Unravelling the Climate Tapestry: Exploring the anomalies that impact the Australian Climate, with a Focus on Melbourne

ADS1002 Semester 2 2023

	Name	Contribution %	Contribution Description
1	Gideon Cher	20%	Report, Coding
2	Joshua Hudson	20%	Report, Coding
3	Stefano Nicholas Rusli	20%	Report, Coding
4	Youjing Gan	20%	Report, Coding
5	Thi do	20%	Report, Coding

Final project report

Part 1 Description of the project:

In today's day and age, civilisation is progressively influenced by the vagaries of climate change, thus it is pivotal to understand the complex intelligence between these climate components. Data science plays a key role in understanding and solving the challenges of climate change. In this project, we use data science techniques to unravel the complex climate network affecting Australia, specifically Melbourne. Analytical data helps us make informed decisions, predict climate patterns, and adapt to changing conditions. This project, Unravelling the Climate Tapestry, aims to illuminate the complex relationship between the three main climate change drivers affecting Australia's climate: the Southern Oscillation index (SOI), the Indian Ocean Dipole (DMI), the Southern Annular Mode (SAM) and our main focus is how and why they play an important role in shaping Australia's climate.

Providing a wealth of information on these climatic factors the Australian Bureau of Meteorology (BoM) gives detailed explanations and index calculation methods that can be found on their official website, making this project both informative and scientifically sound. The three climate drivers we used being SOI, DMI and SAM provided us with the data that would help us predict which anomalies would specifically affect the Australian climate. Thus, suggesting that when the:

- Soi values are below -7 it indicates El Nino.
- When El Nino occurs it often results in heat waves and droughts for Australia.
- When Soi values are above 7 it indicates la Nina, resulting in increased rainfall.
- A positive DMI will likely result in less rainfall and higher temperatures.
- A negative DMI will likely result in higher rainfall and lower temperatures.
- A positive SAM value will result in an increased chance of rain along the east coast and particular colder temperatures.
- A negative SAM result will have a decreased chance of rain and higher temperatures which will also be impacted depending on which season is current.

Climate data for Australian cities, including Melbourne, is also available from the BoM repository. Data including monthly rainfall, average minimum and maximum temperature for Melbourne is published. For other Australian cities, we combined information from the BoM website. Using this information, we created models to predict how these climate drivers would influence the weather in Australia and aim to forecast the weather anomalies based on these drivers.

Considering that the main objective of this project was to model climate anomalies for Melbourne and other Australian cities as a function of SOI, DMI and SAM indices, there were specific tasks we had to undertake to achieve successful results. We also chose Darwin and Brisbane as two cities to compare to Melbourne as they would give us stronger insights to the anomalies affecting the Australian climate. We chose them by recognising the most affected areas by the weather cycles in these three graphs, as the areas which were mostly influenced would give us more noticeable anomalies.

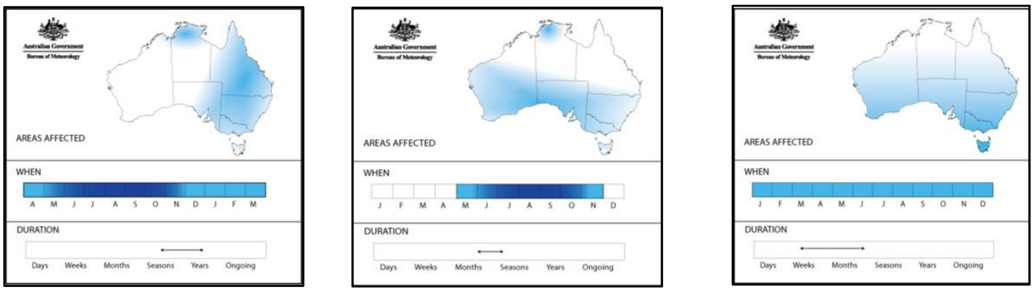


Figure 1.1, 1.2 and 1.3. the three graphs above show the areas affected by the El Nino (left), the DMI (middle) and the SAM (right) , when the effect can occur and how long they last.

This project opens the doors to several avenues of research. Although our initial approach is linear regression, we used other methods such as support vector machines, decision trees, and random forest can provide better modelling accuracy. In addition, we can study the effect of regularization on linear regression, smoothing data or converting outliers into a classification problem.

This project is particularly important at the current moment because recent news is warning of an upcoming El Niño cycle affecting Australia. El Niño is part of ENSO, which is one of the most important climate factors we study. Understanding this effect is crucial and our project aims to provide insight into how ENSO, together with the IOD and SAM, affects the climate of Melbourne and other Australian cities. Our results help to better understand the approaching El Niño and its potential impact on Australia.

Part 2 Details of pre-processing and manipulation of data in Python

Background

The primary objective of this project is to analyse climate patterns and their drivers in Australia. To facilitate this analysis, we were provided with multiple datasets in CSV format. These datasets included climate data comprising rainfall, maximum temperature (max temp), and minimum temperature (min temp) from two different weather stations in Melbourne. Additionally, we had data on climate drivers, specifically the Southern Annular Mode (SAM), the Dipole Mode Index (DMI), and the Southern Oscillation Index (SOI).

Data Pre-processing Steps

1. Data Import

The project commenced with the initial step of importing data. We read the CSV files for each climate variable (rainfall, max temp, and min temp) from both stations, resulting in a total of six distinct datasets.

2. Addressing Repeated Years

A critical observation during data exploration was the presence of repeated years in both station datasets. To ensure data consistency and avoid redundancies, we decided to remove the years 2013 to 2015 from station 2. This step was crucial for preventing duplicate data entries.

3. Concatenation of Data

Following the resolution of the repeated years issue, we proceeded to concatenate the datasets from both stations for each climate variable. This process led to the creation of three consolidated datasets, one for each climate variable.

4. Handling Missing Data

During data inspection, it became apparent that there were missing values in the datasets. Specifically, we identified 21 missing values in the rainfall dataset and 20 missing values in both the max temp and min temp datasets. To address this, we chose to impute the missing values with the mean. The decision to use mean imputation was based on the relatively small number of missing values and the assumption that they would not significantly skew the data distribution.

5. Removal of Redundant Columns

For the sake of data clarity and efficiency in future analysis, we eliminated redundant columns that were not relevant to the project. These columns included “postcode”, “station number”, “annual” the repeated “year” column.

6. Calculation of New Annual Data

Subsequently, we calculated new annual data based on the cleaned datasets that no longer contained missing values. We also set the index of the datasets to be the “year” column. This index setup simplifies future analysis and visualization.

7. Subset Selection

To create a final dataset for the climate variables, we selected data covering the years from 1975 to 2022 because using prior data from earlier will skew the results due to climate change and an increase in temperature over time. This particular date range was chosen to align with the analysis period for the climate drivers, ensuring that the datasets complement each other effectively.

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
1975	20.8	7.0000	48.200000	19.600000	44.400000	29.2000	35.800	109.400000	98.600000	150.200000	75.800000	70.800000	709.800000
1976	24.1	20.0000	33.200000	14.400000	21.200000	29.8000	17.200	56.200000	78.800000	72.400000	95.800000	40.800000	503.900000
1977	75.4	30.2000	27.000000	64.400000	59.600000	109.0000	46.800	24.000000	84.800000	22.200000	36.000000	25.600000	605.000000
1978	52.4	41.0000	79.800000	57.800000	110.600000	25.8000	52.800	93.400000	46.400000	68.600000	128.000000	110.400000	867.000000
1979	36.4	43.0000	27.000000	44.800000	88.600000	16.6000	9.400	63.800000	62.200000	112.000000	27.400000	12.000000	543.200000
1980	62.8	20.2000	15.800000	76.200000	62.800000	87.0000	33.800	46.800000	27.400000	96.000000	44.800000	70.600000	644.200000
1981	40.2	32.6000	44.400000	20.800000	107.200000	56.0000	66.200	81.400000	16.600000	54.600000	54.400000	27.800000	602.200000
1982	51.2	8.8000	55.200000	35.600000	49.000000	30.4000	12.600	28.800000	44.400000	41.200000	17.600000	47.400000	422.200000
1983	29.8	0.6000	46.000000	40.400000	55.200000	32.4000	63.600	58.000000	81.800000	114.200000	70.400000	19.600000	612.000000
1984	70.0	18.8000	38.400000	52.400000	10.200000	23.8000	41.400	52.200000	115.200000	41.400000	55.000000	46.000000	564.800000
1985	16.6	6.2000	49.400000	62.200000	53.800000	62.8000	37.800	80.800000	42.000000	52.800000	77.000000	136.800000	678.200000
1986	22.4	16.0000	17.000000	55.200000	73.800000	28.2000	76.800	48.400000	27.400000	53.800000	38.400000	70.200000	527.600000
1987	57.2	57.2000	50.200000	18.600000	85.200000	51.2000	68.800	23.200000	38.000000	38.800000	82.000000	84.600000	655.000000
1988	55.2	19.2000	29.400000	10.200000	58.000000	50.6000	56.600	53.000000	49.600000	27.400000	165.800000	103.600000	678.600000
1989	47.6	44.2000	81.800000	92.800000	56.000000	90.4000	51.200	77.800000	40.600000	90.400000	31.200000	88.800000	792.800000
1990	3.8	74.4000	32.000000	94.600000	17.200000	37.2000	91.000	49.800000	54.400000	96.600000	47.600000	26.600000	625.200000

Figure 2.1 the figure above is the final dataset for rainfall.

