States* have higher risk in *Northern Territory* and *Queensland* compared to other states in Australia.
- *New Zealand* and *India* have the same value of risks across all states.

However, after gathering all the information based on the Data Visualization, there are some special considerations, which will be added in the future. For *Average Temperature Per Season Per Country*, we used the average value, which can be inaccurate to tell the difference of climate among all six selected countries because any outliers can influence the accuracy of the average temperature and temperature uncertainty. Thus, in the future, we are planning to use the range of the temperature and its uncertainty (the maximum and minimum values). We are also planning to add more countries to make the data wider, which is better for the Australian Migration System.

For *Timeline Graph for Temperature in Australia*, currently we can only go for one residential area, which is very small compared to the entirety of Australia. Therefore, in the future, we are planning to analyse the timeline graph per state and instead of one area, we will add more to see the difference of climate per state from time to time. In addition, we will also use data that are more complete to make the timeline graphs more in sync from one to another.

For *Migration Risk Mapping*, the original data was on each postal code, NOT state area. However, the visualization map cannot load the shape map based on the postal code. At the end, to visualize the data, we had to regrouped it into per state and value based on the mean of the original data and round them on each state. In the future, we will map the risk value in more detail. Instead of mapping the risk on each state, we will go for the original plan, mapping the risk per postal code.

5. Impact and Significance of Results

Phase one

The phase one of our application will construct an app that will provide industries such as government, health organizations or insurance company with the information surrounding the impact of immigration on individuals' health due to the climate differences. As shown in the section above with the PowerBI desecration, at the currently stage the potential health risk due to temperature differences between Australia and the country that individuals migrate from is analysed as shown in the section above, the temperature distance matric were measured with Euclidean distance. This information provided can allow the organizations reduce related health cost in various ways:

In the health organization's perspective, our application provides potential health risk that immigrate could have. Addition health check can be provided by the health organization to the immigrate that could allow them to find the probable health issues that an induvial could have due to the migration in an early stage if any disease is going to occurred.

The insurance company can provide insurance plans to the immigrants that migrates form other countries that identified as higher health risk with addition respiratory and heart checks. This will allow them to identify their customer's potential health risk easier and provide better support to their customer.

The government can have the information provided from the application also work together with health organizations and insurance company to regulate policy that could reduce related health cost and to support the immigrate with health support. Identifying health issues in an early stage and additional health support from insurance plan can highly reduce the cost in public health.

Phase two

While further process to phase two of our application, we create an information portal that individuals from other countries can use to explore a range of options for where they could potentially move to Australia.

The portal we provided would suggest similar climates, also giving the information that whether the average temperatures of certain area in Australia is warmer or cooler also precipitation compared to the city or area that they choose based on the user's selections as well as highlighting the potential health risks. Our portal would also show climate predictions, and how the environment may change over the next century. With this information our portal would suggest a state or a city in Australia that similar to the user's selected place and provide a similarity percentage.

A data visualization of the whole map of Australia is also shown in the interface that could demonstrate the user though out Australia which area will have similar weather conditions with their selected area.

