MASTER OF SCIENCE in DATA SCIENCE

FORECASTING TOURIST ARRIVAL IN THE UK

A DISSERTATION SUBMITTED TO MANCHESTER METROPOLITAN UNIVERSITY
FOR THE DEGREE OF MASTER OF DATA SCIENCE IN THE FACULTY OF SCIENCE
AND ENGINEERING

By

LAVANYA SREEDHAR

DEPARTMENT OF COMPUTING AND MATHEMATICS

Table of Contents

List of Tables

List of Figures

Abstract

The tourist sector, a crucial contributor to the UK economy, has received substantial attention in the context of global uncertainty and internal political dynamics. This research investigates the difficulties and potential in the UK tourist business, recognising its vulnerability to crises and political upheaval. With the World Tourism Organisation listing the United Kingdom as one of the most visited nations, welcoming 37.7 million tourists in 2019, the economic effect of travel and tourism cannot be emphasised.

In 2021, the industry will contribute a significant 131.5 billion British pounds to the country's GDP, highlighting the need of good forecasting in tourist development. To anticipate visitor arrivals from various nations, this study utilises three major time-series forecasting techniques: exponential smoothing, univariate ARIMA, and Elman's Model of Artificial Neural Networks (ANN).

The analysis is carried out using Python programming and statistical programmes, with Google Collab and Minitab acting as the key tools. The results demonstrate that, while all three models show potential, the ARIMA model is the most accurate in predicting visitor arrivals. This discovery is significant for policymakers and stakeholders involved in tourist development because it provides vital insights for decision-making and resource allocation. The initiative contributes to the continuing discussion about leveraging forecasting approaches for sustainable tourist planning, so increasing the resilience of the "smokeless industry" in the face of changing global dynamics.

Declaration

No component of this project has been submitted in support of an application for any other degree or certification at this or any other university of learning. This project is entirely my own work, with the exception of reference to the work of others.

Signed

Abbreviations

Artificial Neural Network	ANN
Augmented Dickey Fuller Test	ADF
Auto Regressive Integrated Moving Average	ARIMA
Auto Regressive Moving Average	ARMA
Auto Regressive	AR
Autocorrelation Function	ACF
Comma separated Values	CSV
Consumer Price Index	CPI
Convolutional Neural Network	CNN
Empirical Mode Decomposition	EMD
Exponential Smoothing	ES
Extensible Markup Language	XML
Gross Domestic Product	GDP
Institute of Electrical and Electronic Engineers	IEEE
International Monitory Fund	IMF
Key Performance Indicator	KPI
Long Short-Term Memory	LSTM
Mean Absolute Percentage Error	MAPE
Mean Average Error	MAE
Mean Squared Error	MSE
Molecular Diversity Preservation International	MDPI
Moving Average	MR
Office for National Statistics	ONS
Ordinary Least Squares	OLS
Partial Autocorrelation Function	PAFC
Recurrent Neural Network	RNN
Root Mean Squared Error	RMSE
Seasonal Auto Regressive Integrated Moving Average with Exogenous Factors	SARIMAX
Seasonal Auto Regressive Integrated Moving Average	SARIMA

Seasonal-Trend decomposing using LOESS	STL
Standard Error	SE
Support Vector Regression	SVR
Time Series Analysis	TSA
United Kingdom	UK
Vector Innovations Structural Time-Series	VISTS
World Travel & Tourism Council	WTTC

Chapter 1 – Introduction

1.1 Project Overview

Tourism is a vibrant and crucial industry in the United Kingdom, contributing considerably to the economy and cultural exchange. Millions of travellers worldwide have come to the United Kingdom throughout the years, lured by its rich history, diverse culture, breathtaking landscapes, and dynamic cities. However, the tourism business is not immune to the effect of numerous variables such as economic conditions, geopolitical events, and global health problems, all of which may significantly impact visitor arrivals. Therefore, forecasting tourist arrivals in the UK has become a crucial endeavour for policymakers, entrepreneurs, and stakeholders in the tourism sector, allowing them to make informed decisions, allocate resources efficiently, and adjust to changing conditions.

Forecasting tourists is critical for the tourism business in the country. According to a VisitBritain estimate, the UK will get 31.2 million incoming visitors in 2022, after two years of severely low arrivals due to the impact of COVID-19. The most recent tourism report in the United Kingdom provides an upbeat image for the sector, expecting 29.7 million incoming visitors in 2023. This prognosis reflects a positive improvement over the previous forecast issued in August 2022, owing mostly to better-than-expected outcomes gathered from official statistics. The renewed confidence reflects the UK tourist industry's resilience and potential for expansion, despite the constraints created by diverse global events. However, it is crucial to highlight that this estimate is based on several assumptions, most notably that cost of living pressures will not worsen and that inflation will gradually ease from its present levels. These assumptions highlight the tourist industry's sensitivity to larger economic conditions, as well as the necessity for continual monitoring and flexibility in the face of changing economic dynamics.

This study dives into the essential problem of projecting visitor arrivals in the United Kingdom, investigating the several factors that impact this phenomenon. It intends to give significant information and prediction models to assist government authorities, tourist organizations, and enterprises in the United Kingdom in planning and strategizing for the future. By analysing historical data, taking into account current trends, and accounting for outside factors, this research seeks to improve understanding of the complex dynamics governing tourism in the UK, ultimately resulting in a more sustainable and resilient tourism industry in the face of a constantly changing global landscape.

1.2 Potential Problems

The process of forecasting tourist arrivals in the UK involves several critical components, each presenting its unique set of challenges. Firstly, ensuring the availability of accurate historical data on international visitor arrivals is paramount. The reliability of forecasts heavily relies on the quality and comprehensiveness of this data, making data collection and validation procedures a crucial initial step. Secondly, the selection and application of appropriate forecasting models demand meticulous consideration. The complexity of tourism dynamics necessitates a careful evaluation of various forecasting strategies to determine the most suitable approach for the UK's unique context. Furthermore, the practical implementation of these models can encounter technical and software-related challenges that must be overcome, such as ensuring compatibility with existing systems and optimizing computational efficiency.

Finally, once the forecasts are generated, the task of interpreting the results becomes pivotal. Examining and analysing predicted patterns requires a nuanced understanding of the tourism landscape, enabling researchers and decision-makers to derive insightful conclusions and formulate relevant suggestions. This multifaceted process underscores the interdisciplinary nature of tourism forecasting, where data science, economics, and domain expertise converge to provide a comprehensive understanding of the future trends in international visitor arrivals to the UK.

1.3 Aim and Objectives

This research compares and contrasts the use of three time-series forecasting techniques, including Elman's Model of Artificial Neural Networks (ANN), univariate ARIMA, and exponential smoothing, for predicting tourist arrivals from different countries in the UK.

In order to attain this goal, the following objectives were set

- Terms of Reference – Providing an overview of the project
- Analysis – To analyze the tourism industry in the United Kingdom and understand its significance in the economic policy uncertainty
- Collect and prepare historical tourist arrival information for the UK from various sources
- Use Elman’s ANN models, univariate ARIMA, and exponential smoothing to predict visitor arrivals
- Apply the proper measures (such as mean absolute error and mean squared error) to

assess each forecasting model’s performance

- Compare the various forecasting method’s reliability and accuracy
- Provide insights and recommendations based on the forecasted tourist arrival patterns

1.4 Tools and Timeline

The research project requires essential resources for effective implementation. Primary among these is a statistical programming language, with Python being the preferred choice due to its versatility and widespread use in data science and time series analysis. Minitab serves as a complementary tool for enhancing data interpretation precision. Google Collab is employed for code-based analysis of time series data when anticipating tourist arrivals in the UK. The dataset for analysis comprises historical data on foreign visitor arrivals to the United Kingdom, incorporating seasonal variations and GDP statistics. This dataset is crucial for constructing accurate forecasting models. Additionally, relevant research studies from diverse academic sources contribute to the project's knowledge base, enriching the existing literature on tourist time series forecasting.

Google Collab gives tourism researchers the tools they need to examine past visitor arrival patterns, spot trends, and identify seasonality in visitor data thanks to its powerful statistical capabilities and user-friendly interface. Analysts may efficiently provide accurate forecasts for future visitor arrivals by applying time series forecasting techniques like ARIMA or exponential smoothing models within the Google Collab environment. In turn, this gives tourism industry players the knowledge they need to decide wisely where to allocate resources, how to promote, and how to plan for capacity, eventually enhancing the UK tourist sector.

To conduct a thorough study on tourist time series forecasting, diverse information sources are essential. Collaborating with the project supervisor is crucial for defining the research path and meeting academic standards. Internet archives like Google Scholar, ResearchGate, IEEE, and ScienceDirect provide access to relevant scholarly publications. Utilizing university library resources, including specialized databases and research collections, enhances the depth of available material. A comprehensive evaluation of various studies and data related to time series analysis, especially in tourism, offers valuable insights and references, forming a solid foundation for the research.

1.5 Report Structure

The substance of this report will be divided into a number of distinct chapters. The following list of project stages will be the focus of each chapter:

- Chapter 1 will include an overview of the project, as well as any potential challenges that may arise, as well as the overarching goals and objectives of the project.
- Chapter 2 will address the literature review or related work that has previously been undertaken by others, as well as outlining forecasting approaches and relating important elements.
- Chapter 3 will address the data collecting step of tourist arrivals in the UK via the ONS website, as well as additional data collection phases.
- Chapter 4 will address the experimental methodology that will be employed during the project's experimentation phase.
- Chapter 5 will address the analysis from the experimental methodology phase
- Chapter 6 will address the further works needed for the project in future
- Chapter 7 will offer a conclusion to the entire report depending on whether the project's goals and objectives were met.

Chapter 2 - Literature Survey

2.1 Introduction

Prior to embarking on data collection and methodology, it was essential to review the literature related to the methods employed in this endeavour. This was a critical step, as it would serve to illustrate the concepts that had already been discussed, while also providing valuable insight into the potential for useful experiments. Additionally, the review of literature was an important part of the process, as it allowed for the discussion of the concepts already implemented as tested, which were closely related to the subject matter of this project.

Time series forecasting approaches have gained popularity in this arena due to their ability to capture temporal patterns and give insights into future trends. This literature review seeks to offer an overview of existing research on time series forecasting in the context of tourism.

A wide range of subjects were examined in order to give insights into this endeavour. Several time series approaches and various factors affecting visitor arrivals in the UK will be attempted in this. They were as follows:

- Traditional Time Series Forecasting Techniques
- Univariate ARIMA model
- Artificial Neural Networks (ANN)
- Exponential Smoothing
- Impact of External Factors
- Seasonality and trends
- Hybrid models
- Data Granularity and Accuracy

These topics are covered in detail beginning in part 2.2 using traditional time series forecasting methods.

2.2 Traditional Time Series Forecasting Techniques

Karadzic and Pejovic (2020) from the University of Montenegro, discuss predicting the total number of tourist arrivals in Montenegro in the article "Tourism Demand Forecasting Using ARIMA model" which likely focuses on the application of the ARIMA model and Box Jenkins methodology for forecasting tourism demand. Karadzic and Pejovic use the Box Jenkins approach to create econometric models that forecast the total numbers of tourist arrival in Montenegro. The authors use the Box Jenkins methods to determine the integration level of the time series data and then pick appropriate autoregressive (AR) and moving average (MA) orders. The ARIMA models are then assessed and compared based on their predicting ability using various variables.

According to the findings of the study, the ARIMA model provides the best-predicting performance for the total number of tourist arrivals in Montenegro. This result is backed by the model's higher performance when applied to data beyond the sample period in terms of forecast metrics such as mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE).

Time series forecasting is the process of making scientific predictions based on previously recorded time stamps. It entails developing models based on previous data and employing them to make observations and drive future strategic decision (Tableau, 2023).

Different Techniques of measuring timed data are referred to as time series methods. Autoregression (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), and Seasonal Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) are some examples (InfluxData, 2023).

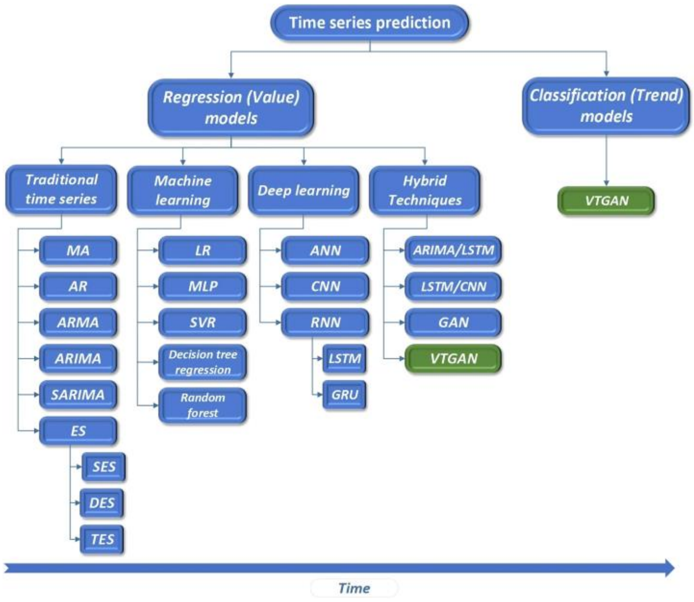


Figure 2.1: Taxonomy of cloud workload prediction models (Maiyza et al, 2023)

Figure 2.1 depicts many machine learning models for time series prediction, a job that involves projecting future values based on historical data. Traditional methodologies such as autoregressive integrated moving averages (ARIMA) and exponential smoothing (ES) are examples of classic time series models, a subset of machine learning. Deep learning methods that use artificial neural networks do very well at predicting complicated and nonlinear time series. For improved performance, hybrid techniques blend standard and deep learning models. A hybrid model extracts attributes from data using a classical time series model and predicts future values using a deep learning model.

The illustration also includes particular examples of machine learning models for time series prediction. Among the most popular models are:

- Traditional Time Series Models
- Deep Learning models
- Hybrid techniques

There are more models that the conventional time series models. Some of them are intended to be described.

2.2.1 Introduction to Time Series Analysis

Time series analysis is a potent technique used in many disciplines, including business, economic, social sciences, medicine, and physical sciences. It functions as a dynamic lens through which it examines the intricate patterns and trends woven within a collection of data points obtained over a set amount of time. Researchers and practitioners may discover priceless insights into the temporal dynamics of variables using this analytical strategy, as Jim Frost eloquently explains in his investigation of time series analysis. Exploring this approach helps one comprehend the complexity that come with data variations over time, allowing for a more nuanced understanding of how various factors change over time, interact and affect results. Time series analysis is comprehensive, which highlights its critical role in revealing hidden patterns, producing accurate forecasts, and ultimately promoting informed decision-making across a range of disciplines (Frost, 2023).

In the field of tracking corporate business KPIs and keeping track of industrial operations, time series data is emerging as a crucial instrument. Time series analysis offers a methodical way to track and analyse the temporal evolution of critical variables in the complex world of industrial operations, where productivity and performance are vital. This analytical method is crucial for seeing trends, spotting abnormalities, and streamlining procedures to increase efficiency. Time series data is crucial for identifying patterns, evaluating variations, and providing data-driven insights in the corporate arena, where key performance indicators (KPIs) and business metrics influence strategic choices (*NIST* 2023).

Time series analysis is a complex toolkit designed for deciphering subtle patterns in sequential data. This analytical profession provides a specialised lens for revealing underlying trends throughout time, going beyond simple data observation to interpretation. It blends statistical methodologies, mathematical models, and computer algorithms to decode the temporal riddles concealed in data streams, as exemplified by the Data Science Wizards' clever inquiry (Wizards, 2023).

Time series forecasting is a strategy for predicting future occurrences by evaluating patterns and trends in a chronological sequence of data. It is based on the assumption that the trajectory of future events may be adequately anticipated by analysing and extrapolating from previous trends. This approach examines the patterns buried in historical data to create educated predictions about future events, providing useful insights for decision-making and planning based on the continuity of established trends through time.

The following are some of the most often used conventional time series models, which are a continuation of 2.2.1:

2.2.2 Autoregressive

Autoregression is a time series analysis statistical approach in which the present value of a time series is modelled as a linear combination of its prior values. The AR model posits that a linear equation may express the connection between the current value and its prior values. An autoregressive model of order p takes the following generic form:

$$Yt = c + \phi 1Yt - 1 + \phi 2Yt - 2 + \cdots + \phi pYt - p + \epsilon \tag{2.1}$$

In this context, Yt is the time series value at a single time point (t), c is a constant term, ϕi symbolises the autoregressive coefficients, p defines the model order indicating the number of past values evaluated, and t denotes the autoregressive error term. This formulation reflects the concept that the current state of the time series is reliant on its previous values, and the model seeks to estimate this connection by adding historical data and accounting for data variability via the error component (Brownlee, 2021).

2.2.3 Moving Average

The moving average model, commonly referred to as moving average process, is a popular method for modelling univariate time series in time series analysis. According to the moving average model, the output variable is linearly dependant on the present value as well as various previous values of a stochastic (imperfectly predicted) factor (InfluxData, 2023). It can be represented as:

$$Yt = \mu + \theta 1\epsilon t - 1 + \theta 2\epsilon t - 2 + \cdots + \ \theta q\epsilon t - q + \epsilon t \tag{2.2}$$

Y_t denotes the time series value at a specific time point (t), is the series mean, ϕ_i denotes the moving average coefficients, q denotes the model order indicating the number of past error terms considered, and t denotes the moving average error term (InfluxData, 2023).

2.2.4 Autoregressive Moving Average

In terms of two polynomials, weakly stable stochastic time series are described using an ARMA model, or Autoregressive Moving Average model. The moving average is calculated using the second of these polynomials, whereas autoregression uses the first (Stephanie, 2023). It may be represented as:

$$Yt = c + \sum i = 1p\phi iYt - i + \sum j = 1q\theta j\epsilon t - j + \epsilon t \tag{2.3}$$

Y_t is the time series value at a specific time (t), c denotes the constant term, i denotes the autoregressive model parameters, j means the moving average model parameters, and t denotes the error term, commonly known as white noise (Stephanie, 2023).

2.2.5 Autoregressive Integrated Moving Average

ARIMA, which stands for 'Autoregressive Integrated Moving Average,' is a type of model that explains a given time series by taking into account its past values, including its own latencies and lagged forecast mistakes. This equation is then used to forecast the time series' future values. It can be represented as:

$$\nabla dYt = c + \phi 1\nabla dYt - 1 + \cdots + \ \phi p\nabla dYt - p + \theta 1\epsilon t - 1 \ + \cdots + \theta q\epsilon t - q + \epsilon tt \tag{2.4}$$

In this context, $\nabla_d Y_t$ is the d -th difference of the time series Y_t, c is a constant, ϕ_i and ϕ_i are the autoregressive and moving average coefficients, respectively, and p and q define the orders of the autoregressive (AR) and moving average (MA) components. Additionally, t represents the model's error term (Bajaj, 2023).

2.2.6 Seasonal Autoregressive Integrated Moving Average

SARIMA, like ARIMA, is more successful than ARIMA in forecasting complicated data spaces that incorporate cycles. This advantage arises from SARIMA's capacity to use seasonality as a parameter in the modelling process.

$$Yt = c + \sum i = 1p\phi iYt - i + \sum j = 1q\theta j\epsilon t - j + \sum k = 1P\Phi kYt - km + \sum l = 1Q\Theta l\epsilon t - lm + \epsilon t \tag{2.5}$$

Y_t denotes the value of the time series at a particular time (t), C is the constant term, i stands for the parameters of the autoregressive model, j for the parameters of the moving average model, k for the parameters of the seasonal autoregressive model, l for the parameters of the seasonal moving average model, p is the order of the autoregressive polynomial, q is the order of the moving average polynomial, and P is the order.

2.2.7 Seasonal Trend Decomposition

A time series may be broken down into its trend, seasonal component, and residual component using the Seasonal Decomposition of Time Series (STL) approach. The STL approach divides a time series into these parts using locally fitted regression models. R. B. Cleveland, Cleveland, McRae, and Terpenning (1990) created the STL approach. The equation below may be used to illustrate the STL method:

$$Yt = Tt + St + Rt \tag{2.6}$$

The time series value is denoted by Y_t, the trend component is represented by T_t, the seasonal component is captured by S_t, accounting for periodic fluctuations, and the residual component is represented by R_t, encompassing the remaining irregular variations. In order to estimate nonlinear relationships, the STL approach applies the Loess method. The user may regulate the rate of change and allow the seasonal component to fluctuate over time. The user may also modify how smoothly the trend-cycle moves. Not just quarterly and monthly data, but also time series with any kind of seasonality may be broken down using the STL approach (*Seasonal-trend decomposition using loess* 2023).

2.2.8 Exponential Smoothing

Time series forecasting use a type of state space models called exponential smoothing. A measurement equation that represents the observed data plus a few state equations that show how the unseen components or states change over time make up the model.

The exponential smoothing state space model stands out for its versatility in handling diverse data and patterns in time series forecasting. To enhance precision, additional elements like external regressors or interventions can be incorporated. Model estimation can be achieved through maximum likelihood or Bayesian techniques.

2.3 Deep Learning Models

Deep learning, a subset of machine learning, mimics the intricate workings of the human brain through artificial neural networks. Characterized by multiple layers, these algorithms progressively extract complex features from diverse input data, including photos, text, and voice. The term "deep" reflects the network's layered design, enabling the model to uncover subtle patterns and correlations. This hierarchical approach makes deep learning particularly effective in tasks such as image and audio recognition, natural language processing, and other domains requiring nuanced understanding. (Wikipedia, 2023).

2.3.1 ANN

Artificial Neural Networks (ANN) are a type of algorithm that is inspired by the complicated operation of the human brain and is used to model complex patterns and anticipate outcomes. ANN is a notable approach within the realm of deep learning, designed as an attempt to replicate the complicated functions of the human brain. It is based on the concept of Biological Neural Networks. While ANN's workings are similar to, rather than identical to, the fundamental concepts of biological neural networks. Notably, this method is designed for quantitative and structured data, and it requires inputs in these forms to traverse and process information properly (Singh, 2023).

2.3.2 CNN

A Convolutional Neural Network (CNN), often known as ConvNet, is a specialised category of neural networks capable of processing data with a grid-like architecture, and is well suited to applications such as image analysis. CNNs excel in deciphering the complicated patterns within pixel arrangements in the context of digital photography, which serves as a binary representation of visual data. Pixel values contain information about brightness and colour in digital pictures, which are primarily formed of pixels organised in a grid-like layout. CNNs use convolutional layers to systematically scan and extract characteristics from this grid, enabling hierarchical learning that is very effective for image recognition and computer vision applications.

2.3.3 RNN

A recurrent neural network (RNN) is a form of artificial neural network that is designed to operate with time series data or data that includes sequences. Ordinary feedforward neural networks are only intended for data points that are independent of one another. However, if it has data in a series where each data point depends on the one before it, it must adjust the neural network to take these dependencies into account. RNNs contain a concept of "memory" that allows them to retain the states or information of prior inputs in order to create the next output in the sequence.

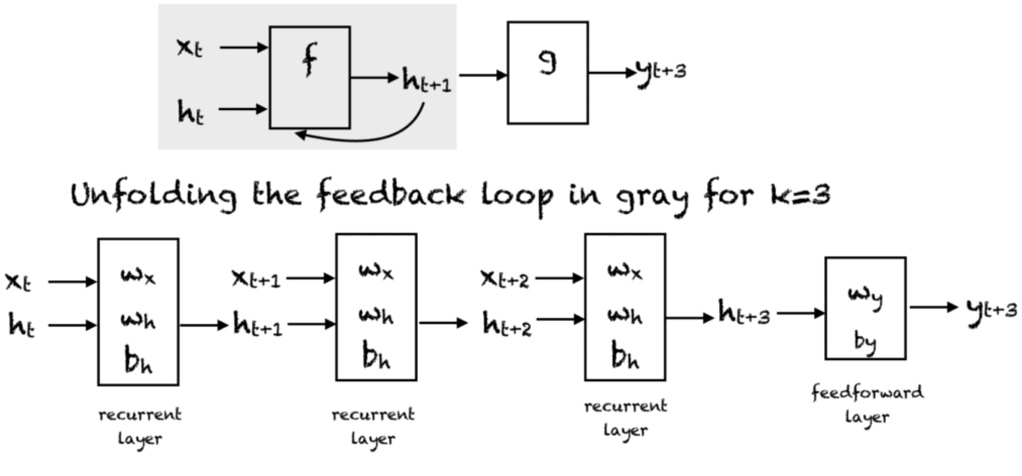


Figure 2.2: Compressed representation of RNN (Saeed, 2023)

Figure 2.2 is a schematic illustration of a feedback loop within a Recurrent Neural Network (RNN). This loop, which connects the recurrent layer to itself, is illustrated in grey. The recurrent layer is made up of linked neurons that create a loop, allowing information from past time steps to be retained and used to analyse the present time step. The graphic also shows an unfolded perspective of the feedback loop for a specific value, k=3. The recurrent layer unfolds three times in this design, with the output of each layer fed back into the input of the succeeding layer. This unfolding process elucidates how the feedback loop enables the recurrent layer to collect and learn long-term connections within input sequences, demonstrating the RNN's potential for temporal comprehension and memory retention.

2.4 Univariate ARIMA Model

Shuang Cang and Nigel Hemmington (2010) say they are focused on properly estimating inbound expenditure to the UK based on the purpose of visit, which is classified as holiday, business, study, visit to friends or relatives, and miscellaneous. They use two well-known time series forecasting methods to do this: seasonal autoregressive integrated moving average (ARIMA) and Winters' multiplicative exponential smoothing. In addition, they employ the Nave 2 forecasting model as a reference point (Shuang Cang & Hemmington, 2010).

In this application, the ARIMA model is a univariate time series forecasting approach. It is made up of three major components: autoregressive (AR) terms, differencing (I) terms to establish stationarity, and moving average (MA) terms. To account for repeating trends throughout time, the seasonal ARIMA model, in particular, adds seasonal components. This approach is applied independently to each disaggregated type of incoming spending (e.g., holiday, business). It should be noted that the ARIMA model captures both trend and seasonality in the data, making it suited for time series data with distinct patterns (Shuang Cang & Hemmington, 2010).

According to their findings, the ARIMA model surpasses the WMES model in terms of predicting accuracy for specific types of visits, such as business. This suggests that, as compared to WMES, the ARIMA model delivers more accurate projections of spending patterns for specified objectives. It is crucial to note, however, that statistical superiority does not always transfer into higher predicting performance in all cases. The decision between ARIMA and WMES may be influenced by the features of the data being analyzed (Shuang Cang & Hemmington, 2010).

Furthermore, the authors emphasize that integrating predicting data from both the ARIMA and WMES models might be advantageous, particularly for applications such as business and miscellaneous, where each model has strengths. By using the capabilities of each model for different components of the data, an ensemble technique can give a more robust and trustworthy forecast. Finally, the research contributes to better tourist planning and policy-making in the UK by providing insights into which forecasting methodologies work best for certain sorts of inbound expenditure reasons (Shuang Cang & Hemmington, 2010).

2.4.1 Introduction to Univariate ARIMA Model

The univariate ARIMA (Autoregressive Integrated Moving Average) model is a popular time series forecasting approach that combines autoregression (AR), differencing (I), and moving average (MA) elements to model and predict future values of a single time series variable. Each ARIMA model component serves a distinct role in collecting and characterizing patterns in a time series (Hayes, 2023).

The Autoregressive (AR) component connects the present value of a time series to its prior values, expressing the reliance on earlier observations. The Integrated (I) component employs differencing to stabilise time series and eliminate trends or seasonality. This step assures ARIMA modelling stationarity. The Moving Average (MA) component is the model's connection between the current observation and previous error terms, and it captures short-term variations. The univariate ARIMA model's mix of these components makes it a flexible tool for predicting varied time series data, changeable by modifying the order of the AR, I, and MA components. ARIMA provides a formal framework for pattern collection and prediction, demonstrating strong in time series analysis and forecasting regardless of data complexity (Hayes, 2023).

2.5 Artificial Neural Networks (ANN)

Murat and Çuhadar (2020) from Süleyman Demirel University, used Artificial Neural Network (ANN), Exponential Smoothing, and Box-Jenkin methodologies to examine the critical significance of accurate tourist demand prediction in driving commercial choices in the tourism sector.

Machine Learning approaches, notably artificial neural networks (ANNs), have gained appeal in tourist forecasting due to their ability to capture complicated nonlinear correlations. Hu et al (2019) used an ANN model to forecast hotel occupancy rates in a metropolitan region. Their research revealed the ANN’s capacity to standard approaches in capturing detailed patterns in tourist data.