Climate Driver Data

The datasets containing climate driver information (DMI, SAM, and SOI) were already found to be in a clean and usable format. We followed a straightforward process, involving the import of the respective CSV files and selecting the data for the years between 1975 and 2022, thus matching our climate dataset.

Year	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1975	-0.367	-0.249	-0.464	0.087	0.066	0.055	-0.039	-0.344	-0.857	-1.056	-0.328	-0.170
1976	0.005	0.077	0.020	0.122	-0.004	0.369	0.594	0.107	-0.208	-0.231	-0.181	0.028
1977	-0.003	-0.261	0.174	-0.029	-0.341	-0.095	-0.171	-0.260	-0.259	0.245	-0.127	-0.020
1978	-0.640	-0.530	-0.582	-0.214	-0.533	-0.174	-0.107	-0.224	-0.261	-0.502	-0.458	-0.260
1979	0.317	-0.158	-0.034	-0.207	-0.444	0.108	-0.315	-0.226	-0.280	-0.314	-0.251	-0.101
1980	-0.158	-0.200	-0.417	-0.095	-0.066	-0.382	-0.661	-0.822	-0.745	-0.657	-0.417	-0.483
1981	-0.201	-0.024	0.027	0.092	-0.018	-0.240	-0.560	-0.628	-0.757	-0.606	-0.328	-0.019
1982	0.144	0.166	0.054	0.119	0.223	0.255	0.265	0.256	0.442	0.623	0.284	-0.162
1983	-0.482	-0.587	-0.752	-0.556	-0.059	0.371	0.525	0.345	-0.069	-0.288	-0.343	-0.101
1984	-0.223	-0.149	-0.122	0.050	-0.305	-0.357	-0.366	-0.498	-0.608	-0.654	-0.418	-0.278
1985	-0.466	-0.627	-0.541	-0.200	-0.223	-0.629	-0.403	-0.459	-0.238	-0.498	0.074	-0.417
1986	-0.113	-0.135	-0.200	-0.286	-0.154	-0.278	-0.546	-0.401	-0.142	-0.048	-0.244	-0.274
1987	-0.081	0.041	-0.137	-0.157	0.165	0.109	0.207	0.297	0.393	0.255	-0.051	0.152
1988	0.353	-0.154	-0.313	-0.128	-0.513	-0.272	-0.150	-0.289	-0.394	-0.525	-0.183	0.156
1989	-0.281	-0.045	-0.319	-0.480	-0.594	-0.780	-0.447	-0.321	-0.225	-0.412	-0.338	-0.167
1990	-0.099	-0.289	-0.143	-0.385	-0.348	-0.568	-0.246	-0.392	-0.183	-0.345	-0.094	0.003
1991	0.065	-0.097	-0.037	0.289	0.379	0.263	0.270	0.058	0.099	-0.030	0.040	0.084

Figure 2.2 the figure above is the final dataset for SAM.

Deseasonalization of Climate Data (indexing)

Before proceeding with exploratory data analysis and modelling, we undertook an important data transformation step by deseasonalizing the climate datasets which are the rainfall, maximum temperature, and minimum temperature dataset. Deseasonalization was achieved by subtracting the monthly mean from each data point for that particular month. This process serves two critical purposes:

Why Deseasonalize Climate Data:

Remove Seasonal Variations: Rainfall data often exhibits seasonality, where certain months experience higher or lower average rainfall due to weather patterns. Deseasonalization effectively removes this seasonal or periodic component, enabling a more focused analysis of underlying trends.

Enhance Anomaly Detection: Deseasonalizing the data simplifies the detection of anomalies or deviations from the expected seasonal patterns. For instance, unusually high rainfall in a typically dry month, or vice versa, may signal an anomaly that warrants further investigation.

Interpreting Deseasonalized Data:

Positive Values: A positive value post-deseasonalized indicates that the observed data for that month surpasses the long-term monthly average. This suggests that there was more of the respective climate variable (e.g., more rainfall, higher minimum temperature, or higher maximum temperature) than expected for that month.

Negative Values: Negative values signify that the observed data for that month falls below the long-term monthly average. This implies that there was less of the respective climate variable (e.g., less rainfall, lower minimum temperature, or lower maximum temperature) than expected for that month.

Zero Values: A deseasonalized value of zero conveys that the observed data matches the long-term monthly average for that specific month. In other words, the respective climate variable (e.g., rainfall, min temp, or max temp) for that month aligns with historical averages and does not deviate from the expected norm.

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1975	-2.544898	0.414583	-1.435417e+00	-1.652083	5.958333e-01	-8.458333e-01	1.442553	-1.593617e+00	-0.012766	-1.423404	-0.046809	-3.212766e-01
1976	-1.444898	0.614583	-3.354167e-01	-0.652083	-8.041667e-01	-4.458333e-01	-0.157447	-7.936170e-01	-1.412766	-3.123404	-1.246809	4.787234e-01
1977	-1.144898	-0.185417	-7.354167e-01	-2.352083	-1.104167e+00	-1.445833e+00	-1.457447	9.063830e-01	-2.312766	0.876596	-1.146809	-8.212766e-01
1978	-2.444898	-1.685417	-1.735417e+00	-1.352083	8.958333e-01	-6.458333e-01	-0.957447	-1.793617e+00	-0.612766	-0.123404	-0.846809	-2.721277e+00
1979	1.555102	-0.285417	8.645833e-01	-1.852083	-1.004167e+00	8.541667e-01	-0.357447	-9.936170e-01	-0.812766	-0.523404	-0.346809	1.478723e+00
1980	-2.144898	-1.885417	-1.435417e+00	0.747917	9.958333e-01	-9.458333e-01	-0.457447	8.063830e-01	1.487234	0.576596	0.353191	2.078723e+00
1981	3.055102	1.414583	-1.135417e+00	1.347917	-4.041667e-01	-6.458333e-01	-0.457447	-8.936170e-01	0.987234	0.176596	-0.846809	-9.212766e-01
1982	0.955102	-0.185417	1.064583e+00	0.147917	-4.041667e-01	-1.345833e+00	-1.257447	2.106383e+00	-0.712766	-0.623404	2.853191	-8.212766e-01
1983	-2.244898	2.314583	-3.541667e-02	-2.252083	-3.041667e-01	-1.045833e+00	-1.057447	-9.361702e-02	-0.912766	-2.123404	-1.746809	-3.212766e-01
1984	-2.544898	-0.585417	-2.035417e+00	-1.052083	6.958333e-01	1.541667e-01	-1.457447	-1.936170e-01	-2.112766	-0.123404	0.153191	-1.121277e+00
1985	-2.644898	-1.785417	1.764583e+00	1.347917	3.958333e-01	-7.458333e-01	-0.857447	-6.936170e-01	-1.612766	0.176596	-0.946809	-9.212766e-01
1986	-3.344898	-0.485417	9.645833e-01	-0.952083	-9.041667e-01	-1.145833e+00	-1.057447	-9.361702e-02	-1.012766	-1.923404	-1.646809	-2.021277e+00
1987	-2.044898	-0.485417	-2.835417e+00	0.247917	-5.041667e-01	1.541667e-01	-0.457447	-5.936170e-01	0.787234	-0.923404	-0.046809	-2.021277e+00
1988	1.655102	-1.985417	5.645833e-01	0.947917	7.958333e-01	4.541667e-01	0.542553	1.063830e-01	0.887234	0.876596	-0.846809	1.787234e-01
1989	-0.744898	0.514583	1.645833e-01	0.547917	4.958333e-01	-7.458333e-01	-0.757447	-1.693617e+00	-0.512766	-0.723404	-0.346809	7.872340e-02
1990	-0.444898	-0.985417	3.645833e-01	-0.152083	-4.166667e-03	-1.458333e-01	0.542553	-1.193617e+00	-0.212766	0.476596	0.353191	1.578723e+00
1991	-0.444898	-0.685417	-2.354167e-01	-0.552083	-6.041667e-01	1.154167e+00	-0.157447	-4.936170e-01	-0.712766	1.276596	-0.346809	-1.421277e+00
1992	-4.044898	-1.785417	2.645833e-01	0.547917	-1.041667e-01	-2.458333e-01	0.242553	-1.093617e+00	-2.712766	-0.423404	-1.646809	7.872340e-02
1993	0.855102	0.314583	-8.354167e-01	2.247917	1.958333e-01	-6.458333e-01	0.642553	1.806383e+00	-0.412766	-0.723404	-1.746809	-1.421277e+00
1994	-1.344898	-0.585417	-1.235417e+00	0.147917	-4.041667e-01	1.541667e-01	1.242553	-6.936170e-01	-1.712766	0.576596	-1.246809	2.578723e+00
1995	0.355102	0.314583	-1.535417e+00	-2.752083	-1.104167e+00	-8.458333e-01	-1.357447	1.406383e+00	-0.812766	-1.223404	-0.446809	-3.821277e+00
1996	-1.044898	-2.185417	-1.354167e-01	-2.352083	-5.041667e-01	4.541667e-01	-0.657447	-2.936170e-01	-0.412766	0.176596	-1.046809	-2.121277e+00

Figure 2.3 the figure above is the dataset for maximum temperature after Deseasonalization.

