GRUBHUB VERSUS UBEREATS TWITTER SENTIMENT ANALYSIS

BYGB 7978-004: Web Analytics
Professor Zixin Mao

Ao Li (Hazel)
Ila Srivastava
Katie Linh Cao
Lakshmi Sravya Katta

Content

1. ***Executive Summary***

2. ***Business Goal Analysis***

3. ***Dataset Description***

4. ***System Design***

5. ***System Implementation***
   - ***Data Analysis***
   - ***Data Cleaning***

6. ***Evaluation***
   - ***Content Analysis***
   - ***Location Analysis***
   - ***Influencer Analysis***
   - ***Word Cloud Visualization***

7. ***Conclusion and Future Direction***

8. ***References***

1. Executive Summary

Online food delivery platforms such as UberEATS and GrubHub serve an important role for sustaining growth in the restaurant industry and present a meaningful opportunity to shift consumer buying behavior online. According to the National Restaurant Association and estimates by Cowen Securities, off-premise sales accounted for $43bn of the $745bn US restaurant market and is expected to reach $43bn by 2022 (LINK). These estimates are backed by commentary from national chain restaurants like McDonalds (partnered with UberEATS) that state online delivery orders are typically 1.5x-2x larger in check size, where ~70% of these sales are incremental business (LINK).

Metropolitan areas such as New York City are particularly attractive markets for online food delivery, due to their population density, young demographic, and relatively high per capita income. The incumbent, GrubHub reportedly accounted for ~86% of online food orders in the greater NYC area in 2017 (LINK) but has since ceded share to innovative competitors including UberEATS, Postmates, Door dash, and Caviar.

Online takeout businesses share many characteristics with other ecommerce platforms such as Expedia (online travel), Wayfair (furniture), Zillow (housing), and Amazon (everything else). Limited product differentiation and structurally lower margins in these distribution business models mean that commercial success is based largely on three factors:

1) The ability to achieve superior unit economics through scale (i.e. maintaining/growing share)
2) Eliminating friction in onboarding productive new vendors/restaurants
3) The strength of the customer relationship (i.e. brand loyalty/retention)

2. Business Goal Analysis

In our project, we are projecting to find our which food delivery service is more well received on Twitter. Previously there are many articles and news comparing different food delivery services including Uber Eats, GrubHub, Seamless and DoorDash. Previous research focus on varies fields including software interface, delivery speed, driver attitude, etc.

Here, we would like to collect feedbacks from customers who order food on food delivery service. We are projecting to collect customers twitter mentioning ‘UberEats’ or ‘GrubHub’. Extract their twitter and relevant metadata, then perform a sentiment analysis to find our which food delivery service is more popular on Twitter.

Based on the user tweets, we would like to analyze the sentiment of each tweet by dividing them in to three categories: Positive, Negative and Neutral. This would enable us to identify the most popular user sentiment associated with the above two services. Through this analysis, we would also like to capture the key points that each delivery service should improve upon and suggest some important recommendations based on user feedback.

By incorporating the changes prescribed, both the delivery services would stand to benefit on customer satisfaction, short term profit, long term wealth gain and sustained brand image improvement down the line.

3. Dataset Description

Screen Name: It is the name of the user that appears in the twitter handle.

Tweet: The text that users has tweeted.

User ID: It is an ID internally generated by twitter.

Location: The base location of the user that the user has added to his/her profile.

Follower Count: The number of followers of the user.

Favorite Count: The count of the number of times the tweets has been liked.

Retweets Count: The number of retweets done by the user.

Verified Account: This field is to demonstrate if the user has a verified account or not. Appears as a blue tick.

