Proposed title of the project

EFFECT OF CLIMATIC CONDITION ON SEASIDE TOURIST CENTRES

RUKAYAT A BALOGUN

Student number

2311143

DR. ANCHAL

Program

MASTERS DEGREE

Department

DATA ANALYTICS AND TECHNOLOGIES

Institution

UNIVERSITY OF BOLTON

Date

15/05/2024

Table of Contents

TABLE OF FIGURES

INTRODUCTION

Tourism, as a global phenomenon, has witnessed remarkable growth and evolution over the years, with international tourist arrivals consistently increasing. Projections suggest a substantial rise to 1.8 billion international tourist by 2030, as indicated by the UNWTO in 2020. However, the onset of the COVID-19 pandemic has disrupted this upward trend, profoundly impacting the tourism industry worldwide. Travel restrictions, lockdown measures, and heightened health and safety concerns have led to significant declines in both international and domestic tourism activities, as evidenced by studies such as Goossling et al.(2020) and Seyfi et al (2022,2023)

While international tourism often receives considerable attention, domestic tourism accounts for the majority of tourism activities, comprising nearly 85% of global travel, as highlighted by Hall (2015). Domestic travelers, whether for leisure, business, or other purposes, constate a significant portion of the tourism industry. Prior to the pandemic, the annual number of tourist trips exceeded the global population, underscoring the immense scale and impact of tourism as a socio-economic phenomenon.

The tourism sector plays a vital role in the global economy, contributing substantially to revenue generation and employment opportunities. In 2022, for instance, the tourism industry contributed approximately 7.6% of the world’s GDP, amounting to an estimated US$8.6 trillion, according to data from the World Travel & Tourism Council (WTTC) in 2023. This undersocres the sector’s significance as a driver of economic growth and development on a global scale.

However, the COVID-19 pandemic has highlighted the vulnerability of the tourism industry to external shocks and crises. The widespread disruptions caused by the pandemic have necessitated a re-evaluation of tourism strategies, and other stakeholders are now tasked with navigating the challenges posed by the pandemic while also planning for a post-pandemic recovery that prioritizes sustainability and resilience in the tourism sector (Ramkissoon, 2023) According to Demeter, Tourism seasonality is a pervasive phenomenon affecting destinations worldwide, with significant implications for economies, communities, and environments. Defined as the systematic variation in tourist demand throughout the year, seasonality manifest through fluctuations in visitor arrivals, overnights stay, and economic activity within tourism sectors. While seasonality is inherent to many destinations, its impact varies across regions, influenced by a myriad of factors such as climate, culture, and infrastructure (Gursoy & Nunkoo, 2019).

In the European Union, islands stand as unique microcosms with diverse geographical, demographic, and economic landscapes. Despite their distinctiveness, islands share common challenges stemming from seasonality, encompassing sectors crucial to their economies, including tourism, transport, culture, and environment. The tourism industry plays a central role in island economies, yet faces persistent challenges related to fluctuating demand, particularly evident in certain months (Hall, 2015).

The irregular distribution of tourist demand poses multifaceted consequences for island destinations, affecting prices, labour markets, corporate returns, local populations, and environmental sustainability (Hadinejad,2019). The seasonality of tourist demand emerges as a central concern for island stakeholders, presenting ongoing challenges for both businesses and local governments alike (Higgins-Desbiolles & Bigby 2023).

A wealth of research has endeavoured to understand the complexities of tourism seasonality, employing various methodologies and indicators to delineate primary dimensions and drivers. From statistical indices measuring fluctuations in demand to qualitative analyses exploring factors and effects, scholars have sought to unravel the nuance of seasonality in tourism destinations (WTTO,2020).

However, despite the abundance of research on tourism seasonality, there remains a dearth of studies comprehensively analysing the performance of seasonality in island present distinct challenges and opportunities regarding tourism seasonality. The existing literature underscores the critical need for research that delves into the nuanced dynamics of seasonality of islands, providing valuable insights for sustainable development and management of tourism sectors.

Therefore, in this study, we would be looking at the effect of temperature variability on the seaside resort. Climatic factor responsible of the atmospheric condition of a place is enormous which we cannot cover in this study. However, this research will focus on tourist centre at the beach side. Our study area is Blackpool, a seaside resort on the Irish Sea coast of England.

BACKGROUND

Tourism stands as a cornerstone of global economic development, contributing substantially to growth, employment, and society advancement. However, despite its significance, a persistent challenge lies in the lack of comprehensive statistical frameworks to represent touristic activities spatially across regions, thereby impeding a thorough understanding of the myriad factors influencing tourism resilience. This limitation becomes particularly pronounced when addressing complex and dynamic phenomena, such as climate-related extreme events, which pose significant threats to tourism sustainability and adaptation efforts. The majority of the amount spent, and number of activities completed during tourism and outdoor leisure visits with Great Britain took place in an urban setting. According to the research carried out by the Office of the National Statistics, it shows the increase in expenditure within urban locations, which has risen 56% from £4.6 billion in 2011 to a high of £7.3 billion in 2019 as shown in Figure 1. Within the same period, all other habitats have seen expenditure fall, despite an increase in the total amount being spent.

