UNIVERSITY OF WESTMINSTER

A study on the relationship between weather and online retail sales

Master Thesis Proposal – May 2018

[redacted]

MSc Business Intelligence and Analytics

Supervisor: Thierry Chaussalet

Table of contents

Introduction……3

Aims……4

Resources……4

Objectives……4

Background……5

Retail sector……6

Methodology……7

Software……7

On-site and off-site analysis……7

Gantt chart……8

Risk assessment……10

Expected outcomes……11

References……11

Appendix: annotated bibliography……12

Introduction

In a report published by the British Retail Consortium in April 2018, Paul Martin, Head of Retail at KPMG, affirms that:

"March was difficult for large parts of the UK retail industry. Seemingly endless **cold weather** dissuaded would-be shoppers from the high street and a number of retailers delivered bad news. […] Retailers with an online presence were far more fortunate, with a marked lift in all categories. The cold weather clearly persuaded shoppers to peruse from the comfort of their own homes, with *beauty* and *clothing* grabbing the most attention." (British Retail Consortium, 2018).

By looking at this affirmation, it results clear that weather affects different businesses in completely different ways.

It is known that consumer behaviour is affected by weather conditions. For example, the exposure to sunlight can improve people's mood and increase their willingness to pay (Murray et al., 2010). Does this mean that customers are more willing to buy goods offline? How does this weather condition affect online retail companies?

This analysis will be conducted in partnership with Shop Direct, UK's second largest online-only retailer. This project was one of the projects of the 'Masters Research Dissertation Programme 2018' made up by the Consumer Data Research Centre (CDRC). I applied for this program and I was selected by Shop Direct to collaborate with them on the study of the relationship between weather and sales of their biggest and fastest growing retail brand (*Very.co.uk*).

At the time of the poster presentation, I was not aware that this company had selected me to conduct this analysis. This is the reason why my poster was about another project (about diabetes), which I had to abandon later on in order to conduct this analysis.

Aims

The main aim of this research is to understand if and how weather significantly affects products demand when it comes to a full season (Spring/Summer, Fall/Winter) or a specific country. The analysis will be conducted on real-world data provided by Shop Direct, UK's second largest online-only retailer. Another purpose of this analysis is to suggest new approaches to improve the products stock levels' forecasting accuracy.

Resources

Shop Direct will provide me with various data sets in order to conduct this research. The data sets that will be made available at the end of May are made of:

- Product and sales information;
- Customer information, including some socio-demographic data and transactional information;
- Weather information: I will be provided with a data set of 5 years history of weather in 20 large cities in the UK (I will get access to the history of weather *actuals* only, not of its *forecasts*).

Objectives

To meet the project's aims, a set of measurable objectives has been set:

- Apply data cleaning and pre-processing tasks to real-world data. It will help me to understand how much this task could be challenging, and to learn new techniques to efficiently complete this task.

- Apply insights gained from the 'Data Visualisation and Dashboarding' module to real-world data: conduct an **exploratory data analysis** by plotting the data, with the aim of detecting the presence of outliers or mistakes in the data set, and to visually understand the underlying patterns in the data and the relationship between different variables. The results of this phase will determine how the whole analysis will be organised.

- Apply business analytics' principles to real-world data, and in particular:

    - Extract samples from the whole population to understand which variables are relevant to the analysis (should the analysis be split per countries, products or single season?). Learn how to apply 'simple random sampling' to a data set in R.
    - For each sample, divide the relative data set in training and test data sets in R. The accuracy of each model built on training sets will be checked by measuring its performance on test data sets by checking confusion matrices, accuracies values and ROC curves.
    - Analyse the relationship between weather and sales through **hypothesis testing**.
    - Conduct a **regression analysis** to understand if Shop Direct should take into consideration weather forecasts throughout its decision-making process, and to list which products among the ones sold by Very (the main brand of Shop Direct) may have a weather driven demand.