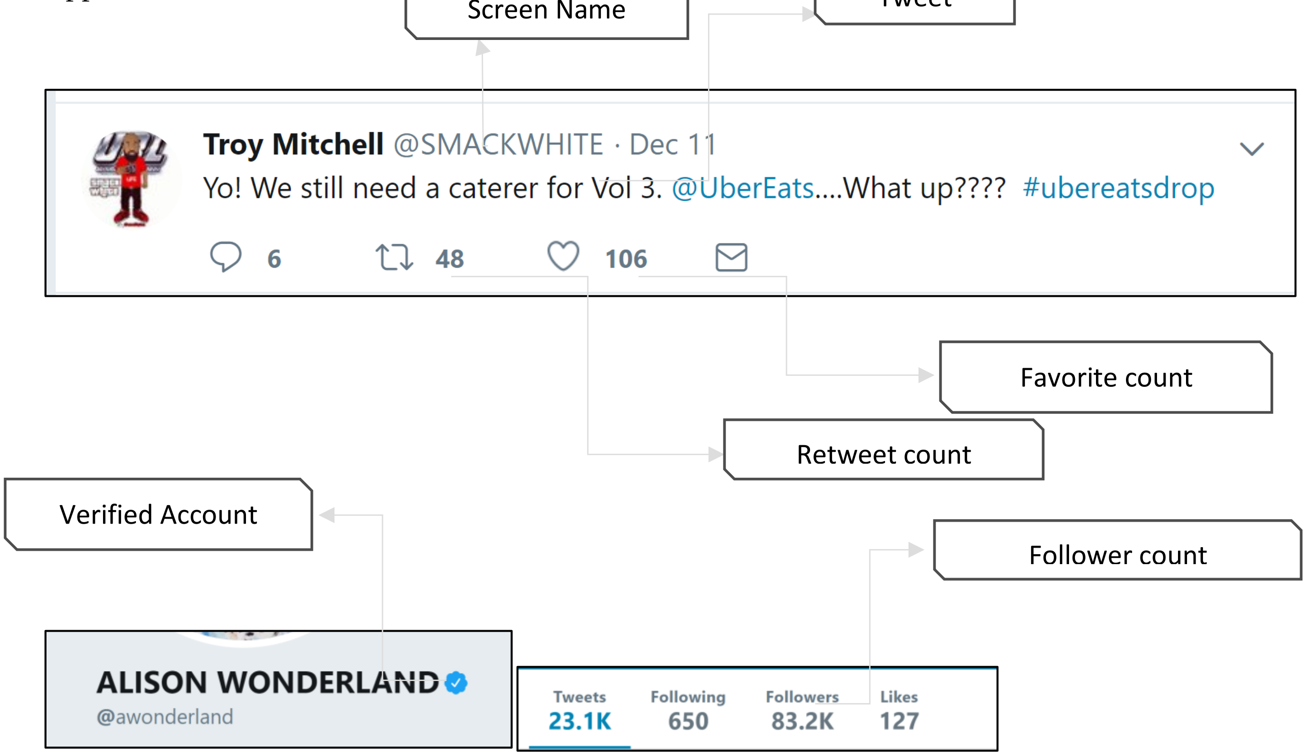

Figure 3.1 Data Description

We wrote a python code using **Twitter API** to extract the above attributes. The code comprise of four functions**:**

get_tweets_uber(): This function extracts all the tweets related to **@UberEats**. We pass @UberEats as the search word and the function reads 100 tweets every time and extract the required user related information and tweet. We also made checks to check if any field is empty. If so we define that field as value 'MISSING'.

get_tweets_grubhub(): This function extracts all the tweets related to **@GrubHub**. We pass @GrubHub as the search word and the function reads 100 tweets every time and extract the required user related information and tweet We also made checks to check if any field is empty. If so we define that field as value 'MISSING'.

For both the above functions we have kept geolocation parameter as '40.6971494,-74.2598755,200km' i.e., we are covering tweets within 125 miles in and areas around New York.

cleaning_tweet(text): This function uses library **re** to identify and remove regular expressions, links and irrelevant strings or characters from the tweet. We pass the tweet which we want to clean as text in the function

Tweet_sentiment(text): This function used the library **TextBlob** to identify the sentiment of a tweet. We pass the tweet which we want to analyze as text in the function. TextBlob is a Python library for processing textual data, to perform sentiment analysis. It calculates the sentiment score i.e., polarity and assigns it a float value. If it is greater than 0.0 then the sentiment is positive, if less than 0.0 then neutral and if equals to 0.0 then neutral.