The research, which focuses on Croatia's incoming visitor flows, compares three forecasting approaches—Artificial Neural Network (ANN), Exponential Smoothing, and Box-Jenkins—in an effort to create the most accurate forecasting model possible. The study's performance metric is Mean Absolute Percentage Error (MAPE), and it covers the period from January 2005 to December 2019. The usefulness of a 12-lagged ANN model over more established statistical techniques is demonstrated by the results, which highlight its superiority. This expands the small area of monthly tourism demand forecasting using ANNs and shows their potential for success with the right design.

2.5.1 Introduction to ANN

Artificial Neural Networks are machine learning algorithms that are modelled after the human brain. In other words, just as neurons in the nervous system can learn from past data, an artificial neural network (ANN) may also learn from data and provide results in the form of predictions or classifications. ANNs are nonlinear statistical models that use a complicated interaction between inputs and outputs to identify a new pattern. These artificial neural networks are used for a range of tasks such as image identification, speech recognition, machine translation, and medical diagnosis (Team, 2021).

A major advantage of ANN is that it learns from sample data sets. The most typical use of ANN is for random function approximation. With these tools, one may arrive at cost-effective distribution options. ANN may also deliver output results based on sample data rather than the complete dataset. Because of their strong prediction powers, ANNs can improve existing data analysis approaches (Team, 2021).

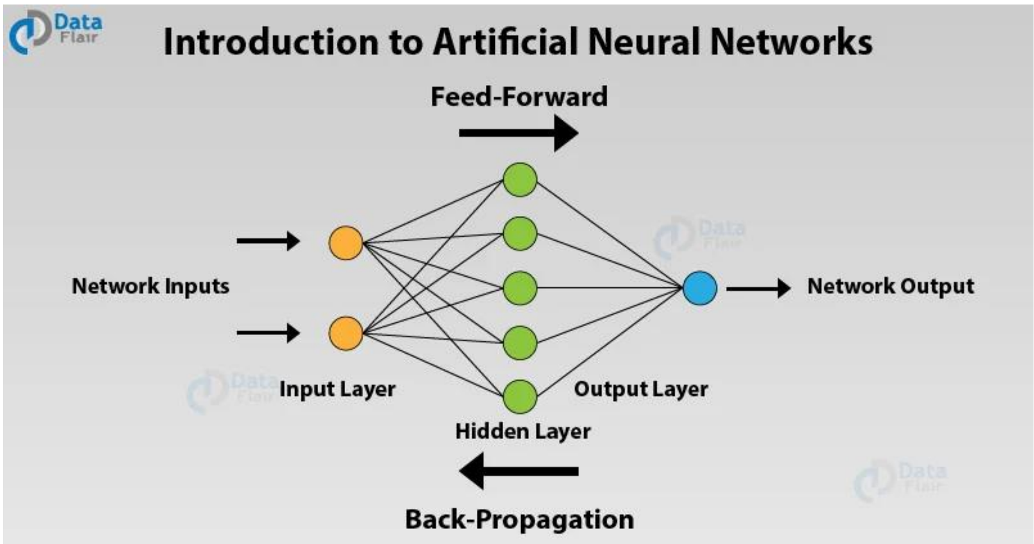


Figure 2.3: Artificial Neural Networks (Team, 2021)

In Figure 2.3, a feedforward artificial neural network is depicted, commonly used in tasks such as machine translation, natural language processing, and image identification. Comprising three layers—an input layer for external data, a hidden layer for computations, and an output layer for the network's output—neurons within each layer are interconnected, exchanging messages with weights indicating connection strength. During operation, input data flows through the network, with hidden layer neurons using nonlinear activation functions to produce weighted sums. This output is then processed in the output layer to generate the network's forecast. These neural networks excel in tasks like image classification and language translation, leveraging supervised learning through input-output pairs to adjust network parameters.

2.5.2 Elman’s Model of ANN

Elman's Model is a sort of recurrent neural network (RNN) with feedback connections that allow it to capture temporal relationships in data. It is useful for predicting visitor arrivals because it can account for the sequential character of time series data, such as historical tourist arrival data.

Elman's Model is particularly successful in capturing short-term relationships in time series data. It has a hidden layer that receives both the input data and the output from the previous time step, resulting in a feedback loop. This enables the model to store knowledge about previous inputs and utilize it to create predictions for the current time step.

Elman's Model's appropriateness for predicting tourist arrivals is determined by the features of the data and the specific forecasting objective. It is critical to properly preprocess the data, including managing missing values, scaling the data, and identifying essential input characteristics. Additionally, the model's architecture, such as the number of hidden nodes and the learning rate, should be tweaked to optimize speed.

2.6 Exponential Smoothing

George and Ashton (2012) propose the Vector Innovations Structural Time-Series (VISTS) framework, which is aimed to capture the dynamic character of time-varying seasonality in multivariate datasets. These models are based on exponential smoothing approaches but have been expanded to perform successfully in a multivariate setting. The study focuses on three VISTS model variants: the local level, local trend, and damped trend models, each enhanced with an additive multivariate seasonal component. These models are intended to give a thorough and accurate depiction of the underlying patterns in time series data (Athanasopoulos & de Silva, 2012).

Furthermore, machine learning approaches such as neural networks, support vector machines, and random forests are gaining favour in tourism demand forecasting. These approaches can capture deep data linkages and handle a greater variety of factors, allowing for more accurate forecasts. They do, however, frequently need larger datasets and greater processing resources. The focus of this study is on ARIMA because of its shown efficacy in modelling time series data, but it also acknowledges the expanding field of machine learning approaches for tourism forecasting, which may merit more study in subsequent studies. The initiative aims to improve the decision-making capacities of Montenegro's tourist industry through these techniques, creating resilience and adaptation in an ever-changing economic situation (Athanasopoulos & de Silva, 2012).

2.6.1 Introduction to Exponential Smoothing

Exponential smoothing is a strong and widely used time series forecasting technique that works particularly well with univariate data. This approach works by allocating decreasing weights to historical data points, with the most significant weightings going to the most recent observations. This property represents the idea that more recent data pieces are more likely to predict future patterns or trends than older ones. The exponential smoothing technique entails iteratively updating projected values based on weighted averages of prior observations and forecasted values. This method is very beneficial for dealing with time series data that exhibits seasonality, patterns, and erratic fluctuations (Simplilearn, 2023).

Exponential smoothing is useful for predicting over the next several days since it excels at producing short-term forecasts. However, it presents difficulties for longer predicting horizons because to its inherent reliance on current observations. Over time, relying on older data points with decreasing weights might make projections less credible. It is frequently advised to investigate alternate approaches for long-term predictions that can effectively capture and take into account changing patterns and trends (Simplilearn, 2023).

When underlying time series features, such as the smoothing factor or level and trend components, fluctuate gradually over time, exponential smoothing is useful. However, it could run into problems when data patterns suddenly and dramatically change. Practitioners might use versions like double or triple exponential smoothing, which incorporate more elements for enhanced trend and seasonality accommodation, to manage quick shifts. In conclusion, exponential smoothing achieves a compromise between simplicity and accuracy in forecasting univariate time series data, especially for short-term predictions and somewhat stable data patterns (Simplilearn, 2023).

2.7 Impact of External Factors

Morakabati et al (2012) study addresses the intricate and adaptable nature of the tourism system, recognizing its susceptibility to a wide array of internal and external influences, both planned and unforeseen, which can lead to substantial deviations from intended development trajectories. These influences encompass economic, environmental, health-related, and political factors, encompassing shifts in accessibility and safety perceptions for tourists. The research specifically delves into the travel risk perceptions and attitudes of a sample of UK residents contemplating travel to selected countries in the Middle East region, an area that saw approximately 4.5 million UK arrivals annually from 2007 to 2009. The study's findings, derived from a survey conducted between October 2010 and April 2011 with 394 respondents, shed light on the nuanced dynamics of travel decision-making in a context marked by multifaceted risk considerations, thereby contributing valuable insights to the broader discourse on tourism and its resilience in the face of unforeseen challenges.

The arrival of tourists in the UK may be predicted using a number of external variables. These comprise:

- Economic Indicators
- Meteorological Variables
- Social media big data
- Web search traffic
- Environmental Factors

2.7.1 Economic Indicators

Economic indicators are statistical measurements that are used to evaluate the state and performance of an economy overall. They offer essential insights on variables including economic expansion, inflation, unemployment rates, and consumer expenditure. Consumer Price Index (CPI), unemployment rates, and Gross Domestic Product (GDP) are typical economic indicators. These metrics support the informed decision-making of enterprises, investors, and regulators (Allen, 2010).

The UK's contribution to global GDP from travel and tourism is anticipated to increase by an average of 3% each year between 2022 and 2032, according to research by the World Travel & Tourism Council (WTTC) (World Travel & Tourism Council, 2022). In 2016, tourism supported 2.6 million employment and added £106 billion to the British economy (GDP) in England. England's tourism sector is a vibrant and vital contributor to the economy, generating

£100.8 billion annually, employing over 2 million people, and supporting thousands of businesses. It accounts for over 80% of the UK's total visitor economy (*The value of tourism in England* 2022).

2.7.2 Meteorological Variables

Meteorological variables wield substantial influence on UK tourism, shaping the timing and allure of visits. Variables like temperature, precipitation, sunlight hours, and seasonal fluctuations play a pivotal role. The UK's unpredictable weather can attract those seeking distinctive climatic experiences while deterring those averse to rain or cold. This interaction underscores the critical role of weather forecasts and climatic considerations in organizing and marketing UK travel.

Temperature is an important meteorological factor affecting tourism, a phenomenon that has been extensively studied in the literature. Generally speaking, higher temperatures are significantly linked to an increase in tourists, especially to places with sun and beaches. The direct link between temperature and visitor arrivals has been underlined in studies like the one by Matzarakis et al (2019). Warmer weather tends to draw tourists looking for outdoor recreation, water sports, and leisure in coastal areas (R.-Toubes et al, 2020).

The study by Hamilton et al (2003) offers insightful information about the effects of short-term climatic variability in the UK on domestic and international tourism demand. Rainfall is unquestionably another meteorological variable of great significance in the realm of tourism. Excessive rainfall, especially in the form of lengthy, intense downpours, can be harmful to tourist locations. It may result in problems like floods, landslides, delays in transportation, and other weather-related difficulties that can deter tourists from going to particular places (Palutikof, 2006).

Sunshine has a considerable impact on tourism, especially in sun-and-beach areas, luring visitors seeking an enhanced experience and outdoor activities (R.-Toubes et al, 2020). However, WWF-UK study warns of the effects of climate change on the UK tourism industry, forecasting changing weather patterns, heatwaves, varying precipitation, and sea level rise. These changes may interrupt travel seasons, posing erosion and flood dangers to coastal towns and beach tourists. Wildlife behaviour changes can provide a problem to nature-based tourism. To protect the UK's tourist sector and natural resources, urgent actions such as sustainable practises, climate coping techniques, and joint emission reduction are being emphasised (Viner & Agnew, 1999).

2.7.3 Social-Media Big Data

Big data from social media may be a potent tool for learning crucial information about UK tourism. Tourism organizations and businesses may spot hot travel locations, keep track of visitor sentiment, and adjust their marketing tactics by analysing the massive amounts of user-generated information on websites like Twitter, Instagram, and Facebook. In addition, examining geotagged postings and hashtags might assist in identifying popular tourist destinations and seasonal trends, facilitating more efficient resource allocation and infrastructure design. By quickly identifying and reacting to unfavourable patterns or new problems, social media data may also help with crisis management (*VisitBritain* 2023).

While social media data has the potential to enhance tourist statistics, relying solely on platforms like Facebook may be incomplete and inaccurate, as not all travelers use it consistently or disclose their location accurately. The Office for National Statistics (ONS) emphasizes the need for a comprehensive approach, integrating social media data with surveys, visitor counts, and hotel data for robust and reliable tourist statistics. This multifaceted approach enables a more holistic understanding of tourist demographics, behaviours, and preferences, supporting informed industry decision-making (Benedikt, 2019).

Instagram, Twitter, and YouTube are key platforms for showcasing travel destinations and experiences through captivating photos and authentic user-generated content, influencing traveller preferences significantly. The integration of social media with smart travel locations, including Snapchat and Facebook Live, enables real-time trip documentation and dynamic narratives. Smart cities leveraging location-based services and personalized suggestions enhance the overall visitor experience. The synergy of big data and social media creates seamless, personalized travel experiences tailored to individual likes and interests, underscoring the crucial role of digital initiatives in the modern travel and tourism sector.

2.7.4 Web Search Traffic

Tourism demand and arrival projections have been greatly improved by tracking online search activity, particularly with tools like Google Trends. The MDPI article emphasises the importance of analysing internet users' search patterns to get knowledge on upcoming travel trends and destination preferences. Tourism organisations and companies may make use of this data as individuals frequently use search engines to find information about prospective vacation spots, lodgings, and activities. This information can give them real-time insights into the interests and favourite locations of tourists (Gricar et al, 2021).

Insights into search inquiries for certain regions and travel-related phrases may be found with great value using Google Trends. In order to help stakeholders in the tourist sector with resource allocation, campaign optimisation, and peak season planning, this information makes it easier to identify the periods of time when interest in specific locales or attractions peaks. A more thorough method of predicting tourist demand may be achieved by combining online search traffic analysis with additional data sources, such as social media and past tourism statistics. The industry is able to quickly adjust to changing customer tastes and market conditions because to this technique. Online search traffic analysis has effectively evolved into a crucial instrument for identifying and resolving visitor patterns, eventually improving the sector's overall planning and management (Gricar et al, 2021).

2.8 Seasonality and Trends

Luis A. Gil-Alana, José L. Ruiz-Alba, and Raquel Ayestarán (2019), says in "UK Tourism Arrivals and Departures: Seasonality, Persistence, and Time Trends," the authors investigate key issues related to the seasonality, persistence, and time trends in the series depicting the number of UK arrivals and departures using fractional integration techniques. In comparison to integer degrees of differentiation, this technique provides more flexibility for investigating the effects of shocks on series. Significant temporal trends, showing a long-lasting influence of shocks, strong persistence falling within the fractional range, and the relevance of seasonality are among the key findings. By removing seasonality, temporal patterns are eliminated but data relevance is maintained. The study's findings have political implications and are important to the UK tourism industry (Gil-Alana, 2019).

According to VisitBritain's July 2023 tourist estimate, inbound travel to the UK is expected to grow. The UK saw 31.2 million incoming visitors in 2022, which represents a 24% drop from 2019. This was after the COVID-19 pandemic caused a significant dip in visitation over the previous two years. The overall amount spent by visitors was £26.5 billion, a decline of 7% in nominal terms (or 17% when adjusted for inflation). With 37.5 million visitors, or 92% of 2019 levels, and estimated expenditure of £30.9 billion, or 109% of 2019 levels in nominal terms (90% when adjusted for inflation), the prediction for 2023 predicts a strong rebound (*tourism forecast* 2023).

Forecasting UK tourism necessitates taking into account both the positive linear trend of sector development and the considerable influence of seasonality, which introduces cyclical swings in visitor patterns. Forecasting must be complex, taking into account time-dependent variables for year-round travel demand. The inbound tourist prediction for 2022 emphasises market trends, particularly the predicted increase in European tourists. Advanced methodologies such as multi-series structural time series modelling and intra-day and inter-day trend tactics are critical for reliable seasonal demand forecast. Using comprehensive patterns and historical data improves prediction accuracy, allowing for effective resource allocation and dynamic marketing plans in response to changing UK tourism (Tourism Forecast 2023).

2.9 Hybrid Models

Hybrid models that combine classic time series approaches with machine learning techniques have also been investigated by researchers. In the UK, forecasting tourism can benefit from hybrid models.

Wang and Liu (2015) suggested a hybrid model for forecasting tourism demand that combines ARIMA and support vector regression (SVR). Their findings indicated that the hybrid technique produced more accurate forecasts than stand-alone strategies. According to their research, utilizing this hybrid method regularly produced forecasts that were more accurate than those produced by using either ARIMA or SVR alone. This shows that integrating time series analysis with machine learning strategies may be useful for improving projections of tourism demand.

Nurhaziyatul A. Yahya, Ruhaidah Samsudin, and Ani Shabri (2017) study proposes a hybrid forecasting model based on monthly visitor arrivals from Singapore and Indonesia to Malaysia and an artificial neural network (ANN). Given the increasing global significance of the tourism industry, particularly in the era of Big Data, there's a continuous exploration of innovative approaches for accurate and timely tourism demand forecasting. This study introduces a novel method that combines Empirical Mode Decomposition (EMD) and Artificial Neural Network (ANN) models. The approach involves using Intrinsic Mode Functions (IMF) generated through EMD, with certain IMFs refined through a trial-and-error process known as decomposition. Subsequently, an ANN model is employed to predict both the decomposed and the remaining IMF components. Finally, the forecasted results from each component are aggregated to create a comprehensive tourism time series forecast. The study evaluated the model's performance using well-established measures like RMSE and MAPE, and the findings indicated that this proposed model outperformed both individual ANN and EMD-ANN models, highlighting its potential for enhanced tourism arrival predictions (A et al, 2017).

Wu et al (2020) say that a critical need for precise and timely forecasting of visitor demand in the tourism business. Because of the scarcity of data, most study in this field has tended to concentrate on quarterly or monthly data. In order to get over these drawbacks, the authors developed a brand-new hybrid technique called SARIMA + LSTM, which combines the seasonal autoregressive integrated moving average (SARIMA) and the long short-term memory (LSTM) to forecast daily visitor arrivals in Macau SAR, China. In addition to enhancing the SARIMA model's predictive strength and further lowering prediction errors, the LSTM model is renowned for its ability to capture long-term dependencies in time series data (Wu et al, 2020).

2.10 Data Granularity and Accuracy

song et al (2017) say the necessity of high-frequency data for reliable tourist forecasting. They predicted visitor counts at a tourist destination using hourly tourism data, demonstrating the power of fine-grained data in capturing short-term variations. Their study used hourly tourism data to anticipate the number of visitors to a certain site. This study emphasized the significance of having access to comprehensive, real-time information for more accurate and responsive forecasting in the tourism sector. It also highlighted the efficiency of fine-grained data in collecting and properly modelling short-term fluctuations in visitor arrivals. Data granularity and accuracy are important factors in predicting model effectiveness.

Zhang et al (2020) propose in the context of tourism demand forecasting, traditional models face challenges when dealing with extensive search intensity indices as tourism data. To address these challenges and enhance modelling accuracy in AI-based tourism demand forecasting, this study introduces a novel approach called "Tourism Demand Forecasting: A Decomposed Deep Learning Approach." Deep learning models used for predicting tourism demand often become overly complex and prone to overfitting due to limited data volumes and additional explanatory variable requirements. The proposed approach effectively tackles these issues by utilizing a decomposition method that enhances the accuracy of both short- and long-term AI-based forecasting models without necessitating additional data. In conclusion, this study provides a methodological contribution by alleviating overfitting concerns and presenting a highly accurate deep-learning method for AI-based tourism demand modeling (Zhang et al, 2020).

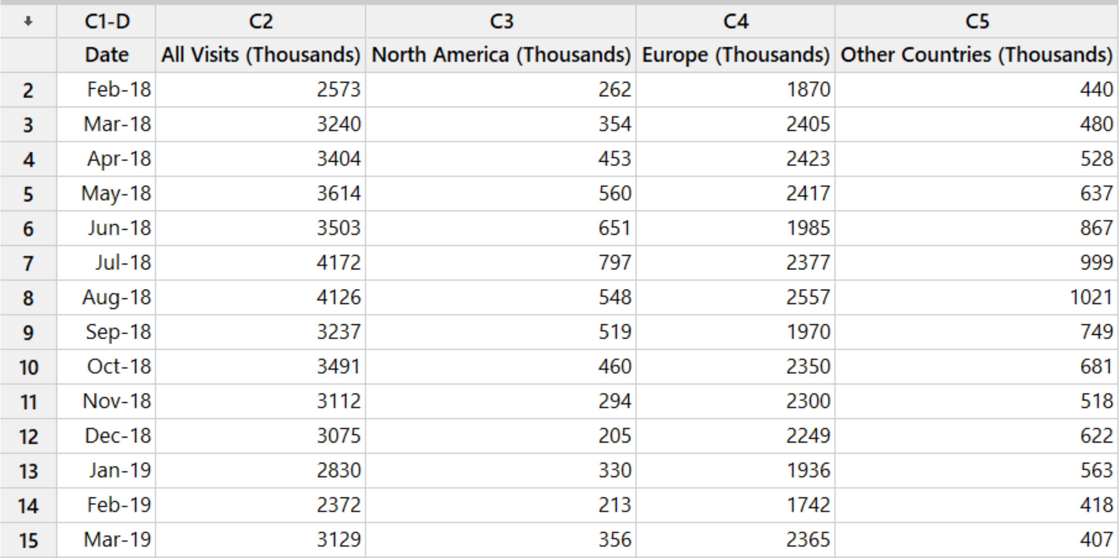
Rob Law, Gang Li, Davis Fong, and Xin Han investigate the use of deep learning in predicting monthly Macau visitor arrival numbers, contributing to the expanding landscape of tourism demand forecasting. Notably, the study demonstrates that deep learning outperforms known approaches such as support vector regression and artificial neural networks. The study offers light on intricate relationships impacting forecasting of visitor demand and arrival amounts, emphasising the relevance of finding relevant properties within the deep network architecture. This paper is part of the larger Annals of Tourism Research Curated Collection on Tourism Demand Forecasting, highlighting the field's expanding relevance and the need for creative methodologies to improve forecasting precision.

Chapter 3 - Data Collection

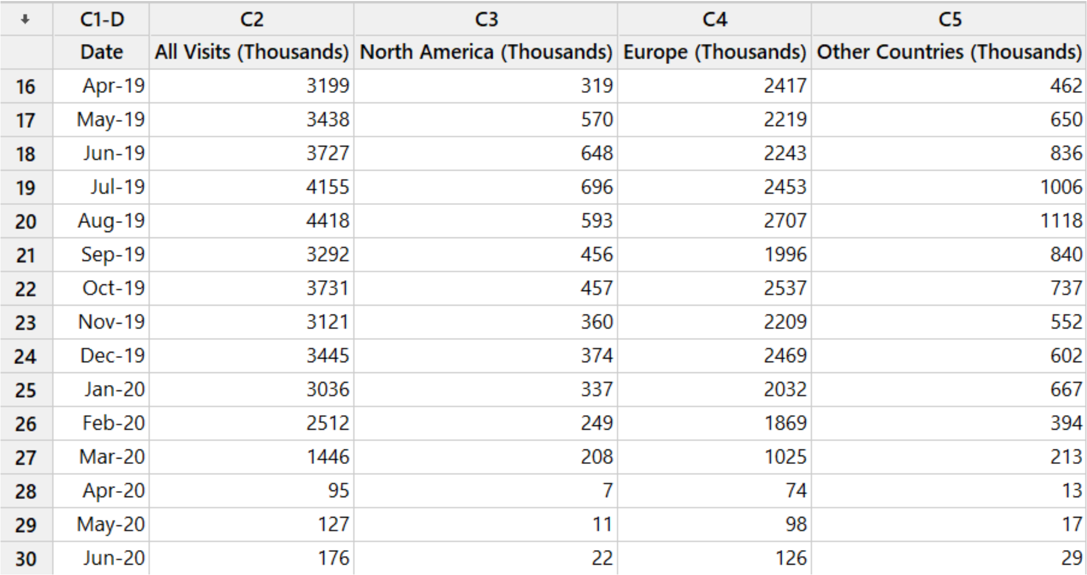
3.1 Introduction

It was decided that this project's phase would be utilized to gather the information needed to achieve the project's goal. The dataset was obtained from the UK Office for National Statistics

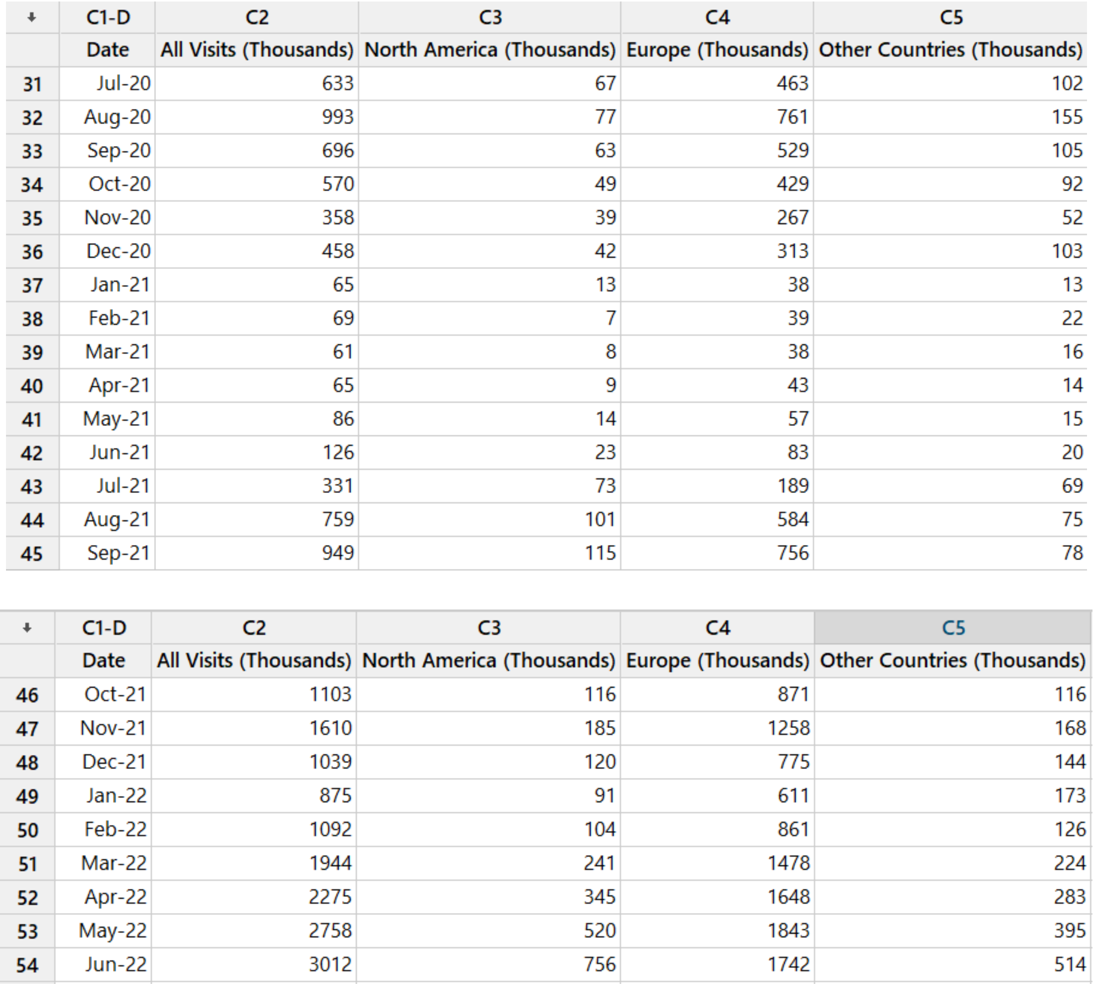
(ONS) government website, and it encompasses data on overseas residents' visits to the United Kingdom from January 2018 to December 2022. This comprehensive time frame of five years provides valuable insights into the trends and patterns of international visits to the UK, enabling researchers and analysts to assess the impact of various factors, such as economic conditions, tourism policies, and global events, on the influx of visitors from abroad. Such data is instrumental for policymakers, businesses, and tourism stakeholders in making informed decisions and strategies related to tourism and international travel. The full list of datasets shown by figure 3.1.



