Matt Held & Kevin Broderick

CS 305

Final Project Group Assignment

Profitability of Walgreens Time Specific Delivery

Dataset Information:

We pulled a dataset from the Walgreens API. This API is available at *developer.walgreens.com* and is available for anyone with a Walgreens Developer account, which can be created at no charge. We specifically used the Store Locator API, with a radius (flag as “r”) set to 10000 (miles) and 10000 items per page (flag as “s”), which are both max values for the free/public API. Our goal was to get that 48000 to 60800 zip code range without missing any data, hence the high items per page number. We chose this zip code range as it included both IL and WI results and included downtown Chicago which has a lot of Walgreens locations in a tighter space. We ended up with 453 store locations from our API data pull. After preprocessing, one extra comment with the dataset is that our dataset included numerical information and location information (see more details in preprocessing). This limited us in the charts we were able to create after preprocessing, due to the tight scope of the data. This is both a positive to have a focused “question” in mind, but also creates conflicts with project requirements. We believe we created graphs that display a high level of understanding of Tableau and the data analytics process overall, even if they weren’t as diverse as the many, we explored in class sessions.

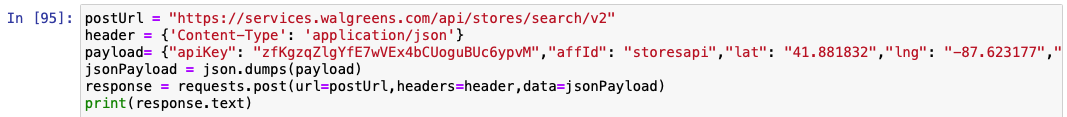


Figure 1: API call in Python to Walgreens Store Locator API

Question:

Our question revolved around the profitability of delivery scheduling. Walgreens delivery scheduling is more like an “Uber Eats” delivery timeline, for example, delivery between 4:30pm to 5:00pm, instead of a “delivery before 8:00pm” timeline. Some locations are configured to do scheduled delivery, while some are not. Delivery charges also are different on a per location basis. With this in mind, our core question is:

*Does the delivery scheduling option result in consumers paying a higher delivery fee?*

If the scheduled delivery causes a higher cost, it could be a more profitable option for Walgreens and to compete with other companies.

Data Preprocessing:

We used Python for data collection (from API), and for data preprocessing. This included importing the following packages, requests, JSON, pandas, and NumPy. All preprocessing work was done inside of Jupyter Notebook.

The first technique used was simply importing only the relevant data/columns that we wanted into the data frame. We did this by a for loop that selected the following information from each store location pulled by the API.

* deliveryScheduling
* deliveryFee
* address
* state
* city
* zip



Figure 2: Only imputing applicable data into the pandas data frame (named walgreensDF)

The next preprocessing technique was fixing data formats. For example, we know that delivery scheduling, delivery fee, and zip are all numerical values, but they are currently objects (string/char). Due to this, we applied the .astype function to change the delivery scheduling to an int, delivery fee to a float, and the zip to an int, as pictured below.

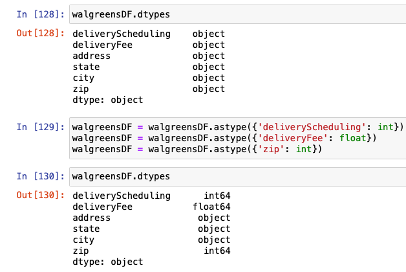


Figure 3: Now we see that termination\_date no longer has N/A values in it

As mentioned previously, another aspect of our target area was to only investigate in zip code range 48000 to 60800. Looking at our walgreensDF with the describe function we can see that there were zip codes below and above this range. Once we found out that there were zip codes we didn’t care about, we used the drop function to remove rows where the zip column was below 48000 or above 60800. After the drop we can see that our number of rows drops from 452 to 409. See code snippets below for additional details.

