

Reporting Date: 23rd February2024

TABLE OF CONTENTS

[ABSTRACT 2](#_Toc159599442)

[INTRODUCTION 3](#_Toc159599443)

[DATA PREPARETION 3](#_Toc159599444)

[EXPLORATORY DATA ANALYSIS 4](#_Toc159599445)

[CONCLUSION 10](#_Toc159599446)

[REFRENCES 11](#_Toc159599447)

# ABSTRACT

The new datasets obtained by AutoBasket was explored in this current project for an upcoming collaboration with HelloFresh Canada. The initial data set, User Interaction Data, revealed parameters like User\_ID, Sessions initiated, Last Purchase and, among others, which serve as inputs for the analysis, revealing the user engagement and choices. A pivotal update involved augmenting the Delivery Preference database with geographic information, incorporating three new columns: A state, latitude, and longitude. This boost in performance also brought ease of spatial representation. As a result, we were able to create effective visualizations such as delivery frequency distribution, address frequency distribution, food category count, sessions-initiated distribution, and purchase history. The Power BI Dashboard integrated the visualizations making them accessible to all members of the team giving a full picture about the behavior of the users and their preferences, the information given is crucial for the decision-making in the campaign with HelloFresh Canada.

# INTRODUCTION

AutoBasket is a company founded by Larry and Veronica Smiles with the purpose of simplifying the grocery shopping experience for individuals and families. Situated in the heart of Toronto, the company's headquarters pulsate with the energy of a city known for innovation and diversity. The company's app automates grocery and recipe lists, linking recipes to required products, and streamlining the shopping process. Today, AutoBasket assists households across the world to save time and energy on their weekly grocery runs, making it easier for busy families and individuals to get everything they need for a great meal.

AutoBasket has hired us as interns to focus on all areas of development for the organization and to help create solutions to common issues within the industry. Our role as interns is not only focused on the knowledge you have learned in school but also on developing your soft skills, including presentations, teamwork, and leadership. This internship will provide us with valuable real-world experience and an opportunity to contribute to the continued success of AutoBasket.

# DATA PREPARETION

The initial stage of the foregoing campaign with HelloFresh Canada is to attain a thorough knowledge of the fresh datasets that were added and acquired. User Interaction Data is the primary source, in which we work with variables User\_ID, Sessions initiated, Questionnaire\_fill, Most Purchased Items, and Last Purchase. These metrics give hints into user engagement, as well as preferences and purchasing patterns. Contrarily, what is Delivery Preference database including of User ID, Delivery Address, Delivery Frequency and Preferred Delivery Day offering? Furthermore, Recipe Preference database has User ID, Preferred Category, and Most Requested Recipes, these fields reveal the user culinary taste and specific needs on recipe.

However, the process of creating graphical representations of data requires a significant adaptation. Recognizing the significance of geographic information, we have updated the Delivery Preference database by incorporating three new columns: State, Latitude, and Longitude. Spatial representation of customers distribution is enabled by use of users’ addresses which we have used as a basis of these additions. By adding geographical coordinates and city information in the data, it becomes more amicable for visualization and therefore you can have maps and location-based insights that are incorporated into the presentation cohesively.

First, the visual defines the spatial distribution of customers’preferences by virtue of home delivery, which is in turn useful to draw out a sketched map of users’ distribution across regions. This information is useful for locating groups where participation is high which can be used for targeted marketing campaign, also, processes may be improved. Besides this, the latitude and longitude tiles will allow for the presentation of location-based visualizations in dynamic and interactive forms, thus depicting the delivery frequencies as well as the preferences of people. This approach of strategic growth is consistent with the mission of creating a well-structured and visually attractive proposal which will have a special focus on our cooperation with HelloFresh Canada. Plus, these elements also manage to provide a higher efficiency of the data preparation and the effect of the visualization process will show this by bringing out both a global efficiency and a local performance.

# EXPLORATORY DATA ANALYSIS

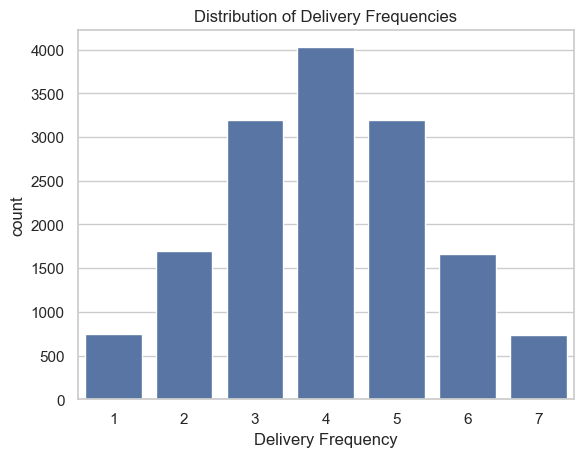


Figure-1: Delivery frequency distribution

This bar chart above shows the number of deliveries a user has order within a week, which represents the distribution of delivery frequency. Every column in the diagram represents a specific frequency, going from 1 to 7 times per week. The height of each column, however, indicates the count of those users who have used the service that particular number of times. The graph gives the clue as to how the service is being used by the users, helping us as to know how much the users are depending on this service. This info can be a valuable tool in driving dispatch time adjustments and increasing customer satisfaction level.

