ORIE 3120
Spring 2021
Homework 4
Due Monday, March 15 at 11:59pm

Submit a pdf to Gradescope. For all queries included, make sure that they can be copied and pasted.

This assignment builds off of this week's recitation (Recitation 4) and uses the same dataset.

1.   Map of UPC 1093 Sales by ZipCode (15 points)

Create a map of sales for UPC 1093 by zipcode. This should be the same as the one you produced in the recitation assignment, except that you will add a filter to include data for this UPC only. You may need to convert UPC to a dimension (and not a measure) to tell Tableau that you want to treat it as a categorical value.

a. Include a screenshot of your map, including the Filters box.
b. What is different about the geographical distribution of sales for this collection of parts from the general sales distribution from recitation?
    [Hint: compare the two maps to where the larger cities are. To see where larger cities are, you can use background maps, whose layers can be modified under Maps > Map Layers in the menu bar. For instance, you can add layers such as cities and major highways, while removing layers such as terrain to enhance visibility. One of the maps has a larger fraction of its sales in areas away from big cities.]
c. Include a few of the part descriptions of parts in UPC 1093. [Hint: Click on a circle on the map and then select the icon with tooltip "View Data". This allows you to view both a summary and the full underlying data table for the points on the map.] What 3-word abbreviation do you see repeated in most of these parts descriptions? What do you think this stands for? [Hint: Use google.]
d. Explain why these types of parts might have a different geographical sales distribution. [Hint 1: People in rural areas drive trucks more than people in urban areas, and also tend to drive longer distances and carry heavier payloads. Hint 2: Larger vehicles that drive longer distances and carry heavier payloads tend to have hotter oil. See, for example, this URL: https://www.hardworkingtrucks.com/engine-oil-coolers/.]
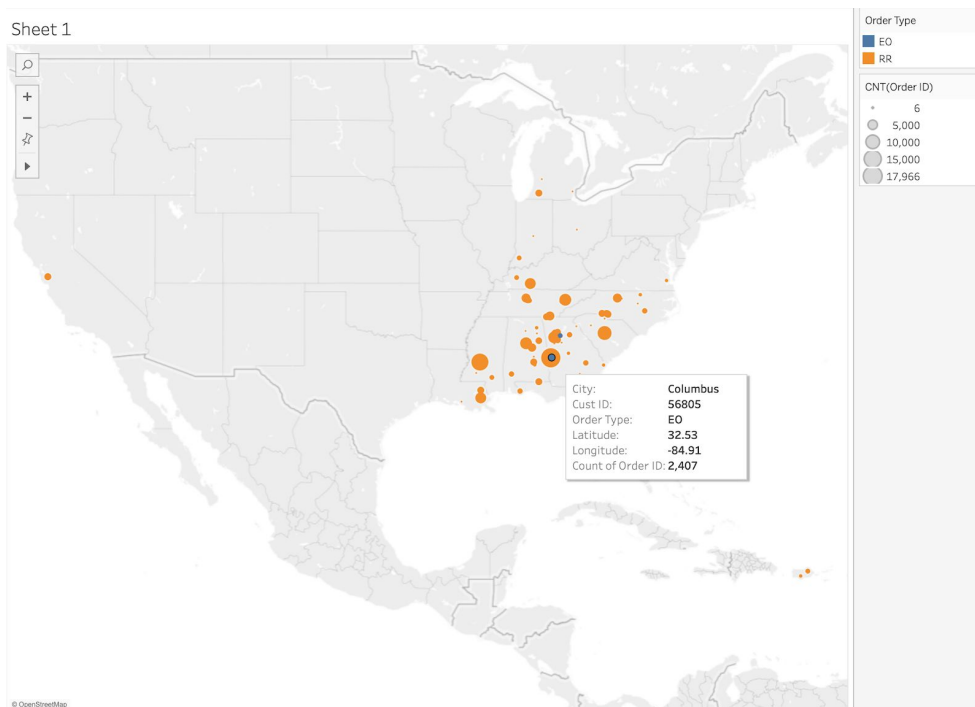
2. Understanding Emergency Orders with Tableau (10 points)

A leading source of costs for the carmaker are emergency orders for parts. These are the orders with type "EO". This is the type of order that a dealership places if they need to get the part very quickly. They are expensive to fulfill because the need to deliver quickly usually prevents combining EO orders together and means that the truck making the deliveries cannot combine the trip with deliveries to other nearby dealerships. In contrast, regular replenishment (orders

with OrderType "RR") are placed several days before the dealership needs the part, and they can be combined together so that a truck can deliver several RR orders together to dealerships that are close together, saving money.