↓	C1-D	C2	C3	C4	C5
	Date	All Visits (Thousands)	North America (Thousands)	Europe (Thousands)	Other Countries (Thousands)
2	Feb-18	2573	262	1870	440
3	Mar-18	3240	354	2405	480
4	Apr-18	3404	453	2423	528
5	May-18	3614	560	2417	637
6	Jun-18	3503	651	1985	867
7	Jul-18	4172	797	2377	999
8	Aug-18	4126	548	2557	1021
9	Sep-18	3237	519	1970	749
10	Oct-18	3491	460	2350	681
11	Nov-18	3112	294	2300	518
12	Dec-18	3075	205	2249	622
13	Jan-19	2830	330	1936	563
14	Feb-19	2372	213	1742	418
15	Mar-19	3129	356	2365	407



↓	C1-D	C2	C3	C4	C5
	Date	All Visits (Thousands)	North America (Thousands)	Europe (Thousands)	Other Countries (Thousands)
16	Apr-19	3199	319	2417	462
17	May-19	3438	570	2219	650
18	Jun-19	3727	648	2243	836
19	Jul-19	4155	696	2453	1006
20	Aug-19	4418	593	2707	1118
21	Sep-19	3292	456	1996	840
22	Oct-19	3731	457	2537	737
23	Nov-19	3121	360	2209	552
24	Dec-19	3445	374	2469	602
25	Jan-20	3036	337	2032	667
26	Feb-20	2512	249	1869	394
27	Mar-20	1446	208	1025	213
28	Apr-20	95	7	74	13
29	May-20	127	11	98	17
30	Jun-20	176	22	126	29



↓	C1-D	C2	C3	C4	C5
	Date	All Visits (Thousands)	North America (Thousands)	Europe (Thousands)	Other Countries (Thousands)
31	Jul-20	633	67	463	102
32	Aug-20	993	77	761	155
33	Sep-20	696	63	529	105
34	Oct-20	570	49	429	92
35	Nov-20	358	39	267	52
36	Dec-20	458	42	313	103
37	Jan-21	65	13	38	13
38	Feb-21	69	7	39	22
39	Mar-21	61	8	38	16
40	Apr-21	65	9	43	14
41	May-21	86	14	57	15
42	Jun-21	126	23	83	20
43	Jul-21	331	73	189	69
44	Aug-21	759	101	584	75
45	Sep-21	949	115	756	78
46	Oct-21	1103	116	871	116
47	Nov-21	1610	185	1258	168
48	Dec-21	1039	120	775	144
49	Jan-22	875	91	611	173
50	Feb-22	1092	104	861	126
51	Mar-22	1944	241	1478	224
52	Apr-22	2275	345	1648	283
53	May-22	2758	520	1843	395
54	Jun-22	3012	756	1742	514

Figure 3.1 Visits to the UK dataset

The Table 3.1 dataset comprises five columns representing monthly data: "Date," "All Visits (in Thousands)," "North America (in Thousands)," "Europe (In Thousands)," and "Other countries (In Thousands)." The "Date" column likely denotes the time period for each data point. The "All Visits" column likely records the total number of visits in thousands for the given month. The subsequent three columns, "North America," "Europe," and "Other countries," likely break down the total visits into geographical regions, each expressed in thousands. This dataset appears to track monthly visitor statistics, allowing for the analysis of visit trends and geographical distribution over time.

3.2 Collection Phase

A comprehensive Google search was conducted to locate a pertinent dataset related to tourism during the data collection phase. Examination of official government travel websites, such as Visit Britain and the ONS, was undertaken. Following meticulous evaluation, the dataset accessible on the ONS leisure website was deemed the most suitable and relevant source for the study's needs. The dataset was made available in multiple formats, including CSV and XML, along with a complementary chart presented in image form. Below, the chart is displayed and its content is elucidated.

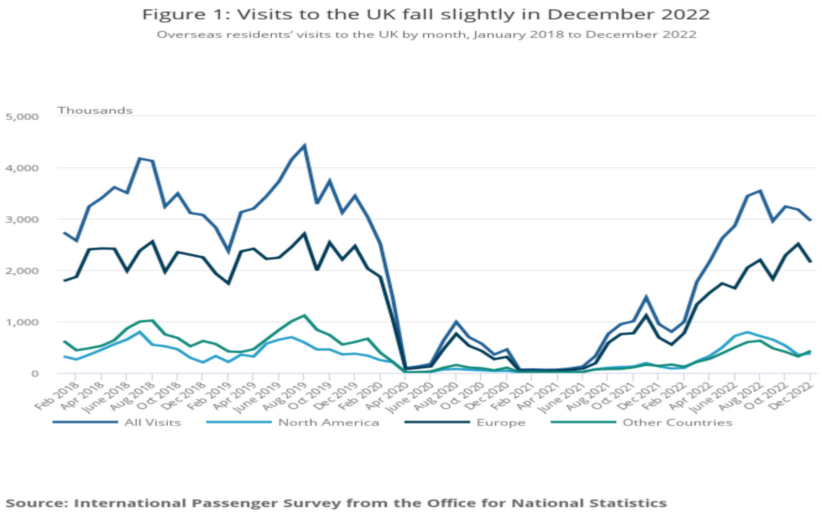


Figure 1: Visits to the UK fall slightly in December 2022

Overseas residents' visits to the UK by month, January 2018 to December 2022

Source: International Passenger Survey from the Office for National Statistics

Figure 3.2: Visits to the UK fall slightly in December 2022 (Osborn, 2023)

Figure 3.2 shows the number of overseas residents' visits to the UK by month, from January 2018 to December 2022. The chart shows that the number of visits has been rising steadily since the start of 2022, but fell slightly in December 2022.

Overseas visitors made 2.8 million trips to the UK in December 2022, up 19% from December 2020 but down 3% from December 2021. This indicates that although the number of visitors to the UK is slowing down at the end of the year, it is still rebounding from the pandemic. The graph also demonstrates that North America and Europe account for the majority of foreign tourists to the UK. In December 2022, 1.5 million visitors from Europe and 1.0 million from North America came to the UK. This shows that even in the winter, travellers from these regions still favour the UK as a vacation destination.

The data indicates that while visits to the UK are still increasing, they are slowing down as the year draws to a close. The chilly weather, the current cost-of-living issue, and the ongoing COVID-19 pandemic are all likely contributing causes to this. The UK tourism business will benefit from the fact that the number of visitors is still higher than it was before the outbreak.

3.3 Implementation

The initial step in the execution phase of this tourist data analysis project is data pretreatment with tools such as the Google Collab and Minitab. The dataset obtained from the ONS in CSV or XML format must be loaded into Google Collab with Python code, and then imported into Minitab exclusively for ARIMA model experimentation and evaluation for further analysis. It is critical to check for any errors such as missing numbers, outliers, or discrepancies during this procedure. Fortunately, there is no missing data according to the information supplied, which streamlines the first data compilation. However, maintaining data consistency and correctness throughout the dataset is critical, and as stated, there are no outliers in the data.

Next, It may use time series analysis techniques to uncover patterns and trends in the arrival of foreign tourists in the UK. In order to generate summary measures like means, standard deviations, and variances over various time intervals, including monthly and yearly intervals, this study uses descriptive statistics. It may develop visualizations like line plots or time series charts to adequately depict these tendencies. By developing these visualisations with Google Collab and Minitab, It may focus on enhancing the clarity and customisation of the chart that depicts the number of visits from international residents from January 2018 to December 2022.

Time series decomposition techniques, such as additive or multiplicative decomposition, may be used in a Google code collaboration to evaluate the seasonality and trend components in the data. Also, Minitab have by isolating the underlying patterns from the noise, this would make it easier to understand the changes in UK travel. Forecasting algorithms like as ARIMA (Autoregressive Integrated Moving Average) are used in Google Collab and Minitab to provide future projections based on existing data. Only in Google Collab are exponential smoothing and ANN used. The factors mentioned in the dataset description, such as the observed increase in tourism from the beginning of 2022 and the modest decline in December 2022, may be accounted for by these models.

Furthermore, knowing where the majority of international visitors originate from—North America and Europe—indicates the potential usefulness of segmenting and examining these geographical trends further. Google Collab and Minitab can help with this by allowing analysts to pick and examine data subsets based on location or other relevant factors. This will provide a more targeted approach to studying the causes driving tourism from certain places.

In summary, the implementation step involves data import, preprocessing, visualisation, time series decomposition, forecasting, and maybe regional analysis, all of which are carried out in Google Collab and Minitab. The objective is to extract meaningful insights from the tourism dataset to aid in decision-making, marketing strategies, and policy recommendations for the UK tourism industry, particularly focusing on the observed recovery trends and the factors influencing them.

Chapter 4 – Experimental Methodology

4.1 Introduction

This chapter focuses on exploring the theories and procedures used for conducting experiments while also providing explanations for the selection of the experiments. Navigating the complexities of model selection is a crucial part of this endeavour. Examining standard deviations, correlations, and the applicability of models like ARIMA, Elman's ANN, and exponential smoothing are all included in this context. The primary objective is to explore these three time series approaches and anticipate tourism arrivals.

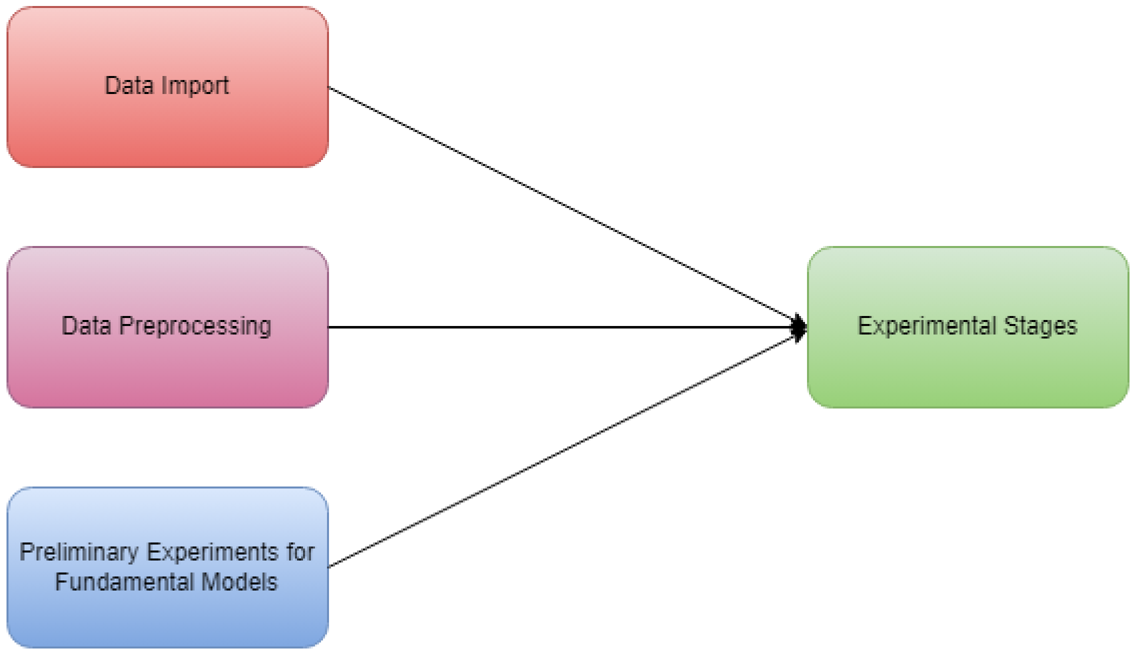


Figure 4.1: Time series model flowchart

In Figure 4.1, depicts the key phases of data preparation, which include turning raw data into a suitable format for analysis. These processes include data import, which involves manually or automatically transferring data from its source to a processing environment. Data Preparation, the most time-consuming phase, including cleaning and modifying data to guarantee consistency and readiness for analysis, addressing tasks such as eliminating duplicates, resolving missing values, and standardizing data formats. Following that, there are experimental stages in which various data processing and analysis techniques are evaluated to discover the most successful ones. In addition, even if it is optional, a period of preliminary tests might be useful for learning more about the data and identifying potential problems. These three fundamental stages serve as a framework for data preparation efforts, even if the precise procedures may vary based on the dataset and its goal.

4.2 Data Preprocessing

Preprocessing entails taking meticulous steps to ensure that the dataset is clean, correct, and ready for in-depth study. The ONS has optimised the dataset by removing headers from the CSV file to improve overall speed. This modification attempted to streamline the dataset structure by removing unnecessary information, resulting in a more efficient and simple representation. The new dataset now exposes information with distinct columns acting as primary identifiers, removing the old format where headers might have added complexity. This restructure not only enhances data processing efficiency but also allows for a more seamless study of the dataset since it aligns with a streamlined and clearly interpretable column-and-row structure.

In the context of a tourism data analysis project, once the dataset has been successfully imported, the next essential task is to scrutinize its quality. In order to identify and address any potential problems that could compromise the validity and accuracy of the future studies, this investigation entails a thorough assessment.

One primary aspect of this data quality check is the identification of missing values within the dataset. Missing data can seriously compromise the accuracy of the study, sometimes producing biased or insufficient results. In this specific case, it's reassuring to note that the dataset obtained from the ONS has been confirmed to contain no missing values, simplifying this particular aspect of the data quality check. Outlier detection and general value distribution. This is helpful in assessing if further relevant features should be created. Preprocessing will then be completed.

4.3 Preliminary Analysis for Base models

As previously stated, approach includes both machine learning and deep learning techniques. The dataset will be subjected to testing in line with the goals of the research. Throughout the testing process, all models will be treated the same to preserve uniformity. In the first portion of the experiment, three time series forecasting models were implemented: univariate ARIMA, Elman's ANN, and Exponential Smoothing. These will be utilised for predicting as well as determining whether there is a substantial variation in the performances of the three models.

4.3.1 ARIMA Model

The ARIMA model was already described in Section 2.4, and stationary data are required for its implementation. Certain limitations in the field of time series analysis call for careful assessment of the data. First, time series analysis (TSA) has difficulties with missing values, necessitating careful management, similar to other modelling techniques. Second, the modelling approach is constrained by the assumption of linearity in the relationships between the data points. Additionally, although useful, the demand for data transformations adds a degree of computational expenditure. Last but not least, it's interesting that time series models mostly work with univariate data, highlighting the necessity of a targeted strategy when working with multivariate datasets. In conclusion, for a thorough and perceptive time series analysis, identifying and resolving these constraints is essential.

In order to check the series of data, first checked the time series data in a line plot. That is shown below.

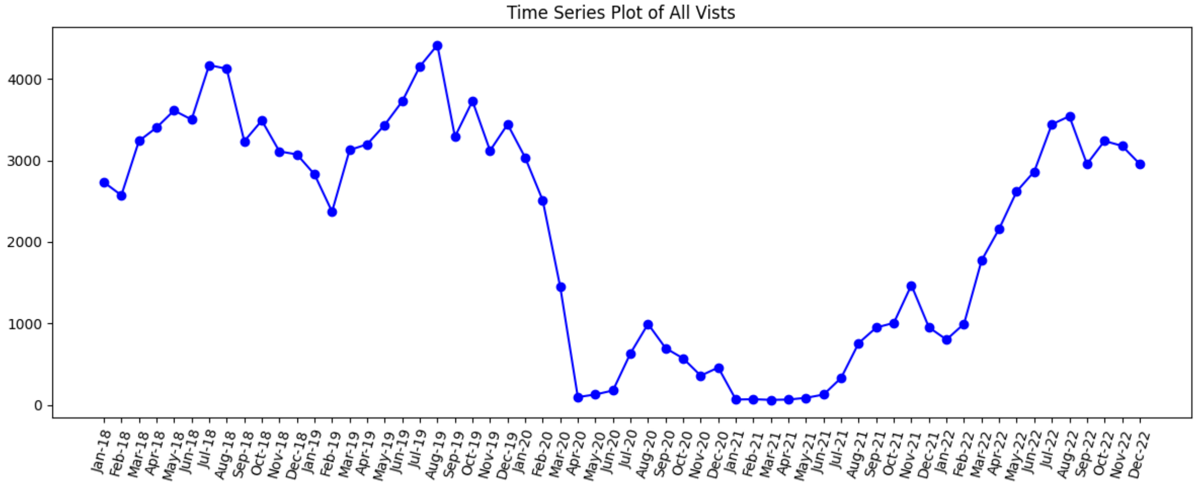


Figure 4.2: Time series plot of All visit column

Figure 4.2 illustrates a time series plot of UK tourism patterns from 2018 to 2022, demonstrating a distinct seasonal cycle, with summer months (June-August) experiencing the biggest peak in visitors and winter months (December-February) seeing the lowest. Furthermore, a major general reduction in visitation is observed between 2020 and 2022; this phenomenon is most likely attributable to the widespread effects of the COVID-19 pandemic.

These examples further explain the plot's structure:

- **Seasonal Variation**: Summertime sees a spike in tourism thanks to things like good weather, school breaks, and cultural activities. Tourists are drawn to the nicer weather and longer daylight hours in the summer, and families have more opportunity to travel during the concomitant school holidays.
- **Decline During the Pandemic**: Between 2020 and 2022, there is a noticeable drop in tourists, which is mostly attributable to the COVID-19 pandemic's effects on travel across the world. owing to limitations put in place by many nations, international travel became difficult and dangerous, and individuals were discouraged from travelling abroad owing to concerns about getting the virus. Tourism suffered a great deal during this recession.

It may be easier to understand the characteristics of the time series data if it is divided into three parts. It's referred to as decomposing. The trend, seasonality, and residual (or irregular) factors make up the decomposition's three primary parts. The residual component takes into account unexplained variances, seasonality captures recurrent patterns, and the trend indicates the data's long-term direction. The creation of a more precise and understandable ARIMA model is facilitated by analysts' ability to extract these individual components from the time series and identify and analyse them independently. The plots of data from Minitab and Google Collaboration are shown here.

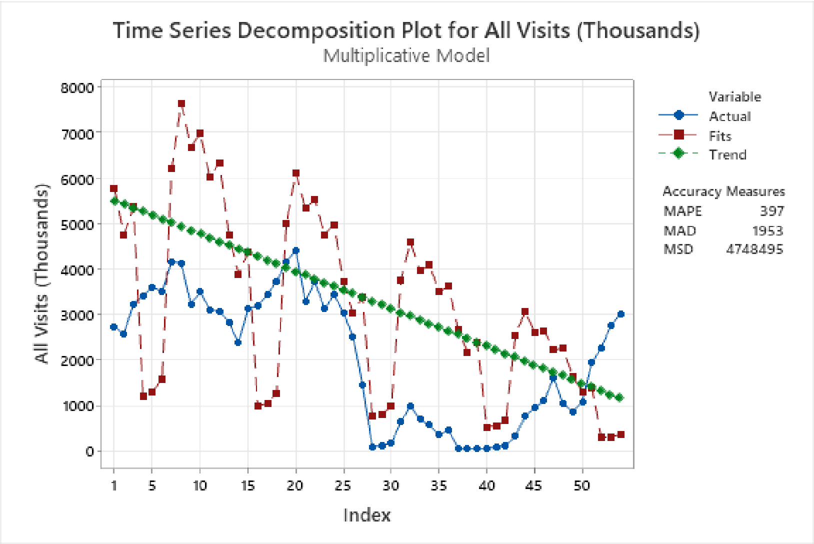


Figure 4.3: Time series Decomposition plot

Figure 4.3, The decomposition plot of the UK tourism time series shows a constant drop in the number of visitors through time, implying an annual loss on average. On the figure, a dashed line represents the decreasing trend.

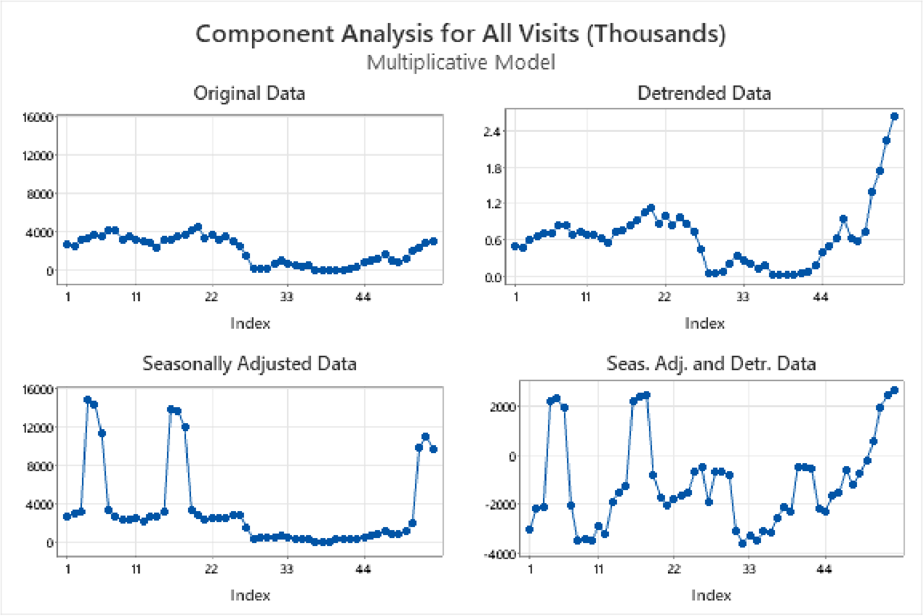


Figure 4.4: Time series Component Analysis plot

Figure 4.4, The component analysis plot of the UK tourism time series displays a unique seasonal pattern, with a peak in the number of visitors occurring during the summer months (June, July, and August) and a trough occurring during the winter months (December, January, and February). A small decreasing trend is also seen over time, showing a slight decrease in the overall number of visitors in recent years.

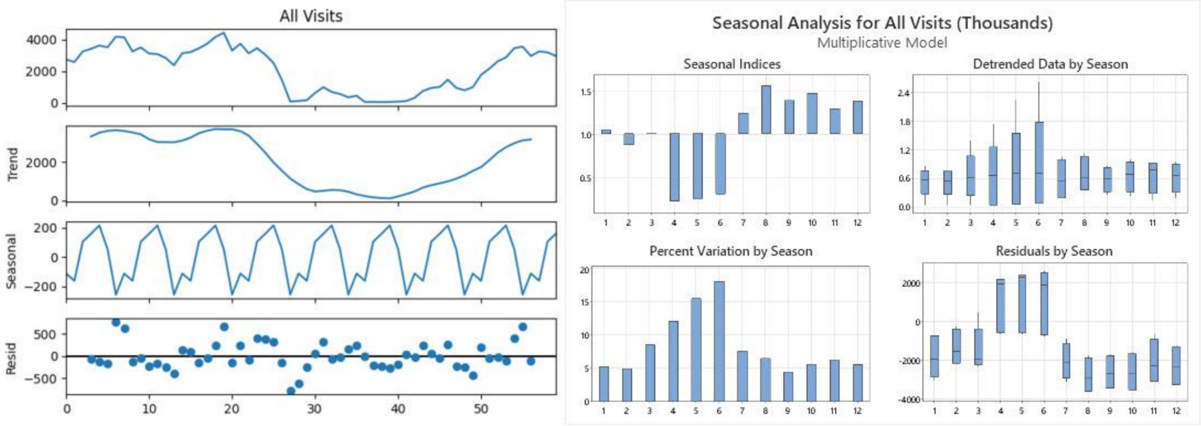


Figure 4.5: Time series Seasonal Analysis and decompose plot

Figure 4.5 depicts the seasonal analysis and breakdown plot of the UK tourism time series, emphasising the seasonal dependency of tourist numbers. Breaking down the elements:

- **Trend** T(t): The long-term trend shows a minor reduction in overall visitors in recent years, which may have been impacted by variables like rising travel expenses, the increasing popularity of substitute locations, or economic downturns.

- **Seasonal** S(t): The recurrent pattern indicates summer peaks and winter troughs. Given the warmer weather and longer daylight hours, summer is the preferred time to travel.
- **Residual** R(t): Unpredictable swings in the time series, attributable to elements like bad weather, tough times economically, or unstable political situations.

It is also obvious that the data is not stationary. Some year-over-year rises and declines are depicted in the plot.

Time series data may be divided into two types: stationary and non-stationary. Stationarity is an important attribute since some models function best with stationary data. Time series data, on the other hand, is frequently non-stationary. As a result, It must grasp how to recognise non-stationary time series and alter them using various approaches such as differencing (Studio, 2022).

Visualisation is the most straightforward way to determine if data is stationary. Figure 4.2 shows that it is not stationary. Other techniques for determining whether or not it is stationary include the ADF approach and the unit root method. Here, the ADF approach is used. To make it stationary, used natural log to convert the data from all visits to log data. ADF was examined again, but it was still not stationary. As a result, the differencing approach was utilised to render it stationary. This is a basic transformation of the series into a new time series that It utilise to eliminate the series' dependency on time and stabilise the mean of the time series, reducing trend and seasonality. Stationary plot is show below in Figure 4.6:

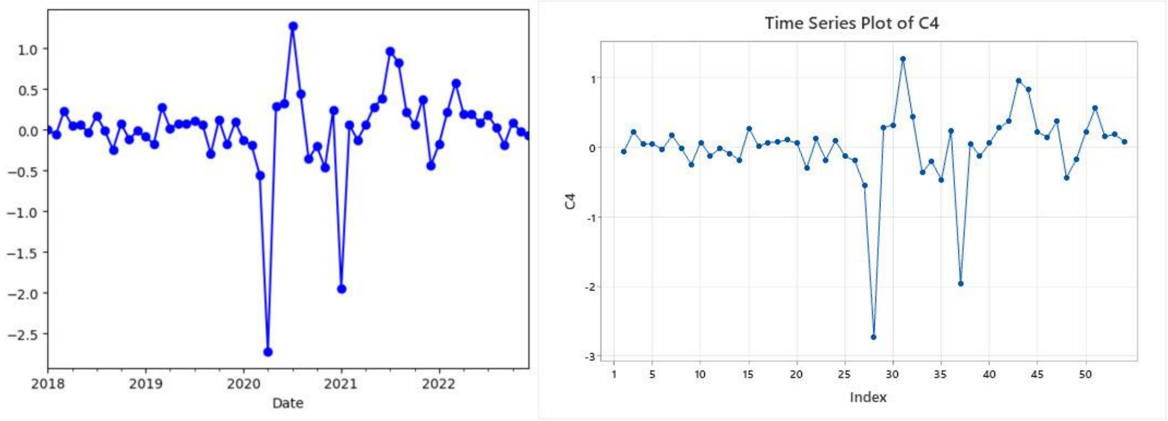
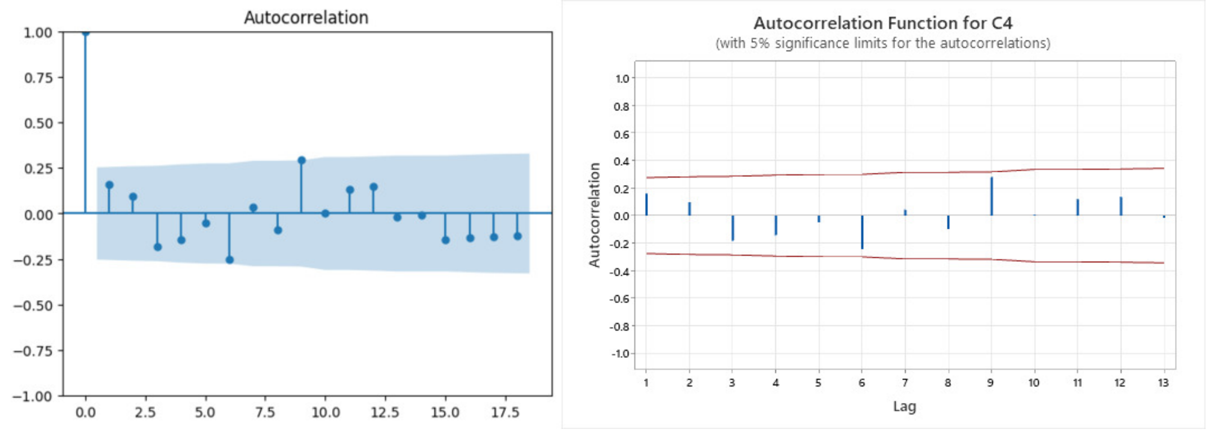

Time Series Plot of C4

Figure 4.6: Time series plot of stationary data from google collab and Minitab

Autocorrelation is a statistical term that assesses the degree of correlation between a time series and its lagged values. In basic terms, it quantifies how a variable is connected to its own historical values. A positive autocorrelation means that high values tend to follow high values and low values tend to follow low values over time. A negative autocorrelation, on the other hand, shows an opposite link.

ACF and PACF are time series analysis statistical tools used to detect and examine the correlation pattern within a dataset. The ACF of a stationary dataset is presented below:


Autocorrelation

Autocorrelation Function for C4

(with 5% significance limits for the autocorrelations)

Figure 4.7: ACF of stationary data from google collab and Minitab

Figure 4.7 shows the autocorrelation function (ACF) of UK tourism data shows a positive relationship between the number of tourists in one period and those in the next one. Effectively, a larger number of tourists visiting the UK in a certain time tends to correspond to a higher possibility of more tourists visiting the UK in the next period.