Table

Description automatically generated

Figure 4: Using describe function on walgreensDF

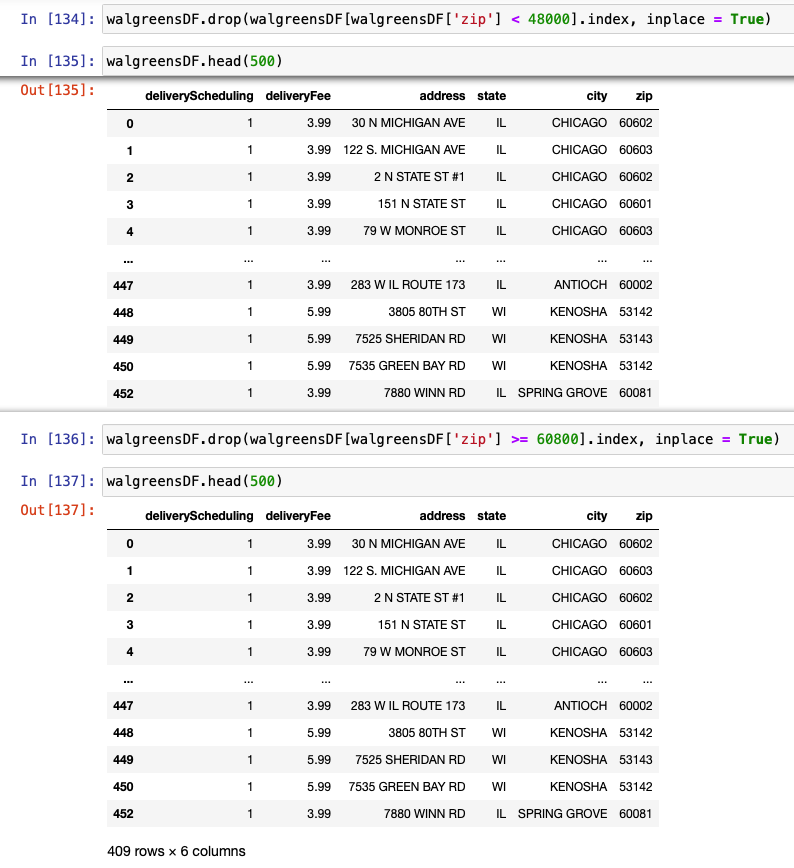


Figure 5: Removing unwanted Zip codes with pandas drop function

The last preprocessing step we looked into was if our data had any outliers. We used the describe function to look closer at this. As we can see, no outliers exist. This is a reasonable conclusion because the Walgreens data is standardized by the amount they charge for delivery, and if there is scheduling or not.

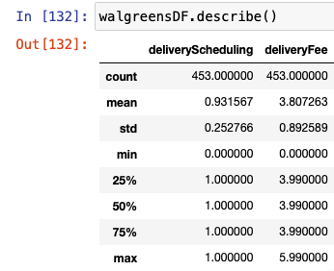


Figure 6: Checking for outliers with describe function

Data Visualization in Tableau:

**Map with Bubbles**

With this chart, we can get a visualization of where the Walgreens are in northeastern Illinois and southwestern Wisconsin. This is how we can get a better understanding of the target market and the general demographics of the area. This chart was less for use in the data analytics process, but more to inform shareholders of the locations and area they asked us to investigate.

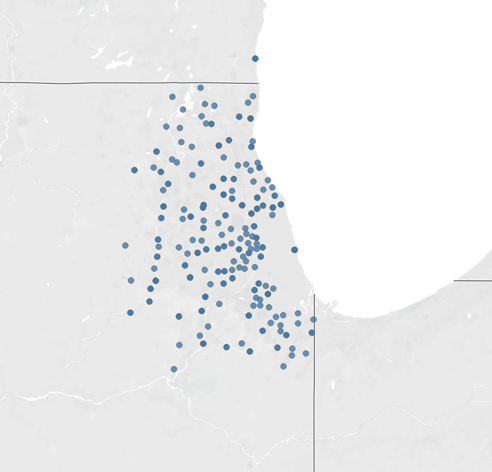


Figure 7: Locations based on the addresses based on Lat/Long coordinates

**Basic Bar Chart**

The second graph we used was a bar chart. This chart compares the delivery scheduling options. On the x-axis being 0 (scheduling not available), and 1 (scheduling available). On the y-axis is the delivery charge average from all locations in our zip code range. This provides a basic view that the delivery scheduling option provides about a $3 advantage over the fee of a non-delivery scheduling option.

Chart, box and whisker chart

Description automatically generated

Figure 8: Bar Chart Comparing Avg. Delivery Cost without (0) and with (1) Delivery Scheduling

**Bubble Chart**

This graph represents which stores (each “bubble” is a zip code) have delivery scheduling and which do not. It gives stakeholders an easy look at this data without complicated numbers. The blue marks represent the Walgreen stores that have delivery scheduling while the turquoise marks represent the ones that do not have delivery scheduling. Part of the experience of Walgreen customers is continuity between stores, having a handful of locations without scheduling creates a disconnect in the customer experience.