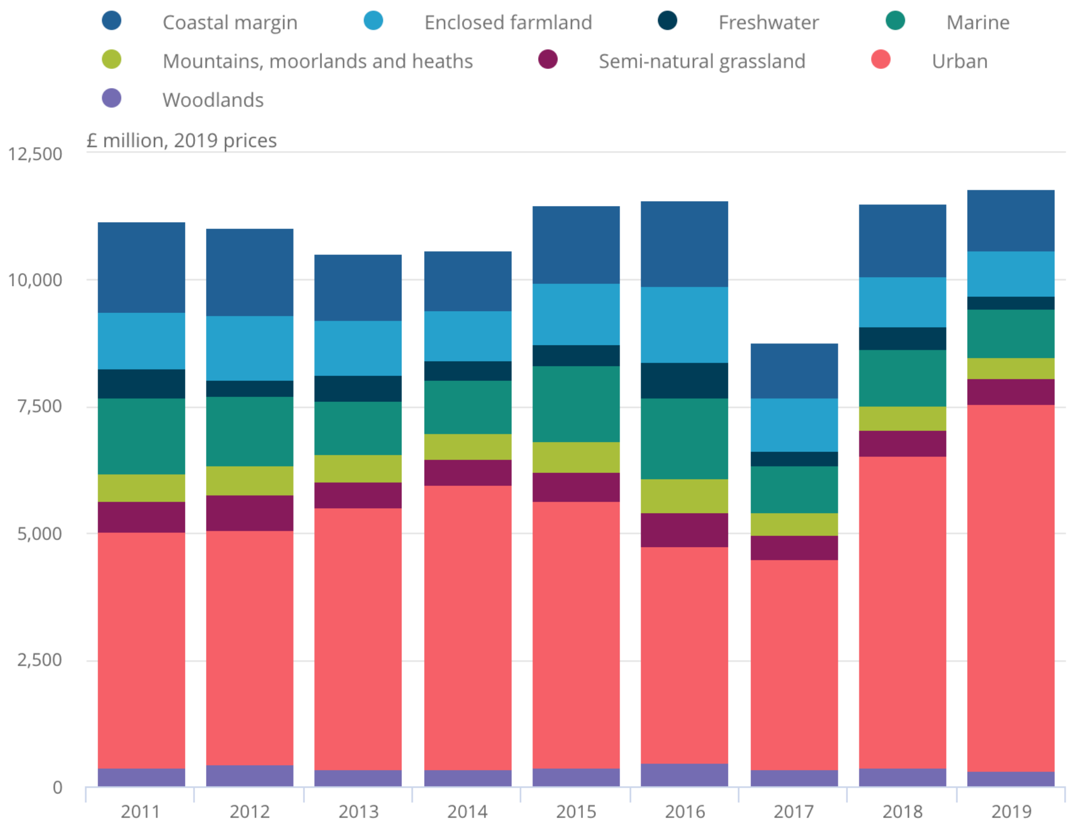

Figure 1: ANNUAL CONTRIBUTION OF THE NATURAL ENVIRONMENT TO TOURISM

In Europe, climate-related disasters have emerged as a major concern for the tourism industry, exemplified by recent events like the wildfires in Greece and drought-induced heatwaves across the continent. These extreme weather events can lead to physical damage to attraction sites, shifts in visitor behaviors and satisfaction, and cascading economic impacts, underscoring the vulnerability of the tourism sector to climate variability and change. Over the long term, climate-related extremes could reshape the regional distribution of physical resources supporting tourism, further exacerbating the sectors's vulnerability.

To address the complex temporal and spartial dimensions of tourism resilence, it is imperative to transcend traditional statistical approaches constrained by limited spatial and temporal dimensions. Recent advancements in tourism-related data availability and big data analytics offer promising avenues for enhancing the geographical localization of touristic destinations and incorporating spatially-relevant information into tourism management strategies. Social media platforms, such as TripAdvisor, play a pivotal role in characterizing attractiveness, identifying attraction density, and evaluating tourism experiences, providing valuable insights for understanding tourism dynamics.

The integration of social media data with statistical and spatial tourism data has emerged as a promising approach for analysing spatial patterns of tourism activity, uncovering key spatiotemporal trends, and enhancing decision-making in tourism management and policy development. A spatially-explicit and consistent tourism-related database olds immense potential beyond tourism management, offering valuable insights for disaster risk management and climate change adaptation. Detailed spatial information pertaining to touristic attractions can enhance preparedness and resilience efforts, enabling more effective emergency response planning and the development of adaptive strategies to reduce vulnerability to natural hazards.

According to the Office of National Statistics, overseas residents made a total of 10.9 million visits to the UK during Quarter 3 (July to September)2023. This is an increase of 1 million visits when compared with the same period in 2022 (9.9 million). Figure 2 shows the number of visits in the third quarter, 2023 is still lower than pre-coronavirus (COVID-19) pandemic levels, where in 2019, 11.9 million visits were made to the UK by overseas residents in the third quarter, 2019. This is a decrease of around 8% between Q3 2019 and Q3 2023.

Figure 1: Overseas residents' visits to the UK increased by around 1 million in Quarter 3 (July to Sept) 2023 compared with Quarter 3 2022

Quarter 1 (Jan to Mar) 2017 to Quarter 4 (Oct to Dec) 2020 (all modes) and Quarter 1 to Quarter 2 (Apr to June) 2021 (air only) with the return to all modes from Quarter 3 2021 to Quarter 3 2023

Source: International Passenger Survey and Overseas travel and tourism from the Office for National Statistics

Figure 2: OVERSEAS RESIDENTS' VISITS TO THE UK

Tourism and travel are significant contributors to global carbon dioxide (CO2) emissions, accounting for 4.9% of emissions in 2010, with more than half attributed to air transport (Gossling & Peeters, 2015). However, the COVID-19 pandemic led to a drastic reduction in air travel, resulting in a decline of between 60 and 80% in air transport emissions, while domestic tourism saw an increase in many countries (Gossling & Higham, 2021). Despite hopes for more sustainable tourism development post-COVOD-19, government investments to save airlines have limited economic effects and may exacerbate climate change.

Prior to the pandemic, a 'business-as-usual' growth scenario forecasted tourism to consume a significant portion of the remaining carbon budget between 2015 and 2100, while contributing only 3.2% of global GDP (Peeters, 2017 ; WTTC, 2018). These forecasts were slightly reduced by COVID-19 lockdowns, according to recent studies (Peeters & Papp, 2023).

Transportation to and from destinations (O/D) accounts for approximately 75% of tourism's energy use and emissions, with air transport contributing over 20% of all trips but about 50% of emissions (Gossling & Peeters, 2015). Long-haul trips particularly those exceeding 6000km, contribute significantly to emissions (Peeters & Landre, 2012). Strategies to reduce emissions include attracting closer markets, encouraging longer stays and fewer trips, and transitioning away from energy-intensive modes, especially flying (Kamb 2021).

The disproportionate focus on international tourism distorts the understanding of the tourism system, while domestic tourism offers opportunities for economic development with fewer emissions from long-haul flights (WTTC, 2018). All stakeholders, including governments, international organizations, the tourism industry, destinations, consumers, and research networks, must collaborate to address the severity of the problem, as emphasized by the Davos Declaration on Climate Change and Tourism (Scott, 2021).

However, changing consumer behavior regarding flying poses a significant challenge due to the high attitude-behavior gap, necessitating substantial policy changes and shifts in the supply of tourism products (Cohen 2016). Addressing these challenges will require concerted efforts from various stakeholders to transition towards more sustainable tourism practices and mitigate the sector's impact on climate change.

STUDY AREA

Figure 3: LANDSCAPE OF BLACKPOOL RESORT

Blackpool, a vibrant seaside resort town in Lancashire, England, has a rich history dating back to its origins as a small hamlet. Its growth began in the mid-eighteenth century, coinciding with the popularity of sea bathing for health reasons. The town's picturesque beach made it an ideal destination, leading to the construction of several hotels by 1781.

