**A black and tan logo

Description automatically generated**

**TECHNOLOGICAL UNIVERSITY OF THE SHANNON: MIDLANDS MIDWEST**

**APPLIED RESEARCH PROJECT**

**Predicting How Weather Affects Customer Preference & Behaviour**

*Submitted by*,

**ÉRIN KENNEDY**

(A00275664)

Under the Guidance of,

**Dr Barry O’Loughlin**

**DEPARTMENT OF ACCOUNTING AND BUSINESS COMPUTING**

**MASTER OF SCIENCE IN DATA ANALYTICS**

**2023-2024**

# Signed Statement

*I have read the University’s code of practice on plagiarism. I hereby certify this material, which I now submit for assessment on the programme of study leading to the award of (Master of Science in Data Analytics ) is entirely my own work and has not been taken from the work of others, only to the extent that such work has been cited within the text of my work.*

*Student ID Number:* A00275664

*Name of Candidate:* Érin Kennedy

A handwritten text on a white paper

Description automatically generated*Signature of Candidate:*

*Date:* 25/Aug/2024

# Acknowledgements

I would like to express my gratitude towards my supervisor, Dr. Barry O’Loughlin, for providing helpful advice and guidance throughout this journey. I would also like to extend my thanks to my lecturers, especially Mr. Noel Tierney and Mr. Jonny O’Dwyer, who taught me throughout the year and helped prepare me for this paper.

I would like to thank the businesses that took part in this study and generously provided their data for examination. Their invaluable contribution contributed greatly to the study and I am grateful for the collaboration.

Thanks should also go towards my classmates and friends who assisted me through moral support and encouragement throughout the academic year and the study.

Lastly, I want to thank my family for their unwavering support, especially my parents and my younger siblings, who have been a source of inspiration for me and provided strength for me throughout this venture.

Contents

[Signed Statement ii](#_Toc175507065)

[Acknowledgements iii](#_Toc175507066)

[List of Illustrations vi](#_Toc175507067)

[Abstract vii](#_Toc175507068)

[1. Introduction 1](#_Toc175507069)

[1.1 Introduction Summary 4](#_Toc175507070)

[2. Literature Review 6](#_Toc175507071)

[2.1 Related Research 6](#_Toc175507072)

[2.2 Factors Affecting Restaurant Sales 7](#_Toc175507073)

[2.3 Weather's Influence on Customer Behaviour 12](#_Toc175507074)

[2.4 Literature Review Summary 19](#_Toc175507075)

[3. Research Methodology 21](#_Toc175507076)

[3.1 Introduction 21](#_Toc175507077)

[3.2 Business Understanding 22](#_Toc175507078)

[3.3 Data Understanding 22](#_Toc175507079)

[3.3.1 Data Collection Methods 22](#_Toc175507080)

[3.3.1.1 Weather Data 22](#_Toc175507081)

[3.3.1.2 Financial Data 25](#_Toc175507088)

[3.3.1.2.1 Sales Prediction 26](#_Toc175507089)

[3.3.1.2.2 Consumer Behaviour Trends 28](#_Toc175507090)

[3.3.1.3 Data Challenges 30](#_Toc175507099)

[3.4 Data Preparation 30](#_Toc175507100)

[3.4.1 Weather Data 30](#_Toc175507101)

[3.4.2 Financial Data 31](#_Toc175507102)

[3.5 Modelling Phase 31](#_Toc175507103)

[3.5.1 Sales Prediction Model 32](#_Toc175507104)

[3.5.1.1 Model Selection 32](#_Toc175507105)

[3.5.1.2 Model Development 34](#_Toc175507106)

[3.5.1.3 Model Creation 34](#_Toc175507107)

[3.5.2 Consumer Behaviour Analysis 34](#_Toc175507108)

[3.5.2.1 Dataset Summary Description 35](#_Toc175507109)

[3.5.3 Limitations of the Analysis 36](#_Toc175507110)

[4. Analysis of Findings 39](#_Toc175507114)

[4.1 Model Accuracy & Results 39](#_Toc175507115)

[4.1.1 Initial Observations 43](#_Toc175507116)

[4.1.2 Matrix: Correlation & Scatterplot 44](#_Toc175507117)

[4.1.3 Correlation Analysis of Product Categories 44](#_Toc175507118)

[4.1.3.1 Beverage 44](#_Toc175507119)

[4.1.3.2 Burger 45](#_Toc175507120)

[4.1.3.3 Chicken 45](#_Toc175507121)

[4.1.3.4 Combination Meals 45](#_Toc175507122)

[4.1.3.5 Fish 46](#_Toc175507123)

[4.1.3.6 Fries 46](#_Toc175507124)

[4.1.3.7 Ice Cream 46](#_Toc175507125)

[4.1.3.8 Kids Meal 46](#_Toc175507126)

[4.1.3.9 Milkshake 47](#_Toc175507127)

[4.1.3.10 Pizza 47](#_Toc175507128)

[4.1.3.11 Portions 47](#_Toc175507129)

[4.1.3.12 Other 47](#_Toc175507130)

[4.2.1 Correlation Analysis with Weather Variables 48](#_Toc175507131)

[4.2.1.1 Sun 48](#_Toc175507132)

[4.2.1.2 Rain 48](#_Toc175507133)

[4.2.1.3 Evaporation 48](#_Toc175507134)

[4.2.1.4 Windspeed 48](#_Toc175507135)

[4.2.1.5 Temperature 49](#_Toc175507136)

[4.2.1.5.1 Average Temperature 49](#_Toc175507137)

[4.2.1.5.2 Minimum Temperature 49](#_Toc175507138)

[4.2.1.5.3 Maximum Temperature 50](#_Toc175507139)

[4.2.1.6 Relative Humidity 50](#_Toc175507140)

[4.3 Scatterplot Matrices Analysis 50](#_Toc175507141)

[4.4 Linear Regression Findings 53](#_Toc175507142)

[4.5 Product – Category Summary 57](#_Toc175507143)

[4.6 Consumer Behaviour Graphs & Trends 57](#_Toc175507144)

[5. Discussion of Findings 63](#_Toc175507145)

[6. Conclusion, Recommendations & Reflection 67](#_Toc175507151)

[7. Bibliography 75](#_Toc175507152)

[8. Appendices 81](#_Toc175507153)

# List of Illustrations

Figures

[Figure 3‑1; The CRISP-DM Cycle 20](#_Toc175504272)

[Figure 3‑2; Correlation Analysis Results; Mullingar & Casement 2021 - 2024 23](#_Toc175504273)

[Figure 3‑3; Boxplo;: Distribution of Total Sales for Restaurant #1 & Restaurant #2 2021-2024 25](#_Toc175504274)

[Figure 3‑4; Paired T-Test; Final\_Total & Restaurant\_Id 25](#_Toc175504275)

[Figure 3‑5; Restaurant #1's Variables & Dimensions 26](#_Toc175504276)

[Figure 3‑6; Restaurant #2's Variables & Dimensions 26](#_Toc175504277)

[Figure 3‑7; Combined Restaurant Data & Dimensions 27](#_Toc175504278)

[Figure 3‑8; Product\_Data Variables & Dimensions 28](#_Toc175504279)

[Figure 3‑9; Summary Statistics for Product Categories (Product\_Data) 34](#_Toc175504280)

[Figure 3‑10; Time-Series Line Graph; Monthly Sales of Product Categories 35](#_Toc175504281)

[Figure 4‑1; Model Performance Metrics; MAE, RMSE & R-Squared 39](#_Toc175504282)

[Figure 4‑2; Comparison of Actual vs Predicted Sales for Random Forest Models 40](#_Toc175504283)

[Figure 4.1.1‑1; Time Series Line Plot; Monthly Sales For Product Categories (Product\_Data) 42](#_Toc175504284)

[Figure 4.1.2‑1; Correlation Matrix: Restaurant #1 Product\_Data 43](#_Toc175504285)

[Figure 4.2.1‑1; Example Scatterplot Matrix of Ice-Cream & Weather Variables 50](#_Toc175504286)

[Figure 4.2.1‑2; Example Scatterplot Matrix of Burger & Weather Variables 51](#_Toc175504287)

[Figure 4.2.1‑3; Final VIF Figures 52](#_Toc175504288)

[Figure 4.2.1‑4; Consumer Behaviour Boxplot Example Trend #1 57](#_Toc175504289)

[Figure 4.2.1‑5; Barchart of Both Restaurants’s Mean Sales on Weekends & Weekdays 58](#_Toc175504290)

[Figure 4.2.1‑6; Dual Axis Barchart Displaying Total & Mean Sales 2021-2024 59](#_Toc175504291)

[Figure 4.2.1‑7; Rainfall & Sales Trends by Restaurant 61](#_Toc175504292)

Tables

[Table 3‑1; Descriptive Statistics of Mullingar Weather Data and Casement Weather Data 23](#_Toc175504293)

[Table 3‑2; Summary Statistics for Item #1 and Item #2 Sales 28](#_Toc175504294)

[Table 3‑3; Comparison between Random Forest & MLR [8], [66] [67], [68] 32](#_Toc175504295)

[Table 4‑1; Model Performance Results: MAE, RMSE & R² 38](#_Toc175504296)

[Table 4‑2; Variable Importance Display of tunedclimatemodel 41](#_Toc175504297)

[Table 4.2.1‑1; Summary of Weather Variable Impacts on Product Category Sales 56](#_Toc175504298)

[Table 8‑8‑1; Daily\_Data Sheet 1 Data Dictionary [82] 80](#_Toc175504299)

[Table 8‑8‑2; Daily\_Data Sheet 2 Data Dictionary [66] 81](#_Toc175504300)

[Table 8‑8‑3; Product\_Category \_Sales Data Dictionary 82](#_Toc175504301)

[Table 8‑8‑4; Product Categories & Weather Correlation Matrix 83](#_Toc175504302)

# Abstract

This study aims to explore the impact of weather on customer behaviour and customer preferences, by utilising machine learning techniques and data analytics. It is well-documented that weather-related variables (temperature, rain, etc.) can affect a customer’s mood and influence their habitual buying patterns, however, research is minimal in countries where ‘negative’ weather, such as cold and rain is considered the norm [1], [2], [3]. A majority of influential weather studies focus on the retail industry and weather-neutral countries, which experience various weather conditions. This study aims to explore this niche area of research by analysing two branches of an Irish fast-food restaurant’s sales experience, hereafter referred to as Restaurant #1 and Restaurant #2, according to different weather variables. This study will utilise a wide-ranging dataset, including variables supplied by the business and alternative additional information gathered, to create a predictive learning model to calculate sales and analyse which variables impact each branch the most. Two separate analyses will be conducted, the first involving the creation of random forest models utilising sales figures, weather variables and additional variables created to provide further information from both branches. The second analysis will be a detailed exploration of the trends in Restaurant #1's product categories, taking into consideration the popularity of certain products, their peak times, and low times in relation to weather and external factors.

# Introduction

Employing customer prediction strategies by utilising weather-related variables, such as humidity, season and rainfall, has been accepted by businesses as a suitable strategy to predict trends within an industry or for a company’s customers since the 1980s, however, it has been discussed as an important factor in determining customer trends since the 1950s [4], [5]. Coca-Cola is a prime example of a large brand leveraging weather strategies to promote its products. By identifying patterns in consumer behaviour concerning weather conditions, Coca-Cola tailored its marketing efforts to align with these trends [6]. However, there remains a notable gap in research concerning the Irish market, particularly within the fast-food sector. Ireland's distinctive climate, characterised by its often cloudy and damp conditions, presents a unique environment for studying the impact of weather on consumer preferences and purchasing behaviour.

Ireland was selected as the study location due to the absence of existing research on this topic in the region, and its unique climatic conditions. The geographical convenience for the author and the willingness of local businesses to share data made Ireland an optimal setting for conducting this research. The fast-food industry was chosen specifically because of the limited research available on how weather influences sales and consumer behaviour in this sector, coupled with businesses' openness to provide access to their data.

Industries, such as the American finance and Chinese retail markets [7], have been studied extensively regarding customer buying patterns and trends however, little has been examined regarding the Irish market, specifically the Irish quick-service market. By leveraging advanced data analytics and machine learning techniques, this research analyses a range of datasets including weather patterns, consumer behaviour metrics, and sales information. This study aims to develop predictive models that provide insight into the probable shifts in consumer preferences and purchasing behaviour by analysing the complex interrelationships between weather variations (such as temperature and seasonal changes).

This study aims to assist businesses in this sector reduce costs based on weather-driven research by:

* optimising their marketing strategies
* increasing efficiency with staff rotation
* improving inventory management
* extending their knowledge of their consumer base’s preferences and behaviours

Additionally, this research will further increase knowledge in this industry. Demographics and cultures vary in their behaviours and attitudes from region to region. This study can help Irish businesses and other associated establishments understand customer buying patterns.

When comparing Ireland to markets with similar weather patterns, it's evident that the Irish population is accustomed to the traditionally cloudy and damp climate. Notably, a correlation emerges when comparing Ireland to colder nations like Sweden and Finland, where there is a significantly high per capita consumption of weather-influenced products such as ice cream. This observation suggests that marketing strategies, commonly effective in milder or warmer climates such as America or Spain, may not yield the same results. In colder climates, products like ice cream are viewed as year-round treats.

These cultural distinctions have the potential to challenge assumptions and established hypotheses regarding the relationship between weather conditions and consumer preferences. As highlighted in a UK study on the retail sector by Rose and Dolega, weather variables, particularly 'wind', have a substantial impact on British consumers, with their influence peaking during the summer and spring months [8]. The study quantifies the relationship between weather conditions and trading outcomes, emphasising the change weather can have on daily retail sales. The thesis topic "Predicting How Weather Affects Customer Behaviour and Preference" expands on the findings of Rose & Dolega by delving into predictive models that forecast customer spending behaviour based on weather conditions. While the authors focus on calculating the effect weather takes on retail sales, this thesis aims to predict customer behaviour and analyse preferences, offering a proactive approach for retailers to tailor their strategies based on anticipated weather effects. The thesis adds a predictive element to the insights provided by the source, enhancing the understanding of how weather influences consumer behaviour in retail settings. Like this study, Irish consumers may follow a similar pattern and, unlike milder climates, be unaffected by precipitation and sunlight, instead becoming influenced by strong wind gusts or hot, humid weather.

Due to these demographics and cultural phenomena, the research questions will be posited as follows;

1. Research Question 1- Is there a relationship between weather and quick-service restaurant sales in Ireland?
2. Research Question 2- Is there a relationship between weather and customer product preference in an Irish quick-service restaurant?

Based on these aims and research questions, the objectives for this study are as follows;

1. Analyse the relationship between weather variables and sales performance in the Irish quick-service restaurant industry.
2. Investigate the influence of weather conditions on consumer product preferences within Irish quick-service restaurants.
3. Develop predictive models to forecast sales trends based on historical weather and sales data.

By addressing these objectives, this research seeks to not only advance our understanding of the role of weather in shaping consumer behaviour and business performance but also provide actionable insights and recommendations for businesses operating in the Irish quick-service market.

* 1. Introduction Summary

This study aims to assist businesses in optimising their marketing strategies, increasing efficiency with staff rotation, improving business inventory management and extending their understanding of consumer preferences and behaviours. The primary research questions address whether there is a relationship between weather and quick-service restaurant sales in Ireland and whether weather conditions influence customer product preferences in these establishments. To achieve these aims, the study sets out the following objectives:

1. Analyse the relationship between weather variables and sales performance in the Irish quick-service restaurant industry.

2. Investigate the influence of weather conditions on consumer product preferences within Irish quick-service restaurants.

3. Develop predictive models based on historical weather and sales data to forecast sales trends.

Ireland was chosen as the study location due to the insufficient research performed concerning the thesis topic and its distinctive climate. Additionally, the convenience of the location for the author and the relative openness of the local business to share data made it an ideal setting for this research. The fast-food industry was selected, similarly, for the reason that there is only minor research on the impact of weather on fast-food sales and customer behaviour.

In the following chapter, *Literature Review,* previous studies and related papers will be discussed and their findings stated. This will be categorised into three groupings: *Related Research*, *Factors Affecting Restaurant Sales* and *Weather’s Influence on Customer Behaviour*. Essential elements of past research that have contributed to today’s knowledge, such as Yoo et al’s application of thermal comfort to consumer behaviour are referenced and detailed [9].

Following this, Chapter Three’s *Research Methodology* will be discussed. This chapter will comprise of an *Introduction*, followed by a brief introduction of phases of the CRoss Industry Standard Process for Data Mining (CRISP-DM) and how they will apply to the following headings. Each phase of CRISP-DM will be utilised to categorise sub-headings of the methodology process and the journey the data undertook. Challenges and limitations encountered will be considered as well as their possible impacts on the analysis results.

In Chapter Four, *Evaluation & Analysis*, the accuracy and outcomes of the models will be examined. This includes initial observations, detailed correlation and scatterplot matrices, and a thorough correlation analysis of product categories and weather variables. Specific sections will explore how various product categories (e.g., beverage, burger, chicken) correlate with sales, and how weather factors (e.g., sun, rain, temperature) influence customer behaviour. The chapter will also present linear regression findings, summarise the analysis of product categories, and highlight other trends through the form of graphs.

Chapter Five, *Discussions & Conclusion*, will interpret the key findings from the research, discuss their implications, and present the conclusions drawn based on the study’s objectives. Finally, in Chapter Six, *Future Work & Recommendations*, suggestions for future research and practical recommendations based on the study’s findings will be provided.

# Literature Review

This chapter informs and reviews the reader about existing research based on related topics to the thesis title, particularly concerning factors influencing restaurant sales and consumer behaviour in the context of specific weather conditions. The chapter is split into four separate sections, including a chapter summary;

* Related Research.
* Factors Affecting Restaurant Sales.
* Weather’s Influence on Customer Behaviour.
* Literature Review Summary.

Key insights and findings from the literature are identified and discussed concerning their connection to consumer behaviour and sales.

## Related Research

Due to a cultural and geographical element, Rose and Dolega’s research can be applied to an Irish perspective on shopping culture. As previously mentioned, the author identified location as a key contributor to retail performance, whilst also noting windspeed as the primary influencer of consumer traffic in British retail [8].

Weather affects consumer preferences, leading to increased demand for distinct categories of takeaway food products through online ordering [10]. Time factors, such as pre and post-holiday periods, can also significantly influence online delivery sales for fast-food restaurants. However, in the area where this study took place (China), it was found that rain, snow and air quality were the primary influencers on online food delivery sales, in contrast to windspeed, which was the main factor in the British study of retail shops [8]. This was thought to be due to uncomfortable or irritating weather leads people to remain indoors and prefer to make their purchases from home. Weekends were also found to be an extremely popular time to purchase deliveries [11]. The study successfully met its goals of examining weather's effect on delivery sales and individual demographic groups, however, it had some limitations. Although the study was performed in 2023, the data used was dated from the year 2016, from a then newly conceived business which poses the risk of concluding outdated results. Also, the authors note, that as 2016 was the beginning of the growth of online food delivery services, the characteristics of the delivery company may have influenced their results and suggested future research to compare multiple food delivery platforms.

## Factors Affecting Restaurant Sales

After a challenging period during the pandemic for the service industry, businesses are adjusting to the new market. Despite the impact of weather on sales, many other factors hold control over the success of sales in a business. Factors such as customer preferences, operational efficiency, and competitive landscape play crucial roles in determining the success of a business i.e. increased sales.

According to Gürsoy & Christina, the COVID-19 pandemic and associated restrictions, such as lockdowns and social distancing measures, led to the temporary closure of many hospitality businesses, impacting their demand and operations [12]. This emphasises how external economic conditions, in this case, the pandemic, can have a significant effect on the hospitality industry, which includes restaurants, affecting their sales and customer traffic. These unforeseen events, for example, ingredient shortages or power outages, can negatively affect the restauranteur despite the circumstance being beyond their control [13].

As per Koo’s research, location is identified as a significant factor affecting shop loyalty, with shop loyalty being directly influenced by location, merchandising, and after-sale service. This study emphasises the importance of location in the retail environment and its impact on consumer behaviour [14]. This study also found merchandising, after-sales services and shop atmosphere to have a significant impact towards a customer’s attitude towards the retail business, contrasting the findings of Yoo et al. Yoo et al.'s previous study concluded that store location alone significantly influenced a customer's attitude. However, Koo's research found that location had no significant impact on a customer's attitude towards the shop [15]. Although both studies were conducted in Korea, Yoo et al.'s study had a smaller sample size and was carried out in two large department stores in Ulsan. In contrast, Koo's study had a larger sample size and was conducted across six major discount shops in Daegu. It could be argued that Koo's study better represented Korea's retail customers.

Similarly, other related location factors can also affect customer footfall and sales outcomes. In a 2006 study in Britain conducted by Forman et al., the disclosure of reviewer location can significantly impact the geographic distribution of sales, emphasising the influence of geography on electronic commerce. The study suggests that shared geographical location enhances the relationship between reviewer disclosure and product sales, indicating that shop location can affect sales by influencing consumer behaviour based on reviewer identity and location information [16]. This finding accentuates the importance of considering shop location as a factor affecting sales in Britain but also in neighbouring Ireland. By extension, the insights gained from this research have implications for understanding consumer perceptions and purchasing decisions across both regions. A more recent study, similar to Forman et al’s study confirmed Forman et al’s hypothesis that identity cues and information within reviews are more relevant to a consumer when making decisions [17]. These studies highlight the importance of external factors on sales such as shop location and public identity to persuade consumers to purchase within.

While external conditions like the pandemic can drastically alter the operational landscape of a hospitality business, internal factors also play a role in determining sales success. A 2017 study highlighted the critical success factors in selling, such as learning orientation, customer orientation, intrinsic motivation, hard work, and technical expertise, and explored the mediating role of social intelligence in sales performance. This research stresses the importance of managing social interactions in influencing sales outcomes and its function between key selling success factors and sales performance [18]. However, it is important to note that this study specifically focuses on sales representatives within business-to-business firms. In the context of a fast-food business, cashiers and managers can be considered as sales representatives, as they engage in upselling during promotional periods and provide guidance on products. Sales promotions and effective marketing techniques play a crucial role in driving sales within the restaurant and fast-food industry [19], [20], [21]. By leveraging promotions such as discounts, coupons, limited-time offers, and loyalty programs, businesses can incentivise customers to make purchases and increase their frequency of visits. Moreover, strategic marketing campaigns can create awareness, generate interest, and cultivate brand loyalty among consumers. A 2013 study on the Egyptian fast-food industry revealed that customers who benefited from promotional campaigns exhibited higher loyalty to fast-food restaurants compared to those who did not benefit from promotions [21]. More recent research based in Indonesia came to the same conclusion. Discount and promotional programmes were found to have a direct effect on customer loyalty. Although the business analysed was not a fast-food outlet, it provided additional support for the 2013 study that customer satisfaction and loyalty can be influenced through promotional discounts and attractive offers. Targeted advertising on social media platforms or through email newsletters can reach specific demographics and encourage them to try new menu items or visit during off-peak hours.

Additionally, engaging in community events or partnering with local organisations can enhance brand visibility and build positive associations with the restaurant or fast-food establishment.

