Customer Purchasing Habits Based on Weather Conditions

Phuoc Le, Phuc Le, Nirmit Jasapara, and Aidan Jones

Abstract–This research paper aims to investigate relationships between weather conditions and consumer purchasing behavior by running an unbiased general data analysis on the relationships between the two domains of data. Weather data, such as temperature, wind speed, and precipitation, is compared with purchasing history across various categories of products to determine if there exist buying patterns that are not yet apparent or exploited by the current market and quantify the extent of all relationships. By examining the correlation between weather conditions and sales data across various product categories, this research seeks to shed light on whether consumers tend to purchase more of a certain type of product in specific weather conditions and by how much.

For the exploratory analysis, a comprehensive dataset will be acquired by combining various data sources such as retail purchase records of various famous stores, resupply records and the corresponding historical weather data. The data will be comprehensive and diverse enough to see general trends that are not specific to any area or time. What this means is that the data will be collected from diverse geographical locations, consider seasonal variations and even changing trends across the years. Statistical techniques, such as regression analysis and correlation tests, will be used to find and quantify any significant relationships between weather conditions and consumer purchases.

I. Introduction

This research could have great implications for various retail businesses. It could help businesses resupply in a more granular manner, ensuring maximum profits while keeping costs down through smarter foreshadowing. Accurate supply prediction can also help to reduce waste by reducing access inventory and preventing unwanted products to be left in storage and eventually be thrown away. This also helps consumers by ensuring the products they need are always in stock and can be acquired through local channels that are faster than online channels.

II. Related Works

The relationship between weather and shopping has been a topic that researchers have been speculating for quite a while. There have been numerous research papers written that each tackle a specific domain of the theory. For example, Arunraj et al. from the Deggendorf Institute of Technology in Germany analyzed two food retail stores and a fashion retail store located in Lower Bavaria in order to find the impact of weather on store traffic and the sales [1]. Since the target geographical area is small, all data was gathered from local sources [1]. This study successfully showed that seasonal weather such as snowfall and rain had a big impact on fashion sales [1]. However, due to the limited nature of the study, it would be difficult to prove if this was a weather phenomenon or seasonal habits specific to the area the stores were located. A successor to this study was later created by Badorf et al. from the Kühne Logistics University. They built on Arunraj's research and conducted the study on 673 retail stores across Germany [3]. They realized that seasonal changes indeed impacted the relationship and that it needed to be taken into account for any model created to predict weather impact on sales [3]. Rose et al. also showed how much seasonal changes impacted the impact of weather on sales [5]. They also showed how much the change in region would affect how much impact weather had on sales [5].

Another study by Xin Tian et al. from University of Chinese Academy of Sciences in Beijing shed further light on the subject by doing a study on all of China. Unlike the study in Germany, only a single convenience store chain was picked, however, the chain had 146 stores all over China [2]. Their research showed that weather impact was a little more complex than realized. In "good" weather, people would make small purchases continuously, while "bad" weather prompted people to make huge bulk purchases [2]. Also, while the team showed some evidence on how different categories of products were impacted, due to the fact that only a single convenience chain was picked, the results were limited to food and beverages [2].

However, all previous research didn't focus on how different products were impacted by weather. Okayama et al. conducted research to take into account product categories for weather impact on sales [4]. They used nonnegative tensor factorization to create item clusters with high sensitivity to weather [4]. They showed that product type has a huge influence on how weather impacts it [4]. However, the research does not iterate which types of products are impacted, how they are impacted, when they are impacted or by how much.

All the research mentioned above shows that there is a big correlation between weather and sales, however, they all go into the problem with various assumptions. All the models used are heavily influenced by common sense such as people will buy more drinks in hot weather or that snow will cause people to buy more related fashion while slowing down other sales. However, none of these studies try to find out how much each category of product is truly impacted by looking at the problem with an unbiased lens.