The establishment of a railway station in 1846 further fueled Blackpool's expansion by increasing accessibility to the resort. Throughout the nineteenth century, Blackpool continued to attract more visitors, becoming a bustling borough in 1876 with a population of 141,000 according to the office of National Statistics from census in 2021. Its growth was closely linked to the Lancashire cotton-mill tradition of annual factory maintenance shutdowns, knows as wakes weeks, during which many workers opted to vacation by the seaside.

However, in the late 20th century, evolving holiday preferences and the rise of overseas travel posed challenges to Blackpool's status as a premier resort destination. Despite economic hurdles, the town's identity and economy remained rooted in tourism. Today, Blackpool's iconic seafront boasts attractions such as the renowned Blackpool Tower, dazzling illuminations, thrilling Pleasure Beach amusement park, and historic Winter Gardens, continuing to attract millions of visitors each year.

Blackpool's resilience and ability to adapt to changing trends underscore its enduring appeal as a must-visit destination, blending rich history with modern attractions to offer a memorable experience for all who venture to its shores.

Blackpool experiences a temperature maritime climate, classifirs according to the Koppen system. This climate typically entails cool summers, frequent overcast skies, and minor fluctuations in annual temperature ranges.

Extreme temperature records in Blackpool reflect this temperate climate patten. The lowest temperature recorded was -15.1 ^{o}C(4.8 ^{o}F) in December 1981, with an even colder reading of -18.3 ^{o}C(-0.9 ^{o}F) noted in January 1881. On the other end of the spectrum, the highest temperature on record soared to 37.2 ^{o}C(99.0 ^{o}F) during a heatwave in 2022 while an average summer typically sees the warmest temperatures reaching around 28.5 ^{o}C(83.3 ^{o}F) between 1991 and 2020.

In terms of precipitation, Blackpool receives an average of slightly less than 900mm (35 in) annually, with measurable precipitation occurring on approximately 147 days throughout the year. These climatic characteristics contribute to the overall ambiance and weather patterns experienced by residents and visitors alike in Blackpool.

Climate data for Blackpool (BLK),[a] elevation: 10 m (33 ft), 1991–2020 normals, extremes 1960–present [hide]													
Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
Record high °C (°F)	14.3 (57.7)	18.4 (65.1)	19.4 (66.9)	24.4 (75.9)	28.6 (83.5)	31.3 (88.3)	37.2 (99.0)	32.2 (90.0)	30.0 (86.0)	26.2 (79.2)	16.8 (62.2)	15.0 (59.0)	37.2 (99.0)
Mean maximum °C (°F)	10.8 (51.4)	11.1 (52.0)	14.8 (58.6)	19.0 (66.2)	23.4 (74.1)	25.5 (77.9)	26.3 (79.3)	25.2 (77.4)	22.9 (73.2)	18.7 (65.7)	14.1 (57.4)	11.7 (53.1)	28.5 (83.3)
Mean daily maximum °C (°F)	7.3 (45.1)	7.8 (46.0)	9.7 (49.5)	12.6 (54.7)	15.7 (60.3)	18.1 (64.6)	19.8 (67.6)	19.5 (67.1)	17.6 (63.7)	14.1 (57.4)	10.4 (50.7)	7.9 (46.2)	13.4 (56.1)
Daily mean °C (°F)	4.8 (40.6)	5.0 (41.0)	6.6 (43.9)	8.9 (48.0)	11.8 (53.2)	14.5 (58.1)	16.4 (61.5)	16.3 (61.3)	14.1 (57.4)	11.1 (52.0)	7.7 (45.9)	5.2 (41.4)	10.2 (50.4)
Mean daily minimum °C (°F)	2.3 (36.1)	2.2 (36.0)	3.4 (38.1)	5.1 (41.2)	7.9 (46.2)	10.9 (51.6)	12.9 (55.2)	13.0 (55.4)	10.6 (51.1)	8.0 (46.4)	5.0 (41.0)	2.5 (36.5)	7.0 (44.6)
Mean minimum °C (°F)	−3.8 (25.2)	−3.1 (26.4)	−2.4 (27.7)	−0.4 (31.3)	2.6 (36.7)	6.7 (44.1)	9.2 (48.6)	7.9 (46.2)	4.3 (39.7)	1.5 (34.7)	−1.9 (28.6)	−4.5 (23.9)	−5.9 (21.4)
Record low °C (°F)	−11.5 (11.3)	−13.2 (8.2)	−9.7 (14.5)	−6.1 (21.0)	−1.9 (28.6)	−1.0 (30.2)	3.3 (37.9)	1.9 (35.4)	−0.7 (30.7)	−4.3 (24.3)	−7.0 (19.4)	−15.1 (4.8)	−15.1 (4.8)
Average precipitation mm (inches)	77.8 (3.06)	64.0 (2.52)	54.4 (2.14)	48.7 (1.92)	54.0 (2.13)	63.1 (2.48)	66.0 (2.60)	79.9 (3.15)	83.5 (3.29)	101.4 (3.99)	94.7 (3.73)	99.1 (3.90)	886.6 (34.91)
Average precipitation days (≥ 1.0 mm)	14.4	11.4	11.2	9.9	9.9	10.1	10.9	12.2	11.6	14.4	15.7	15.6	147.3
Mean monthly sunshine hours	55.0	80.4	119.3	175.5	217.9	210.1	201.1	182.6	141.8	98.0	60.7	49.3	1,591.7

Figure 4: AVERAGE WEATHER CONDITION OF BLACKPOOL FROM 1991 -2020

PROBLEM STATEMENT

The impact of climatic conditions on tourist centres is a critical issue that requires thorough examination and strategic planning to ensure the sustainability and resilience of tourism destinations. Various studies have shed light on the intricate relationship between weather patterns, events, and environmental factors, and their influence on tourist behaviour and destination preferences. Understanding these dynamics is essential for effectively managing tourist centres and addressing challenges posed by changing climatic conditions.

For instance, research by (Kumar et al, 2023) underscores the pivotal role of weather data in forecasting travel demand. By integrating meteorological information into predictive models, researchers aim to enhance the accuracy of travel forecasts. This is particularly important in destinations where weather plays a significant role in attracting visitors, such as coastal resorts or ski destinations. For example, the success of beach resorts like Blackpool in Lancashire, England relies heavily on favourable weather conditions for activities such as sunbathing and swimming.