Use Tableau to calculate the number of orders with type EO. Also calculate the number of orders with type RR.

Then use Tableau to make the following map

Each circle indicates a dealership (customer) and a type of order and has a size that indicates the number of orders of this type placed by this dealership. The "detail" associated with each point should show the city, customer ID, OrderType, latitude, longitude, and the number of orders of this type placed by this customer.

The map only includes those customers whose total sales value across all orders is $100,000 or larger.

Include a screenshot of your Filters and Marks boxes along with your map.

Observe that customer 56805 in Columbus is placing a substantial number of EO orders. This is likely to be expensive. Rather than placing a lot of EO orders, dealerships that have a large sales volume should invest in a place to store parts that they can use in most situation when they need parts, and that they restock using RR orders.

3. Understanding Emergency Orders with SQL (20 points)

This builds on the previous question. While Tableau is great for many things, some data analysis is hard to do purely in Tableau Desktop. In situations like this, you can use SQL. To do this question, the data from the zip files you used in Recitation 4 have been placed into an sqlite database called HW4.sqlite. This database is nearly identical to the csv's used in the recitation, except that some names of the fields and tables have been tweaked to remove spaces. This makes them easier to use in SQLite since you don't need to remember to use square brackets.

a. Create a view, Q3a, that produces the same 545,370 records that you produced in Tableau in Recitation 4.

To do this, use inner joins to join together customers, orders, zipcodes, states, and parts. Join on the following conditions, which are the same ones you used in the recitation:
  - customers.Zip = zipcodes.Zip
  - zipcodes.State = states.StateAbbreviation
  - customers.CustID = orders.CustID
  - orders.PartID = parts.PartID

Include the following fields in your view:
  - OrderType
  - CustID (from either customers or orders, since they will be the same)
  - SalesValue, which is equal to the product of orders.quantity and parts.price

It would be easy to include more fields, but we'll only need the ones above for our analysis.

For part (a), include the query that you used to create your view and a screenshot of the first few records. If you like, you can run SELECT COUNT(*) FROM Q3a to confirm that you get 545,370 records.

b. Create a view, Q3b, that calculates the number of EO orders placed by each dealership. It should return two fields: CustID and NumEO. If a dealership did not place any EO orders, it won't appear in the view's results.

Include your query and a screenshot of the first few records.

c. Create a view, Q3c, that calculates the total number of orders placed by each dealership and their total sales value. It should return three fields: CustID, NumOrders, and TotalSalesValue.

Include your query and a screenshot of the first few records.

d. Create a view Q3d that joins together the results from Q3b and Q3c to produce, for each customer, the following records:
  - TotalSalesValue, i.e., the sum of the sales value of all orders placed by this customer

- NumEO, i.e., the number of EO orders. If the customer placed no EO orders, this should be 0.
- NumOrders, i.e., the number of orders
- FractionEO, the fraction of orders that are EO orders, rounded to two digits of precision

Include your query and a screenshot of the first few records.

Hint: Because Q3b doesn't include a record for every dealership, you'll need to use a left join and think carefully about NULLs.

e. Write a SELECT statement that returns all of the records from Q3d that have TotalSalesValue larger than 100,000, sorted in decreasing order of FractionEO.

Include your query and a screenshot of the first few records.

You should see that dealer number 56805 has 100% of its orders placed as EO orders. This is the same dealer that we noticed was problematic in question 2. What the parent company would do next is to talk to this dealership to let them know that they need to change their business practices. They would also talk to other dealerships high on the list for part e to understand if EO orders could be reduced.

4. Coverage of Pilot Distribution Center in Birmingham, AL (15 points)

Suppose we locate a pilot distribution center in Birmingham AL (zipcode=35216). We want to figure out how many dealers, and how much sales, are within 150 miles of this location.

Use Tableau to create a map of those dealers within 150 miles of Birmingham, AL, with circles at these dealers whose size indicates the total dollar value of the dealers' sales. [Hint #2: You may assume that 1 degree of latitude corresponds to 69 miles, and 1 degree of longitude at a latitude of lat corresponds to 69 * cos(lat) miles, where cosine is computed assuming that its input is in degrees. For details, see this URL. You can hard-code cos(lat) using the latitude of Birmingham, since cos(lat) varies only a small amount over the zipcodes in the dataset.]

Tableau functions that you might want to use include: POWER, SQRT, and (optionally) COS. If you use Tableau's COS function, keep in mind that it assumes its input is in radians. You can read about these functions at help.tableau.com.

Include in your answer a screenshot of your map and the formulas inputted into Tableau for the calculated values in your pdf.

Here's an example of a correct map: