Analysis of Time series and Logistic regression Models

Bolormaa Mendbayar- x23176725

National College of Ireland

MSc in Data Analytics

Statistics for Data Analytics

TABA- Semester 2023/2024

***Abstract*—1. Time Series: The examination of past weather observations from Dublin Airport, with a focus on the variable of grass minimum temperature, is presented in this report. This measure, which represents the temperature below ground, is important for weather-related operations and agricultural planning. Visualizations were used in the initial data exploration phase to identify trends in the ground-level cold across time. By using historical data to estimate future temperatures, a variety of forecasting models were used, allowing the most accurate prediction technique to be determined. Grass Minimum Temperature is important because it helps determine when to plant crops and provides information about how plants grow. Making educated decisions about weather-sensitive activities is made easier when one is aware of these temperature trends.**

2. Logistic regression: The goal of this project is to create a predictive model that calculates the probability of heart problems using logistic regression. The model-building process involves careful selection and evaluation, emphasizing the relationships between attributes and the occurrence of cardiac conditions. For healthcare applications, heart state prediction is essential since it enables early detection and intervention. This estimate helps with specific treatment plans, proactive healthcare planning, and risk-reduction activities to lower the probability of heart conditions. Healthcare practitioners can benefit greatly from understanding and forecasting cardiac problems based on participant data. This can help them provide enhanced care for patients and make smarter choices that will enhance the health of patients.

I. INTRODUCTION

1. Time Series: This research examines the meteorological data that Dublin Airport has gathered over many years. The dataset contains 29889 rows and 9 columns, information about the temperature, amount of precipitation, and other weather-related aspects. We are studying one specific area—[specified variable, such as the grass minimum temperature (C)].

Why do this action? Numerous tasks, such as determining when to plant crops or how cities should adapt to weather changes, depend on an understanding of how the weather varies over time. Our objective is to identify trends in the Grass minimum temperature and apply statistical analysis to create informed predictions regarding its future evolution.

The results of this study have the potential to improve our ability to forecast the weather, which might have a huge impact on a variety of daily activities.

First, we will check out the data to see its patterns using plots and graphs, because want to identify any consistent changes that occur over time. Next three categories explore in detail:

- **Simple Time Series Models**: By identifying the fundamental trends and patterns observed in the data, these simple models assist in providing the foundation.
  1. Average method: In this case, all future values are predicted to be equal to the historical data's average, or "mean."
  2. Naïve method (Random Walk) involves setting all forecasts to equal the value of the most recent observation.
  3. Seasonal Naïve method: For highly seasonal data, a similar strategy works well. In this instance, we set each forecast to match the most recent figure that was observed during the same year's season.
  4. Drift method: An alternative approach to the Naïve method involves allowing the forecasts to fluctuate over time. This fluctuation is referred to as the drift, and it is determined by averaging the changes observed in the previous data.
- **Exponential smoothing**: This approach uses various weights to previous observations in order to observe short-term fluctuations and trends.
  1. A simple exponential smoothing model:

     Suits a time series without a trend or seasonal component, but with a constant level and an irregular component at time. Only level is used as a smoothing parameter.
  2. A Holt exponential model: Fits level and a trend for a time series
  3. A Holt-Winters exponential smoothing:

     Adjusts for level, trend, and seasonal components in a time series.
- **Models ARIMA/SARIMA**: These complex models attempt to capture both short- and long-term trends

in the data by taking into consideration both past observations and their relationships with one another. The projected values in the autoregressive integrated moving average (ARIMA) forecasting method are a linear function of the most recent actual values and the most recent prediction mistakes (residuals).

1. Seasonal Time Series: Repeating patterns at regular intervals, like seasons or months.
2. Non-Seasonal Time Series: Lack recurring patterns or cycles.

2. Logistic Regression: This study examines the dataset which contains information about 100 rows, including age, weight, gender, fitness score, and a categorization of heart disease as "Present" or "Absent." The principal aim is to build a binary logistic regression model that helps understanding of the association between these variables and the existence or non-existence of cardiac problems.

We start our investigation by going over every variable in the dataset in detail. We then use descriptive statistics and graphics to get a better understanding of participant characteristics and how they are distributed throughout the dataset.

Each of the variables of the final logistic regression model will be succinctly stated to make sure all necessary assumptions are satisfied. An assessment of the model's fit and performance will be included with this report to confirm that participant variables accurately predict the presence or absence of heart problems.

In order to provide important insights for healthcare implications, the ultimate goal of this study is to identify the correlations between participant features and the probability of heart problems.

II. Methodology

Time Series Analysis:

1. Data Collection and Understanding:

Meteorological data from Dublin Airport, containing temperature-related attributes, was collected.
Descriptive statistics and visualizations (time series plots, seasonal plots, decomposition) were used to comprehend the data.

2. Data Cleaning:

Identified and handled missing values and outliers.

3. Data Modeling:

- Employed different time series models such as Simple Time Series Models (Average, Naïve, Seasonal Naïve, Drift), Exponential Smoothing (Simple, Holt, Holt-Winters), and ARIMA/SARIMA models.
-Evaluated models using measures like RMSE, MAE, MAPE, and assessed their forecasting accuracy.

Logistic Regression:

In approaching the research questions, the adopted methodology followed the Knowledge Discovery in Databases (KDD) process, a comprehensive framework enabling systematic data analysis and knowledge extraction.

1. Data Understanding and Preprocessing:

-The dataset containing participant information for cardiac conditions was explored.
-Descriptive statistics and visualizations were utilized to understand variable distributions.
-Balanced the dataset and handled missing values, outliers, and transformed variables.

