CS 329E Project 5, due Thursday, 03/06.

In this project, we will build the mart layer of the warehouse, a layer that sits on top of the intermediate layer and takes the intermediate tables as input. The purpose of the mart layer is to serve dashboards and reports that help business leaders understand the health of their business so they can decide where to make investments. The mart layer is arguably the most important one in the warehouse because it is user facing.

The mart layer consists of one or more marts. Each mart is a table or view that is designed to meet the reporting needs of a specific business unit.

Methodology

- Imagine the personas of your end users and what reporting needs they may have that can be answered by your warehouse. Come up with 3-5 business questions that will drive the design of your marts. Document your business questions in a new notebook.
- Build 10 marts to address aspects of your business questions. Certain questions may translate into only one mart while others may be covered by multiple marts.
- Each mart should come from a query that comprises multiple intermediate tables joined together and aggregated.
- The mart layer should be built from at least 2/3 of your intermediate tables and should include all of your data sources.
- **Do NOT mutate the intermediate area**. All mutations should be applied only to the mart area of the warehouse.

Implementation Plan

- Create a Colab notebook that implements 10 different marts. Name your notebook `5-[your-domain]-mrt.ipynb`.
- Create the marts in a new dataset in BigQuery. Name your BQ dataset `[your-domain]-mrt` where `mrt` is short for mart. For example, air_travel_mrt.
- Document your business questions as short Markdown comments in your notebook using section headers.
- Document any non-trivial design choices, especially if you came up with broad business questions that you want to explore from different angles.
- As you craft your mart queries, make note of any outstanding data issues that you encounter (e.g. missing values, incorrect values, etc.) as Markdown comments in your notebook.
- Choose descriptive names for your marts.
- Use lowercase letters for table names and column names in this layer.

- Create a new folder in your repo and name it `project5`. Store your notebook for this project in the `project5` folder. You do not need to create an ERD.
- 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 5 Rubric
Due Date: 03/06/25

`5-[your-domain]-mrt.ipynb` is thorough and meets all requirements **-10** for each empty or missing mart found in the mart dataset **-5** for each missing business question **-7** for each mart created without multiple joins or aggregation **-7** for each missing data source **-7** marts don't cover 2/3 of intermediate tables **-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

`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"`
`}`
 	
Example:
 	
`{`
`  "commit-id": "dab96492ac7d906368ac9c7a17cb0dbd670923d9",`
`  "project-id": "some-project-id"`
`}`
	Required

Total Credit:	**100**