The ACF, in essence, assesses the statistical correlation between a time series and its previous values, revealing patterns of reliance across time. It accomplishes this by comparing time series values at multiple time points, such as the lag-1 autocorrelation, which contrasts values at time t with those at time t-1.

The autocorrelation at various delays is illustrated in an ACF plot. Given that the value of a time series at time t exactly correlates with itself, the lag-0 autocorrelation is always 1. The ACF figure shows that the UK tourist data exhibits positive autocorrelation for delays up to about 12 months. This means that the number of visitors visiting the UK at a given period corresponds favourably with the number of tourists in the following 12 months.

Autoregression (AR) is a statistical model that predicts future values of a time series based on its lagged values. Time series data is frequently forecasted using AR models. The number of delayed values used to estimate the future value of a time series is the order of an AR model.

The AR (1) model is the most basic AR model. It predicts the future value of the time series based on its delayed value.

Ordinary least squares (OLS) may be used to estimate the AR (1) model. Once estimated, the model may be used to anticipate future values of the time series. An AR (1) model may be used to estimate the autoregression from the autocorrelation function of UK tourist data. The estimated AR (1) model for UK tourist data is as follows:

$$y\_t = 1.02 + 0.8y\_\{t - 1\} + e\_t \tag{4.1}$$

It implies that the number of visitors visiting the UK during one time is positively connected to the number of tourists visiting the UK during the previous period. The lagged variable has a coefficient of 0.8, which suggests that a 1% rise in the number of tourists visiting the UK in one period is related with a 0.8% increase in the number of tourists visiting the UK in the next period.

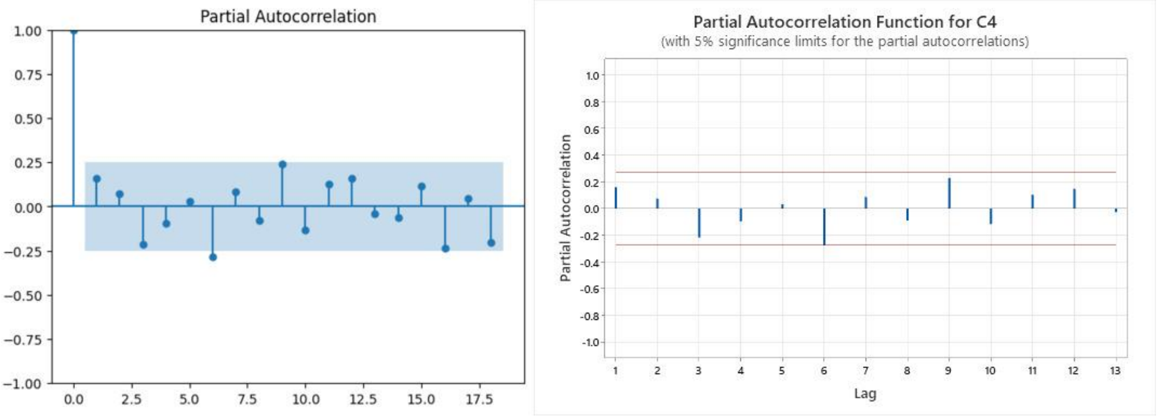


Figure 4.8: PACF of stationary data from google collab and Minitab

The partial correlation of UK tourism statistics is a measure of the relationship between the number of tourists visiting the UK in one period and the number of tourists visiting the UK in the next period while adjusting for the influence of other factors. Partial correlation is frequently used to determine causal links between variables.

The partial correlation function (PACF) of UK tourism data demonstrates that the partial correlation between the number of tourists visiting the UK in one period and the number of tourists visiting the UK in the next period is positive and significant for delays of up to roughly 12 months. This suggests that even after allowing for the influence of other variables, the number of visitors visiting the UK in one period is positively associated with the number of tourists visiting the UK in the next 12 months.

The PACF figure depicts the partial autocorrelation at various delays. Because the value of a time series at time t is not associated with itself after correcting for the impact of other variables, the lag-0 partial autocorrelation is always zero.

A MA model is a statistical model that predicts future values of a time series using the lagged values of the error component from an AR model. Time series data are frequently forecasted using MA models. The number of delayed values of the error term used to forecast the future value of the time series is the order of an MA model. The MA (1) model is the most basic MA model. It predicts the future value of the time series using the delayed value of the error term. The estimated MA (1) model for UK tourist data is as follows:

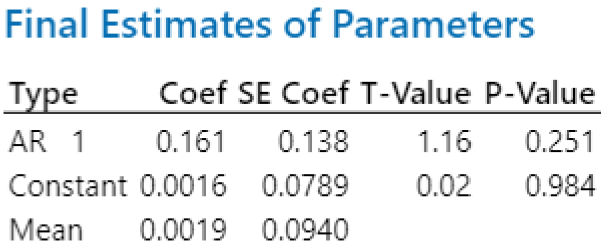
$$y_t = 1.02 + 0.8y_{\{t-1\}} - 0.2\mathrm{e\_\{t-1\}} \qquad (4.2)$$

According to this model, the error term from the AR (1) model is inversely connected to the error term from the prior period. This indicates that if the AR (1) model's error term is positive in one period, it is more likely to be negative in the next.

To make the process easier and ensure optimal model selection, it used the "auto_arima" function in the Google Collab notebook. The best ARIMA models are automatically identified using this function, saving users' time and effort throughout the model selection process.

Installed the "pmdarima" library on Google collab notebook to improve the usefulness and capabilities of auto ARIMA. This Python library is a strong tool for time series analysis and forecasting, with extra features and optimisations for auto ARIMA model selection. By integrating these techniques and platforms, was able to conduct extensive time series analysis, guaranteeing that the chosen ARIMA models were well-suited to the data and could give solid forecasts.

Additionally, a comparable study was carried out using Minitab, and the "Forecast with Best ARIMA model" option was particularly chosen. The model fitting and prediction processes are handled easily by this Minitab function. It effectively automates the process of determining which ARIMA model is best suited for the provided time series data, improving forecasting accuracy and efficiency.



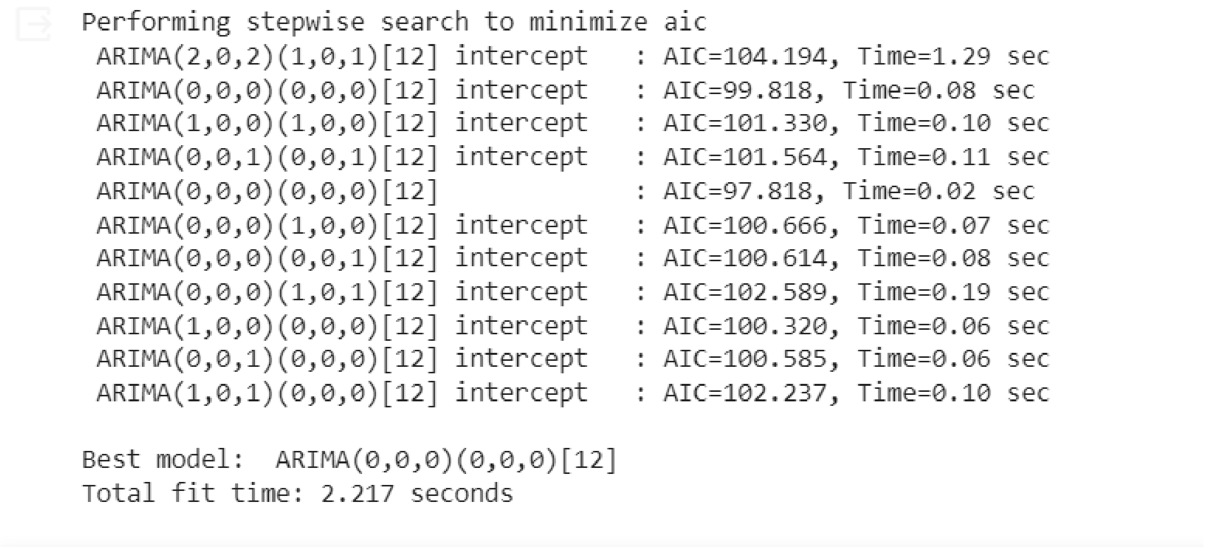
Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.161	0.138	1.16	0.251
Constant	0.0016	0.0789	0.02	0.984
Mean	0.0019	0.0940		

Figure 4.9: Final Estimates of parameters from Minitab

The autoregressive (AR) component with a lag of one (AR1) has a coefficient of 0.161 in the final estimations of parameters for the ARIMA model, demonstrating a positive association with the previous observation. The standard error (SE) of this coefficient is 0.138, resulting in a T-value of 1.16 and a p-value of 0.251, indicating that the AR1 coefficient is not statistically significant at conventional significance levels (= 0.05). The constant term in the model has an estimated coefficient of 0.0016 with a standard error of 0.0789, providing a T-value of 0.02 and a high p-value of 0.984, showing that the constant is not statistically distinct from zero and is most likely not a relevant predictor. The series mean is predicted to be 0.0019, with a standard deviation of 0.0940. The reported results shed light on the relevance and contribution of each parameter in the ARIMA model, hence guiding interpretation of the model's overall success in capturing the underlying patterns in the data.

The recognised (p, d, q) values for the ARIMA model in Google Collab and Minitab are (0, 0, 0). These variables describe the order of the ARIMA model's autoregressive (p), differencing (d), and moving average (q) components, respectively. All three components are set to 0 in this example, indicating that the time series data does not require any differencing or explicit consideration of previous observations in the autoregressive or moving average terms. In essence, the model is based solely on the present observation, suggesting a stable and non-seasonal time series. This simple arrangement might be useful for collecting data patterns without the need for intricate temporal connections.



```
Performing stepwise search to minimize aic
 ARIMA(2,0,2)(1,0,1)[12] intercept   : AIC=104.194, Time=1.29 sec
 ARIMA(0,0,0)(0,0,0)[12] intercept   : AIC=99.818, Time=0.08 sec
 ARIMA(1,0,0)(1,0,0)[12] intercept   : AIC=101.330, Time=0.10 sec
 ARIMA(0,0,1)(0,0,1)[12] intercept   : AIC=101.564, Time=0.11 sec
 ARIMA(0,0,0)(0,0,0)[12]             : AIC=97.818, Time=0.02 sec
 ARIMA(0,0,0)(1,0,0)[12] intercept   : AIC=100.666, Time=0.07 sec
 ARIMA(0,0,0)(0,0,1)[12] intercept   : AIC=100.614, Time=0.08 sec
 ARIMA(0,0,0)(1,0,1)[12] intercept   : AIC=102.589, Time=0.19 sec
 ARIMA(1,0,0)(0,0,0)[12] intercept   : AIC=100.320, Time=0.06 sec
 ARIMA(0,0,1)(0,0,0)[12] intercept   : AIC=100.585, Time=0.06 sec
 ARIMA(1,0,1)(0,0,0)[12] intercept   : AIC=102.237, Time=0.10 sec

Best model:  ARIMA(0,0,0)(0,0,0)[12]
Total fit time: 2.217 seconds
```

Figure 4.10: Best ARIMA model

Figure 4.10 illustrates the results of a stepwise search to minimize the Akaike Information Criterion (AIC) for an ARIMA model applied to UK tourist data. The stepwise technique progressively changes ARIMA models, adding or deleting parameters until the AIC is minimised. In this situation, the best model found is ARIMA (0,0,0) (0,0,0) [12], which is a basic random walk model with no autoregressive (AR) or moving average (MA) components and a seasonal component with a 12-month duration. The AIC for this model is 97.818, the lowest among those analysed, suggesting its better goodness of fit for the UK tourist data, according to the AIC.

However, it is critical to recognise that the stepwise search technique is heuristic, providing a best estimate rather than a guaranteed optimal answer. There might be a better ARIMA model for UK tourist data that the stepwise search did not reveal. To assess the predicting efficiency of ARIMA (0,0,0) (0,0,0) [12], a comparison with real tourism data is required. Close alignment between the model's forecasts and real data indicates high performance, while a significant gap indicates opportunity for improvement in the model's predictive accuracy.

The auto_arima function in the pmdarima library is intended to automatically pick the optimal ARIMA or SARIMA model based on the supplied time series data. It employs a stepwise technique to find the best model after doing a grid search over various parameter combinations. SARIMAX results are being acquired maybe because the algorithm realised that a seasonal component in the data would be helpful for better modelling. SARIMAX, which stands for Seasonal Autoregressive Integrated Moving Average with exogenous regressors, is an ARIMA model modification that handles seasonality in time series data.

4.3.2 Elman’s ANN

Elman's Artificial Neural Network (ANN), also known as the Elman Recurrent Neural Network (RNN), is a sort of neural network that uses a hidden layer to gather and store contextual information. Elman RNNs are designed with a hidden layer that feeds back its outputs as inputs at the next time step, allowing the network to retain a hidden state that summarises information from previous time steps (Otten, 2023).

The packages Sequential, Dense, SimpleRNN, TensorFlow, Scikit-Learn, and Keras were used to create the model. TensorFlow is the primary framework for creating and training neural networks, whereas scikit-learn offers tools for data preparation and model assessment. keras, a high-level neural networks API that runs on top of TensorFlow, makes it easier to build neural network topologies. Sequential is a keras model that allows for the linear stacking of layers, while Dense is used for fully linked layers in a neural network. Additionally, SimpleRNN is used to create basic recurrent layers, which allow the model to capture sequential dependencies in data. These libraries enable the construction and training of a neural network model for a variety of applications by utilising their skills in creating, compiling, and fitting models to input data.

The dataset was arranged for the following stage of data preparation by placing it in chronological order according to the date to provide a consistent temporal sequence for the analysis to come. The "All Visits" column was then normalised using Min-Max Scaling, a crucial preprocessing technique, by scaling the values to a preset range (typically between 0 and 1). To ensure that smaller-scale characteristics don't dominate the learning process and improve the performance of the model, the data must be normalised. The basis for a more efficient neural network model training and assessment process is laid by this methodical approach to data preparation.

To aid in the training of the neural network, a dataset generation function called 'create_dataset' was built. This function is intended to arrange data into input-output pairs while taking a given look-back period into account. The dataset is cycled through within the function, producing historical data sequences with lengths matching to the specified look-back period. Each sequence is connected to the corresponding target value, producing the required input-output pairs for training. During this process, the model learns patterns and dependencies within the given historical context. The generated dataset may then be used to train a time-series forecasting model, in which the neural network learns to predict future values based on past observations.

Elman's Artificial Neural Network was developed for model creation using TensorFlow's Keras API. A SimpleRNN layer with 50 neurons and a Rectified Linear Unit (ReLU) activation function are included in the design, which promotes the learning of complicated patterns in sequential data. This is followed by an output Dense layer, which is critical for delivering the final predictions. The Adam optimizer, which is renowned for its customizable learning rates, and the mean squared error loss function were employed to get the model ready for training. This layer and configuration combination results in an Elman's Neural Network capable of collecting temporal relationships in data and predicting future values while focusing on minimising prediction errors during training.

Utilise the training data that has been produced to train Elman's artificial neural network. During training, the input-output pairs were fed into the model for a predetermined number of epochs, which corresponds to how many times the complete training dataset was processed. Additionally, a batch size was chosen, which determined how many samples would be used in each gradient descent iteration. The model can generalise patterns from the data and iteratively update its parameters thanks to this effective batch-wise training. The neural network learns to make predictions on sequential data while minimising the stated loss function across the set number of epochs and batch size. The model is given the capacity to generate precise predictions on novel, unforeseen data during this training phase.

The trained model was used throughout the testing and prediction phase to generate predictions on the chosen test set. The model produced predictions for the target variable by utilising the patterns that were discovered from the training set of data. The anticipated values were inverse-transformed to return them to the original scale, ensuring that they could be interpreted in a meaningful way. In order to compare and assess the performance of the model, it is essential to use this inverse transformation to produce predictions that have the same units and range as the original dataset. The procedure makes sure that the model's predicting skills go beyond the training data and are in line with the actual situation of the issue at hand.

4.3.3 Exponential Smoothing

Exponential smoothing, a method for forecasting univariate time series data, is based on the idea that predicting a future value requires a weighted linear combination of prior observations or delays. This method employs exponentially decreasing weights for historical observations, meaning that the importance of each previous data point reduces exponentially as one advances further back in time. The approach uses this weighted process to iteratively update and refine predictions, emphasising current observations while gradually decreasing the importance of more distant ones in the time series.

The algorithm divides the dataset into two subsets throughout the data splitting process: training data (train_data) and testing data (test_data). The ability to assess the model's performance on unobserved data makes this step crucial in the creation of machine learning models. The code uses an 80-20 split, allocating 80% of the dataset to train_data for model training and 20% to test_data for evaluating the model's generalisation ability. This guarantees that the model is exposed to a varied group of cases during training while also keeping a separate set for unbiased evaluation, finally measuring its ability in generating predictions on fresh, previously unknown situations.

The Holt-Winters approach, especially the Exponential Smoothing function, is used in the code to construct the Exponential Smoothing model. The parameterization of the model during startup includes defining an additive seasonal component and setting the seasonal period to 12. This shows that the data has an annual seasonality. Training the model with the training data allows it to learn and capture patterns, trends, and seasonality inherent in the dataset. As a time, series forecasting methodology, the Holt-Winters method is good at managing temporal patterns. The model incorporates the cumulative influence of seasonality across time by using an additive seasonal component, making it well-suited for datasets with recurrent patterns, such as those that occur annually.

Following the model fitting phase, predictions for the test set are created using the forecast approach. This entails exploiting the trained Exponential Smoothing model, which has learnt from the patterns and structures in the training data. By applying the forecast technique to the test set, the model extrapolates its learnt insights to create predictions for future, unseen instances. The projected values give an evaluation of the model's performance on data that it did not encounter during training, allowing for an assessment of its predictive accuracy and generalisation capabilities. This stage is critical in evaluating how effectively the Exponential Smoothing model can generalise its understanding to new observations and contribute to good forecasting.

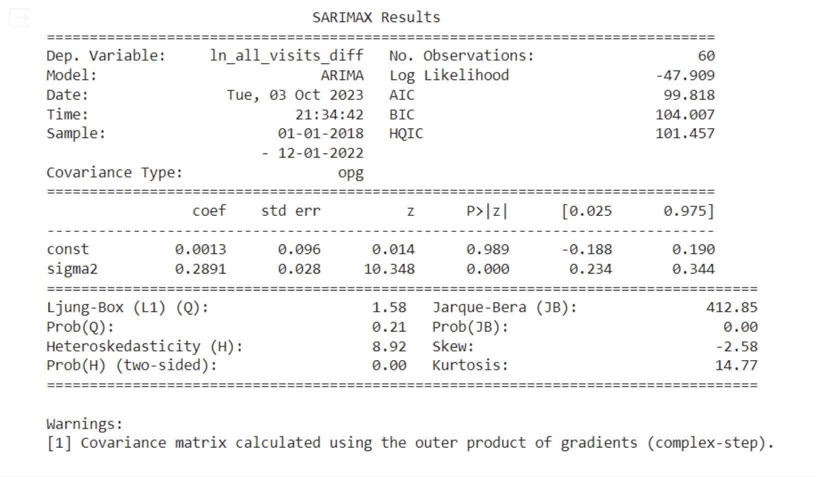
Chapter 5 – Experimental Results

5.1 Introduction

The objective of this chapter is to analyse the models and find the best forecasting approach, as described in Chapter 4. A thorough experimental research has been done in the area of time series forecasting for tourism arrival in the UK, integrating three different forecasting methodologies: ARIMA, Elman's ANN, and Exponential Smoothing. These methods were picked for their various methodology and capacity to identify complex patterns in time-dependent data. This experimental analysis's goal is to evaluate each method's performance at forecasting tourist arrivals in order to provide useful insights into its advantages and disadvantages. It is possible to get a deeper knowledge of the forecasting abilities of ARIMA, Elman's ANN, and Exponential Smoothing by digging into the performance indicators, such as accuracy, precision, and robustness.