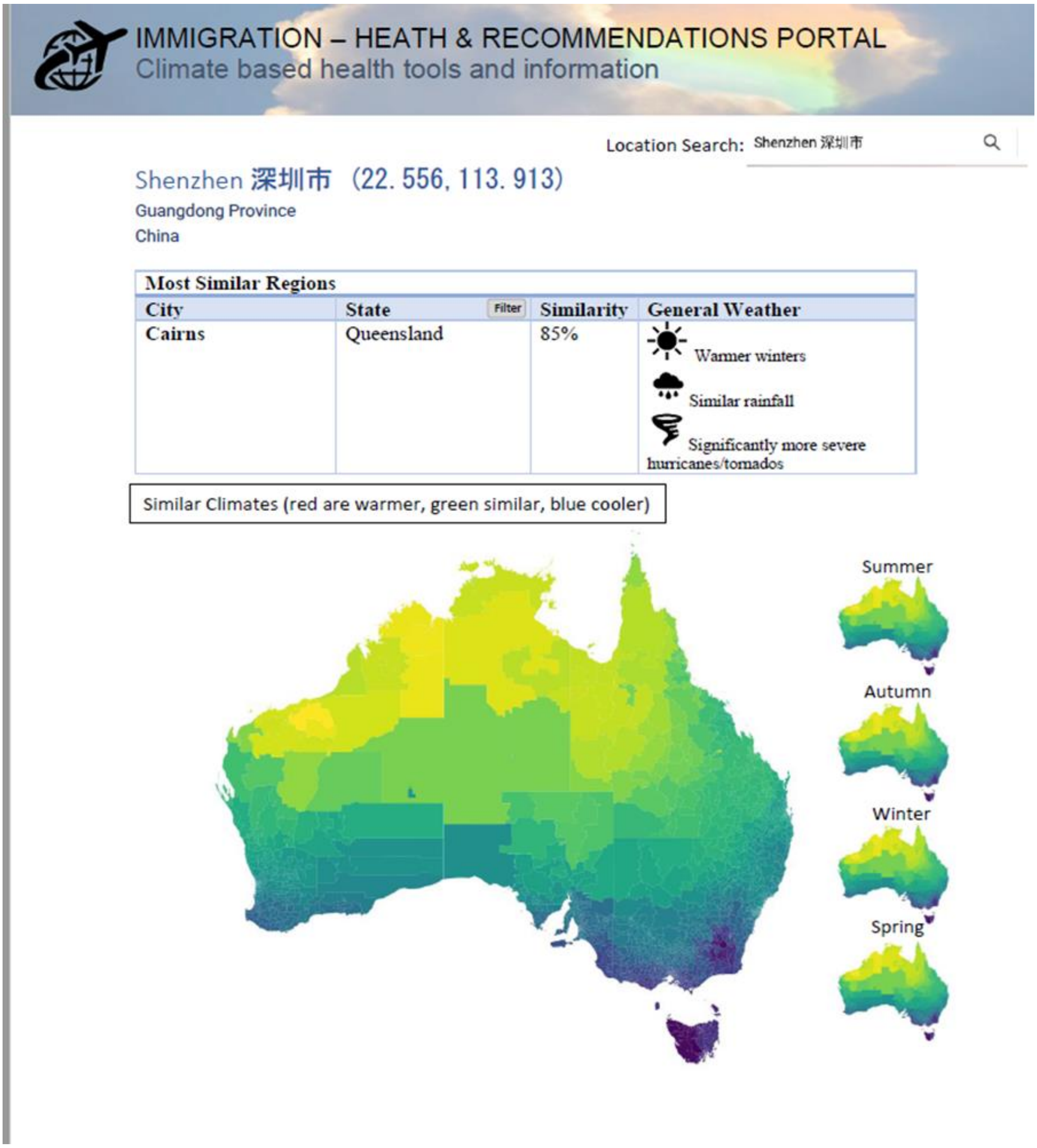


Figure 26 Prototype of Web based information portal for phase 2

This portal could enhance the application in our phase one development, with the feedback from the portal we are able to provide the government an idea of which countries will have higher tendency for migration to relative sectors of Australia. This portal can be used by the government, health organizations or insurance companies to provide information and suggestions to their potential customers. The government can formulate policies to encourage people to migrate to Australia and the insurance company can provide related insurance plans also the health organization can also provide health support to potential health risk to immigrate.

6. Project Management

Agile Methodology & Jira

The group agrees to adopt agile methodology in this project. The length of the sprint is set as a week. There were 7 sprints in total for this two and half months as there were some sprints extended for several reasons, e.g., lots of other course's assignment in the week. The extension was considered different aspects, such as availability of group members and project schedule.

The team uses Jira as project management tools. Jira provides a handy dashboard for handling sprints (See Figure 27). A JIRA, which can be a task, user story or bug, can be drag-and-drop from backlog and incomplete JIRAs will be automatically shifted to a new sprint with completing a current sprint.