III. Data

The customer purchase history dataset that was utilized for our analysis was taken from the Tableau website from a user named Michcael Martin. The dataset, titled "Sample - Superstore Sales (Excel).xls," comprises ten thousand entries, providing insights into customer purchasing patterns, including location, product category, price, and additional details [6]. This data set is a subset from a larger Superstore Sales dataset from 2017. For our analysis, the sample dataset is large enough to allow our team to find the general trends correlation between customer buying habits and weather conditions.

The sample Superstore Sales dataset did not come with any weather condition data, so our team needed a way to get the relevant weather data. To get the necessary weather data, our team utilized an open weather api called Meteostat. Using the location and date data from each entry in the sample Superstore Sales dataset, our team was able to retrieve and append the corresponding weather data into the data. The weather data that our team has decided to pull and analyze is the average temperature, precipitation, and average wind speed for each entry.

IV. Data Analysis I

Prior to data analysis, our team needs to preprocess and clean the data to make it workable. Our team first drops any unwanted or unrelated columns from the data set. Some of these columns include “Row ID', 'Order ID', 'Ship Mode', 'Customer ID', 'Customer Name', and more. These columns are not needed for our analysis and therefore can be dropped without consequences. During this step, our team noticed that there were some very extreme outliers in the data, and leaving in the data might have unwanted influence on the result of the analysis. To solve this problem, our team chose to remove entries that might have a too extreme of an outlier. Removing an entire entry is the best way to deal with extreme outliers in this particular situation because some of the columns are connected in some way to one or more other columns in the dataset. Any data that is removed cannot be imputed because of their connection to other columns. An example of this is for the sales column, where its value combined with the quantity sold, gives the profit amount, therefore imputing the sales value would invalidate that entire entry. Our team came to the conclusion that removing the entry if it contains any abnormal outliers is the best option.

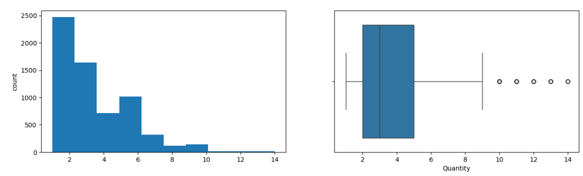

Fig. 1: Distribution of Quantity Sold

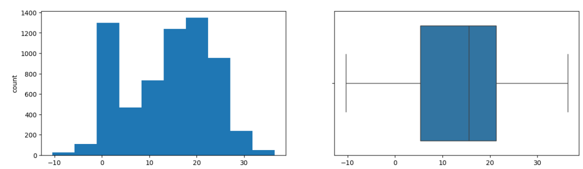

Fig. 2: Distribution of Average Temperature

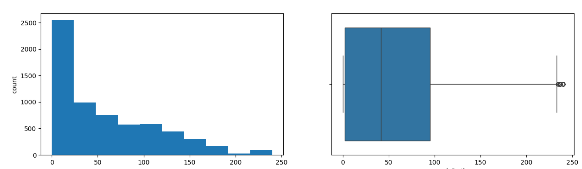

Fig. 3: Distribution of Precipitation

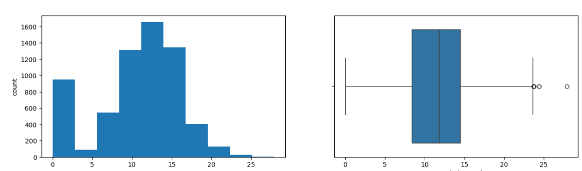

Fig. 4: Distribution of Average Wind Speed

To analyze the data to find customer buying trends during certain weather conditions, our team has decided to focus our analysis on sub category, quantity sold, average temperature, precipitation, and average wind speed data feature of the dataset. Figure 1 to 4 shows the histogram and boxplot of relevant columns in the dataset our team used for our analysis. It shows the general trends and distribution for each category.