4. System Design

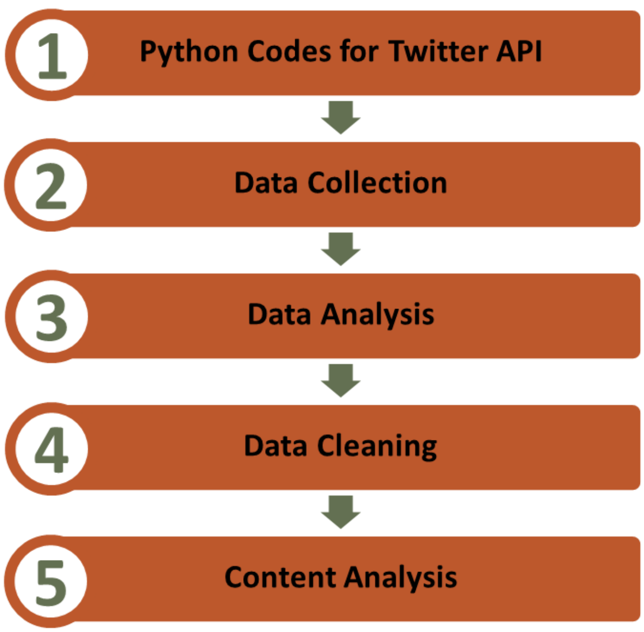

Figure 4.1 Project Phases

Above is a flowchart listing out the five major phases of the project, including writing Python code for Twitter API, data collection, data analysis, data cleaning, and content analysis.

5. System Implementation

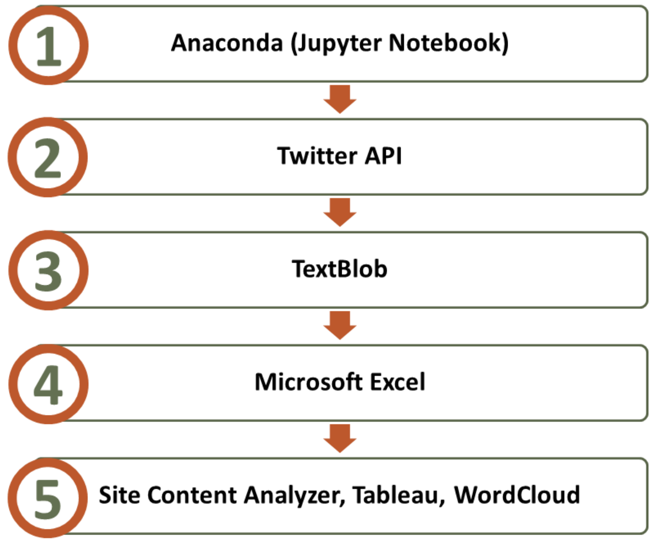

Figure 5.1 Project Tools Implementation

Data Analysis: TextBlob

Sentiment analysis as the task of detecting, extracting and classifying opinions and sentiments concerning different topics, as expressed in textual input. Sentiment analysis can be performed in many different ways. Many brands and marketers use keyword-based tools that classify data (i.e. social, news, review, blog, etc.) as positive/negative/neutral.

TextBlob is a Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more. It give you a "Polarity-score" and a "Subjectivity-score" for your text. Polarity can take on a range from -1 to 1, where -1 is the most negative and 1 is the most positive. Subjectivity ranges from 0 to 1, where 1 is high subjectivity and 0 is low subjectivity (in other words high objectivity).

Here, we have used TextBlob to assign the polarity score for each tweet thereby dividing them into three categories: Negative (< 0), Neutral (= 0), Positive (> 0) thereby, enabling our sentiment analysis.