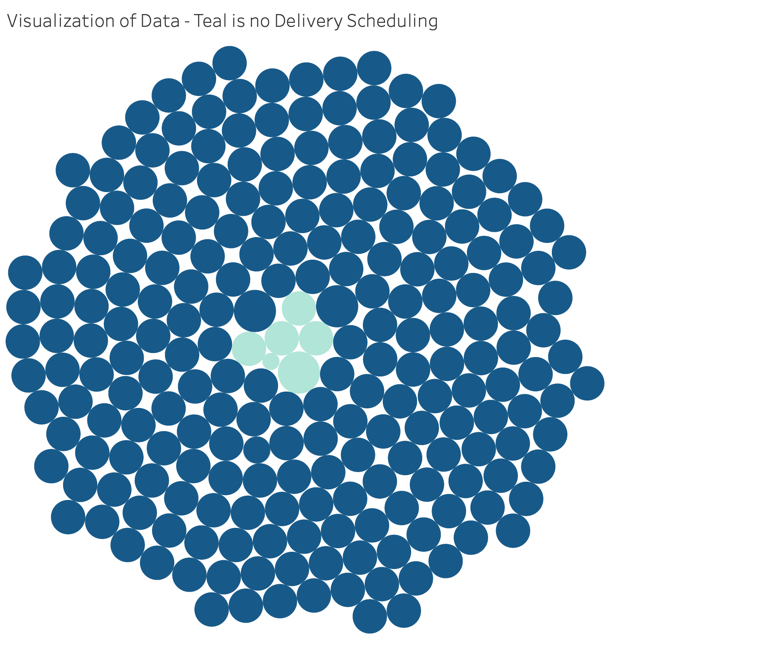


Figure 9: Bubble Chart of Each Location by Zip Code

**Map Chart for Locations Without Delivery Scheduling**

This graph shows the locations of the cities in our dataset that do not have delivery scheduling. Lighter colors means that the delivery fee is low while the darker shades of blue means that the delivery fee is more expensive in that area. We included the delivery scheduling statistics for that area to show that these areas do not have scheduling

Graphical user interface, application

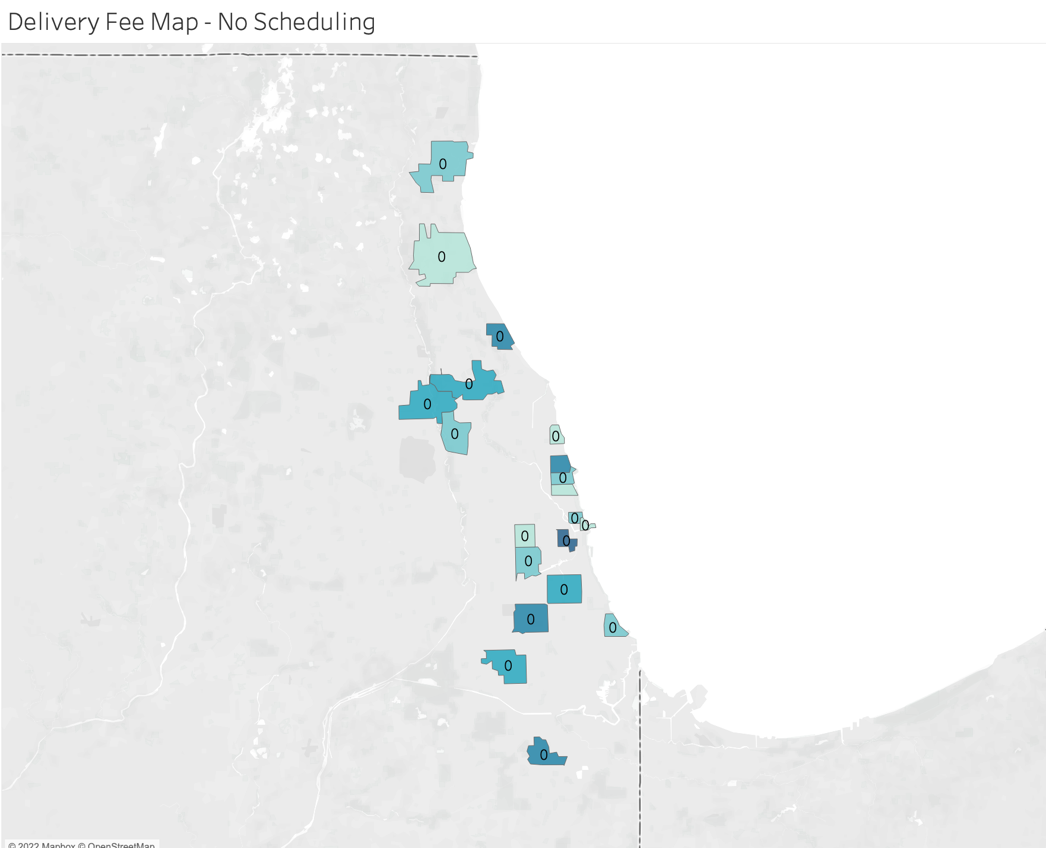
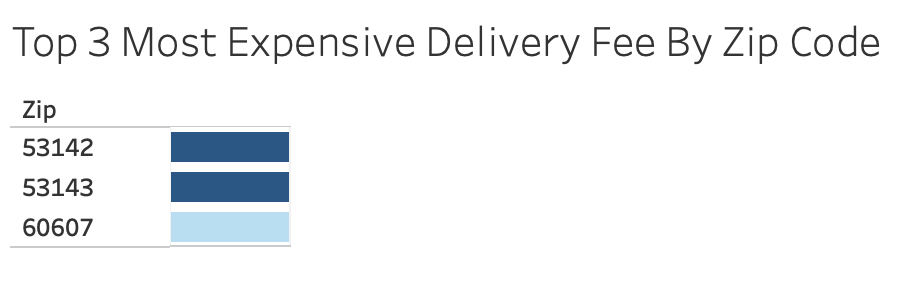
Description automatically generated