Part 3 Summary of exploratory data analysis and any significant conclusions

Averages in rainfall

The data is now ready to be used after it has been pre-processed and manipulated, so we immediately calculated the average rainfall for each month from 1975 to 2022. Refer to Figure 3.1.

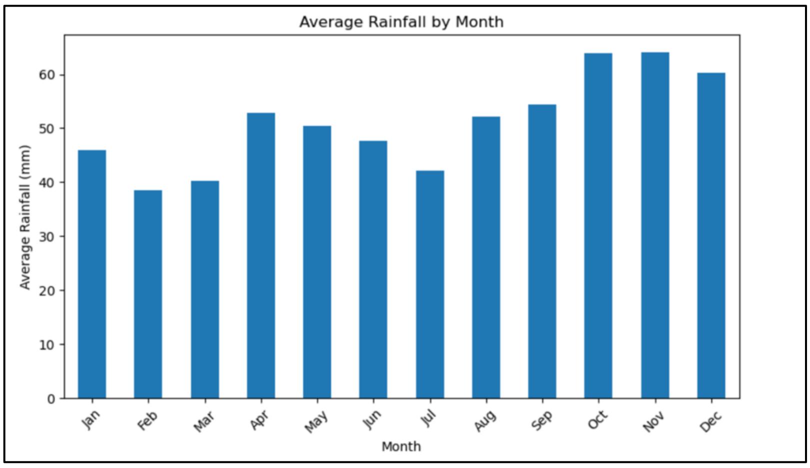


Figure 3.1 the figure showcases the average rainfall by Month.

The numbers on the left (y-axis) represents how many millimetres of rainfall that accumulates over the entire month, and the mean of these average rainfalls is approximately 51. Here, we can see that the average rainfall increases during transitional seasons (Spring and autumn) Notably, November and October boast an average of 64, while February has a significantly lower average of 36. From personal observations, the year 2023 follows these averages quite well, as we are currently experiencing more rain in spring rather than in summer, which was 8 months ago. Therefore, it is safe to say that we can extrapolate this data for years or even decades into the future ensuring a high degree of accuracy. The three main reasons for this are: seasonal changes, climate influences, and climate patterns.

1. **Seasonal changes**. October is part of the spring season, when levels of humidity are higher, accompanied by the moist air. Chance of rainfall increases during this season due to spring being in the middle of transitioning from winter to summer, alongside the fact that the weather becomes a lot more unpredictable. Conversely, February is part of the summer season, which is usually has lower humidity, warmer conditions, and drier air. With the sun being higher in the sky, evaporation increases leading to less clouds and therefore lower rainfall.

2. **Climate influences**. The interaction between the temperature of the Southern Ocean and the surrounding air can influence the formation of weather systems that bring rain. In October, the temperature of the Southern Ocean is rather low from the winter months, and when warmer air passes over the ocean during this time, rainfall is more likely. The Southern Ocean will warm up by February, reducing the contrast of temperature between the ocean and the air. This ultimately decreases the frequency of those same rain-bearing weather systems, leading to less rainfall.

3. **Climate patterns.** The climate drivers Indian Ocean Dipole (DMI) and Southern Annular Mode (SAM) may also affect the pattern of rainfall. A negative DMI, where the eastern Indian Ocean is warmer than the western Indian Ocean can cause increased rainfall due to lower temperature. Moreover, when SAM is negative, the winds expand, bringing moisture-filled air from the Southern Ocean to Melbourne, which increases the chances of rainfall.

CORRELATIONS

We created two heatmaps (figure 3.2 and 3.3), one for each month, to find the correlations between the individual climate drivers and climate variables to find any strong linear relationships.

However, correlations are either moderate or weak when comparing the climate drivers against the variables, with the highest correlations being between DMI and maximum or minimum temperature for February. This is mainly because these climate drivers are not everyday occurrences, therefore it is theoretically unlikely for them to greatly sway the general weather.

Comparing the variables against each other, we can see that the minimum temperature and maximum have a correlation of at least 0.8 for both months, making it highly correlated.

Showcasing this using a time-series graph (figure 3.4 and 3.5) allows us to be able to visualize trends more easily. Using November as an example, we can see that figure 3.4 looks very similar to figure 3.5, proving it to be consistent with the fact that the correlation between minimum and maximum temperatures are high – an increase in maximum temperature results in an increase in minimum temperature, and a decrease in maximum temperature results in a decrease in minimum temperature.

The trend for both figures suggest that both Melbourne's minimum and maximum temperatures slightly rise over time, further implying that the overall temperature has risen mainly from 1995-2022. The main reasons for this are: global warming, urbanization, climate variability, and ocean currents.

1. **Global warming**. Burning of fossil fuels, deforestation, and industrial processes increased the amount of greenhouse gasses in the atmosphere of the Earth. This enhances the greenhouse effect, capturing additional warmth resulting in global warming. Consequently, temperatures have been steadily increasing on a global scale, including Melbourne as well.

2. **Urbanization.** Urban growth in Melbourne – population increase, infrastructure development, etc – has substantially increased over the past few decades. This results in a phenomenon called 'urban heat island effect' in which the temperature of urban areas is notably higher than the rural areas surrounding them. This occurs when cities with a high amount of urban infrastructure, including concrete and asphalt, absorb and radiate heat, increasing local temperatures and further expands the disparity of temperatures between urban and rural areas. This ultimately results in both higher maximum and minimum temperatures in the city.

3. **Climate variability.** The climate diver Indian Ocean Dipole (DMI) may also play a role in the increase of temperature over time. A positive DMI is when the western Indian Ocean has a warmer sea surface temperature than the eastern Indian Ocean. As the chances of rainfall reduces greatly due to less moisture-filled air brought from the Southern Ocean, the overall number of clouds greatly decreases in Melbourne. This allows for the UV rays of the sun to shine more directly onto the area, leading to higher temperatures.

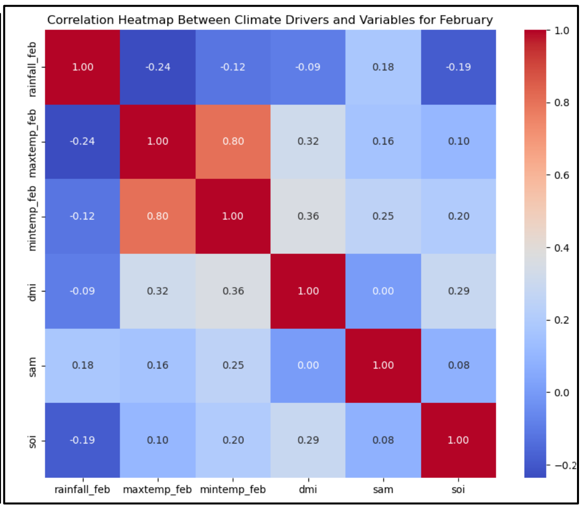


Figure 3.2 and 3.3 the two graphs show the correlations between climate drivers and variables for November and February

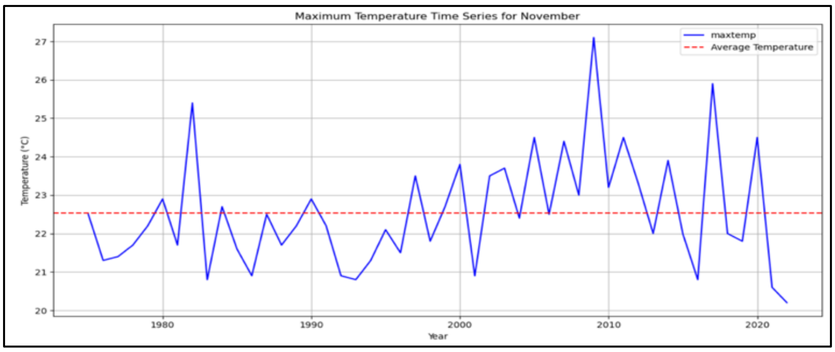


Figure 3.4 the graph shoes the maximum November temperature throughout the years.