Data Cleaning: Excel

After collection of data we have stored the data in a text file with tab delimiter. Further we removed the duplicate data from our file in Microsoft Excel. We have also done manual cleansing of few tweets to ensure that we work upon relevant data and avoid duplicate tweets.

For Location field especially, there was a lot of irrelevant data that do not describe any geographical areas. This has been categorized as missing. Also, there was a divergence in the location description where some fields were describing only street or city or state. This data has been uniformized to represent the state details.

6. Evaluation

a. Content Analysis

Word	Total
ubereats	428 [4.32%]
new	360 [3.63%]
york	289 [2.91%]
usa	202 [2.04%]
order	164 [1.65%]
food	149 [1.50%]
philadelphia	114 [1.15%]
delivery	109 [1.10%]
brooklyn	100 [1.01%]
support	100 [1.01%]
nyc	81 [0.82%]
just	80 [0.81%]
uber	80 [0.81%]
jersey	75 [0.76%]
florida	63 [0.64%]
service	62 [0.63%]
ordered	59 [0.59%]
time	56 [0.56%]
city	53 [0.53%]
app	53 [0.53%]
amp	52 [0.52%]
eats	50 [0.50%]
driver	48 [0.48%]
delivered	45 [0.45%]
open	45 [0.45%]
deliver	43 [0.43%]
queens	39 [0.39%]
bronx	36 [0.36%]
what	35 [0.35%]
missing	35 [0.35%]
restaurant	35 [0.35%]
customer	34 [0.34%]
got	33 [0.33%]
use	32 [0.32%]
hungry	32 [0.32%]
want	31 [0.31%]
grubhub	31 [0.31%]
manhattan	31 [0.31%]
hour	30 [0.30%]
people	29 [0.29%]
code	28 [0.28%]
money	28 [0.28%]
need	27 [0.27%]
today	26 [0.26%]
connecticut	26 [0.26%]
really	26 [0.26%]
georgia	25 [0.25%]
help	25 [0.25%]
waiting	24 [0.24%]
day	24 [0.24%]
guys	23 [0.23%]
call	23 [0.23%]
ordering	22 [0.22%]
place	22 [0.22%]
tonight	22 [0.22%]
first	22 [0.22%]
doordash	21 [0.21%]
lol	21 [0.21%]
cold	20 [0.20%]
drivers	20 [0.20%]
take	19 [0.19%]
philly	19 [0.19%]
way	19 [0.19%]
dinner	19 [0.19%]
til	18 [0.18%]
ave	18 [0.18%]
long	18 [0.18%]
stop	18 [0.18%]
account	18 [0.18%]
getting	18 [0.18%]
house	18 [0.18%]
fuck	17 [0.17%]
make	17 [0.17%]
know	17 [0.17%]
know	17 [0.17%]
back	17 [0.17%]
good	17 [0.17%]
street	17 [0.17%]
please	17 [0.17%]
free	17 [0.17%]
worst	16 [0.16%]
away	16 [0.16%]
hours	16 [0.16%]
eat	16 [0.16%]
minutes	16 [0.16%]
chicken	15 [0.15%]
shit	15 [0.15%]
refund	15 [0.15%]
used	15 [0.15%]
island	15 [0.15%]
phone	15 [0.15%]
think	15 [0.15%]
thanks	15 [0.15%]
better	14 [0.14%]
using	14 [0.14%]
mcdonald	14 [0.14%]
drive	14 [0.14%]
door	14 [0.14%]
come	14 [0.14%]
night	14 [0.14%]
fee	14 [0.14%]
issue	14 [0.14%]
ago	14 [0.14%]
lunch	14 [0.14%]
germantown	14 [0.14%]
newark	13 [0.13%]
two	13 [0.13%]
guy	13 [0.13%]
tell	13 [0.13%]
available	13 [0.13%]
delivering	13 [0.13%]