- Make suggestions to improve Shop Direct's forecasting accuracy. Apply academic and professional marketing experience to set some advertising strategies tailored per segment of customers or per products, according to the results obtained by the whole analysis (weather-based marketing).

Background

Many researches have been conducted in fields such as finance or psychology, which proved that weather changes can affect human behaviour (Murray et al., 2010). It was only after a while that weather effects started being analysed on the retail sector as well. In fact, Xavier Brusset (2016) states that even if the relationship between weather conditions and sales demand has been known for a long time, managers still struggle in understanding its potential and they do not recognise the competitive advantage they could get by simply evaluating this relationship per each product sold. He also affirms that the only firms that are aware of the importance of this relationship are the ones which belong to the energy and agricultural sectors, which are known to be considerably affected by seasonal weather conditions. He also highlights that, with climate change, the connection between sales and unseasonal weather conditions has an increased need to be assessed: the more frequent weather swings are having a significant impact on sales.

It is evident that retail sales will always be influenced by these weather swings: for example, clothing sales will always be affected by an increase or decrease of temperature (Brusset, 2016). However, it is not possible anymore to blame the weather for drops in the total amount of sales. In fact, weather awareness in spreading in all companies (Elin, 2015): some firms are currently buying insurances in order to protect themselves against weather unpredictability (Brusset, 2016).

When studying the relationship between weather conditions and sales, it is necessary to distinguish between seasonal and unexpected weather conditions. The former type of weather conditions is common in every sector and it can be relatively easily forecasted from one year to another. On the other hand, unexpected weather conditions cannot be forecasted in any way and companies from any type of sectors should in fact focus their attention mainly on the study of the relationship between sales demand and this type of weather conditions.

Retail sector

This analysis will focus mainly on the retail sector and on how different retailers could deal with weather uncertainty.

Retailers are usually positioned at the end of the supply chain, and they are the ones with a higher ability to promptly react to situations (such as weather changes) that affect customers' behaviour. It is essential for them to forecast products demand in an accurate way because it affects the products' stock levels (Elin, 2015). A poorly accurate forecast could lead to a surplus of unsold items in the inventories or to a shortage of products, which represents lost revenue for any type of company.

The results of a study on the effects of weather on retail sales conducted by Starr-McCluer in 2000, show that unusual weather significantly affects monthly fluctuations in consumer spending. However, weather influence usually washes out at a quarterly frequency. This analysis, as the one conducted by Elin (2015), was based on offline retail data: the purpose of my research is in fact to reproduce the same analysis on a data set with online only retail information, to discover if the online retail sales show the same patterns as the offline one, and to understand how to address the different needs of online retailers.

Another analysis relevant for my project is a research published on 'The Journal of Retailing and Consumer Services' (Murray et al., 2010). The authors of this research conducted three different regression analysis which allowed them to prove that different weather variables do effect customer's willingness to pay. For example, sunlight was the variable which had the highest effect on customers' mood and therefore on customers' overall spending. They also proved that bad weather decreases customers' willingness to go out and shop in store. According to the authors, this is known to be one of the three general categories of effects that weather can have on consumer behaviour. This affirmation was also previously stated by Parsons (2001), who explained that bad weather, such as rain, snow and low temperatures, has a negative correlation with total sales and store traffic.

It is evident that weather conditions do influence retail sales, however very few researches have been conducted on **online-only retailers**. It will then be interesting to analyse if customers behaviour changes in regard to online shopping. Will bad weather lead to an increase of online retail sales, or will it still negatively affect customers' willingness to buy goods?

According to the amount of literature available, it could be stated that weather awareness is currently spreading to the retail sector as well, and it is not ignored in the marketing literature anymore. The effect of weather on consumer behaviour has been explained by marketing scholars in the past, but only at an aggregate level: this topic has then been analysed further by Moon et al. (2018) in a paper which analyses the impact that individual differences in weather sensitivity has on consumers' weather-related purchase decisions. This type of approach will be used also in my analysis, in order to identify any discrepancy between different segments of customers.