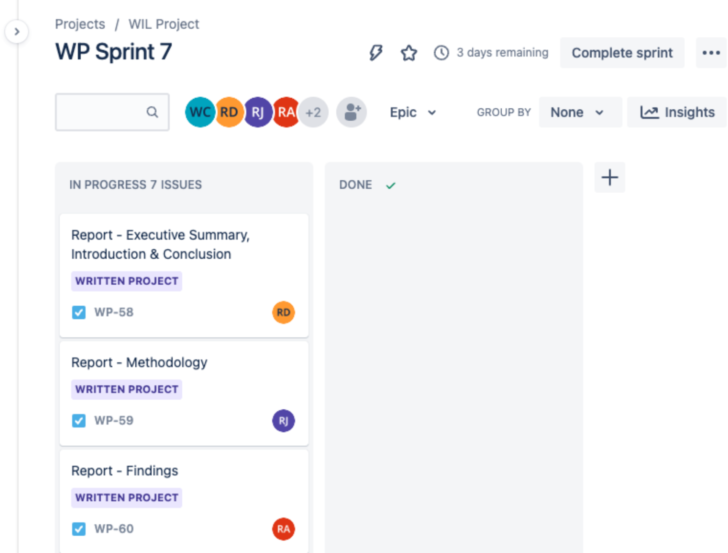


Figure 27 Jira Sprint Dashboard

Gantt Chart and Epics

Besides sprints, there are 4 epics. They are milestone 1, MVP, video presentation and written report. Those epics clearly divided JIRAs into different aspects. There is a roadmap in Jira which shows as a Gantt chart (See Figure 28). It shows a clear plan for the project. Groupmates can check the current sprint and which epics that a particular JIRA is in.

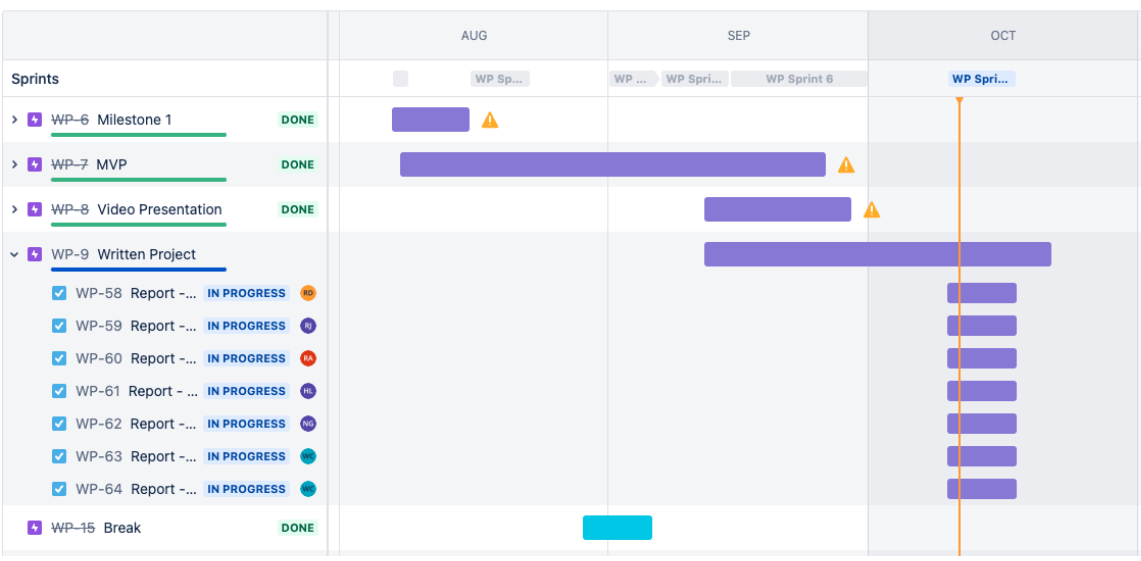


Figure 28 Jira Roadmap - Gantt Chart group by Epics

Documentation & Confluences

There are a lot of documentation work for a project, especially for a data science project. The team starts to save files in SharePoint folder which is integrated with Microsoft Teams. But the files were messy in format and hard to search with contents. A content collaboration tool as a web application called Confluence is then used. And both Jira and Confluence are developed by Atlassian. They are in a package and already integrated which means Jira can easily link to Confluence, vice versa. In milestone 1, we create a JIRA with the user story (e.g., as an immigrant, I would like to know how the weather of a city in Australia so that I can plan to move to which city) and assign somebody to do some research work. The assignee posted the result in Confluence and made a link to Jira. Other groupmates can easily find the research by searching the user story in Jira, vice versa. (See Figure 29)