2. Model Building and Evaluation:

-Developed logistic regression models considering all features and later with selective features (backward model).
-Assessed model performance using metrics like accuracy, precision, recall, F1-score, ROC curves, confusion matrix, and significance tests.

III. Implementation

3.1 Time Series Analysis

3.1.1 Overview

The methodology systematically explores the determinants of Grass Minimum Temperature (GMT) through a structured time series analysis. It begins with cleaning, and formatting to create a coherent time series dataset.

3.1.2 Data Understanding

In this analysis carry out the requirement of variable gmin.Grass.Minimum.Temperature...degrees.C.(gmin). After importing the dataset into the preferred R environment, an initial examination was carried out to comprehend its format and information. The dataset consists of diverse weather attributes and dates, comprising a mix of 9 numerical and categorical variables ready for analysis.

To understand the target variable following descriptive statistics were calculated.

Varia -ble	Min	Median	Mean	Max
date	1942-01-01	1982-12-01	1982-12-01	2023-10-31
gmin	-15	4.5	4.315	17.9

Table 1: Summary of necessary variables

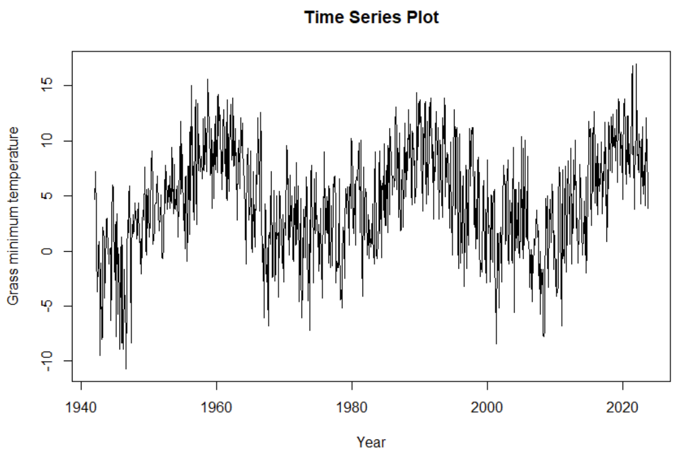


Figure 1: Time Series plot

In Figure 1, the Grass minimum temperature varies across the recorded time period, a clear upward trend over the 80 years cyclical nature, with faster and slower periods displaying seasonal patterns.

3.1.3 Data Cleaning

The first step in data cleaning was to identify missing values. We found that 7 missing values in the dataset, and dropped it. The second step was analyzing the outliers, as a result #29857 rows remained. The plots below display the 'gmin' variable before and after removing outliers.

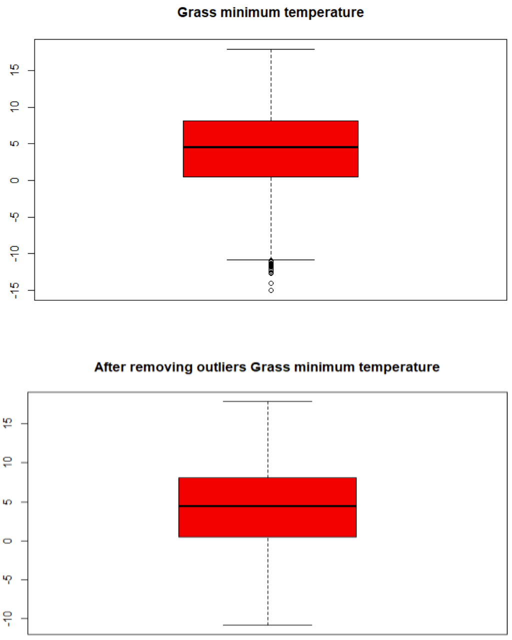


Figure 2. Comparison of the before and after outliers

3.1.4 Visualizations

Before we create any visualizations, stored the data in time series object monthly.

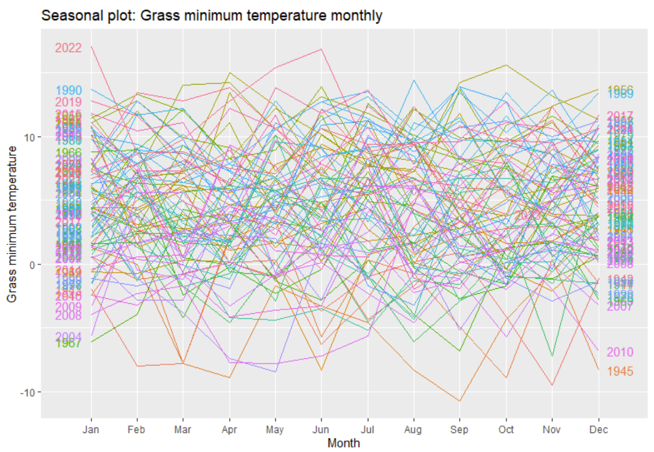


Figure 3. Seasonal plot monthly

In Figure 3, how the Grass minimum temperature changes month by month across different years and we can see that between 0 to 10 degrees is very density.