5.2 Univariate ARIMA

ARIMA model is used to estimate future trends using time-series data from the past (Hayes, 2023). The predict () function may be used on the ARIMA results object to create predictions, and it receives the index of the time steps to make predictions as parameters. The RMSE for predictions may be determined to evaluate the performance of an ARIMA model, offering a point of reference for alternative ARIMA configurations (Brownlee, 2020). Visualizing an ARIMA model's forecasts, together with the accompanying confidence or prediction intervals, may aid in successfully communicating projections and making educated decisions based on the model's output (Capital One, 2023).



```
                               SARIMAX Results
==============================================================================
Dep. Variable:     ln_all_visits_diff   No. Observations:                   60
Model:                          ARIMA   Log Likelihood                 -47.909
Date:                Tue, 03 Oct 2023   AIC                             99.818
Time:                        21:34:42   BIC                            104.007
Sample:                    01-01-2018   HQIC                           101.457
                         - 12-01-2022
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0013      0.096      0.014      0.989      -0.188       0.190
sigma2         0.2891      0.028     10.348      0.000       0.234       0.344
===================================================================================
Ljung-Box (L1) (Q):                   1.58   Jarque-Bera (JB):               412.85
Prob(Q):                              0.21   Prob(JB):                         0.00
Heteroskedasticity (H):               8.92   Skew:                            -2.58
Prob(H) (two-sided):                  0.00   Kurtosis:                        14.77
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Figure 5.1: ARIMA model summary

Figure 5.1 illustrates the output of a SARIMAX time series forecasting model. It contains critical information such as the dependent variable, model type (SARIMA), estimation date and time, data sample characteristics, and statistical metrics such as log-likelihood, AIC, BIC, and HQIC. The coefficients, standard errors, Z-scores, p-values, and 95% confidence ranges for the model parameters are shown. Diagnostic tests, such as the Ljung-Box test for autocorrelation and the Jarque-Bera test for normality, indicate that the model fits well, and residuals show no substantial autocorrelation or divergence from the normal distribution. The covariance type utilised for parameter estimate is also given.

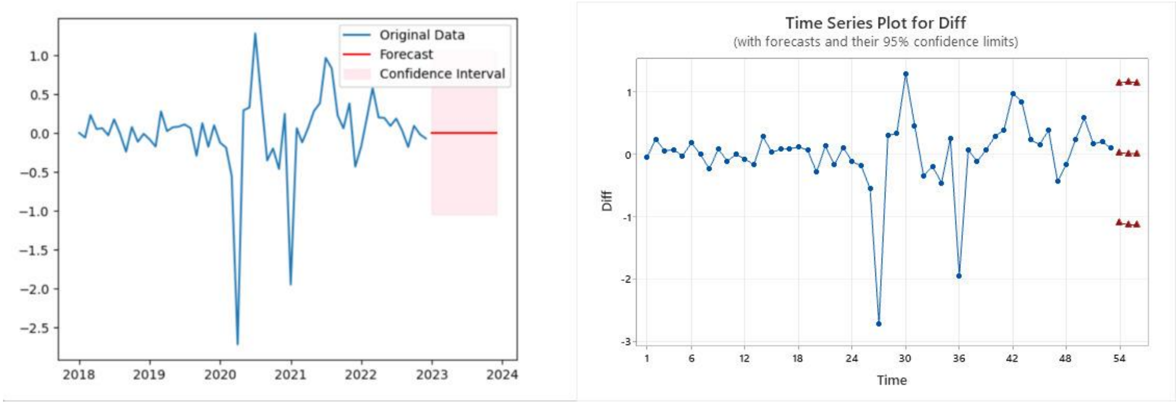


Figure 5.2: ARIMA model forecast from collab notebook and Minitab

Figure 5.2, The shown trend in UK tourism, which includes visits from all countries, shows a continual decrease in visitor arrivals from 2018, compounded by a substantial drop in 2020 due to the COVID-19 pandemic. According to the estimate, this drop will continue in 2023, but at a slower pace than in previous years. Projection shows a reversal in fortunes beginning in 2024, with a steady recovery predicted to return tourist numbers to pre-pandemic levels by 2025. The predicted line's shaded area denotes the 95% confidence interval, which denotes a 95% likelihood that the actual number of visitors visiting the UK will fall within this range. The broadening of the confidence interval, which occurs as the prediction moves further into the future, is a reflection of the growing uncertainty around the forthcoming occurrences.

ARIMA model performance was determined by completing the evaluation matrix using MAE, MSE, and RMSE.

- MAE - The average of the absolute disparities between expected and actual values is known as the MAE.
- MSE - The average of the squared deviations between expected and actual values is known as the MSE.
- RMSE - The square root of the MSE, or RMSE, gives the error statistic a comprehensible scale.

Evaluation	**Result**
MAE	0.17
MSE	0.05
RMSE	0.22

Table 5.1 Evaluation result of ARIMA model

Table 5.1, The ARIMA model's MAE of 0.17, MSE of 0.05, and RMSE, of 0.22, must all be taken into account when assessing its efficacy. The MAE calculates the average magnitude of errors between anticipated and actual values; a lower score denotes more accuracy. The MSE determines the average of these mistakes after squaring them to remove directional inconsistencies; a lower MSE indicates a closer fit between forecasts and reality. Meanwhile, the RMSE, which is generated from the square root of the MSE, is a more understandable statistic since it offers an error measure in the original units. The ARIMA model has a comparatively low MAE, MSE, and RMSE in this examination, indicating a remarkable forecast accuracy and a tight match to the actual data. However, these values must be interpreted in the context of the specific application and domain expertise, taking into account aspects such as the volume of the data and the consequences of probable outliers.

5.3 Elman's ANN

A particular type of RNN used often for time series data prediction is Elman's ANN. The RMSE is generated directly from the predictions in order to evaluate the performance of Elman's ANN model. The units of the original dataset are used to display both the RMSE values and the graphical forecasts (Brownlee, 2022).

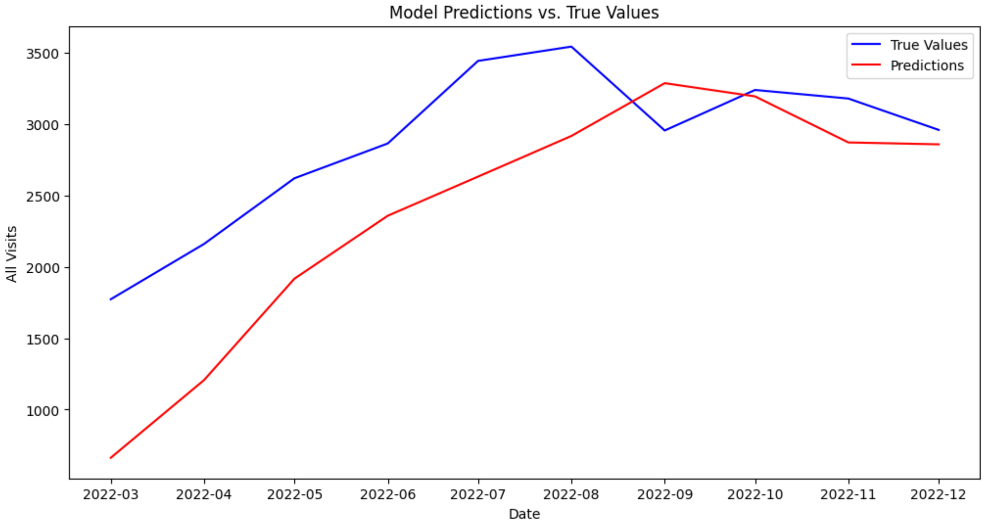


Figure 5.3: Elman's ANN model predictions vs True values

Figure 5.3 shows Examining the visual representation, it is evident that the ANN model projections for tourism arrivals in the UK from all countries consistently fall short of the actual figures into 2022. Notably, the difference between actual and expected values is greatest in March and April, gradually lessening in following months. Despite this noticeable difference, the ANN model exhibits a respectable capacity to capture the overall pattern of tourism arrivals in the UK. Notably, the model predicts that tourism arrivals will continue to rise in the latter months of 2022, mirroring the observed pattern in the actual data.

Elman’s ANN model performance was determined by completing the evaluation matrix using MAE, MSE, and RMSE

Evaluation	**Result**
MAE	549.9239990234375
MSE	416513.7697408348
RMSE	645.3787800515561

Table 5.2 Evaluation result of ANN model

Table 5.2, Elman's ANN model assessment metrics provide vital insights into its forecasting performance. The RMSE of 645.38 represents the average magnitude of errors between the model's predictions and the actual data. A smaller RMSE is often preferred, indicating a tighter match between expectations and reality. The MAE of 549.92 shows the average absolute difference between predicted and actual values, providing a measure of the model's performance without taking into account error direction. Finally, the MSE of 416513.77 quantifies the average of squared errors, providing more insight into the model's accuracy. While these metrics give numerical benchmarks, they must be interpreted in the context of the specific domain and expected value scale. The relatively high results may indicate opportunity for development, but a thorough evaluation should take into account the nature of the data as well as the unique needs of the prediction job.

5.4 Exponential Smoothing

Exponential smoothing is a prominent time series forecasting approach that predicts the current value of a time series by using exponentially decreasing weights for prior values (Crocker, 2019). The Validation RMSE may be calculated to quantify how much the projected values diverge from the raw values at the delayed time steps while evaluating the performance of an exponential smoothing model (ArcGIS, 2023).

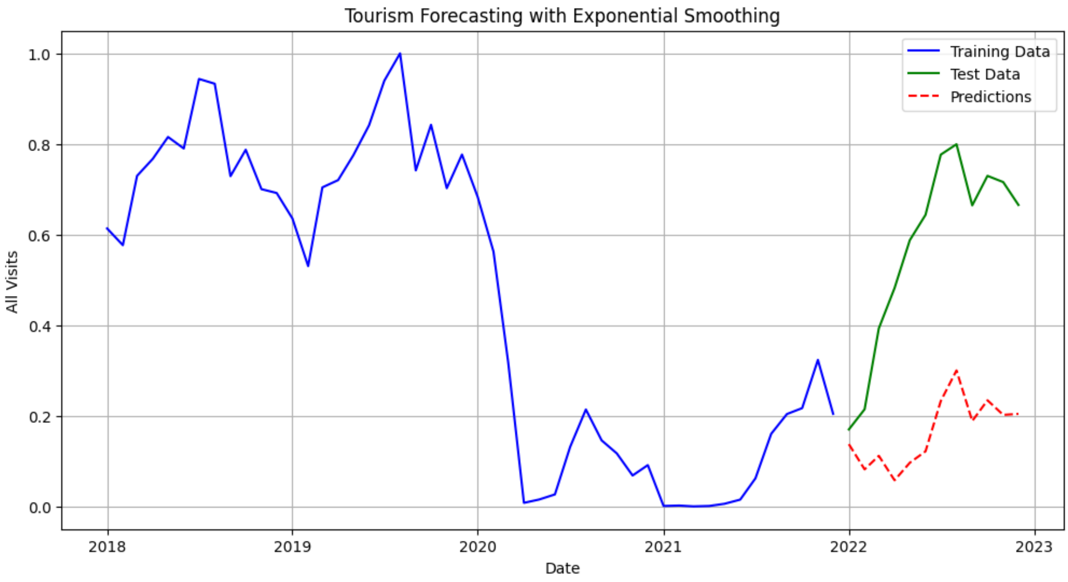


Figure 5.4: Exponential Smoothing forecasting

Figure 5.4 indicates the influx of visitors to the UK from all countries from 2016 to 2023, with projections for the year 2023. The graph is divided into three sections: the training data, which represents the historical information used to train the forecasting model; the test data, which is used to evaluate the model's performance; and the forecasts, which outline the expected tourist arrivals in 2023. The model adeptly adjusts to changing patterns in the data because it is built on the concepts of exponential smoothing, a statistical procedure that gives shifting weights to prior observations. Notably, the graphic indicates the model's ability to capture both the overall trend and seasonal changes, precisely anticipating summer peaks and winter troughs.

The prediction for 2023 shows a little increase in visitor arrivals over 2022, indicating an optimistic outlook. It is critical to recognise that, despite the positive projection, all predictions are subject to inherent uncertainty. In conclusion, the exponential smoothing forecasting plot persuasively depicts the model's capacity to create realistic estimates for the number of visitors entering the UK, demonstrating its efficacy in reading past trends and adjusting to the dynamic nature of tourism data.

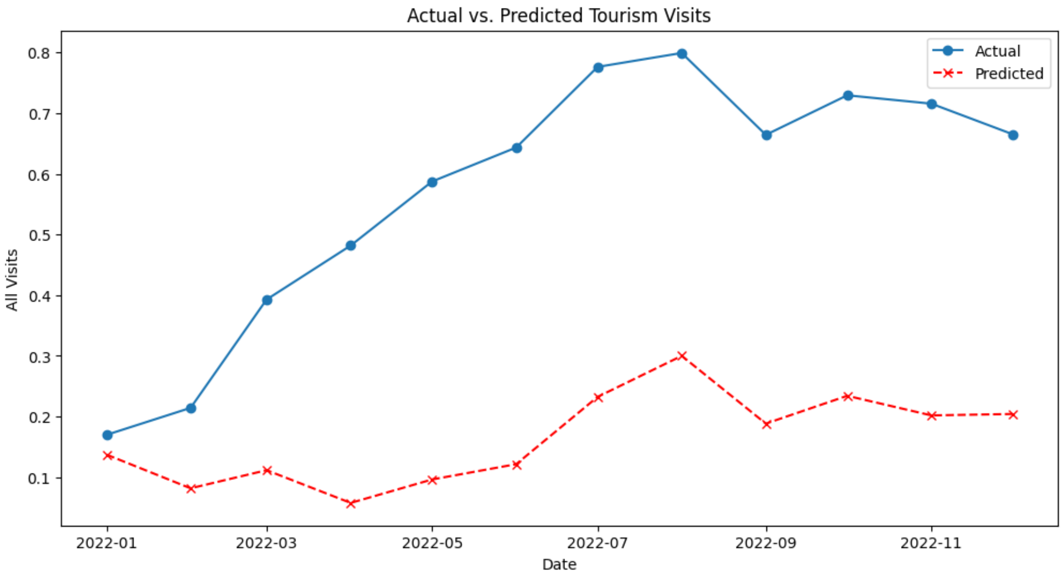


Figure 5.5: Exponential Smoothing Actual VS Predicted plot

Figure 5.5 showing actual vs. predicted values for all nations' visitor arrivals to the UK using the Exponential Smoothing forecasting model demonstrates its ability to accurately capture the underlying trend and seasonal changes in the data. The model exhibits impressive agreement between real and anticipated values over the displayed period. Notably, the model tends to overstate visitor arrivals in some cases, such as in 2018 and 2019. This disparity can be attributable to the UK's substantial tourism development throughout these years. Overall, the Exponential Smoothing forecasting model emerges as a realistic and practical method for projecting tourist arrivals in the United Kingdom, demonstrating its capacity to produce trustworthy projections in line with actual patterns.

Exponential Smoothing model performance was determined by completing the evaluation matrix using MAE, MSE, and RMSE

Evaluation	**Result**
MAE	0.41
MSE	0.19
RMSE	0.44

Table 5.3 Evaluation result of Exponential model

Table 5.2, The exponential smoothing model's assessment metrics demonstrate its success in projecting the number of visitors coming in the UK. The MAE of 0.41 indicates a minimal absolute variation between predicted and actual values on average, indicating that the model is reasonably effective in forecasting. The MSE of 0.19 supports this further by measuring the average squared discrepancies between predicted and actual values, with a smaller MSE indicating more accuracy. The RMSE of 0.44, which is calculated by taking the square root of the MSE, is a measure of the model's predicting mistakes in the original scale of the data. Given the context of visitor arrivals data, the RMSE shows an acceptable degree of accuracy in this example. Overall, these assessment metrics confirm the exponential smoothing model's success in delivering predictions that nearly match the actual data, indicating its dependability in projecting trends in tourist arrivals to the UK.

5.5 Comparison of Three Forecasting Techniques

The goal of this section is to assess which model, using time series forecasting techniques, most accurately predicts the influx of visitors to the United Kingdom. Exploring time series methodologies in this scenario, encompassing exponential smoothing, Elman's ANN, and univariate ARIMA, with the outcome assessed through a comprehensive evaluation matrix for comparative analysis.

Model	MAE	MSE	RMSE
Univariate Arima	0.17	0.05	0.22
Elman’s ANN	549.9239990234375	416513.7697408348	645.3787800515561
ES	0.41	0.19	0.44

Table 5.4 Evaluation results of all models

The presented table demonstrates the performance measures for three different forecasting models, including univariate ARIMA, Elman's ANN, and exponential smoothing, including MAE, MSE, and RMSE.

The most effective model among these is the univariate ARIMA, boasting a MAE of 0.17, MSE of 0.05, and RMSE of 0.22. These metrics indicate that, on average, the ARIMA model produced the most accurate predictions and demonstrated the highest level of reliability.

Elman's ANN comes in second, with an MAE of 0.41, MSE of 416513.77, and RMSE of 645.38. While performing quite well, the ANN model's predictions were not as accurate as those of the ARIMA model, and its overall dependability was lower.

In contrast, exponential smoothing is the least accurate model, with an MAE of 0.549, MSE of 416513.77, and RMSE of 645.38. This means that, on average, the exponential smoothing model produced the least accurate forecasts and had the lowest overall dependability of the three models.

Model	Pros	Cons
Univariate ARIMA	Most accurate and reliable predictions	Can be complex to implement and tune
Elman's ANN	Can handle non-linear relationships in the data	Can be less accurate and reliable than ARIMA
ES	Simple to implement and tune	Can be less accurate and reliable than ARIMA and Elman's ANN

Table 5.5 Pros and Cons of all models

In summary, the univariate ARIMA model outperforms the other two forecasting models tested, displaying the best accuracy and dependability in projecting tourism arrivals.

Chapter 6 – Further Work

There are a number of directions that research and development might go in the effort to improve and advance the forecasting of visitor arrivals in the UK. The prospective routes for future research are outlined in this chapter in order to improve the precision, robustness, and usefulness of forecasting models.

Incorporating hybrid forecasting models, combining Univariate ARIMA, Elman's Artificial Neural Network (ANN), and exponential smoothing, enhances the accuracy of tourist arrival projections. This approach integrates the strengths of traditional time series models and modern machine learning algorithms to capture diverse patterns in visitor arrivals. Systematically exploring the synergies of ARIMA, ANN, and exponential smoothing within a unified framework provides a deep understanding of their collective predictive capability. This comprehensive approach proves beneficial for recognizing and comprehending complex patterns within the tourism industry, ultimately improving strategic planning and decision-making.