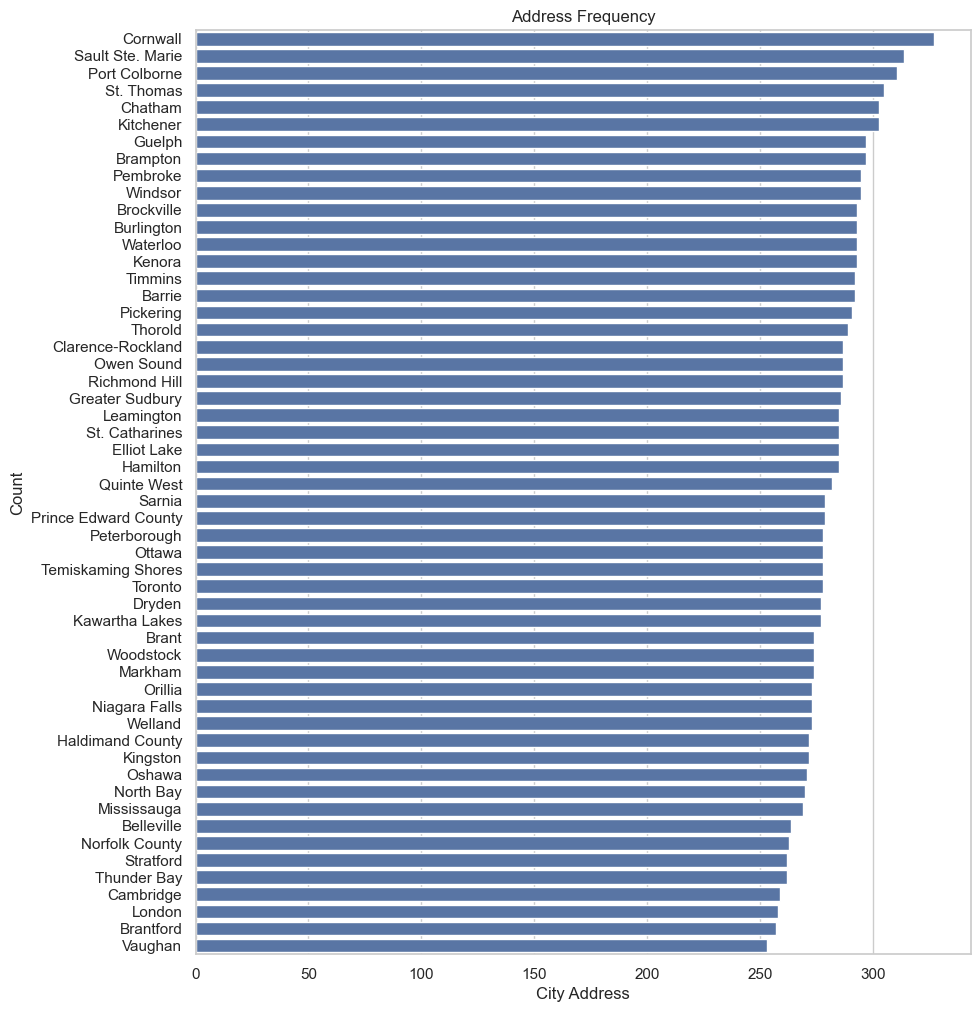


Figure-2: Address frequency distribution

The horizontal bar graph presented above offers a comprehensive view of the frequency of addresses across various cities. Each bar corresponds to a specific city, with its length indicating the number of addresses associated with that city. The cities are listed on the Y-axis, while the X-axis represents the frequency count.

From the graph, it’s evident that Vaughan has the highest address frequency, followed by London and Bradford. Conversely, cities like Sault Ste. Marie and Cornwall have some of the lowest frequencies. This graph provides valuable insights into the geographical distribution of addresses, which can be instrumental for logistics planning and resource allocation.