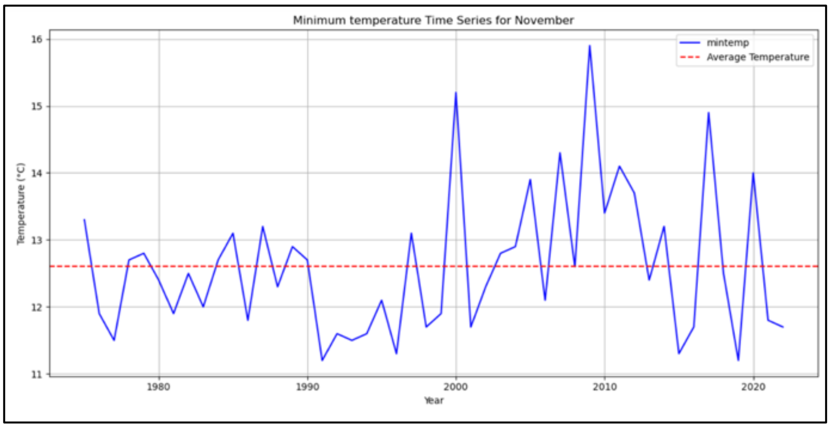


Figure 3.5 the graph shows the minimum November temperature throughout the years.

Showcasing the November and February rainfall using a time-series graph (figure 3.6 and 3.7) allows us to visualise that the rainfall has decreased in November starting from 1988. Why was there a Rainfall peak in November during 1988? La Nina during this time played a great role on the southeast half of the country from April 1988 to July 1989, where rainfall levels were the highest in the recorded data for the majority of New South Wales, South Australia, northern regions of Victoria, and certain areas in the southern and eastern parts of Queensland (Bureau of Meteorology, n.d.). There were a lot of floods during this La Nina, with the most significant being widespread flash flooding in Melbourne during November and December 1988 (refer to figure 3.8).

On the other hand, rainfall for February increased starting from 1975, having a peak during 2005, then decreased up until 2022 (Bureau of Meteorology, n.d.). The peak during 2005 was due to a particularly strong or severe low-pressure system which formed over the Eastern Bass Strait on February 2, 2005. This system was formed due to several meteorological conditions in the environment – the main one being the convergence of air masses (collision of warm and moist air from the Tasman Sea and cold air from the Southern Ocean) - which eventually led to continuous rainfall for over 30 hours.

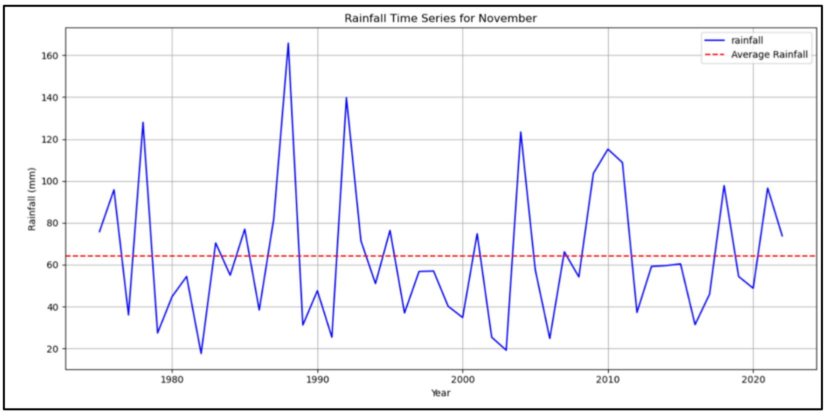


Figure 3.6 the graph shows the November rainfall throughout the years.

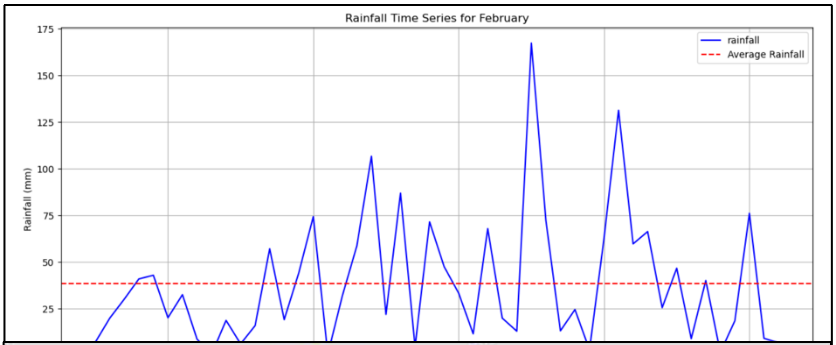


Figure 3.7 the figure shows February rainfall throughout the years.

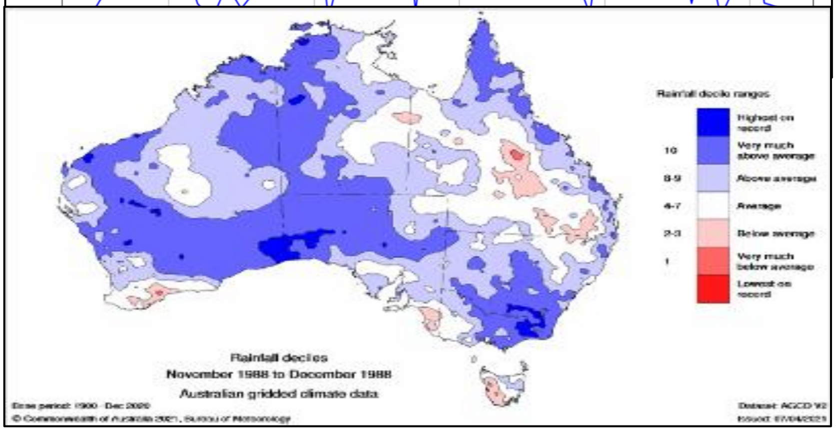


Figure 3.8 the figure shows the intensity of the rainfall in Australia during 1988 of November and December

(Bureau of Meteorology, n.d.)

DATA AFTER INDEXING

Using the previous data, we 'indexed' the data by subtracting the monthly averages for each month throughout the whole time-period. This not only aids in visualizing the dataset but also helps assess the extent of its fluctuations to be able to find the anomalies or outliers more effectively and efficiently. These outliers will then be used later during the modelling process to check whether their presence is caused by the climate drivers. Figure 4.1, 4.2 and 4.3 shows the indexed data for the climate variables from 1976 – 2023. There is also quite a bit of index variability in these 3 graphs, with rainfall having a variability of about 175, minimum temperature of about 5.5, and maximum temperature of about 8.

Some notable features in figure 4.1 are February in 2005 and December in 1993. These two bars represent the top two anomalies as both deviate the most from the average with at least 100 millimetres of rainfall. Referring to figure 3.1, it is unsurprising that December, the month with the 3rd highest rainfall, also happens to have a year where the rainfall deviates the average the most. On other hand, February – the month with the least average rainfall – in fact has a year where the rainfall also deviates significantly (the 2nd highest). As mentioned previously, this is because of the formation of a low-pressure system.

Similarly, the same reason was why December had the highest rainfall to date.

There was heavy rainfall over two days, 92.2 mm over the course of two days, totalling to 195mm of rainfall in just that month ("Record rain takes toll on sports," 1993). This surpassed the previous record which was 182mm in 1863 for any given month in a year. Although this disrupted major sports events, there was no major structural damage reported.

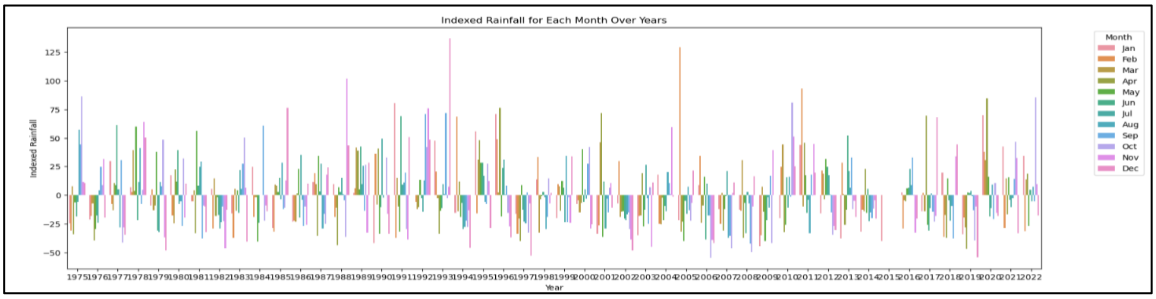


Figure 4.1 the figure shows the indexed rainfall data for each month over years.

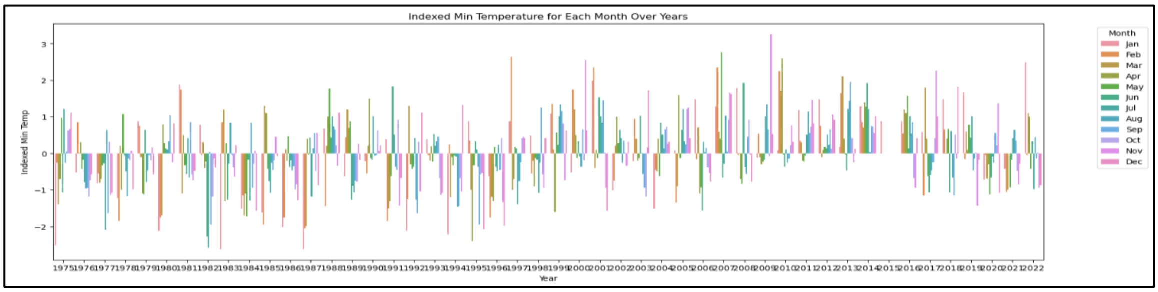


Figure 4.2 the figure shows the indexed minimum temperature data for each month over years.