Figure 10: Map Chart that Shows Each Location Without Delivery with Color Scale of Delivery Fee

**Bar Chart with Color Gradient**

With this graph, we discovered that the locations with the highest delivery fees ($5.99) were in Kenosha and Chicago. Kenosha has delivery scheduling while Chicago does not. Delivery scheduling has a major impact on the delivery fee due to both Kenosha zip codes having delivery scheduling and as we previously discussed, the high fee. Since central Chicago is a part of a city hub, we see that this type of environment causes a higher fee, so it’s not a great approach to judge delivery scheduling on that direct area.



A picture containing table

Description automatically generated

Figure 11: Bar Chart of Top 3 Zip Codes with Most Expensive Delivery Fees with Color Bar for if Area has Delivery Scheduling

**Tableau Dashboard**

While we explained our graphs in previous steps, our dashboard focuses on the “story highlights” of delivery scheduling in its current context at Walgreens. We chose to show the most direct graph first showing that cost resulted in a $3 fee increase for when delivery scheduling is offered. We then go into showing that only a few locations do not offer delivery scheduling. However, we emphasize that the top two locations with the highest delivery fee also offer delivery scheduling. We then tap into consumer consistency with a few IL. locations that do not offer delivery scheduling, yet still have a high delivery fee. This creates a disconnect in the consumer experience and not only is a loss of profit, but a loss of consumer satisfaction.

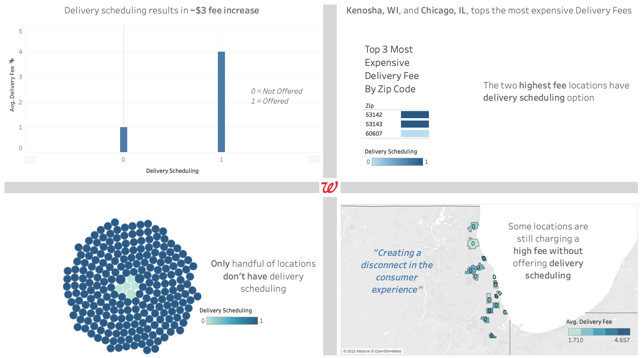


Figure 12: Our Tableau Dashboard

Conclusion (Overview):

Overall, we can see in our charts that delivery fees tend to be higher when delivery scheduling is included. This concludes that Walgreens should include delivery scheduling options in the remaining store in the selected zip code range. We also considered the consumer continuity aspect with only having a handful of locations in the zip code range that do not have delivery scheduling.

Division of Labor:

We collaborated in each aspect of this project, however, here is what we each took a lead/focused on.

|  |  |
| --- | --- |
| **Matt Held** | **Kevin Broderick** |
| * Preprocessing * Data Collection (API) * Tableau Dashboard | * Tableau * Adding location data on from API pull |

We peer reviewed each other’s work for the parts of the report we worked on. Here are the parts of the report that we focused on.

|  |  |
| --- | --- |
| **Matt Held** | **Kevin Broderick** |
| * Dataset Information * Preprocessing * Conclusion | * Question * Data Visualization |

References

“Walgreens Developer Program.” *Walgreens Developer*, developer.walgreens.com. Accessed 29 Apr. 2022.