Figure 6.1 Keyword Frequency in Uber Eats Sample

Word	Total
grubhub	401 [5.04%]
new	283 [3.55%]
york	235 [2.95%]
order	162 [2.03%]
philadelphia	130 [1.63%]
usa	123 [1.54%]
delivery	92 [1.16%]
food	91 [1.14%]
just	79 [0.99%]
nyc	79 [0.99%]
brooklyn	72 [0.90%]
amp	71 [0.89%]
care	61 [0.77%]
city	56 [0.70%]
app	52 [0.65%]
down	50 [0.63%]
seamless	47 [0.59%]
time	43 [0.54%]
jersey	42 [0.53%]
ordered	40 [0.50%]
missing	39 [0.49%]
what	39 [0.49%]
philly	36 [0.45%]
ubereats	35 [0.44%]
first	32 [0.40%]
restaurant	31 [0.39%]
florida	30 [0.38%]
today	30 [0.38%]
use	28 [0.35%]
need	26 [0.33%]
deliver	25 [0.31%]
got	25 [0.31%]
queens	25 [0.31%]
call	25 [0.31%]
manhattan	24 [0.30%]
want	24 [0.30%]
night	24 [0.30%]
open	23 [0.29%]
got	25 [0.31%]
queens	25 [0.31%]
call	25 [0.31%]
manhattan	24 [0.30%]
want	24 [0.30%]
night	24 [0.30%]
open	23 [0.29%]
good	23 [0.29%]
bronx	22 [0.28%]
friday	22 [0.28%]
doordash	22 [0.28%]
uber	22 [0.28%]
hour	22 [0.28%]
ordering	22 [0.28%]
tonight	21 [0.26%]
place	21 [0.26%]
right	20 [0.25%]
ave	20 [0.25%]
cold	20 [0.25%]
late	19 [0.24%]
driver	19 [0.24%]
island	19 [0.24%]
delivered	19 [0.24%]
help	19 [0.24%]
service	18 [0.23%]
dinner	18 [0.23%]
bar	18 [0.23%]
georgia	18 [0.23%]
restaurants	18 [0.23%]
people	17 [0.21%]
know	17 [0.21%]
really	17 [0.21%]
sorry	16 [0.20%]
eats	16 [0.20%]
come	16 [0.20%]
orders	16 [0.20%]
customer	15 [0.19%]
orders	16 [0.20%]
customer	15 [0.19%]
code	15 [0.19%]
day	15 [0.19%]
bell	15 [0.19%]
pizza	15 [0.19%]
working	15 [0.19%]
make	15 [0.19%]
lunch	15 [0.19%]
guys	14 [0.18%]
big	14 [0.18%]
eat	14 [0.18%]
connecticut	14 [0.18%]
taco	14 [0.18%]
take	13 [0.16%]
minutes	13 [0.16%]
street	13 [0.16%]
thanks	13 [0.16%]
thank	13 [0.16%]
using	13 [0.16%]
hours	13 [0.16%]
special	12 [0.15%]
tried	12 [0.15%]
hello	12 [0.15%]
staten	12 [0.15%]
last	12 [0.15%]
available	12 [0.15%]
park	12 [0.15%]
united	11 [0.14%]
week	11 [0.14%]
grub	11 [0.14%]
states	11 [0.14%]
shit	11 [0.14%]
online	11 [0.14%]
lovash	11 [0.14%]
hey	11 [0.14%]
support	11 [0.14%]
cheese	11 [0.14%]

Figure 6.2 Keyword Frequency in GrubHub Sample

Location for Uber Eats and GrubHub

Philadelphia, Brooklyn, NYC, Jersey, Florida, queens, Bronx, Manhattan. When take a closer look at word frequency, word frequency mentioning New York is 730, and in GrubHub this number is 597. Based on our tweet sample, Uber Eats has a higher popularity.

Competitors

Customers kept mention other food delivery software in the tweet: In Uber Eats Content, people mention GrubHub and Door Dash. In GrubHub Content, people mention Uber Eats and Seamless. In the tweet sample, GrubHub and UberEats are usually come in pairs. Restaurants are giving promotions on both platform, and customers are using each other as a substitute.