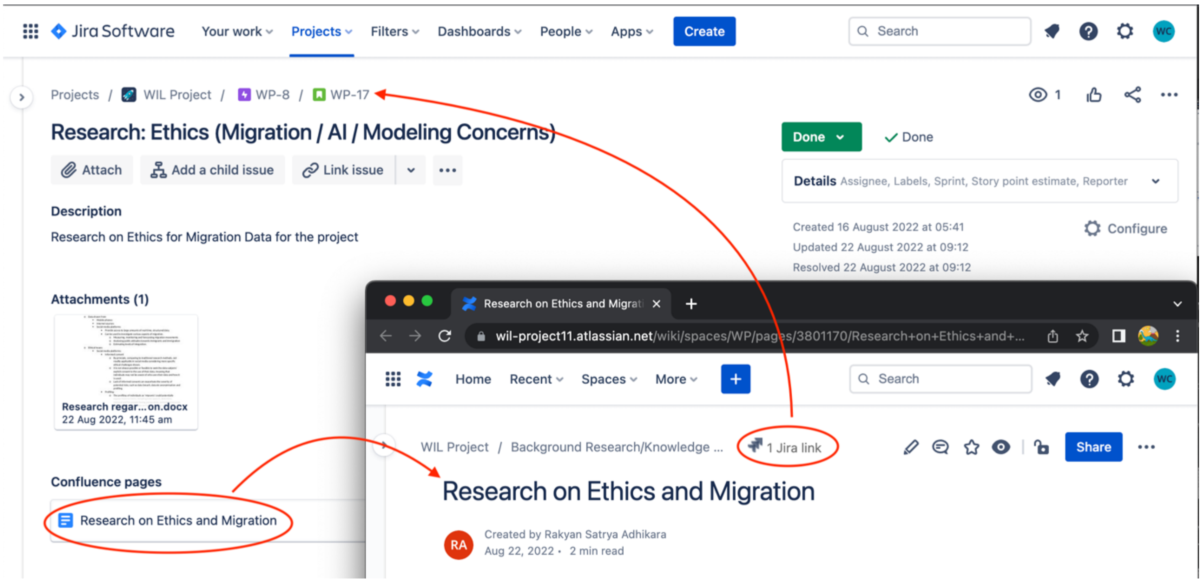


Figure 29 Linking between Jira and Confluences

Jira dashboard

Besides Roadmap, Jira also provides a customized dashboard for the project manager to keep track of the project status. (See Figure 30) Users can create some filters by Jira default parameters or by Jira Query Language (JQL). JQL provides a flexible way to create any filters and its syntax seems like SQL. The dashboard requires you to apply some Jira filters. Users can save the filters after searching by JQL and can be used for displaying some customized results in the dashboard. The dashboard can also be shared to the teams. Any team member can use the dashboard to know the status of the project as well as their own tasks.

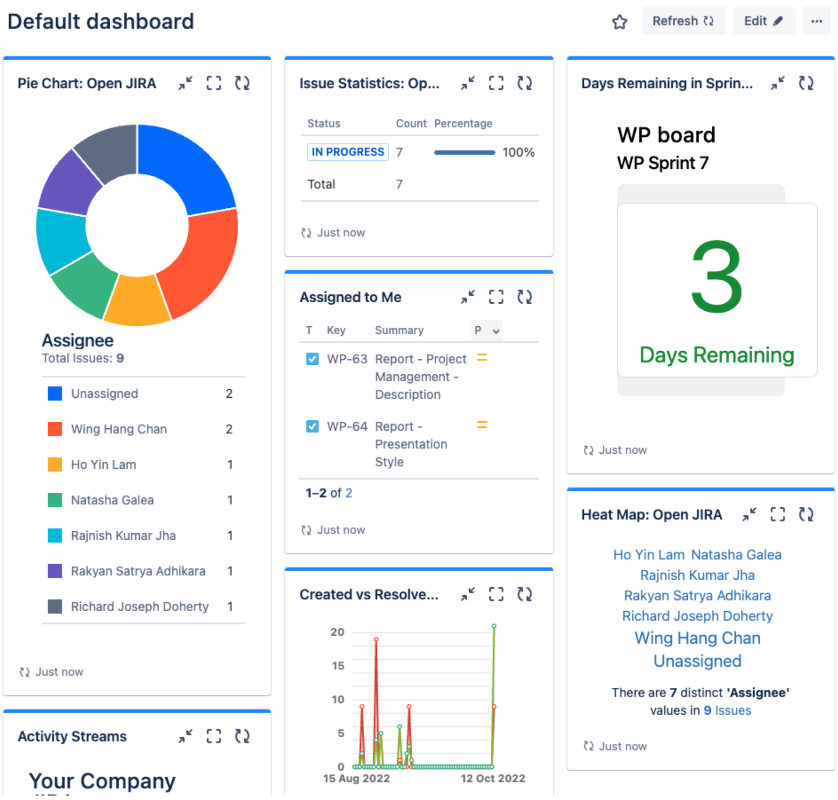


Figure 30 Jira Customized Dashboard for project management

Version control - Git

The team uses Git as this project version control. This is a private project posted in GitHub (See Figure 31). As it is not a big project, and each member has different roles. A particular file of source code is usually developed by the same person. The team did not have the conflict issue but there would be some conflicts in the local environment. It can be easily solved by replacing files from the internet or merging with local changes. Also, there is no branch created for the project as the project is not deployed to production. There is no need to create a branch for another version.