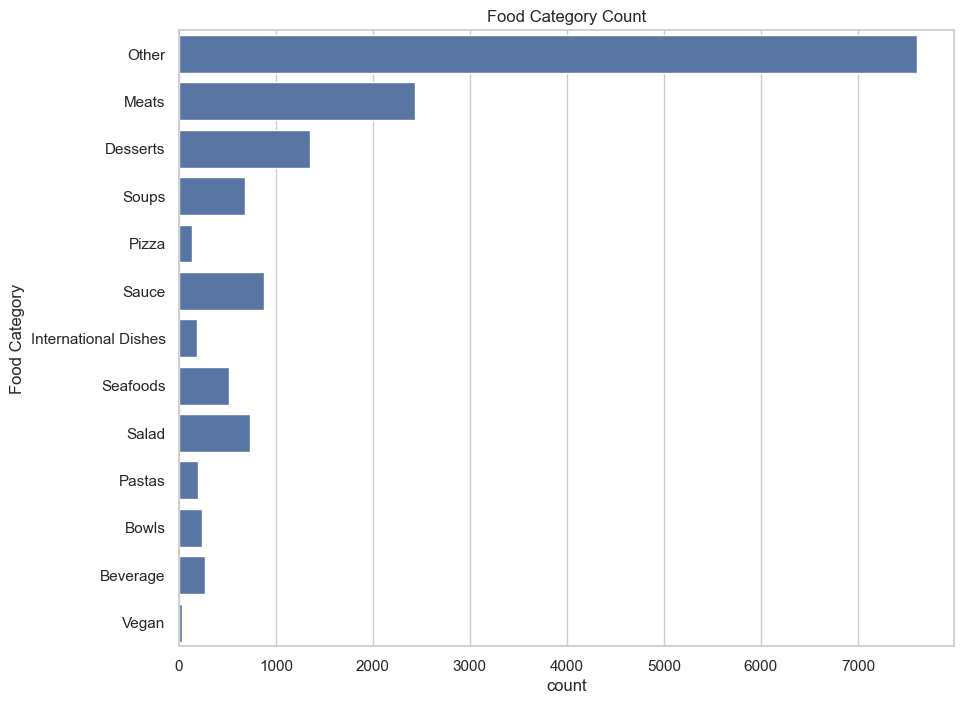


Figure-3: Food category count

This bar chart shows us the different food groups and the numbers that stand for each one of them. Each one of the bars is a food category and the length of the bars indicates how much of the category is there is. This graph is a very useful aid in the order to prepare a good menu, as well as the right inventory. This knowledge shall help us to further to please our customers' taste even better.

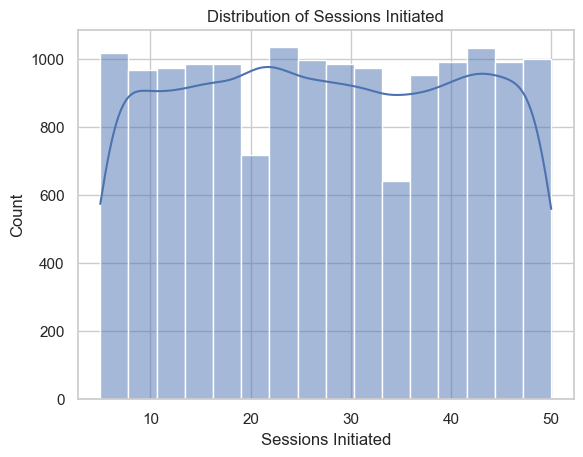


Figure-4: Sessions initiated distribution

The image is a graph which is made of both a bar and a line graph in order to illustrate the distribution of sessions induced by users through time or under other circumstances. The graph is useful to assess user engagement and user retention, at the same time to notice the patterns or outliers in user behavior.

The table displays that the number of sessions opened up ranges between 0 and 50, being bumpy. The sessions initiated metrics had a peak value of around 10 and more than 1000 users starting sessions. The minimum sessions begun are 0, around 800 users not initiating any sessions. The graph shows that the sessions initiations are also down around 25, and around 50 they are up. The graph shows that there is a wide range of user-engagement since some users conducted several sessions while others didn’t conduct any. The graph however means that there are chances of improving customer loyalty since there are so many users who stay in only one or two sessions.