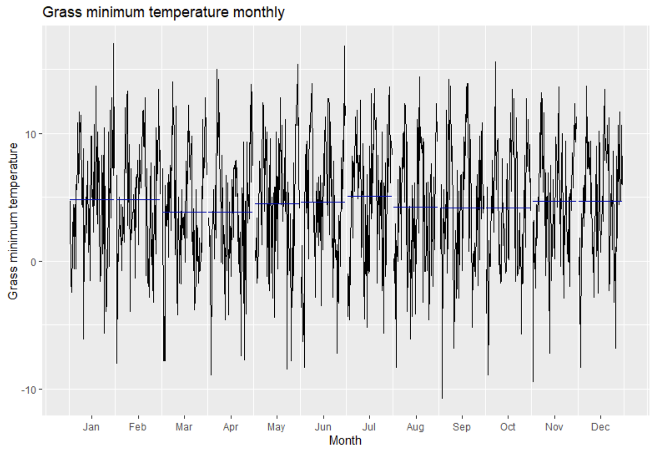


Figure 4. Subseries plot monthly

In Figure 4, these representations aid in our understanding of fluctuations and trends in the 'Grass minimum temperature' over various data segments by allowing us to see temperature patterns within shorter timeframes.

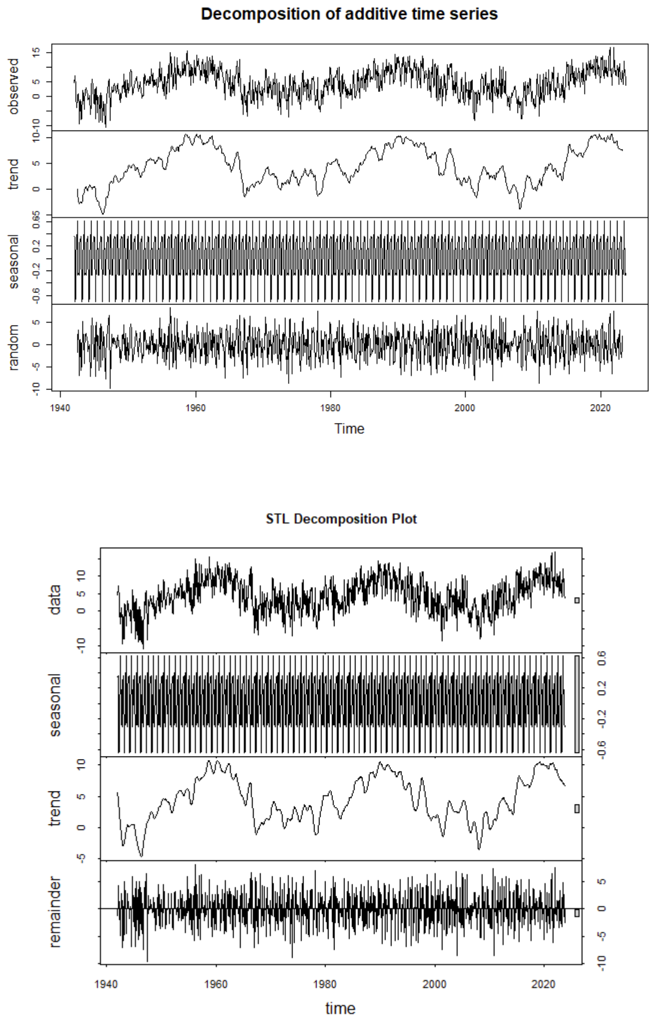


Figure 5. Decomposition of additive and STL time series

Additionally, first plot of Figure 5, the additive method is used for seasonal decomposition separating the time series into additive components of seasonal, trend and random. This plot suggests an alternative view of seasonal and trend patterns within the data.

The second plot which STL(Seasonal and Trend decomposition using Loess) displays each component separately overall time series. As we see time series has repeated up down trend over the period, and approximately every 30 years this seasonal trend appears again.

3.1.5 Modelling

3.1.5.1 Simple Time Series

The time series needs to be divided into two parts before modeling can begin: the test set will represent the data starting in 2023 and the training set comprises the years 2019 to 2022.

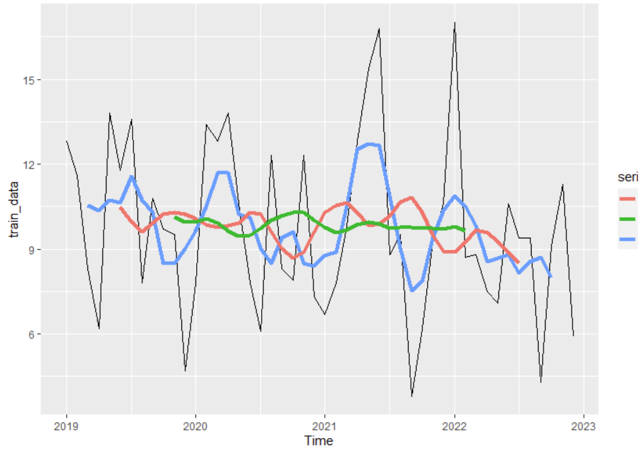


Figure 6. Smoothed Series Comparison

As above Figure 6 shows how different moving average window size impact the representation of the data. We can see that MA(20) is most smoothness line similar to straight line, among from those MA(5) is most appropriate window size for analyzing the underlying trends in the time series data

Simple time series modelling, I built 4 different approaches of data using train set. As below shown in Table 2 Average method has the lowest error compared to other methods.