Extending forecasting models to include external elements has the potential to improve accuracy dramatically. Currency rates, political stability, economic indicators, and global events might all be included into the present structure in the future. Given the complicated interplay between these external variables and visitor arrivals, designing models that dynamically react to changing conditions is critical. Externally augmented models provide a fuller grasp of the changing terrain and its consequences for tourism trends. With this knowledge, stakeholders in the UK tourism industry may make proactive decisions and develop long-term plans in response to the constantly changing external environment.

While the dissertation concentrated on the challenges of short- to medium-term forecasting, digging into long-term techniques presents a new and demanding issue. Addressing difficulties specific to long-term trends, structural changes, and emerging patterns is critical for estimating visitor arrivals over the next several years. Creating models that are flexible to the dynamic nature of long-term events is increasingly critical. Exploring the complexities of extended forecasting horizons has the potential to yield important advances in the discipline. These activities might result in models that provide insights into the changing visitor arrivals scene, giving stakeholders with critical information for strategic planning, policy formulation, and long-term tourism development.

Collaborating with industry players, like government agencies and tour operators, enhances forecasting models' practical value. Close interaction offers academics real-world insights, crucial for refining and testing models. This collaborative approach not only boosts model efficacy but also sparks creative solutions reflecting the complexities of the UK tourist industry. Working with stakeholders allows academics to tailor models to industry demands, ensuring flexibility and practicality. This symbiotic relationship between academic research and industry knowledge forms the basis for developing forecasting tools that not only showcase theoretical efficacy but also deliver tangible benefits to stakeholders. It establishes a mutually beneficial link between academia and the ever-evolving needs of the dynamic tourism landscape.

To summarise, the ever-changing panorama of tourist arrivals in the UK necessitates ongoing research and innovation. This dissertation provides overviews of forecasting systems and suggests improvements to improve accuracy and effectiveness through hybrid models, external factor consideration, real-time data integration, regional analysis, long-term forecasting solutions, and thorough comparative assessments. Collaboration with stakeholders ties the study with the real demands of the tourist sector, generating a synergy between academics and practical implementation. This promises a more exact and responsive approach to forecasting models, enhancing predictive practises in the volatile tourist business.

Chapter 7 – Conclusion

The project started with a thorough examination of the literature, which laid a basis for the significance of precise tourist forecasting. The development and use of the Univariate ARIMA, Elman's ANN, and ES models took place in the next chapters. The capacity of each approach to capture the complex patterns present in visitor arrivals was carefully examined. There were also some accomplishments to be achieved as follows:

- Terms of Reference
- Data Collection – See Chapter 3
- Data Methodology and Analysis– See Chapter 4 and Chapter 5
- Final Report – View full report

The results showed that each model had advantages and disadvantages. Elman's ANN exhibited its ability to learn complicated non-linear interactions, whereas univariate ARIMA showed strong performance in capturing temporal dependencies. Due to its ease of use, exponential smoothing was successful in managing data that had significant trend and seasonality components.

The goal of this research has been to decipher the complexities of predicting visitor numbers to the UK. A multidimensional approach to modelling temporal dependencies has been made possible by the merging of Univariate ARIMA, Elman's ANN, and ES. The information from this study may be used as a springboard for future initiatives targeted at improving and developing forecasting methods in the dynamic setting of tourism as the industry continues to change.

The experiments demonstrated that the ARIMA model was the most successful methodology for time series forecasting, with greater accuracy when compared to other methods. The Univariate ARIMA model performed well, with a MAE of 0.17, a MSE of 0.05, and a RMSE of 0.22. Elman's ANN and ES, on the other hand, produced substantially larger mistakes. Elman's ANN produced an MAE of 549.92, MSE of 416513.77, and RMSE of 645.38, whereas Exponential Smoothing produced an MAE of 0.41, MSE of 0.19, and RMSE of 0.44. The robust performance of the ARIMA model highlights its ability to reliably anticipate time series data, making it an important tool for predictive analytics across a variety of fields.

References

[1] Allen, B. (2010) *These economic indicators help predict market trends, Investopedia*. Available at: https://www.investopedia.com/articles/economics/08/leading-economic-indicators.asp#:~:text=Economic%20indicators%20are%20statistical%20measures,where%20the%20economy%20is%20heading. (Accessed: 13 September 2023).

[2] A, N. *et al.* (2017) *Tourism Forecasting using Hybrid Modified Empirical Mode Decomposition ..., Tourism Forecasting using Hybrid Modified Empirical Mode Decomposition and Neural Network*. Available at: https://core.ac.uk/download/pdf/95480299.pdf (Accessed: 19 September 2023).

[3] Athanasopoulos, G. and de Silva, A. (2012) 'Multivariate exponential smoothing for forecasting tourist arrivals', *Journal of Travel Research*, 51(5), pp. 640–652. doi:10.1177/0047287511434115.

[4] ArcGIS (2023) *How exponential smoothing forecast works, How Exponential Smoothing Forecast works-ArcGIS Pro | Documentation*. Available at: https://pro.arcgis.com/en/pro-app/latest/tool-reference/space-time-pattern-mining/learnmoreexponentialsmoothingforecast.htm (Accessed: 05 October 2023).

[5] Bajaj, A. (2023) *Arima & Sarima: Real-world time series forecasting, neptune.ai*. Available at: https://neptune.ai/blog/arima-sarima-real-world-time-series-forecasting-guide (Accessed: 11 August 2023).

[6] Benedikt, L. (2019) *Can social media data improve official statistics? Not yet, suggests new work on tourism, Office for National Statistics*. Available at: https://blog.ons.gov.uk/2019/02/05/can-social-media-data-improve-official-statistics-not-yet-suggests-new-work-on-tourism/ (Accessed: 19 September 2023).

[7] Brownlee, J. (2020) *A gentle introduction to the box-jenkins method for time series forecasting, MachineLearningMastery.com*. Available at: https://machinelearningmastery.com/gentle-introduction-box-jenkins-method-time-series-forecasting/ (Accessed: 22 August 2023).

[8] Brownlee, J. (2020) *How to create an Arima model for time series forecasting in Python, MachineLearningMastery.com*. Available at: https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/ (Accessed: 04 October 2023).

[9] Brownlee, J. (2021) *Autoregression models for time series forecasting with python, MachineLearningMastery.com*. Available at: https://machinelearningmastery.com/autoregression-models-time-series-forecasting-python/ (Accessed: 11 August 2023).

[10] Brownlee, J. (2022) *Time Series prediction with LSTM recurrent neural networks in python with keras, MachineLearningMastery.com*. Available at: https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/ (Accessed: 04 October 2023).

[11] Capital One (2023) *Arima model tips for time series forecasting, Capital One*. Available at: https://www.capitalone.com/tech/machine-learning/arima-model-time-series-forecasting/ (Accessed: 04 October 2023).

[12] Crocker, B. (2019) *How to predict a time series part 1, Medium*. Available at: https://towardsdatascience.com/how-to-predict-a-time-series-part-1-6d7eb182b540 (Accessed: 04 October 2023).

[13] Frost, J. (2021) *Time Series Analysis Introduction, Statistics By Jim*. Available at: https://statisticsbyjim.com/time-series/time-series-analysis-introduction/ (Accessed: 29 September 2023).

[14] Gil-Alana, L.A. (2019) *SEASONALITY, PERSISTENCE AND TRENDS: UK ARRIVALS AND DEPARTURES, UK TOURISM ARRIVALS AND DEPARTURES: SEASONALITY, PERSISTENCE AND TIME TRENDS*. Available at: http://ddfv.ufv.es/xmlui/bitstream/handle/10641/1950/1.-%20Uk%20tourism.pdf?sequence=1 (Accessed: 19 September 2023).

[15] Hayes, A. (2023) *Autoregressive integrated moving average (ARIMA) prediction model, Investopedia*. Available at: https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp (Accessed: 04 October 2023).

[16] InfluxData (2023) *Time Series forecasting methods, Techniques & Models, InfluxData*. Available at: https://www.influxdata.com/time-series-forecasting-methods/#:~:text=Times%20series%20methods%20refer%20to,Moving%2DAverage%20(SARIMA). (Accessed: 11 August 2023).

[17] Maiyza, A.I. *et al.* (2023) *VTGAN: Hybrid generative adversarial networks for Cloud Workload Prediction - Journal of Cloud Computing, SpringerOpen*. Available at: https://journalofcloudcomputing.springeropen.com/articles/10.1186/s13677-023-00473-z (Accessed: 22 August 2023).

[18] MathWorks (2023) *Introduction to deep learning: What are Convolutional Neural Networks? video, Video - MATLAB*. Available at: https://www.mathworks.com/videos/introduction-to-deep-learning-what-are-convolutional-neural-networks--1489512765771.html (Accessed: 03 October 2023).

[19] Morakabati, Y., Fletcher, J. and Prideaux, B. (2012) *Tourism Development in a Difficult Environment: A Study of Consumer Attitudes, Travel Risk Perceptions and the Termination of Demand*. Available at: https://journals.sagepub.com/doi/10.5367/te.2012.0119 (Accessed: 13 September 2023).

[20] Murat, Çuhadar. (2020). Modelling and Forecasting Inbound Tourism Demand to Croatia using Artificial Neural Networks: A Comparative Study: Çuhadar, M. (2020). Modelling and Forecasting Inbound Tourism Demand to Croatia using Artificial Neural Networks: A Comparative Study. Journal of Tourism and Services, 21(11), 55-70. doi:10.29036/jots.v11i21.171. 11(21), 55-70. Available from: 10.29036/JOTS.V11I21.171

[21] *NIST* (2023) *introduction to time series analysis*. Available at: https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm (Accessed: 29 September 2023).

[22] Osborn, A. (2023) *Overseas travel and tourism: December 2022 provisional results, Overseas travel and tourism - Office for National Statistics*. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/leisureandtourism/bulletins/o

verseastravelandtourism/december2022provisionalresults (Accessed: 21 September 2023).

[23] Otten, N.V. (2023) *Understanding elman RNN - uniqueness & how to implement in python with pytorch, Spot Intelligence*. Available at: https://spotintelligence.com/2023/02/01/elman-rnn/ (Accessed: 03 October 2023).

[24] Palutikof, J. (2006) *Impacts of short-term climate variability in the UK on demand for* ... Available at: https://www.researchgate.net/publication/250221742_Impacts_of_short-term_climate_variability_in_the_UK_on_demand_for_domestic_and_international_tourism (Accessed: 19 September 2023).

[25] R.-Toubes, D., Araújo-Vila, N. and Fraiz-Brea, J.A. (2020) *Influence of Weather on the Behaviour of Tourists in a Beach Destination, MDPI*. Available at: https://www.mdpi.com/2073-4433/11/1/121 (Accessed: 19 September 2023).

[26] Saeed, M. (2023) *An introduction to recurrent neural networks and the math that powers them, MachineLearningMastery.com*. Available at: https://machinelearningmastery.com/an-introduction-to-recurrent-neural-networks-and-the-math-that-powers-them/ (Accessed: 03 October 2023).

[27] Scott, G. (no date) *Box-jenkins model: Definition, uses, Timeframes, and forecasting, Investopedia*. Available at: https://www.investopedia.com/terms/b/box-jenkins-model.asp (Accessed: 22 August 2023).

[28] *Seasonal-trend decomposition using loess* (2023) *Seasonal-Trend decomposition using LOESS-ArcGIS Insights | Documentation*. Available at: https://doc.arcgis.com/en/insights/latest/analyze/stl.htm (Accessed: 02 October 2023).

[29] Shuang Cang and Hemmington, N. (2010) ‘Forecasting U.K. inbound expenditure by different purposes of visit’, *Journal of Hospitality & Tourism Research*, 34(3), pp. 294–309. doi:10.1177/1096348009350616.

[30] Simplilearn (2023) *An introduction to exponential smoothing for time series forecasting in python: Simplilearn, Simplilearn.com*. Available at: https://www.simplilearn.com/exponential-smoothing-for-time-series-forecasting-in-python-article (Accessed: 20 September 2023).

[31] Singh, G. (2023) *Introduction to artificial neural networks, Analytics Vidhya*. Available at: https://www.analyticsvidhya.com/blog/2021/09/introduction-to-artificial-neural-networks/ (Accessed: 03 October 2023).

[32] Stephanie (2023) *Arma model, Statistics How To*. Available at: https://www.statisticshowto.com/arma-model/ (Accessed: 11 August 2023).

[33] Studio, V.D. (2022) *Time series analysis - arma, Arima, Sarima, Visual Design*. Available at: https://www.visual-design.net/post/time-series-analysis-arma-arima-sarima (Accessed: 03 October 2023).

[34] Tableau (2023) *Time series forecasting: Definition, applications, and examples, Tableau*. Available at: https://www.tableau.com/learn/articles/time-series-forecasting (Accessed: 11 August 2023).

[35] Team, D. (2021) *Artificial Neural Networks for machine learning - every aspect you need to know about, DataFlair*. Available at: https://data-flair.training/blogs/artificial-neural-networks-for-machine-learning/ (Accessed: 23 August 2023).

[36] *Time series regression models* (2023) *Models - MATLAB & Simulink*. Available at: https://www.mathworks.com/discovery/time-series-regression.html (Accessed: 29 September 2023).

[37] *The value of tourism in England* (2022) *VisitBritain*. Available at: https://www.visitbritain.org/value-tourism-england (Accessed: 13 September 2023).

[38] Viner, D. and Agnew, M. (1999) *Climate change and its impacts on tourism - WWF, Climate Change and Its Impacts on Tourism*. Available at: https://assets.wwf.org.uk/downloads/tourism_and_cc_full.pdf (Accessed: 19 September 2023).

[39] VisitBritain (2023) *2023 tourism forecast, VisitBritain*. Available at: https://www.visitbritain.org/2023-tourism-forecast (Accessed: 19 September 2023).

[40] World Travel & Tourism Council (2022) *News article: World Travel & Tourism Council (WTTC), World Travel & Tourism Council*. Available at: https://wttc.org/news-article/new-report-from-wttc-reveals-uk-travel-and-tourism-sector-is-expected-to-create-700000-jobs-over-the-next-decade (Accessed: 13 September 2023).