The increased importance attributed to weather is contributing to the creation of new fields that deal with weather variables. An example is 'weather-based marketing', a new discipline whose aim is to create weather responsive campaigns. Weather Unlocked, a website which proposes weather driven solutions for advertising, has made available an extensive guide about this topic (Weather Unlocked[1], no date). Companies are now able to gain competitive advantage and to optimise their revenue by setting weather ads, which are more effective than standard ads thanks to their correlation with actual

[1] http://www.weatherunlocked.com/resources/the-complete-guide-to-weather-based-marketing

weather conditions, or by selecting the more effective keywords thanks to Google AdWords' Weather-Based Management script.

Methodology

Software

I will be working mainly with R for the exploratory data analysis and for the regression analysis. I will use Excel for the pre-processing tasks, and Tableau to visualise data in the exploratory data analysis and at the end of the whole project to visually see its results and highlight the most important points of my analysis.

On-site and off-site analysis

The plan of execution of this analysis will follow how the various objectives have been previously structured and listed. In particular, the first part of the analysis will be conducted on-site working at Shop Direct's office in London.

1. The first phase (**on-site**) will start at the end of May and it will last approximately eight weeks. It will include the preparation task, the exploratory data analysis and the regression analysis. It involves the use of sensitive information, and this is the reason why this first phase cannot be completed off-site.

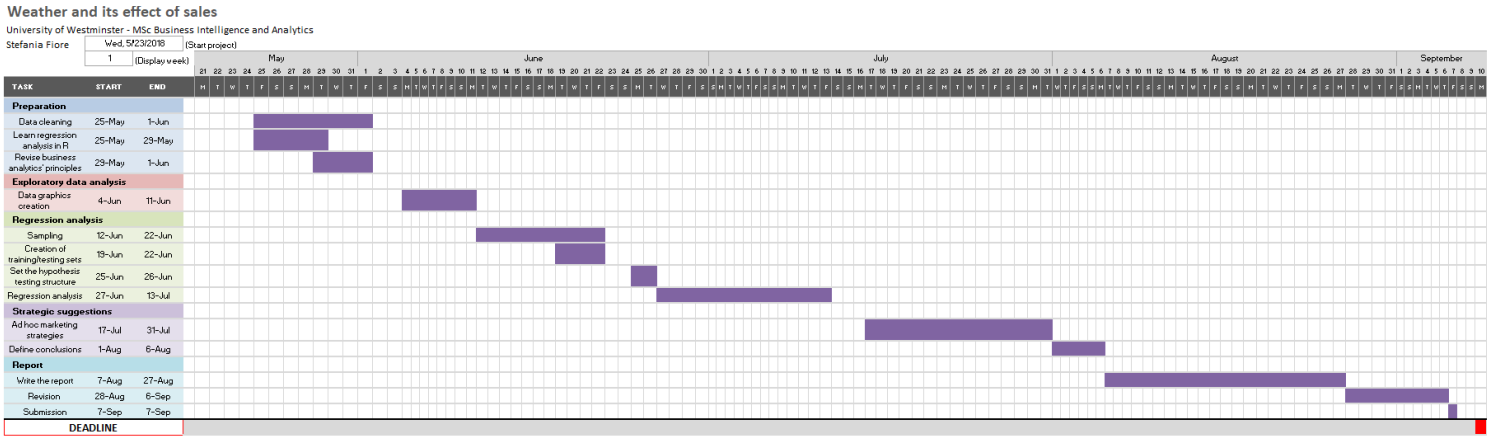
   To be more specific, this phase will start with the cleaning of the data: each data set will be uploaded to R and the presence of outliers will be checked by plotting boxplots for each variable available in the data set. Using appropriate functions in R, the variables will be divided between categorical and numeric; an exploratory data analysis will then be conducted by writing some codes in R, following both a univariate and multivariate approach (for example, a table of correlations will be created to easily understand how variables influence each other and the explanatory variable, which will be "sales"). The data will then be plotted by using Tableau, to quickly spot some underlying patterns in the data sets. Since the data will be a time series, line charts will be created to see its trend over the last 5 years.
   The way the regression analysis will be planned will be based on the results obtained from this first analysis; the regression analysis will be conducted in R as well.
   If time allows to do that, neural networks will be created in R in order to estimate sales values for the next years according to different weather conditions; this analysis will follow Taghizadeh's (2017) approach.