Simple time series			
	RMSE	MAE	MAPE
Average method	3.05	2.46	29.623
Naïve method	3.73	3.01	36.006
Seasonal-Naïve method	4.46	3.70	41.041
Drift method	3.73	3.00	35.654

Table 2. Error measures for forecasting

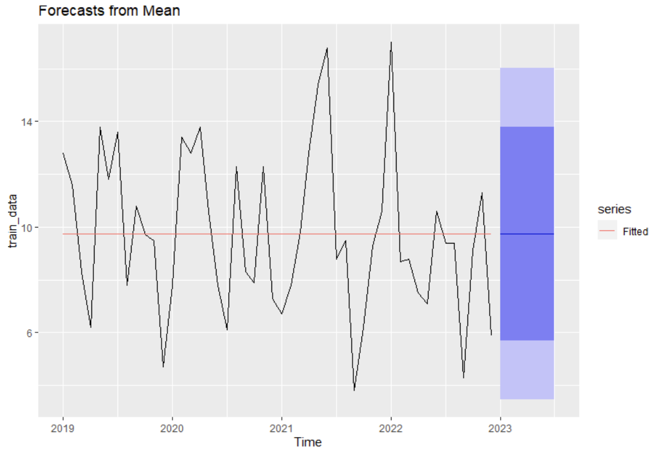


Figure 7 . Forecasts from the average method on train set

Figure 7 displays the forecasted values overlaid with the fitted values from the average method model, and fitted value is close to 10.

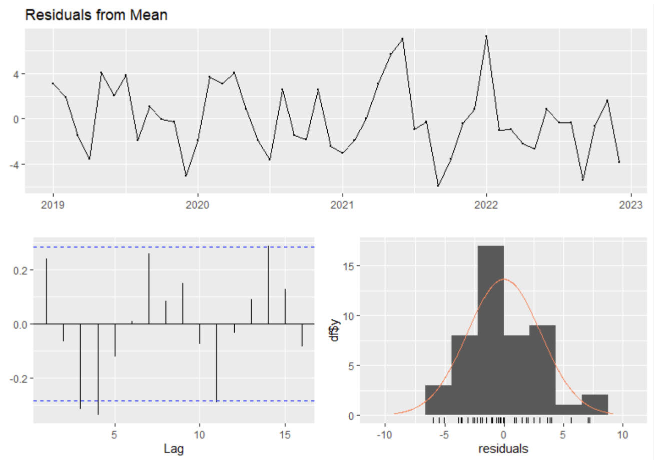


Figure 8. Residuals from average method

In Figure 8, assessed the quality of forecasts by examining the residuals, which are differences between the predicted values and actual values., generally the assumptions made about the residuals in the forecasting process are reasonable.

Cross-validation is a crucial technique in machine learning and time series analysis for assessing the predictive performance of models. For Drift model, applied Cross-validation and resulting roughly 5.678 which is higher than forecasting model, otherwise a higher RMSE from cross-validation frequently indicates a more conservative and dependable model that is less likely to overfit and can generalize better to new, unseen data, even though a lower RMSE in a single instance could seem more desirable.

The fitted four models' average method had lowest errors, so I carried on it to evaluation step using test set. Compared to forecasting modelling error, both metrics were lower, it suggests these observations are positive signs, also exhibited strong performance in the evaluation phase when applied to test set.

Evaluation on test set		
MAE	RMSE	MAPE
2.080	2.481	33.705

Table 3. Error measures for evaluating

3.1.5.2 Exponential Smoothing

Exponential Smoothing modelling built using different methods such as Holt's method, Simple Exponential Smoothing, Holt's-Winter method. The AIC or Akaike Information Criterion is a metric used to compare statistical models. It balances the goodness of fit of a model with its complexity, aiming to select the most appropriate models among set of candidates. AIC value of each modelling shows below Table 4.

Exponential Smoothing	
Model	AIC
Holt's linear trend ANN	298.97
Holt's linear trend AAN	301.34
Holt's linear trend AAA	314.33
Holt's linear trend ZZZ	298.974
SES(Simple Exponential Smoothing)	298.974
Holt-winter additive seasonal	314.331
Holt-winter multiplicative seasonal	311.588

Table 4. AIC value for each modelling

Error measures for low AIC		
Holt's Linear ANN and ZZZ, SES		
MAE	RMSE	MAPE
2.461	3.053	29.630

Table 5. Error measures for low AIC models

From Table 4, the lowest AIC value which is 298.97 of modelling is Holt's Linear trend ANN and ZZZ, and Simple Exponential Smoothing. When comparing models, a lower AIC value indicates a better trade-off between model complexity and goodness of fit, making it the preferred option. And in Table 5 shows out of sample errors for low AIC models.

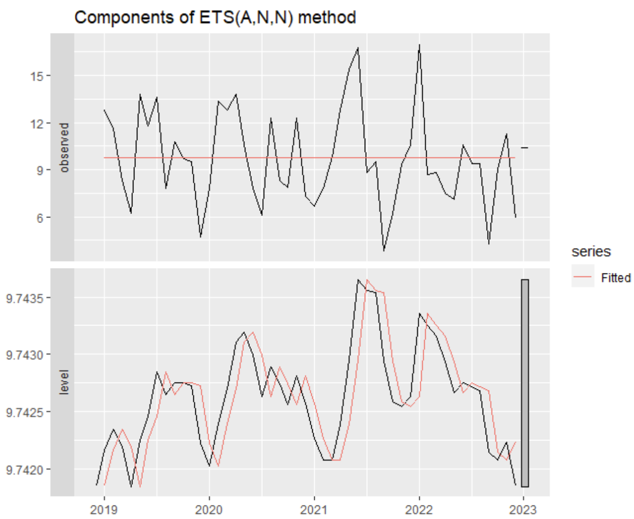


Figure 9. ANN smoothing model and fitted values

The purpose of creating Figure 9 is to provide a comprehensive view of the time series data, the forecasted values generated by the Holt's Linear trend ANN, and fitted values obtained from the model. Second plot of Figure 9, red line is almost aligning forecasted values.