Figure 1 displays the quantity amount customers tend to purchase of a certain item. Here we can see that customers tend to buy items in smaller amounts, and might indicate that this dataset skews towards small purchases and not any bulk buying of items. In figure 2, the histogram shows that the bulk of the sales happened when the average temperature is neither too high nor too low. This follows our expectations where people would rather stay indoors when temperatures are too extreme. Another worthy point to mention is the trend seen in figure 3, where most purchases are made when precipitation is low. This also conforms with our team's expected trend and pattern of purchases, as people tend to stay indoors more when it is raining. As for average wind speed, the trend is similar to that of the average temperature, where most purchases are made when wind speeds are not too low or high.

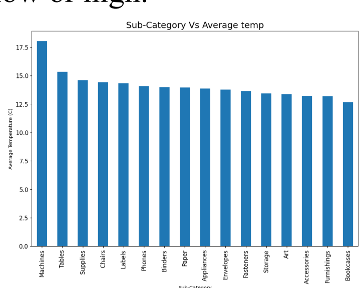

Fig. 5: Bar Plot of Sub-Category and Average Temperature

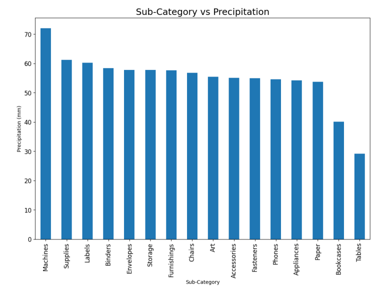

Fig. 6: Bar Plot of Sub-Category and Precipitation

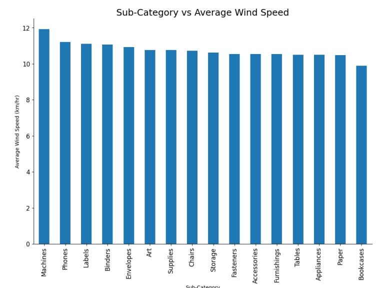

Fig. 7: Bar Plot of Sub-Category and Average Wind Speed

Figure 5 to 7, shows the average weather condition for each product subcategory. For each weather condition, our team found that the subcategory machines tend to be bought on average at more extreme conditions. This is an expected result from our team because machines are more prone to breakdowns in hasher conditions due to increased stress on their components or reduced operational efficiency.

V. Data Analysis II

Upon our team's analysis of the dataset by subcategories and quantity sold, we have found some interesting and unexpected results.