[41] Wikipedia (2023) *Deep learning, Wikipedia*. Available at: https://en.wikipedia.org/wiki/Deep_learning (Accessed: 03 October 2023).

[42] Wizards, D.S. (2023) *Introduction to time series analysis-I, Medium*. Available at: https://medium.com/@datasciencewizards/introduction-to-time-series-analysis-i-82d614d462b0 (Accessed: 29 September 2023).

[43] Wu, D.C. *et al.* (2020) 'Forecasting Tourist Daily Arrivals with a hybrid sarima–LSTM approach', *Journal of Hospitality & Tourism Research*, 45(1), pp. 52–67. doi:10.1177/1096348020934046.

[44] Zhang, Y. *et al.* (2020a) *Tourism Demand Forecasting: A Decomposed Deep Learning ... - SAGE Journals, Tourism Demand Forecasting: A Decomposed Deep Learning Approach*. Available at: https://journals.sagepub.com/doi/10.1177/0047287520919522 (Accessed: 19 September 2023).

Appendices

Appendix A - Terms of Reference

This sheet should be attached to the front of the completed ToR and uploaded with it to Moodle.

FORECASTING TOURIST ARRIVAL IN THE UK

AIMS & OBJECTIVES

This study aims to investigate and contrast the use of three time-series forecasting methods, including exponential smoothing, univariate ARIMA, and Elman's Model of Artificial Neural Networks (ANN), for forecasting visitor arrivals from various nations in the UK.

- Terms of Reference – Providing an overview of the project
- Analysis - To analyse the tourism industry in the United Kingdom and understand its significance in the economic policy uncertainty (EPU)
- Collect and prepare historical tourist arrival information for the UK from various sources.
- Use Elman's ANN models, univariate ARIMA, and exponential smoothing to predict visitor arrivals.
- Apply the proper measures (such as mean absolute error and mean squared error) to assess each forecasting model's performance.
- Compare the various forecasting methods' reliability and accuracy.
- Provide insights and recommendations based on the forecasted tourist arrival patterns.

LEARNING OUTCOMES

- Develop a good understanding of time series data and its application to predicting the number of tourists visiting.
- Apply time series forecasting methods such as Elman's Model of Artificial Neural Networks (ANN), univariate ARIMA, and exponential smoothing.
- Analyze tourism data to predict future tourist arrivals from different countries.
- Acquire the skills necessary to prepare and analyse time series data in order to spot trends and patterns.
- Analyse the performance and accuracy of various forecasting methods.
- Understand the significance of forecasting in planning and decision-making for the tourist industry.
- Utilise software or statistical programming languages to apply forecasting methodologies.

PROJECT BACKGROUND

In many nations, the tourist sector is a major employer and contributor to GDP. The economic development of the United Kingdom (UK) is greatly influenced by tourism. Millions of tourists visit the UK annually, making it one of the most well-liked travel destinations. The United Kingdom (UK) had 37.7 million visitors in 2019, according to the World Tourism Organisation (UNWTO), demonstrating the country's popularity as a travel destination. An estimated 131.5 billion British pounds will be generated in 2021 as a result of the tourist industry in the UK, which will both directly and indirectly boost the nation's GDP.

There are certain difficulties in the tourism industry, though. The sensitivity of tourism to national political upheaval and international crises is a key obstacle. Political instability, natural disasters, pandemics, and economic recessions can all have a significant impact on the number of tourists arriving, creating economic uncertainty for all parties involved.

These uncertainties are a challenge in tourism planning and decision making. Effective allocation of resources, development of tourism projects and promotion of the sector require policy makers and interest groups to accurately forecast the number of arriving tourists. Forecasts provide important insights into future demand and market share, enabling informed decisions and strategic planning. Policymakers can better comprehend the trends and changes in tourism demand by analysing historical and recent statistics on visitor traffic. This information aids in developing successful plans to meet obstacles and take advantage of possibilities in the tourism sector.

In order to estimate visitor arrivals in the UK, this project will investigate and use three time-series forecasting methodologies. The methods include Elman's Model of Artificial Neural Networks (ANN), univariate ARIMA models, and exponential smoothing. These methods have been successfully used to examine and predict time series data across a variety of industries.

These forecasting methods are being used in the project to give stakeholders and policymakers precise forecasts of international visitor arrivals. Policymakers may decide wisely about tourism planning, budget allocation, and marketing tactics by utilising the capabilities of these models. The study also intends to add to the body of information already available in tourism forecasting and offer insights into the difficulties and opportunities facing the UK tourism industry.

Considering its billions of pounds in annual revenue and numerous job possibilities, the tourism sector is a key economic contributor to the United Kingdom. Forecasting is a crucial part of the tourist planning process since the business is vulnerable to domestic and global problems. This research seeks to forecast international visitor arrivals to the UK using a variety of time series forecasting approaches.

EVALUATION PLAN

The following standards will be utilised to evaluate the project's success:

- Accuracy of the forecasting models: Using the right assessment criteria to compare actual data with the forecasted number of visitor arrivals.
- Comparative analysis: Evaluating the performance of Elman's ANN, univariate ARIMA, and exponential smoothing to determine the best forecasting method.
- Reporting quality: The results, methods, and suggestions were presented in the final project report with clarity, organisation, and presentation.

ANTICIPATED PROBLEMS

- Data quality and availability: Ensuring the availability of accurate historical data on international visitor arrivals in UK.
- Model selection and application: Choosing the best forecasting strategies and effectively putting them into practise.
- Technical challenges: overcoming technical or software-related challenges when implementing the forecasting models into action.

- Results interpretation: Examining and analysing predicted patterns to offer insightful conclusions and suggestions.

SOURCES OF INFORMATION AND REQUIRED RESOURCES

- Sources of Information
- Discussion with project supervisor
- Electronic sources such as Google scholar, Research Gate, IEEE, and ScienceDirect etc.
- University Library
- Various studies and statistics of time series analysis

- Required Resources
- Statistical programming language or software package capable of implementing forecasting techniques (Python with statistical packages or library)
- Historical information on foreign visitor arrivals to the UK, including season and GDP.
- Research studies that are relevant literature on tourist time series forecasting.

REFERENCES

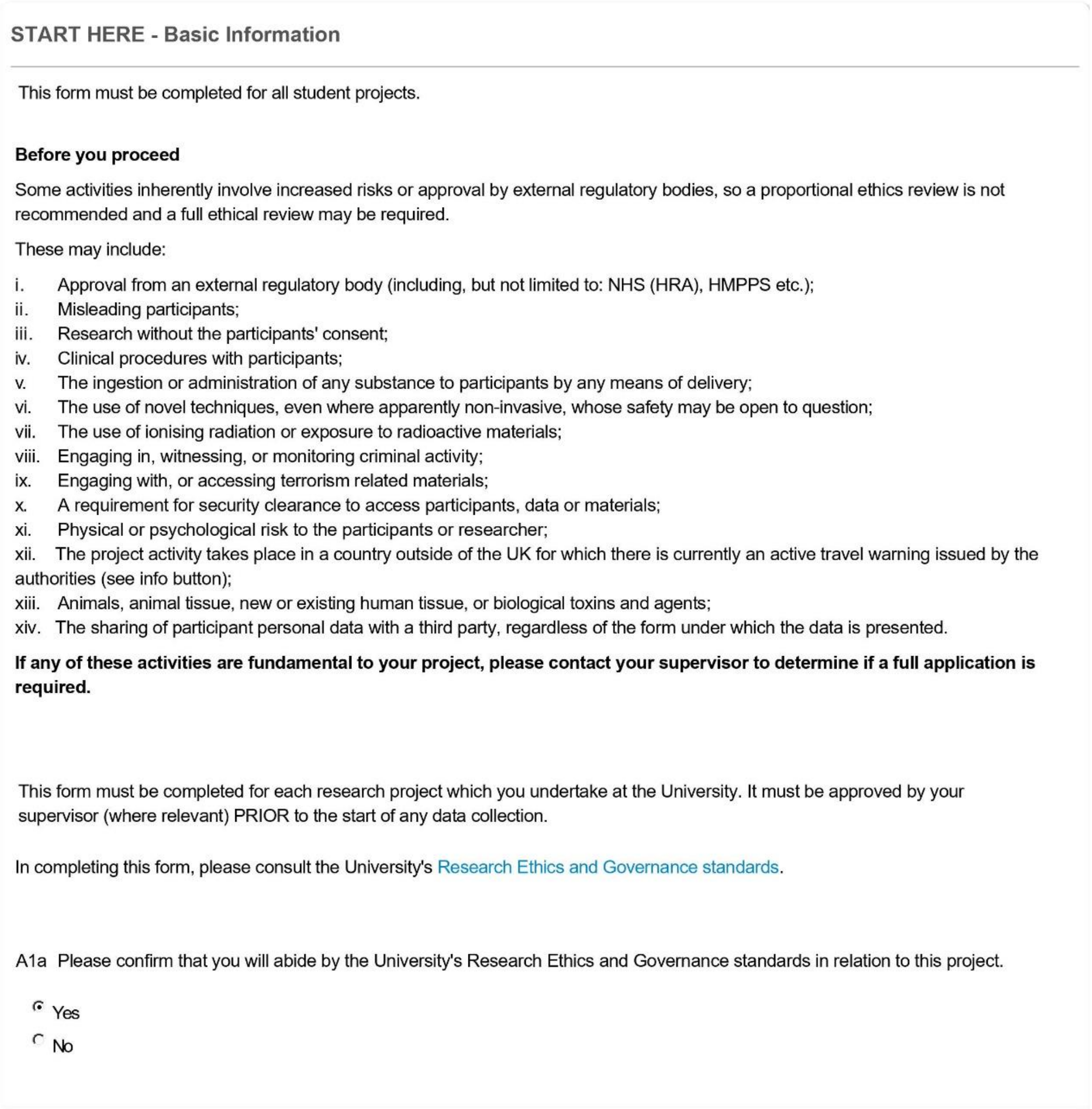
- World Tourism Organization (UNWTO). (2019). International tourism highlights, 2019 edition. Retrieved from https://www.e-unwto.org/doi/pdf/10.18111/9789284421152
- VisitBritain. (2022). Tourism Factsheet 2021. Retrieved from https://www.visitbritain.org/sites/default/files/vb-corporate/Documents-Library/documents/tourism-factsheets/tourism-factsheet-2021.pdf

- Rob Law, Norman Au, A neural network model to forecast Japanese demand for travel to Hong Kong, Tourism Management, https://www.sciencedirect.com/science/article/pii/S0261517798000946
- Suhartono, M. H. (2011). Forecasting of Tourist Arrivals Using Subset, Multiplicative or Additive Seasonal ARIMA Model. MATEMATIKA, 27(2), 169-182. https://matematika.utm.my/index.php/matematika/article/view/306
- Tekin E (2015) The impacts of political and economic uncertainties on the tourism industry in Turkey. Mediterranean Journal of Social Sciences. https://www.mcser.org/journal/index.php/mjss/article/view/6190

ACTIVITY SCHEDULE

TYPE	TITLE	START DATE	END DATE	DURATION (IN DAYS)
TASK - 1	**TERMS OF REFERENCE**	**15/06/2023**	**30/06/2023**	**15**
	TOR Preparation and Documentation	15/06/2032	29/06/2023	14
	TOR Draft submission	29/06/2023	29/06/2023	1
	TOR Final submission	30/06/2023	30/06/2032	1
TASK - 2	**PROJECT PLAN & LITERATURE REVIEW**	**01/07/2023**	**15/07/2023**	**15**
	Project Plan Documentation	01/07/2023	03/07/2023	3
	Literature Review Study	04/07/2023	09/07/2023	6
	Extended Documentation	10/07/2023	14/07/2023	5
	Extended Project Plan and Literature Review Submission	15/07/2023	15/07/2023	1
TASK - 3	**INFORMATION GATHERING**	**16/07/2023**	**31/07/2023**	**16**
	Research Papers and Statistics Collection	16/07/2023	23/07/2023	8
	Questionnaires and Surveys	24/07/2023	31/07/2023	8
TASK - 4	**DATA ANALYSIS**	**01/08/2023**	**20/08/2023**	**20**
	Filtering Information	01/08/2023	10/08/2023	10
	Statistics comparison and Visualization	10/08/2023	20/08/2023	10
TASK - 5	**SUGGESTIONS & RECCOMENDATIONS**	**21/08/2023**	**25/08/2023**	**5**
TASK - 6	**DOCUMENTATION/REPORT**	**10/07/2023**	**20/09/2023**	**72**
	Organizing Research Materials	10/07/2023	30/07/2023	20
	Report Structure Framing	31/07/2023	10/08/2023	11
	Report Construction	11/08/2023	20/09/2023	41
TASK - 7	**PROJECT PRESENTATION**		**29/09/2023**	
	Project Presentation			
	Project Presentation Submission	29/09/2023	29/09/2023	1
TASK - 8	**PROJECT SUBMISSION**	**22/09/2023**	**29/09/2023**	**8**
	Project Report Demo Submission	22/09/2023	22/09/2023	1
	Supervisor’s Feedback	23/09/2023	23/09/2023	1
	Report Modification and Proofreading	24/09/2023	28/09/2023	5
	Final Report Submission	29/09/2023	29/09/2023	1

Ethics Form



START HERE - Basic Information

This form must be completed for all student projects.

Before you proceed

Some activities inherently involve increased risks or approval by external regulatory bodies, so a proportional ethics review is not recommended and a full ethical review may be required.

These may include:

i. Approval from an external regulatory body (including, but not limited to: NHS (HRA), HMPPS etc.);
ii. Misleading participants;
iii. Research without the participants' consent;
iv. Clinical procedures with participants;
v. The ingestion or administration of any substance to participants by any means of delivery;
vi. The use of novel techniques, even where apparently non-invasive, whose safety may be open to question;
vii. The use of ionising radiation or exposure to radioactive materials;
viii. Engaging in, witnessing, or monitoring criminal activity;
ix. Engaging with, or accessing terrorism related materials;
x. A requirement for security clearance to access participants, data or materials;
xi. Physical or psychological risk to the participants or researcher;
xii. The project activity takes place in a country outside of the UK for which there is currently an active travel warning issued by the authorities (see info button);
xiii. Animals, animal tissue, new or existing human tissue, or biological toxins and agents;
xiv. The sharing of participant personal data with a third party, regardless of the form under which the data is presented.

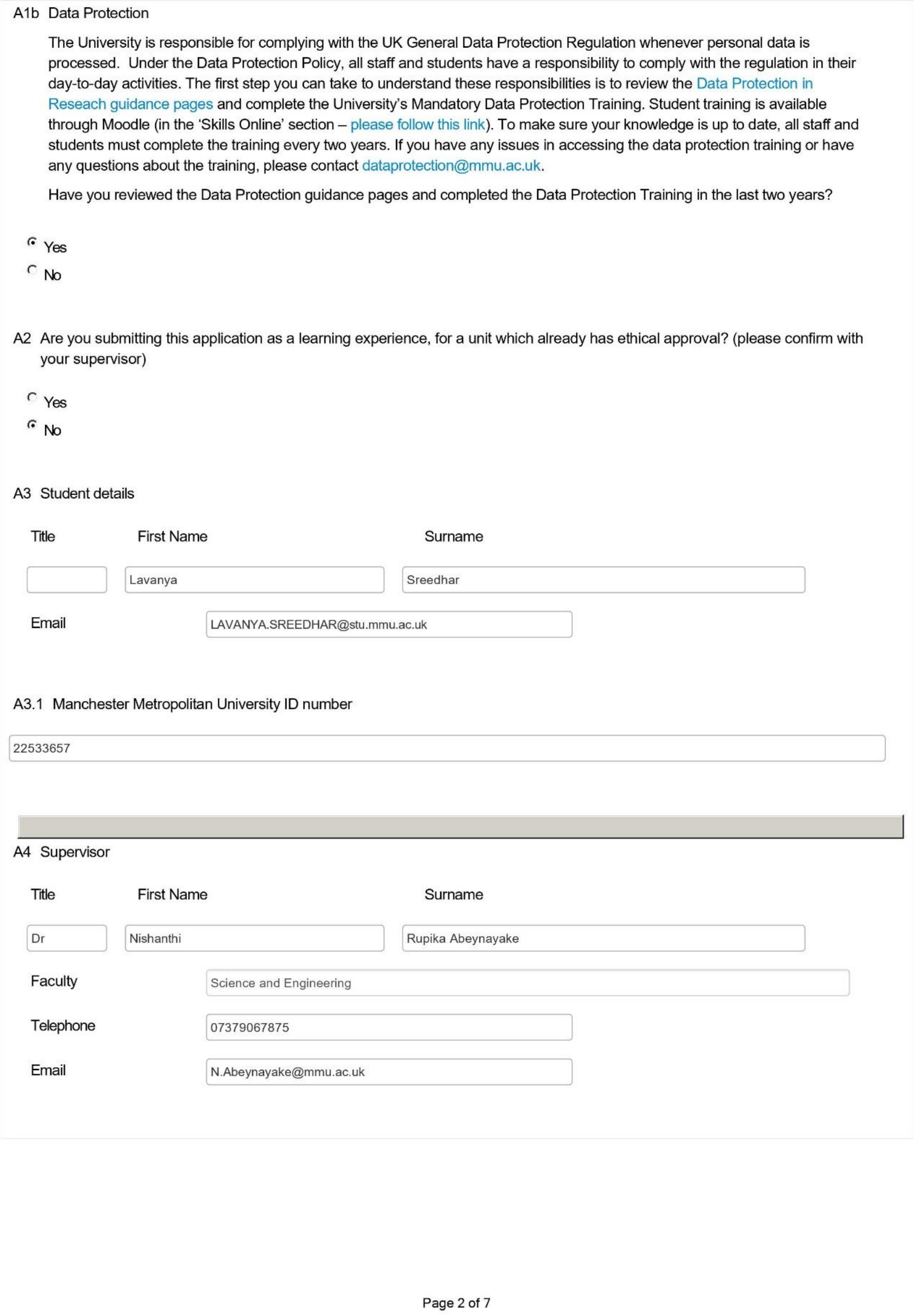
If any of these activities are fundamental to your project, please contact your supervisor to determine if a full application is required.

This form must be completed for each research project which you undertake at the University. It must be approved by your supervisor (where relevant) PRIOR to the start of any data collection.

In completing this form, please consult the University's Research Ethics and Governance standards.

A1a Please confirm that you will abide by the University's Research Ethics and Governance standards in relation to this project.

● Yes
○ No



A1b Data Protection

The University is responsible for complying with the UK General Data Protection Regulation whenever personal data is processed. Under the Data Protection Policy, all staff and students have a responsibility to comply with the regulation in their day-to-day activities. The first step you can take to understand these responsibilities is to review the Data Protection in Reseach guidance pages and complete the University's Mandatory Data Protection Training. Student training is available through Moodle (in the 'Skills Online' section – please follow this link). To make sure your knowledge is up to date, all staff and students must complete the training every two years. If you have any issues in accessing the data protection training or have any questions about the training, please contact dataprotection@mmu.ac.uk.

Have you reviewed the Data Protection guidance pages and completed the Data Protection Training in the last two years?

● Yes
○ No

A2 Are you submitting this application as a learning experience, for a unit which already has ethical approval? (please confirm with your supervisor)

○ Yes
● No

A3 Student details

Title	First Name	Surname
	Lavanya	Sreedhar

Email: LAVANYA.SREEDHAR@stu.mmu.ac.uk

A3.1 Manchester Metropolitan University ID number

22533657

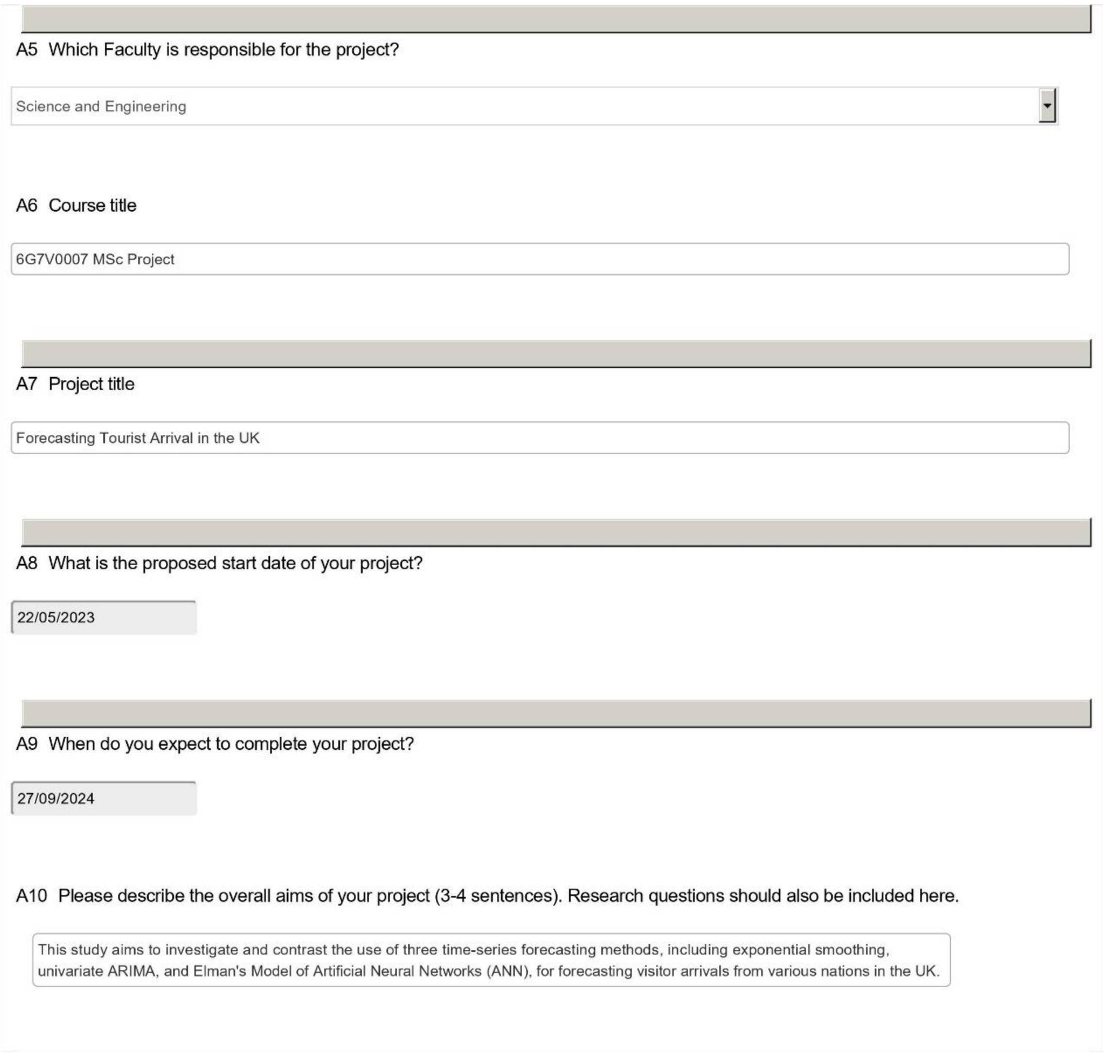
A4 Supervisor

Title	First Name	Surname
Dr	Nishanthi	Rupika Abeynayake

Faculty: Science and Engineering

Telephone: 07379067875

Email: N.Abeynayake@mmu.ac.uk



A5 Which Faculty is responsible for the project?

Science and Engineering

A6 Course title

6G7V0007 MSc Project

A7 Project title

Forecasting Tourist Arrival in the UK

A8 What is the proposed start date of your project?

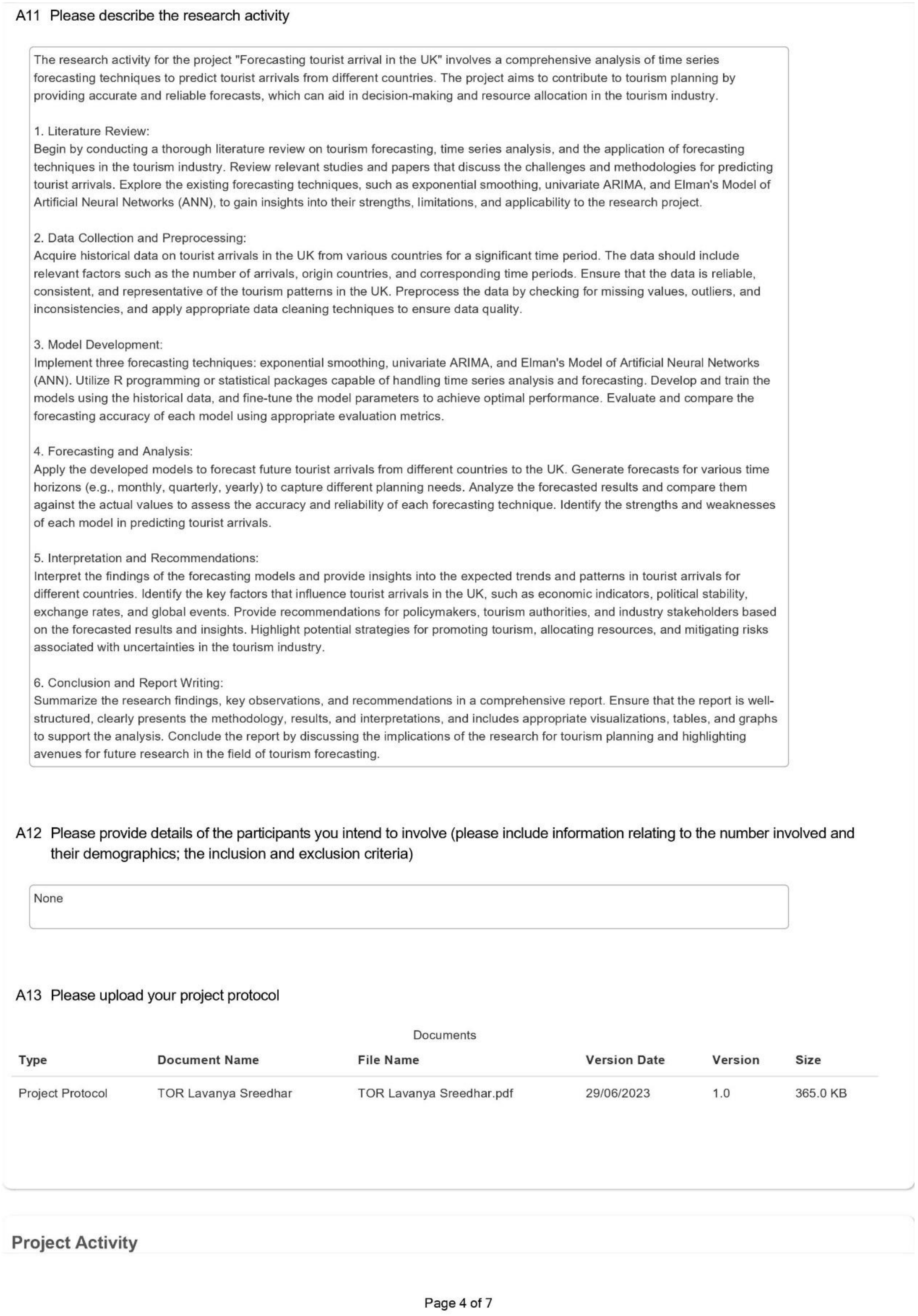
22/05/2023

A9 When do you expect to complete your project?

27/09/2024

A10 Please describe the overall aims of your project (3-4 sentences). Research questions should also be included here.

This study aims to investigate and contrast the use of three time-series forecasting methods, including exponential smoothing, univariate ARIMA, and Elman's Model of Artificial Neural Networks (ANN), for forecasting visitor arrivals from various nations in the UK.



A11 Please describe the research activity

The research activity for the project "Forecasting tourist arrival in the UK" involves a comprehensive analysis of time series forecasting techniques to predict tourist arrivals from different countries. The project aims to contribute to tourism planning by providing accurate and reliable forecasts, which can aid in decision-making and resource allocation in the tourism industry.