Similarly, (Hendricks et al., 2011) explores the influence of events on travel demand, advocating for the inclusion of event data in forecasting models. Events such as music festivals, sports tournaments, and cultural celebrations can significantly impact tourist activity and visitor numbers. For instance, the annual Oktoberfest in Munich, Germany, attracts millions of visitors from around the world, highlighting the importance of considering event-related factors in tourism planning and management.

Moreover, studies conducted by (Shah et al., 2023) examines the specific effects of daily weather fluctuations on consistent snowfall and favourable weather conditions for attracting visitors during the winter season. Understanding how weather patterns impact ski resort visitation is crucial for optimizing resource allocation and promoting sustainable tourism practices in these regions.

In addition to weather-related impacts, (Lionetti,2023) explores the seasonal variations in weather and their effects on foreign visitor arrivals in region like India. For example, the monsoon season in India can have a significant impact on tourist arrivals, with intense rainfall and warmer temperatures affecting travel patterns and tourism-related activities. By anticipating and preparing for these seasonal variations, destinations can develop strategies to mitigate potential disruptions and ensure a positive visitor experience.

Furthermore, research by (Rutty et al., 2015) examines the relationship between temperature, travel, and the spread of COVID-19, highlighting the complex interplay between climatic conditions public health, and tourism. Destinations around the world have faced challenges in managing tourist flow and mitigating the risk of virus transmission, particularly in areas with extreme weather conditions or seasonal variations. Understanding how climatic conditions influence the spread of infectious diseases is essential for developing effective public health and tourism management strategies.

AIMS

Therefore, the aim of this research is to determine the effect of temperature variability on the seaside resort during the month of May in Blackpool.

OBJECTIVES

1. Discover the impact of temperature on tourist attraction centres.

2. Predict the temperature during the month of May using statistical and machine learning models.

3. Compare the statistical and machine learning models used for prediction using evaluation metrics.

4. Determine the impact of the prediction result in helping stakeholders make timely decision.

REPORT CONTENT

The content of this report in divided into chapters with chapter one including the introduction and background of the study, chapter 2 details the methodology adopted in the report while the last chapter details the implementation, result discussion and implication of study.

METHODOLOGY

MACHINE LEARNING AND STATISTICS MODEL

For our prediction task, we will employ three machine learning algorithms, namely Linear Regression, Decision Tree, Random Forest, alongside two statistical modelling techniques: Linear Regression and ARIMA (AutoRegressive Integrated Moving Average) model.

STATISTICAL TECHNIQUES

ARIMA MODEL

ARIMA (AutoRegressive Integrated Moving Average) model is a statistical technique used for time series forecasting. It models the relationship between a dependent variable and its lagged values, as well as the errors or residuals. ARIMA is particularly useful for analysing and forecasting time series data with trends, seasonality, and autocorrelation. Autoregressive models assume that Y_t is a linear function of the preceding values and is given by the equation

$$Y_t = \alpha_1 Y_{t-1} + \varepsilon_t$$

An ARIMA model is labelled as an ARIMA model (p, d, q), wherein:

p is the number of autoregressive terms;

d is the number of differences; and

q Is the number of moving averages.

To apply the ARIMA model, several key steps must be followed:

Model Identification

Initially, we converted our data to time series data. Then, we asses if the data is stationary, implying a constant mean throughout the time series. Stationarity can be verified through visual inspection or by conducting the Augmented Dickey-Fuller (ADF) test. This test determines if the series is non-stationary, indicating fluctuations in mean over time. If the data is non-stationary, differencing is performed until stationarity is achieved using the ‘diff’ function in R.

Identification of Parameters (ACF and PACF)

Once stationarity is established, we proceed to identify the appropriate parameters for the ARIMA model. This involves examining the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. A correlogram displays autocorrelation, indicating the correlation between a variable at time (t) and the same variable at time t-k. The partial correlogram removes shorter autocorrelation lags when calculating correlation at longer lags. Additionally, data decomposition may be necessary to remove seasonal trends, with the frequency determined based on the data.

Selection of the Best ARIMA Model:

Next, we select the most suitable ARIMA model for prediction. This can be achieved manually by analyzing ACF and PACF plots or automatically using functions like auto.arima(). The goal is to find the model that best fits the data and provides accurate predictions.

Forecasting

Once the best ARIMA model is identifies, forecasting for future periods can be conducted. This involves generating predictions based on the model parameters and historical data. Forecast accuracy should be evaluated to assess the reliability of the model.

MACHINE LEARNING TECHNIQUES

LINEAR REGRESSION

Linear regression analysis serves the purpose of describing the relationship between two or more variables, with one variable being dependent and the others independent. The dependent variable is the outcome to be explained, while the independent variable(s) influence or impact the dependent variable. This relationship is represented in a scatter plot while the distribution will show its linearity as shown in Figure 5. This relationship is depicted through a mathematical formula:

$$E(Y) = \beta_o + \beta_1 X$$

Where

E(Y) represents the expected value of the dependent variable

X denotes the independent variable

βo is the y-axis intercept, and

β_1 is the regression coefficient, signifying the slope of the line (Schneider, Hommel & Blettner, 2010)

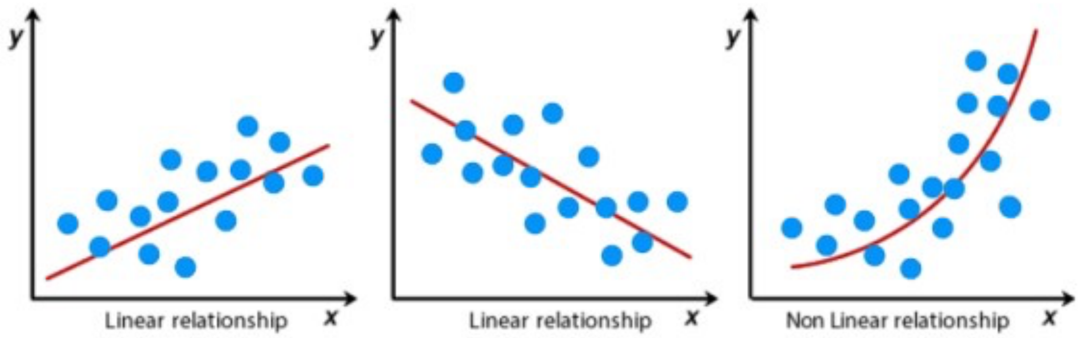

Figure 5: LINEAR RELATIONSHIP GRAPH

The figure 5, shows a positive relationship, negative relation and a non linear relationship. Regression analysis is vital for the following reasons:

1. describing and analyzing the strength of association between dependent and independent variables (Schneider, Hommel & Blettner, 2010)
2. Adjusting for the effect of covariates
3. Estimating the value of the dependent variable from observed independent variable values
4. Identifying significant risk factors affecting the dependent variable(Kumari & Yadav, 2018)

Several factors influence regression analysis, including sample size, missing data, and the nature of the sample. Assumptions of regression include linearity, normal distribution of data, homogeneity of variance, and independence (Chai & Draxler, 2014).