It’s worth noting that while effective marketing can drive foot traffic and increase brand awareness, pricing strategy is equally critical in converting this foot traffic into sales and maximising revenue potential. Pricing decisions can significantly influence consumer perceptions of value and affordability, ultimately impacting purchasing decisions [22] [23]. For instance, offering competitive pricing compared to competitors can attract price-sensitive customers and increase market share. Alternatively, implementing dynamic pricing strategies, such as surge pricing during peak hours or seasonal pricing for limited-time menu items, can capitalise on fluctuations in demand and maximise revenue potential. Moreover, bundling menu items or offering combo deals can encourage upselling and increase the average transaction value per customer [24]. However, in Liu et al’s paper, beyond price considerations, emphasis is put on business readiness for changes in supply and demand, and the efficiency of delivery systems, especially during turbulent times such as the pandemic. This contrasts with Beristain and Zorilla’s 2011 paper, which was written post-2008 economic crash and highlights the broader and long-term use of social media competence to enhance sales, without the pandemic focus [22], [23]. Whilst Liu et al. address immediate adaptations required for survival throughout the COVID-19 era, these insights can also apply to external, sudden market challenges that the business may face in the future. Meanwhile, Beristain and Zorilla’s findings provide valuable guidance for implementing broader, social-media-orientated strategies that are essential for long-term growth. A more recent study from 2021, similar to that of Beristain and Zorilla’s study, in regards to the rise of South Korean ‘grocerants’ (grocery restaurants), corresponds with the author's conclusion that store image, quality and brand equity generate perceived value amongst customers [25].

Internal shop stimuli, such as the layout, appearance and atmosphere of the restaurant, can also influence a customer’s purchasing behaviour, increasing impulse buying [26]. A survey of Indian fast-food outlets in Shiraz, India found that a customer’s likelihood of making a purchase and/or purchasing larger quantities of items increased once the restaurant created a pleasant experience, for example, environmental stimuli such as relaxed, slow background music and visual stimuli such as well-dressed, clean staff [27].

Customer satisfaction and experience within the restaurant and online (reviews, social media pages, advertisements, etc.) can be a lasting contributor to sales. A study by Maitlo et al. presents a model of customer experience and purchase intention in the online environment, highlighting the direct relationship between customer online experience and purchase intention. The findings suggest that customers with a positive online experience are more likely to make frequent purchases online, emphasising the importance of enhancing customer experience to drive purchase intentions [28]. This research highlights the importance of customer satisfaction driving sales, which may contribute to a restaurant’s food delivery or online collection sales. To encourage customer orientation within a business, managerial coaching is found to positively impact sales performance through results orientation [29].

Corporate Social Responsibility (CSR) is another increasingly key factor in the success of businesses, including those in the restaurant and fast-food industry. By engaging in CSR activities, such as sourcing ingredients sustainably, reducing waste, supporting local communities, and ensuring fair labour practices, businesses can enhance their reputation and build trust with consumers, [29]. A strong commitment to CSR can differentiate a brand in a competitive market, attracting socially conscious customers who prioritise ethical considerations in their purchasing decisions [32]. Additionally, CSR initiatives can foster customer loyalty and positive word-of-mouth, further boosting sales and brand equity [33] [34]. Implementing effective CSR strategies demonstrates a company's dedication to more than just profit, aligning business operations with broader societal values and expectations.

Another critical determinant of success in the restaurant and fast-food industry is operational efficiency. Streamlining operations can lead to faster service, reduced costs, and improved customer satisfaction. This includes optimising staffing levels, reducing waste, and utilising technology to enhance order accuracy and speed [35]. For instance, employing advanced point-of-sale systems, kitchen automation, and inventory management software can help manage resources more effectively and minimise delays [36]. Efficient operations not only enhance the dining experience but also allow businesses to serve more customers, thereby increasing sales and profitability. In an industry where margins can be tight, operational efficiency can provide a significant competitive advantage by ensuring that resources are used optimally, and customer expectations are consistently met.

## Weather's Influence on Customer Behaviour

Retailers have long been aware of the significant impact that weather can have on their sales and customer behaviour [8], [37], [38]. According to Linden’s 1962 research, unusual weather conditions can have a significant impact on consumer behaviour [37]. These conditions can cause alterations in the timing of purchases, lead to purchases that may not have otherwise occurred, or, in some cases, result in a permanent loss of demand. This highlights the importance of considering external factors such as weather patterns in forecasting consumer demand and making informed business decisions. Studies have found that weather factors like temperature, precipitation, and wind can influence consumer purchasing decisions, impacting the number of items purchased and the average price spent per order. The effect of weather on retail sales can be quite substantial, with some estimates suggesting a weather-related impact of up to 23% on daily sales depending on shop location and up to 40% based on product category [8], [38]. In consumer behaviour, weather plays a significant role in shaping purchasing decisions. For instance, on a hot summer day, many consumers are inclined to choose a refreshing cold drink over a hot beverage like tea. This preference is not only driven by physiological needs for cooling but also by cultural and habitual associations with seasonal comfort. These decisions are observable in everyday life, illustrating how weather directly impacts consumer choices [39].

Smart restaurant owners and marketing managers leverage these weather-driven preferences by implementing weather-responsive strategies in their advertising campaigns. By aligning their messaging and product offerings with current weather conditions, marketers can enhance relevance and appeal to consumers' immediate needs and desires. For example, during hot weather spells, advertisements may emphasise chilled beverages, outdoor dining options, or cooling products, thereby resonating more effectively with consumer preferences at that moment.

This approach not only improves the effectiveness of marketing efforts but also demonstrates an understanding and responsiveness to consumer behaviour dynamics influenced by weather. Such strategic adaptations can lead to increased consumer engagement, higher sales volumes, and strengthened brand affinity, illustrating the power of integrating weather insights into marketing strategies.

A comprehensive study on the quantitative impact of weather on UK retail sales found that weather effects are most pronounced in the summer and spring months, with wind consistently being the most influential weather variable. The authors also observed that sales in out-of-town stores exhibited a more complex relationship with weather compared to traditional high street locations, with the London and South-East regions experiencing the greatest levels of weather-related influence. In addition, the research explores how weather influences consumer mood. It suggests that more exposure to sunlight can lower negative effects and increase consumer spending. This finding has been corroborated by other well-known researchers like Lambert and Murray, whose individual studies will be discussed shortly. However, there may be a potential gap in the literature in terms of needing a more thorough analysis of how specific weather conditions impact different product categories and customer preferences beyond the UK context.

Since the 1970s, research has indicated that sunlight is an influencer of customer behaviour. It has been found that exposure to sunlight can boost a person’s mood by increasing the neurotransmitters dopamine and serotonin. [2], [1]. Howarth and Hoffman conducted a study investigating the relationship between mood and weather, analysing ten mood variables concerning eight weather variables [2]. The research highlighted the significant impact of humidity, temperature, and hours of sunshine on mood, with humidity being the most influential predictor. This study provides valuable insights into how weather conditions can affect human emotions and behaviour, laying a foundation for understanding the potential impact of weather on customer behaviour and preferences in the thesis topic. Another related study is Bruyneel et al’s research of why lottery ticket sales increase after sundown. It is understood that consumers purchase these to avoid depletion in mood or engage in mood repair due to bad weather. These attempts result in insufficient energy to resist the temptation to play the lottery [40]. Moon et al. emphasised that individual differences in weather sensitivity affect purchase intentions, with personal traits playing a crucial role in how consumers respond to weather changes. This implies that people’s unique characteristics and sensitivity to weather conditions influence their buying behaviour, affecting decisions on what, when, and how they purchase products. Additionally, it was illustrated how weather conditions, such as sunny or unfavourable days, can affect consumer spending by influencing mood states. These findings stress the importance of considering individual differences and emotional responses in understanding the effects of weather on consumer behaviour [41]. Furthermore, Liao drew attention to the role of weather updates, particularly negative ones, in altering customer valuations [42]. This perspective shifts the focus from solely examining the direct impact of weather on purchase decisions to the significance of real-time weather information in shaping consumer perceptions and choices. Understanding how weather updates influence consumer behaviour can offer valuable insights for businesses seeking to tailor their strategies based on weather forecasts.

The academic literature has explored the effect of weather on sales in-depth, with studies covering a range of retail settings from brick-and-mortar stores to convenience chains [7], [38]. A 2020 study by Badorf and Hoberg found that weather has a complex, non-linear effect on sales, with extreme weather conditions often leading to underestimation or overestimation of the true impact using traditional forecasting models, supporting the use of machine learning models such as Random Forests [38]. The ability to accurately predict how the weather will impact consumer behaviour and preferences can provide retailers with a significant strategic advantage. Some research has indicated that incorporating weather forecast data into sales forecasting models can improve accuracy up to seven days in advance. [38]. By understanding the nuanced relationships between weather and purchasing patterns, retailers can better optimise their inventory management, marketing, and operational decisions to adapt to changing climate conditions and maximise revenues [7], [43]. Weather conditions have also been studied in diverse contexts beyond consumer behaviour, such as in supply chain management. Appelqvist et al. investigated how weather variations impact demand and supply chain performance in the distribution of sporting goods, revealing the intricate relationship between weather fluctuations and business operations [44]. To effectively optimise its supply chain strategies in response to weather-related fluctuations, a business needs to first understand these dynamics. Taking into account weather variables such as temperature, precipitation, and air quality can provide retailers with valuable insights into consumer behaviour and preferences. This in turn enables them to make more informed and strategic decisions that drive sales and profitability. The study's insights can be used to develop predictive models for understanding how weather impacts customer choices and supply chain dynamics alternate in various industries.

Sandqvist & Siliverstovs investigated the aggregate impact of abnormal weather on consumer spending, highlighting the seasonal change channel as a key driver of purchasing behaviour [45]. By elucidating how weather effects manifest through seasonal transitions, this study contributes to a deeper understanding of the nuanced ways in which weather influences consumer expenditure patterns. Recognising these patterns can enable businesses to optimise their marketing efforts in alignment with seasonal weather variations. Moreover, Ma et al. explored the impact of weather on consumer reviews, revealing how weather conditions can influence consumers' evaluations of products and services [46]. By uncovering the interplay between weather and consumer perceptions, this study provides valuable insights into the holistic effects of weather on consumer decision-making processes. Understanding how weather influences consumer evaluations can assist businesses in effectively managing their reputations and customer relationships.

Climate change has led to various consequences, including a rise in abnormal weather patterns [47]. While Ireland is not susceptible to experiencing severe weather conditions, such as hurricanes, precipitation levels are expected to decrease in the near future, causing extended periods of heat [48]. It is anticipated that the frequency and severity of Atlantic storms in Ireland will also fluctuate, thereby necessitating the preparation of businesses for potential impacts [49]. A study focused on the UK retail market identifies this trend and ranks various retail sectors by their sensitivity to weather variability, modelling the impact of climate change on sales resilience [43].

Weather pattern correlations with consumer behaviours have been greatly researched within the finance and retail markets, yet there are few studies performed in the restaurant industry. Li et al. concluded that favourable weather conditions, such as sunny weather, positively influenced a consumer’s mindset, leading to increased impulse purchases, more expensive products, and larger quantities of items. In contrast, unpleasant weather such as rainfall, reflected negatively upon consumers’ mindsets which resulted in smaller purchases, a preference for online shopping or abandoning the shopping plan [50]. While Li et al.'s analysis primarily focused on mobile promotional advertisements, these findings suggest broader implications for consumer behaviour influenced by the weather. In the context of fast food restaurants, it can be inferred that during unfavourable weather, consumers may opt for takeaways or choose cheaper menu options. In contrast, during favourable weather conditions, consumers may be inclined to purchase larger quantities, including additional items such as desserts or premium meal choices. Li et al.’s conclusion supported Murray et al’s previous research in a North American retail shop [5]. Their study explores the effect of weather on consumer spending and the underlying psychological mechanisms. Contrasting this with the thesis topic, which aims to predict how weather influences customer behaviour, incorporating findings could enhance the thesis by providing insights into the psychological drivers of consumer spending in response to weather conditions. This could be a topic of discussion for future research.

The impact of weather on consumer preferences and decision-making extends to online and offline shopping behaviours, especially in the context of the Covid-19 pandemic. Nizma & Siregar analysed consumer preferences for shopping at online and offline stores during the pandemic, identifying external factors that influence these preferences and developing models to predict consumer choices between online and offline shopping [51]. This research underscores the evolving nature of consumer preferences in response to external factors like weather and pandemics, emphasising the need for businesses to adapt their strategies to changing environmental conditions. Further research by Bujisic et al. explored how specific weather factors drive the valence of consumer comments, emphasising the connection between weather, mood, and consumer behaviour. This study highlights the importance of understanding how weather influences the sentiment of consumer feedback, indicating a direct link between weather conditions and customer responses [52].

In their study of retail convenience shops in South Korea, Yoo Et al, found extreme weather temperatures to be the most influential variables for customer buying behaviour, providing support for Fanger’s hypothesis of thermal comfort in retail settings. Fanger’s study discusses the variations in clothing habits based on the outdoor climate in field comfort studies worldwide. As per Fanger’s research, people may seek comfort where their physiological and environmental needs are met, enabling them to escape their primary residence. The thermal comfort hypothesis postulates that the presence of a temperature-controlled environment, which induces a comfortable feeling, can attract and retain customers who are willing to expend their resources [53], [9]. Further factors posited that impact a customer’s purchase intentions are the ability to personalise general products. McDonald’s introduction of self-service kiosks showed a significant improvement in customer experience and customer satisfaction which can partly be attributed to improved order accuracy and reduced urgency to order on behalf of the customer [54], [55]. Customers appreciate the ability to personalise their order with ease, however, it is found that a preference for personalisation is negated when the product is intended for someone else [56]. Personalisation preference can be increased when the customer’s anticipation of regret is accounted for and businesses have an apt return or exchange policy in place. Additionally, Govers and Schoormans delved into how product personality influences consumer preference through a congruence effect, emphasising the psychological aspects of consumer decision-making [57], however in a more recent study, it was found that interest in the brand itself proved to be a more influential factor regarding purchase intention and brand attitude [58]. These studies shed light on the multifaceted role of product characteristics and brand reputation in shaping consumer preferences, indicating that aligning product personality with consumer expectations can enhance customer attraction and loyalty. Future research into this topic could include insights into how external factors, such as weather, interact with product characteristics to influence consumer behaviour. This would complement the findings on product personality congruence from Govers and Schoorman's research and could be considered a promising area for future research.

## Literature Review Summary

This chapter explores various factors influencing restaurant sales and customer behaviour, with a particular focus on the weather’s impact. Key points include:

Factors Affecting Restaurant Sales:

* External factors: Economic conditions, pandemics, and unforeseen events.
* Internal factors: Location, customer service, shop atmosphere, marketing strategies, operational efficiency, pricing strategies, CSR (impact on brand reputation), sales promotions, internal shop stimuli (such as layout and atmosphere), customer satisfaction, customer experience (both in-restaurant and online), return policies.

Weather's Influence on Customer Behaviour:

* Impact of temperature, precipitation, and wind on purchasing decisions.
* Sunlight exposure's effect on mood and spending patterns.
* Influence of weather conditions on consumer reviews and evaluations.
* Weather's role in shaping online vs. offline shopping preferences.
* Extreme weather temperatures' impact on customer buying behaviour.
* Relationship between weather, seasonal changes, and consumer expenditure patterns.
* Weather forecasting's importance for business strategy and sales prediction.
* Climate change implications and the need for businesses to prepare for weather variability.

The chapter emphasises the interplay between the various factors affecting restaurant sales, with a special focus on how weather conditions influence customer behaviour. It highlights the need for businesses to consider both internal operational factors and external elements like weather patterns in their strategic planning and decision-making processes. Additionally, it acknowledges that while other variables may contribute to high variance in data results, weather remains a significant contributing factor.

In the following chapter, the research methodology and layout of the data utilised will be discussed and explained. *Research Methodology* will be classed by a brief introduction, a discussion of the data collection methods and an examination of the data analysis process.

# Research Methodology

## Introduction

Diagram of a diagram of data

Description automatically generatedThis thesis will adopt the Cross Industry Standard Process for Data Mining (CRISP-DM) to maintain a clear and concise structure and approach. CRISP-DM consists of six phases that provide a structured approach for the remainder of this paper. It is highly applicable in data analytics projects due to the concept of agility within data science [59]. CRISP-DM can offer adequate support through its six phases, shown below in *Fig*. 3-1 [60].

Figure 3‑1; The CRISP-DM Cycle

The six phases consist of business understanding, data understanding, data preparation, modelling, evaluation and deployment. Each will be identified and discussed individually as to how they pertain to this topic. As illustrated in *Fig*. 3-1, transitioning between these phases requires the adoption of a flexible approach, as the outcome of each phase determines what has to be performed next [61]. This approach guarantees a step-by-step advancement through the necessary stages to accomplish the research goals.

## Business Understanding

The initial phase of CRISP-DM seeks to create a comprehensive evaluation of the subject. It includes setting objectives and goals and developing a project plan. Please refer to this paper’s initial *Introduction* to view objectives, aims and research questions set forth.

A successful criterion for this analysis may be defined as a model that accurately predicts sales data using historical weather data and a clear explanation of consumer purchase trends from Restaurant #1’s product categories dataset. This would allow businesses to make informed decisions and assumptions based on the predictions generated by the model.

## Data Understanding

The second stage of CRISP-DM involves the acquisition and appropriate handling of the data. Data exploration is conducted and familiarisation with the figures is essential. The data utilised in this study was collected from two fast-service restaurants, operating under the same franchise, and publicly accessible weather data. Machine learning techniques were employed using quantitative methodology, consistent with prior research. Given the constraints posed by the fast-service restaurant's limited reporting, data underwent adjustment and amalgamation to extract the requisite information effectively. Examination of data occurred at daily, weekly, and annual intervals, in alignment with the availability of reports for scrutiny.

## Data Collection Methods

### Weather Data

Local weather data was collected from Met Éireann’s online resources, offering detailed daily information on multiple weather variables from the nearest weather station of the restaurants; Mullingar. Variables that were provided included date, precipitation amount, mean windspeed and evaporation (refer to *Table 8-8-1*).

* + - 1. Data Challenges

Despite comprehensive data availability from the Mullingar weather station, there was a notable absence of sunlight exposure and duration data. To address this gap, sunlight data was sourced from the Casement Weather station in Dublin. A thorough assessment was conducted to ensure the suitability of this substitute data for analysis.

In this regard, reference was made to the Köppen Climate Classification [15], which examines significant differences in climate patterns across land areas. This analysis revealed no substantial variability in climate between the Midlands and Dublin suggesting the climates to be sufficiently similar, supporting the potential applicability of Casement’s sunlight data to the Mullingar context.

As an additional measure, long-term averages, medians, minimums and maximums of primary weather variables from each station were compared, and a correlation analysis was conducted for variables relevant to cloud cover and sunlight duration (rain and windspeed). There were no major differences between the primary variables such as min temp and max temp. However, a noticeable distinction was observed in the rain and wind speed variables, particularly in the differences between their maximum and median figures.

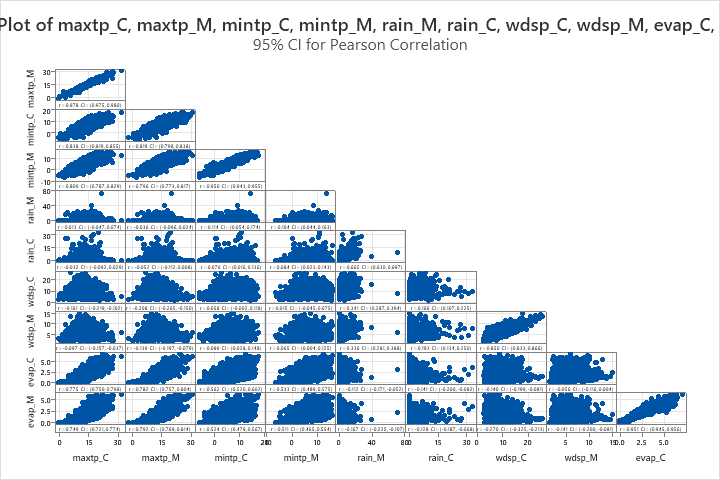
Table 3‑1; Descriptive Statistics of Mullingar Weather Data and Casement Weather Data

A comparison of statistics and statistics

Description automatically generated

The correlation analysis revealed, as visually seen in the descriptive statistics above, that the temperatures have a high correlation (.978 and .95). Evaporation also recorded a high correlation (.951). Regarding windspeed, the correlation analysis yielded compelling evidence that both stations recorded similar windspeeds and supported similar climates (.85). The least correlated was rainfall (.665) which, whilst suggesting a moderately strong correlation, suggests a substantial difference in rainfall and an indirect effect on sunshine duration due to cloud cover.

Figure 3‑2; Correlation Analysis Results; Mullingar & Casement 2021 - 2024



**Correlations**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **maxtp\_C** | **maxtp\_M** | **mintp\_C** | **mintp\_M** | **rain\_M** |  | **rain\_C** | **wdsp\_C** | **wdsp\_M** | **evap\_C** |
| maxtp\_M | 0.978 |  |  |  |  |  |  |  |  |  |
| mintp\_C | 0.838 | 0.819 |  |  |  |  |  |  |  |  |
| mintp\_M | 0.809 | 0.796 | 0.950 |  |  |  |  |  |  |  |
| rain\_M | 0.013 | -0.036 | 0.114 | 0.104 |  |  |  |  |  |  |
| rain\_C | -0.032 | -0.052 | 0.076 | 0.084 | 0.665 |  |  |  |  |  |
| wdsp\_C | -0.161 | -0.208 | 0.058 | 0.015 | 0.341 |  | 0.166 |  |  |  |
| wdsp\_M | -0.097 | -0.139 | 0.089 | 0.065 | 0.336 |  | 0.193 | 0.850 |  |  |
| evap\_C | 0.775 | 0.782 | 0.562 | 0.533 | -0.112 |  | -0.141 | -0.140 | -0.056 |  |
| evap\_M | 0.749 | 0.792 | 0.524 | 0.511 | -0.167 |  | -0.128 | -0.270 | -0.141 | 0.951 |

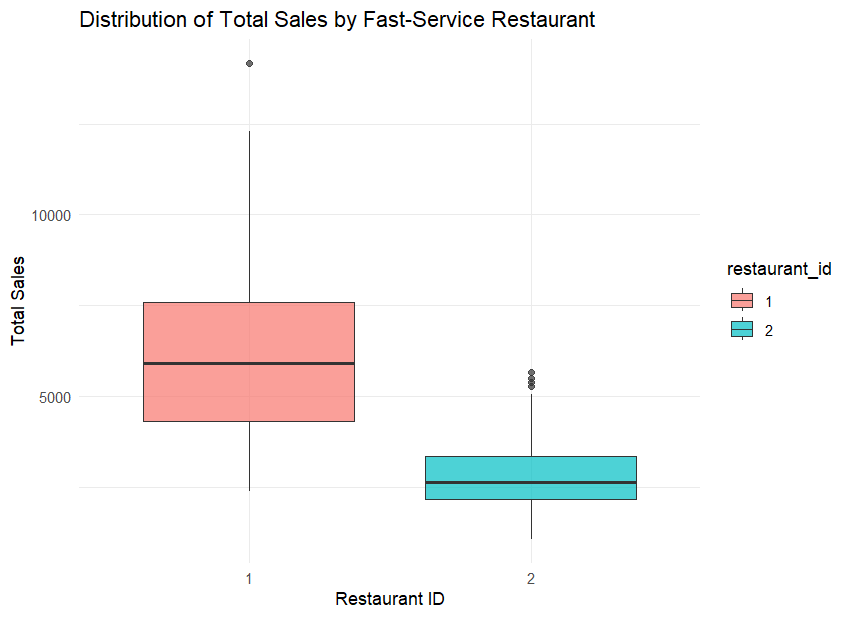
Considering these results, in this researcher’s opinion, it was thought appropriate to include Casement’s data due to the high correlation with Mullingar’s figures, the informative nature of Casement’s data and the proximity of the stations. While the strategic incorporation of Casement’s data enhances the completeness and accuracy of the weather analysis for the Mullingar region, it is also essential to acknowledge the potential limitations stemming from the differences observed between the two weather stations. Despite efforts to ensure the suitability of Casement’s data through evaluation and reference to climate classifications, notable disparities were identified, particularly in the rainfall variable. These differences may introduce uncertainties in the analysis, potentially impacting the reliability and reproducibility of the findings. Therefore, it is necessary to interpret the results with caution.