2. The second phase (**off-site**) will start in the middle of July, and it includes the creation of marketing strategies to propose to Shop Direct (based on the results of the previous phase) and the definition of the conclusion to the whole project. According to the results reached in the first phase of the project, some marketing strategies will be proposed, and they will be tailored per segment of customers. This phase will also include the writing and revision of the final report.

This plan of execution is shown in more details in the following Gantt chart.

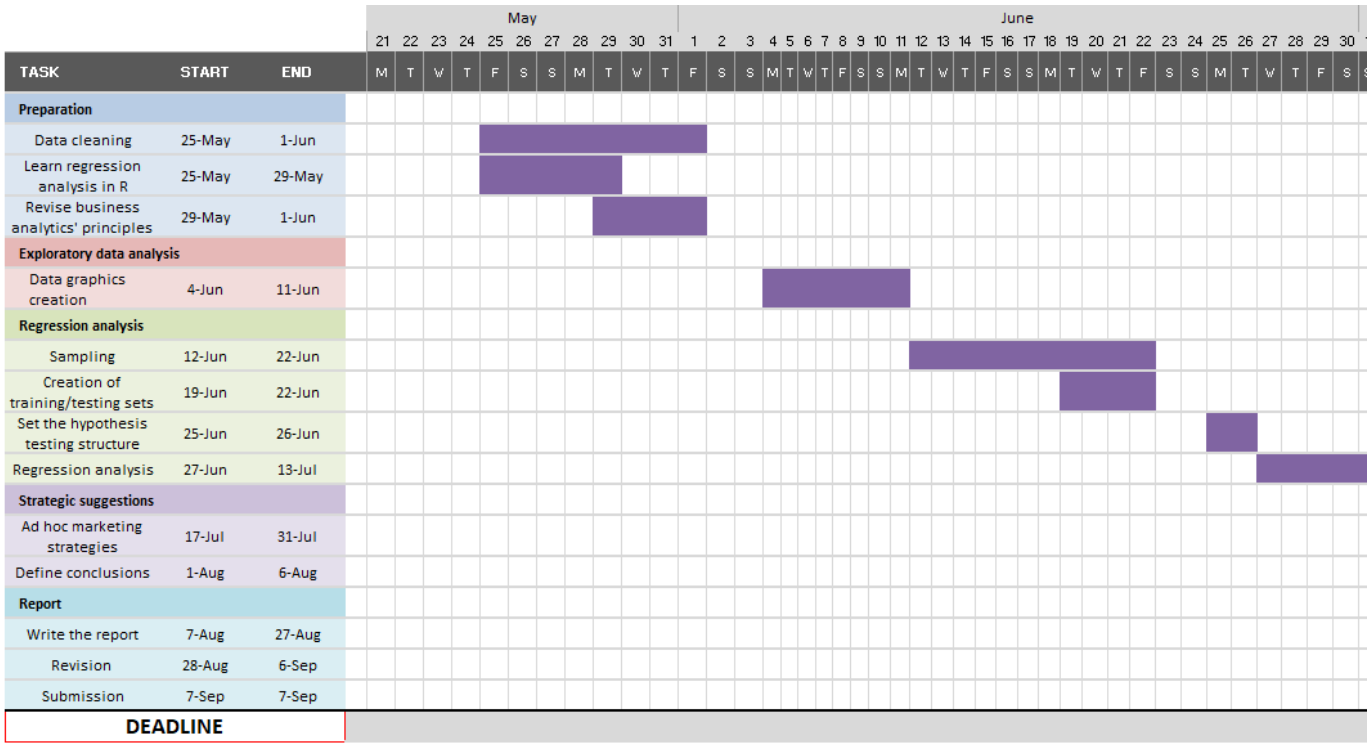
Gantt chart

The workload will start at the end of May and it will finish at the start of September. It is described thoroughly in the following Gantt chart, which includes the flow of each single task of the plan of execution mentioned above.