Different types of regression analysis exist, such as

1. linear regression,
2. logistic regression and
3. Multiple regression.

In this study, we focus on linear regression, specifically univariate linear regression, which examines the linear relationship between a dependent variable and one independent variable. The dependent variable must be continuous, while the independent variable(s) can be continuous or categorical. Initial assessment of the relationship between two continuous variables should be based on a scatter plot to determine linearity before conducting linear regression (Schneider, Hommel & Blettner, 2010).

RANDOM FOREST

Random Forest (RF) is a robust ensemble learning method developed by Breiman, blending bagging sampling and random feature selection to create a collection of decision trees with controlled variation. This technique is widely used for both classification and regression tasks due to its adaptability and effectiveness (Fawagreh, Gaber & Elyan, 2014). Unlike linear regression, which assumes linearity, RF excels in capturing nonlinear relationships present in the data, leading to more accurate predicitons (Schonlau & Zou, 2020).

At the core of the random forest algorithm are decision trees, which recursively split the dataset into subset based on certain criteria until a stopping condition is met. These trees (as shown in figure 6) are constructed for both classification and regression tasks, partitioning the data into leaf nodes to make predictions (Schonlau & Zou, 2020). However, decision trees are susceptible to overfitting, resulting in poor generalization accuracy. Random forests address this issue by aggregating multiple trees trained on boostrapped datasets, reducing overfitting through a process called bagging (Au,2018)

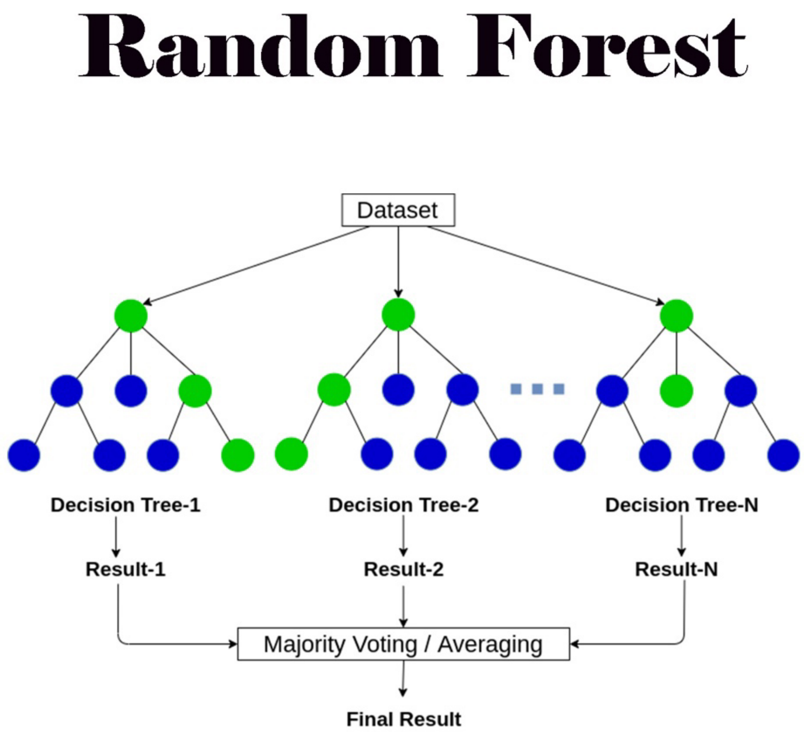

Figure 6: RANDOM FOREST

The predictive power of random forests lies in their ability to combine the outputs of multiple individual trees. Each tree is built on a boostrapped dataset, and at each node, a random subset of variables is considered for splitting, adding diversity to the trees (Couronne, Probst & Boulesteix, 2018). When making predictions, the class predicted by each tree is aggregated using a simple majority vote classification tasks, resulting in a final prediction.

DECISION TREE

Decision tree is a machine learning algorithm used for both classification and regression tasks. It is a versatile and intuitive method that recursively splits the dataset into subsets based on the most significant feature, aiming to create homogeneous groups with respect to the target variable.

The decision tree algorithm consists of nodes and branches. Each node represents a feature or attribute, and each branch represents a decision or rule based on that feature. The Decision tree as (shown in figure 7) starts with a root node and grows by recursively splitting the dataset into subset based on the feature that provides the best split, typically chosen to maximize information gain or minimize impurity.