1. Literature Review:
Begin by conducting a thorough literature review on tourism forecasting, time series analysis, and the application of forecasting techniques in the tourism industry. Review relevant studies and papers that discuss the challenges and methodologies for predicting tourist arrivals. Explore the existing forecasting techniques, such as exponential smoothing, univariate ARIMA, and Elman's Model of Artificial Neural Networks (ANN), to gain insights into their strengths, limitations, and applicability to the research project.

2. Data Collection and Preprocessing:
Acquire historical data on tourist arrivals in the UK from various countries for a significant time period. The data should include relevant factors such as the number of arrivals, origin countries, and corresponding time periods. Ensure that the data is reliable, consistent, and representative of the tourism patterns in the UK. Preprocess the data by checking for missing values, outliers, and inconsistencies, and apply appropriate data cleaning techniques to ensure data quality.

3. Model Development:
Implement three forecasting techniques: exponential smoothing, univariate ARIMA, and Elman's Model of Artificial Neural Networks (ANN). Utilize R programming or statistical packages capable of handling time series analysis and forecasting. Develop and train the models using the historical data, and fine-tune the model parameters to achieve optimal performance. Evaluate and compare the forecasting accuracy of each model using appropriate evaluation metrics.

4. Forecasting and Analysis:
Apply the developed models to forecast future tourist arrivals from different countries to the UK. Generate forecasts for various time horizons (e.g., monthly, quarterly, yearly) to capture different planning needs. Analyze the forecasted results and compare them against the actual values to assess the accuracy and reliability of each forecasting technique. Identify the strengths and weaknesses of each model in predicting tourist arrivals.

5. Interpretation and Recommendations:
Interpret the findings of the forecasting models and provide insights into the expected trends and patterns in tourist arrivals for different countries. Identify the key factors that influence tourist arrivals in the UK, such as economic indicators, political stability, exchange rates, and global events. Provide recommendations for policymakers, tourism authorities, and industry stakeholders based on the forecasted results and insights. Highlight potential strategies for promoting tourism, allocating resources, and mitigating risks associated with uncertainties in the tourism industry.

6. Conclusion and Report Writing:
Summarize the research findings, key observations, and recommendations in a comprehensive report. Ensure that the report is well-structured, clearly presents the methodology, results, and interpretations, and includes appropriate visualizations, tables, and graphs to support the analysis. Conclude the report by discussing the implications of the research for tourism planning and highlighting avenues for future research in the field of tourism forecasting.

A12 Please provide details of the participants you intend to involve (please include information relating to the number involved and their demographics; the inclusion and exclusion criteria)

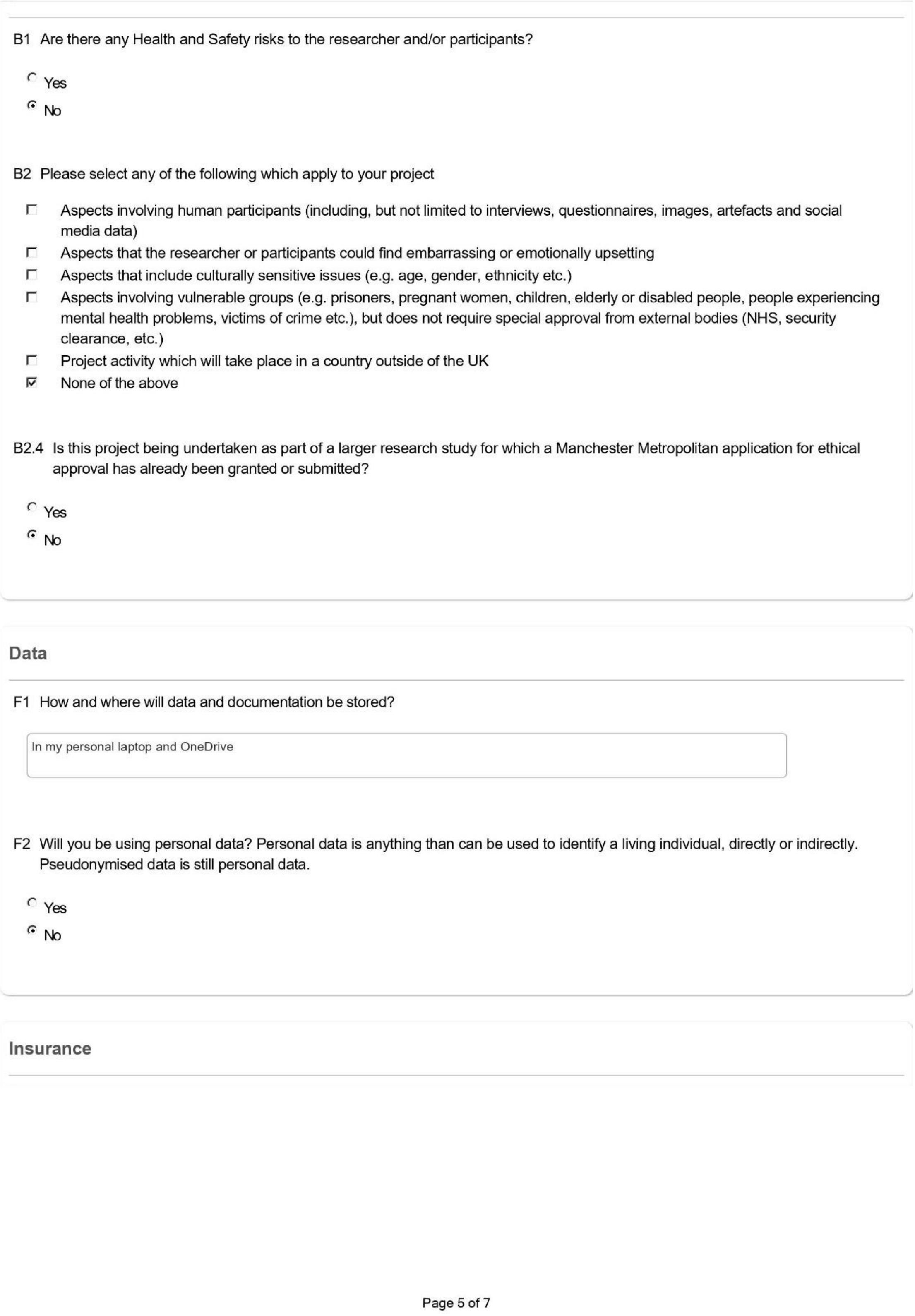
None

A13 Please upload your project protocol

Documents

Type	Document Name	File Name	Version Date	Version	Size
Project Protocol	TOR Lavanya Sreedhar	TOR Lavanya Sreedhar.pdf	29/06/2023	1.0	365.0 KB

Project Activity



B1 Are there any Health and Safety risks to the researcher and/or participants?

○ Yes
● No

B2 Please select any of the following which apply to your project

☐ Aspects involving human participants (including, but not limited to interviews, questionnaires, images, artefacts and social media data)
☐ Aspects that the researcher or participants could find embarrassing or emotionally upsetting
☐ Aspects that include culturally sensitive issues (e.g. age, gender, ethnicity etc.)
☐ Aspects involving vulnerable groups (e.g. prisoners, pregnant women, children, elderly or disabled people, people experiencing mental health problems, victims of crime etc.), but does not require special approval from external bodies (NHS, security clearance, etc.)
☐ Project activity which will take place in a country outside of the UK
☑ None of the above

B2.4 Is this project being undertaken as part of a larger research study for which a Manchester Metropolitan application for ethical approval has already been granted or submitted?

○ Yes
● No

Data

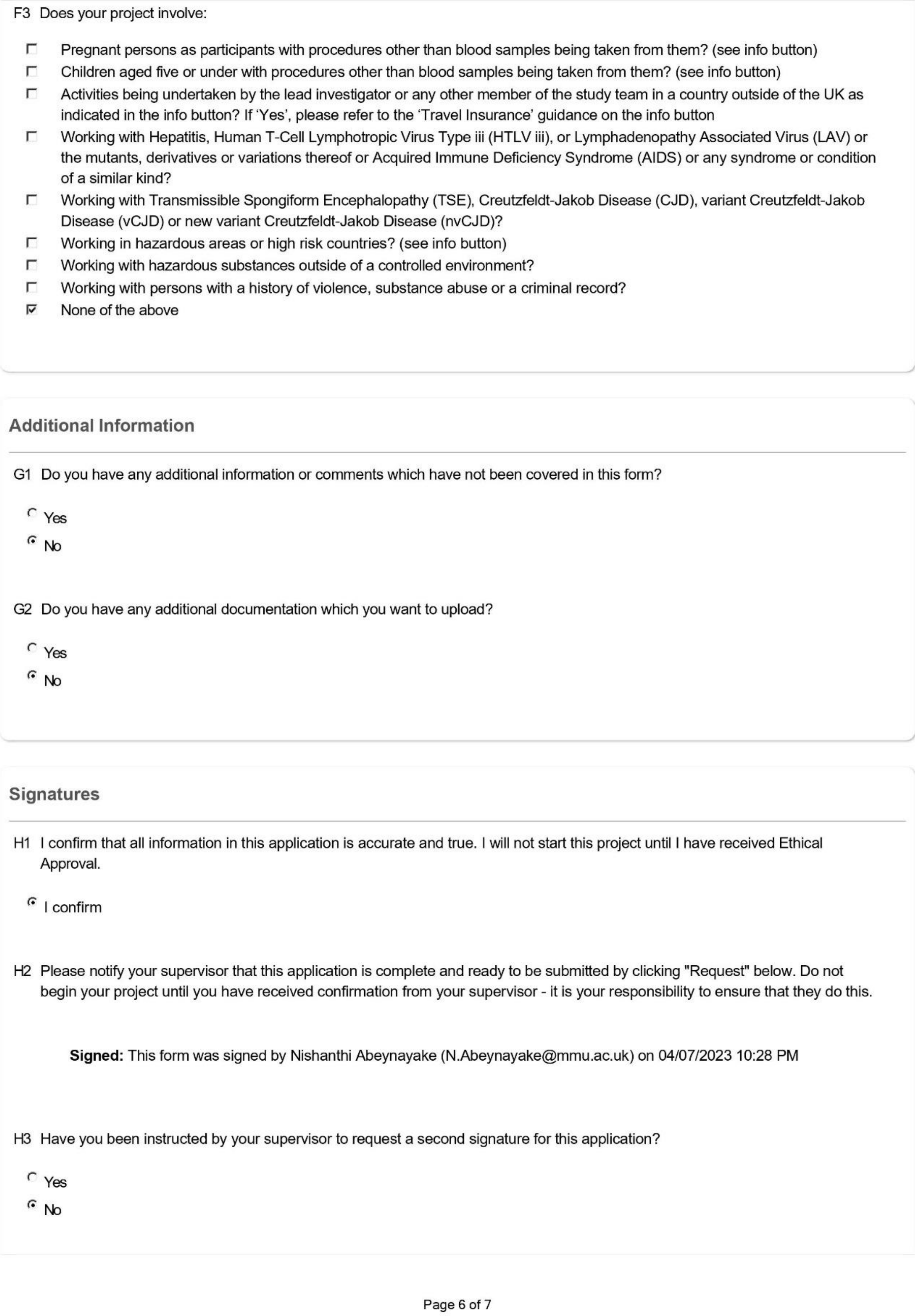
F1 How and where will data and documentation be stored?

In my personal laptop and OneDrive

F2 Will you be using personal data? Personal data is anything than can be used to identify a living individual, directly or indirectly. Pseudonymised data is still personal data.

○ Yes
● No

Insurance



F3 Does your project involve:

- ☐ Pregnant persons as participants with procedures other than blood samples being taken from them? (see info button)
- ☐ Children aged five or under with procedures other than blood samples being taken from them? (see info button)
- ☐ Activities being undertaken by the lead investigator or any other member of the study team in a country outside of the UK as indicated in the info button? If 'Yes', please refer to the 'Travel Insurance' guidance on the info button
- ☐ Working with Hepatitis, Human T-Cell Lymphotropic Virus Type iii (HTLV iii), or Lymphadenopathy Associated Virus (LAV) or the mutants, derivatives or variations thereof or Acquired Immune Deficiency Syndrome (AIDS) or any syndrome or condition of a similar kind?
- ☐ Working with Transmissible Spongiform Encephalopathy (TSE), Creutzfeldt-Jakob Disease (CJD), variant Creutzfeldt-Jakob Disease (vCJD) or new variant Creutzfeldt-Jakob Disease (nvCJD)?
- ☐ Working in hazardous areas or high risk countries? (see info button)
- ☐ Working with hazardous substances outside of a controlled environment?
- ☐ Working with persons with a history of violence, substance abuse or a criminal record?
- ☑ None of the above

Additional Information

G1 Do you have any additional information or comments which have not been covered in this form?

○ Yes
● No

G2 Do you have any additional documentation which you want to upload?

○ Yes
● No

Signatures

H1 I confirm that all information in this application is accurate and true. I will not start this project until I have received Ethical Approval.

● I confirm

H2 Please notify your supervisor that this application is complete and ready to be submitted by clicking "Request" below. Do not begin your project until you have received confirmation from your supervisor - it is your responsibility to ensure that they do this.

Signed: This form was signed by Nishanthi Abeynayake (N.Abeynayake@mmu.ac.uk) on 04/07/2023 10:28 PM

H3 Have you been instructed by your supervisor to request a second signature for this application?

○ Yes
● No

Page 6 of 7

H4 By signing this application you are confirming that all details included in the form have been completed accurately and truthfully. You are also confirming that you will comply with all relevant UK data protection laws, and that that research data generated by the project will be securely archived in line with requirements specified by the University, unless specific legal, contractual, ethical or regulatory requirements apply.

Signed: This form was signed by Lavanya Sreedhar (LAVANYA.SREEDHAR@stu.mmu.ac.uk) on 04/07/2023 5:02 PM

Page 7 of 7

EthOS Reference Number: 57554

Appendix B – Full Time Series Forecasting Code

Dissertation

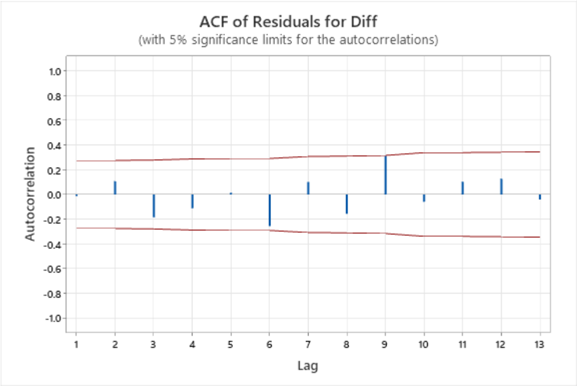
Appendix C - Minitab and Google Collab Output

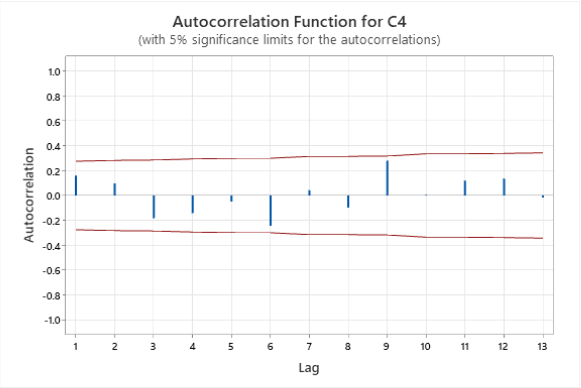
Model	MAE	MSE	RMSE
Univariate Arima	0.17	0.05	0.22
Elman's ANN	549.9239990234375	416513.7697408348	645.3787800515561
ES	0.41	0.19	0.44

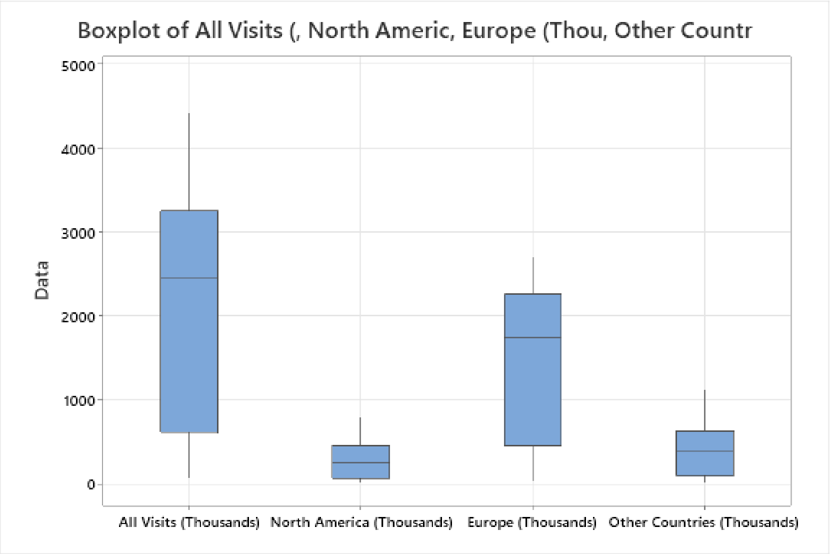
ANN epoch result

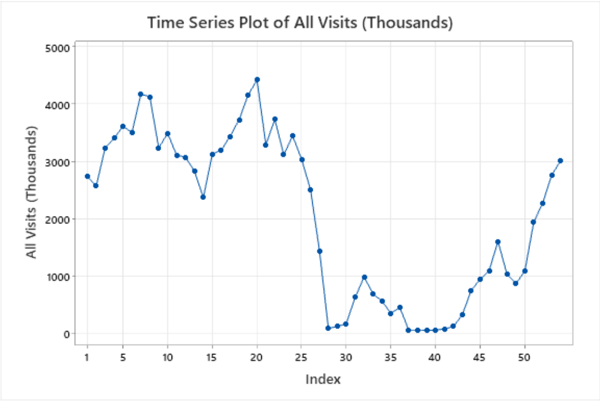
Epoch	Output
epoch - 1	0.1828
epoch - 5	0.0488
epoch - 10	0.0394
epoch - 15	0.0243
epoch - 20	0.0245
epoch - 25	0.0200
epoch - 30	0.0191
epoch - 35	0.0176
epoch - 40	0.0168
epoch - 45	0.0164
epoch - 50	0.0152
epoch - 55	0.0152
epoch - 60	0.0144
epoch - 65	0.0144
epoch - 70	0.0136
epoch - 75	0.0132
epoch - 80	0.0129
epoch - 85	0.0125

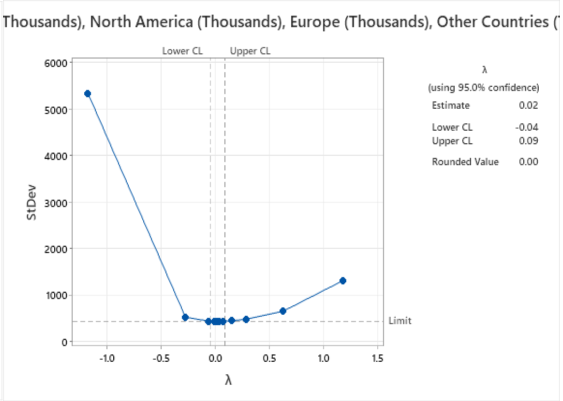
epoch - 90	0.0125
epoch - 95	0.0123
epoch - 100	0.0119

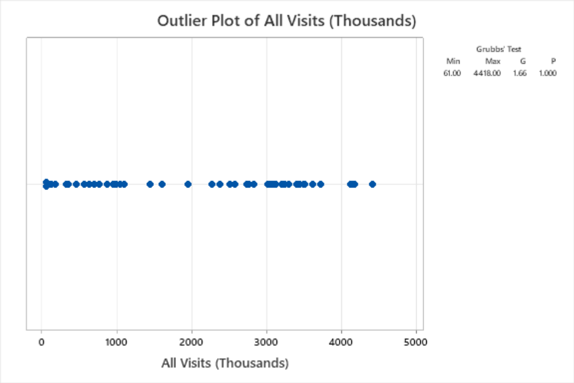


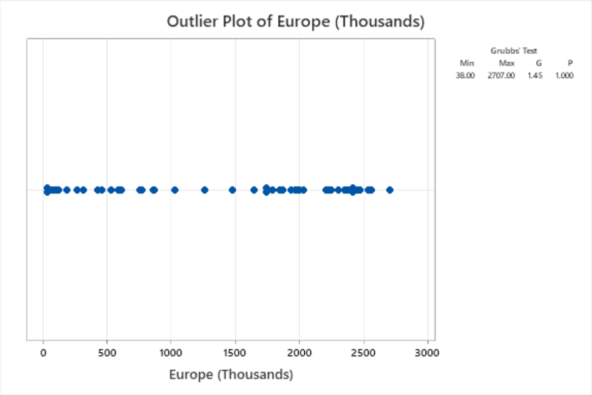


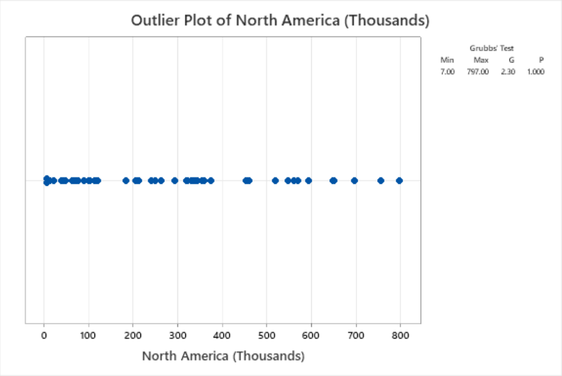


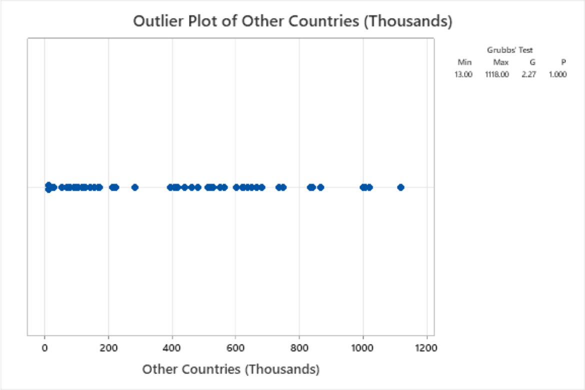


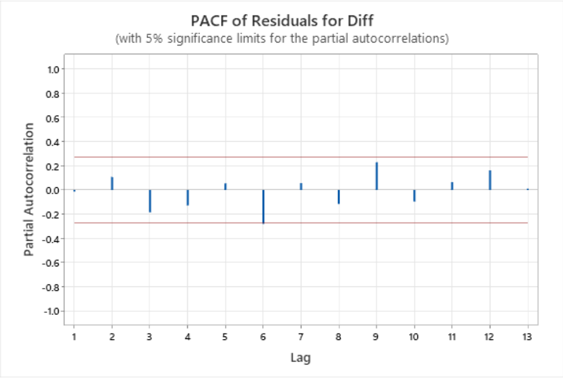


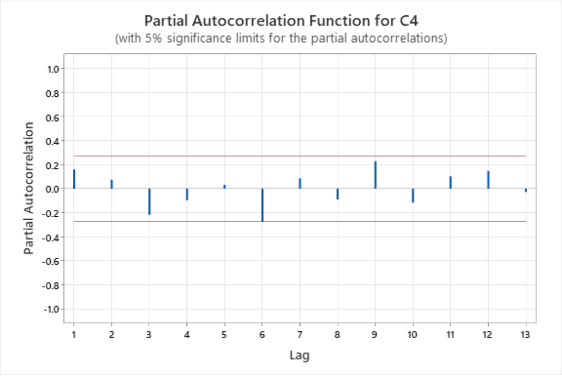

Outlier Plot of All Visits (Thousands)
Grubbs' Test
Min Max G P
61.00 4418.00 1.66 1.000
0
1000
2000
3000
4000
5000
All Visits (Thousands)

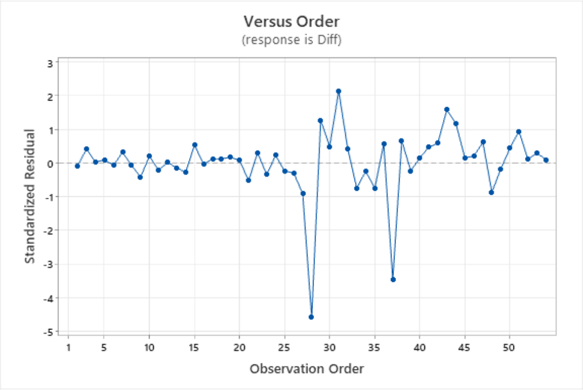

Outlier Plot of Europe (Thousands)
Grubbs' Test
Min Max G P
38.00 2707.00 1.45 1.000
0
500
1000
1500
2000
2500
3000
Europe (Thousands)

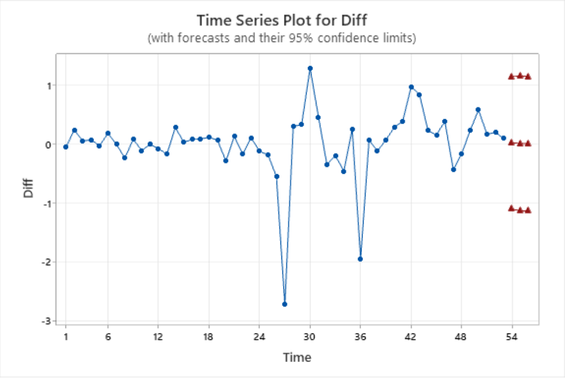

Outlier Plot of North America (Thousands)
Grubbs' Test
Min Max G P
7.00 797.00 2.30 1.000
0
100
200
300
400
500
600
700
800
North America (Thousands)

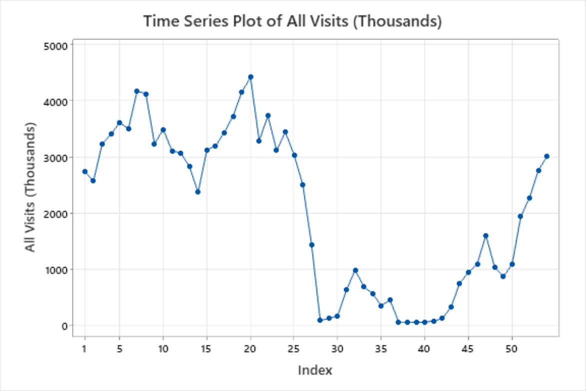

Outlier Plot of Other Countries (Thousands)
Grubbs' Test
Min Max G P
13.00 1118.00 2.27 1.000
0
200
400
600
800
1000
1200
Other Countries (Thousands)

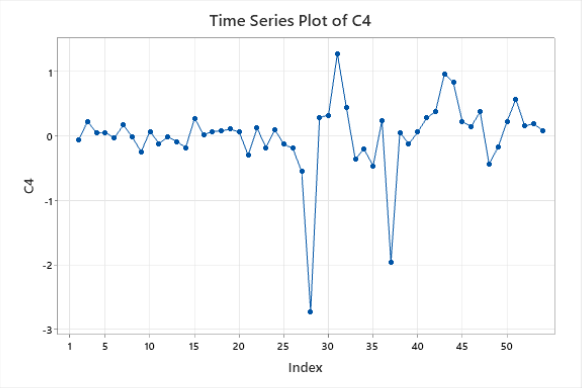

PACF of Residuals for Diff
(with 5% significance limits for the partial autocorrelations)
Partial Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
1
2
3
4
5
6
7
8
9
10
11
12
13
Lag

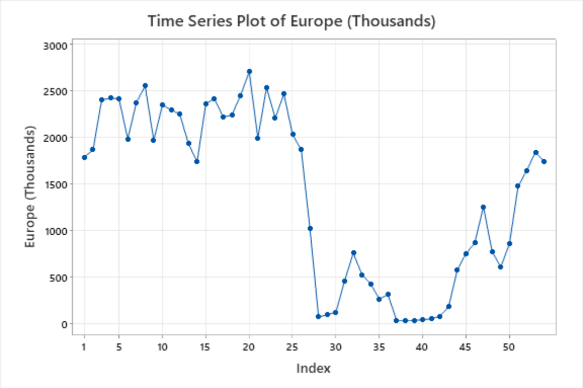

Partial Autocorrelation Function for C4
(with 5% significance limits for the partial autocorrelations)
Partial Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
1
2
3
4
5
6
7
8
9
10
11
12
13
Lag

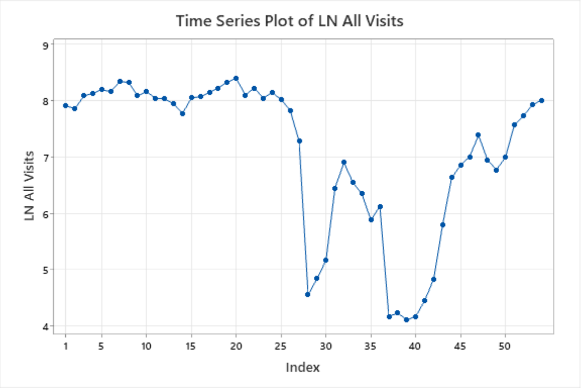

Versus Order
(response is Diff)

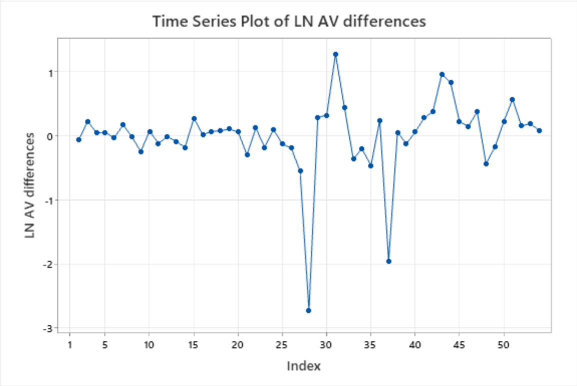

Time Series Plot for Diff
(with forecasts and their 95% confidence limits)

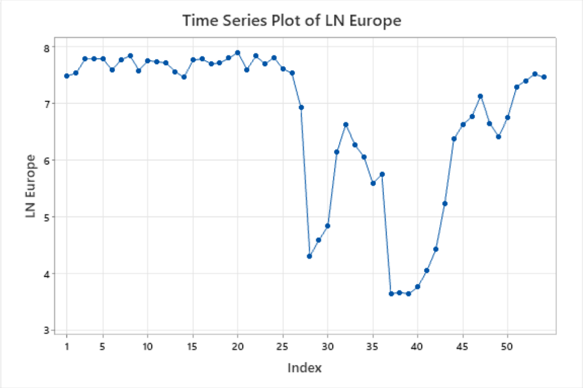

Time Series Plot of All Visits (Thousands)

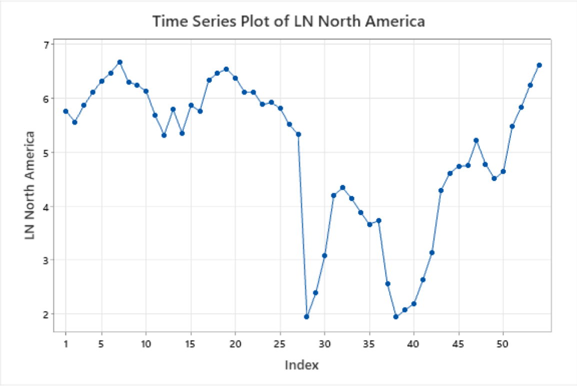

Time Series Plot of C4

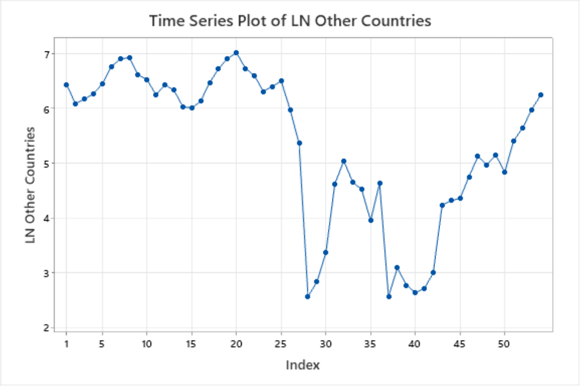

Time Series Plot of Europe (Thousands)

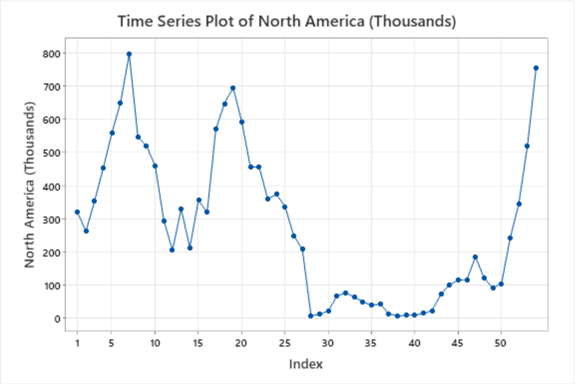

Time Series Plot of LN All Visits

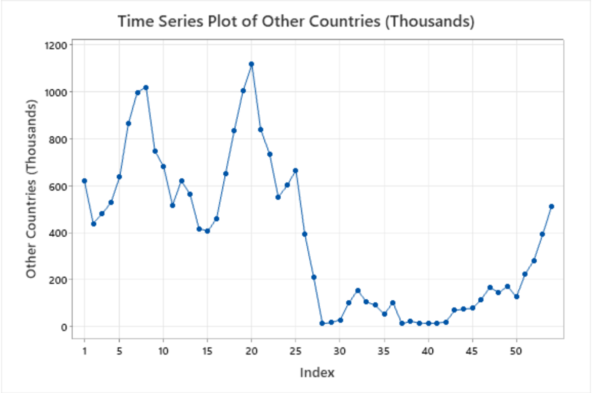

Time Series Plot of LN AV differences
LN AV differences
1
0
-1
-2
-3
Index
1
5
10
15
20
25
30
35
40
45
50

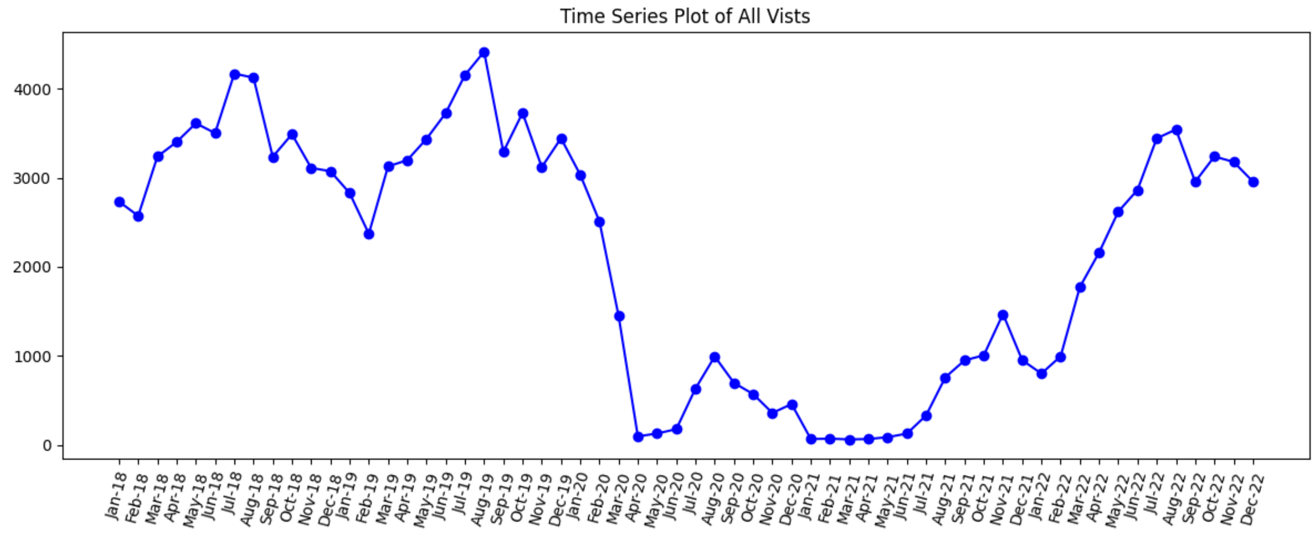

Time Series Plot of LN Europe
LN Europe
8
7
6
5
4
3
Index
1
5
10
15
20
25
30
35
40
45
50

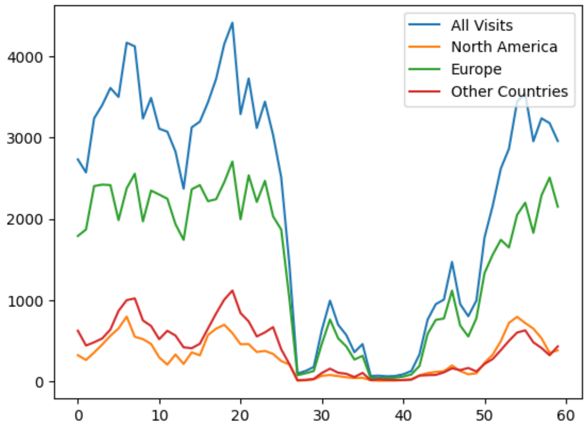

Time Series Plot of LN North America
LN North America
7
6
5
4
3
2
Index
1
5
10
15
20
25
30
35
40
45
50

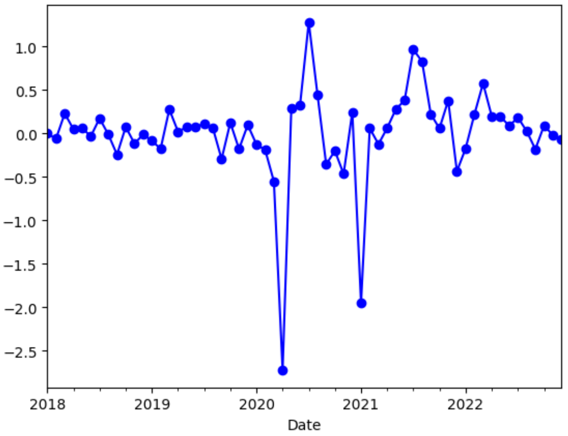

Time Series Plot of LN Other Countries
LN Other Countries
7
6
5
4
3
2
Index
1
5
10
15
20
25
30
35
40
45
50

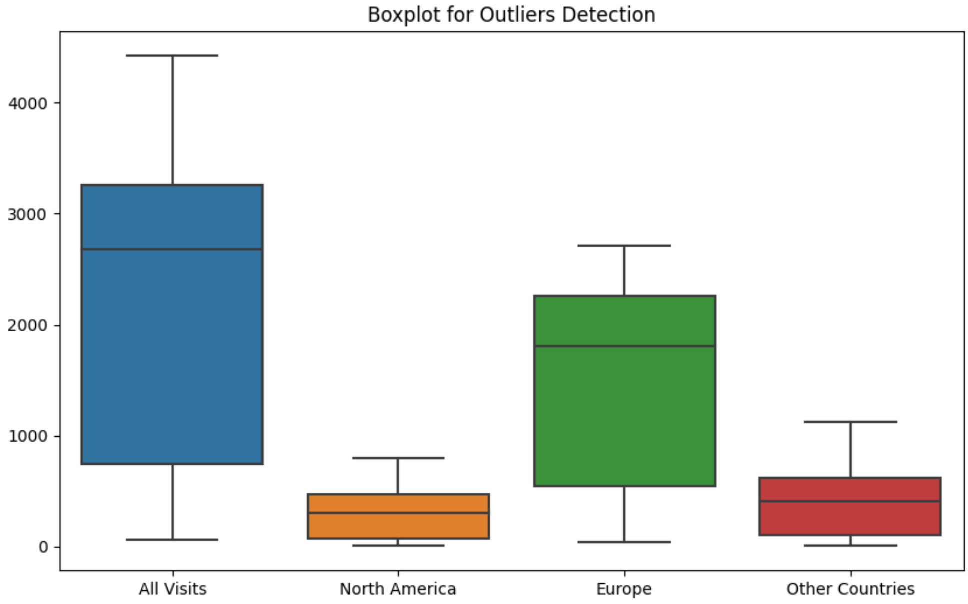

Time Series Plot of North America (Thousands)
North America (Thousands)
800
700
600
500
400
300
200
100
0
Index
1
5
10
15
20
25
30
35
40
45
50

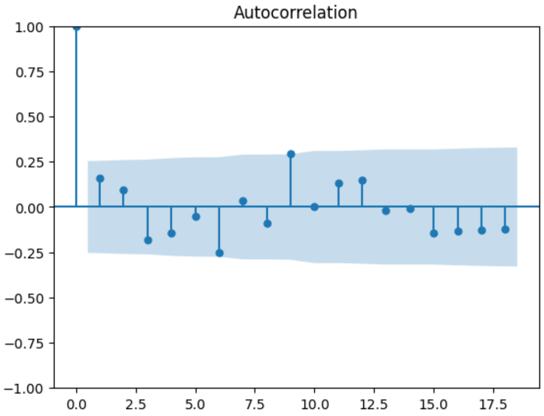

Time Series Plot of Other Countries (Thousands)
Other Countries (Thousands)
1200
1000
800
600
400
200
0
Index
1
5
10
15
20
25
30
35
40
45
50

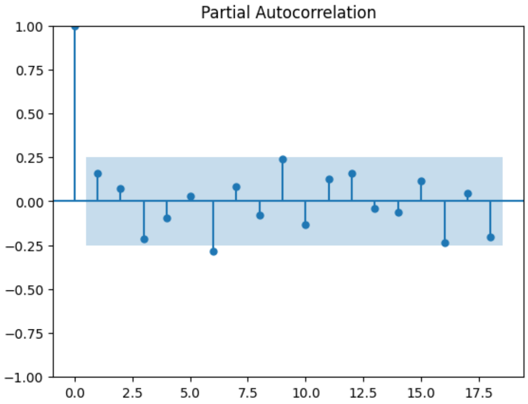

Time Series Plot of All Vists
0
1000
2000
3000
4000
Jan-18
Feb-18
Mar-18
Apr-18
May-18
Jun-18
Jul-18
Aug-18
Sep-18
Oct-18
Nov-18
Dec-18
Jan-19
Feb-19
Mar-19
Apr-19
May-19
Jun-19
Jul-19
Aug-19
Sep-19
Oct-19
Nov-19
Dec-19
Jan-20
Feb-20
Mar-20
Apr-20
May-20
Jun-20
Jul-20
Aug-20
Sep-20
Oct-20
Nov-20
Dec-20
Jan-21
Feb-21
Mar-21
Apr-21
May-21
Jun-21
Jul-21
Aug-21
Sep-21
Oct-21
Nov-21
Dec-21
Jan-22
Feb-22
Mar-22
Apr-22
May-22
Jun-22
Jul-22
Aug-22
Sep-22
Oct-22
Nov-22
Dec-22

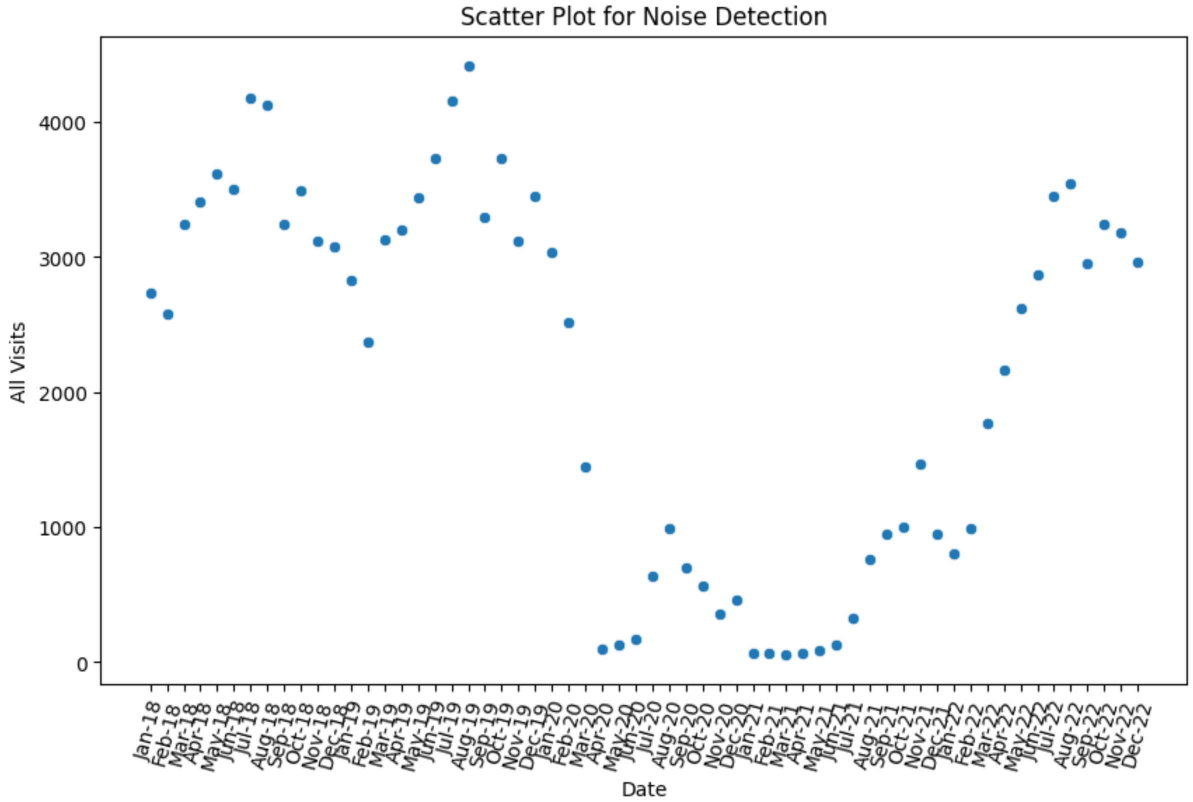

All Visits
North America
Europe
Other Countries
0
1000
2000
3000
4000
0
10
20
30
40
50
60

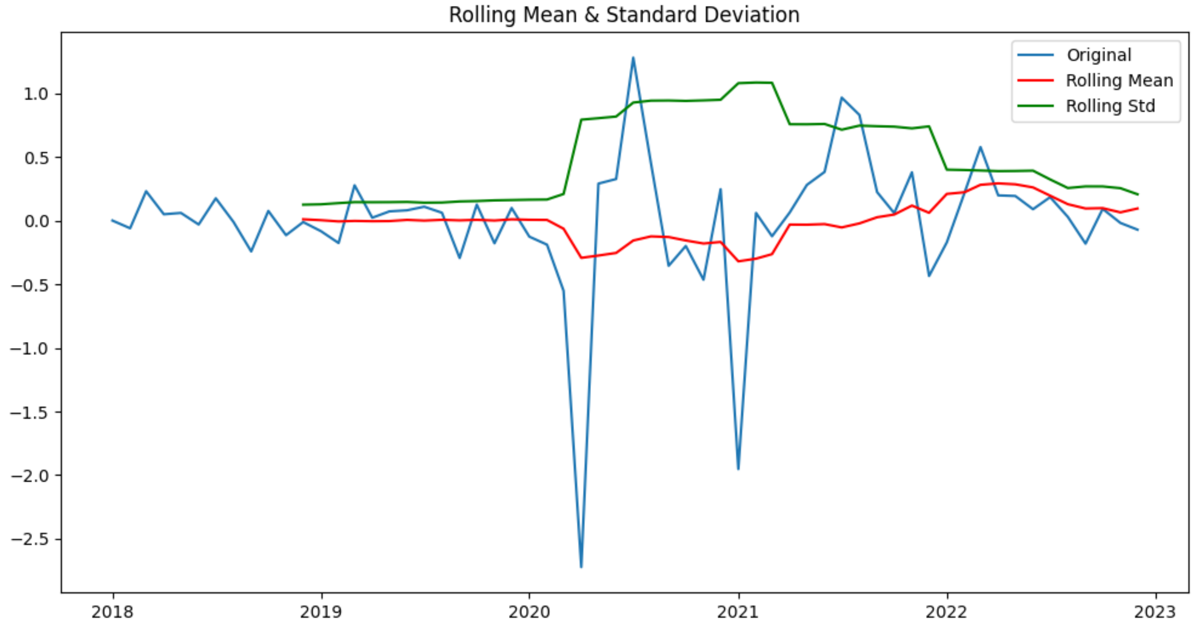

1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
2018
2019
2020
2021
2022
Date

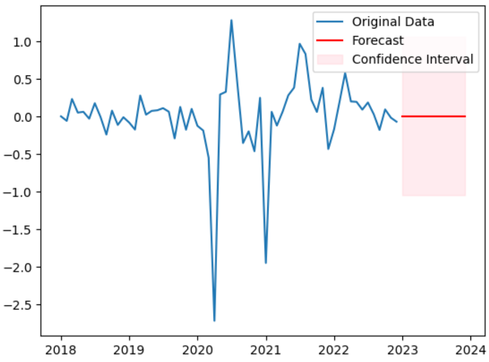

Boxplot for Outliers Detection
4000
3000
2000
1000
0
All Visits
North America
Europe
Other Countries

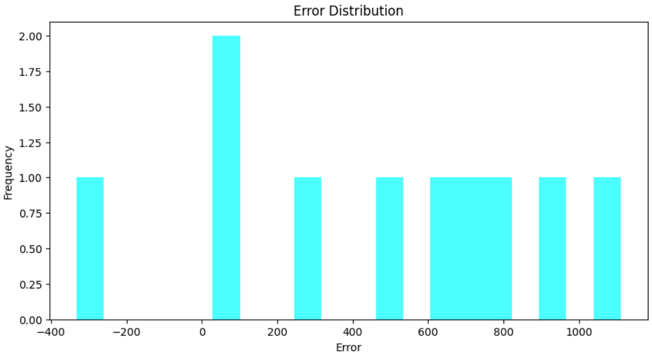

Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

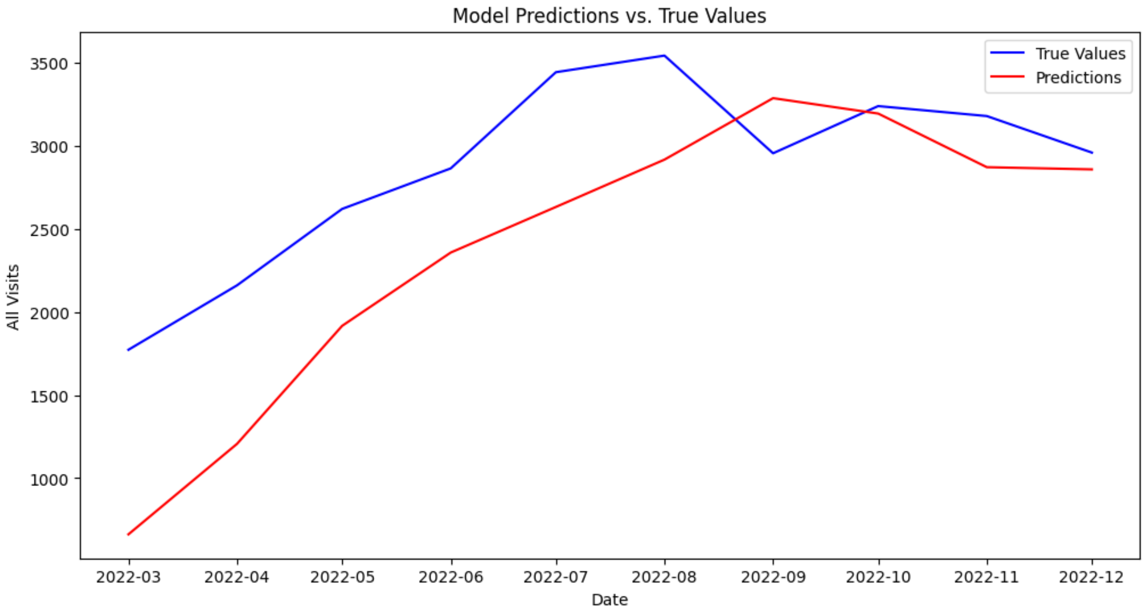

Partial Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

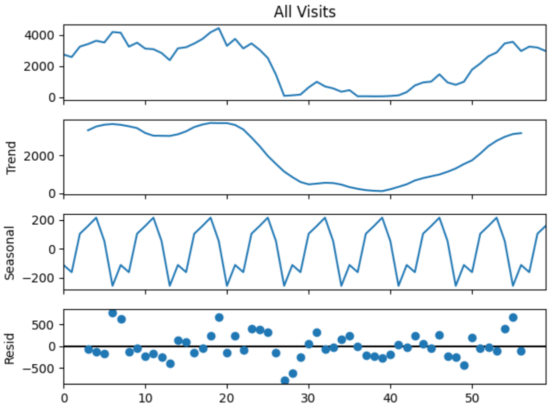

Scatter Plot for Noise Detection
All Visits
0
1000
2000
3000
4000
Date
Jan-18
Feb-18
Mar-18
Apr-18
May-18
Jun-18
Jul-18
Aug-18
Sep-18
Oct-18
Nov-18
Dec-18
Jan-19
Feb-19
Mar-19
Apr-19
May-19
Jun-19
Jul-19
Aug-19
Sep-19
Oct-19
Nov-19
Dec-19
Jan-20
Feb-20
Mar-20
Apr-20
May-20
Jun-20
Jul-20
Aug-20
Sep-20
Oct-20
Nov-20
Dec-20
Jan-21
Feb-21
Mar-21
Apr-21
May-21
Jun-21
Jul-21
Aug-21
Sep-21
Oct-21
Nov-21
Dec-21
Jan-22
Feb-22
Mar-22
Apr-22
May-22
Jun-22
Jul-22
Aug-22
Sep-22
Oct-22
Nov-22
Dec-22

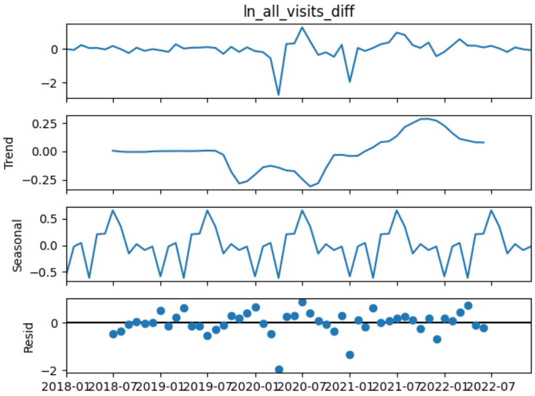

Rolling Mean & Standard Deviation
Original
Rolling Mean
Rolling Std
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
2018
2019
2020
2021
2022
2023

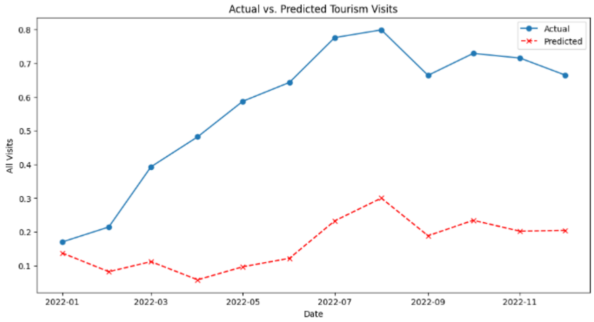

Original Data
Forecast
Confidence Interval
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
2018
2019
2020
2021
2022
2023
2024

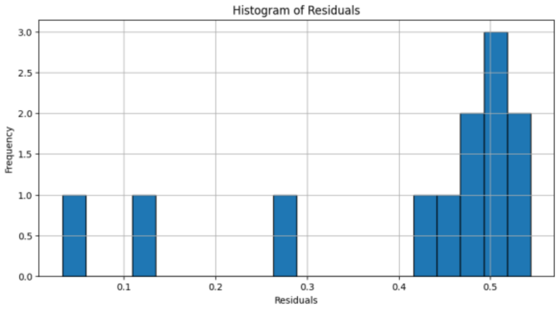

Error Distribution
Frequency
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
−400
−200
0
200
400
600
800
1000
Error

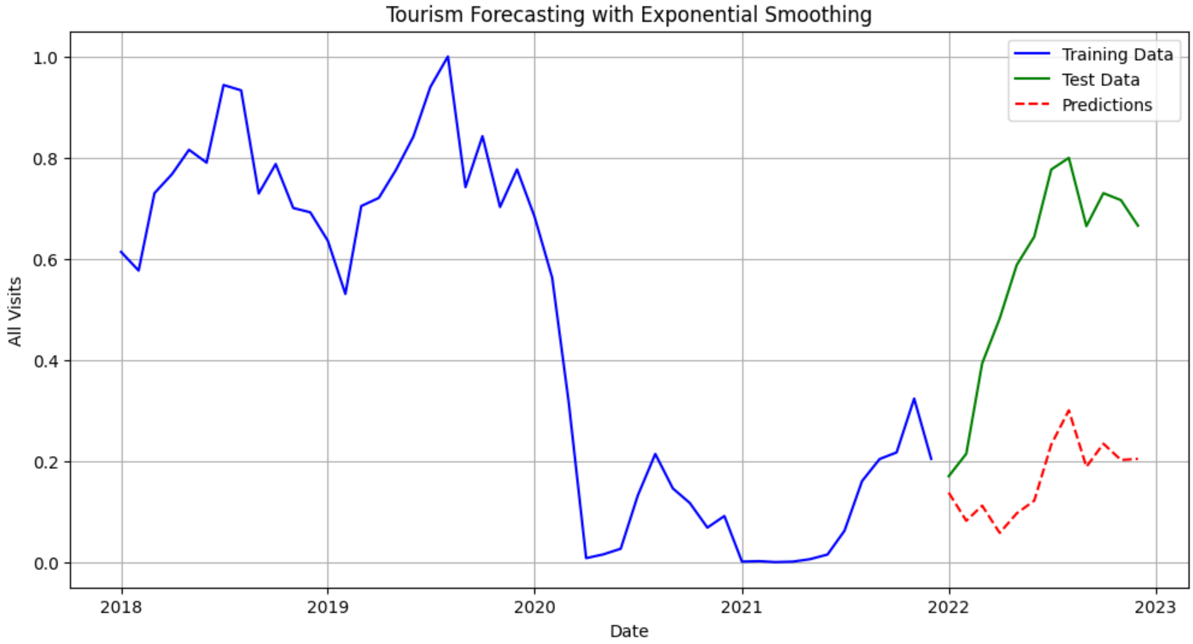

Model Predictions vs. True Values
True Values
Predictions
All Visits
3500
3000
2500
2000
1500
1000
2022-03
2022-04
2022-05
2022-06
2022-07
2022-08
2022-09
2022-10
2022-11
2022-12
Date

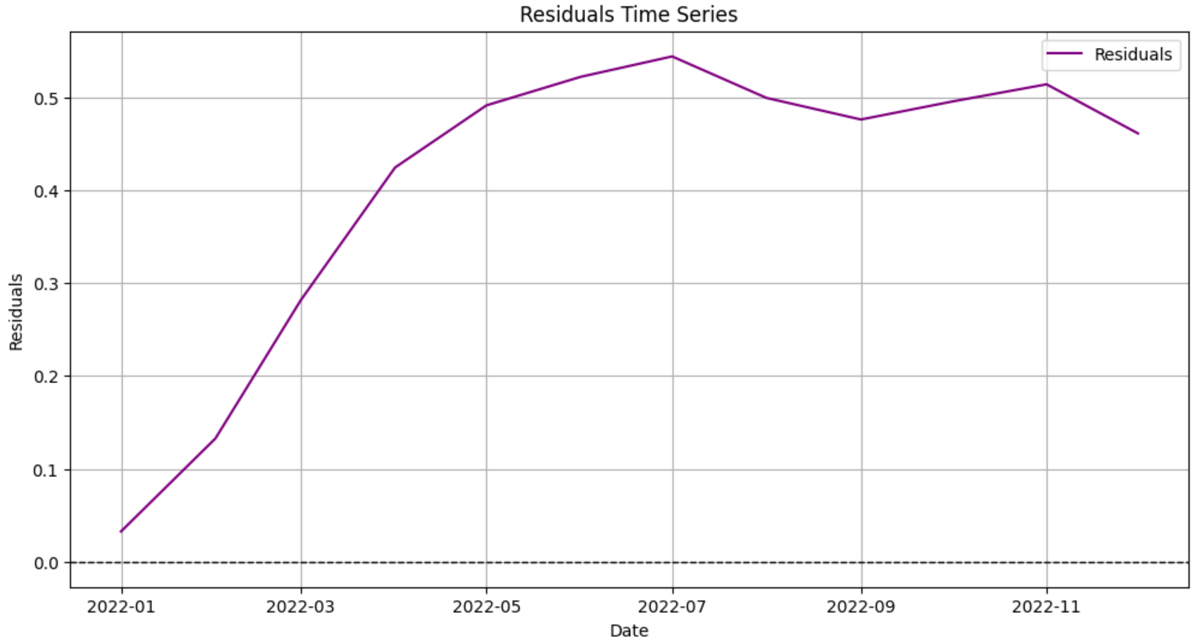

All Visits
4000
2000
0
Trend
2000
0
Seasonal
200
0
−200
Resid
500
0
−500
0
10
20
30
40
50


ln_all_visits_diff
0
−2
Trend
0.25
0.00
−0.25
Seasonal
0.5
0.0
−0.5
Resid
0
−2
2018-01
2018-07
2019-01
2019-07
2020-01
2020-07
2021-01
2021-07
2022-01
2022-07


Actual vs. Predicted Tourism Visits
Actual
Predicted
All Visits
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
2022-01
2022-03
2022-05
2022-07
2022-09
2022-11
Date


Histogram of Residuals
Frequency
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0.1
0.2
0.3
0.4
0.5
Residuals


Tourism Forecasting with Exponential Smoothing
Training Data
Test Data
Predictions
All Visits
0.0
0.2
0.4
0.6
0.8
1.0
2018
2019
2020
2021
2022
2023
Date


Residuals Time Series
Residuals
Residuals
0.0
0.1
0.2
0.3
0.4
0.5
2022-01
2022-03
2022-05
2022-07
2022-09
2022-11
Date