Group 3 - Philly Food Delivery

Ming Yuen Thomas Cheong, Benjamin Jacobson, Christopher Sedayao
Northeastern University

Abstract

Most mature delivery restaurants lose money on their delivery orders. This is because of the premiums introduced by do-it-all delivery services like GrubHub or DoorDash. This project aims to leverage data that we get from our undisclosed partner in the logistics industry to provide insight and analysis on delivery cost efficiency. The goal is to enable restaurants to make better informed operational decisions that will be beneficial for maintaining profit margins in the delivery market. Restaurant owners do not necessarily have the time to dig into nuanced financial information, so our project is an attempt to provide insightful content at-a-glance. Based on our dataset of the Philadelphia delivery market, most restaurants have the ability to save significantly by outsourcing delivery to delivery-only logistics companies.

Link to Github Repository:
https://github.com/NEU-DS-4200-F20/project-group-3-philly-food-delivery

1 Introduction

The project will be about working with an undisclosed food delivery logistics company to leverage customer ordering habits to help restaurants/businesses make better informed operational decisions. Our partner is differentiated in the market by offering only delivery services, not a customer ordering interface. Our main focus will be to find trends in their data that will be beneficial for restaurants to convince them to use our partner instead of other competitors' delivery branches. Restaurants currently pay, on average, 25% of an order's total value to have it delivered by the larger delivery companies like DoorDash. This premium accounts for having a large delivery radius to customers. If restaurants are able to source delivery through other means, they can rather pay, on average, 5% of an order's total value. Hence, we see value in showing restaurants why they should consider alternative methods of fulfilling delivery, especially if their customer-base is located within close proximity to a given restaurant.

Given the direction to show why restaurants should switch order deliveries to our partners instead of other competitors, we decided that the main factor would be the cost savings each restaurant could realize. In order to deduce a cost saving analysis, we introduced information regarding distributions of order origin, such as Grubhub and Snackpass, and whether or not these order costs could be saved. We determined that these costs could be saved if they had competitor orders that were within our partner's delivery zones. This is because our partner has delivery zones which are delivery radiuses that our partner operates in. These delivery zones have a maximum of 10 minutes which allow our partner company to operate efficiently. With this idea, we can determine cost savings if restaurants are using competitors to deliver orders that are within our partner's delivery zones. To make cost savings analysis realistic, we are seeing how much savings restaurants can save given a conversion of 50% of competitor orders within delivery zones to our partner.

Link to Github Page:
https://neu-ds-4200-f20.github.io/project-group-3-philly-food-delivery/

Link to Demo Video:
https://github.com/NEU-DS-4200-F20/project-group-3-philly-food-delivery/blob/gh-pages/DSproject.mp4

Link to Presentation Slides:
https://github.com/NEU-DS-4200-F20/project-group-3-philly-food-delivery/blob/gh-pages/Presentation%20Slides.pdf

2 Related Work

The paper on food delivery routing decisions [4] covers an optimization model for delivering hot (mid-day) meals primarily to schools. It uses computational analysis and a visualization tool to visualize and describe its models which we can use as a reference. Several studies [1], [5], [9] cover different possible reasons why people order food online. These articles go further to examine the relationship between attributes like customer attitudes, experiences, and behavioral intentions in relation to delivery services. Some other attribute relationships that those articles discussed include convenience factors, an individual's hedonic motivation, price saved, time saved, and consumer attitudes. This can guide us in describing trends we find in our data. Next, Jaiswal [6] depicts that smart delivery allows for customers to receive orders without human interaction. It is especially useful to see the impacts of smart delivery systems on delivering items given the effects of COVID-19. The study on "Optimized Food Delivery Network Based on Spatial Crowdsourcing" [7] investigates to see if taxis can support the demand from online food orders. One of the variables they were analyzing is to minimize cost while minimizing distance traveled which is relevant to our partner's operational "milk run" model. Similarly, Correa [3] evaluated online food delivery and it's traffic conditions by analyzing performance metrics. This can be helpful given that delivery time may be a key attribute we investigate. Relating to driver routes, Steever [8] analyzed models in which a single customer may be able to order from multiple restaurants at the same time. This is relevant as our partner demonstrated interest in a "milk-run" model that can be efficient for the company going forward. Lastly, Chen [2] and Zhao [10] argued that the advancement of technology has changed food delivery businesses drastically. Chen's study [2] focuses on food delivery services and how traffic conditions may impact performance on the delivery service. This will be relevant as our partner focuses on customers within proximity and we can see if there is some correlation between distance and performance.

We did change our final state of the project throughout the course. In the beginning, we wanted to look at the trend of delivery drivers as well to provide insight for our partner company but decided to just focus on restaurant insights. A lot of the visualization from the research articles were complex and did not fit our target audience so we went for simple graphs given that our partner stressed to keep visualizations clean and simple. Our project related to articles that described how technology has changed food delivery businesses and how contactless deliveries have changed the restaurant market. Given the impact of COVID, our project focused on restaurant transactions within 2020, we focused on how restaurants were losing money by relying on big-box delivery services to source and deliver food to customers. This project adds to some projects in the related work as it dives deeper into restaurant's profit margins and how it relates to the means of completing the delivery service of an order.

3 Partner

Our partner has requested to remain undisclosed. They are a food delivery logistics company with operations in various primary and secondary cities. They offer specialized, cost-effective logistics solutions for restaurants. Their motivation for this project is to find out how they can relay to restaurants the cost saving benefits of leveraging their logistics service in comparison to all-in-one delivery services.

3.1 Interview Reflection

We enjoyed meeting our clients and learning about their company, goals, and how we can collectively put out something that mutually benefits both of us. There was a mutual excitement for the project and it seems like we will be able to put something together that is leverageable by the intended user. During the call, we narrowed the scope down into two focuses: How can we better enable the undisclosed company to better pay their drivers, or how can we provide useful, actionable data to partnered restaurants in regard to delivering orders. Throughout the discussion, we learned about our partner's business model, their interests in terms of data visualizations. We were most surprised by the way that they think about solving issues. They are very methodical and they leverage their company data effectively. This interview solidified our motivating question, along with our partner, we think we can make interesting insights regarding how restaurants can make better operating decisions.