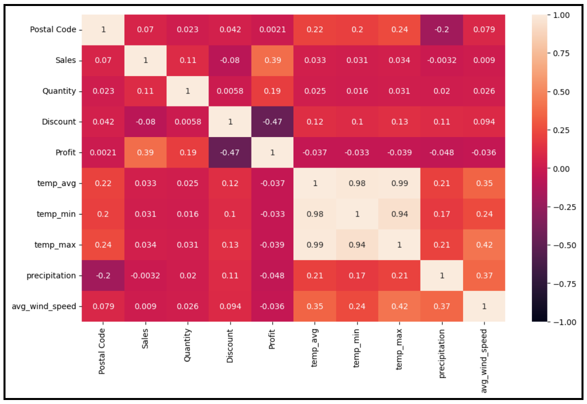

Fig. 8: Heat Map of the data Set

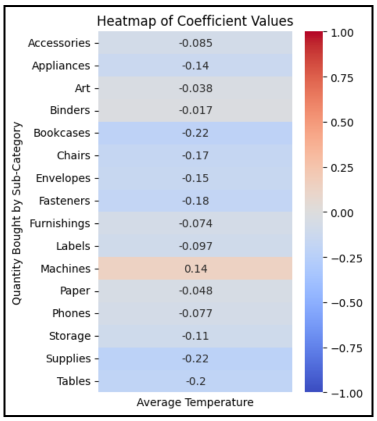

Fig. 9: Heatmap of Coefficient Values and Average Temperature

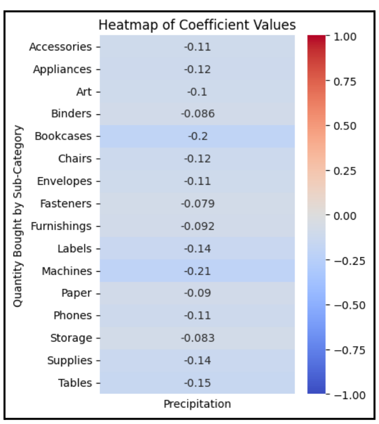

Fig. 10: Heatmap of Coefficient Values and Precipitation

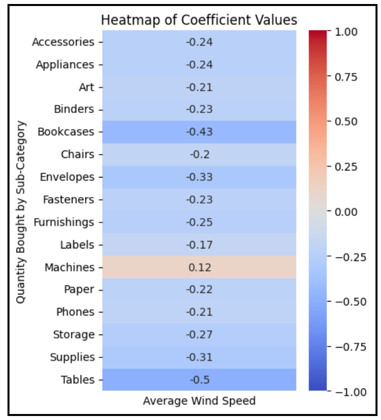

Fig. 11: Heatmap of Coefficient Values and Average Wind Speed

Our team first looked at the heat map for the overall data set, which is shown in figure 8. Here we can see all of the correlation coefficients of each feature to another. Our team was surprised to see that there are little to no correlations between the different sales data to the weather data.

Our team wanted to dig deeper into the relationship between purchasing trends and weather conditions, so we decided to analyze subcategories of products sold by quantity for each weather condition. We did this by grouping each subcategory with the weather condition and summing them by quantity, and creating a heat of the relationship. The result of which can be seen in figure 9 to 11, which showcase the correlation coefficients of each subcategory sold by quantity during certain weather conditions.

Our findings were unexpected as there is almost no correlation found between quantity sold by subcategory versus average temperature, or precipitation as seen in figure 9 and 10. Our team did find an interesting relationship in figure 11, where the subcategories bookcase and table have a moderate negative correlation with the average wind speed at -0.43 and -0.5 respectively. This means that for higher wind speeds customers tend to buy less quantity of bookcases and tables.

VI. Conclusion

Most prior research on the relationship between weather and sales has been done with logical scaffolding in place. However, when we decided to look at all the correlation values between weather and different categories, we found some interesting values. The most prominent one being the negative correlation between tables and bookcases vs wind speed. There is absolutely no psychological cue as to why this could happen. However, it is a strong indicator that there are definitely undiscovered buying patterns that we have not yet discovered. The human brain is extremely complicated and trying to use existing psychological models blinds us to patterns that using an unbiased approach such as ours could reveal. This research shows that we need to re-examine the relationship between weather and buying patterns while looking at all possible correlations. This could help suppliers maximize their profit in unexpected ways and help consumers get products that they would usually find sold out.

Reference

[1] N. S. Arunraj and D. Ahrens, "Estimation of non-catastrophic weather impacts for the retail industry," International Journal of Retail & Distribution Management, vol. 44, no. 7, pp. 731–753, 2016. DOI: 10.1108/IJRDM-07-2015-0101.

[2] 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, pp. 102583, 2021. DOI: 10.1016/j.jretconser.2021.102583.

[3] 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, pp. 101921, 2020. DOI: 10.1016/j.jretconser.2019.101921.

[4] S. Okayama, H. Yamashita, K. Mikawa, M. Goto, and T. Yoshikai, "Relational analysis model of weather conditions and sales patterns based on nonnegative tensor factorization," International Journal of Production Research, vol. 58, no. 8, pp. 2477–2489, 2020. DOI: 10.1080/00207543.2019.1692157.

[5] N. Rose and L. Dolega, "It's the Weather: Quantifying the Impact of Weather on Retail Sales," Applied Spatial Analysis and Policy, vol. 15, no. 1, pp. 189–214, 2022. DOI: 10.1007/s12061-021-09397-0.

[6] M. Martin, "Sample - Superstore Sales (Excel).xls," Tableau. [Online]. Available: https://community.tableau.com/s/question/0D54T00000CWeX8SAL/sample-superstore-sales-excelxls. Accessed: October 29, 2023