### Financial Data

The financial data utilised in this analysis is sourced from two Midlands-based branches of a fast-service delivery and restaurant franchise, ranging from May 11th 2021 to May 31st 2024. These dates remained consistent for both restaurants to prevent a skew in the data. The reports provided by the businesses included ‘Sales by Type’, ‘Sales by Weather’, ‘Daily Sales’, and ‘Weekly Product Sales’ (Restaurant #1 only). It is important to note Restaurant #2 did not support delivery and is significantly smaller than the other due to its less central location, as seen in both branches' distribution of sales below distribution.

Figure 3‑3; Boxplo;: Distribution of Total Sales for Restaurant #1 & Restaurant #2 2021-2024



The first restaurant, being larger and easily accessible, has a much wider range of daily sales. It consistently brings in between five and ten thousand euros, even on slower sales days, making the maximum amount Restaurant #2 has ever achieved in the available data. The second restaurant consistently has lower daily sales than the first, but its sales remain relatively stable.

A paired T-Test was performed to compare the *Final\_Total’s* of the two fast-food restaurants to determine if they are significantly different.

Figure 3‑4; Paired T-Test; Final\_Total & Restaurant\_Id

A screenshot of a test

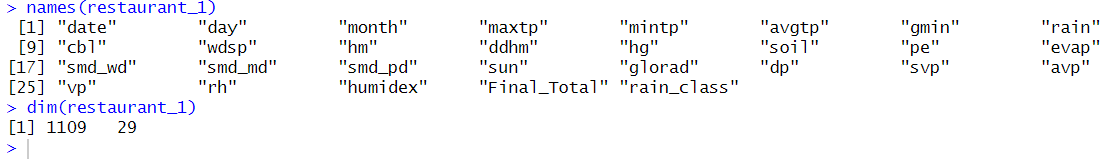
Description automatically generated

The paired T-Test revealed the means of each branch to be significantly different (p-value < .05) from the other. The null hypothesis is rejected. Restaurant #1 has considerably higher sales with a mean of 6023, whilst Restaurant #2 has a substantially lower mean of 2785. It is reasonable to consider that different weather variables may affect each restaurant differently.

### Sales Prediction

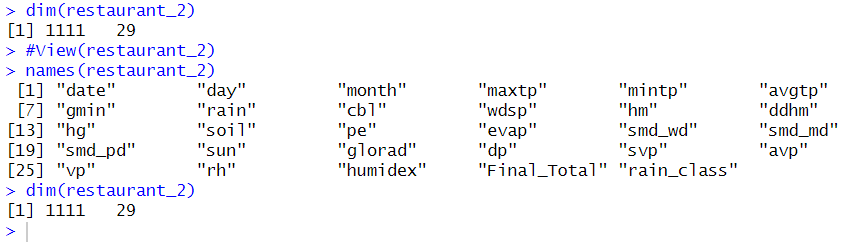
To allow for a comparison analysis of both restaurants each restaurant was allocated their unique ID within a combined dataset, however, they were also analysed individually. Restaurant #1’s dataset contained 1109 rows and 29 columns.

Figure 3‑5; Restaurant #1's Variables & Dimensions



Restaurant #2’s dataset contained 1111 rows and 29 columns. The data was thoroughly checked to explain the discrepancy between both branches. It was discovered that the difference in the number of rows is because Restaurant #2 is open for more days than Restaurant #1, although Restaurant #2 does have a shorter working week.

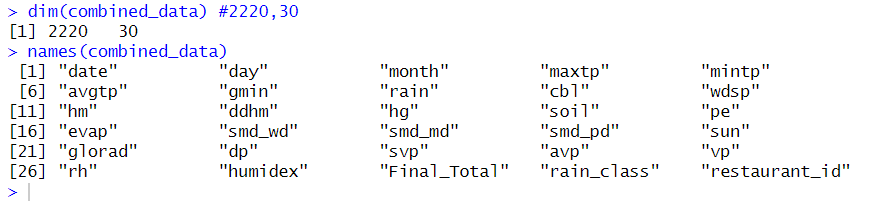
Figure 3‑6; Restaurant #2's Variables & Dimensions



Many of the columns in both datasets are not immediately necessary but are used to calculate other contributory variables, such as relative humidity. These columns are retained to potentially uncover variables that may unexpectedly and indirectly impact sales. For instance, variables like rain\_class and relative humidity (*rh*) are kept for further analysis, even though they are primarily supporting other variables (see *appendix*).

A comparison analysis was conducted using both separate data sheets and a combined dataset. *Final\_Total* indicates the final in-house figures of the day for both restaurants, also representing the dependent variables for analysis.

Figure 3‑7; Combined Restaurant Data & Dimensions

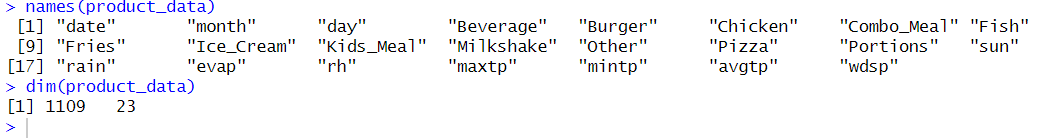


By combining figures from both restaurants, there is a total of 2230 rows. An additional numeric column called *restaurant\_id* has been created. This column is designed to accurately identify individual restaurants and ensure the integrity of the data. The *restaurant\_id* column will prevent the model from merging sales data from different restaurants as a single entity.

### Consumer Behaviour Trends

To begin answering the second research question, a separate dataset ‘product\_data’ was used to find the most popular or trending product category, and compare the effect of the weather to these trends. Due to time constraints, only Restaurant #1 was included in this analysis. This dataset contained 1109 rows and 23 columns, similar to Restaurant #1’s daily sales dataset dimensions. Additionally, as part of confidentiality measures to anonymise the business, certain products were renamed or described more broadly in terms of their food type. This ensured the protection of sensitive business information while allowing for meaningful analysis of sales data. Key variables included were Beverage, Pizza, Combo Meal, date, six primary weather variables (*sun, evap, rain, mintp, maxtp, avgtp*) and seasonal stand-in variables, such as month and day.

Figure 3‑8; Product\_Data Variables & Dimensions



The product categories largely remained similar to how the business phrased them, with some changes made to product headings to maintain anonymity. However, throughout the available data, the product category names did not remain continuous, largely due to two specific product categories. An initial visual and descriptive analysis suggests that this discontinuity may be due to poor sales in both categories. After consulting with the business's staff, it was confirmed that the product and all its accompanying sides are no longer available in the restaurant, explaining the disappearance of the data. This issue is resolved by completely removing both categories. Neither category made any sales after 14/06/2022, and they both performed poorly in terms of sales, as indicated by their summarised statistics in their individual sales historical data.

Table 3‑2; Summary Statistics for Item #1 and Item #2 Sales

|  |  |  |
| --- | --- | --- |
| Summary Statistics for Item #1 and Item #2 Daily Sales | | |
| **Statistic** | **Item #1** | **Item #2** |
| Min | 0.000000 | 0.000000 |
| 1st Qu. | 1.560000 | 0.000000 |
| Median | 3.120000 | 0.000000 |
| Mean | 3.710831 | 3.455021 |
| 3rd Qu. | 5.500000 | 5.050000 |
| Max | 15.600000 | 25.230000 |









### Data Challenges

Working with the financial data for this analysis presents several significant challenges. A notable hurdle continues to be the company's software, which does not support exporting data directly to Excel. This limitation resulted in the use of manual methods, such as printing to PDF and ultimately scanning into Excel, followed by extensive manual formatting. This process is not only time-consuming but also prone to errors, which may affect the accuracy and reliability of the analysis. Adding to the complexity, the business underwent a system change in 2021. Unfortunately, this transition meant access is lost to records predating May 2021. As a result, crucial historical data that would provide valuable insights for comparison and trend analysis over time is not included. The second restaurant had a larger amount of data predating 2021. However, to avoid skewing the data and creating an inaccurate model, only data comparable to Restaurant #1 was selected.

Maintaining confidentiality posed another challenge. To protect sensitive business information, anonymisation is employed by renaming or broadly categorising certain products. This anonymisation process, while necessary, poses difficulties in maintaining the level of detail required for a detailed analysis of sales data.

## Data Preparation

Within the data preparation stage, data is cleaned and formatted. Additional variables are added to the selected datasets that bring new information or knowledge to the results.

## Weather Data

Additional variables, such as average temperature, acquired from public online resources were added [62]. Actual vapour pressure, saturated vapour pressure, vapour pressure and relative humidity were appended to the dataset to allow calculations for further variables. The Antoine formula was utilised to calculate vapour pressure, where the Antoine constants for water were employed [63], [64].

Where A = 8.1122, B = 1592.864, C = 226.184 and T = average temperature (C°).

Inspired by Yao et al’s use of a human comfort index, a humidex (humidity index) was included as a predictor variable [11]. The humidex measures how uncomfortable the weather feels to a human being. It employs relative humidity, dewpoint and temperature figures to determine this. These values are obtained from Weather Underground [62]. The formula was acquired from ChemEurope [65].

Where rh represents relative humidity and e represents vapour pressure.

## Financial Data

Much of the financial data preparation involved acquiring and formatting data because Restaurant #1’s software primarily connects and shares reports with other software within the company. As a result, the options are limited to printing the report or manually inputting figures. Restaurant #2, despite being considerably smaller, has a more adaptable system that is faster. Despite the improved speed, there was still not enough time to acquire product reports and allow for formatting. After this process was completed, data analysis could begin. When analysing the comparison datasets, the dates the business was closed for bank holidays were excluded from the random forest’s training and test sets to maintain accuracy.

## Modelling Phase

In the model development phase of CRISP-DM, a particular modelling technique is selected to construct, train, test, and evaluate accuracy using metrics. In line with the objectives of this thesis, the model aims to accurately forecast the sales of a fast-food restaurant using weather variables. This can be considered a suitable evaluation metric for the models.

## Sales Prediction Model

## Model Selection

When selecting a model to analyse the restaurant’s daily sales dataset, two techniques were considered due to the quantitative nature of the data: multiple linear regression (MLR) and Random Forest. Ultimately, the Random Forest technique was chosen due to several key advantages. Firstly, Random Forests are particularly well-suited for handling both categorical and continuous variables, which aligns well with the structure of the dataset which includes a mix of categorical factors (such as day of the week and rainfall category) and continuous variables (such as daily sales figures and various weather conditions). This capability allows the model to integrate and utilise diverse types of data effectively, whereas MLR often requires additional preprocessing and encoding of categorical variables, potentially complicating the model.

Table 3‑3; Comparison between Random Forest & MLR [8], [66] [67], [68]

|  |  |  |
| --- | --- | --- |
| Criteria | Multiple Linear Regression (MLR) | Random Forests |
| Interpretability | Simple and highly interpretable | Complex; harder to interpret but offers feature importance insights |
| Computational Efficiency | Fast and efficient, especially for smaller datasets | Computationally intensive, especially with large datasets |
| Model Assumptions | Requires strict assumptions (e.g., linearity, normality) | No strict assumptions, flexible with data |
| Handling of Relationships | Best for linear relationships | Excels at modelling non-linear relationships |
| Sensitivity to Outliers | Highly sensitive | Robust to outliers due to averaging across trees |
| Handling of High-Dimensional Data | Prone to overfitting with too many predictors | Reduced risk of overfitting (ensemble averaging) |
| Feature Importance | Provides clear statistical significance for each predictor | Identifies key features, but lacks transparency in how decisions are made |
| Scalability | Less suitable for large datasets or complex interactions | Scales well with large datasets and multiple features |
| Handling of Missing Data | Can struggle with missing data | Handles missing data effectively through imputation in trees |

Additionally, familiarity with Random Forests and their interpretations influenced this decision. Previous research, in similar contexts, has demonstrated the effectiveness of Random Forest models, capturing patterns and relationships that MLR might miss, further supporting this decision [8], [69]. For instance, Rose & Dolega’s study highlighted the superior ability of Random Forests to account for variability within extensive datasets and their capacity to handle both categorical and continuous variables effectively [8].

Moreover, random forest techniques have proved to be extremely accurate in comparison to alternative supervised learning techniques, such as linear regression or K-Nearest Neighbour, reaching a mean accuracy rate of 96.06% in a 2021 study [70].

## Model Development

The models were built in RStudio using relevant packages, and training and test sets were developed for both datasets. For each model, the training sets and test sets were split 80:20.

Several models are created to determine the ideal mtry values and the most significant predictors. For the combined restaurant datset, four main models are developed, along with experimental models based on observations from the important factors in the first four models. The typical structure followed:

* + 1. allmodel; utilised all the variables available.
    2. tuned\_all\_model; utilised all variables available, with optional mtry.
    3. climate\_model; utilised climate and spatial variables
    4. tuned\_climate\_model; utilised climate and spatial variables.

Due to consistently low importance and bottom ranking in variable importance in both Restaurant #1 and Restaurant #2’s initial models, humidex\_class and rain\_category were removed from both datasets to facilitate a smoother analysis.

## Model Creation

Four random forest models were developed to identify patterns or trends and create a working predictive model. Optimal mtry was utilised to fine-tune the model and experimentation with ntree was conducted.

## Consumer Behaviour Analysis

The product dataset was analysed by creating visual graphs of the products within the dataset, summary descriptions, and a correlation matrix. A correlation and scatterplot matrix were created to find any underlying relationships between weather variables. Additionally, linear regression analysis was conducted to identify the most influential weather variables regarding individual product categories, as seen in Badorf & Hoburg’s investigation into weather on retail sales [38].

## Dataset Summary Description

The product\_data dataset comprises 1109 rows and 23 columns from Restaurant #1, with 12 columns dedicated to product categories. A descriptive analysis was performed to identify the top-performing product categories and their popularity. The descriptive metrics employed include the mean, minimum, maximum, range, and interquartile range.

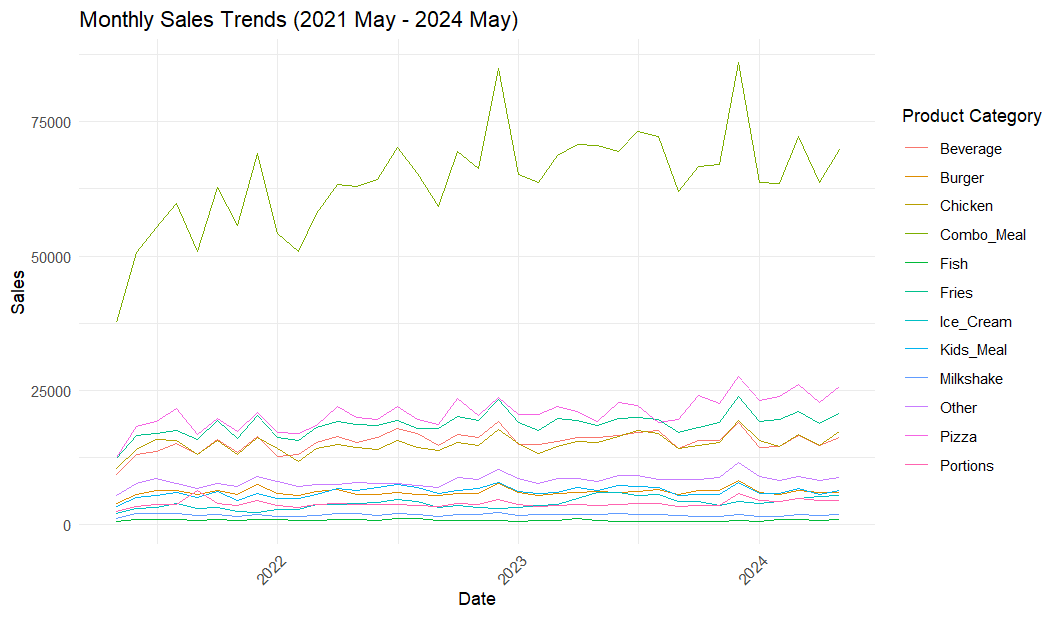
humidexFigure 3‑9; Summary Statistics for Product Categories (Product\_Data)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Mean** | **Minimum** | **Q1** | **Median** | **Q3** | **Maximum** | **IQR** |
| Beverage | 513.74 | 196.28 | 385.53 | 511.15 | 638.00 | 1222.13 | 252.47 |
| Burger | 200.14 | 59.27 | 147.84 | 192.61 | 247.03 | 558.35 | 99.19 |
| Chicken | 500.68 | 146.65 | 354.59 | 487.71 | 635.07 | 1237.39 | 280.48 |
| Combo\_Meal | 2145.6 | 789.5 | 1575.7 | 2097.5 | 2677.1 | 5054.2 | 1101.4 |
| Fish | 26.025 | 0.000 | 8.770 | 22.390 | 35.550 | 308.210 | 26.780 |
| Fries | 621.30 | 240.28 | 449.61 | 601.33 | 775.93 | 1461.38 | 326.32 |
| Ice\_Cream | 129.94 | 18.01 | 90.36 | 121.22 | 159.56 | 374.80 | 69.20 |
| Kids\_Meal | 202.10 | 29.04 | 130.55 | 194.31 | 263.47 | 593.39 | 132.92 |
| Milkshake | 58.996 | 5.390 | 38.380 | 56.610 | 78.115 | 197.650 | 39.735 |
| Other | 272.52 | 0.00 | 203.37 | 264.29 | 335.64 | 713.47 | 132.27 |
| Pizza | 694.38 | 0.00 | 442.62 | 647.32 | 926.23 | 1612.37 | 483.61 |
| Portions | 130.13 | 0.00 | 91.19 | 126.15 | 158.86 | 844.27 | 67.67 |

The dataset reveals notable variations in daily sales across different product categories. *Combo\_Meal* stands out with the highest mean sales of 2145.6 and a maximum of 5054.2, indicating consistently high demand compared to other items. Similarly, *Pizza* and *Fries* also display high average sales, with means of 694.38 and 621.30, respectively, and substantial maximum values, reflecting their strong popularity. In contrast, *Milkshake* and *Fish* exhibit lower average sales, with means of 58.996 and 26.025, respectively, suggesting less frequent demand. *Ice\_Cream* and *Kids\_Meal* fall in-between, with moderate means and noticeable variability. The ranges for *Combo\_Meal* and *Pizza* highlight occasional spikes in sales, while the lower categories like *Milkshake* and *Fish* show a narrower range, indicating more consistent but lower sales.

In *figure 3-10*, a time series scatterplot illustrates the monthly sales, revealing a notable difference in popularity between the *Combo\_Meals* category and the *Fish* or *Ice\_Cream* category. However, the plot becomes unclear when four years of data are grouped monthly. Nevertheless, it is evident that consumers strongly prefer combination meals, which could be influenced by multiple factors such as pricing and ease of ordering.

Figure 3‑10; Time-Series Line Graph; Monthly Sales of Product Categories



A steadily increasing non-linear trend can be seen from May 2021, due to the negative impacts of Covid-19, the business suffering losses from unexpected closures and adjustments. Dramatic peaks can be noticed across all product categories towards the end of each calendar year, although in regards to the lesser popular products, these jumps are only slight.

## Limitations of the Analysis

Visualisations, correlation matrices, and linear regression play a crucial role in data analysis but come with inherent limitations that can impact the accuracy and interpretability of results. Visualisations and descriptive graphics can help to illustrate a clear understanding of numerical data in a visual format, making data easier to understand, and allowing for better analysis. However, graphics can sometimes oversimplify trends and may not effectively highlight outliers or differences in the data. Additionally, there is potential for graphics to be misinterpreted, leading to incorrect conclusions by researchers or readers, which can confuse [71].

An important limitation to mention regarding correlation is that "Correlation is Not Causation" [72]. A high or low correlation between variables may indicate a relationship may indicate a relationship, however, this could be coincidental. Despite this, a correlation matrix is a useful tool for exploring relationships between variables in a dataset, such as between product category sales and weather variables. Its main advantage is its ability to quickly reveal the strength and direction of linear relationships, helping identify key variables and detect potential multicollinearity issues. However, the correlation matrix has notable limitations:

* It cannot capture non-linear relationships.
* Lack of temporal context.
* Does not establish causation.

Additionally, in datasets with many variables, the matrix may become complex and difficult to interpret, potentially obscuring important patterns. To address these concerns, a linear regression analysis and time series analysis will be conducted alongside the correlation matrix for a more comprehensive understanding of the data [73].

Linear regression, while a widely used analytical method, presents several limitations that can impact its effectiveness when applied to datasets involving weather conditions and sales. One major limitation is the assumption of linearity, which presumes a straight-line relationship between predictors (such as weather variables) and the outcome variable (product categories). This assumption may fail to capture complex, non-linear interactions, leading to potentially misleading conclusions. Additionally, linear regression is sensitive to outliers—extreme values in sales or unusual weather conditions can disproportionately influence the model’s results, distorting the accuracy of predictions [74]. The model also assumes homoscedasticity, meaning that residuals should have constant variance across all levels of the predictors. If variability in sales changes with extreme weather conditions, this assumption might be violated, resulting in less reliable estimates. Furthermore, linear regression presumes that residuals are independent of each other. If past weather patterns or sales trends affect current observations, autocorrelation might occur, which can undermine the model's validity. Multicollinearity, where independent variables are highly correlated, can also be problematic, as it complicates the isolation of each predictor's effect. The assumption of normally distributed residuals is another limitation; skewed or heavy-tailed data can affect the accuracy of statistical tests [75].



# Analysis of Findings

In the fifthstage of CRISP-DM, which is *Evaluation*, the process involves using the created models, assessing the results they achieve, and conducting a final review. Typically, a comprehensive deployment plan is created to outline how the model will be integrated into a business model, however, in this circumstance, the *Deployment* stage will not take place.

## Model Accuracy & Results

The four models exhibited high variability and achieved acceptable MAE and RMSE results.