Evaluated models for the lowest AIC value of models on test set.

Evaluation on test set		
Holt's Linear ANN and ZZZ, SES		
MAE	RMSE	MAPE
2.080	2.481	33.705

Table 6. Error measures for evaluating

Forecasting RMSE and MAE metrics were higher evaluating models, and values were exact same as *3.1.E.1 best Simple Time Series* modelling.

3.1.5.3 ARIMA/SARIMA

The first step of modelling was to determine the number of differences needed for stationarity, resulting no needed to do differencing. Following this performed Duckey-Fuller test and resulting alternative hypothesis was stationary.

Prior to constructing models, it is essential to determine the values of 'p' and 'q' for the ACF/PACF plots. This is crucial because the order numbers (p, d, q) for ARIMA modeling are identified from the ACF/PACF plots.

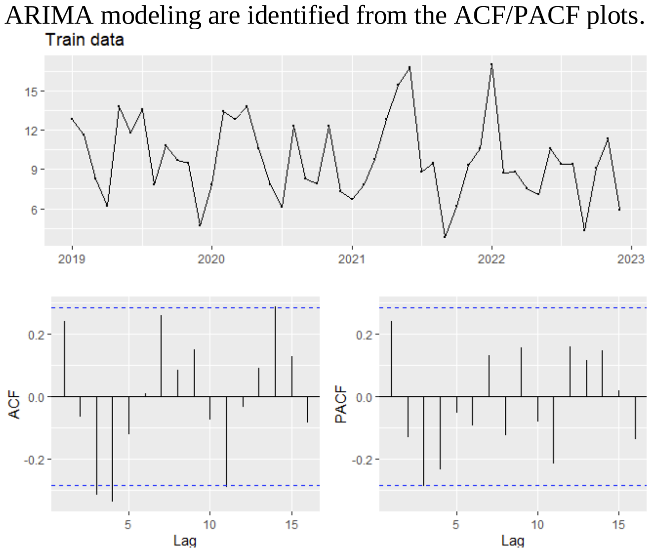


Figure 10. Time series plot

From PACF plot of Figure 10 indicates that the most suitable model is AR(1), we can see that p=1, on the other hand from ACF plot q=0, so we will first fit an ARIMA(1,0,0) model. Will also fit close variations of this model, such as ARIMA(0,0,1), ARIMA(0,0,0), ARIMA(1,0,1)

.

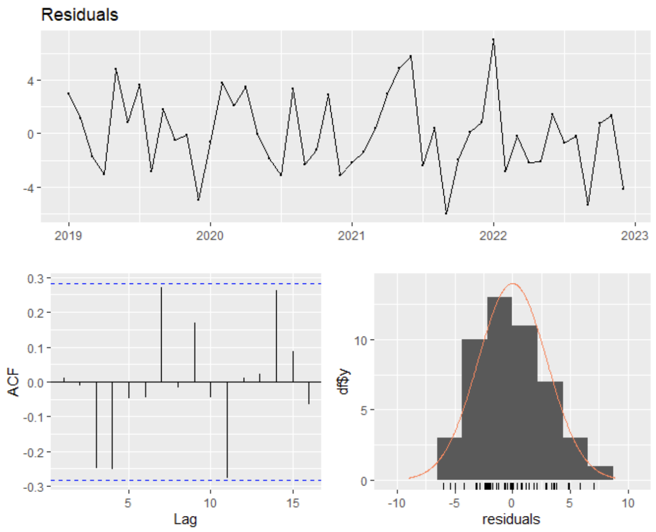
ARIMA modelling				
	AIC	RMSE	MAE	MAPE
ARIMA(1,0,0)	246.44	2.9594	2.4185	29.112
ARIMA(0,0,1)	246.19	2.9591	2.3839	28.775
ARIMA(0,0,0)	247.37	3.0531	2.4611	29.623
ARIMA(1,0,1)	248.15	2.9502	2.3928	28.852

Table 7. Error measures for forecasting

From Table 7, the ARIMA(0,0,1) model has the lowest AIC. Let's use the *auto.arima()* function from *forecast* library to automatically determine the model order. Automatic arima suggested ARIMA(0,0,1) as the most suitable model.

We can now plot the residuals to see if they significantly differ from white noise. Note that this also performs the Ljung-Box test for serial correlation. It has the following hypotheses:

- H0: The distribution of the series is independent.
- H1: There is serial correlation in the series, not independent distribution.




```
        Ljung-Box test

data:  Residuals
Q* = 13.296, df = 10, p-value = 0.2076

Model df: 0.   Total lags used: 10
```

Figure 11. Diagnostic tests

In Figure 11, the lack of significant autocorrelation in the residuals up to the lag order of 10 according to the Ljung-Box test results suggests that the model appropriately represents the temporal relationships found in the data and result of this test indicates no serial correlation.

Next, we produced a forecast using the fitted model, will forecast 7 months into the future.