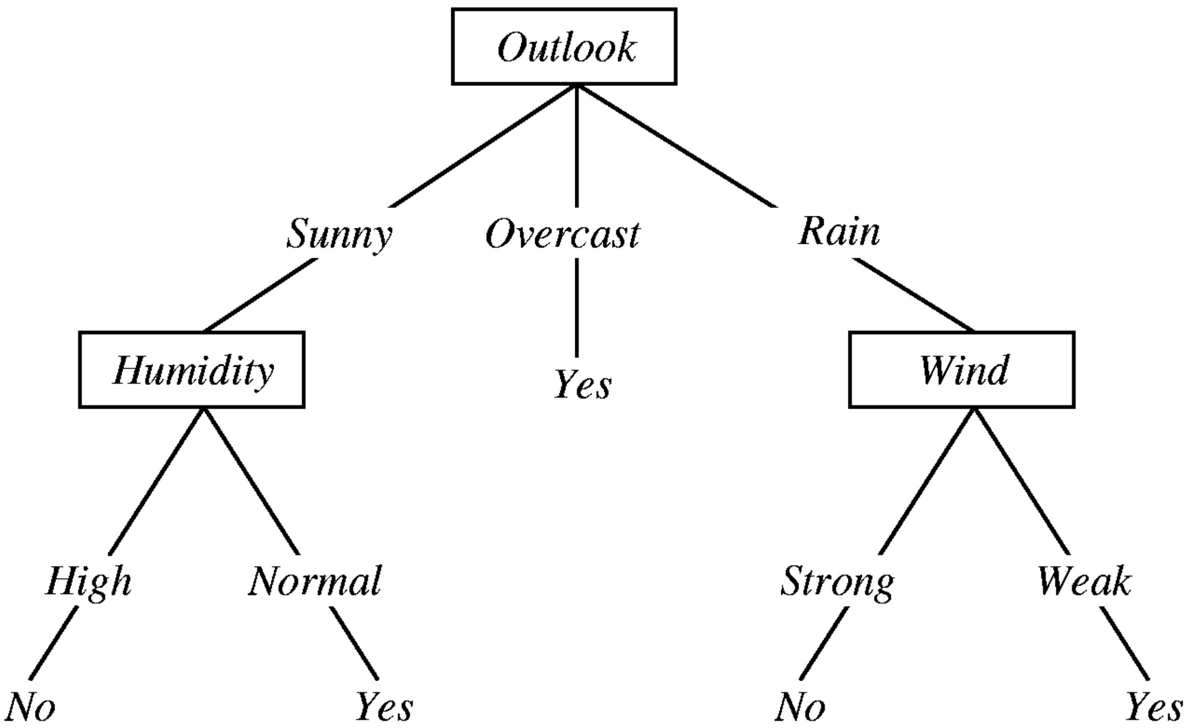

Figure 7: DECISION TREES

As the tree grows, it forms a hierarchy of nodes, with each internal node representing a decision based on a feature and each leaf node representing a final prediction or outcome. The splitting process continues until a stopping criterion is met, such as a maximum depth, minimum number of samples per leaf, or no further improvement in impurity reduction.

Decision trees are easy to interpret and visualize, making them useful for understanding the decision-making process. However, they are prone to overfitting, particularly when the tree depth is too large or the stopping criteria are not properly tuned. Overfitting occurs when the model captures noise or irrelevant patterns in the training data, leading to poor generalization performance on unseen data.

To address overfitting, various techniques can be employed, such as pruning, which removes unnecessary branches from the tree to present it from becoming too complex, or using ensemble methods like random forests, which combine multiple decision trees to improve predictive accuracy and robustness.

SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) is a supervised learning model used for both classification and regression tasks. When applied to regression, it is referred to as a Support Vector Regression (SVR). SVR can be linear or non-linear, depending on the kernel functions used. Linear SVR employs linear kernel functions, similar to linear SVMs

SVR, like its classification counterpart, utilizes kernels, sparse solutions, and VC control of the margin and number of support vectors. Despite being less popular than SVM, SVR has demonstrated effectiveness in estimating real-value functions.

One notable advantage of SVR is its computational efficiency, which remains consistent regardless of the dimensionality of the input space. Additionally, SVR exhibits excellent generalization capabilities, leading to high prediction accuracy in regression tasks (Marriette Awad and Rahul Khanna, 2015).

EVALUATION METRICS

MEAN ABSOLUTE ERROR (MAE)

MAE measures the average of the sum of absolute differences between observed values and predicted values and corresponds to the expected loss for the L1 loss function. The range of MAE is (0, +∞+∞) (Steurer, Hill & Pfeifer, 2021).

Mean Absolute Error (MAE) = Sum of Absolute Errors/ Number of Predictions

According to Chicco, Warrens & Jurman, (2021), MAE can be used if the outliers represent a corrupted part of the data as it doesn’t penalize too much the training data outliers this providing a generic and bounded performance measure for the model. But if the data also contains too many outliers, the model performance will be greatly depreciated. The MAE assess the average magnitude of errors in a group of predictions without taking into account their direction. It assess the precision of continuous variables (Ajala et al., 2022). With MAE, the smaller the valye, the higher the accuracy of the prediction model and one of the advantages of MAE is the unit is the same as the original data and it is easy to calculate and understand (Chicco, Warrens & Jurman, 2021).

MEAN SQUARED ERROR (MSE)

Mean squared error measures the squared difference between the estimated values and the true values (Steurer, Hill & Pfeifer, 2021). While according to Jierula et al., (2021), it is the sum, over all the data point, of the square of the difference between the predicted and actual target variables, divided by the number of data points. MSE is often referred to as quadratic loss since the penalty is related to the square of the error rather than the error itself (Ajala et al., 2022).

Mean Squared Error (MSE) = Sum of Squared Errors / Number of Predictions

Due to the squaring of the errors, the outliers are given more weight resulting in a smooth gradient for small errors and the negative values and positive values do not cancel each other out . Also, with an increase in error, MSE grows exponentially. The MSE value of a good model should be close to zero (Ajala et al., 2022). Therefore, the MSE unit is the square of the original unit and the range of MSE is (0, +∞); the smaller the MSE value is, the higher the accuracy of the prediction model. The perfect value of MSE is 0, indicating that the prediction model is perfect. MSE has defaulted as the loss function of linear regression in machine learning (Jierula et al., 2021)

ROOT MEAN SQUARED ERROR (RMSE)

RMSE measures the average magnitude of error between the predicted value and the actual value. Thus, RMSE is the average distance measured vertically from the actual value to the corresponding predicted value on the fit line. Simply, it is the square root of MSE (Jierula et al., 2021)

Root Mean Squared Error (RMSE = $\sqrt{MSE}$

RMSE is often preferred over MSE since it generates smaller values that are more easily compared across methods and hence easier to interpret for the user (Steurer, Hill & Pferifer, 2021). MSE is measured in units that are the square of the target variable, while RMSE is measured in the same units as the target variable.

RMSE is used when the error is highly non-linear. The RMSE indicates the number of errors in the predicted data on average and is a good measure of the prediction accuracy (Kaliappan et al., 2021). In the same manner as MSE, the range of RMSE is 0, +∞+∞);the smaller the RMSE value is, the higher the accuracy of the prediction model (Jierula et al., 2021).

For this study, we will be selecting RMSE as our metric for evaluation. MSE and RMSE are very sensitive to outliers therefore they penalize large errors and consequently prove to be very effective in improving model performance. Another advantage of RMSE over MAEs is that RMSEs avoid the use of absolute value and under the circumstances of calculating mode error sensivities or data assimilation applications, MAEs are definitely not prefereed over RMSEs. They can minimize greater errors even though it means accepting more frequent, smaller errors.