Table 4‑1; Model Performance Results: MAE, RMSE & R²

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Mean Absolute Error (MAE) | Root Mean Squared Error (RMSE) | R-Squared (R²) |
| All\_Model | 624.1952 | 966.8824 | 83.96 |
| Tuned\_All\_Model | 536.8051 | 908.8939 | 87.25 |
| Climate\_Model | 616.5841 | 956.8538 | 84.52 |
| Tuned\_Climate\_Model | 531.048 | 904.6512 | 87.42 |

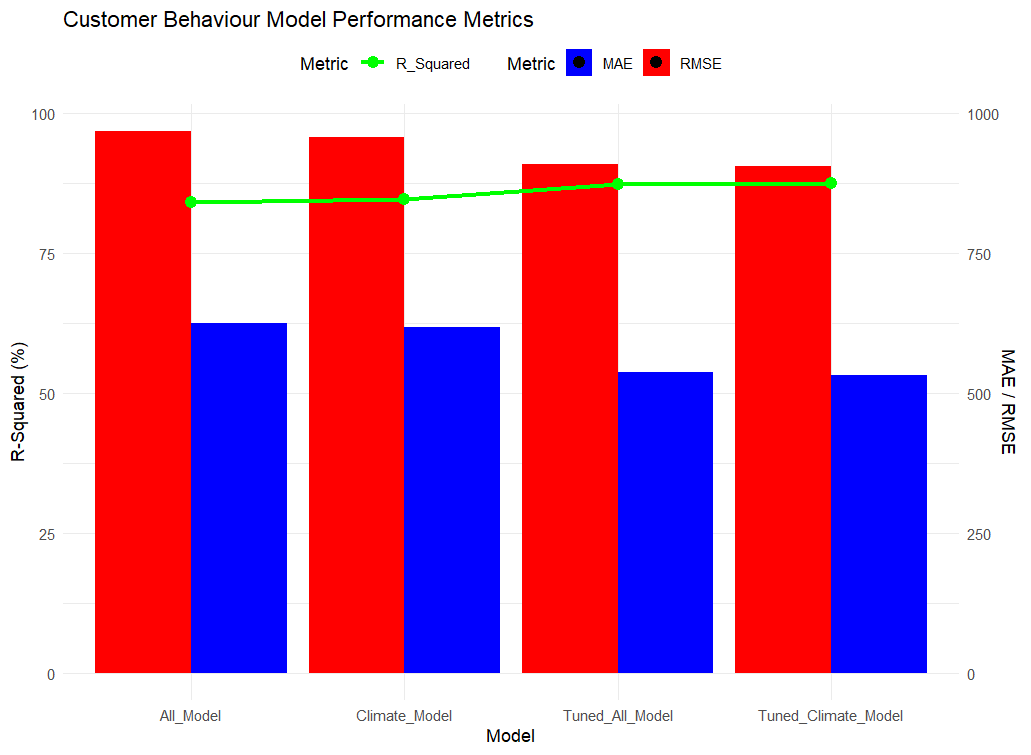
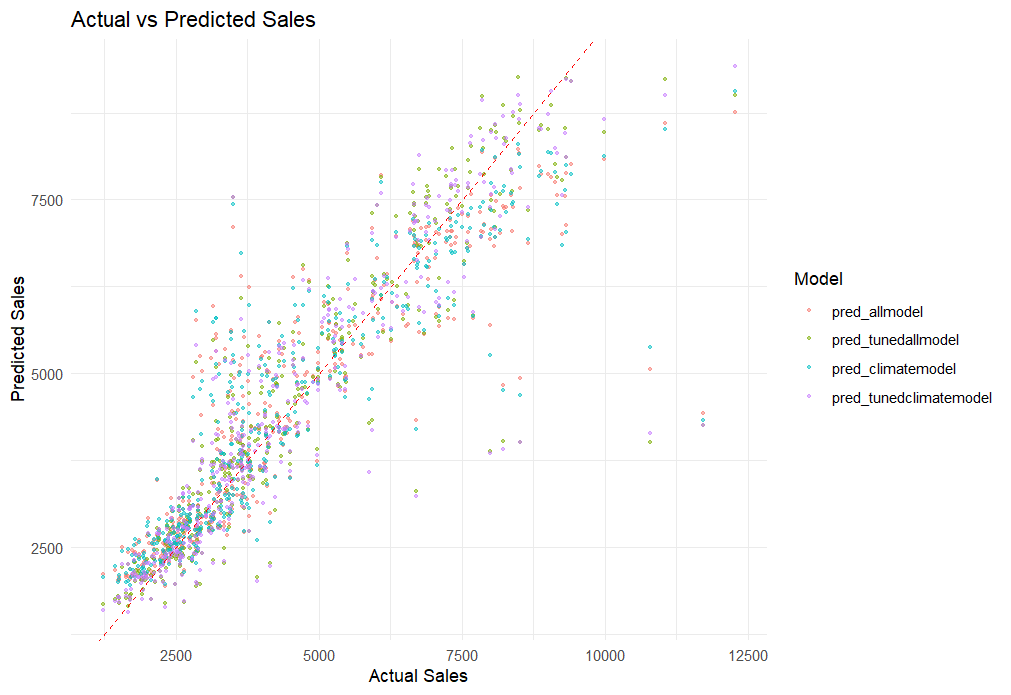
The findings suggest that irrespective of the selection of predictor variables employed, fine-tuning the Random Forest model generally enhances its performance. With the best overall performance, the *Tuned\_Climate\_Model* indicates that Irish fast-food sales predictions can be made with high accuracy using a carefully chosen and fine-tuned subset of weather-related predictors. The R² values for each model demonstrate that the chosen predictors can account for a sizable percentage of the variability in sales.

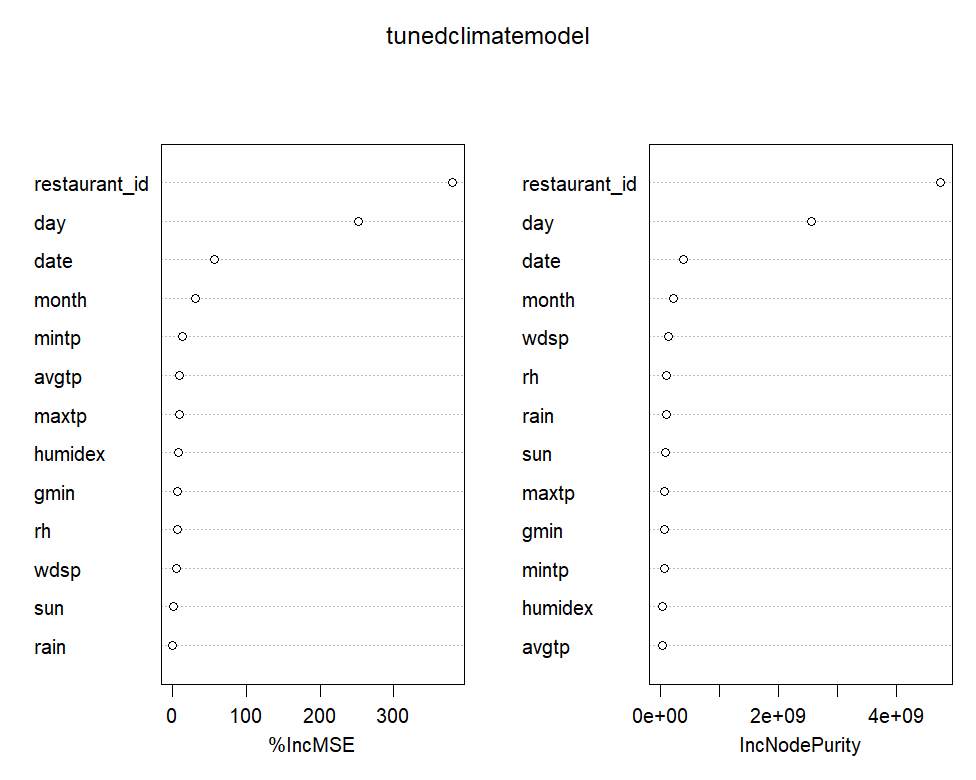
Figure 4‑1; Model Performance Metrics; MAE, RMSE & R-Squared

The dual-axis plot (fig. 4-1) provides a comprehensive view of how different metrics compare across models. The Tuned\_Climate\_Model consistently shows the lowest values for both MAE and RMSE, indicating more accurate predictions, fewer errors, and lower deviations compared to other models. The Tuned\_Climate\_Model stands out with the lowest values for both MAE and RMSE, reflecting its superior performance in minimising prediction errors, alongside a slightly higher R-squared figure, representing variability accounted for.

Figure 4‑2; Comparison of Actual vs Predicted Sales for Random Forest Models

The model's prediction results are seen in the provided graph (*see fig. 4-2*). Overall, it can be concluded that all four models perform accurately using weather variables, notably between the ranges of 1250 and 3750. The models follow a consistent trajectory, with a few outliers. One potential reason for predicting lower sales for larger actual figures may be the lower likelihood of these higher numbers. These high figures can be influenced by various factors within the shops, such as a large number of takeaways for events, multiple parties on site, and purchased vouchers.

Table 4‑2; Variable Importance Display of tunedclimatemodel



The varaible importance results from the *tunedclimatemodel* show that the most important factor in predicting sales is the restaurant ID. This suggests that unique aspects of each restaurant, like its location or the type of customers it attracts, play a critical role in forecasting sales. Time-related factors, such as the day of the week, specific dates, and the month, are also key drivers of sales. The day and date, in particular, stand out, emphasising how patterns like weekday versus weekend and seasonal trends significantly influence sales performance. When it comes to weather, windspeed is a noteworthy factor, though it doesn't have as much impact as time-based variables. This indicates that while weather does affect sales, its influence is more subtle. Other weather elements like relative humidity, maximum temperature, and minimum temperature also play a role, though to a lesser extent. Sunshine and rain, on the other hand, have a minimal impact on sales, suggesting that these weather conditions aren't as crucial as other factors.

# Initial Observations

Figure 4.1.1‑1; Time Series Line Plot; Monthly Sales For Product Categories (Product\_Data)

A graph of different colored lines

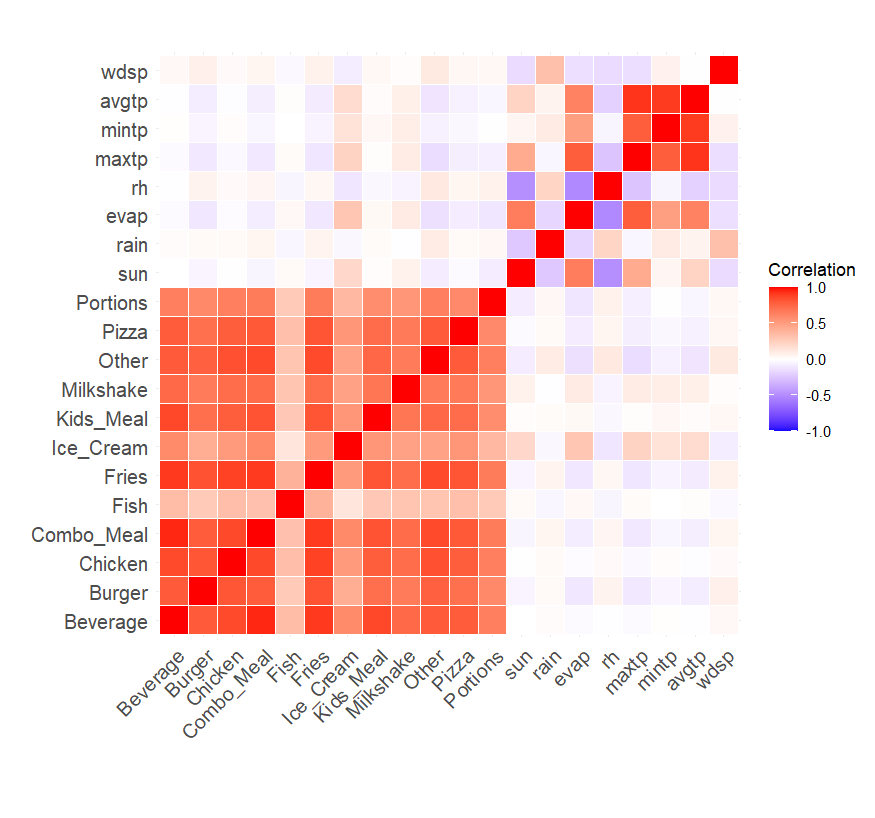
Description automatically generated

Upon examining fig. 3-10 and fig. 4.1. 1-1, it is evident that there are slight increases in consumer activity in the summer months, followed by a sharp drop in September, likely corresponding with the start of national schools, secondary schools, and colleges. This trend begins to slowly recover after the end of the summer holidays.

# Matrix: Correlation & Scatterplot

Some weak positive and negative relationships can be seen between the product categories and weather variables. As well as between the food and weather, relationships between the weather variables can be seen clearly, for example, relative humidity has a strong inverse relationship with the duration of hours of sunlight, whilst evaporation also has a strong positive relationship with maximum temperature achieved.

Figure 4.1.2‑1; Correlation Matrix: Restaurant #1 Product\_Data



# Correlation Analysis of Product Categories

### Beverage

The *Beverage* category shows moderate positive correlations with *Ice\_Cream* (0.59) and weak positive correlations with *Fish* (0.34). These relationships suggest that *Beverage* sales increase in conjunction with *Ice\_Cream* sales, possibly indicating a preference for both items being purchased together. However, there are strong correlations positive correlations between popular products including *Combo\_Meal* (.95), *Fries* (.91) and *Kids\_Meal* (.87), which can indicate frequent pairings among customer purchases. In contrast, *Beverage* sales show negligible correlations with weather variables, indicating that weather conditions have minimal impact on Beverage sales.

### Burger

Burger sales exhibit a strong positive correlation with *Fries* (0.83), reflecting the common pairing of burgers and fries in meal orders. This strong association reinforces the idea that customers frequently purchase these items together. There is a weak correlation between *Fish* (.28) and *Portions* (.60). Other correlations with weather variables are very weak, suggesting that weather conditions have little to no effect on *Burger* sales.

### Chicken

Chicken sales are strongly positively correlated with *Fries* (0.89), indicating that these items are often ordered together. This strong correlation underscores a typical consumer preference for pairing chicken with fries. There is a moderate correlation with *Ice\_Cream* (0.52), indicating that it is frequently purchased alongside. The weak or negligible correlations with weather variables further confirm that chicken sales are not significantly influenced by changes in weather conditions.

### Combination Meals

*Combo\_Meals* also show a strong positive correlation with *Fries* (0.91) and Beverage (.95), suggesting that *Combo\_Meals* frequently include fries and a drink. The correlations with weather variables are very weak, indicating that weather has minimal effect on the sales of *Combo\_Meals.*

### Fish

*Fish* demonstrates a moderate correlation with *Fries* (.40) but shows weak relationships with other product categories and weather variables. This weak association suggests that *Fish* and *Fries* are somewhat related in sales but not strongly. it generally has a more isolated sales pattern. Weather variables show very weak correlations with Fish sales, indicating that weather conditions do not significantly influence Fish sales.

### Fries

*Fries* have a strong positive correlation with both *Ice\_Cream* (0.52) and *Kids\_Meal* (0.82). This suggests that *Fries* are often bought with *Ice\_Cream* and *Kids\_Meal*, indicating a trend where *Fries* go well with other popular items. *Beverage* (.91), *Burger* (.83), *Chicken* (.89), and *Combo\_Meal* (.91) show the strongest correlations among the products, indicating that fries are commonly purchased regardless of the choice of the main item. The correlations with weather variables are generally weak, suggesting that *Fries* sales are largely unaffected by weather conditions.

### Ice Cream

*Ice\_Cream* shows a moderate positive correlation with *Kids\_Meal* (0.54), reflecting a trend where both items are often purchased together. Additionally, *Ice\_Cream* has a positive correlation with evaporation rates (0.30), indicating that higher temperatures or greater evaporation might drive increased *Ice\_Cream* sales. The weak negative correlation with relative humidity (rh) suggests that high humidity does not significantly deter *Ice\_Cream* purchases.

### Kids Meal

The data shows a strong positive correlation between *Kids\_Meal* and *Milkshake* (0.69), indicating that milkshakes are a popular addition to *Kids\_Meal* orders. There is also a significant correlation between *Beverage* (0.87) and *Combo\_Meal* (0.83), suggesting that parents may choose these to keep costs down. On the other hand, the correlations with weather variables are very weak, suggesting that weather conditions do not significantly influence *Kids\_Meal* sales.

### Milkshake

Milkshake sales show a strong positive correlation with multiple products such as *Beverage* (.74), *Burger* (.67), *Chicken* (.73) and *Combo\_Meal* (.73) as well as *Fries* (.73), suggesting that milkshakes are frequently ordered alongside other items. Weak positive correlations with weather variables indicate that *Milkshake* sales are only slightly influenced by changes in weather conditions.

### Pizza

*Pizza* shows significant positive correlations with several products: *Beverage* (0.80), *Burger* (0.72), *Chicken* (0.79), and *Combo\_Meal* (0.81). This suggests that pizza is often bought together with these items, indicating that these products might be part of a larger meal or popular combination. The weak correlations with weather variables indicate that weather conditions do not significantly impact Pizza sales.

### Portions

*Portions* exhibit moderate positive correlations with *Pizza* (0.60) and weak positive correlations with other items such as *Burger* (0.60), *Beverage* (.65) and *Chicken* (0.65). This suggests that *Portions* are somewhat associated with the sales of these products but are less strongly correlated compared to other items. The weak correlations with weather variables imply that weather conditions have minimal influence on *Portion* sales.

### Other

The *Other* category has moderate positive correlations with *Milkshake* (0.67) and *Pizza* (0.80). This suggests that the *'Other'* category, which might include various additional items, often accompanies *Milkshakes* and *Pizza*. The correlations with weather variables are very weak, indicating that weather does not significantly affect the sales of items categorised as *'Other'*.

# Correlation Analysis with Weather Variables

### Sun

Sun exposure has a weak positive correlation with *Ice\_Cream* (0.20) and *Milkshake* (0.07), suggesting that sunny weather might have a minor positive impact on these products. This minor effect could be attributed to the fact that cold products are typically consumed throughout the year in colder, damper climates, and therefore have less impact when the sun is shining.

### Rain

Rainfall has very weak correlations with all product categories, indicating that rain does not significantly impact sales. *Ice\_Cream* (-0.04) displays a weak inverse relationship with rain, suggesting that ice cream sales insignificantly decrease during periods of wet weather. Similarly, *Milkshake* (-0.1) also displays a similar relationship.

### Evaporation

Evaporation rates show a moderate positive correlation with *Ice\_Cream* (0.30), indicating increased ice cream sales in warmer conditions. Sales of *Burgers* (-0.10), *Fries* (-0.10), and *Pizza* (-0.08) insignificantly decrease in warmer weather conditions.

### Windspeed

Windspeed exhibits relatively weak correlations with all product categories. The strongest positive correlation is with the *'Other'* category (0.11), suggesting a minor association between higher wind speeds and sales of miscellaneous items. However, the correlation values with other product categories, including *Beverage* (0.04), *Burger* (0.08), *Chicken* (0.02), and *Combo\_Meal* (0.05), are notably low. This indicates that variations in windspeed do not significantly impact sales of these products. The overall minimal influence of windspeed on product sales suggests that it is not a substantial factor in driving consumer behaviour for the items analysed.

### Temperature

### Average Temperature

The average temperature is weakly positively correlated with several product categories. The strongest correlation is with the *Ice\_Cream* category (0.18), indicating that higher average temperatures are slightly associated with increased *Ice\_Cream* sales. Other notable correlations include a moderately positive association with *Combo\_Meal* (0.07) and a very weak negative correlation with *Other* (-0.11). The correlations with other popular products, including *Beverage* (0.01) and *Fries* (0.08), are minimal. This suggests that while average temperature has a modest effect on *Ice\_Cream* sales, it generally does not have a substantial impact on other product categories.

### Minimum Temperature

The minimum temperature shows stronger correlations with several product categories in comparison to other weather variables. The strongest positive correlation is with *Ice\_Cream* (0.15), and there is also a weak positive correlation with *Milkshake* (0.09), indicating that higher minimum temperatures are linked to increased sales of cold products like ice cream. This pattern is also observed in other categories, such as *Combo\_Meal* (0.04) and *Milkshake* (0.10), where higher minimum temperatures are associated with slightly higher sales. The correlations with *Beverage* (0.01) and *Chicken* (0.01) are weaker, suggesting that minimum temperature has a significant impact on ice cream sales and a minor influence on other product categories.

### Maximum Temperature

Maximum temperature shows moderate correlations with various product categories. The strongest positive correlation is with *Ice\_Cream* (0.23), indicating an association between higher maximum temperatures and increased *Ice\_Cream* sales. Other correlations include a negative relationship with Combo\_Meal (-0.10) and a moderate correlation with Milkshake (0.10). The correlations with Beverage (-.03) and Chicken (-.03) are weak. This suggests that maximum temperature plays a substantial role in driving Ice\_Cream sales and has a minor effect on other products, particularly in warmer conditions.

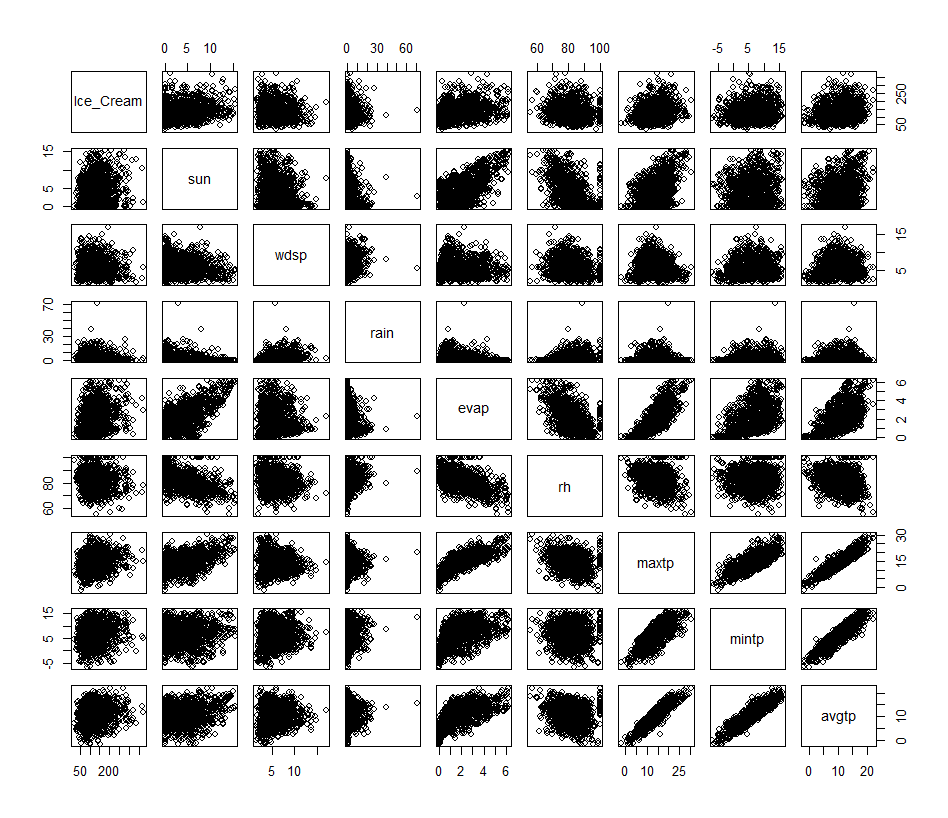
### Relative Humidity

Relative humidity is generally weakly correlated with product categories. Some notable correlations include a weak negative trend with ice cream (-0.11) sales, suggesting that higher humidity might slightly reduce ice cream sales. Additionally, there is a weak positive trend with Other (0.12) and Portion (0.07) sales, indicating slight increases in sales during higher humidity weather conditions.