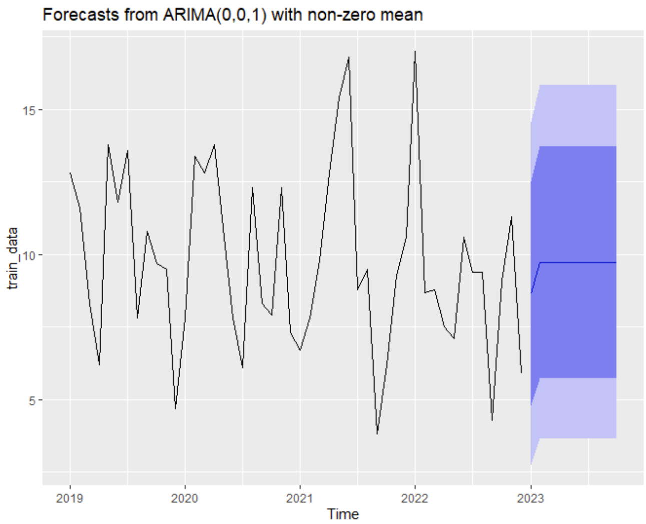


Figure 12. Forecast Plot: Model fit001 Predictions with Uncertainty

Figure 12 observing the forecasted values against the actual data in the plot aids in identifying any trends, seasonal patterns, or deviations from the historical observations.

Lastly evaluated ARIMA(0,0,1) model got MAE, RMSE, and MAPE metrics values, these are measures of out of sample error.

Evaluation on test set		
MAE	RMSE	MAPE
3.019	3.573	0.312

Table8. Error measures for evaluating

3.2 Logistic regression

3.2.1 Data Understanding

The cardiac.csv dataset, consists 100 rows and 5 variables such as age, weight, gender, fitness_score, and cardiac_condition. The cardiac_condition is target variable indicating the presence or absence of a cardiac condition among the features.

Building a binary logistic regression model to investigate the associations between the and the probability that a participant would have a heart problem requires an understanding of these variables. The aim of this analysis is to determine the potential impact of these factors on the participants' presence or absence of cardiac problems.

To understand the data the following descriptive statistics were calculated.

Variable	**Min**	**Median**	**Mean**	**Max**
Age	30.00	41.10	39.00	74.00
Weight	50.00	79.66	79.24	115.42
Fitness Score	27.35	43.63	42.73	62.50

Table 9: Descriptive statistics

Several visualizations were created to illustrate the distributions within the dataset, as an example distribution of cardiac condition across gender below shown.

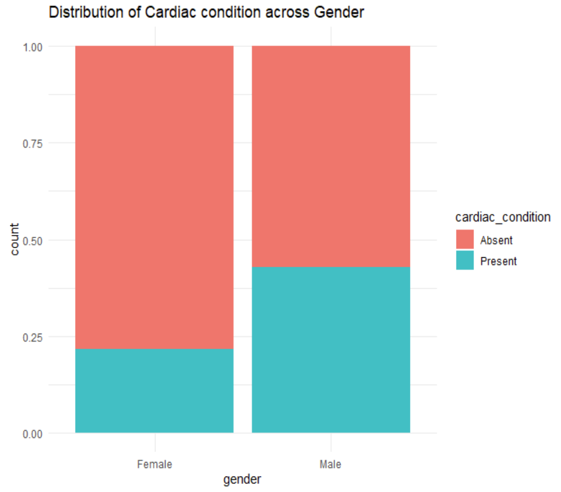


Figure 13. Distribution of gender

In figure 13, absent is consisting of female and male both more than 50%. That means dataset is imbalanced, for the logistic regression balanced dataset is one of crucial step.

When checks the table for target variable which is cardiac_condition, level of absent has 65, present has 35 values respectively.

3.2.2 Data Preprocessing

Before cleaning the dataset, it was balanced so that the 'Absent' and 'Present' tables contained 35 values each. Consequently, the dataset retained a total of 70 rows.

The first step in data cleaning was to identify missing values. We found that no missing values in the dataset. After this used effect coding for gender variable's level.

The second step was analyzing the outliers, as a result #67 rows remained. As an example, variable age before and after removing outliers are shown in figure 14.

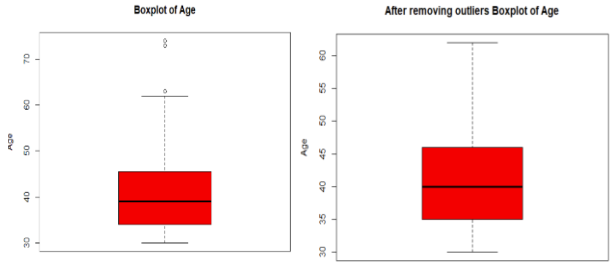


Figure 14. Comparison of the before and after outliers

Note, variable gender is not numeric, its factor, but we did effect coding, so in correlation matrix it's excluded.

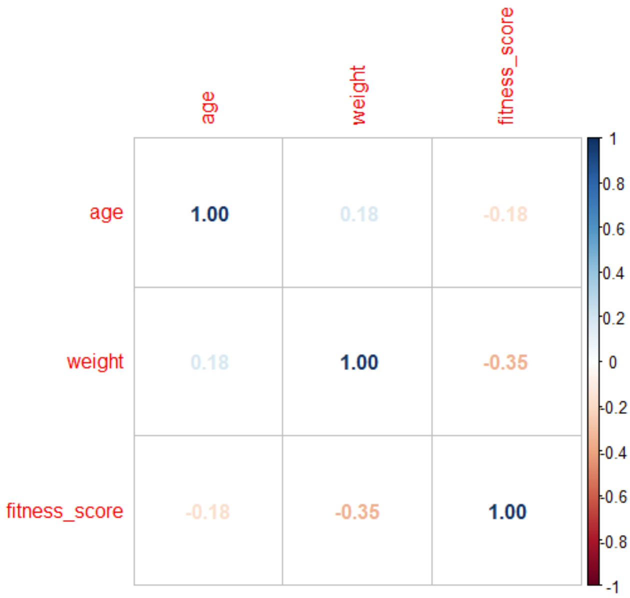


Figure 15: Correlation matrix – excluding gender

3.2.3 Data Transformation

Inspection of the data showed that the following needs to be done:

- Transformation of independent variables is required to normalize.