While MSE is a helpful metric, it is also difficult to interpret as it involves squaring of the error terms and it doesn't have the same unit as the value to be predicted. RMSE is not ambiguous in its meaning, uses the same unit as the target variable and it is more appropriate to use when model errors follow a normal distribution. The sensitivity of the RMSE to outliers is of its most important characteristics. In fact, the existence of outliers and their probability of occurrence is well described by the normal distribution underlying the use of RMSE (Chai & Draxler, 2014)

DISCUSSION OF FINDINGS

The primary objective of this study is to examine how variations in temperature over time effect the climatic condition of seaside tourist centres. I began with a dataset containing approximately 5451 rows of latitude and longitude coordinates, along with corresponding climate measures recorded at 3hours intervals. For this extensive dataset, I systematically selected a sequential subset of 248 rows for my analysus. Following the selection process, I conducted data cleansing procedures, which involved removing duplicates, missing values, eliminating characters and dealing with outliers.

I then proceeded to gain understanding of the data. I reshaped the data and restructured it to be in a more interpretable format. I proceeded to apply statistical analysis where I used linear interpolation to replace NA values. After dealing with the outliers, I applied linear regression to see the columns that were correlated and I observed Temperature and Humidity has strong correlation. The highest temperature experienced in this region is 14 degrees Celsius which is the typical temperature in the month of May.

I applied multicollinearity to show the relationship with several features in the data . I then converted the datetime to its format as it was in character format. After ensuring the data is clean and ready for analysis, I applied 4 machine learning algorithm to get the best fit for the dataset, Random Forest came out the best with the lowest RSME. I prepared my data for time series analysis using appropriate measures like tsoutliers to check for outliers and ARIMA for Autoregression. This also implies that our data is stationary and that the mean and variance are almost constant over time. We decomposed our data into seasonal, trend, random, and observed components. Seasonality is the result of regular and predictable fluctuations over time. For the seasonality, we noticed a consistent pattern of rising and falling, which is due to the fact that the temperature rises at sunrise and falls at sunset. There might also be fluctuations throughout the month.

We needed to check the stationarity of the dataset. One of the assumptions for stationarity is that if the p value is < 0.05, then our data is stationary; otherwise, it is non-stationary. Therefore, after running the ADF test to check for stationarity, our p value was less than the assumed value; therefore, our data is stationary and does not require any differencing. We then split our dataset into an 80% training set and a 20% test set for further analysis. We trained our data with the training set and predicted with the test set. After this process, we calculated our RMSE, as this is going to be our evaluation matrix. We also calculated the accuracy of the ARIMA model, which indicates that the training RMSE is close to the test RMSE, and this might be a good fit for prediction as predictions is close from original values

Before we began the development of the remaining models, we needed to create a new table of time steps. Rather than using the date and time columns, we would be using the time step column for our model development. The time steps column is the difference or interval in the time component, which represents time as a continuous numeric value. After column creation, we then divided our dataset into training and testing, as we did in ARIMA, with the same percentage.

The next model for development is linear regression. We fitted our training data with a training set and predicted with a test set, then output the summary of the prediction model. The output shows different parameters, such as residuals, which show the difference between the observed temperature value and the predicted value of the model, and R squared, which indicates how well the model fits the data. The higher the value, the more it fits, but in our case we have a very low value, which indicates that time has a very low impact on temperature variability. The F-statistics p value is also greater than 0.05, which shows that the relationship is not significant. After the summary had been carried out, we then calculated the RMSE value in order to compare it with other models.

In working with support vector machines, we also followed the same process as linear regression, but this time we tested three different kernel parameters for the study. The linear kernel, polynomial kernel, and radial kernel. The output of the summary indicates that the linear kernel utilised 185 support vectors, the radial 181, and the polynomial with 182 support vectors. We then computed the RMSE values for each of the SVR models. We observed that the model with the linear kernel has the lowest RMSE value, which implies that it provides the most accurate predictions of the three kernels.

In the development of random forest models, we also utilised three different "ntree" parameters (100, 200, and 500) for our model development. We trained each model with the training data and made predictions using our test data. We computed the RMSE value for each of the ntree parameters, and we observed that the ntree with 200 has the lowest RMSE value. It is slightly lower than that of 500. This indicates that the ntree with 200 provided a more accurate prediction compared with others.

IMPLICATIONS

As the study area is Blackpool Seaside Resort in the United Kingdom. Travellers will leverage the favourable temperature to plan their visit to this time of the year as it allows them to enjoy seaside and outdoor events. As we have mentioned earlier, the WTTO records the higher immigrants record during this time of year and the yearly statics by the National Bureau of Statistics in their report have shown that a large population of tourist visit the seaside during this time. Kamb (2021) in this research also discussed the social impact of climatic change with tourist business. He further reinstated that tourist the temperature is neither too hot nor too cold during this season. On the other hand, a decrease in temperature may lead to a faster heartbeat, striated muscles shivering, breathing becomes deeper, urine flow is increased and the elevation of biological oxidations which results in cows requiring more energy (Kamb , 2021). Based on this observation of the study area the minimum temperature observed is 9 degree Celsius while the maximum temperature recorded is about 18 degree Celsius. Also has we have observed through the time series plot that there are fluctuations in temperature across the day from sunset to sunrise. Using this metric we can conclude that Tourist Stakeholders in this sector needs to plan more for the negative effect of cold weather on the tourist seaside centres Also according to Pedersen, (n.d.), the thermal neutral zone which vary not only between species, but also with age. For example when the temperature drops to a very low point, tourist engage in hiking, mountain climbing while others will prefer indoor games keeping warm rather than being exposed to cold and been put to high risk of infection. Therefore, Tourist Stakeholders especially in seaside resort can use the result of predictions using our most accurate model, which is the random forest to prepare and make appropriate provision against the negative effect of cold weather on Tourists.

REFERENCES

Ajala, S., Muraleedharan Jalajamony, H., Nair, M., Marimuthu, P. & Fernandez, R.E. (2022) Comparing machine learning and deep learning regression frameworks for accurate prediction of dielectrophoretic force. *Scientific Reports*. 12 (1). doi:10.1038/s41598-022-16114-5.

Au, T.C. (2018) Random Forests, Decision Trees, and Categorical Predictors: The 'Absent Levels' Problem. *Journal of Machine Learning Research*.19. https://www.stat.berkeley.edu/.

Chai, T. & Draxler, R.R. (2014) Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*. 7 (3), 1247–1250. doi:10.5194/gmd-7-1247-2014.

Cohen, S. A., Higham, J., Gossling, ¨ S., Peeters, P. M., & Eijgelaar, E. (2016). Finding effective pathways to sustainable mobility: Bridging the science-policy gap. *Journal of Sustainable Tourism*, 24(3), 317–334. https://doi.org/10.1080/ 09669582.2015.1136637

Couronné, R., Probst, P. & Boulesteix, A.L. (2018) Random forest versus logistic regression: A large-scale benchmark experiment. *BMC Bioinformatics*. 19 (1). doi:10.1186/s12859-018-2264-5.