## Scatterplot Matrices Analysis

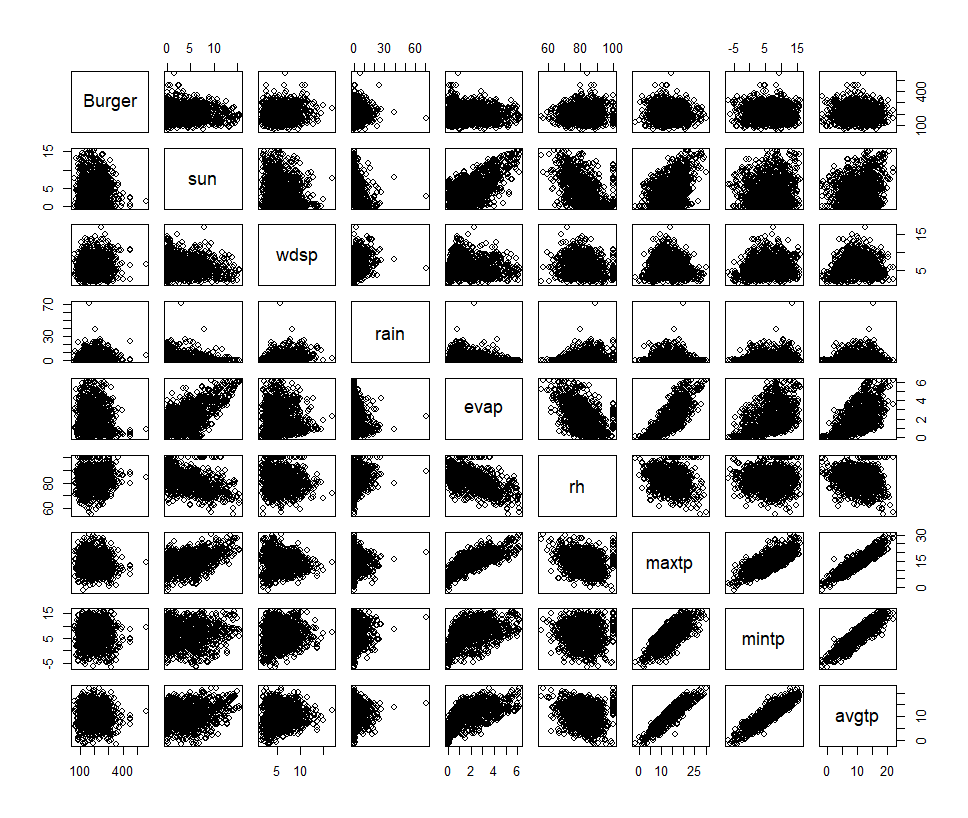
The pairs() function is used to visualise the matrix of scatterplots for each product category against its primary weather variables. It provides a clearer and more certain view of observations seen before, such as the weakened effect of cold weather on the sales of cold items.

Figure 4.2.1‑1; Example Scatterplot Matrix of Ice-Cream & Weather Variables



The data presented in *Fig*. 4.2. 1-1 reveals various relationships between weather factors and ice cream sales. A weakened positive correlation is evident between sunshine duration and ice cream sales, indicating that longer sunshine hours slightly boost ice cream product consumption. Conversely, there are weak negative correlations between ice cream sales and both wind speed and rainfall amount, suggesting that higher wind speeds and increased rainfall slightly decrease ice cream product sales. Additionally, a strong linear correlation exists between sunshine duration and evaporation, showing that longer sunshine hours significantly enhance evaporation rates. Moreover, there is a strong linear negative correlation between sunshine duration and relative humidity, meaning that increased sunshine hours substantially reduce relative humidity levels. This observation aligns with the positive linear relationship found between evaporation and temperature variables, reinforcing support of the correlation between these weather-specific factors.

Figure 4.2.1‑2; Example Scatterplot Matrix of Burger & Weather Variables



The scatterplot matrices reveal a weak inverse relationship between burger product sales and increasing temperatures. This trend suggests that customers are less likely to purchase hot items, such as burgers, during higher temperatures. Additionally, a weak positive relationship exists between burger product sales and increasing evaporation rates. Furthermore, the data shows a weak positive relationship between burger product sales and windspeed. Aside from high temperatures, burger product sales appear to be relatively unaffected by other weather variables, making them a consistently popular item regardless of weather conditions.

## Linear Regression Findings

Individual linear regression models are developed for each product category. The primary goal of these models is not to create predictive tools but to conduct an exploratory analysis to identify further relationships between weather variables and product categories within the restaurant. It is expected that variables with the highest correlation will be the best predictors (*refer to Table 8-8-4*). The variables included in each model are selected based on their significant impact on human body sensations and overall comfort levels (*refer to Table 4.2. 1-2*).

Upon creation of the models, a VIF (variation inflation factor) analysis was performed which showed high multicollinearity between the temperature variables. To reduce multicollinearity and prevent this from complicating the model’s ability to isolate the individual features, the *mintp* and *maxtp* variables were substituted for *avgtp* alone. For all product models, the VIF remained consistent.

Figure 4.2.1‑3; Final VIF Figures

A black text on a white background

Description automatically generated

* + 1. Burger Findings

Among the weather variables tested, windspeed emerged as the only significant factor, displaying a positive relationship with burger sales (coefficient = 2.0600, p-value = 0.0344). This finding suggests that higher wind speeds may be associated with a slight increase in sales, possibly because consumers seek comfort from warmer foods during cooler, windier conditions, as observed in Yoo et al’s study documenting the sales responses of various Korean businesses to extreme temperatures [9]. Other weather variables, including sunlight, rainfall, evaporation, relative humidity, and average temperature, did not exhibit significant effects on burger sales.

* + 1. Beverage Findings

The results reveal that no predictors—*sun, rain, evap, rh, maxtp, mintp, avgtp, and wdsp*—show statistical significance. All p-values exceed the 0.05 threshold, suggesting that, within the scope of this model, none of the predictors have a statistically significant influence on the *Beverage* variable. The predictor closest to significance is windspeed (*wdsp*, p =0.429). This implies that beverage sales trends are not significantly influenced by weather conditions, despite beverages being a popular category.

* + 1. Chicken Findings

None of the weather variables tested emerge as statistically significant at the conventional 5% level. The predictor closest to being significant is windspeed (*wdsp*, p = 0.432). This predictor still is yet to meet the significance threshold, however, it is still worth noting windspeed is the closest factor of influence. These results indicate that chicken sales are not significantly influenced by weather variables.

* + 1. Combination Meal Findings

Using the 5% level of significance, no weather variables tested were found to be statistically significant. However, the maximum temperature came close, with a p-value of 0.0531. This suggests that maximum temperature may have some influence on combination meal sales. The coefficient for maximum temperature is -32.708, which hints at a slight negative relationship—higher temperatures could be linked to a decrease in combo-meal sales. Customers may not be inclined to opt for multiple hot food items in warmer temperatures.

* + 1. Fish Findings

Similar to the results of the combination meals model, no weather variables tested showed significant effects on fish sales at the 5% level. The coefficients for these variables are small and not statistically significant, with p-values well above the conventional threshold. For example, windspeed has a p-value of 0.3962, and relative humidity has a p-value of 0.3022, both of which suggest no strong relationship with fish sales.

* + 1. Fries Findings

In the *fries* linear regression model, windspeed shows a significant influence on chip product sales (*wdsp*, p = 0.0647). The remaining predictors do not reach any form of significance within the *fries* model, suggesting windspeed positively affects chip product sales in a slight manner (estimate = 5.1455).

* + 1. Ice-Cream Findings

Evaporation (*evap*, p = 7.95e-09, estimate =11.7407) is revealed to be a highly significant positive factor in ice cream product sales. One reason this may be is due to the association between high temperatures and increased evaporation rates. Ice cream products may be purchased as a consumer behavioural response in reaction to ease discomfort.

* + 1. Kids-Meal Findings

In the Kids Meal model, none of the weather variables reach the 5% significance level. While minimum temperature (*mintp*, p = 0.0789) initially shows a slight degree of significance, this effect diminishes when the average temperature is included in place of maximum and minimum temperatures. However, this may be due to the high correlation between sun and temperature variables. Ultimately, none of the variables demonstrate statistical significance.

* + 1. Milkshake Findings

The milkshake model did not result in any significant influencing variables. The variable closest to a significance threshold was evaporation level (*evap*, p=0.1431).

* + 1. Pizza Findings

The pizza regression model revealed no significant variables also. The closest significant variable is windspeed (*wdsp*, p = 0.165572). The variable results suggest weather factors alone do not significantly affect pizza sales.

* + 1. Portions Findings

Although *portions* is a category primarily based on what consumers have already ordered, a regression model is constructed to analyse the influential weather variables. Evaporation (*evap*, p = 0.0299) is revealed to be a significant influencing weather variable in portion sales, having a slight negative impact on sales (estimate = -4.69511). This may be due to the fact that digestion slows down during warmer weather and customers do not feel the need to add extras or add-ons to their meal [76], [77].

* + 1. Other Findings

The *other* category is comprised of miscellaneous products that do not fit into the other categories, however, various significant relationships are identified. Windspeed (*wdsp*, p = 0.000482) demonstrates the strongest positive correlation with sales, with a highly significant p-value, indicating that sales increase notably as windspeed rises ( estimate = 3.9528). Relative humidity (*rh*, p = 0.002058) also shows a strong positive impact on sales, though to a lesser extent than windspeed ( estimate = 1.2654). Lastly, average temperature (*avgtp*, p = 0.016351) exhibits a weaker but still statistically significant relationship, with sales tending to decrease as temperatures rise (estimate = -1.8746).

## Product – Category Summary

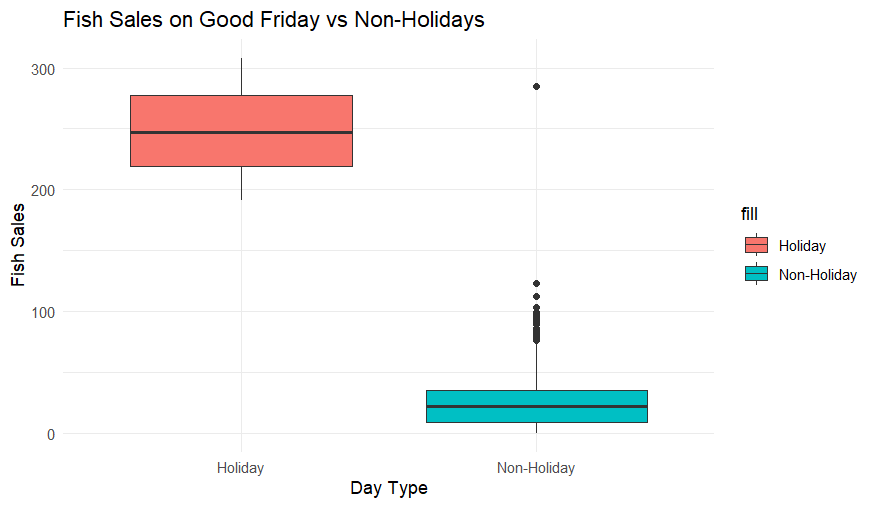
Table 4.2.1‑1; Summary of Weather Variable Impacts on Product Category Sales

|  |  |  |  |
| --- | --- | --- | --- |
| Product Category | Most Significant Weather Variable (p-value) | Impact (Positive/Negative) | Closest Variable to Significance  (if none significant) |
| Burger | Windspeed  (p = 0.0344) | Positive | N/A |
| Beverage | None | N/A | Windspeed  (p = 0.429) |
| Chicken | None | N/A | Windspeed  (p = 0.432) |
| Combination Meal | None | N/A | Maximum Temperature  (p = 0.0531) |
| Fish | None | N/A | Windspeed  (p = 0.3962) |
| Fries | Windspeed (p = 0.0647) | Positive | N/A |
| Ice-Cream | Evaporation (p = 7.95e-09) | Positive | N/A |
| Kids-Meal | None | N/A | Minimum Temperature  (p = 0.0789) |
| Milkshake | None | N/A | Evaporation  (p = 0.1431) |
| Pizza | None | N/A | Windspeed  (p = 0.165572) |
| Portions | Evaporation  (p = 0.0299) | Negative | N/A |
| Other | Windspeed  (p = 0.000482) | Positive | N/A |

## Consumer Behaviour Graphs & Trends

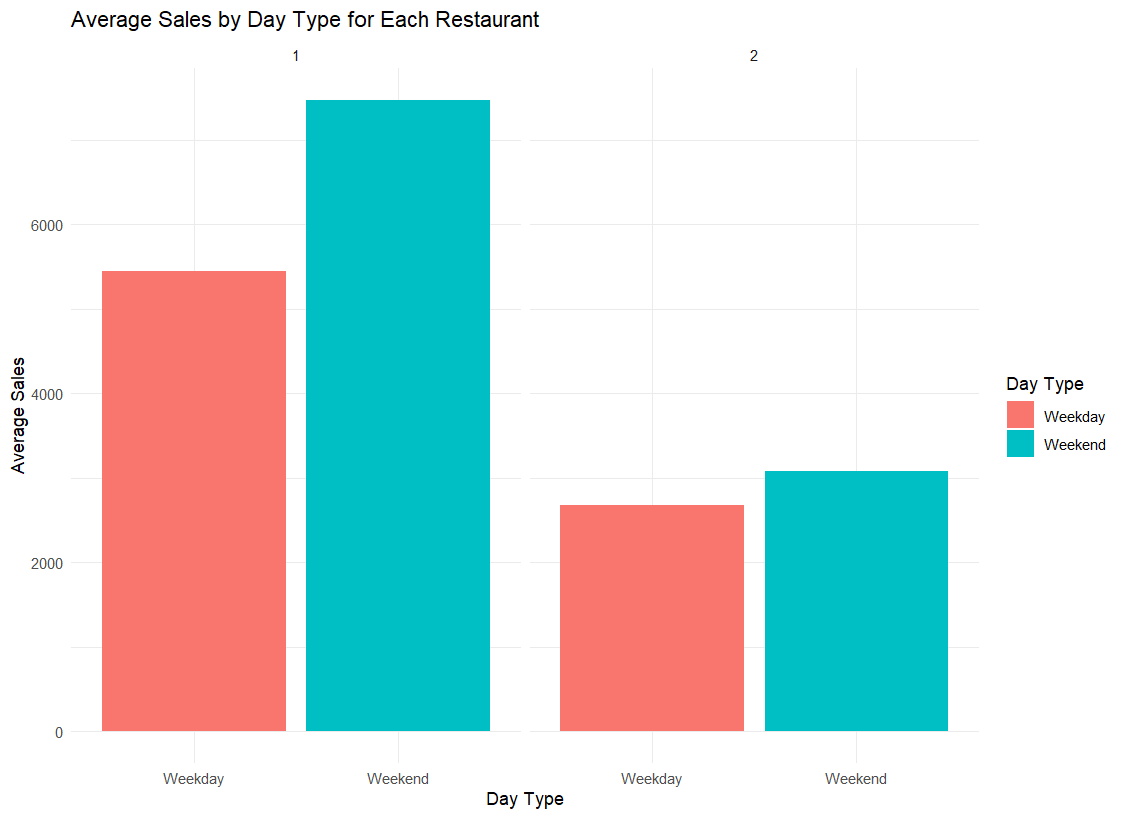
Consumer behaviours and preferences are influenced by culture, social norms and traditions [78]. An examination of some of the most celebrated Irish holidays’s daily product category sales figures shows above-average sales for certain products emphasising the role of tradition and routine in customers’ lives.

Figure 4.2.1‑4; Consumer Behaviour Boxplot Example Trend #1



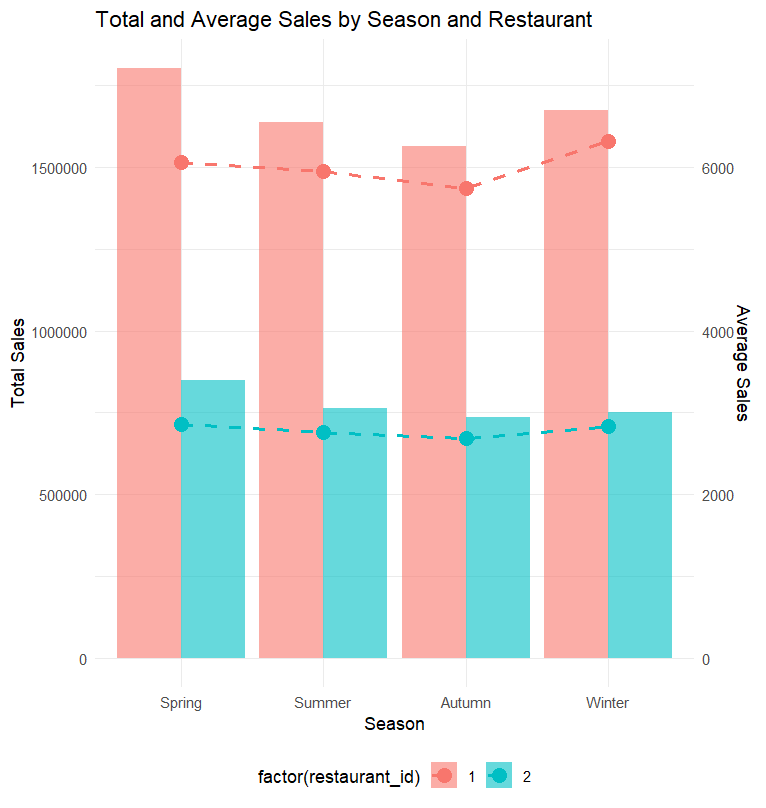
Fish, an unpopular food category, experiences a 90% increase in its average daily sales. Good Friday, a Catholic tradition followed by many Irish residents, involves consuming fish instead of other meat. This tradition ensures that the fish category will remain on the menu despite its low annual sales.

Figure 4.2.1‑5; Barchart of Both Restaurants’s Mean Sales on Weekends & Weekdays



The analysis shows that spatial variables such as date, day, and month are highly important. It was found that customers tend to prefer purchasing fast food on weekends (Saturday and Sunday) compared to weekdays (Monday to Friday). This preference is particularly strong for restaurant #1, which is centrally located and experiences high foot traffic. The trend coincides with traditional patterns of discretionary spending and shopping trips, as shown in figures 4.2 1-4. However, this trend has less impact on restaurant #2, likely due to its less central location. This could be because customers passing by on their way home may choose to stop there, especially considering the limited dining options in the area. Furthermore, the restaurant provides a more economical choice for those who don't want to cook and are looking to avoid higher delivery fees from other centrally located restaurants.

Figure 4.2.1‑6; Dual Axis Barchart Displaying Total & Mean Sales 2021-2024



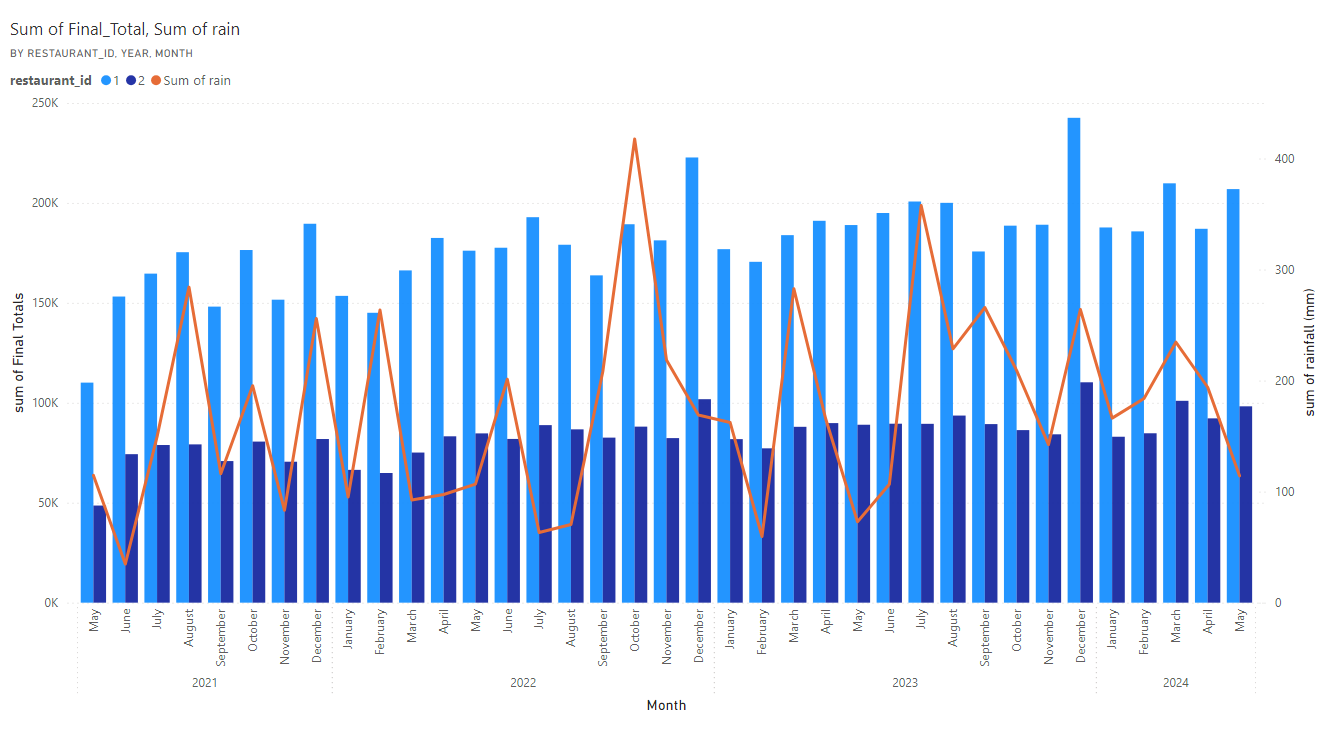
The analysis of sales data in the Irish quick-service restaurant industry shows distinct seasonal trends in consumer behaviour. The seasons are grouped as follows: Spring (March-May), Summer (June - August), Autumn (September - November), and Winter (December - February). In the Spring months, the data indicates that sales for restaurant #1 peak and the average sales for restaurant #2 are the highest. This suggests that consumers are more inclined to make purchases during this period, likely due to more favourable weather conditions, such as milder temperatures and extended daylight hours, which encourage outdoor activities and dining out.

During the Summer, while total sales remain substantial, they are slightly lower than in Spring. This reduction in total sales may be attributed to variations in consumer behaviour related to summer activities, such as holidays or different dining preferences. The warmer weather might not consistently drive higher sales as expected, possibly due to shifts in consumer habits or preferences during this season.

Autumn shows a consistent decline in sales compared to both Spring and Summer for both restaurants. This decline could be a result of the transition to cooler temperatures and shorter days, which might reduce consumer activity and outdoor dining. The steady but lower level of sales during autumn indicates that while consumer preferences are shifting, there is still notable purchasing activity in the fast-food sector.