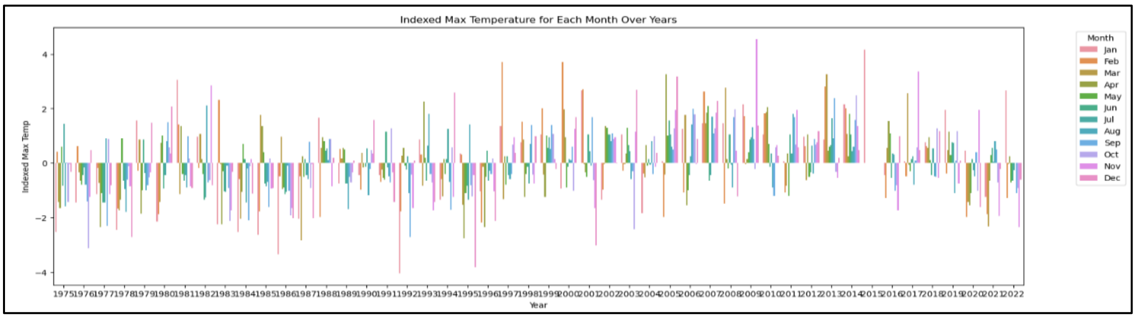


Figure 4.3 the figure shows the indexed maximum temperature data for each month over years.

Part 4 Summary of any undertaken modelling and any significant conclusions

The goal of the modelling is to see whether anomalies in Australian weather a result of the climate are drivers changes. Before modelling, we filter the rainfall, and temperature data frames to only show data where the climate is abnormal. This is done by selecting the top and bottom 20% of the indexed data, representing the anomalies in the climate. We initially started by only observing how the weather events affected Melbourne. Initially running a heat map to see the correlations between the variables: rainfall, max temperature, min temperature and the climate drivers: SAM, SOI, and DMI. Of note DMI had a correlation of 0.28 with rainfall, and 0.18 and 0.14 for SOI and SAM respectively. For Minimum temperature SOI had a correlation of 0.42 and DMI had a correlation of 0.3 for Maximum temperature. Next a single variable linear regression model was run.

When running each variable against rainfall, we see that SAM and SOI receive very low scores, and DMI returns an r squared score of 0.1. This is expected as DMI has an effect on rainfall and temperature, unlike SAM. However, SOI being so low is surprising as SOI measures El Nino, which often causes droughts. When running the models, this time measuring for minimum temperature, SOI receives a r squared score of 0.22, and the other variables do not return relatively strong scores. For Maximum temperature there are surprisingly no values that have a good r squared value, indicating that none of the relationships might be linear.

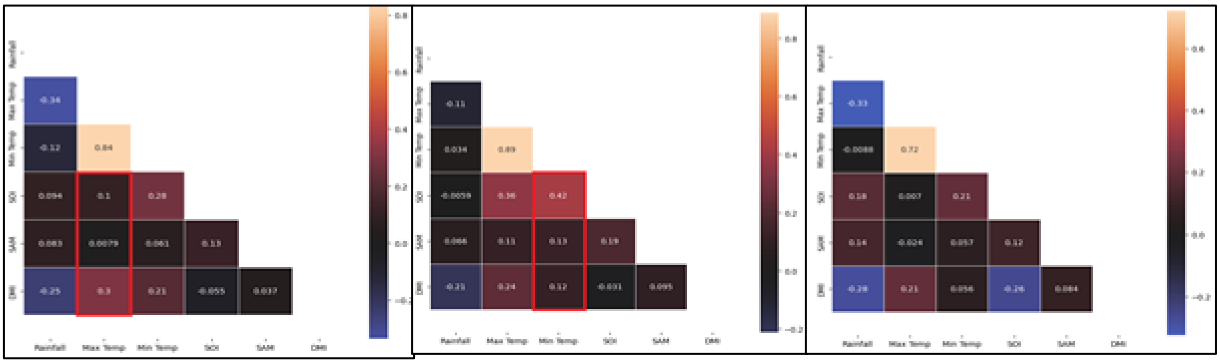


Figure 4.4 the figure shows the heat maps for the correlation of the three variables, DMI, SAM, SOI, respectively.

['DMI']

	R^2	RMSE	MAE
train	0.101001	38.408864	32.751829
test	-0.196396	37.238962	34.136029

['SAM']

	R^2	RMSE	MAE
train	0.018955	40.123266	35.666053
test	-0.116333	35.971384	33.798734

['SOI']

	R^2	RMSE	MAE
train	0.035346	39.786676	34.92758
test	-0.111939	35.900506	33.24532

['DMI']

	R^2	RMSE	MAE
train	0.008824	1.555354	1.492816
test	0.025944	1.547257	1.450557

['SAM']

	R^2	RMSE	MAE
train	0.031192	1.537704	1.458934
test	-0.050724	1.606996	1.496340

['SOI']

	R^2	RMSE	MAE
train	0.219692	1.380025	1.209884
test	-0.021891	1.584793	1.370200

['DMI']

	R^2	RMSE	MAE
train	0.072080	1.617336	1.419626
test	0.076379	1.732184	1.589069

['SAM']

	R^2	RMSE	MAE
train	0.005376	1.674458	1.513132
test	-0.127969	1.914237	1.788611

['SOI']

	R^2	RMSE	MAE
train	0.004655	1.675065	1.526206
test	-0.054695	1.851017	1.737466

Figure 4.5 shows the single variable linear regression model for each variable against, Rainfall, Min temp, Max temp.

From this a multivariable regression was run, to see if all the climate drivers together could determine whether Melbourne's climate outliers are caused by the events. For Rainfall the model receives an r squared score of 0.135, and a low testing score of -0.16. This is an improvement from the best single variable model, however, as seen in the graph, a few extreme outliers in the training data, has likely skewed the data and resulted in a low testing score. For minimum temperature, a r squared score 0.232, and for maximum temperature 0.087. These r squared scores are relatively low, and is likely because of a couple reasons. Firstly, the weather is very complicated and much more complex models are used to predict the weather. Secondly, the climate drivers can affect the rainfall and temperature differently based on seasonality and where in Australia you are.

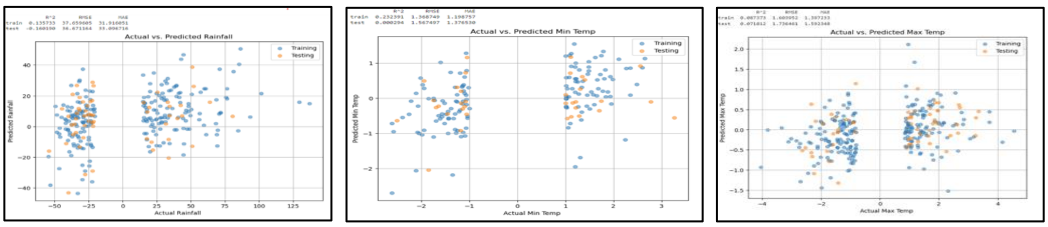


Figure 4.6 shows scatterplot modelling graphs of the actual versus the predicted of the three variables.

To improve this multivariate model, we can separate the data into seasons to see if the climate drivers have more of an impact on the weather in Melbourne during a certain season. As seen in the table, during the seasons of Winter and Spring, the models receive much higher r squared scores then the non-seasonal regression. From this we can conclude that during the seasons of Spring and Winter, the climate drivers of DMI, SAM and SOI are likely to cause anomalies in Melbourne's weather. For minimum temperature all seasons receive much higher r squared scores, except for autumn. For maximum temperatures, whose models have previously performed very poorly, we see an increase in the scores. This coupled with minimum temperature indicates that temperature outliers in Melbourne are likely to occur based on seasonality. From all these models it highlights that many of the climate drivers will affect Melbourne differently based on the season and time of year, and therefore this model is far better to use than the other models. To conclude Melbourne's modelling, DMI has an influence on Rainfall, which was expected from prior research, whereas Temperature is influenced by all the climate variables but on a seasonal basis.

```
['Dec', 'Jan', 'Feb']
Performance of predicting: Rainfall
R^2 (Train): 0.0309
R^2 (Test): 0.0493
RMSE (Train): 45.3179
RMSE (Test): 34.3807
MAE (Train): 39.7870
MAE (Test): 31.5417

['Mar', 'Apr', 'May']
Performance of predicting: Rainfall
R^2 (Train): 0.0688
R^2 (Test): -0.0962
RMSE (Train): 37.4679
RMSE (Test): 31.6081
MAE (Train): 33.2198
MAE (Test): 28.6466

['Jun', 'Jul', 'Aug']
Performance of predicting: Rainfall
R^2 (Train): 0.2326
R^2 (Test): -0.1674
RMSE (Train): 26.6481
RMSE (Test): 31.1576
MAE (Train): 21.5874
MAE (Test): 28.8547

['Sep', 'Oct', 'Nov']
Performance of predicting: Rainfall
R^2 (Train): 0.2515
R^2 (Test): 0.4137
RMSE (Train): 36.9987
RMSE (Test): 27.9033
MAE (Train): 32.6570
MAE (Test): 25.9284
```

```
['Dec', 'Jan', 'Feb']
Performance of predicting: Min Temp
R^2 (Train): 0.4329
R^2 (Test): 0.2083
RMSE (Train): 1.3093
RMSE (Test): 1.3493
MAE (Train): 1.0527
MAE (Test): 1.1250

['Mar', 'Apr', 'May']
Performance of predicting: Min Temp
R^2 (Train): 0.1839
R^2 (Test): -1.5802
RMSE (Train): 1.3455
RMSE (Test): 2.2285
MAE (Train): 1.2197
MAE (Test): 1.6961

['Jun', 'Jul', 'Aug']
Performance of predicting: Min Temp
R^2 (Train): 0.3109
R^2 (Test): 0.1004
RMSE (Train): 1.1202
RMSE (Test): 1.3454
MAE (Train): 0.9179
MAE (Test): 1.1872

['Sep', 'Oct', 'Nov']
Performance of predicting: Min Temp
R^2 (Train): 0.4056
R^2 (Test): 0.4759
RMSE (Train): 1.1980
RMSE (Test): 1.1969
MAE (Train): 0.8739
MAE (Test): 1.0693
```

```
['Dec', 'Jan', 'Feb']
Performance of predicting: Max Temp
R^2 (Train): 0.2698
R^2 (Test): -0.1722
RMSE (Train): 1.1563
RMSE (Test): 1.4535
MAE (Train): 0.9936
MAE (Test): 1.1246

['Mar', 'Apr', 'May']
Performance of predicting: Max Temp
R^2 (Train): 0.1812
R^2 (Test): -0.3515
RMSE (Train): 1.0214
RMSE (Test): 1.1253
MAE (Train): 0.8746
MAE (Test): 0.9136

['Jun', 'Jul', 'Aug']
Performance of predicting: Max Temp
R^2 (Train): 0.0681
R^2 (Test): -0.0229
RMSE (Train): 1.0827
RMSE (Test): 0.8247
MAE (Train): 0.9378
MAE (Test): 0.6561

['Sep', 'Oct', 'Nov']
Performance of predicting: Max Temp
R^2 (Train): 0.1497
R^2 (Test): 0.0083
RMSE (Train): 0.9469
RMSE (Test): 0.8452
MAE (Train): 0.7616
MAE (Test): 0.6852
```

Figure 4.7 shows the performance of predicted variables when split into seasons.

However, different places around Australia may be influenced differently by the climate drivers. The first place we explored was Brisbane. The R^2 values for all climate variables are relatively low, indicating that the climate drivers for predicting rainfall and minimum temperature, much like Melbourne are very low, however SAM and SOI have significantly higher r squared scores for maximum temperature compared to Melbourne. This indicates that Maximum Temperature may be influenced more by the climate drivers than Melbourne. RMSE values indicate the average prediction error, with the RMSE being higher for Rainfall than Max Temp and Min Temp. This suggests that the model's accuracy is lower for Rainfall. MAE values provide information on the average magnitude of prediction errors. Similar to RMSE, the MAE is highest for Rainfall, indicating that Rainfall predictions have a more significant average absolute error. The scatter plots visually show no clear linear relationship between the model's predictions and the actual values, confirming the limited effectiveness of the linear regression model in capturing the complex relationships between climate drivers and climate variables. In summary, using these climate drivers, the linear regression model demonstrates limited effectiveness in predicting Rainfall, Max Temp, and Min Temp in Brisbane. The visual representation in the scatter plots further emphasises the absence of a linear relationship between the predictions and the actual values, suggesting that more sophisticated modelling approaches or consideration of additional factors may be necessary to improve predictive accuracy.

Running the multivariable model for Brisbane's rainfall, reveals much like Melbourne that the climate drivers collectively do not have a linear relationship with abnormal rainfall. Minimum temperature, unlike Melbourne, does not perform too well on the multivariable model. This is probably due to the hot and humid climate of Brisbane, leading to a stagnant minimum temperature that is not as influenced by the climate drivers. However, the maximum temperature model performs much better, with an r squared score of 0.3, indicating that the climate drivers together are likely to be the cause of maximum temperature anomalies.

```
['DMI']
            R^2       RMSE        MAE
train  0.010863  63.304273  54.042206
test  -0.003687  82.266954  62.792276

['SAM']
            R^2       RMSE        MAE
train  0.002574  63.568947  54.612044
test   0.005608  81.885127  62.248492

['SOI']
            R^2       RMSE        MAE
train  0.035944  62.496535  52.037733
test   0.065398  79.385224  60.333275
```

```
['DMI']
            R^2      RMSE       MAE
train  0.028455  1.899276  1.783881
test  -0.149872  2.120253  2.004095

['SAM']
            R^2      RMSE       MAE
train  0.018715  1.908771  1.792623
test  -0.115867  2.088666  1.966002

['SOI']
            R^2      RMSE       MAE
train  0.051451  1.876663  1.730023
test  -0.063691  2.039251  1.868642
```

```
['DMI']
            R^2      RMSE       MAE
train  0.096873  1.925232  1.713852
test   0.185259  1.892687  1.649998

['SAM']
            R^2      RMSE       MAE
train  0.140895  1.877723  1.677262
test   0.007299  2.089191  1.899405

['SOI']
            R^2      RMSE       MAE
train  0.118281  1.902277  1.667134
test   0.175956  1.903461  1.670036
```

Figure 4.8 shows the multivariable model for Brisbane's rainfall.

```
            R^2      RMSE       MAE
train  0.297033  1.698541  1.466085
test   0.271243  1.790029  1.465867
```

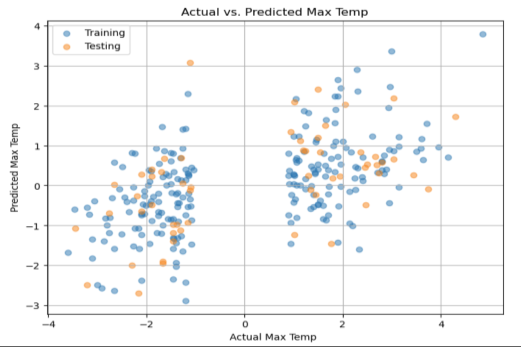


Figure 4.9 this figure shows the actual vs predicted max temp in the form of a scatterplot.

```
['Dec', 'Jan', 'Feb']
Performance of predicting: Rainfall
R^2 (Train): 0.0777
R^2 (Test): -0.1458
RMSE (Train): 65.4996
RMSE (Test): 90.3532
MAE (Train): 56.0064
MAE (Test): 75.2312

['Mar', 'Apr', 'May']
Performance of predicting: Rainfall
R^2 (Train): 0.0411
R^2 (Test): 0.0013
RMSE (Train): 62.6694
RMSE (Test): 98.1911
MAE (Train): 51.0076
MAE (Test): 58.2680

['Jun', 'Jul', 'Aug']
Performance of predicting: Rainfall
R^2 (Train): 0.2056
R^2 (Test): -2.0880
RMSE (Train): 33.4163
RMSE (Test): 29.6286
MAE (Train): 23.2833
MAE (Test): 25.6941

['Sep', 'Oct', 'Nov']
Performance of predicting: Rainfall
R^2 (Train): 0.2385
R^2 (Test): 0.0773
RMSE (Train): 48.2202
RMSE (Test): 80.9522
MAE (Train): 41.3412
MAE (Test): 61.3297
```

```
['Dec', 'Jan', 'Feb']
Performance of predicting: Min Temp
R^2 (Train): 0.2581
R^2 (Test): -0.6267
RMSE (Train): 1.5486
RMSE (Test): 1.5330
MAE (Train): 1.3031
MAE (Test): 1.3619

['Mar', 'Apr', 'May']
Performance of predicting: Min Temp
R^2 (Train): 0.0398
R^2 (Test): -0.1792
RMSE (Train): 1.9979
RMSE (Test): 2.1635
MAE (Train): 1.8499
MAE (Test): 1.9533

['Jun', 'Jul', 'Aug']
Performance of predicting: Min Temp
R^2 (Train): 0.2213
R^2 (Test): 0.4018
RMSE (Train): 1.8024
RMSE (Test): 1.6220
MAE (Train): 1.5104
MAE (Test): 1.4811

['Sep', 'Oct', 'Nov']
Performance of predicting: Min Temp
R^2 (Train): 0.0932
R^2 (Test): 0.2535
RMSE (Train): 1.6687
RMSE (Test): 1.8981
MAE (Train): 1.5220
MAE (Test): 1.8243
```

```
['Dec', 'Jan', 'Feb']
Performance of predicting: Max Temp
R^2 (Train): 0.2092
R^2 (Test): -0.1194
RMSE (Train): 0.8548
RMSE (Test): 1.3876
MAE (Train): 0.7087
MAE (Test): 1.2111

['Mar', 'Apr', 'May']
Performance of predicting: Max Temp
R^2 (Train): 0.0447
R^2 (Test): 0.1309
RMSE (Train): 1.3811
RMSE (Test): 1.0926
MAE (Train): 1.0530
MAE (Test): 0.8283

['Jun', 'Jul', 'Aug']
Performance of predicting: Max Temp
R^2 (Train): 0.1142
R^2 (Test): 0.0609
RMSE (Train): 1.2186
RMSE (Test): 1.0828
MAE (Train): 0.8740
MAE (Test): 0.8017

['Sep', 'Oct', 'Nov']
Performance of predicting: Max Temp
R^2 (Train): 0.2026
R^2 (Test): 0.0595
RMSE (Train): 1.0496
RMSE (Test): 1.1031
MAE (Train): 0.8310
MAE (Test): 0.7872
```

Figure 5.0 this figure shows the performance of predicting max temp. Including the R^2 score, RMSE and MAE.