Min-Max Normalization, also known as Min-Max Scaling, is a method used to rescale numerical data within a predetermined range, typically between 0 and 1. This technique is applied to standardize various features or attributes of a dataset, ensuring that all values lie within the specified range. Following Table 10 shows normalized features.

	age	**weight**	**Fitness_ score**
Min.	0.000	0.000	0.000
Median	0.338	0.483	0.443
Mean	0.313	0.459	0.421
Max.	1.000	1.000	1.000

Table 10: Normalization of the features

3.2.4 Modelling

Before constructing the model, the dataset was divided into training and testing data, utilizing the 'set seed' function to ensure reproducibility by fixing the randomization process based on the student number.

3.2.4.1 Model 1

First model built using all normalized features and gender. As below shown essential metrics for logistic regression in Table 11.

Logistic regression Model 1	
Accuracy	0.833
Precision	1
Recall	0.667
F-1 score	0.8

Table 11: Evaluation metrics

In Table 11, roughly, 83.3% of the model’s predictions across all classes were correct. The other metrics are more than 0.65 which means with higher values indicating better performance.

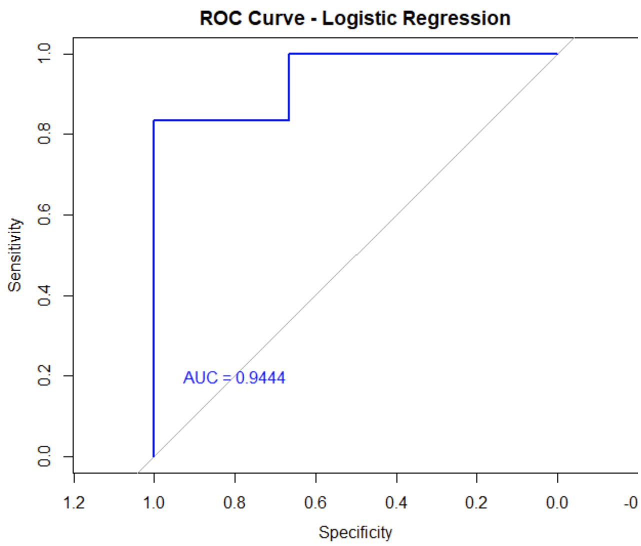


Figure 16. ROC curve of model 1

Figure 16 displays the corresponding Area Under the Curve (AUC), which is an essential metric for evaluating the performance of a binary classification model, such as the logistic regression model. As in the plot AUC value is about 0.94, in this situation indicates that the model has good discriminatory power to separate the two classes.

Assessment of model 1			
Omnibus test		Df	Chi
	Age	1	0.0002056
	Weight	1	0.1461058
	Gender	1	0.2343554
	Fitness_score	1	0.0144264
Wald Statistic		Pr	
	Age	0.00313	
	Weight	0.33957	
	Gender	0.01525	
	Fitness_score	0.02847	
Odds ratios	(Intercept)	2.569961e+00	
	age	2.777956e+02	
	weight	1.584853e-01	
	gender	4.524484e+00	
	fitness_score	2.604186e-03	
Confusion matrix	Predicted		
	Actual 0 1		
	Absent 6 0		
	Present 2 4		
-2 log likelihood	52.93		
Pseudo R-square measures	Cox&Snell R squared(r2cu)		
	0.46		

Table 12: Asssessment of model 1

From Table 12, Wald statistic test indicates which of our predictors are statistically significant and resulting except weight is significant. Odd ratios, the change in odds being in one of the categories of outcome when the value of a predictor increase by unit, and for every one-unit increase in age, the odds of the outcome increase by a factor approximately 278 times.

3.2.4.2 Backward model

Initially, a backward model was constructed using the first model, wherein all features exhibited high AIC values, recorded at 62.94. Subsequently, by excluding the 'weight' variable, the AIC value decreased to 61.89.

So, excluding the weight new logistic regression is built, result is displayed in Table 13.

Backward model	
Accuracy	0.75
Precision	0.8
Recall	0.6667
F-1 score	0.7272

Table 13: Evaluation metrics

In Table 13, 75% of the model's predictions across all classes were correct. The other metrics are more than 0.65 which is similar to model 1 result.

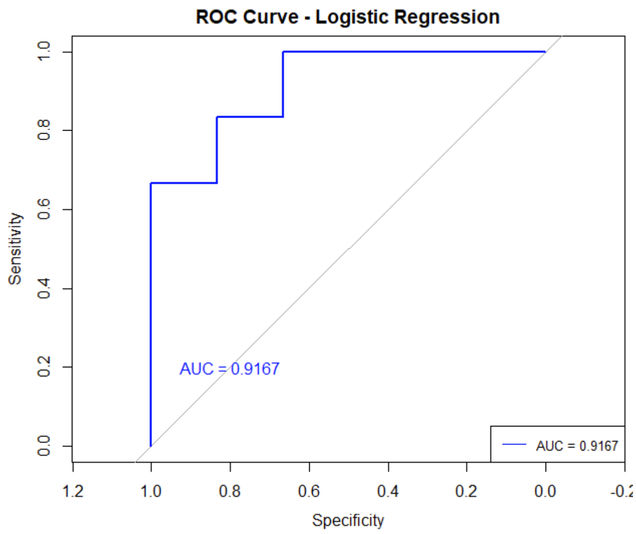


Figure 17. ROC curve of backward model

Figure 17 as in the plot AUC value is about 0.917, in this situation indicates that the model has good performance and it is lower than model 1AUC value.