In Winter, the data reveals a significant increase in the average sales for restaurant #1, even though the total sales are only slightly higher than those observed in Summer and Autumn. This pattern suggests that, despite a potential decrease in the frequency of visits during the colder weather, certain periods of the Winter-time experienced a sharp increase in their daily sales, for example, Christmas Eve, New Year’s Eve and the dates running up to these busy times.

Figure 4.2.1‑7; Rainfall & Sales Trends by Restaurant

[](https://app.powerbi.com/MobileRedirect.html?action=OpenReport&groupObjectId=a4e3fa10-a5fd-4f2c-848c-f561a38c5438&reportObjectId=6701a1a7-96d9-4d8f-887c-47f74a85b586&ctid=068b196a-2d57-407f-a70d-3c0571c3266a&reportPage=9e9b8a2c4d1e49d92d0e&pbi_source=copyvisualimage)

Although there was a low correlation between individual product categories, a general trend between the two restaurants can be observed. The more centrally located restaurant appears to have higher monthly sales during drier months, while restaurant two does not seem to be affected by rainfall amounts, maintaining a more consistent stable, monthly sales figure. A possible reason for this occurrence is restaurant #2’s primary demographic of licence-holding customers who drive to their location, whilst restaurant #1 has the benefit of foot traffic and a small car park.

# Discussion of Findings

This study aimed to assist businesses within the Irish quick-service restaurant industry by exploring the relationships between various weather variables and their daily takings, with the long-term aim of reducing costs through optimised marketing strategies, increased staff efficiency, improved inventory management, and a greater understanding of consumer preferences and behaviours. In addition to this, the research sought to contribute to already existing knowledge in this industry by investigating whether Irish consumers are influenced by specific weather conditions, such as strong wind gusts or hot, humid weather, in a manner similar to their British counterparts.

**Research Question 1; Is there a relationship between weather and quick-service restaurant sales in Ireland?**

The analysis confirmed that weather does indeed have an impact on sales in the Irish fast-food market, although this impact is relatively minor. The predictive models, particularly the *tunedclimatemodel* developed using the Random Forest technique, demonstrated a high level of accuracy. This model achieved an R-squared value of 87.42%, indicating that it could explain a significant portion of the variance in daily sales figures. The model's MAE and RMSE were 531.05 and 904.65, respectively, suggesting that the model can predict daily sales with an average error of between approximately €531 and €904. Restaurant #1 has a minimum daily sales intake of approximately 2400, whilst restaurant #2 has a minimum daily sales intake of approximately 1100. This model would be best suited for restaurant #1 due to their higher earnings, although restaurant #2 has a mean intake of €2800, which may allow it to use the model to estimate its future daily sales within a confidence interval of 530 to 900 euros.

The importance of various variables was also assessed using the variable importance plot from the *tunedclimatemodel*. It revealed that the most significant predictor of sales was the restaurant ID, which indirectly highlights the importance of factors unique to each restaurant, such as location and customer demographics. Spatial variables, like the day of the week, date, and month, also played critical roles in predicting sales, highlighting the influence of time-based patterns, such as weekend versus weekday sales and seasonal trends. This finding was consistent across all four models. The weekend versus weekday observations (refer to *Fig.* 4.2 1-4) indicate that consumer behaviour is likely influenced more by habitual or temporal patterns rather than weather patterns. However, this may be an instance of local demographics coming into play and affecting the data. Two restaurants in the Midlands cannot be representative of all of Ireland, which needs to be taken into consideration, however the findings may be generalisable and transferable in a similar context if particular examination and appropriation of variables is addressed accordingly.

Among the weather-related variables, wind speed emerged as a noteworthy factor, though its influence was less pronounced than that of the temporal variables. These results of the importance plot, aligned similarly to Rose & Dolega’s results where it was found location and windspeed to be the topmost influential factors to various retail category sales across England [8]. Upon further examination, windspeed is shown to be quite weaker, yet remains a significant weather variable in this circumstance. Relative humidity, maximum temperature, and minimum temperature also contributed to the model but with lower importance scores, suggesting that while weather conditions do influence sales, their impact is relatively modest compared to time-based and restaurant-specific factors. Other weather variables like sunshine and rain had minimal influence on category sales. Rain’s insignificant influence is to be expected from residents of a damp climate. Intense weather conditions such as storms were not present in the analysed data and it would be an interesting avenue of future research to analyse would the rain variables, as well as other variables’ influence increase as the weather intensity increases.

**Research Question 2; Is there a relationship between weather and customer product preference in an Irish quick-service restaurant?**

The analysis of consumer behaviour in the Irish fast food market reveals significant similarities with British consumer behaviour, as identified in Rose & Dolega’s study [8]. The findings suggest that weather conditions do influence customer product preferences, though the impact is relatively minor. According to the VIF analysis, multicollinearity was kept to a minimal degree within the linear regression models. The highest VIF value was for the variable *evaporation*, at 3.56, which is still below the commonly accepted threshold for concern (typically VIF > 5). Therefore, multicollinearity does not pose a significant issue in the model.

Seven out of the twelve food categories showed no significant influence from weather. These categories included beverages, chicken, combination meals, fish, kids meals, pizza, and milkshakes. Although milkshakes are not typically thought of as seasonal items, it was anticipated that there would be a subtle weather-related trend in milkshake sales due to popular marketing campaigns introducing new flavours during the Spring holiday seasons. However, this did not happen. This could have occurred due to stable sales throughout the year, indicating customer preference remained consistent. Alternatively, it’s possible that the impact of new flavours introduced during the Spring holiday seasons was not strong enough to create a noticeable shift in sales, suggesting that these campaigns did not significantly alter consumer behaviour.

Of the five influenced categories, wind speed emerged as the most significant weather-related factor, consistent with the findings from the British study. However, while wind speed did have some influence, its effect was minimal. Wind speed had a minor positive impact on several product categories, including *Burger* and *Fries*. A significant impact was seen on the miscellaneous category *Other*. This suggests that on windier days, there may be a slight increase in sales for these items, regarding products within the *Other* category this effect would be more pronounced. Windspeed also became the most common weather variable closest to being statistically significant (where no significance threshold was met). This result contrasts with the more prominent weather influence seen within the English consumer base of Rose & Dolega’s research [8].

Evaporation significantly influenced the weather, particularly in the ice cream category, where it had a positive effect, although the estimated change was minimal. This correlation is surprising, however, it makes sense, as higher evaporation rates typically correlate with warmer, drier weather, conditions under which consumers are more inclined to purchase cold treats like ice cream. The significance of increasing evaporation rates on ice cream sales rather than increasing temperature rates suggests Irish consumers are more affected by humidity than temperature when seeking sweet cold products typically marketed for warm climates.

Despite Ireland's generally cooler climate, where a lower impact on ice cream sales might have been expected, the findings reveal that ice cream remains a year-round favourite among consumers. Sales remain stable throughout the year, with only slight increases during periods of higher evaporation. This suggests that Irish consumers, accustomed to cooler weather, do not necessarily wait for warm conditions to enjoy cold food items, which are typically marketed for warmer seasons. Instead, ice cream continues to be a consistent choice, with minor upticks in demand when weather conditions become drier and warmer. Conversely, higher evaporation rates had a minor negative impact on portion sales, indicating a slight decrease in demand for this category under similar weather conditions. This subtle shift may indicate that, in drier, warmer weather, consumers might favour fewer additions to their food.



# Conclusion, Recommendations & Reflection

The objectives of this study were to analyse the relationship between weather variables and sales performance in the Irish quick-service restaurant industry, investigate the influence of weather conditions on consumer product preferences within Irish quick-service restaurants and develop a predictive model to forecast sales trends based on historical weather and sales data.

The study’s objectives were largely met.

1. **The relationship between weather variables and Irish quick-service restaurant industry sales performance.**

The relationship between weather variables and sales performance was thoroughly analysed, uncovering some minor influences, particularly regarding wind speed. The analysis revealed that wind speed had some impact on consumer behaviour, possibly affecting the likelihood of customers venturing out to quick-service restaurants. This finding is consistent with other studies conducted in similar geographic regions, where wind speed has been observed to influence retail foot traffic. However, the extent of this impact in Ireland was less pronounced than anticipated. The findings suggest that Irish fast-food customers are more likely to be influenced by internal and spatial factors, like location and date, rather than external factors such as weather. This is supported by the variable importance rankings of the model, where restaurant ID, date, and day consistently ranked the highest in explaining variability.

1. **The influence of weather conditions on consumer product preferences within Irish quick-service restaurants.**

The influence of weather on consumer product preferences was also investigated, although the findings were less conclusive. The goal was to identify whether certain weather patterns led to shifts in consumer choices, such as preferring hot foods on cold days or opting for lighter meals during warmer weather. The investigation into product preferences under different weather conditions produced ambiguous results. While there were some indications that higher temperatures might lead to increased sales of cold, sweet food categories like ice cream, these trends were not strong enough to draw definitive conclusions. Another related food category, milkshake, did not show significant results in high temperatures, although a correlation may be revealed with additional data and further research. The data did not consistently demonstrate a clear pattern of weather-induced changes in product preferences across the Irish fast-food industry.

1. **Development of a predictive model to forecast sales trends.**

This study’s primary objective was to develop a predictive model capable of forecasting sales based on past sales and weather data. The models were successfully developed and the objective was successfully achieved, through the use of the Random Forest technique. After excluding business closing days for both businesses, 2230 rows of data were used to train and test models based on various weather-related and time-related varibales, obtained through online and primary sources. The Random Forest algorithm was selected for its ability to handle large datasets with multiple input variables, including both continuous and categorical data. It is effective in capturing complex, non-linear relationships between these variables. The best-performing model, tunedclimatemodel, demonstrated strong predictive capabilities by accurately forecasting sales in response to changing weather conditions, with 87% variability explained. Its predictions typically fall within a range of 531 to 905 euros off-target. This model can be considered successful due to its high r-squared value and moderate RMSE and MAE figures.

The success of the predictive model suggests that fast-service restaurants in Ireland could benefit from integrating weather forecasts into their sales predictions. By doing so, rota management, inventory management and marketing strategies can be optimised and aligned with expected sales volume within a certain confidence interval. This approach can improve customer satisfaction by ensuring the availability of popular food items during peak times.

**Limitations**

There are areas where the study could have been improved. Given the cultural and habitual similarities between the United Kingdom and the Republic of Ireland, wind speed was anticipated to play a similarly significant role within the Irish market. The limited impact of weather variables suggests that a more extensive dataset or more thorough research methods may result in stronger insights. The potential data shortage and the need to consider other influential factors, such as economic conditions or marketing strategies, indicate that future research could expand on this work by incorporating a broader range of variables. Moreover, using data from an alternative weather station (Casement Weather Station) may have affected the data in ways that were not previously recognised. The sun data utilised in the study might not accurately represent the weather conditions experienced by the restaurants.

Another limitation is that while this study primarily focused on weather-related factors, other external factors could indirectly influence sales and potentially impact the model's understanding of the data. Factors such as economic conditions, marketing strategies, and local competitors' activities were not explored in this analysis. Investigating these factors and integrating them may lead to a more comprehensive understanding, clearer insight into customer product preferences, and a more accurate model.

One significant limitation of this study is the assumption that consumer behaviour and product preferences remain consistent over time. However, these behaviours and preferences are subject to change due to various factors, such as lifestyle shifts, health trends, and evolving concerns within the food industry, which could alter the relationship between weather and sales [79]. For instance, evolving health and dietary trends can significantly influence consumer choices, leading to shifts in demand for different types of products offered by quick-service restaurants. A notable example is the introduction of McDonald’s limited-time vegan burgers in 2021, which, due to popular demand and changing dietary preferences, became a permanent menu item across many franchises in Europe [80]. Such shifts could modify the relationship between weather and sales, as consumers make decisions based on these evolving preferences rather than external factors like the weather. Moreover, lifestyle changes, such as the increase in remote work and the rise of food delivery services, have a profound impact on consumer behaviour. With more people working from home, there may be reduced foot traffic in certain areas, lessening the influence of weather on in-person dining at fast-food restaurants. Additionally, the convenience offered by food delivery services may further diminish the weather's impact on sales, as consumers can now order from the comfort of their homes, regardless of outdoor conditions.

**Future Work**

There are many avenues of interest for future research to take place concerning this study. In future research, it would be beneficial to use a more comprehensive dataset that includes additional variables. This could include economic indicators (e.g. unemployment rates), detailed marketing data (e.g. promotional campaign periods), and more accurate weather data (e.g. localised sun duration figures). This study covered four years of data from each restaurant, which may not fully account for long-term sales trends. A larger dataset, spanning several years' worth of sales data, could lead to a more detailed analysis and help uncover trends that might otherwise go unnoticed. Liu et al’s use of additional temporal-based dummy variables (e.g. Monday, Tuesday, etc.) may prove to be an influential addition to a future model [11]. These variables may provide some insight into which exact days, exact months or periods of the holiday season are the highest in sales volume and provide useful information to the business.

From a methodological perspective, while the Random Forest algorithm proved effective, exploring other advanced machine learning techniques could further enhance the predictive accuracy of the model. Techniques such as multiple linear regression could provide a more straightforward interpretation of relationships between variables, particularly in understanding how individual weather factors impact sales. Additionally, gradient boosting machines, which are similar to Random Forest but more powerful, could better capture complex interactions and manage multicollinearity within the variables. Gradient boosting machines might offer improved performance, especially when integrated with other influential factors like economic trends, seasonal variations, and promotional periods.

**Reflection**

Using the Gibbs cycle as a guide, which involves the six stages: description, feelings, evaluation, analysis, conclusion and action plan, I will reflect on my experience writing my first thesis and what I would change in repeating this venture [81].

1. **Description**

This study involved investigating the relationship of weather’s effect on sales performance within the Irish quick-service industry. The purpose of this study was to build a predictive model using machine learning techniques to help effectively forecast sales figures using weather and time-based information. The technique I chose to perform this was Random Forest, both because of its suitability and my familiarity with the technique, which made me more comfortable interpreting the model’s output.

1. **Feelings**

At the start of the academic year, I felt daunted by the idea of writing a thesis, especially as the first in my family to pursue a master’s degree. I initially considered opting for a more familiar field, like business but ultimately chose to follow my instincts, knowing I would have regretted not doing so. Forming an idea for my project proposal was challenging, but I found guidance by reviewing example reports from businesses, which helped me shape my project. Unfortunately, my topic was already a popular one however it also allowed me to look toward past research as a guide and test out their results in comparison to my updated results by location and industry sector. Once my idea began to take shape, the project planning and decision-making became easier. I received great advice and assistance from my lecturers and classmates throughout this period of uncertainty. Soon afterwards, my topic was accepted and finalised, which boosted my confidence.

The project was not without its challenges, particularly during the data pre-processing stage. Importing data from unfamiliar software was time-consuming, but once completed, the analysis became much smoother, easing my stress as the deadline slowly approached. A significant hurdle to overcome was the initial shortage of data at the beginning of the data collection. Restaurant #1 underwent a system change in 2021 and could not provide further data before that year, leading me to urgently seek additional data from a similar restaurant. This situation incurred a brief period of confusion and self-doubt for me where I felt lost and unable to approach my thesis paper. I began to question myself and how I would approach the new data. Regular meetings with my supervisor helped me stay focused, and eventually, I secured the necessary data, though differences between the two restaurants contributed to the study’s limitations.

Other challenges included coding difficulties, but, the abundance of online tutorials and helpful guides made it easier to find solutions or identify the problem, which I am very thankful for. I found it helpful when I came across a difficult tutorial to talk through complex code as if explaining to someone else, rather than bottling up the confusion, which prevented me from becoming overwhelmed. Knowing I could seek advice from my supervisor or classmates also provided reassurance.

1. **Evaluation**

The experience taught me the value of perseverance and self-confidence. Despite initial doubts and challenges, I managed to overcome internal and external obstacles like procrastination and the unfortunate challenges caused by the software changes by the businesses as well as the missing sun data from Mullingar Weather Station. I feel more comfortable building random forest models as well as linear regression models compared to before. I attribute this to the fact that I was creating them outside of the academic environment and depending on myself.

1. **Analysis**

In hindsight, I am grateful for the challenges faced during the data preprocessing. These experiences were critical learning experiences that I have learned from and can now reflect on. One of the key takeaways was the importance of thorough preparation before diving into the data. I initially made the mistake of not fully assessing the availability and completeness of the data, which I thought would save time but ended up costing more time in the long run. These difficulties highlighted areas I need to improve, such as technical skills and stress management. Despite the challenging aspects of my thesis journey, the experiences provided me with a clearer understanding of how to proceed next time and the steps I need to take to improve in the future.

1. **Conclusion**

Overall, the process of writing this thesis was both challenging and enlightening. It pushed me out of my comfort zone and forced me to overcome obstacles independently and think on my feet. A willingness to adapt, motivation and planning have shown to be very important in this journey. I’ve come to realise setbacks and difficulties, while frustrating, are often key drivers in personal growth and development. I hope that the younger members of my family will look at this achievement and feel inspired to pursue their own academic goals.

1. **Action Plan**

If I were to undertake a similar project in the future, I would begin by performing a thorough analysis of the data available, ascertaining its quality and how much is available for use. This step would help avoid delays and ensure a smoother workflow with fewer obstacles.

Additionally, I plan to engage in continuous learning to maintain and enhance my data analysis skills. I will do this by participating in online courses through platforms like LinkedIn Learning and revisiting publicly available video tutorials from past lecturers. This will keep the knowledge I have accumulated during my studies fresh, even when I’m not actively using the software.

# Bibliography

[1] J. L. Sanders and M. S. Brizzolara, ‘Relationships between Weather and Mood’, *The Journal of General Psychology*, Jul. 1982, Accessed: Apr. 19, 2024. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/00221309.1982.9709917

[2] E. Howarth and M. S. Hoffman, ‘A multidimensional approach to the relationship between mood and weather’, *British Journal of Psychology*, vol. 75, no. 1, pp. 15–23, 1984, doi: 10.1111/j.2044-8295.1984.tb02785.x.

[3] A. Furnham and R. Milner, ‘The impact of mood on customer behavior: Staff mood and environmental factors’, *Journal of Retailing and Consumer Services*, vol. 20, no. 6, pp. 634–641, Nov. 2013, doi: 10.1016/j.jretconser.2013.06.001.

[4] A. T. Steele, ‘Weather’s Effect on the Sales of a Department Store’, *Journal of Marketing*, vol. 15, no. 4, pp. 436–443, Apr. 1951, doi: 10.1177/002224295101500404.

[5] K. B. Murray, F. Di Muro, A. Finn, and P. Popkowski Leszczyc, ‘The effect of weather on consumer spending’, *Journal of Retailing and Consumer Services*, vol. 17, no. 6, pp. 512–520, Nov. 2010, doi: 10.1016/j.jretconser.2010.08.006.

[6] ‘Temperature-Sensitive Billboards’, TrendHunter.com. Accessed: Jan. 26, 2024. [Online]. Available: https://www.trendhunter.com/trends/cocacola-billboard

[7] X. Tian, S. Cao, and Y. Song, ‘The impact of weather on consumer behavior and retail performance: Evidence from a convenience store chain in China’, *Journal of Retailing and Consumer Services*, vol. 62, p. 102583, Sep. 2021, doi: 10.1016/j.jretconser.2021.102583.

[8] N. Rose and L. Dolega, ‘It’s the Weather: Quantifying the Impact of Weather on Retail Sales’, *Appl. Spatial Analysis*, vol. 15, no. 1, pp. 189–214, Mar. 2022, doi: 10.1007/s12061-021-09397-0.

[9] J. Yoo, J. Eom, and Y. Zhou, ‘Thermal comfort and retail sales: A big data analysis of extreme temperature’s impact on brick-and-mortar stores’, *Journal of Retailing and Consumer Services*, vol. 77, p. 103699, Mar. 2024, doi: 10.1016/j.jretconser.2023.103699.

[10] D. Liu, W. Wang, and Y. Zhao, ‘Effect of weather on online food ordering’, *Kybernetes*, vol. ahead-of-print, Feb. 2021, doi: 10.1108/K-05-2020-0322.

[11] W. Yao, H. Zhao, and L. Liu, ‘Weather and time factors impact on online food delivery sales: a comparative analysis of three Chinese cities’, *Theoretical and Applied Climatology*, vol. 153, pp. 1–14, Jun. 2023, doi: 10.1007/s00704-023-04542-w.

[12] D. Gursoy and C. G. Chi, ‘Effects of COVID-19 pandemic on hospitality industry: review of the current situations and a research agenda’, *Journal of Hospitality Marketing & Management*, Jul. 2020, Accessed: Jun. 10, 2024. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/19368623.2020.1788231

[13] ‘The Grimsey Review 2’, High Streets Task Force. Accessed: Jun. 10, 2024. [Online]. Available: https://www.highstreetstaskforce.org.uk/resources/details/?id=2a3d7ab0-fc51-43aa-9192-488fc34ffded

[14] D. Koo, ‘Inter‐relationships among store images, store satisfaction, and store loyalty among Korea discount retail patrons’, *Asia Pacific Journal of Marketing and Logistics*, vol. 15, no. 4, pp. 42–71, Jan. 2003, doi: 10.1108/13555850310765033.

[15] C. Yoo, J. Park, and D. Macinnis, ‘Effects of Store Characteristics and In-Store Emotional Experiences on Store Attitude’, *Journal of Business Research*, vol. 42, pp. 253–263, Jul. 1998, doi: 10.1016/S0148-2963(97)00122-7.

[16] C. Forman, A. Ghose, and B. Wiesenfeld, ‘Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets’, Jul. 26, 2006, *Rochester, NY*: 1026893. doi: 10.2139/ssrn.1026893.

[17] J. Li and X. Liang, ‘Reviewers’ Identity Cues in Online Product Reviews and Consumers’ Purchase Intention’, *Front Psychol*, vol. 12, p. 784173, Jan. 2022, doi: 10.3389/fpsyg.2021.784173.

[18] S. Aydin, E. Koc, and N. Kaya, ‘The Influence of Critical Success Factors on B2B Sales Performance and the Mediating Role of Social Intelligence’, *Business Management and Strategy*, vol. 8, no. 2, Art. no. 2, Aug. 2017, doi: 10.5296/bms.v8i2.11563.

[19] V. Bondarenko and B. Vyshnivska, ‘PROMOTIONAL MARKETING AS A METHOD OF INCREASING SALES’, *Three Seas Economic Journal*, vol. 4, no. 2, Art. no. 2, Jun. 2023, doi: 10.30525/2661-5150/2023-2-3.

[20] R. Guesalaga, ‘The use of social media in sales: Individual and organizational antecedents, and the role of customer engagement in social media’, *Industrial Marketing Management*, vol. 54, pp. 71–79, Apr. 2016, doi: 10.1016/j.indmarman.2015.12.002.

[21] H. H. S. Abdelhamied, ‘THE EFFECTS OF SALES PROMOTION ON POST PROMOTION BEHAVIORS AND BRAND PREFERENCES IN FAST FOOD RESTAURANTS’, *tourismos*, vol. 8, no. 1, Art. no. 1, Mar. 2013, doi: 10.26215/tourismos.v8i1.330.

[22] R. Liu, Y. Sun, and J. Wang, ‘Influential Factors of Sales Revenue in the Fast-Food Industry’, presented at the 2022 7th International Conference on Financial Innovation and Economic Development (ICFIED 2022), Atlantis Press, Mar. 2022, pp. 2832–2838. doi: 10.2991/aebmr.k.220307.459.