Interesting Findings

Some words need attention in Uber Eats: driver, delivery, missing, refund, Mc Donald. These keywords convey several concerns of customers. Customers care much about their food delivery speed, and whether they can get their refund if they are not satisfied. The most popular food get mentioned on Uber Eats is Mc Donald, and the most popular food on GrubHub is Taco Bell and pizza.

It is reasonable for customers to use food delivery service as a fast food delivery. When you use Uber Eats, the delivery fee for fast food is the cheapest, from $2.99 to $3.99, and very often free delivery fees during promotion period. Fast food price is also comparatively affordable compare to other restaurants, and the food preparing time is also the shortest, only 15-20 mins preparation time.

Business can periodically report interesting findings and targeting existing problems in their delivery service. For example, since in our content analysis, customers pointing out their dissatisfaction of refund and delivery speed, than Uber Eats and Grub Hub can provide better refund policy for customers and explore methods to speed up delivery.

b. **Location Analysis:**

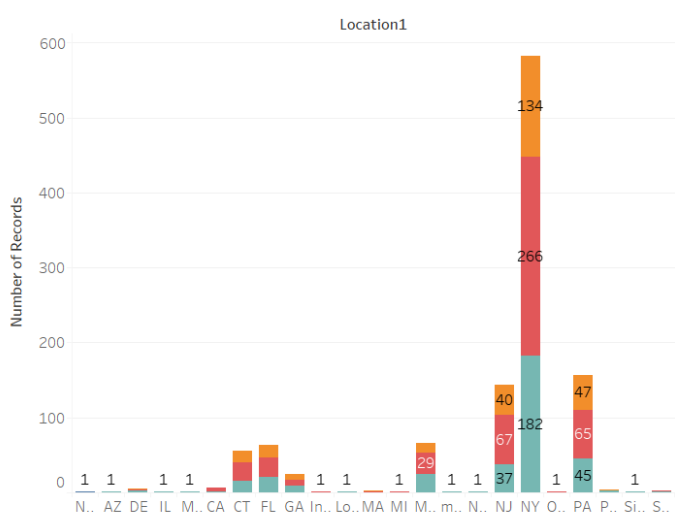

Figure 6.3 Sentiment Distribution by Location, Uber Eats[1]

