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 closer space. We ended up with 453 store locations from our API data pull.

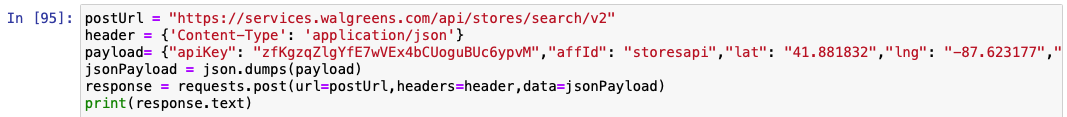


Figure 1: API call in Python to Walgreens Store Locator API

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 four 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.

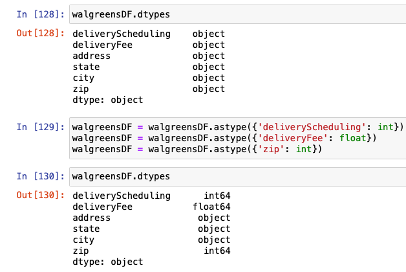


Figure 3: Now we see that termination\_date no longer has N/A values in it

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

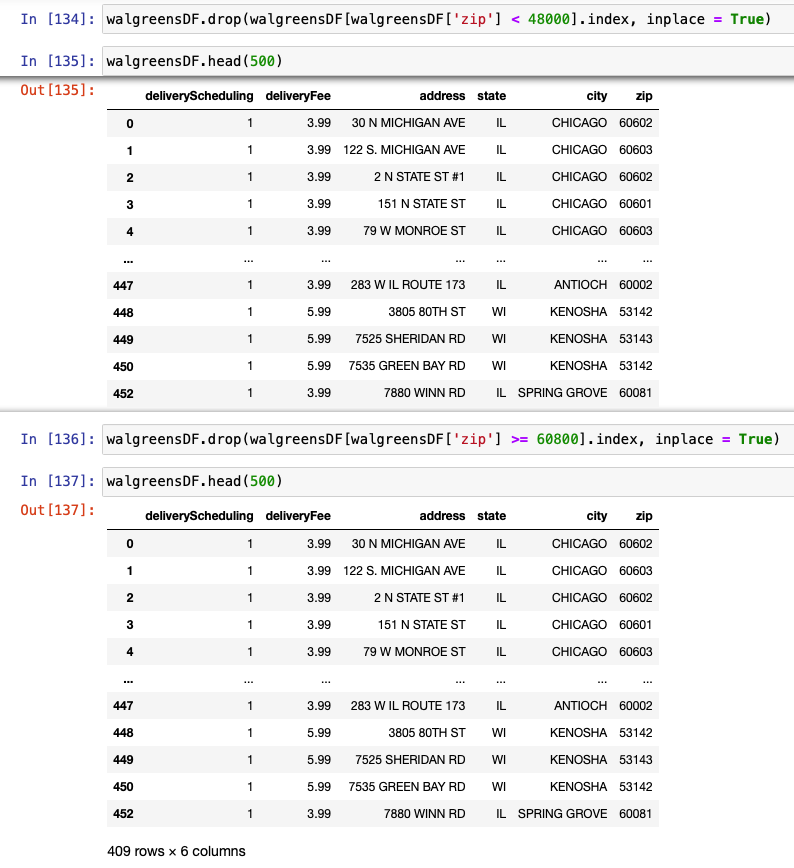


Figure 5: Removing unwanted Zip codes with pandas drop function

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.

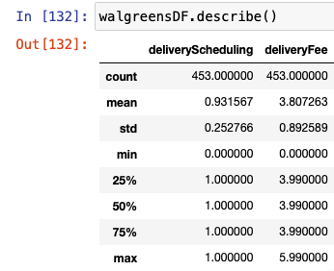


Figure 6: Checking for outliers with describe function

Data visualization in Tableau:



Figure 7: EXAMPLE TEXT FOR IMAGE CAPTION

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** |
|  |  |

We each 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** |
| * Preprocessing |  |

References