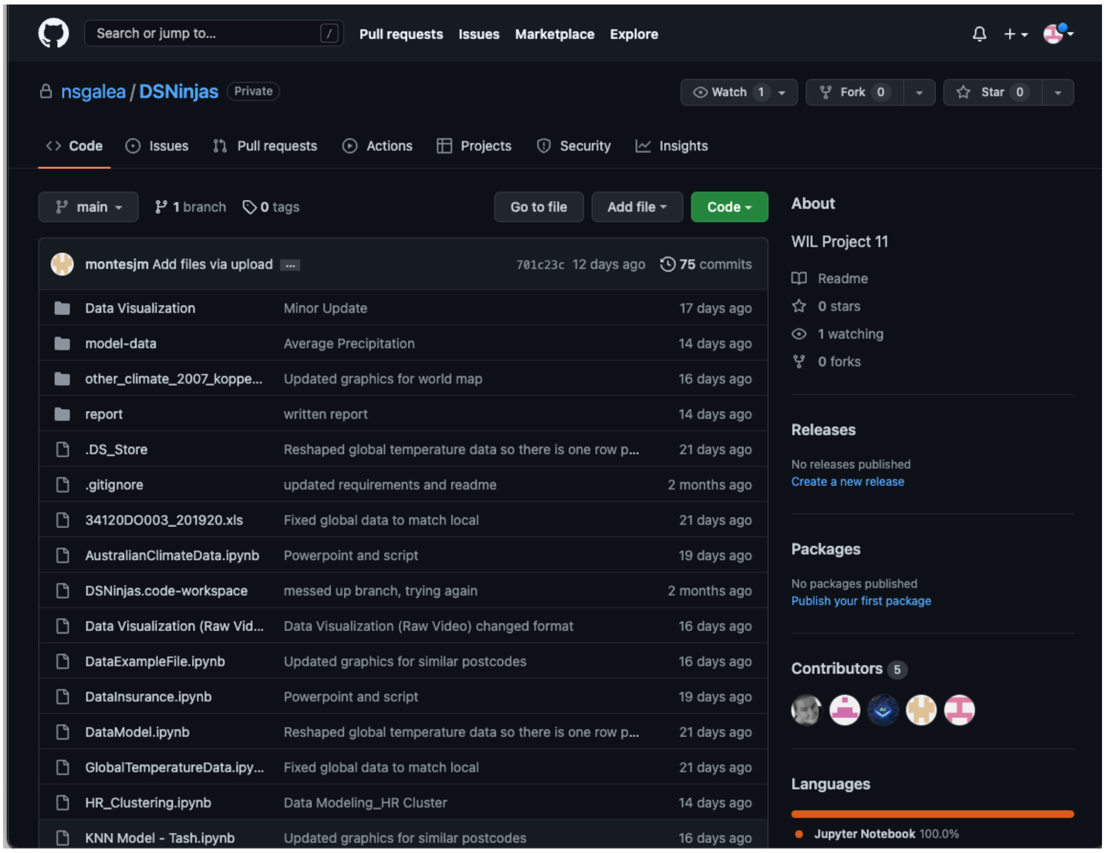


Figure 31 Github page for the project

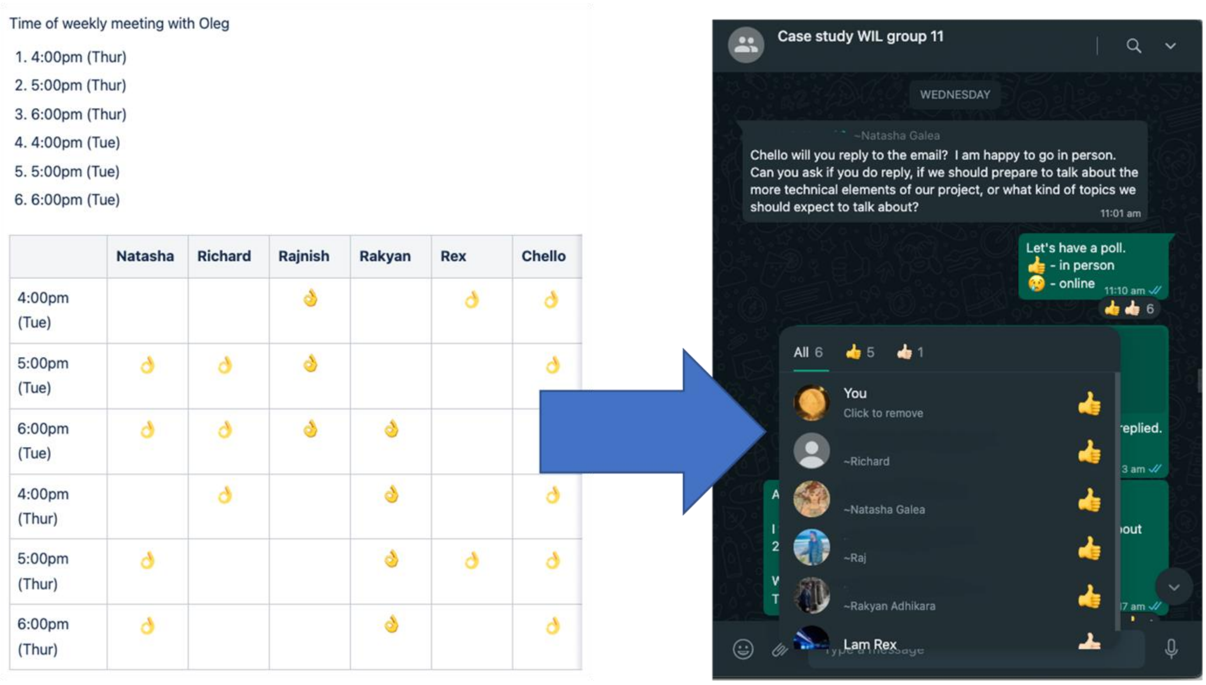
Communication

There are several communication channels for the team, such as email, messaging in Microsoft Teams (MS Teams), WhatsApp, etc. It would be quite confusing as the team did not choose one for the primary channel at the beginning. After a few communication issues occurred, the group decided to use WhatsApp as primary channel and use other channels as supplementary. The team keeps using MS Teams for conferencing. Online conferencing is flexible, but it is not as efficient as face-to-face meetings. However, it is hard to find a time for every member to meet as some members have full time jobs. So, we usually do it in hybrid mode, i.e., some group members meet at the library and some join with MS Team. There is a benefit for using MS Team compared with face-to-face meetings. That is recording. We usually recorded the meeting, especially meeting with the mentor. Team members who cannot join can watch the playback. Other than MS Teams, we use WhatsApp for instant messaging. The team would discuss some topics in WhatsApp group chat if it were urgent. There is another function which used with WhatsApp to solve a problem which will be discussed in the next section.

Issues and resolution of issues

Response of Polling

It is a team project with six members. The group always needs to poll for some topics, for example the next meeting date and choosing which user stories. Some polls can be done within the meeting. But there would be some ad-hoc poll which cannot be discussed in the meeting. It was done by a Confluences page at the beginning. But it takes time for every teammate to respond. It is then changed the poll in WhatsApp by responding to the chat with emojis. (See Figure 32) It solves the slow reply issue. If there is any complex poll such as many options, it would be done within meetings or divided a complex poll into 2 or more polls.