The analysis of Brisbane's weather data, considering key climate drivers including the Southern Oscillation Index (SOI), the Southern Annular Mode (SAM), and the Indian Ocean Dipole (DMI), reveals intriguing insights into how these factors impact the city's weather. Brisbane's climate can be categorised into four seasons: Summer, Autumn, Winter, and

Spring. During the summer months (December-February), Brisbane experiences hot and rainy conditions. The analysis suggests that SOI has a limited impact on rainfall but has a slightly positive correlation with maximum temperatures. SAM does not strongly influence Brisbane's summer weather, and DMI shows a negligible correlation with the city's climate during this season. In autumn (March-May), SOI continues to have a limited impact on rainfall, while SAM exhibits only a slight negative correlation with maximum temperatures. DMI shows minimal influence on weather conditions during this season. The drivers are considered to have a relatively minor effect on Brisbane's autumn climate. The winter season (June-August) is characterised by mild and dry weather. The analysis reveals that SOI has a minimal impact on rainfall during winter but a slightly positive correlation with both maximum and minimum temperatures. SAM exhibits a slight negative correlation with maximum temperatures during this season. DMI also has a negligible effect on Brisbane's winter climate. Spring (September-November) in Brisbane witnesses mild and pleasant conditions. SOI has limited influence on rainfall but a slightly positive correlation with maximum temperatures. SAM shows a moderate negative correlation with maximum temperatures. DMI exhibits a negligible impact on Brisbane's spring weather.

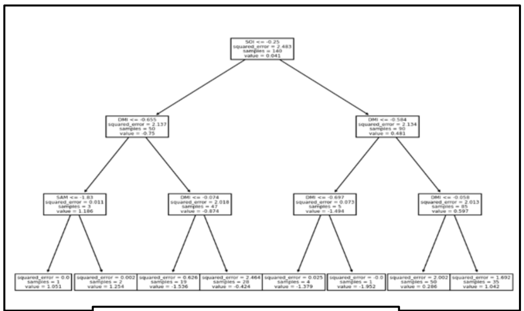
Furthermore, we can look at Darwin to compare how the climate drivers may affect a different climate of Australia. Notable things seen from Darwin is that when looking at rainfall seasonally, we see that during Summer and Spring the model performs quite well. But notably an r^2 score of 1 is given for the train and -170 is given for the test. This is quite extreme, and upon further analysis this occurred because that in Darwin during this season there were only 3 rainfall anomalies, meaning that the model cannot perform properly due to a low data set size. Also, the models for rainfall anomalies perform very well, with an r^2 scores of 0.3, 0.4, and 0.5 for the seasons. The difference between what the climate drivers are influencing depending on where you are and the season, demonstrates that the best model to run to predict outliers in the climate is dependent on location and the season.

```
['Jun', 'Jul', 'Aug']
Performance of predicting: Rainfall
R^2 (Train): 1.0000
R^2 (Test): -170.6127
RMSE (Train): 0.0000
RMSE (Test): 95.6307
MAE (Train): 0.0000
MAE (Test): 80.9338
```

```
['Dec', 'Jan', 'Feb']
Performance of predicting: Min Temp
R^2 (Train): 0.2974
R^2 (Test): 0.0298
RMSE (Train): 1.0526
RMSE (Test): 0.9976
MAE (Train): 0.8777
MAE (Test): 0.9234

['Mar', 'Apr', 'May']
Performance of predicting: Min Temp
R^2 (Train): 0.0418
R^2 (Test): 0.1981
RMSE (Train): 1.4666
RMSE (Test): 1.0532
MAE (Train): 1.3648
MAE (Test): 0.9707

['Jun', 'Jul', 'Aug']
Performance of predicting: Min Temp
R^2 (Train): 0.4294
R^2 (Test): 0.1989
RMSE (Train): 1.2026
RMSE (Test): 1.3232
MAE (Train): 0.9533
MAE (Test): 1.1204

['Sep', 'Oct', 'Nov']
Performance of predicting: Min Temp
R^2 (Train): 0.5321
R^2 (Test): 0.1946
RMSE (Train): 0.9707
RMSE (Test): 1.0108
MAE (Train): 0.8246
MAE (Test): 0.8906
```

Figure 5.1 and 5.2 these two figures show the performance of predicting Rainfall and the performance of predicting Min temp divided into seasons.

Many of the relationships are not linear, as seen with the scatterplots and the low r squared scores for many of the relationships. To see if a better model can be achieved, decision trees and forest regression can be used. Many of the models did not perform well when running the decision trees, but some did. Notable trees are Melbourne's Forest of predicting minimum temperature, getting an r^2 scores of 0.27.



```
0.27247647084817095
```

Figure 5.3 this figure shows the decision tree of predicting minimum temperature and Melbourne's Forest score juxtaposed.

From the tree it is highlighted that DMI and SOI have the biggest influence on the minimum temperature of Melbourne. The tree highlights that if DMI and SOI are negative it will likely result in lower-than-average minimum temperatures. Looking at the model, but seasonality, due to similar r^2 scores in all the seasons it is likely that minimum temperature is influenced all year round.

```
['Dec', 'Jan', 'Feb']
0.2731816288183595

['Mar', 'Apr', 'May']
0.291621872874133

['Jun', 'Jul', 'Aug']
0.2763727474401495

['Sep', 'Oct', 'Nov']
0.2789212437387858
```

Figure 5.4 shows the r^2 scores for minimum temperature seasonally.

Comparing this to the linear model which got an r^2 score of 0.23, the non-linear model performs much better. We see a similar improvement in Brisbane's model for maximum temperature seasonally getting an improved r^2 score for the months. Similar results are also seen for Darwin, receiving similar results for the non-seasonal regression, but improved r^2 score for the seasonal. Overall, the use of decision trees allows for an improved model for predicting anomalies in the climate seasonally, caused by the climate drivers.

```
['Dec', 'Jan', 'Feb']
0.19320214281835446

['Mar', 'Apr', 'May']
0.21040223460863172

['Jun', 'Jul', 'Aug']
0.2031252191286943

['Sep', 'Oct', 'Nov']
0.21308417591586593
```

```
['Dec', 'Jan', 'Feb']
0.29107820599995127

['Mar', 'Apr', 'May']
0.28502295215522

['Jun', 'Jul', 'Aug']
0.2828532987671051

['Sep', 'Oct', 'Nov']
0.29105893607718947
```

Figure 5.5 and 5.6 shows the r^2 scores for Darwin min temp and Brisbane max temp seasonally, predicting the magnitude of anomalies, respectively.

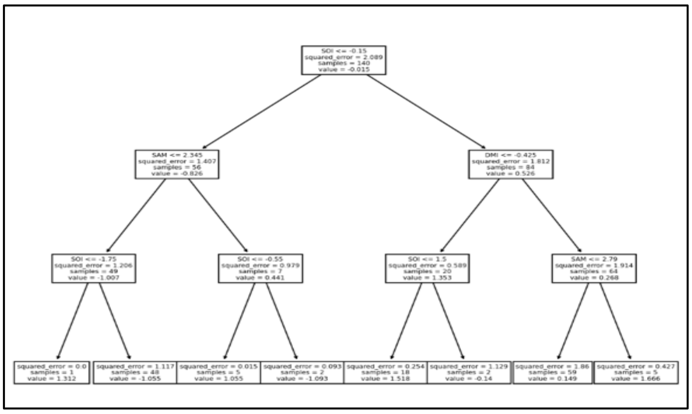


Figure 5.7 shows the decision tree for Darwin min temp.

