Assignment 2: Analytic or Machine Learning Model

Assignment 2: Analytic or Machine Learning Model

- Please share your completed assignments with Salaar & Thomas: s.liaqat@mail.utoronto.ca ; thomas.rosenthal@utoronto.ca
- Submissions can be in the form of a zip folder, Dropbox link, Google Drive link, etc
- Due on Saturday, January 28 at 11:59pm
- Weight: 40% of total grade

Q1

Create an analytic or machine learning model using our farmersmarket.db as the source of data.

For either, start with a question you want to answer. These are examples to get you thinking:

- Have sales for a vendor increased/decreased over time?
- Do some products sell better on rainy/snowy days?
- Will a customer who has purchased from a vendor return to that vendor and purchase again in the next 30 days?

If you choose to make an analytic model, imagine that I have just taken over the administration of the farmers market but don't know any SQL. What do I need to know to be a good administrator?
Focus on the following:

- Clean your data: e.g. if you have nulls, they should probably be coalesced.
- Include at least two tables (I'm not sure how you'd manage with fewer anyways). No one wants to remember IDs, so you should replace them.
- Include a performance metric associated with various time periods (e.g. years, quarters, etc).
- Discuss: Is the data model sufficient? Do we need other tables? Could we collect other data that might be helpful? If so, how would we gather this, and how would we store it (i.e. describe new tables in a logical or physical model)?
- Describe: How does your analytic model help me run the market? What insights can I gather from it?

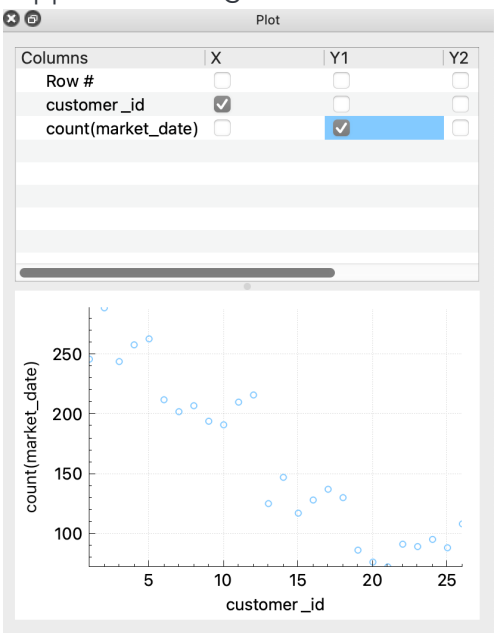
If you choose to make an ML model, pick a simple algorithm. This assumes you have some comfort in R or python. There is no expectation that you produce a statistically significant or highly accurate model.
Focus on the following:

- Make at least one "feature" in SQL that serves your question (e.g. CASE statement, windowed function, etc).
- Include at least two tables in the query (I'm not sure how you'd manage with fewer anyways). IDs don't make for good features, so you should replace them.
- Report the accuracy or significance of your model.
- Discuss: How would you monitor this data? Can you collect "ground truth" data for your model? If yes, how would it flow back into SQL, and where would we store it (i.e. describe new tables in a logical or physical model)?
- Describe: Is your model supervised? Unsupervised? What type of output variable are you trying to solve for? Why did you pick this algorithm and why was it appropriate for the problem you were trying to describe?

Q2

Produce some sort of graphic from your SQL dataset.

If you have no experience making graphics, keep it simple (e.g. count by a variable). SQLite can support this using the **Plot** window. See below:



If you feel more adventurous, use a tool like ggplot (R), seaborn (python), etc. If you have access to Excel or Tableau or another BI platform, export a csv from SQL and build there.

Describe why you picked this graphic. What does it tell us about your dataset?

Q3

Reflect: How can knowing SQL help you in your day-to-day work/life? Highlight a couple of different ways. This doesn't have to be strictly professional -- feel free to think about SQL in your personal life, if applicable.