[23] J. Juan Beristain and P. Zorrilla, ‘The relationship between store image and store brand equity: A conceptual framework and evidence from hypermarkets’, *Journal of Retailing and Consumer Services*, vol. 18, no. 6, pp. 562–574, Nov. 2011, doi: 10.1016/j.jretconser.2011.08.005.

[24] K. Kim, M.-J. Kim, and J.-K. Jun, ‘Small Queuing Restaurant Sustainable Revenue Management’, *Sustainability*, vol. 12, no. 8, Art. no. 8, Jan. 2020, doi: 10.3390/su12083477.

[25] H. M. Jeon and S. R. Yoo, ‘The relationship between brand experience and consumer-based brand equity in grocerants’, *Serv Bus*, vol. 15, no. 2, pp. 369–389, Jun. 2021, doi: 10.1007/s11628-021-00439-8.

[26] A. S. Mattila and J. Wirtz, ‘The role of store environmental stimulation and social factors on impulse purchasing’, *Journal of Services Marketing*, vol. 22, no. 7, pp. 562–567, Jan. 2008, doi: 10.1108/08876040810909686.

[27] M. Esmaeilpour and Z. Mohseni, ‘Effect of Customer Experiences on Consumer Purchase Intention’, no. 73, 2019.

[28] M. Z. Maitlo, N. Jugwani, and R. G. Gilal, ‘The Model of Customer Experience and Purchase Intention in Online Environment’, *Sukkur IBA Journal of Management and Business*, vol. 4, no. 1, Art. no. 1, Apr. 2017, doi: 10.30537/sijmb.v4i1.101.

[29] J. Coimbra and T. Proença, ‘Managerial coaching and sales performance: the influence of salesforce approaches and organisational demands’, *International Journal of Productivity and Performance Management*, vol. 72, no. 10, pp. 3076–3094, Jan. 2022, doi: 10.1108/IJPPM-10-2021-0620.

[30] K. H. Kang, S. Lee, and C. Huh, ‘Impacts of positive and negative corporate social responsibility activities on company performance in the hospitality industry’, *International Journal of Hospitality Management*, vol. 29, no. 1, pp. 72–82, Mar. 2010, doi: 10.1016/j.ijhm.2009.05.006.

[31] S. R. Miller, L. Eden, and D. Li, ‘CSR Reputation and Firm Performance: A Dynamic Approach’, *Journal of Business Ethics*, vol. 163, no. 3, pp. 619–636, May 2020, doi: 10.2307/45283718.

[32] H. Ramkissoon, F. Mavondo, and V. Sowamber, ‘Corporate Social Responsibility at LUX\* Resorts and Hotels: Satisfaction and Loyalty Implications for Employee and Customer Social Responsibility’, *Sustainability*, vol. 12, no. 22, Art. no. 22, Jan. 2020, doi: 10.3390/su12229745.

[33] Y. Inoue and S. Lee, ‘Effects of different dimensions of corporate social responsibility on corporate financial performance in tourism-related industries’, *Tourism Management*, vol. 32, no. 4, pp. 790–804, Aug. 2011, doi: 10.1016/j.tourman.2010.06.019.

[34] O. Ozdemir, E. Erkmen, and M. Kim, ‘Corporate social responsibility and idiosyncratic risk in the restaurant industry: does brand diversification matter?’, *International Journal of Contemporary Hospitality Management*, vol. 32, no. 9, pp. 2925–2946, Jan. 2020, doi: 10.1108/IJCHM-03-2020-0167.

[35] C. Martin-Rios, C. Demen-Meier, S. Gössling, and C. Cornuz, ‘Food waste management innovations in the foodservice industry’, *Waste Management*, vol. 79, pp. 196–206, Sep. 2018, doi: 10.1016/j.wasman.2018.07.033.

[36] P. Moreno and P. Tejada, ‘Reviewing the progress of information and communication technology in the restaurant industry’, *Journal of Hospitality and Tourism Technology*, vol. 10, no. 4, pp. 673–688, Jan. 2019, doi: 10.1108/JHTT-07-2018-0072.

[37] F. Linden, ‘Consumer markets: merchandising weather’, in *The Conference Board Business Record*, 1962, pp. 15–16.

[38] F. Badorf and K. Hoberg, ‘The impact of daily weather on retail sales: An empirical study in brick-and-mortar stores’, *Journal of Retailing and Consumer Services*, vol. 52, p. 101921, Jan. 2020, doi: 10.1016/j.jretconser.2019.101921.

[39] admin, ‘How The Weather Factor Influence Purchase Decisions’, HYPERVSN. Accessed: Jun. 15, 2024. [Online]. Available: https://hypervsn.com/blog/the-weather-factor-unveiling-the-influence-on-consumer-purchase-decisions.html

[40] S. Bruyneel, S. Dewitte, P. H. Franses, and M. G. Dekimpe, ‘Why Consumers Buy Lottery Tickets When the Sun Goes Down on Them.’.

[41] S. Moon, J. Kwon, S.-U. Jung, and Y. H. Bae, ‘The impact of individual differences in weather sensitivity on weather-related purchase intentions’, *International Journal of Market Research*, vol. 60, no. 1, pp. 104–117, Jan. 2018, doi: 10.1177/1470785317744855.

[42] R. Liao, ‘Alibaba splits into six in biggest overhaul in 24 years’, TechCrunch. Accessed: Oct. 27, 2023. [Online]. Available: https://techcrunch.com/2023/03/28/alibaba-splits-into-six-in-biggest-overhaul-in-24-years/

[43] J.-L. Bertrand and M. Chabot (Parnaudeau), ‘Ranking the Impact of Climate Variability on UK Retail Sectors: A Path to Resilience’, Oct. 19, 2015, *Rochester, NY*: 2675965. doi: 10.2139/ssrn.2675965.

[44] P. Appelqvist, F. Babongo, V. Chavez-Demoulin, A.-P. Hameri, and T. Niemi, ‘Weather and supply chain performance in sport goods distribution’, *International Journal of Retail & Distribution Management*, vol. 44, no. 2, pp. 178–202, Jan. 2016, doi: 10.1108/IJRDM-08-2015-0113.

[45] A. P. Sandqvist and B. Siliverstovs, ‘Is it good to be bad or bad to be good? Assessing the aggregate impact of abnormal weather on consumer spending’, *Empirical Economics: Journal of the Institute for Advanced Studies, Vienna, Austria*, vol. 61, no. 6, pp. 3059–3085, Dec. 2021, doi: 10.1007/s00181-020-02006-y.

[46] J. Ma, D. Huang, and Y. Wang, ‘The congruent and comparative impacts of weather on consumer reviews: evidence from an online forum and experiment’, *International Journal of Internet Marketing and Advertising*, Jan. 2021, Accessed: Aug. 25, 2024. [Online]. Available: https://www.inderscienceonline.com/doi/10.1504/IJIMA.2021.112788

[47] W. Kron, P. Löw, and Z. W. Kundzewicz, ‘Changes in risk of extreme weather events in Europe’, *Environmental Science & Policy*, vol. 100, pp. 74–83, Oct. 2019, doi: 10.1016/j.envsci.2019.06.007.

[48] ‘Climate Change - Met Éireann - The Irish Meteorological Service’. Accessed: Apr. 19, 2024. [Online]. Available: https://www.met.ie/climate/climate-change

[49] ‘Why are we seeing so many storms this season? - Met Éireann - The Irish Meteorological Service’. Accessed: Feb. 09, 2024. [Online]. Available: https://www.met.ie/why-are-we-seeing-so-many-storms-this-season

[50] C. Li, X. Luo, C. Zhang, and X. Wang, ‘Sunny, Rainy, and Cloudy with a Chance of Mobile Promotion Effectiveness’, *Marketing Science*, vol. 36, no. 5, pp. 762–779, Sep. 2017.

[51] C. Nizma and D. A. Siregar, ‘Analysis of Shopping Consumer Preferences Online and Offline Stores in the Pandemic Time of Covid-19’, presented at the 2nd International Conference on Science, Technology, and Modern Society (ICSTMS 2020), Atlantis Press, Sep. 2021, pp. 223–229. doi: 10.2991/assehr.k.210909.050.

[52] M. Bujisic, V. Bogicevic, H. G. Parsa, V. Jovanovic, and A. Sukhu, ‘It’s Raining Complaints! How Weather Factors Drive Consumer Comments and Word-of-Mouth’, *Journal of Hospitality & Tourism Research*, vol. 43, no. 5, pp. 656–681, Jun. 2019, doi: 10.1177/1096348019835600.

[53] P. O. Fanger, ‘Assessment of man’s thermal comfort in practice’, *Br J Ind Med*, vol. 30, no. 4, pp. 313–324, Oct. 1973, doi: 10.1136/oem.30.4.313.

[54] B. Stanley, Y. Pratama, and A. Subakti, ‘The Impact of Self-Order Kiosk and Service Quality on Customer Experience in McDonald’s Citra Garden 6 Jakarta’, *E3S Web of Conferences*, vol. 426, Sep. 2023, doi: 10.1051/e3sconf/202342602073.

[55] ‘The Benefits of McDonald’s Self Service Kiosk - Restaurant Self Service’. Accessed: Jul. 10, 2024. [Online]. Available: https://www.wavetec.com/blog/mcdonalds-leveraging-self-service-technologies/

[56] Jieun Choi, Doo-Hee LEE, and C. R. Taylor, ‘The Influence of Purchasing Context and Reversibility of Choice on Consumer Responses Toward Personalized Products and Standardized Products’, *Psychological Reports*, vol. 118, no. 2, pp. 510–526, Apr. 2016, doi: 10.1177/0033294116637730.

[57] P. C. M. Govers and J. P. L. Schoormans, ‘Product personality and its influence on consumer preference’, *Journal of Consumer Marketing*, vol. 22, no. 4, pp. 189–197, Jan. 2005, doi: 10.1108/07363760510605308.

[58] A. A. Skeie, ‘Brand and spokesperson congruence in product advertising : How the level of congruence affects the consumers brand attitude and purchase intention, and how brand involvement moderates this relationship’, Master thesis, 2015. Accessed: Jul. 10, 2024. [Online]. Available: https://openarchive.usn.no/usn-xmlui/handle/11250/2386964

[59] N. Hotz, ‘Agile Data Science’, Data Science Process Alliance. Accessed: Jun. 27, 2024. [Online]. Available: https://www.datascience-pm.com/agile-data-science/

[60] A. VK, ‘How The CRISP-DM Method Can Help Manage Your Next Data Science Project’, AIM. Accessed: Jun. 24, 2024. [Online]. Available: https://analyticsindiamag.com/crisp-dm-data-science-project/

[61] N. Hotz, ‘What is CRISP DM?’, Data Science Process Alliance. Accessed: Jun. 24, 2024. [Online]. Available: https://www.datascience-pm.com/crisp-dm-2/

[62] ‘Dublin, Ireland Weather History | Weather Underground’. Accessed: Jun. 12, 2024. [Online]. Available: https://www.wunderground.com/history/monthly/ie/dublin/EIDW/date/2024-5

[63] ‘Vapor Pressure Calculation by Antoine Equation (ethanol)’. Accessed: Jun. 24, 2024. [Online]. Available: http://ddbonline.ddbst.com/antoinecalculation/antoinecalculationcgi.exe?component=ethanol

[64] ‘antoine equation pressure - Google Search’. Accessed: Jun. 24, 2024. [Online]. Available: https://www.google.com/search?client=firefox-b-d&sca\_esv=0119435a3a07c38d&sca\_upv=1&sxsrf=ADLYWIIml5Gb8XVVAXTEBcec1wPujE8xbg:1719236994767&q=antoine+equation+pressure&tbm=vid&source=lnms&fbs=AEQNm0Aa4sjWe7Rqy32pFwRj0UkWd8nbOJfsBGGB5IQQO6L3J\_86uWOeqwdnV0yaSF-x2joQcoZ-0Q2Udkt2zEybT7Hdf5FBKg6QdtJ\_mF8k5Wx\_fK47VnnH0hqC26evHUklzukhRGDp8I9R6sObxD5rXV9iKTeMw0OsLcinUtCw7tu06Z-vfaM&sa=X&ved=2ahUKEwjEj-7ysPSGAxV4XUEAHU2jBk0Q0pQJegQIExAB&biw=766&bih=701&dpr=1.25#fpstate=ive&vld=cid:0d09f81a,vid:HxByfAVpYZk,st:0

[65] ‘Humidex’. Accessed: Jun. 18, 2024. [Online]. Available: https://www.chemeurope.com/en/encyclopedia/Humidex.html

[66] ‘Machine Learning with R : Expert Techniques for Predictive Modeling’. Accessed: Aug. 20, 2024. [Online]. Available: https://eds.p.ebscohost.com/eds/ebookviewer/ebook/bmxlYmtfXzIxMDYzMDRfX0FO0?sid=da3167b8-5484-4456-963e-5851103c3269@redis&vid=2&format=EB&rid=4

[67] ‘Introduction to Random Forest in R’, Simplilearn.com. Accessed: Aug. 20, 2024. [Online]. Available: https://www.simplilearn.com/tutorials/data-science-tutorial/random-forest-in-r

[68] ‘Multiple Linear Regression in R: Tutorial With Examples’. Accessed: Aug. 20, 2024. [Online]. Available: https://www.datacamp.com/tutorial/multiple-linear-regression-r-tutorial

[69] X. Wang, D. Lin, W. Fan, and T. Wang, ‘Research on Sales Forecast of Fresh Produce Considering Weather Factors’, 2018.

[70] S. Raizada and J. R. Saini, ‘Comparative Analysis of Supervised Machine Learning Techniques for Sales Forecasting’, *IJACSA*, vol. 12, no. 11, 2021, doi: 10.14569/IJACSA.2021.0121112.

[71] ‘What are the advantages and disadvantages of data visualization? | Tableau’. Accessed: Jul. 29, 2024. [Online]. Available: https://www.tableau.com/data-insights/data-visualization/advantages-disadvantages

[72] ‘What are potential limitations of correlational research methods? | TutorChase’. Accessed: Jul. 29, 2024. [Online]. Available: https://www.tutorchase.com/answers/ib/psychology/what-are-potential-limitations-of-correlational-research-methods

[73] R. J. Janse *et al.*, ‘Conducting correlation analysis: important limitations and pitfalls’, *Clinical Kidney Journal*, vol. 14, no. 11, pp. 2332–2337, Nov. 2021, doi: 10.1093/ckj/sfab085.

[74] ‘The Disadvantages of Linear Regression’, Sciencing. Accessed: Jul. 29, 2024. [Online]. Available: https://sciencing.com/disadvantages-linear-regression-8562780.html

[75] ‘Assumptions of Multiple Linear Regression’, Statistics Solutions. Accessed: Jul. 29, 2024. [Online]. Available: https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/assumptions-of-multiple-linear-regression/

[76] S. Allan, ‘Does Hot Weather Affect Digestion?’, Canadian Digestive Health Foundation. Accessed: Aug. 15, 2024. [Online]. Available: https://cdhf.ca/en/does-hot-weather-affect-digestion/

[77] G. Zheng, K. Li, and Y. Wang, ‘The Effects of High-Temperature Weather on Human Sleep Quality and Appetite’, *International Journal of Environmental Research and Public Health*, vol. 16, no. 2, Art. no. 2, Jan. 2019, doi: 10.3390/ijerph16020270.

[78] V. Melnyk, F. A. Carrillat, and V. Melnyk, ‘The Influence of Social Norms on Consumer Behavior: A Meta-Analysis’, *Journal of Marketing*, vol. 86, no. 3, pp. 98–120, May 2022, doi: 10.1177/00222429211029199.

[79] ‘Changing consumer preferences and food consumption patterns’. Accessed: Aug. 23, 2024. [Online]. Available: https://openknowledge.fao.org/server/api/core/bitstreams/0aa558d4-57c7-498d-87f7-b9e37577882f/content/src/html/changing-consumer-preferences-and-food-consumption-patterns.html

[80] ‘McDonald’s Says McPlant Trial Was Unsuccessful in the US, Meanwhile Ongoing Expansion Continues in Europe - vegconomist - the vegan business magazine’. Accessed: Aug. 23, 2024. [Online]. Available: https://vegconomist.com/gastronomy-food-service/fast-food/mcdonalds-us-mcplant-trial-not-successful/

[81] ‘Gibbs’ Reflective Cycle’, The University of Edinburgh. Accessed: Aug. 23, 2024. [Online]. Available: https://www.ed.ac.uk/reflection/reflectors-toolkit/reflecting-on-experience/gibbs-reflective-cycle

[82] ‘Historical Data - Met Éireann - The Irish Meteorological Service’. Accessed: Dec. 06, 2023. [Online]. Available: https://www.met.ie/climate/available-data/historical-data

[83] ‘Diabetes Prediction using R’, GeeksforGeeks. Accessed: Aug. 25, 2024. [Online]. Available: https://www.geeksforgeeks.org/diabetes-prediction-using-r/

[84] Y. Holtz, ‘Dual Y axis with R and ggplot2’. Accessed: Aug. 25, 2024. [Online]. Available: https://www.r-graph-gallery.com/line-chart-dual-Y-axis-ggplot2.html

[85] stikpet, *R - Dual Axis Bar Chart with ggplot*, (Jul. 03, 2022). Accessed: Aug. 25, 2024. [Online Video]. Available: https://www.youtube.com/watch?v=EhK9K9jwsco

[86] B. Lantz, *Machine learning with R. [electronic resource]*. Packt Publishing, 2013.

# Appendices

Table 8‑8‑1; Daily\_Data Sheet 1 Data Dictionary [82]

|  |  |
| --- | --- |
| **Variable** | **Description** |
| date | Date of observation (format; DD-MM-YY) |
| day | Day of the month |
| month | Month of the year |
| maxtp | Maximum air temperature (°C) |
| mintp | Minimum air temperature (°C) |
| avgtp | Average temperature (°C) |
| gmin | 09 UTC grass minimum temperature (°C) |
| rain | Precipitation amount (mm) |
| cbl | Mean cloud base level pressure (hPa) |
| wdsp | Mean wind speed (kt) |
| hm | Highest ten-minute mean wind speed (kt) |
| ddhm | Wind direction at max ten-minute mean (deg) |
| hg | Highest gust (kt) |
| soil | Mean 10 cm soil temperature (°C) |
| pe | Potential evapotranspiration (mm) |
| evap | Evaporation rate (mm) |
| smd\_wd | Soil moisture deficit on a well-drained day (mm) |
| smd\_md | Soil moisture deficit on a moderately-drained day (mm) |
| smd\_pd | Soil moisture deficit on a poorly-drained day (mm) |
| sun | Sunshine duration (hours) |
| glorad | Global radiation (J/cm²) |
| dp | Average dew point (°C) |
| svp | Saturation vapour pressure (hPa) |
| avp | Actual vapour pressure (hPa) |
| vp | Vapour pressure (hPa) |
| rh | Relative humidity (%) |
| humidex | Humidity index |
| Trans\_Count | Transaction count |
| Net\_Total | Net total amount |
| Final\_Total | Final daily total amount for Restaurant #1 |
| Avg\_Amt | Average amount spent per transaction |
| rain\_class | Rain classification (e.g., No-Rainfall, Light) |
| rain\_category | Rain category (e.g., Non-Rainy, Storm) |
| humidex\_class | Humidex classification (e.g., No discomfort) |

Table 8‑8‑2; Daily\_Data Sheet 2 Data Dictionary [66]

|  |  |
| --- | --- |
| **Variable** | **Description** |
| date | Date of observation (format; DD-MM-YY) |
| day | Day of the month |
| month | Month of the year |
| maxtp | Maximum air temperature (°C) |
| mintp | Minimum air temperature (°C) |
| avgtp | Average temperature (°C) |
| gmin | 09 UTC grass minimum temperature (°C) |
| rain | Precipitation amount (mm) |
| cbl | Mean cloud base level pressure (hPa) |
| wdsp | Mean wind speed (kt) |
| hm | Highest ten-minute mean wind speed (kt) |
| ddhm | Wind direction at max ten-minute mean (deg) |
| hg | Highest gust (kt) |
| soil | Mean 10 cm soil temperature (°C) |
| pe | Potential evapotranspiration (mm) |
| evap | Evaporation rate (mm) |
| smd\_wd | Soil moisture deficit on a well-drained day (mm) |
| smd\_md | Soil moisture deficit on a moderately-drained day (mm) |
| smd\_pd | Soil moisture deficit on a poorly-drained day (mm) |
| sun | Sunshine duration (hours) |
| glorad | Global radiation (J/cm²) |
| dp | Average dew point (°C) |
| svp | Saturation vapour pressure (hPa) |
| avp | Actual vapour pressure (hPa) |
| vp | Vapour pressure (hPa) |
| rh | Relative humidity (%) |
| humidex | Humidity index |
| Shop\_Total | Daily sales for Restaurant #2 |
| rain\_class | Rain classification (e.g., No-Rainfall, Light) |
| rain\_category | Rain category (e.g., Non-Rainy, Storm) |
| hum\_class | Humidex classification (e.g., No discomfort) |

Table 8‑8‑3; Product\_Category \_Sales Data Dictionary

|  |  |
| --- | --- |
| **Variable** | **Description** |
| date | Date of observation (format; MM/DD/YYYY) |
| Beverage | Sales amount for beverages |
| Breakfast | Sales amount for breakfast items |
| Burger | Sales amount for burgers |
| Chicken | Sales amount for chicken items |
| Combo Meal | Sales amount for combo meals |
| Fish | Sales amount for fish items |
| Fries | Sales amount for fries |
| Ice-Cream | Sales amount for ice-cream |
| Kebabs | Sales amount for kebabs |
| Kids Meal | Sales amount for kids meals |
| Milkshake | Sales amount for milkshakes |
| Other | Sales amount for other items |
| Pizza | Sales amount for pizzas |
| Portions | Sales amount for portions |
| sun | Sunshine duration (hours) |
| rain | Precipitation amount (mm) |
| evap | Evaporation rate (mm) |
| rh | Relative humidity (%) |
| maxtp | Maximum air temperature (°C) |
| mintp | Minimum air temperature (°C) |
| avgtp | Average temperature (°C) |

Table 8‑8‑4; Product Categories & Weather Correlation Matrix

|  |
| --- |
| Beverage Burger Chicken Combo\_Meal Fish Fries Ice\_Cream Kids\_  Meal  Beverage 1.00 0.80 0.86 0.95 0.34 0.91 0.59 0.87  Burger 0.80 1.00 0.82 0.80 0.28 0.83 0.43 0.72  Chicken 0.86 0.82 1.00 0.86 0.34 0.89 0.52 0.79  Combo\_Meal 0.95 0.80 0.86 1.00 0.32 0.91 0.60 0.83  Fish 0.34 0.28 0.34 0.32 1.00 0.40 0.13 0.29  Fries 0.91 0.83 0.89 0.91 0.40 1.00 0.52 0.82  Ice\_Cream 0.59 0.43 0.52 0.60 0.13 0.52 1.00 0.54  Kids\_Meal 0.87 0.72 0.79 0.83 0.29 0.82 0.54 1.00  Milkshake 0.74 0.67 0.73 0.73 0.30 0.73 0.49 0.69  Other 0.80 0.78 0.84 0.86 0.30 0.86 0.48 0.75  Pizza 0.80 0.72 0.79 0.81 0.33 0.82 0.54 0.73  Portions 0.65 0.60 0.65 0.66 0.28 0.66 0.37 0.58  sun 0.00 -0.05 -0.01 -0.04 0.03 -0.05 0.20 0.02  rain 0.02 0.03 0.03 0.05 -0.04 0.06 -0.04 0.02  evap -0.02 -0.10 -0.02 -0.08 0.03 -0.10 0.30 0.03  rh -0.01 0.06 0.02 0.05 -0.04 0.04 -0.11 -0.03  maxtp -0.03 -0.10 -0.03 -0.10 0.02 -0.11 0.23 0.01  mintp 0.01 -0.05 0.01 -0.04 0.00 -0.05 0.15 0.04  avgtp -0.01 -0.08 -0.01 -0.07 0.01 -0.08 0.18 0.02  wdsp 0.04 0.08 0.02 0.05 -0.03 0.08 -0.08 0.04  Milkshake Other Pizza Portions sun rain evap rh maxtp mintp  Beverage 0.74 0.80 0.80 0.65 0.00 0.02 -0.02 -0.01 -0.03 0.01  Burger 0.67 0.78 0.72 0.60 -0.05 0.03 -0.10 0.06 -0.10 -0.05  Chicken 0.73 0.84 0.79 0.65 -0.01 0.03 -0.02 0.02 -0.03 0.01  Combo\_Meal 0.73 0.86 0.81 0.66 -0.04 0.05 -0.08 0.05 -0.10 -0.04  Fish 0.30 0.30 0.33 0.28 0.03 -0.04 0.03 -0.04 0.02 0.00  Fries 0.73 0.86 0.82 0.66 -0.05 0.06 -0.10 0.04 -0.11 -0.05  Ice\_Cream 0.49 0.48 0.54 0.37 0.20 -0.04 0.30 -0.11 0.23 0.15  Kids\_Meal 0.69 0.75 0.73 0.58 0.02 0.02 0.03 -0.03 0.01 0.04  Milkshake 1.00 0.67 0.67 0.54 0.07 -0.01 0.10 -0.05 0.10 0.09  Other 0.67 1.00 0.80 0.65 -0.08 0.10 -0.13 0.12 -0.14 -0.06  Pizza 0.67 0.80 1.00 0.60 -0.02 0.03 -0.08 0.05 -0.07 -0.04  Portions 0.54 0.65 0.60 1.00 -0.08 0.04 -0.10 0.07 -0.07 0.00  sun 0.07 -0.08 -0.02 -0.08 1.00 -0.24 0.65 -0.48 0.44 0.05  rain -0.01 0.10 0.03 0.04 -0.24 1.00 -0.17 0.22 -0.04 0.10  evap 0.10 -0.13 -0.08 -0.10 0.65 -0.17 1.00 -0.51 0.79 0.50  rh -0.05 0.12 0.05 0.07 -0.48 0.22 -0.51 1.00 -0.25 -0.04  maxtp 0.10 -0.14 -0.07 -0.07 0.44 -0.04 0.79 -0.25 1.00 0.79  mintp 0.09 -0.06 -0.04 0.00 0.05 0.10 0.50 -0.04 0.79 1.00  avgtp 0.08 -0.11 -0.06 -0.04 0.23 0.06 0.63 -0.20 0.93 0.91  wdsp 0.02 0.11 0.04 0.04 -0.15 0.33 -0.14 -0.15 -0.14 0.07  avgtp wdsp  Beverage -0.01 0.04  Burger -0.08 0.08  Chicken -0.01 0.02  Combo\_Meal -0.07 0.05  Fish 0.01 -0.03  Fries -0.08 0.08  Ice\_Cream 0.18 -0.08  Kids\_Meal 0.02 0.04  Milkshake 0.08 0.02  Other -0.11 0.11  Pizza -0.06 0.04  Portions -0.04 0.04  sun 0.23 -0.15  rain 0.06 0.33  evap 0.63 -0.14  rh -0.20 -0.15  maxtp 0.93 -0.14  mintp 0.91 0.07  avgtp 1.00 0.00  wdsp 0.00 1.00 |

**Rcode for Model Development & Data Visualisation**

library(dplyr)

library(ggplot2)

library(randomForest)

library(caret)

library(tidyr)

library(readxl)

library(reshape2)

combined\_data <- read\_excel("~/EK\_Predictions\_RStudio/Data\_Files/Comparison\_Daily\_Sales.xlsx",

sheet=3, skip=31)

names(combined\_data)

#View(combined\_data)

str(combined\_data)

combined\_data$date <- as.Date(combined\_data$date)

combined\_data <- combined\_data %>%

filter(Final\_Total != 0)

combined\_data <- na.omit(combined\_data)

dim(combined\_data) #2220,30

#-------------------------------------------- summary statistics

summary\_stats <- combined\_data %>%

group\_by(restaurant\_id) %>%

summarise(

Final\_Total\_mean = mean(Final\_Total, na.rm = TRUE),

Final\_Total\_median = median(Final\_Total, na.rm = TRUE),

Final\_Total\_sd = sd(Final\_Total, na.rm = TRUE)

)

print(summary\_stats)

#------------------------------------------------------ line plot visuals pg 24 [83]

#combined\_data$restaurant\_id <- as.factor(combined\_data$restaurant\_id)

ggplot(combined\_data, aes(x = restaurant\_id, y = Final\_Total, fill = restaurant\_id)) +

geom\_boxplot(alpha = 0.7) +

labs(title = "Distribution of Total Sales by Fast-Service Restaurant",

x = "Restaurant ID",

y = "Total Sales") +

theme\_minimal()

#------------------------------------------------------------ random forest pg 39

set.seed(100)

train <- sample(nrow(combined\_data), 0.8\*nrow(combined\_data),

replace = FALSE)

data\_train <- combined\_data[train,]

data\_test <- combined\_data[-train,]

names(combined\_data)

dim(data\_train)

head(data\_train)

dim(data\_test)

head(data\_test)

#--------------------------------------------------------Model 1 - allmodel

allmodel <- randomForest(Final\_Total~.,ntree = 500, importance = TRUE,

data= data\_train)

allmodel #83.96%

importance(allmodel)

varImpPlot(allmodel) #rest\_id, day, date, glorad, month, cbl, wdsp

pred\_test <- predict(allmodel, newdata = data\_test)

pred\_test

mae <- mean(abs(pred\_test - data\_test$Final\_Total))

rmse <- sqrt(mean((pred\_test - data\_test$Final\_Total)^2))

print(paste("MAE:", mae)) #"MAE: "MAE: 623.62"

print(paste("RMSE:", rmse)) #RMSE: 964.98"

plot(data\_test$Final\_Total, pred\_test,

xlab = "Actual", ylab = "Predicted",

main = "Actual vs Predicted Sales")

abline(0, 1, col = "red")

#------------------------------------------------------- Model 2 tunedallmodel

optimal\_mtry <- best\_mtry[which.min(best\_mtry[, 2]), 1]

tunedallmodel <- randomForest( Final\_Total ~ ., mtry = optimal\_mtry, ntree = 500,

importance = TRUE, data = data\_train,)

tunedallmodel #86.21%

importance(tunedallmodel)

varImpPlot(tunedallmodel) #rest\_id, day, date, month, cbl, glorad, wdsp

pred\_test <- predict(tunedallmodel, newdata = data\_test)

pred\_test

mae <- mean(abs(pred\_test - data\_test$Final\_Total))

rmse <- sqrt(mean((pred\_test - data\_test$Final\_Total)^2))

print(paste("MAE:", mae)) #"MAE: 536.96"

print(paste("RMSE:", rmse)) #"RMSE: 909.4"

plot(data\_test$Final\_Total, pred\_test,

xlab = "Actual", ylab = "Predicted",

main = "Actual vs Predicted Sales")

abline(0, 1, col = "red")

#------------------------------------------------ Model 3 climate model

names(combined\_data)

climatemodel <- randomForest(

Final\_Total ~ maxtp + mintp + avgtp + gmin + rain +

wdsp + sun + rh + humidex +month +day +date +restaurant\_id,

data = data\_train, ntree = 500, importance = TRUE)

climatemodel #84.52%

importance(climatemodel)

varImpPlot(climatemodel) #rest\_id, day, date, wdsp, month, rh, maxtp

pred\_test <- predict(climatemodel, newdata = data\_test)

pred\_test

mae <- mean(abs(pred\_test - data\_test$Final\_Total))

rmse <- sqrt(mean((pred\_test - data\_test$Final\_Total)^2))

print(paste("MAE:", mae)) #[1] "MAE: 627.53"

print(paste("RMSE:", rmse)) #"RMSE: 968.47"

plot(data\_test$Final\_Total, pred\_test,

xlab = "Actual", ylab = "Predicted",

main = "Actual vs Predicted Sales")

abline(0, 1, col = "red")

#-----------------------------------------------tunedclimatemodel

optimal\_mtry <- best\_mtry[which.min(best\_mtry[, 2]), 1]

tunedclimatemodel <- randomForest(

Final\_Total ~ maxtp + mintp + avgtp + gmin + rain +

wdsp + sun + rh + humidex +month +day +date +restaurant\_id,

mtry = optimal\_mtry, ntree = 500, importance = TRUE, data = data\_train)

tunedclimatemodel #87.42

importance(tunedclimatemodel)

varImpPlot(tunedclimatemodel) #rest\_id, day, date, month, wdsp, rh, rain

pred\_test <- predict(tunedclimatemodel, newdata = data\_test)

pred\_test

mae <- mean(abs(pred\_test - data\_test$Final\_Total))

rmse <- sqrt(mean((pred\_test - data\_test$Final\_Total)^2))

print(paste("MAE:", mae)) #[1] "MAE: 529.33"

print(paste("RMSE:", rmse)) #"RMSE: 904.65"

plot(data\_test$Final\_Total, pred\_test,

xlab = "Actual", ylab = "Predicted",

main = "Actual vs Predicted Sales")

abline(0, 1, col = "red")

#----------------------------------------

data\_test$pred\_allmodel <- predict(allmodel, newdata = data\_test)

data\_test$pred\_tunedallmodel <- predict(tunedallmodel, newdata = data\_test)

data\_test$pred\_climatemodel <- predict(climatemodel, newdata = data\_test)

data\_test$pred\_tunedclimatemodel <- predict(tunedclimatemodel, newdata = data\_test)

predictions <- melt(data\_test, id.vars = "Final\_Total",

measure.vars = c("pred\_allmodel", "pred\_tunedallmodel",

"pred\_climatemodel", "pred\_tunedclimatemodel"))

#---------------------------------------------------- Actual vs Predicted plot pg 41

ggplot(predictions, aes(x = Final\_Total, y = value, colour = variable)) +

geom\_point(alpha = 0.5, size = 0.7) +

geom\_abline(slope = 1, intercept = 0, colour = "red", linetype = "dashed") +

labs(title = "Actual vs Predicted Sales",

x = "Actual Sales",

y = "Predicted Sales",

colour = "Model") +

theme\_minimal() +

theme(legend.position = "bottom")

#---------------------------------- Average Sales Bar Chart Weekdays/Ends pg 59

combined\_data$DayType <- ifelse(combined\_data$day %in%

c('Saturday', 'Sunday'), 'Weekend', 'Weekday')

average\_sales <- combined\_data %>%

group\_by(DayType) %>%

summarise(AverageSales = mean(Final\_Total, na.rm=TRUE))

ggplot(average\_sales, aes(x=DayType, y=AverageSales, fill=DayType)) +

geom\_bar(stat='identity') +

labs(title='Average Sales: Weekdays vs Weekends', x='Day Type', y='Average Sales') +

theme\_minimal()

#----------------------------------------------seasonal dual axis with total/average sales pg 60 [84], [85]

get\_season <- function(month) {

if (month %in% c('Mar', 'Apr', 'May')) {

return('Spring')

} else if (month %in% c('Jun', 'Jul', 'Aug')) {

return('Summer')

} else if (month %in% c('Sep', 'Oct', 'Nov')) {

return('Autumn')

} else {

return('Winter')

}

}

combined\_data$Season <- sapply(combined\_data$month, get\_season)

seasonal\_summary <- combined\_data %>%

group\_by(Season, restaurant\_id) %>%

summarise(

TotalSales = sum(Final\_Total, na.rm=TRUE),

AverageSales = mean(Final\_Total, na.rm=TRUE)

) %>%

ungroup() %>%

mutate(Season = factor(Season, levels = c('Spring', 'Summer', 'Autumn', 'Winter')))

ggplot(seasonal\_summary, aes(x=Season, y=TotalSales, fill=factor(restaurant\_id))) +

geom\_bar(stat='identity', position='dodge', alpha=.6) +

geom\_line(aes(y=AverageSales \* 250, group=factor(restaurant\_id),

colour=factor(restaurant\_id)), size=1, linetype="dashed") +

geom\_point(aes(y=AverageSales \* 250, colour=factor(restaurant\_id)), size=4) +

scale\_y\_continuous(

name = "Total Sales",

sec.axis = sec\_axis(~./250, name = "Average Sales")

) +

labs(title="Total and Average Sales by Season and Restaurant", x="Season") +

theme\_minimal() +

theme(legend.position = "bottom"

)

#---------------------------------------------------- wkend/wkday barchart avg sales pg 59

average\_sales <- combined\_data %>%

group\_by(restaurant\_id, DayType) %>%

summarise(avg\_sales = mean(Final\_Total, na.rm = TRUE),

.groups = 'drop')

# average\_sales$restaurant\_id <- as.factor(average\_sales$restaurant\_id)

ggplot(average\_sales, aes(x = DayType, y = avg\_sales, fill = DayType)) +

geom\_bar(stat = "identity", position = "dodge") +

facet\_wrap(~ restaurant\_id) +

labs(title = "Average Sales by Day Type for Each Restaurant",

x = "Day Type",

y = "Average Sales",

fill = "Day Type") +

theme\_minimal()

#--------------------------------------------------- Correlation Matrix & Heat graph pg 43 [83]

library(corrplot)

colnames(product\_data)

product\_data\_numeric <- product\_data %>%

select(where(is.numeric))

#product\_data\_numeric <- na.omit(numeric\_data)

cor\_matrix <- cor(product\_data\_numeric, use = "complete.obs")

cor\_matrix<- round(cor\_matrix,2)

cor\_matrix

cor\_matrix <- cor(product\_data\_numeric)

cor\_matrix\_df <- as.data.frame(as.table(cor\_matrix))

#----

ggplot(data = cor\_matrix\_df, aes(x = Var1, y = Var2, fill = Freq)) +

geom\_tile(colour = "white") +

scale\_fill\_gradient2(low = "blue", high = "red", mid = "white",

midpoint = 0, limit = c(-1, 1), space = "Lab",

name = "Correlation") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust = 1,

size = 12, hjust = 1),

axis.text.y = element\_text(size = 12)) +

coord\_fixed() +

labs(x = "", y = "")

#-------------------------------------------------------------------------------------- pairs () scatterplot matrix pg 50/51 [86]

pairs(product\_data\_numeric[c("Beverage", "sun", "wdsp", "rain", "evap", "rh", "maxtp","mintp","avgtp")])

pairs(product\_data\_numeric[c("Burger", "sun", "wdsp", "rain", "evap", "rh", "maxtp","mintp","avgtp")])

pairs(product\_data\_numeric[c("Chicken", "sun", "wdsp", "rain", "evap", "rh", "maxtp","mintp","avgtp")])

pairs(product\_data\_numeric[c("Combo\_Meal", "sun", "wdsp", "rain", "evap", "rh", "maxtp","mintp","avgtp")])

pairs(product\_data\_numeric[c("Fish", "sun", "wdsp", "rain", "evap", "rh", "maxtp","mintp","avgtp")])

pairs(product\_data\_numeric[c("Fries", "sun", "wdsp", "rain", "evap", "rh", "maxtp","mintp","avgtp")])

pairs(product\_data\_numeric[c("Ice\_Cream", "sun", "wdsp", "rain", "evap", "rh", "maxtp","mintp","avgtp")])

pairs(product\_data\_numeric[c("Kids\_Meal", "sun", "wdsp", "rain", "evap", "rh", "maxtp","mintp","avgtp")])

pairs(product\_data\_numeric[c("Milkshake", "sun", "wdsp", "rain", "evap", "rh", "maxtp","mintp","avgtp")])

pairs(product\_data\_numeric[c("Other", "sun", "wdsp", "rain", "evap", "rh", "maxtp","mintp","avgtp")])

pairs(product\_data\_numeric[c("Pizza", "sun", "wdsp", "rain", "evap", "rh", "maxtp","mintp","avgtp")])

pairs(product\_data\_numeric[c("Portions", "sun", "wdsp", "rain", "evap", "rh", "maxtp","mintp","avgtp")])

#------------------------------------------------------------------------------------------------- MAE, RMASE, Rsquared Dual Axis Bar Chart

accuracydf <- data.frame(

Model = c("All\_Model", "Tuned\_All\_Model", "Climate\_Model", "Tuned\_Climate\_Model"),

MAE = c(624.1952, 536.8051, 616.5841, 531.048),

RMSE = c(966.8824, 908.8939, 956.8538, 904.6512),

R\_Squared = c(83.96, 87.25, 84.52, 87.42)

)

accuracydf\_long <- accuracydf |>

pivot\_longer(cols = -Model, names\_to = "Metric", values\_to = "Value") |>

mutate(

Metric = factor(Metric, levels = c("RMSE", "MAE", "R\_Squared")),

Scaled\_Value = ifelse(Metric == "R\_Squared", Value, Value / 10)

)

ggplot(accuracydf\_long, aes(x = Model, y = Scaled\_Value, fill = Metric)) +

geom\_col(data = accuracydf\_long |> filter(Metric %in% c("RMSE", "MAE")),

aes(fill = Metric), position = "dodge") +

geom\_line(data = accuracydf\_long |> filter(Metric == "R\_Squared"),

aes(group = 1, colour = Metric), size = 1.2) +

geom\_point(data = accuracydf\_long |> filter(Metric == "R\_Squared"),

aes(colour = Metric), size = 3) +

scale\_y\_continuous(

name = "R-Squared (%)",

sec.axis = sec\_axis(~ . \* 10, name = "MAE / RMSE")

) +

labs(title = "Customer Behaviour Model Performance Metrics", x = "Model") +

theme\_minimal() +

scale\_fill\_manual(values = c("RMSE" = "red", "MAE" = "blue")) +

scale\_colour\_manual(values = c("R\_Squared" = "green")) +

theme(legend.position = "top")

#--------------------------------------------------------------------------------------------- regression models for products

product\_data <- read\_excel("~/EK\_Predictions\_RStudio/Data\_Files/Product\_Sales\_Transposed.xlsx", sheet = 1)

product\_data <- product\_data %>% filter(Beverage != 0)

#------------------------------------------------------------------------ Burger

burgermodel <- lm(Burger ~ sun + rain + evap + rh + avgtp + wdsp, data = product\_data)

summary(burgermodel)

vif(burgermodel)

#------------------------------------------------------------------------ Beverage

beveragemodel <- lm(Beverage ~ sun + rain + evap + rh + avgtp + wdsp, data = product\_data)

summary(beveragemodel)

vif(beveragemodel)

#------------------------------------------------------------------------ Chicken

chickenmodel <- lm(Chicken ~ sun + rain + evap + rh + avgtp + wdsp, data = product\_data)

summary(chickenmodel)

vif(chickenmodel)

#------------------------------------------------------------------------ Combo\_Meal

combomodel <- lm(Combo\_Meal ~ sun + rain + evap + rh + avgtp + wdsp, data = product\_data)

summary(combomodel)

vif(combomodel)

#------------------------------------------------------------------------ Fish

fishmodel <- lm(Fish ~ sun + rain + evap + rh + avgtp + wdsp, data = product\_data)

summary(fishmodel)

vif(fishmodel)

#------------------------------------------------------------------------ Chips

friesmodel <- lm(Fries ~ sun + rain + evap + rh + avgtp + wdsp, data = product\_data)

summary(friesmodel)

vif(friesmodel)

#------------------------------------------------------------------------ Ice\_Cream

icecreammodel <- lm(Ice\_Cream ~ sun + rain + evap + rh + maxtp + mintp + avgtp + wdsp, data = product\_data)

summary(icecreammodel)

vif(icecreammodel)

#------------------------------------------------------------------------ Kids\_Meal

kidsmealmodel <- lm(Kids\_Meal ~ sun + rain + evap + rh + avgtp + wdsp, data = product\_data)

summary(kidsmealmodel)

vif(kidsmealmodel)

#------------------------------------------------------------------------ Milkshake

milkshakemodel <- lm(Milkshake ~ sun + rain + evap + rh + avgtp + wdsp, data = product\_data)

summary(milkshakemodel)

vif(milkshakemodel)

#------------------------------------------------------------------------ Other

othermodel <- lm(Other ~ sun + rain + evap + rh + avgtp + wdsp, data = product\_data)

summary(othermodel)

vif(othermodel)

#------------------------------------------------------------------------ Pizza

pizzamodel <- lm(Pizza ~ sun + rain + evap + rh + avgtp + wdsp, data = product\_data)

summary(pizzamodel)

vif(pizzamodel)

#------------------------------------------------------------------------ Portions

portionsmodel <- lm(Portions ~ sun + rain + evap + rh +avgtp + wdsp, data = product\_data)

summary(portionsmodel)

vif(portionsmodel)

#----------------------------------------------------------------------------------------------------- Holiday Sales Boxplot

paddys\_day <- as.Date(c("2021-03-17", "2022-03-17", "2023-03-17", "2024-03-17"))

good\_friday <- as.Date(c("2021-04-02", "2022-04-15", "2023-04-07", "2024-03-29"))

easter\_monday <- as.Date(c("2021-04-05", "2022-04-18", "2023-04-10", "2024-04-01"))

halloween <- as.Date(c("2021-10-31", "2022-10-31", "2023-10-31"))

christmas <- as.Date(c("2021-12-24", "2022-12-24", "2022-12-25", "2023-12-24"))

analyse\_holiday <- function(holiday\_dates, holiday\_name, data, category) {

holiday\_data <- data %>% filter(date %in% holiday\_dates)

non\_holiday\_data <- data %>% filter(!date %in% holiday\_dates)

plot <- ggplot() +

geom\_boxplot(data = holiday\_data, aes\_string(x = "'Holiday'", y = category, fill = "'Holiday'")) +

geom\_boxplot(data = non\_holiday\_data, aes\_string(x = "'Non-Holiday'", y = category, fill = "'Non-Holiday'")) +

labs(title = paste(category, "Sales on", holiday\_name, "vs Non-Holidays"),

x = "Day Type",

y = paste(category, "Sales")) +

theme\_minimal()

print(plot)

}

category <- "Milkshake"

analyse\_holiday(paddys\_day, "St.Patrick’s Day", product\_data, category)

analyse\_holiday(good\_friday, "Good Friday", product\_data, category)

analyse\_holiday(easter\_monday, "Easter Monday", product\_data, category)

analyse\_holiday(halloween, "Halloween", product\_data, category)

analyse\_holiday(christmas, "Christmas", product\_data, category)