Assessment of backward model			
Omnibus test		Df	Chi
	Age	1	0.0002056
	Gender	1	0.0963819
	Fitness_score	1	0.0160560
Wald Statistic		Pr	
	Age	0.00307	
	Gender	0.01411	
	Fitness_score	0.03324	
Odds ratios	age	218.78	
	gender	3.082	
	fitness_score	0.0128	
Confusion matrix	Predicted		
	Actual 0 1		
	Absent 5 1		
	Present 2 4		
-2 log likelihood	53.88		
Pseudo R-square measures	Cox&Snell R squared(r2cu)		
	0.445		

Table 14: Asssessment of backward model

From Table 14, Wald statistic test indicates all predictors are statistically significant. In general, comparing Table 12 and Table 14, deviance (-2 log likelihood) was high in first modelling, that means first model is better the fit. For Omnibus test, we reduced the Ch-square value all predictors, and degree of freedom still remained the same. Pseudo R-square measures refer to the amount of variation in the dependent variable, which is predicted by the predictor variables collectively, and maximum value of this, in theory 1.00 if the relationship is perfect, if it is 0 no relationship. First and backward modelling R-squared were almost similar values they had.

IV. Conclusion

In conclusion. Time series and logistic regression are two important fields that this study covered in detail. The analysis of historical meteorological data from Dublin Airport revealed patterns and variations in the gmin, offering important new information about the temporal development of ground-level cold. By means of thorough investigation with multiple models, such as Exponential Smoothing and ARIMA/SARIMA, the forecasting potential of gmin was improved. The models that were identified exhibited unique capabilities. Specifically, the Average Method demonstrated the best predictive performance in the Simple Time Series modeling. The Exponential Smoothing models had competitive forecasting ability, closely correlating with the top-performing Simple Time Series models. This was especially true for Holt's Linear trend ANN and ZZZ, as well as SES. For ARIMA/SARIMA order (0,0,1) was most suitable modelling and Automatic arima model also suggested as well.

Error measures such as RMSE, MAE, and MAPE were same in training period for best Simple Series model and Exponential Smoothing. In contrast, Arima has the lowest out of sample errors among those 3 models.

And in the test period Simple Time series and Exponential Smoothing was the lower than ARIMA/SARIMA, but there was no huge difference at all. Additionally, the lowest AIC was 246.19 in ARIMA/SARIMA modelling. So, we can say that ARIMA/SARIMA model is an optimum model.

In addition, the Logistic Regression study examined how participant characteristics such as age, weight, gender, and fitness scores may be used to predict the chance of heart problems. There were notable correlations found between the characteristics and the likelihood of heart issues in both models—Model 1, which included all features, and the Backward Model, which included a more specialized collection of predictors. Although the first, more complete model included more variables, the improved model—which did not include "weight"—performed similarly well, but with somewhat less predictive ability. The model coefficients exhibited significance for various predictors such as age, fitness_score among others. These coefficients indicate their significant influence on cardiac condition.

Future work, consider incorporating more historical weather data or additional meteorological attributes to improve the robustness of the models.

The integration of different methods other than logistic regression may be investigated in future studies in an attempt to increase the anticipated accuracy and analytical depth. In particular, the application of techniques such as Random Forest and Decision trees exhibits potential as a supplementary means to enhance the current study.

V. References

[1] R.J. Hyndman and G. Athanasopoulos, "Forecasting: principles and practice," OTexts, 2018.

[2] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, and G.M. Ljung, "Time series analysis: forecasting and control," John Wiley & Sons, 2015.

[3] J.D. Cryer and K.S. Chan, "Time series analysis: with applications in R," 2nd ed., Springer, 2008.

[4] P.J. Brockwell and R.A. Davis, "Introduction to time series and forecasting," 3rd ed., Springer, 2016.

[5] D.W. Hosmer Jr., S. Lemeshow, and R.X. Sturdivant, "Applied logistic regression," 3rd ed., John Wiley & Sons, 2013.

[6] A. Agresti, "Foundations of linear and generalized linear models," John Wiley & Sons, 2015.

[7] J. Fox and S. Weisberg, "An R companion to applied regression," 3rd ed., Sage Publications, 2018.

[8] S. Menard, "Applied logistic regression analysis," 2nd ed., Sage Publications, 2002.

[9] C. Chatfield and H. Xing, The Analysis of Time Series: An Introduction with R (Chapman & Hall/CRC Texts in Statistical Science). CRC Press, 2019.

[10] J. D. Cryer and K. S. Chan, Time Series Analysis: With Applications in R (Springer Texts in Statistics). Springer, 2008.

[11] J. M. Hilbe, Logistic Regression Models (Chapman & Hall/CRC Texts in Statistical Science). CRC Press, 2009.

[12] R. Krispin, Hands-On Time Series Analysis with R: Perform time series analysis and forecasting using R. Packt Publishing, 2019.

[13] S. W. Menard, Logistic Regression: From Introductory to Advanced Concepts and Applications. SAGE Publications, 2010.

[14] R. Shumway and D. Stoffer, Time Series: A Data Analysis Approach Using R (Chapman & Hall/CRC Texts in Statistical Science). CRC Press, 2019.

[15] P. Teetor, R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics. O'Reilly Media, 2011.

[16] W. A. Woodward, H. L. Gray, and A. C. Elliott, Applied Time Series Analysis with R. CRC Press, 2017.