Mini-Projects 2 and 3: Driving Decisions with Data

Read the whole description before starting any work.

Only use sqlite3 and python3 or higher for your code. Any other flavors or versions will not be accepted.

1. Identify a dataset you would like to work with. This must be non-normalized. You cannot pick an already normalized dataset.
2. Identify a stakeholder who must make decisions using your dataset. This is whom you will be addressing in this component. You will write a report which the stakeholder will use to drive their decisions.

2.1. For instance, a retail dataset from an E-Commerce outlet would yield insights that help a CEO or Product Manager to make decisions.

3. Normalize the dataset.

3.1. Detail your normalization and adherence to up to 3NF in your report's technical appendix. The technical appendix is a separate section which talks about all the technicalities of your work. This is not the same as the "body" of your report. The body is the section which your stakeholder peruses - they don't necessarily have to know about technicalities.

4. Ask yourself any five questions about the dataset that the stakeholder would want to know about: eg: year-on-year sales, or median monthly sales or any such metrics which have a sound logic to driving a decision. For instance, lowering yearly sales of sports goods might prompt a change in marketing, or dropping that category altogether.
5. Find answers to these five questions using SQL queries written using SQLITE3.

(Note: At least two of these queries need to use joins.)

6. Corroborate your answers by implementing the same queries, on the same tables using pandas in Python. Pandas also defines a join functionality - you will have to leverage it.

—-Explain your queries in your report's technical appendix. —-

7. For the body of the report, in three pages or less (as long as you convey your insights, the length doesn't matter - just keep the writing under three pages), write in non-technical language, your findings, and the decisions that can potentially arise from them. These should be addressed to your stakeholder - as in - the body of your report should be something that reaches the stakeholder's desk and helps them make decisions.

These are all your deliverables for Part 1 of the Mini-Project.

For part -2- follow all instructions of mini-project 1 *with the following twist*.

Write a public facing version of a story/journalistic piece which talks about your five findings from above, with visuals supporting your findings produced using python.

Note: The visuals do not have to be from the same query - as long as they support the same line of thinking, the visuals are acceptable.

This piece must be published on medium.

Extra credits for sharing on LinkedIn, just as earlier.

Submission link for both parts will be posted on Piazza soon.