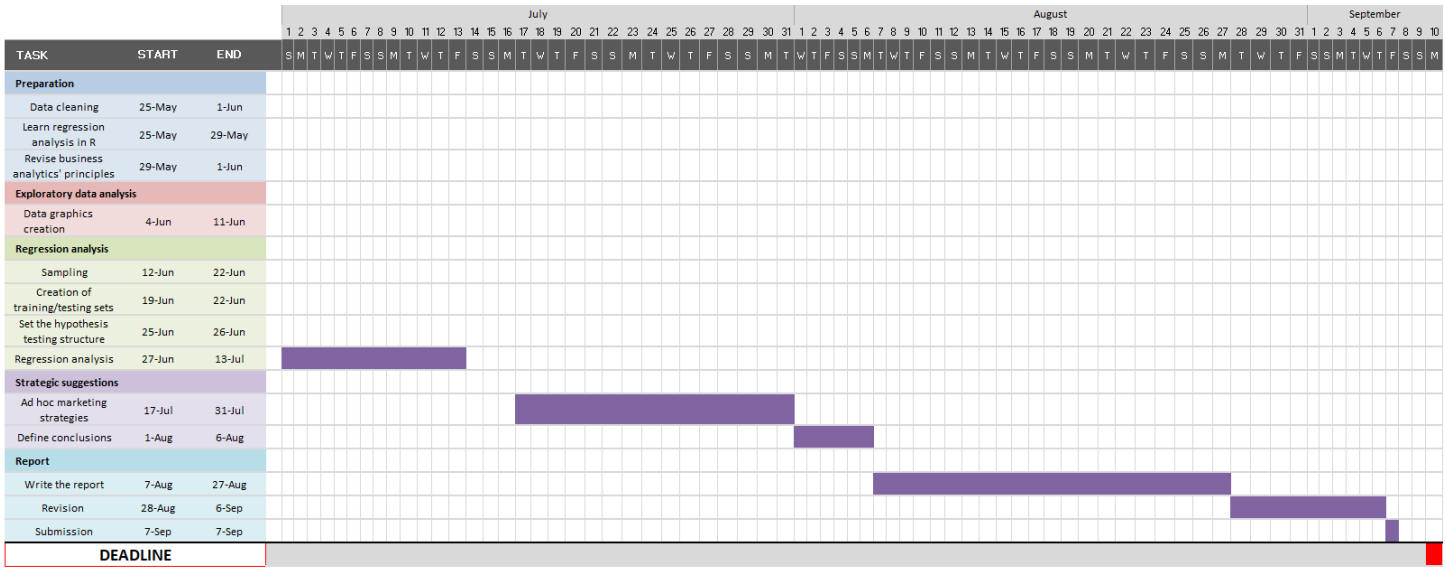


To simplify its understanding, the Gantt chart was divided into two charts showing different months:

- **May & June:**



- **July, August & September:**



Risk assessment

Knowing that a project will always involve some levels of uncertainty and risk, I have created a table in which all possible risks are listed along with their eventual solutions.