Demeter, C., Fechner, D., & Dolnicar, S. (2023). Progress in field experimentation for environmentally sustainable tourism*: A knowledge map and research agenda. Tourism Management,* 94, 104633.

Fawagreh, K., Gaber, M.M. & Elyan, E. (2014) Random forests: From early developments to recent advancements. *Systems Science and Control Engineering*. 2 (1), 602–609. doi:10.1080/21642583.2014.956265.

Gossling, ¨ S., & Higham, J. (2021). The low-carbon imperative: Destination management under urgent climate change. *Journal of Travel Research, 60, 1167–1179*. https://doi.org/10.1177/0047287520933679

Gossling, ¨ S., & Peeters, P. M. (2015). Assessing tourism's global environmental impact 1900–2050. *Journal of Sustainable Tourism, 23(5),* 639–659. https://doi.org/10.1080/09669582.2015.1008500

Gursoy, D., & Nunkoo, R. (Eds.). (2019). The Routledge Handbook of Tourism Impacts: Theoretical and Applied Perspectives. Abingdon: Routledge

Hall, C. M. (2015*). On the mobility of tourism mobilities. Current Issues in Tourism*, 18(1), 7–10.

Hendricks, E. A., Peng, M. S., Ge, X., & Li, T. (2011). Performance of a dynamic initialization scheme in the Coupled Ocean–Atmosphere Mesoscale Prediction System for tropical cyclones (COAMPSTC*). Weather and forecasting, 26(5)*, 650-663.

Hadinejad, A., Moyle, B. D., Scott, N., Kralj, A., & Nunkoo, R. (2019). Residents' attitudes to tourism: *A review. Tourism Review, 74(2*), 150–165.

Higgins-Desbiolles, F., & Bigby, B. C. (Eds.). (2023). The Local Turn in Tourism: *Empowering Communities. Bristol:* Multilingual Matters.

Lionetti, S., Pfäffli, D., Pouly, M., vor der Brück, T., & Wegelin, P. (2021). Tourism Forecast with Weather, Event, and Cross-industry Data. *In ICAART (2)* (pp. 1097-1104).

Jierula, A., Wang, S., Oh, T.M. & Wang, P. (2021) Study on accuracy metrics for evaluating the predictions of damage locations in deep piles using artificial neural networks with acoustic emission data. *Applied Sciences (Switzerland)*. 11 (5), 1–21. doi:10.3390/app11052314.

Kaliappan, J., Srinivasan, K., Mian Qaisar, S., Sundararajan, K., Chang, C.Y. & Suganthan, C. (2021) Performance Evaluation of Regression Models for the Prediction of the COVID-19 Reproduction Rate. *Frontiers in Public Health*. 9. doi:10.3389/fpubh.2021.729795.

Kamb, A., Lundberg, E., Larsson, J., & Nilsson, J. (2021). Potentials for reducing climate impact from tourism transport behavior. *Journal of Sustainable Tourism, 29, 1365–1382*. https://doi.org/10.1080/09669582.2020.1855436

Lionetti, S., Pfäffli, D., Pouly, M., vor der Brück, T., & Wegelin, P. (2021). Tourism Forecast with Weather, Event, and Cross-industry Data. *In ICAART (2)* (pp. 1097-1104).

Kumar, M., Kaur, M., & Sharma, N. (2023, August). A Novel ZeroShot Learning Approach for Multi-Classification of Orange Sooty Mold Disease Severity Levels. *In 2023 3rd Asian Conference on Innovation in Technology (ASIANCON) (pp. 1-5).* IEEE.

Kumari, K. & Yadav, S. (2018) Linear regression analysis study. *Journal of the Practice of Cardiovascular Sciences*. 4 (1), 33. doi:10.4103/jpcs.jpcs_8_18.

Peeters, P. M. (2017). Tourism's impact on climate change and its mitigation challenges. *How can tourism become 'climatically sustainable'?* Delft: Delft University of Technology.

Peeters, P. M., & Landr´e, M. (2012). The emerging global tourism geography – *an environmental sustainability perspective.* Sustainability, 4(1), 42–71. https://doi. org/10.3390/su4010042

Peeters, P., & Papp, B. (2023). Envisioning Tourism in 2030 and Beyond. *The changing shape of tourism in a decarbonising world. T. Foundation.*

Scott, D. (2021). *Sustainable tourism and the grand challenge of climate change*. Sustainability, 13(4), 1966.

Ramkissoon, H. (2023). Perceived social impacts of tourism and quality-of-life: *A new conceptual model. Journal of Sustainable Tourism*, 31(2), 442–459.

Rutty, M., Scott, D., Johnson, P., Jover, E., Pons, M., & Steiger, R. (2015). The geography of skier adaptation to adverse conditions in the Ontario ski market. *The Canadian Geographer/Le Géographe Canadien,* 59(4), 391-403.

Schneider, A., Hommel, G. & Blettner, M. (2010) Lineare regressionsanalyse - *Teil 14 der serie zur bewertung wissenschaftlicher publikationen. Deutsches Arzteblatt.1*07 (44) pp.776–782. doi:10.3238/arztebl.2010.0776.

Schonlau, M. & Zou, R.Y. (2020) The random forest algorithm for statistical learning. *Stata Journal*. 20 (1), 3–29. doi:10.1177/1536867X20909688.

Seyfi, S., Hall, C. M., & Saarinen, J. (2022). Rethinking sustainable substitution between domestic and international tourism: A policy thought experiment. *Journal of Policy Research in Tourism, Leisure and Events*. https://doi.org/ 10 .1080/ 19407963 .2022 .2100410.

Seyfi, S., Hall, C. M., & Shabani, B. (2023). COVID-19 and international travel restrictions: *The geopolitics of health and tourism*. Tourism Geographies, 25(1), 357–373.

Shah, R., Pawar, A., & Kumar, M. (2023, June). Enhancing Machine Learning Model Using Explainable AI*. In International Conference on Data & Information Sciences* (pp. 287-297). Singapore: Springer Nature Singapore.

Steurer, M., Hill, R.J. & Pfeifer, N. (2021) Metrics for evaluating the performance of machine learning based automated valuation models. *Journal of Property Research*. 38 (2), 99–129. doi:10.1080/09599916.2020.1858937.

United Nations World Tourism Organization (2020). UNWTO World Tourism Barometer (English version). https://www.e-unwto.org/loi/wtobarometereng