CS 378 Project 7, due Thursday, 10/30.

Objectives

Build the mart layer of the lakehouse

In-Scope

Data marts

Out-of-Scope

Dashboards and other visualizations

Work Items

- Decide on the personas of the business users for your lakehouse
- Come up with 5 user journeys that your target personas would want to use your lakehouse for
- Design and implement 5 data marts such that:
 - Each mart is tailored to the target persona and enables them to make a business decision
 - Each mart is implemented with a summary query that is comprised of joins and aggregations
 - The marts should exhibit broad coverage of your lakehouse data with at least 2/3 of your int tables being used as sources across the 5 marts
 - Make note of data quality issues and identify future work items to resolve them

Code Samples

See snippets repo for code samples

Implementation Details

- Create a new folder in your repo and name it project7. Store your artifacts for this project in the project7 folder.
- Create a Colab notebook named project7-data-marts.ipynb.
- Annotate your notebook with section headers and short Markdown comments throughout your work.
- Import the provided <u>notebook</u> into your Colab and follow the same steps.
- The data mart queries should be read-only and formulated on top of your int layers. **Do** not mutate your int layer.
- Create a [your_domain]_mrt dataset in BigQuery and materialize the query results as new tables in this dataset.
- Choose descriptive names for your marts with table names and column names all in lower case. As usual, use an underscore character between words (instead of hyphens).
- Optional: you can choose to make an ERD for this layer, but it is not required because data marts are not expected to be connected. A more common way to document this

- layer is with a data dictionary. Please create a data dictionary if you have extra time, but I'm keeping this item as optional to keep the scope light.
- Create a submission.json file and upload it to Canvas by the deadline. Only one person per group needs to do this step.

CS 329E Project 7 Rubric

Due Date: 10/30/25

7-[your-domain]-mrt.ipynb is thorough and meets all requirements -10 for each empty or missing data mart found in the mart dataset -10 marts are too narrow and don't span across 2/3 of intermediate tables -7 for each missing mart that is missing a user journey -7 for each mart created without the use of joins or aggregation -5 for each mart that is missing future work -5 notebook lacks Markdown annotations and is hard to follow -5 did not follow naming convention for dataset, tables or columns -100 missing file	100
<pre>submission.json submitted into Canvas. Your project will not be graded without this submission. The file should have the following schema: { "commit-id": "your most recent commit ID from Github", "project-id": "your project ID from GCP" }</pre>	Required
Example:	
<pre>{ "commit-id": "dab96492ac7d906368ac9c7a17cb0dbd670923d9", "project-id": "some-project-id" }</pre>	
Total Credit:	100