4 Data

We have been granted access to a variety of data related to consumer ordering habits in the Philadelphia area. This will not include locations of customers but rather distances from customers to the restaurant. Similarly, we have worked on a solution to anonymize sensitive customer and restaurant data. The dataset contained a couple hundred thousand rows of basic transaction data. It included information such as the order was created, the cuisine type, order distance in miles, order total, etc… While there were hundreds of restaurants, we only wanted to focus on the top 10 restaurants based on their order volume to see if there is a trend there. We thought this small sample size would be enough to showcase overall trends.

5 Task Analysis

Task ID #	**Domain Task**	**Analyze Task** (high-level)	**Search Task** (mid-level)	**Analytic Task** (low-level)
1	Restaurant insight - Compare the distribution of proportions of orders being delivered by competitors	Discover	Locate	Compare
2	Restaurant insight - Examine customer distances from restaurants	Discover	Explore	Identify
3	Restaurant insight - Identify the distribution of orders within partner zones	Discover	Explore	Identify

Our 3 domain tasks seek to find relationships regarding the proportion of orders from delivery services per restaurant, the order distances for each restaurant, and whether or not these orders are within our partner operating zones. These tasks will help guide our project to determine if trends exist for our project value proposition.

Our visualizations will primarily be developed to communicate and present information because we are trying to create valuable insights for our partner to use in their daily operations. Our restaurant insights visualizations will communicate to restaurants why they should join our partner and not another logistics company.

Our primary consumer of these visualizations will be for both restaurants as well as our partner company. As explained above, these restaurant insights will convince restaurant owners to switch the logistics side of delivery from other delivery companies to our partner. It will also allow our partner company to look into restaurant insights to make better decisions to market themselves.

6 Execution & Design Process

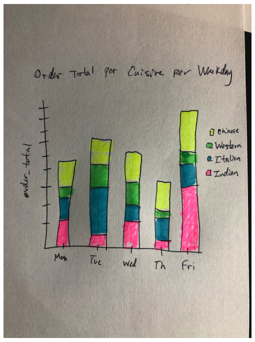


Figure 1: Order Total per Cuisine per Weekday Stacked Bar Chart

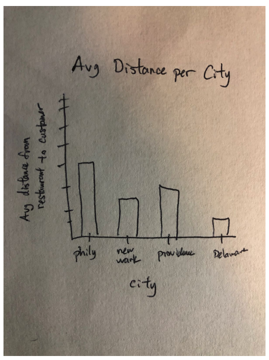


Figure 2: Average Order Distance per City

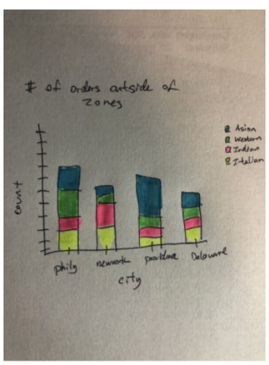


Figure 3: Number of Orders outside of partner zones

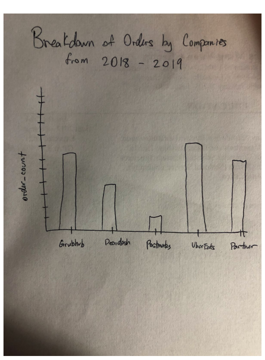


Figure 4: Breakdown of Orders by Order Origin

The 4 figures shown above were initial design sketches we created. While our final visualization looks drastically different, similar ideas can be seen transcended down. Figures 1 and 2 were initial drawings trying to convey order distance as well as the number of orders. They were initially implemented to show the number of orders by each cuisine for each weekday as well as the average order distance per city. We ended up deviating from them to only show data from the top 10 restaurants based on order volume to avoid too much clutter and confusion. Figures 3 and 4 were our initial sketches of showing the distribution of orders outside of our partner zones as well as the distribution of order origin for total orders. Similarly, we were able to simplify it into two pie charts to make it easier to read and compare. The pie charts also allow readers to understand the distributions as a whole. Our final visualizations reflect our partners' demand of making it simple. They emphasized that simple visualizations are vital given the busy schedules of restaurant owners.

Our greatest concern for our visualizations were from our pie charts. In theory, pie charts are not the best form of visualization because humans are bad at interpreting angle. However, with the addition of pie chart percentages, we think we were able to mitigate the downsides of using pie charts. Furthermore, it is an easier representation to exemplify our claim that restaurants should switch to our partner company given how large delivery companies dominate the market share while 93% of orders from the top 10 restaurants are actually within our partner company zones.

For all visualizations, we used color encoding to differentiate between different aspects of the data to differentiate them for the user. For example, color was used in both pie charts to separate the different fields in comparison. For the two bar charts, we added tool-tips as well as color change during hovering to help usability. The color change reinforces which bar is hovered while the tool tip gives the user more information for the bar.

The main feedback we were given during usability testing from classmates were some confusion regarding what cost savings are, how the top 10 restaurants were chosen, separating large numbers with commas, lowering the saturation of colors in the bar chart, and that the categorical colors in pie charts should be the same when mousing over restaurants. These were all minor changes that we were able to address. We added the description of what top 10 restaurants were, what partner zones were, and what cost savings mean above the visualization to help guide readers before they dive into our visualization. Next, we were able to separate large numbers with commas, lower the saturation of our bar charts, and provide consistency between categorical colors in the pie chart for each order origin. It was very important to get feedback from our classmates and our professor as they provide a fresh set of eyes to a project we have been working on for the semesters. While some things may seem intuitive to us, given that we worked on this project almost every day, it may not be as easy to understand for new readers. Overall, we had a well thought out interactive visualization with no bugs and only minor fixes to clean up.