RISK	SOLUTION
Data quality The company I will be working for will provide me with the data needed to conduct this analysis, therefore I will not worry about collecting the data because it will come from a trusted source. However, it is possible that the data will present some mistakes (it has been collected over the past 5 years presumably from different employees). Also, weather data would probably come from a third-party source, and it could not be completely accurate.	I would select a random subset of the whole data set to check for its quality. I could also compare it with another random subset to check for data consistency. In regard to weather data, I could compare a subset to other external sources (such as Met Office[2]).
Data understanding Products and customer information should be straightforward; however, weather data will probably not be easy to understand due to a lack of knowledge in this area.	I would see what type of weather information was used in the researches mentioned in the literature review, and then I would try to get more information only about the weather variables which could be relevant to this analysis.
Data accessibility The first part of the analysis will be conducted on-site to facilitate the understanding of the data (an industry supervisor will guide me through this project). I will not be in London for one month, and I was informed that it is possible to access the data off-site, but I could have some troubles in doing that.	To avoid having troubles in accessing the data, I will work hard on the practical analysis meanwhile I am in London. I will also try to access the data set off-site while being in the city, in order to get help from my industry supervisor in case something does not work properly.
Regression analysis in R This will be the first time I conduct a regression analysis by writing codes in R, and it is possible that I will often incur into error messages and that I will have to solve them before being able to proceed with my analysis.	If I incur into many message errors, I will also conduct the regression analysis in Excel by using StatTools, an Excel's add-in that I have already used in the past. I will then compare the results obtained with both software to check for consistency.

[2] https://www.metoffice.gov.uk/public/weather/climate-historic/#?tab=climateHistoric

Expected outcomes

By conducting this analysis, I expect to discover the relationship between weather and customers' willingness to buy products sold on Very.co.uk, Shop Direct's main online-only retail brand.

To understand if the project has been a success or not, every model built in R will be tested on the test data sets created from the available historical data. The performance of each model will be evaluated by computing its MAE, MAPE and MSE. The best model will then be used by Shop Direct to forecast future products stock levels.

I expect to learn how to apply theoretical concepts studied during this academic year on real-world data, to overcome challenges related to the use of Big Data and to improve my consultancy and communication skills in regard to the relationship that I will build with the company I will work for.

References

British Retail Consortium (2018). *Roller coaster ride in run-up to Easter*. Available from: https://brc.org.uk/retail-insight-analytics/retail-sales-reports/retail-sales-monitor/reports

Brusset, X. (2016). *White paper on the impact of the weather on sales.*

Elín (2015). *Use of weather data in supply chain management*.

Moon, S., Kwon, J., Jung, S., Bae, Y.H. (2018). The impact of individual differences in weather sensitivity on weather-related purchase intentions. *International Journal of Market Research.* 60 (1), 104-117. Available from http://journals.sagepub.com/doi/full/10.1177/1470785317744855 .

Murray, K.B., Di Muro, F., Finn, A., Popkowski Leszczyc, P. (2010). The effect of weather on consumer spending. *Journal of Retailing and Consumer Services.* 17 (6), 512-520. Available from https://www.sciencedirect.com/science/article/pii/S0969698910000822 .

Parsons, A.G. (2001). The Association Between Daily Weather and Daily Shopping Patterns. *Australasian Marketing Journal (AMJ).* 9 (2), 78-84. Available from https://www.sciencedirect.com/science/article/pii/S1441358201701772 .

Starr-McCluer, M. (2000). *The effects of weather on retail sales.* Washington, DC: Board of Governors of the Federal Reserve System.

Taghizadeh, E. (2017). *Utilizing artificial neural networks to predict demand for weather-sensitive products at retail stores*. Available from http://arxiv.org/abs/1711.08325 .

Appendix: annotated bibliography

1. British Retail Consortium (2018). *Roller coaster ride in run-up to Easter*. Available from: https://brc.org.uk/retail-insight-analytics/retail-sales-reports/retail-sales-monitor/reports

Article produced by BRC and KPMG, which provides information about what happened in the retail sector in the last few months (for example, it explains if sales increased or not in comparison to the previous year).

2. Brusset, X. (2016). *White paper on the impact of the weather on sales.*

Generic paper which introduces the topic about the impact of weather on sales comparing various sectors and giving some examples on companies (such as Bosch) which are currently buying insurances to protect themselves against risks connected to weather.

3. Davies, J. and Elliott, D. (2015). Analysing the impact of weather and climate on official statistics time series. *Survey Methodology Bulletin.* 39.

This paper presents two case studies; the first one is about an investigation on the effects that temperature has on retail sales for clothing.

4. Elín (2015). *Use of weather data in supply chain management*.

MSc thesis submitted at the Reykiavik University (Iceland). The main topic is the analysis of the influence of weather in supply chain management, based on the case study of the leading Iceland retail and service company that provides fuel, supplies and refreshments. A regression analysis was done in order to understand the influence of weather, and the structure of this analysis is similar to the one that I am going to do for my project.

5. Gardner, M.P. (1985). Mood states and consumer behavior. *Journal of Consumer Research.* 12 (3), 281-300. Available from http://www.econis.eu/PPNSET?PPN=258978864 .

This paper reviews how mood states are known to affect consumer behaviour according to the psychological literature, and it discusses how mood-related approaches could be helpful for marketing research.

6. Moon, S., Kwon, J., Jung, S., Bae, Y.H. (2018). The impact of individual differences in weather sensitivity on weather-related purchase intentions. *International Journal of Market Research.* 60 (1), 104-117. Available from http://journals.sagepub.com/doi/full/10.1177/1470785317744855 .

This is an analysis made from a marketing point of view, which highlights how individual differences in weather sensitivity play a role in consumer's purchase decisions.

7. Murray, K.B., Di Muro, F., Finn, A., Popkowski Leszczyc, P. (2010). The effect of weather on consumer spending. *Journal of Retailing and Consumer Services.* 17 (6), 512-520. Available from https://www.sciencedirect.com/science/article/pii/S0969698910000822 .

Really good paper, which thoroughly describes three studies done by the authors with the aim of understanding the influence of specific weather variables on consumer's behaviour.

8. Parsons, A.G. (2001). The Association Between Daily Weather and Daily Shopping Patterns. *Australasian Marketing Journal (AMJ).* 9 (2), 78-84. Available from https://www.sciencedirect.com/science/article/pii/S1441358201701772 .

This paper is about a regression analysis done in 2001 to understand the correlation between daily changes of weather and daily variations in shopping patterns.

9. Regnier, E. (2008). Doing something about the weather. *Omega.* 36 (1), 22-32. Available from https://www.sciencedirect.com/science/article/pii/S0305048305001805 .

This article gives a general overview of the initial integration of weather forecasts in the decision-making process.

10. Starr-McCluer, M. (2000). *The effects of weather on retail sales.* Washington, DC: Board of Governors of the Federal Reserve System.

The aim of this paper was to understand if monthly fluctuations in consumer spending can be attributed to weather. It was discovered that weather could be a reason for the monthly variations in sales, but it was discovered that the effect of weather tends to disappear quarterly.

11. Steele, A.T. (1951). Weather's effect on the sales of a department store [in Des Moines, Iowa]. *Journal of Marketing.* 15 436-443. Available from.

Regression analysis on the effects of weather on retail sales of the case study taken under consideration.

12. Taghizadeh, E. (2017). Utilizing artificial neural networks to predict demand for weather-sensitive products at retail stores. Available from http://arxiv.org/abs/1711.08325 .

Interesting paper, which proposes a way to forecast demand for weather-sensitive products by building neural networks. I will probably follow the same approach for the last part of my research.

13. Weather Unlocked (no date). *How to Increase Advertising ROI using Weather.*

A guide created by 'Weather Unlocked', which explains how to increase advertising ROI taking advantage of weather forecasts.

14. Weather Unlocked (no date). *The Complete Guide to Weather Based Marketing.*

A guide created by 'Weather Unlocked', which introduces basic topics about weather-based marketing.