Classification Modelling

Since none of the previous models had very high accuracy, we determined that the climate indicators may cause anomalies in the climate to occur, but not have a direct impact on the degree of the anomaly, leading to a low accuracy when using the model. From this we created a new model with dummy variables of -1, 0, and 1 for rainfall, minimum temperature, and maximum temperature. -1 represents the anomalies that are below the normal climate, where the rainfall or temperature is one standard deviation below the mean. 1 represents the anomalies above the normal climate, where the rainfall or temperature is one standard deviation above the mean. 0 represents all the other values where the climate is normal (within one standard deviation). (Zaveri et al., 2020). Decision tree classification was then used on this data to determine the influence of the climate drivers. Running these classifications, we can see a large increase in accuracy compared to all previous models. The results shown here are for Melbourne. For rainfall there is an accuracy of 0.69. The flow diagram shows that the number of samples with low anomaly rainfall (-1) is highest when DMI is more than -0.284 and SOI is less than -2.3. The number of samples with normal rainfall (0) is highest when DMI is between -0.284 and -0.103 and SOI is less than -2.3. The number of samples with high anomaly rainfall (1) is highest when DMI is greater than -0.103 and SOI is greater than or equal to -2.3. Similar results are seen with minimum and maximum temperature, with an accuracy of 0.72 and 0.69 being achieved respectively. SOI is the most important climate driver for all three variables, followed by DMI and then SAM. The SOI measures El Nino and La Nina conditions, which are known to influence droughts/floods and heatwave risk in Victoria. We can also see that this model when split up into seasons performs better. This is highlighted with Rainfall in winter receiving a 0.86 accuracy. From the seasons we can see that the climate drivers during the seasons will affect the climate of Melbourne differently based on the time of year. These seasons show that during certain seasons, such as winter and summer, the climate drivers are more likely to be the cause of climate anomalies. Overall, the climate driver of SOI has the biggest impact on the weather anomalies in Melbourne. This analysis demonstrates that the climate drivers are more able to predict the dryer/wetter or warmer/colder states, rather than the actual magnitude of the anomalies. This is likely because the weather is very complex and more than just these three climate drivers are used to predict the weather and therefore, the climate drivers are likely to cause anomalies in the weather to occur, but it can’t be used to predict the exact temperature or rainfall of the anomalies due to the complex nature of the weather.

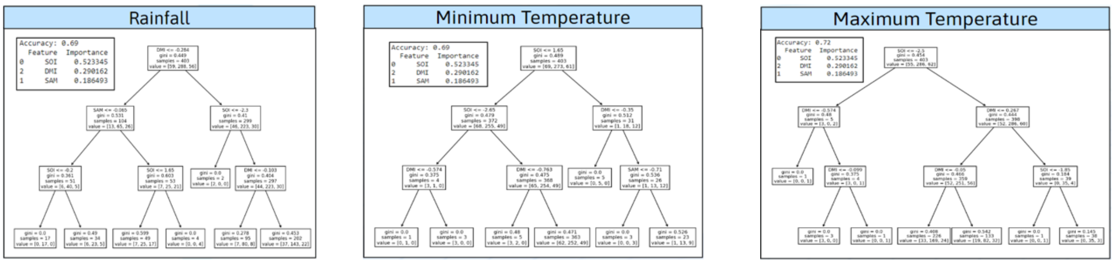


Figure 5.8 shows the decision tree for the three variables Rainfall, Min temp and Max temp and their respective accuracy scores as well as feature of importance.

```
Months: ['Dec', 'Jan', 'Feb']
Accuracy: 0.57
Feature Importance:
  Feature  Importance
1     SAM    0.374272
2     DMI    0.334455
0     SOI    0.291273

Months: ['Mar', 'Apr', 'May']
Accuracy: 0.64
Feature Importance:
  Feature  Importance
1     SAM    0.374752
2     DMI    0.341882
0     SOI    0.283367

Months: ['Jun', 'Jul', 'Aug']
Accuracy: 0.86
Feature Importance:
  Feature  Importance
1     SAM    0.462892
0     SOI    0.279569
2     DMI    0.257540

Months: ['Sep', 'Oct', 'Nov']
Accuracy: 0.59
Feature Importance:
  Feature  Importance
2     DMI    0.345389
1     SAM    0.329798
0     SOI    0.324813
```

```
Months: ['Dec', 'Jan', 'Feb']
Accuracy: 0.68
Feature Importance:
  Feature  Importance
1     SAM    0.379727
2     DMI    0.312443
0     SOI    0.307830

Months: ['Mar', 'Apr', 'May']
Accuracy: 0.73
Feature Importance:
  Feature  Importance
0     SOI    0.362322
1     SAM    0.319226
2     DMI    0.318452

Months: ['Jun', 'Jul', 'Aug']
Accuracy: 0.68
Feature Importance:
  Feature  Importance
1     SAM    0.417132
2     DMI    0.304878
0     SOI    0.277990

Months: ['Sep', 'Oct', 'Nov']
Accuracy: 0.73
Feature Importance:
  Feature  Importance
1     SAM    0.391541
0     SOI    0.326967
2     DMI    0.281491
```

```
Months: ['Dec', 'Jan', 'Feb']
Accuracy: 0.59
Feature Importance:
  Feature  Importance
1     SAM    0.348831
0     SOI    0.333725
2     DMI    0.317444

Months: ['Mar', 'Apr', 'May']
Accuracy: 0.73
Feature Importance:
  Feature  Importance
1     SAM    0.365416
2     DMI    0.320938
0     SOI    0.313646

Months: ['Jun', 'Jul', 'Aug']
Accuracy: 0.82
Feature Importance:
  Feature  Importance
1     SAM    0.423812
0     SOI    0.368721
2     DMI    0.207467

Months: ['Sep', 'Oct', 'Nov']
Accuracy: 0.70
Feature Importance:
  Feature  Importance
2     DMI    0.387908
1     SAM    0.331489
0     SOI    0.280603
```

Figure 5.9 shows the accuracy scores and feature of importance for Rainfall, Min temp and Max temp, respectively.

Part 5 Conclusions in the relation to the original problem

In conclusion, our comprehensive project, "Unravelling the Climate Tapestry," has offered nuanced insights into the complex interplay between major climate drivers and their broad-ranging impacts on Australia's diverse meteorological patterns. From meticulous data pre-processing to innovative analytical techniques, our journey underscores the multifaceted nature of climate behaviour and its sensitivity to both global phenomena and localized influences.

Our initial foray into linear regression models revealed the inadequacy of linear methods in capturing the dynamism of weather patterns, influenced by climate drivers like the Southern Oscillation Index (SOI), the Dipole Mode Index (DMI), and the Southern Annular Mode (SAM). This realization prompted a strategic pivot toward more complex, non-linear models, including decision trees and forest regression techniques. However, even these sophisticated approaches encountered limitations in predicting the precise magnitude of climate anomalies, reflecting the intricate and somewhat elusive relationships between climatic factors and weather outcomes.

The transformative moment arrived with our adaptation to classification modelling. By re-conceptualizing our target variables into categorical terms representing normal, lower, or higher than average climatic conditions, we unveiled the more influential role of climate drivers. This approach dramatically enhanced model accuracy, particularly in discerning broader climatic trends rather than exact numerical forecasts. Notably, the classification models identified SOI as a predominant influence, affirming its criticality in understanding Australia's susceptibility to extreme weather events associated with El Niño and La Niña episodes.

This journey through various analytical realms underscores a vital lesson: the sheer unpredictability inherent in our climate system. Rather than seeking absolute predictions, our research suggests the greater importance of understanding climate extremes' directional tendencies, influenced by key climatic drivers. Such knowledge is invaluable in strategic planning, from urban development adapting to rising temperatures to disaster readiness in the face of increasing weather anomalies.

Our findings advocate for a continued evolution in our analytical methodologies, favouring flexibility and adaptiveness in model selection and application. Future research endeavours would do well to integrate these learnings with emerging technologies and predictive techniques, possibly harnessing the power of artificial intelligence and deep learning. This continuous advancement is not just about keeping pace with climate change but also about proactively shaping our response strategies, ensuring that societies are not mere bystanders but active, informed participants in combating and adapting to our ever-changing environment.

Bibliography (if needed)

1. Bureau of Meteorology. (n.d). Australian rainfall during El Nino and La Nina events. Retrieved from http://www.bom.gov.au/climate/history/enso/
2. Bureau of Meteorology. (n.d). Severe Weather Event Victoria, 2-3 February 2005. Retrieved from **http://www.bom.gov.au/weather-services/severe-weather-knowledge-centre/thunderstorm-reports/Severe_Weather_Event_Victoria_2-3_Feb_2005.pdf#:~:text=The%20system%20brought%20abnormally%20low%20temperatures%20and%20severe,most%20significant%20feature%20produced%20by%20the%20low-pressure%20system.**
3. Record rain takes toll on sports. (1993, December 29). *The Canberra Times*, pp. 15–16. **https://trove.nla.gov.au/newspaper/article/127257202**
4. Zaveri, E., Damania, R., & Russ, J. (2020). Rainfall anomalies are a significant driver of cropland ... - PNAS. https://www.pnas.org/doi/abs/10.1073/pnas.1910719117