7 Visualization Design

For our visualization design, we decided to implement two bar charts as well as two pie charts. The bar charts help visualize the top 10 restaurants and their related order volume data as well as their average distance (in miles) for all orders. We added hover-linking through both bar charts that will change the color of the hovered bar to show the relation between the two charts. It will also link the same restaurant data together between the two bar charts. There is also a tool-tip included for the user to find the specific order volume or average order distance for that specific restaurant when hovering. Given that information, we implemented details on demand pie charts for the bar the user ends up hovering over. The pie charts help distinguish the distribution of order origin (grubhub, snackpass, etc...) as well as the distribution of orders outside delivery zones for that specific restaurant. If no bars are hovered, then the pie charts represent data for all 10 restaurants, whereas, if a bar is hovered, only that specific restaurant data is shown. To elevate our project further, we decided to include a cost reduction aspect to our visualizations. The cost reduction/saving represents the 20% premium restaurants pay for competitors to provide logistics. The number shown is what restaurants can save if they convert competitor orders to our partner company. In order to make our findings realistic, our cost savings calculation assumes that restaurants can convert 50% of competitor orders within our partner's delivery zones to our partner. Similar to the pie chart, the cost savings is a detail on demand where if no restaurant bar is hovered, it represents total cost savings by all 10 restaurants.

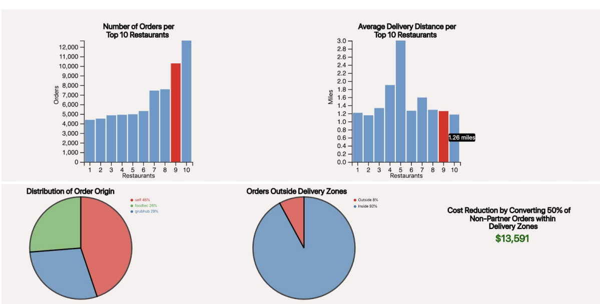


Figure 5: Current state of visualization

Our final visualization turned out exactly as what we had planned in our digital sketches as shown in appendix D, except with the addition of the cost savings analysis. The top row of the visualization with the two bar charts are interactive whereas the bottom row with the two pie charts and the cost savings analysis will change based on the interaction on the top row. We had this set up as the top row visualizes basic restaurant information whereas specific details show up in the second row. We did this in order to make this usable for busy restaurant owners - quick, insightful details on demand. We also implemented some minor UI fixes from the testing done by our peers.

8 Discussion

The biggest takeaway from this visualization is that restaurants have the potential to protect their profit margins if they can better understand their customer geography. Our first major finding comes from the distribution of order origin pie chart. All 10 restaurants primarily use Grubhub as their delivery partner. This pie chart shows how dominating other providers are in terms of taking in delivery orders. Our second finding is that the orders outside delivery zones pie chart show that all these restaurants that primarily use grubhub as their delivery partner can save significantly, as the majority of their orders come from within our partner's own delivery zones. In fact, 93% of total orders from the top 10 restaurants come from within our partner's delivery zones. If restaurants were able to convert just half of delivery orders to be fulfilled by our partner, these restaurants can be saving about 20% of each order's total value. Given what our visualization shows, we hope that this benefits restaurants into making more effective operating decisions.

9 Conclusion

We have been able to create a final visualization for our partner that started off as an idea. We met with our partner multiple times throughout the intermediate steps of our project to understand and get a better sense of a problem they would like to solve. Once we were able to finalize a direction, we began our data analysis with the dataset they provided to see if we could make initial inferences or find trends within the data.

One key aspect that our partner wanted to know was how to exemplify why restaurants are better off joining our partner rather than their competitors. Given the rows of transaction data above, we decided that the most impactful data we could use to explain why restaurants should switch to our partner company is to show cost savings for restaurants if they were to switch their delivery logistics to our partner. While there were hundreds of restaurants, we only wanted to focus on the top 10 restaurants based on their order volume to see if there is a trend there. We thought this small sample size would be enough to showcase overall trends. We made initial sketches of possible visualizations and iteratively improved on them as the project progressed. Once we were able to finalize our visualizations, we created the same visualizations on our website.

We believe that our visualization will convince restaurant owners to switch to our partner company given the cost savings they will be able to realize. While our current stage of the project is sufficient to exemplify why restaurants should switch to our partner, we realize that we can make improvements in the future to further our visualization. Our partner organization has logistic services in different cities so it may be valuable to see if similar trends are happening outside of Philadelphia. In addition, it may also be important to analyze actual restaurant profits given that they do switch to our partner company. This will help solidify our claim that restaurants are better off switching to our partner in this case given that 93% of orders within the top 10 restaurants based on order volume are within our partner zones.

10 Acknowledgements

We would like to acknowledge the guidance and support from professor Cody Dunne as well as David Saffo throughout the project. We would not have been able to create our final project without their feedback and guidance through visualization best practices as well as the implementation of interactive visualizations. They have been a pivotal part of this project and we appreciate everything they have done.

8 References

[1] Chandra, Yakob Ultama, and Cadelina Cassandra. “Stimulus Factors of Order Online Food Delivery.” *Stimulus Factors of Order Online Food Delivery - IEEE Conference Publication*, 2019. doi: 10.1109/ICIMTech.2019.8843715

[2] Chen, Li-Wen. “Impact Assessment of Food Delivery on Urban Traffic.” *Impact Assessment of Food Delivery on Urban Traffic - IEEE Conference Publication*, 6 Nov. 2019. doi: 10.1109/SOLI48380.2019.8955108

[3] Correa, Juan C., et al. “Evaluation of Collaborative Consumption of Food Delivery Services through Web Mining Techniques.” *Journal of Retailing and Consumer Services*, Pergamon, 29 May 2018. doi:10.1016/j.jretconser.2018.05.002

[4] Dayama, Niraj Ramesh, and Mohan Krishnamoorthy. “Facility Location and Routing Decisions for a Food Delivery Network.” *IEEE Xplore Login*, 2016. doi: 10.1109/IEEM.2016.7797843.

[5] Hingle, Melanie, et al. “Collection and Visualization of Dietary Behavior and Reasons for Eating Using Twitter.” *Journal of Medical Internet Research*, JMIR Publications Inc, 24 June 2013. doi: 10.2196/jmir.2613.