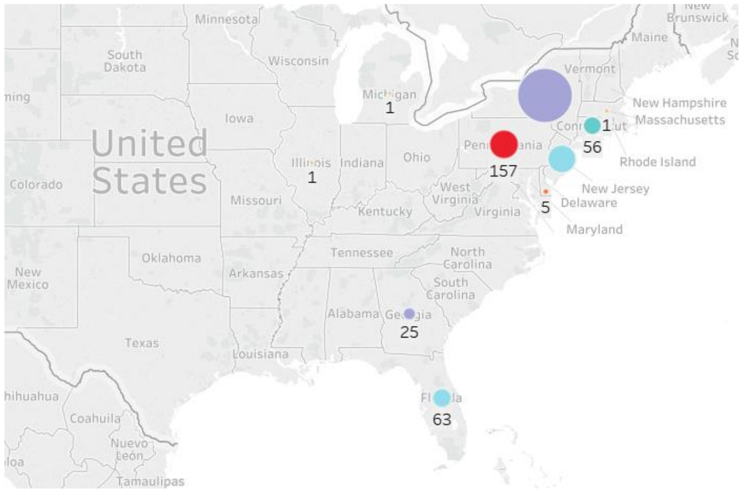

Figure 6.4 Number of Tweets by Location, Uber Eats

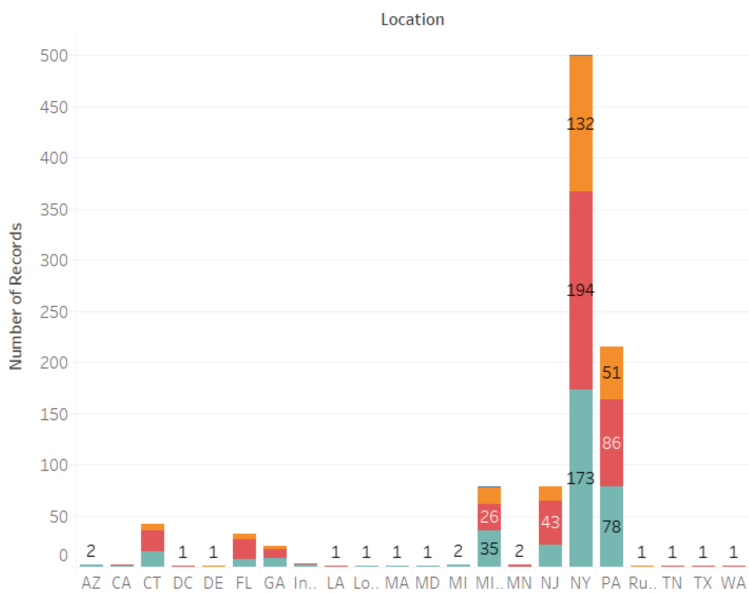

Figure 6.5 Sentiment Distribution by Location, GrubHub

[1] Orange as negative, red as neutral, blue as positive

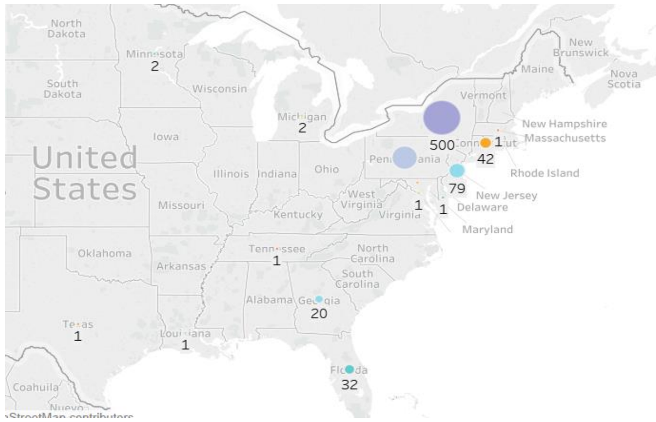

Figure 6.6 Number of Tweets by Location, GrubHub

Based on the above generated Tableau charts, we can identify that the majority of the Uber Eats users come from US Northeast, particularly from the New York metropolitan area.

Although most of the users come from the tri state area, there are a few outliers from southern states like Georgia, Florida and a few from Midwest.

Grub Hub follows a similar pattern as well with most of the users from around New York and surroundings.

As we have put the geo location as within 125 miles/ 200 km from New York, it makes sense that most of the users come from those location. But we explain the reason for users from South and other far off locations as due to them travelling to New York or surroundings at the time of their respective tweets.

According to our analysis, we further confirm that food delivery service is more popular in metropolitan cities rather than in surrounding county area. Businesses can try to design different promotion plans by location. Provide pop-up free delivery fees for customers in urban area to boost total number of orders, and promotions like $5 dollar off if order more than $25 in surrounding county area to increase the revenue per order.

c. Influencer Analysis:

- **UberEats**

Screen Name	Follower count
YahooFinance	830102
FOX29philly	592611
yo	228375
EXPandAMP	97613
cryptodemedici	92907
SniperZeroXI	67237
KYWNewsradio	53893
kmin	37896
NjTank99	31968
dberkowitz	31107

- **Grubhub**

Screen Name	Favourite count
BoobPunchTina	506397
cushbomb	292529
Sherlake	264180
ketaminh	253385
rosegoldmamii	220027
rosegoldmamii	220027
JJFan18	194681
dervogelfamber	165578
lifeisastory_	158290
106th	151468

Figure 6.7 Influencer Analysis

As we know, Social media marketing is the trend of the day. Advertising through influencers has gained tremendous popularity rather than just using traditional celebrities.

