Here is the list of questions. The list is divided by topic and is also divided up into 5 sections, so we each can pick one.

Everything in red is speculative; we are not likely to find answers in the data set, but we can think about them.

There are questions at the bottom that I thought were a little too advanced to do for exploratory data analysis. They seem more like questions to improve the business rather than just looking at the data. We can plan on looking at these later, but you guys can also do these now if you have time.

I made comments about people’s questions I didn’t understand.

**Person 1 – Laura**

* Customers
  + Where are the customers located?
    - Are they concentrated in certain cities/zip codes?
    - Why are there more customers in these areas? Is anything done different in these cities with more sales? If yes, what?
  + What type of areas are the zip codes in?
    - Are they residential? Commercial? Industrial?
  + How many returning customers?
    - If not, why don’t they return?
  + Who are the customers that haven’t purchased in the longest time?
    - Do we know why?
  + Who are the top customers?
  + Do we have average number of customers in the areas where there is highest shipping?

**Person 2 – Sumit Sapra**

* Products
  + What are the most expensive and least expensive products?
    - How does the price relate to the amount sold? –
  + What product categories are sold the least and most? – pie chart
    - Are product sales related to the seller? – correlation table and reg plot
  + Do certain cities tend to purchase certain product categories more?
  + Are particular products ordered at certain times?
    - What can we infer from that? Is it due to specific events, seasons?
  + How are product descriptions and numbers of photos having an impact on their sales?
  + Can we find evidence of more sales of a category in the culture of the city?

**Person 3**

* Orders
  + What periods of time (or dates) do we see more orders? (Can evaluate in terms of month, season, etc.)
    - Are certain items ordered more during certain periods?
    - Is this peak period associated with a special event that could justify higher sales?
  + About how long does it take to deliver products?
    - Is that related to customer satisfaction?
  + What is the relationship between purchasing and wait times?
    - Do the lowest purchasing cities have the longest wait times?
    - Why are they the lowest purchasing cities?
  + What is the average transaction value?
    - Total revenue / number of transactions. This metric gives you a general idea of how much people are spending. A high dollar amount could mean that shoppers are purchasing your more expensive products or they’re buying larger quantities.

Having a low average dollar per transaction could indicate that you need to rethink your pricing

* + Analyze relationship between order delivery dates reported by carrier vs. reported by customers, vs. estimated. Why the discrepancy? Could it be hurting?

**Person 4 -- Meng**

* Reviews
  + What items/orders have the highest reviews?
    - Who sells those items?
  + What is the highest review score? The lowest? The most common?
  + Who are the most unsatisfied customers?
* Payments
  + Do customers who use a certain payment spend more than other customers?
  + What is the main payment type? (credit card)
  + What are the payments installments in the data? (average installment numbers ) What do they suggest about most customers’ habits?
  + What is the relationship between payment value and installments? (correlation)\_
    - Do installments encourage larger orders?

**Person 5**

* Seller – Olivia
  + Which sellers make the most money?
    - Why? Is it related to pricing? How does their pricing compare to other sellers?
    - Where is the top seller located?
    - Which sellers have the most orders?
  + Which sellers do not perform well?
  + Where are the sellers located?
    - Which city has the most sellers? Which city has the least sellers?
    - Which state has the most/least sellers?
* Customer satisfaction
  + Clustering analysis of customer segments – We have just covered that topic in class so I am not sure how to go about it but I will look online. If anyone has an idea please let me know!.

**Future problems to think about**

* Questions for later:
  + Is our model for the estimated delivery correct? What is the error rate?
    - Can we improve it?
  + Does Olist want to offer a store credit card?
    - <https://www.thebalanceeveryday.com/which-retail-stores-have-credit-card-offers-2892559>
  + What is the prefect order rate?
    - Product delivered on time to the buyer's requested delivery date; shipped completed; invoiced correctly, undamaged in transit.
  + Can we plot out last mile delivery efficiency?
  + What are the areas we can invest in hub lockers or any self-served locker facilities?
  + Based on the sellers and customers states, zip code, latitude, longitude, and cities, can we decide what locations are best to invest in cross dock distribution centers?
  + Should Olist own their own logistics or outsource?
    - * <https://www.ebayinc.com/stories/news/ebay-to-launch-managed-delivery-an-end-to-end-fulfillment-service-for-sellers/>
  + Reverse logistics process