Time of weekly meeting with Oleg

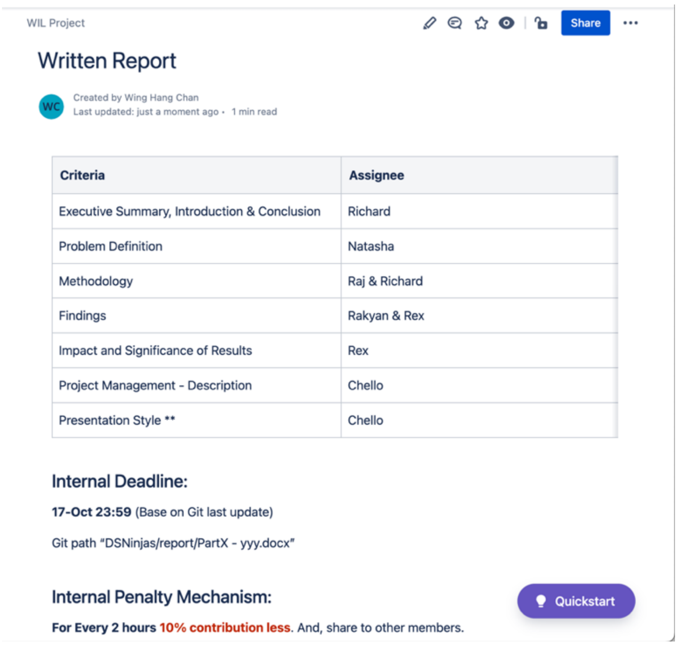
1. 4:00pm (Thur)
2. 5:00pm (Thur)
3. 6:00pm (Thur)
4. 4:00pm (Tue)
5. 5:00pm (Tue)
6. 6:00pm (Tue)

	Natasha	Richard	Rajnish	Rakyan	Rex	Chello
4:00pm (Tue)			👌		👌	👌
5:00pm (Tue)	👌	👌	👌			👌
6:00pm (Tue)	👌	👌	👌	👌		
4:00pm (Thur)		👌		👌		👌
5:00pm (Thur)	👌			👌	👌	👌
6:00pm (Thur)	👌			👌		👌

Figure 32 Screen capture of Confluence page and WhatsApp for polling

Internal Hand-in Mechanism

There are some hard deadlines for the project. But the team tried to make the product perfect and therefore we edited the product at the very last minute. To tackle this problem, we introduce an internal deadline with an internal penalty mechanism. (See Figure 33) There will be contribution deduction for the latecomer. There is a special consideration mechanism. If there are more than half of the group members voted for the applicant of special consideration. It will still equally contribute. The internal deadline also solves a problem that teammates who do not seek help or ask for help at the very last minute. The mechanism is clearly stated in a Confluences page and agreed by all team members.


WIL Project

Written Report

Created by Wing Hang Chan
Last updated: just a moment ago • 1 min read

Criteria	Assignee
Executive Summary, Introduction & Conclusion	Richard
Problem Definition	Natasha
Methodology	Raj & Richard
Findings	Rakyan & Rex
Impact and Significance of Results	Rex
Project Management - Description	Chello
Presentation Style **	Chello

Internal Deadline:

17-Oct 23:59 (Base on Git last update)

Git path "DSNinjas/report/PartX - yyy.docx"

Internal Penalty Mechanism:

For Every 2 hours 10% contribution less. And, share to other members.

Figure 33 Confluences page for Internal Hand-in Mechanism

Improvement

We used many tools as mentioned before to manage the project. However, there is still room to improve. For example, we can integrate Jira with GitHub. Teammates can know each other works of source code change through Jira. It can increase groupmates' intention to update Jira status.

Contribution

Name	Student ID	Contribution
Ho Yin Lam	3889140	100%
Natasha Sheree Galea	3540515	100%
Rajnish Kumar Jha	3821892	100%
Rakyan Satrya Adhikara	3935019	100%
Richard Joseph Doherty	3863706	100%
Wing Hang Chan	3939713	100%

Table 3 Contribution table

7. Conclusion

The minimum viable project for *monitoring Migration Health Risk in Relation to Climate Change was created by ideation and synergizing of past, current and potential future migration to Australia in relation to similarities in climate* was able to successfully demonstrate the possibility of developing an informative and useful tool. The website's potential lies in directing migration to likely similar climate zones within Australia. Whereafter a pool of useful data could inform government and healthcare agencies; extrapolating data for potential healthcare policies and reform shared with insurance agencies.

The website itself represents the different climate zones by visualization and interactive application based on underlying machine- learning algorithms and models that are widely accepted within the industry.

Further ideation and sectional developments of the web-based application have the potential to utilize additional climate data in this model. The reach of this machine learning algorithm could be extended to:

1) Clustering of data pertaining to climate and migration.
2) Further data driven insights and realizations.
3) Optimizing user interface to broaden the acceptability and widen the potential reach to the ordinary man on the street (potential migrant).
4) Greater capabilities would allow targeted comparisons of climate-based zones and further identification of potential health risks related to potential migration from one zone to another.

However, the model had several flaws. One of the primary issues that needs to be developed in the next iteration, was clearly identifying the factors that go into "climate similarity". For example, is temperature variability or mean temperature more important for a climate? Adding precipitation would also allow for more accurate climate assessment, however more accurate may not be more reflective of the "feel" of the climate. T focus of the initial prototype MVP was on developing the ability to compare climates and assessing the accuracy of such a model. However, future models need to focus more on user "perception" and health facts, depending on the audience. When targeting individuals considering moving, their primary goal may be comfort and how good the climate "feels", however government and businesses may be more concerned with the health effects of the climate similarity. Therefore, future roadmap enhancements, would primarily be focused on user interface and providing information in a clear way, that the user can use to make their own decision about the risks.

Overall, the project was a success, but there is a lot of room for development in the future. It proved that it is possible to develop an accurate model for climate comparison. However, the model still required development in the areas of visual communication, as well as the addition of more features including: precipitation, humidity, seasonality, climate, extreme weather events, social facility access and language access.

Reference

[1] “Map explorer,” 6 10 2021. [Online]. Available: https://www.climatechangeinaustralia.gov.au/en/projections-tools/map-explorer/.

[2] “Overseas Arrivals and Departures,” Australian Bureau of Statistics, July 2022. [Online]. Available: https://www.abs.gov.au/statistics/industry/tourism-and-transport/.

[3] K. Vilkman, S. Pakkanen, T. Lääveri, H. Siikamäki and A. Kantele, “Travelers’ health problems and behavior: prospective study with post-travel follow-up,” *BMC Infectious Diseases,* vol. 16, no. 1, 2016.

[4] M. Lee, F. Nordio, A. Zanobetti, P. Kinney, R. Vautard and J. Schwartz, “Acclimatization across space and time in the effects of temperature on mortality: a time-series analysis,” *Environmental Health,* vol. 13, no. 1, 2014.

[5] S. Vardoulakis, K. Dear, S. Hajat, C. Heaviside, B. Eggen, and A. J. McMichael, “Comparative assessment of the effects of climate change on heat- and cold-related mortality in the United Kingdom and Australia,,” *Environmental Health Perspectives,* vol. 122, no. 12, pp. 1285-1292, 2014.

[6] R. Basu and J. M. Samet, “Relation between elevated ambient temperature and mortality: A review of the epidemiologic evidence,,” *Epidemiologic Reviews,* vol. 24, no. 2, pp. 190-202, 2002.

[7] “Overseas Arrivals and Departures,” July 2022. [Online]. Available: https://www.abs.gov.au/statistics/industry/tourism-and-transport/.

[8] C. F. Wright, S. Clibborn, N. Piper, and N. Cini, “Economic Migration and Australia in the 21st Century,” Think Asia,, 01 Mar 2017. [Online]. Available: https://think-asia.org/handle/11540/6737.

[9] S. i. Australia, “Overseas student health cover,” Studies in Australia, 18 Oct 2022. [Online]. Available: https://www.studiesinaustralia.com/studying-in-australia/how-to-study-in-australia/student-visas/overseas-student-health-cover.

[10] F. Murtagh and P. Contreras, Methods of Hierarchical Clustering, Dublin,, Ireland: University of London, 2011.

[11] S. Kumar, Agglomerative Clustering and Dendrograms — Explained, Medium, 2020.

[12] T. Li, A. Rezaeipanah and E. M. Tag El Din, An ensemble agglomerative hierarchical clustering algorithm based on clusters clustering technique and the novel similarity measurement, Vols. 34, Issue 6, Part B, Journal of King Saud University - Computer and Information Sciences, 2022, pp. 3828-3842.

[13] H. E. Beck, N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg and E. F. Wood, Present and future Köppen-Geiger climate classification maps at 1-km resolution, vol. 5, Scientific Data, 2018.

[14] SciJinks, “What Are the Different Climate Types?,” SciJinks, 6 October 2022. [Online]. Available: https://scijinks.gov/climate-zones/.

[15] F. Murtagh, “Methods of Hierarchical Clustering,” Deblin, 2011.

[16] S. Kumar, “Agglomerative Clustering and Dendrograms - Explained,” Medium, 2022. [Online]. Available: https://towardsdatascience.com/agglomerative-clustering-and-dendrograms-explained-29fc12b85f23#:~:text=Dendrograms%20are%20a%20diagrammatic%20representation,of%20hierarchical%20(agglomerative)%20clustering.