[6] Jaiswal, Rahul, et al. “Smart Solution for Reducing the COVID-19 Risk Using Smart City Technology.” *Smart Solution for Reducing the COVID-19 Risk Using Smart City Technology - IET Journals & Magazine*, 3 Aug. 2020. doi: 10.1049/iet-smc.2020.0043

[7] Liu, Yan, et al. “FooDNet: Toward an Optimized Food Delivery Network Based on Spatial Crowdsourcing.” *FooDNet: Toward an Optimized Food Delivery Network Based on Spatial Crowdsourcing - IEEE Journals & Magazine*, 2018. doi: 10.1109/TMC.2018.2861864.

[8] Steever, Zachary, et al. “Dynamic Courier Routing for a Food Delivery Service.” *Computers & Operations Research*, Pergamon, 26 Mar. 2019. doi: 10.1016/j.cor.2019.03.008

[9] Yeo, Vincent Cheow Sern, et al. “Consumer Experiences, Attitude and Behavioral Intention toward Online Food Delivery (OFD) Services.” *Journal of Retailing and Consumer Services*, Pergamon, 29 Dec. 2016. doi: 10.1016/j.jretconser.2016.12.013

[10] Zhao, Yuyang, and Fernando Bacao. “What Factors Determining Customer Continuingly Using Food Delivery Apps during 2019 Novel Coronavirus Pandemic Period?” *International Journal of Hospitality Management*, Pergamon, 9 Sept. 2020. doi: 10.1016/j.ijhm.2020.102683

9 Group Charter

9.1 Group Purpose

Through this partnership, our group's purpose is to provide valuable insights to restaurants who would leverage the data from our undisclosed company. This project shares the duality of the company being better equipped to present the benefits of their logistics solution to potential restaurants, as well as allowing restaurants to make better informed operational decisions based on consumer habits.

9.2 Group Goals

The goal is to leverage data that we get from our partner and use this opportunity to provide insight and analysis to help restaurants/businesses to make better decisions that will be beneficial for their company. We also want to learn to make quality visualizations that are rid of biases that convey valuable insights that can help guide companies forward. We will have two main goals throughout our project. (1) To provide visualizations to help business owners make more informed operational decisions. (2) To lend insight into consumer behavior in the online food delivery market.

9.3 Group Member Roles/Responsibilities

Benjamin Jacobson will be the communications director for this project. He has connections with our partner and has already been in communication with them and will continue doing so for the remainder of the project. Thomas Cheong will be the meeting facilitator and will keep each other on tasks and set reminders for everyone. He will check in to make sure the project is going smoothly and will keep track of deadlines. Christopher Sedayao will be the note taking liaison who will be in charge of taking notes during meetings, sending notes to Thomas, and will also be the person who will double-check the project to make sure everything is set up correctly. Overall, we will all be accountable for looking over the project deliverables to make sure we are on the right track and will be responsible for our own contributions. We will all bring up any difficulties that we face and will do our best to figure it out before reaching out to the group.

9.4 Ground Rules

Our group will meet once per week on Sunday over Zoom. Every week we will discuss what we have done in the previous week, then what we will be doing that week, and finally if we have any potential blockers preventing us from getting work done. We will conduct discussions in a group context with everyone being free to chip in ideas and suggestions. Everyone on the team is fairly comfortable with each other and don't believe we will have any issues with differing or dissenting opinions. We will hold each other accountable by having deliverables due before each meeting every week. We expect to give full participation and commitment to this project and a willingness to get the job fully done.

9.5 Potential Barriers and Coping Strategies

The largest potential barrier we're anticipating is the inability to work in person. With Covid-19 cases rising nationally, we don't anticipate much in-person interaction. We all live in Boston within 15 minutes from each other, so there is definitely the possibility of in-person collaboration if we deem it safe enough. With all computer science projects, there is a significant advantage to work in person, so we will gauge as time goes on. Another barrier we anticipate is having to manage time with interviewing for co-ops. The three of us are all actively searching for our final co-op, so will need to be flexible with each other when interviews come around. Despite these two significant barriers, the three of us are all friends from N.U.in 2017. We are very comfortable with each other in a working and casual environment, so I think that if issues arise, we will all feel the necessity to approach each other and discuss pain-points quickly and effectively.

9.5 Project 5 - Check-up

We have all been abiding by our agreed-upon guidelines. We are all responsible for what our own works and have been accomplishing them in a timely manner while updating each other throughout the process. We are all comfortable with our group roles. While we each have our own roles, we work together cohesively throughout as a single unit and do not confine each other to our designated roles. There have not been many problems throughout this partnership. We have been working well and have enjoyed working together.

Appendix A: Interview

Our partner's target users are themselves as well as for other restaurants who have not joined our partner's company. Our partner will be able to leverage runner/delivery driver insights to better inform them of future hires. Also, restaurants will be able to use insights to visualize why joining this food delivery logistics company is better than other competitors. Their mission is to help restaurants by driving more direct volume and save on margins as big delivery companies have large fees. Our partner company has a lot of data to work with. Due to this, we can conceive the direction of this project, and then have the dataset fields "fit" to the project spec. The following attributes are data that we will be using, with further fields to be discussed soon: order value metrics, cuisine type, order timing, average order volume, driver performance metrics, driver neighborhoods, and deliveries by sequence. A major motive for our partner in having visualizations is to see how they can make their platform and food delivery as a whole more sustainable for restaurants and drivers. Our partner understands that food delivery as a system right now is not efficient. There isn't a direct solution to this but having some visual insights may help our partner with seeing the effects of their business model and the potential for expansion. Two key insights our partner will be looking for in the data is by getting valuable information regarding restaurant and driver insights. Our partner has already completed various data analysis and visualizations before. To list some of many:

1. Understand order source breakdown
2. Customer heat map
3. Top 50 customers and where they're ordering
4. Monthly order source
5. Average deliveries per day by week
6. Total potential profit by source
7. Amount paid to other logistics companies(anonymous) in the last x months

<u>Notes:</u>

2 focuses for the partner → drivers and restaurants

<u>Interested in restaurant owners</u>

- Too occupied to look AND analyze data
- Restaurant owners busy → need easy visualizations → little to no user input
- Want them to be doing as little as possible
    - I.e. What percentage of your customers order x, y, and z?
- Want something **simple** so that an average restaurant owner can understand
- Easy ways to see how their customers are behaving

<u>Interested in the drivers</u>

- How can this be sustainable for them?
- Have applicant zip codes → where best/successful drivers come from?
- Sufficient driver supply or market supply

<u>They have applicant zip codes</u>

- Where do our best drivers come from
- Where are they coming from in other markets(?)
- Regionally adjusted/market adjusted
- Sufficient driver supply and sufficient demand
    - This is the best market for them
    - From their data and public census data
- How to prioritize who to target → density of routes → high frequency customers and market to them?

<u>Client wants economic justice for drivers and restaurants</u>

- Goal: economically equitable decisions

<u>Client wants something that definitely is impactful</u>

- Not just a school project

<u>Data (columns to consider)</u>

- Which restaurants have highest AOV (average order value)
- Cuisine type
- Order frequency
- Deliveries by sequences
- They use the milk run model
    - Operations model for minimal cost at higher efficiency
    - Milk runs = delivery will pick up 10-15 meals and delivery them in a sequence and for the right restaurants, it works well
    - If restaurants can guarantee delivery volume, like a subscription, it can change how restaurants operate
    - What foods travel well (indian food)
    - Mumbai dabbawala (example)
    - If can inform them which restaurants can do this
    - Right areas with pent up demand

High level (tasks)

Runner supply insight → funnel best runners

- Best runners
    - Performance metrics (obviously)
    - Neighborhoods
    - Are men or women better
        - Non intuitive
    - Batch frequency

Restaurant insight → case study of how habitat is better and how to save money with Habitat

- Key fields
    - Average order value
    - Customer order frequency
    - Cuisine type
    - Prep time(?)
        - Biggest problem with food delivery
    - Delivery fees
        - How it affects customer behavior
    - Tips
        - How that varies by cuisine or customer type
    - Time of day

- Food delivery right now is really broken so it may be cool to see how to make it sustainable for restaurants, runners, and the whole ecosystem and what are key things to visualize. Batch economics and where orders go

Another possible goal

- How to make food delivery sustainable?

Appendix B: Data Exploration

Data Types

The following columns in the dataset are categorical:

Driver data

- id (driver_id)
- name (location_name)
- age
- current_stage (rejected or accepted)

Restaurant data

- created_at(when order was created)
- cuisine _type
- Daas_type (payment type ie. credit, cash...)
- day_name (weekday of order)
- habitat (geographic radius to which certain restaurants are contained, max 10 mins for diameter of zone)
- Out_of_zone (is order outside of habitat zone)
- Partner (where the order came from ie. Grubhub)
- Territory (city)

The following columns in the dataset are quantitative:

Driver data

- Completed_orders (driver_completed_orders),

Restaurant data

- Assigned_to_arrive (minutes between dispatcher assigns runner and runner arrives at restaurant)
- Duration_mins_a2b (minutes from restaurant to customer)
- Distance_miles_a2b (distance from restaurant to customer)
- Order_total
- Prep_time (prep time estimated by kitchen)
- Requested_to_assigned (minutes from delivery requested to dispatcher assigned)
- Tip (driver tip)
- total_delivery_time

Our partner gave us two csv datasets. One pertains to driver data. This includes information such as location, age, accepted or rejected application, and the number of completed orders. The other csv dataset pertains to restaurant data such as order total, prep time, cuisine, order day of the week, payment type, time to complete delivery etc. Both datasets were generated organically through the operations of our partner company.

Potential Issues

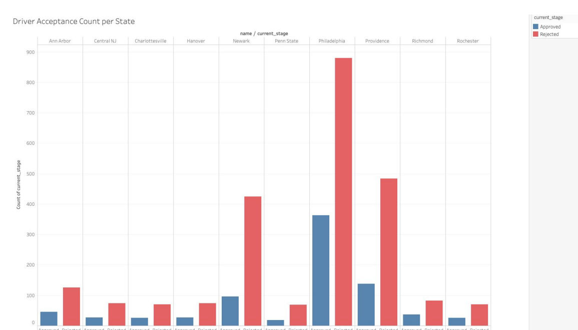
Some potential issues regarding our dataset are duplicated columns and some missing data within columns. We have just under one million rows of data, so management of the dataset could be cumbersome. The other area we are raising is with the driver dataset. There seems to be fewer identifiers of a “good” or “bad” driver than we anticipated, so this is something we are discussing currently with our partner.

Insights

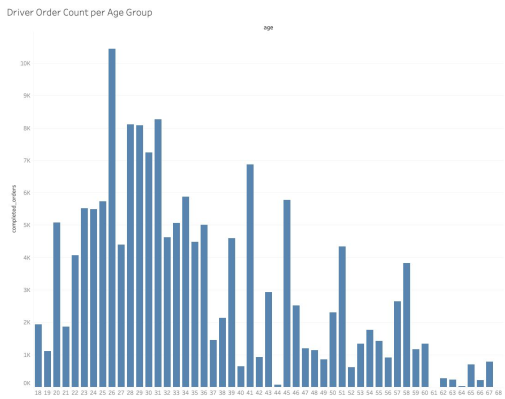
For driver data, we explored the following visualizations: Driver order count per state, driver order count per age group, and driver acceptance/rejection count per state. There were no surprising trends given that our partner’s entry market was in Philadelphia so the delivery order count should be highest in that region and the acceptance/rejection count should be highest there too. Next, there is a unimodal graph that is skewed to the right given the delivery order counts per age group which is also expected given that we expect less drivers orders as they get older.

For the restaurant data, we explored the following visualizations: Cuisine volume by day, order volume per habitat by territory, order cost to tip given, delivery time by habitat. I think one thing we all found really interesting was the cyclicality of customer ordering patterns. From Monday to Sunday, the order volumes in each category exponentially increase. Once Monday comes again, the orders drop back down and repeat the cycle.