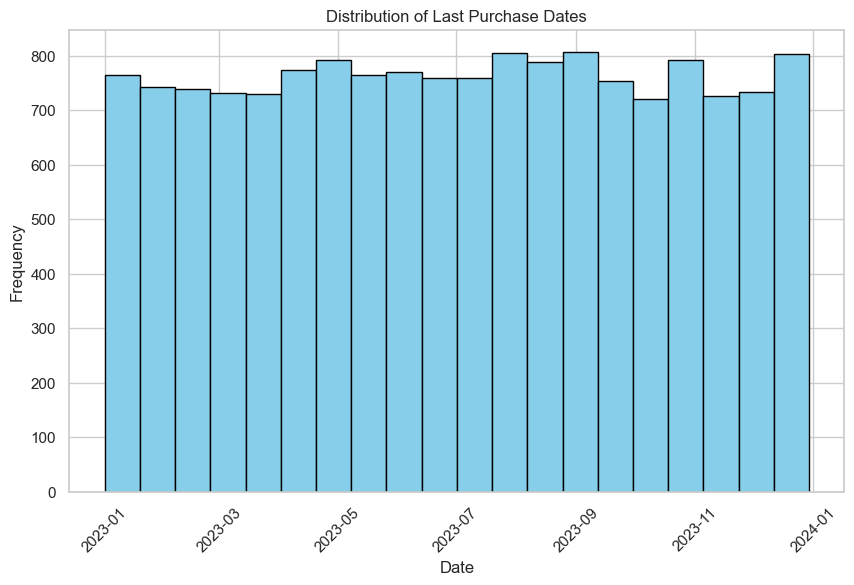


Figure-5: Purchase History

The image is a bar chart that is used for the managing of the frequencies of the customers' purchases on the different dates in between January 2023 and January 2024. The graph is very useful for checking the loyalty of the customer and/or engagement rate over time. The line demonstrates that sales kept extremely high and with some fluctuations, but almost constant from January to December. Apex of the highest happened in the July 2023. At the same time, the lowest point was witnessed in December 2023. The graph indicates that customers are on the whole quite pleased with products/ services they buy and they depend very little if at all on seasonal preferences. This chart speaks for the increasing possibility of marketing or bringing incentive as people purchase the product in more frequency.

A screenshot of a computer

Description automatically generated

Figure-6: Power BI Dashboard

Delivery Locations Map: The top left section of the dashboard identifies the delivery locations geographically by the latitude and longitude coordinates. This visualization then allows to see the distribution of deliveries around the map.

* **Delivery Frequency**: Under the map, many numbers, that is 61K, stand for the total number of deliveries. This counts for the overall customer traffic.
* **Preferred Food Categories**: Place the pie chart with a heading "Preferred Food Categories" on the right side of the map. This graph gives the clues about the most preferred food groups by the users.
* **Quarterly Data**: On the right-hand side of the pie chart is a table which depicts data for each quarter and the total of sessions initiated, per quarter, and the average time ranged in minutes. Such information can be used to see how the user engagement changes over time.
* **Monthly Delivery Frequencies**: The below the table is a bar graph illustrating the total of frequency of deliveries and sessions that are initiated by month. The graph may provide clue on any trends or pattern that occur during different seasons as regards to delivery frequency.
* **Last Purchase by City**: Below the dashboard line graph showing the count of last purchases by city. The graph is offering a glance at the delivery service demand in different city areas.

With every component being integral to a wholesome understanding of user behavior and preferences, overall, easier decision-making and strategy planning is ensured for the delivery service.

# CONCLUSION

To sum up, the exploratory data analysis and visualization of the enrich data facilitate the collaborative efforts with HelloFresh Canada by allowing to understand the customers' behavior, preferences, and the geographic distribution. Integration of new data, such as User Interaction, Delivery Preference, and Recipe Preferences, enabled us to get a clear picture of how our customers relate to us, how they get their meals delivered, and their culinary preferences. The deliberate integration of geographic data, latitude, and longitude, has advanced the capacity to visualize spatial data, featuring interactive maps and location-specific insights

The figures 1 to 6 that are given in the form of graphs provide a detailed and multifaceted view of all the aspects such as delivery frequency distribution, address frequency distribution, food category preferences, session initiation patterns, transaction history and master Power BI Dashboard which contains all the key indicators in one place. Such visualizations not only give a simple representation of current trends but also offer an excellent foundation for future decision-makers and strategists on their plans.

The programming the strategic growth roadmap of updating the Delivery Preference Indicator with geo-information matches the aim of preparing a neatly arranged and neat proposal for the collaboration with HelloFresh Canada. These visuals provide a powerful imagery for the purpose of knowledge of user engagement, improvement of logistical operations, menu adaptation according to customers' preferences, and identification of areas that require modifications. Equipped with such data acumen and illustrative visuals, AutoBasket can articulate advantageously, merging the local’s performance with a global efficiency’s concept. This analysis will inform the subsequent steps towards partnership collaborations ensuring that the insight-driven decisions are made, and the joint-venture is successful.

# REFRENCES

1. (2024, 2 16). Retrieved from faker: https://faker.readthedocs.io/en/master/
2. *synthetic-data-generation-techniques*. (2024, 2 16). Retrieved from turing: https://www.turing.com/kb/synthetic-data-generation-techniques