So, we have tried to identify who the influencers are in order to under stand the impact of their reviews. We have looked at their follower count to identify the top 10 most popular users who have reviewed each delivery service. Through sponsorships and brand collaborations with these influencers, we can better promote our product, special offers and discounts by reaching out to their massive follower strength effectively.

d. Word Cloud

Figure 6.8 Word Cloud

We have also input all the texts to generate Word Cloud for each service in order to better visualize the word frequency based on their weight, how often the words are mentioned in the tweets. It is visible that most users tweet about the same content.

7. Conclusion and Future Direction

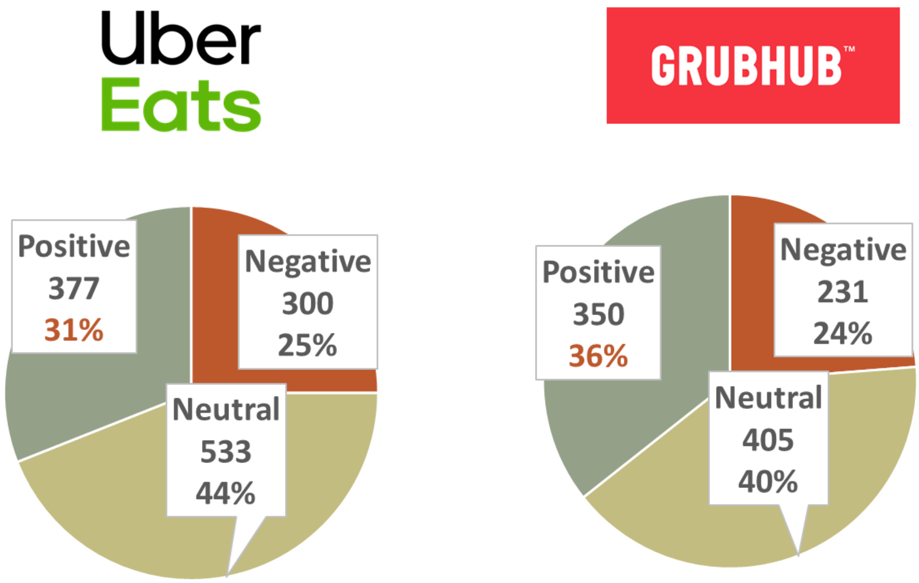

Figure 7.1 Sentiment Analysis Pie Chart

The Twitter sentiment analysis result indicates that GrubHub is more popular and well-received than UberEats, with a 36% of positive sentiment, compared to that of UberEats (31%). It is also worth noting that while the positive sentiment of UberEats isn't as high, the neutral sentiment takes 44% of all Twitter mentions.

As the industry matures, and competition intensifies, it is more important than ever for these businesses to execute on these core strategies to remain relevant. To do so, these digital platforms can leverage the vast wealth of first, and third-party data (i.e. social media, location tags, cookies) to create better user experiences, while driving operational efficiencies. For instance, negative customer feedback relating to delivery times on Twitter, can drive decisions to add optimal driver capacity in certain locales, and/or hours of the day. Similarly, Facebook's social graph data can be leveraged to better-identify local tastes and demands, informing companies on which new restaurants to onboard, in certain geographies.

To take our project a further step, we could focus on several things: First, perform a social network analysis for top influencers. We believe it would be interesting to find out the network relationship of influencers mentioning food delivery service. This will further help us providing suggestions to businesses. Second, our program is restricted by processing speed, therefore our number of tweets is limited in a short time. We are looking to find a better Python package in getting tweets to increase our sample. Last but not least, language restriction. New York is a very diverse city and we have huge number of residents tweeting using Spanish, Hindi, Chinese, Vietnamese, etc. We are looking to improve the language handling ability of our program.

8. References

https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/

https://labsblog.f-secure.com/2018/01/26/how-to-get-tweets-from-a-twitter-account-using-python-and-tweepy/

https://onlinehelp.tableau.com/current/pro/desktop/en-us/buildexamples_bar.htm

https://www.grubhub.com/

https://www.ubereats.com/

https://www.twitter.com/