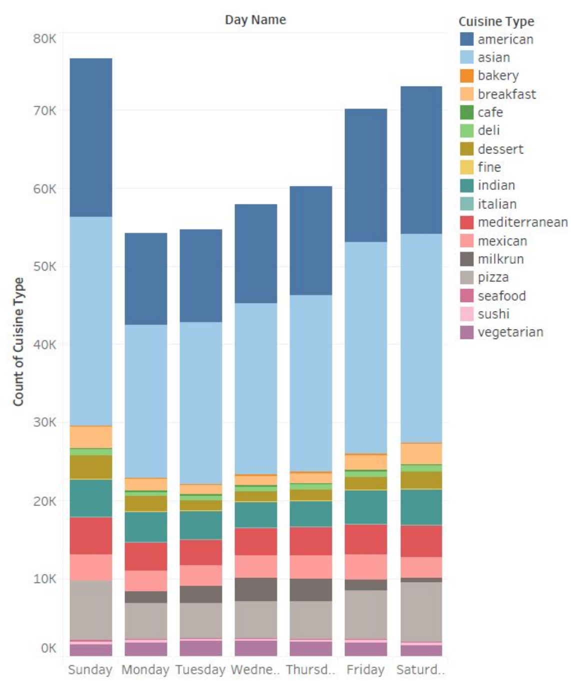
Screenshots



Visualization 1: Driver Acceptance/Rejection Count per State. The data we were exploring was relating to driver data using the driver state, and driver acceptance/rejection. We then used the count of accepted/rejected drivers as the y-axis. The visual encodings used were color and grouping. Color was used to distinguish between accepted and rejected bars and grouping was used to distinguish between the different states. The trend this visualization shows is that for each state, there is nearly double the number of rejected drivers than there are accepted drivers.



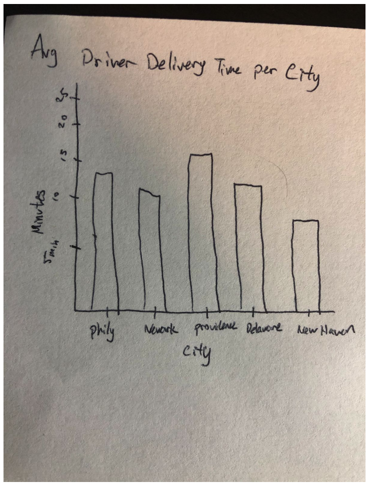
Visualization 2: Driver Order Count per Age Group. The data we were exploring was related to driver data using driver age, and driver orders completed. The visual encoding used is just grouping for grouping the drivers in the same age groups together. The trend this visualization shows is that the distribution of order deliveries for each age group is heavily weighted towards younger drivers.



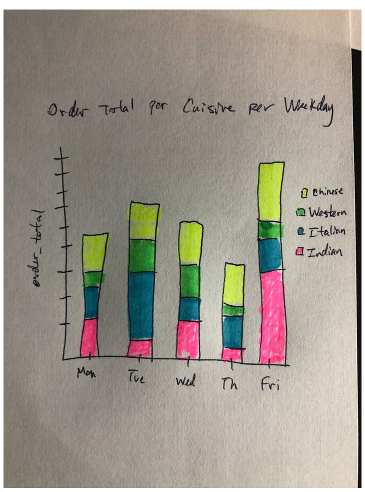
Count of Cuisine Type for each Day Name. Color shows details about Cuisine Type. The view is filtered on Cuisine Type, which excludes Null.

Visualization 3: Order volume by day, partitioned by cuisine type. The data we explored was the count of each order type, and the day of week that it was ordered on. The visual encoding is grouping, as we grouped the orders, and their types, into the day of the week. We used a color encoding to differentiate between the cuisine types, as indicated in the legend. The trend in this visualization shows that there is a cycle in consumer ordering patterns - as the week progresses from Monday to Sunday, general order volume increases.

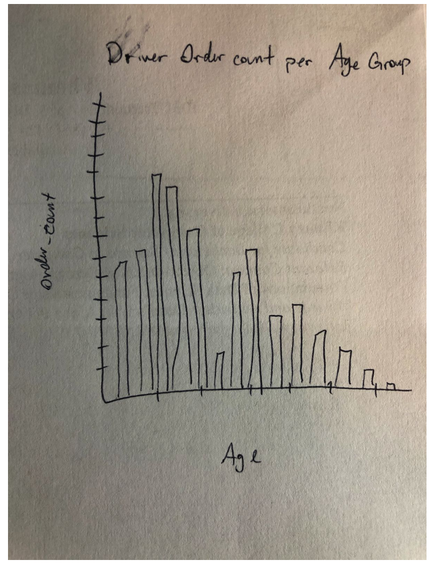
Appendix C: Design Sketches



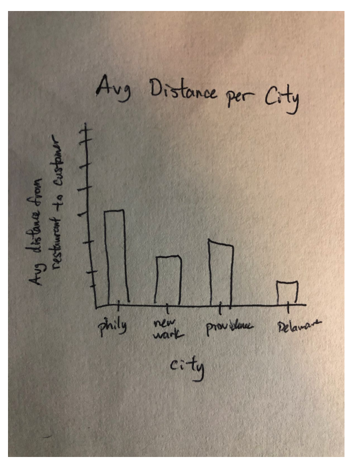
Artist: Thomas Cheong
Related Task: 3
I wanted to uncover, at a high-level, which cities have market potential. This is key for restaurants trying to enter the market as it enables an understanding of the potential benefits of faster ordering services. A bar chart was most fitting for this chart as it lends to understanding the total order time (in minutes) to get from point A to point B. I used lines/areas as a marking to represent the bar charts as I thought it would be easy to compare bar charts per city. The channel used was horizontal and vertical for the positioning of the bar as well as length and area to portray the height for each bar.



Artist: Thomas Cheong
Related Task: 2
I wanted to figure out how each cuisine performs per weekday and I thought it would be best represented as a stacked bar chart. If this was represented as regular bar charts would be too clustered. The marks used are lines and areas given the use of the bars to represent each data. The channel used is both horizontal and vertical for each bar as well as color to differentiate each cuisine section within a bar. Lastly, the size of length and area was used to represent the height of each bar.

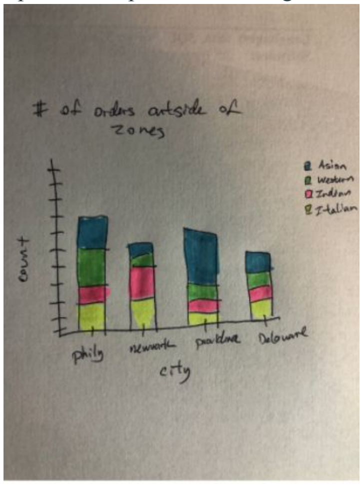


Favorite 1
Artist: Thomas Cheong
Related Task: 1
This visualization is to help our partner understand what characteristics of their drivers make them successful. As a preliminary concept, I wanted to see if there was a relationship between age and orders delivered. I decided to use a bar chart as it best fits representing the total orders by age group. The mark used here would be lines and areas to represent each bar. The channel used would be both horizontal and vertical for each bar as well as length and area to represent each bar's height.

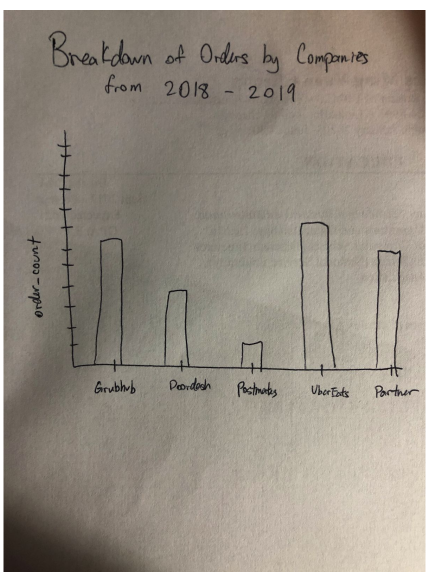


Artist: Christopher Sedayao
Related Task: 2
In order to get restaurant and runner insight on distances for their various cities, I chose to use average order distance from restaurant to customer as quantitative data and cities as categorical variables in a bar chart. The marks used are line and area to represent each bar and positioning of

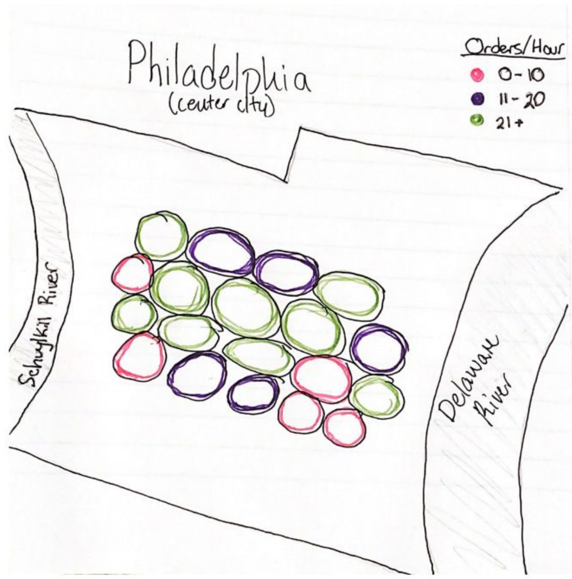
horizontal and vertical along with length and area to represent the position and height for each bar.



Artist: Christopher Sedayao
Related Task: 2 and 3
I want to use a stacked bar chart with the number of orders outside of each habitat per cuisine type by city in order to help restaurants and runners see how far they have to go to deliver their food. Also, breaking down what sorts of cuisine types by color gives both runners and restaurants specificity in their work. The marks used are lines and areas to represent the bar. Channels used are both horizontal and vertical as well as length and area to represent the position and height of the bar. Lastly, color is used to differentiate the different cuisine types.



Favorite 2
Artist: Christopher Sedayao
Related Task: 2
Breaking down which logistics group our partners' companies leverage assists in seeing who is currently dominating the market and how our partner can help their restaurants adjust to a better pricing model for logistics. Using a bar chart is an easy and intuitive way to break down the proportions of the competition as it is easy for the user to compare bar heights. The marks used are simply areas given the pie chart visualization. Channels used are color, tilt, and area to represent each different section within the graph.



Favorite 3
Artist: Benjamin Jacobson
Related Task: 2
I chose to do a heatmap on Center City Philadelphia. Each circle represents an ordering zone. For each ordering zone, I chose one of three colors to represent the average order volume per day. This will help restaurants understand how their proximity to a zone may influence order volume. Zones are preset sizes, so there is no encoding within each circle. For total attributes: one mark (point of the circle on the map), and two channels (color to represent the order volume, shape to represent the area evaluated).

Artist: Benjamin Jacobson
Related Task: 3
This is a pointed line chart uncovering the average tip received by a driver, by every week day. I chose a line as my means of visualization as it lends to the general ordering pattern of customers - more orders get placed as the week goes by. This can potentially be developed into a

model that can be locationally generated and based on different times of each day. For total attributes: two marks (points and lines representing how tips fluctuate throughout the week) and one channel (relationship between Tip and Day of Week, both vertical and horizontal).

Artist: Benjamin Jacobson
Related Task: 2
Understanding the impact of faster delivery is very important to the top line. When restaurants have the ability to cater to their closest customers, they're able to capture revenues more effectively. I chose to create this pie chart in order to show the proportion of orders completed within 10 minutes, differentiating between the groups by color. For total attributes: one mark (area) and three channels (color, tilt, area).

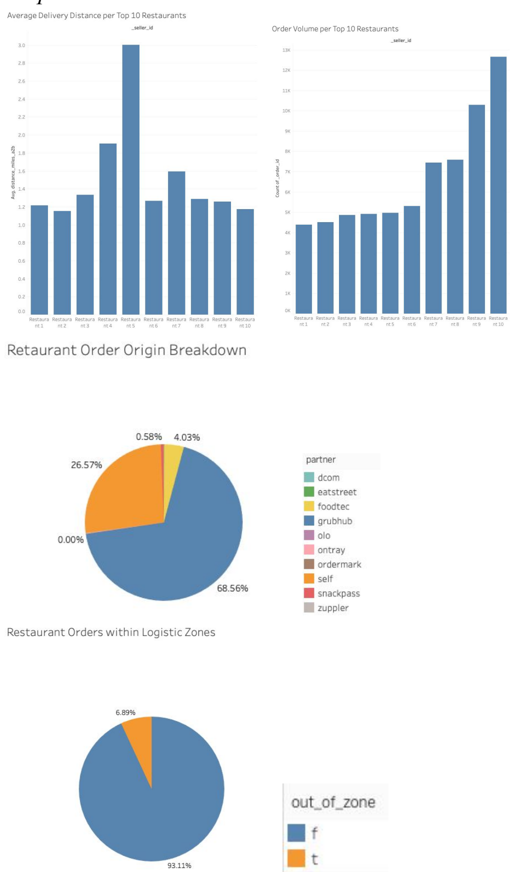
Favorites
1. (*Refer to Favorite 1 Visualization*) We chose this visualization as one of our favorites because it explores one attribute that may correlate with what makes a good delivery driver. By looking at the age group distribution and the number of orders delivered in each group, it can help inform our partner company which age groups may be better than others. While this is one attribute that may help, we will be exploring other factors as well but this graph is a step in the right direction.This visualization is appropriate because we are comparing age group categorical data so a bar chart will allow us to see variations between each group.

2. (*Refer to Favorite 2 Visualization*) We chose this visualization as a favorite because it will be useful in comparing the direct competition of our partner. By comparing the amount of orders with other competitors, our partner can use this visualization to see which competitors are dominating the market. In our future designs, we will be implementing an option to make the bar charts stacked to show the distribution of orders by city for each competitor. This data is useful for our partner as they can see how much restaurants are being charged by other competitors. From there, they can reach out to restaurants and adjust their operations.

3. (*Refer to Favorite 3 Visualization*) We chose this as a favorite as it really can evolve into a useful resource for restaurants. When looking at orders per area, at a high-level, it shows basic customer traffic flows through each area. At a more nuanced level, this chart can be incorporated into a chain of visualizations that can show, more granularly, what cuisine types are chosen most, what areas have the most order volume from different vendors, and potentially, what food items are ordered. These would be key indicators for a restaurant wanting to enter the order-by-internet market, or even a restaurant wanting to know how to navigate their online sales more effectively. I think that this visualization would also be incredibly useful because of its simplicity. Restaurant owners don't want to have to find their own insights - they just want items to act on. While this visualization isn't necessarily for information-at-request, it can be developed into a tool that accomplishes that goal.

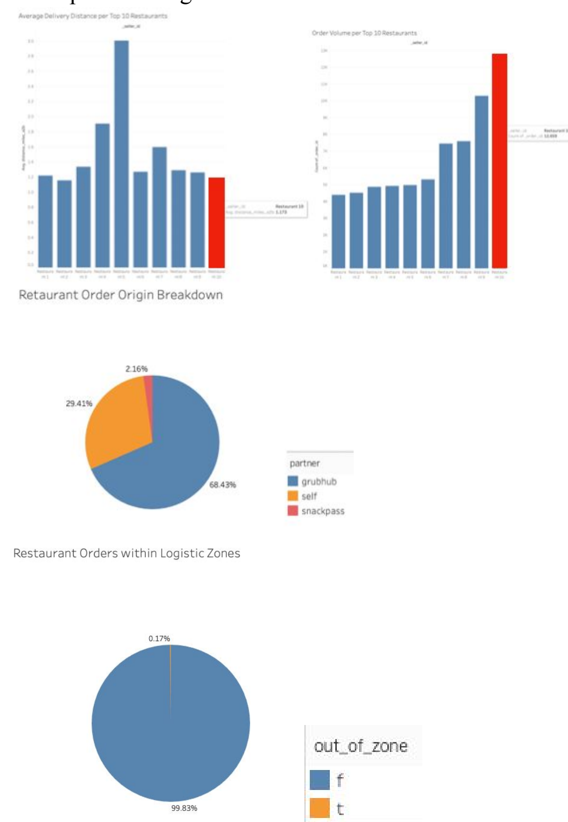
Appendix D: Digital Sketches

Graphs Without Interaction



Graphs Without Interaction:

Hover/brush over either bar chart to reveal specific restaurant data (barchart data) along with two data on demand pie charts updating the breakdown of restaurant order origin as well as their percentage of orders in or out of our partner's logistics zones.



Our visualizations address the first two task analyses. We chose to prioritize the first two tasks as those would help the restaurants best understand their customers' location, which therefore allows the restaurants to better choose logistics services. If their customers' distance is within a delivery zone, but they are using our partner's competitors (UberEats and Grubhub), their profit margins will be diminished due to the premium they have to pay competitor services. We have discussed with potential end users, and this strategy seems to be very impactful operationally. Therefore, we have shifted from our runner data insights to be primarily restaurant focused. Our final digital sketches have also differed from our initial sketches given our need to change focus. We have also changed our task analyses to better fit our restaurant insights and the limitations of the data set.

Appendix E: Reflections

Reflection - Benjamin Jacobson

Throughout the entirety of this project, our group's communication was excellent. We all held each other equally accountable for keeping up with the progression of the project, making communication very easy. We usually held calls on Facetime or through Zoom, meeting two to three times a week. With our partner, we've maintained a great connection, communicating with them through email, slack, and Zoom. I think we all wanted this to be successful, so we were all working together effectively. One thing that could have been improved was if we had more concrete communication with our partner early on. While we had 2-3 calls with them, it would have been good to get more progress done with them early on.

Reflection - Thomas Cheong

The communication process throughout the project was through zoom and iMessage. It has been going well by pair programming throughout zoom and working with each other through screen sharing and working through problems at the same time. There is not much that could be improved upon as we are all working together at the same time and we work cohesively together. One minor thing that could be improved possibly is to split up work but that may create problems such as merge conflicts.

Reflection - Christopher Sedayao

The communication was very thorough and well done throughout the project. We used iMessage to schedule meeting times and Zoom for the actual meetings. We worked together on the project through pair programming and constant collaboration. We have used Slack and email to communicate with our partner outside of meetings which are also on Zoom. I thought we worked cohesively and smoothly on this project with no real conflicts. I think one thing that could have been improved on was having everyone join every meeting, but due to differences in schedules, this did not always happen. If someone did miss time, we updated that person through text or in the next Zoom meeting.

Appendix F: Slides

Link to presentation slides:
https://github.com/NEU-DS-4200-F20/project-group-3-philly-food-delivery/blob/gh-pages/Presentation%20Slides.pdf