CS 329E Project 4, due Thursday, 02/22.

In this project, we create a new layer in our database for consuming the modeled data. This layer, known as the consumption layer, takes as input the final tables from the staging layer and extends their schema to model historical changes to the data. Although our dataset currently includes only active records, we will simulate changes in a future project. The concept of capturing changes is a common practice in data warehousing and is known as slowly changing dimensions of type 2.

Objectives

- Review your schema design with Prof. Cohen during today's class (02/16/23)
- Make the suggested changes based on Prof. Cohen's feedback
- Create the tables in the consumption layer
- Update the ERD and data dictionary to reflect the tables in the consumption layer

Implementation Guidelines

You can start developing the consumption layer prior to your design review. Once the review takes place, implement the action items from the review and publish the changes to your repo.

Please follow these guidelines when creating the consumption layer:

- Tables should be stored in their own dataset in BigQuery. The name of the dataset should follow the convention **[domain]_*csp*** where [domain] is the name of your data domain and ***csp*** is short for consumption. For example, airline_csp.
- Tables should be created in a notebook named **csp-layer.ipynb**.
- To model change history, the schema of each table should be extended with three fields:
  - **effective_time<TIMESTAMP>** set to the current timestamp for all records.
  - **discontinue_time<TIMESTAMP>** set to null for all records.
  - **active_flag<BOOL>** field set to True for all records.
- The naming convention for tables and columns remains the same as in the staging layer.
- Update the ERD and data dictionary from the staging layer to capture the new fields in the consumption layer. Denote the logical relationships between entities with dotted lines.
- Publish to your repo: **csp-layer.ipynb, erd-csp.pdf**, and **data-dict-csp.xlsx**.

CS 329E Project 4 Rubric
Due Date: 02/22/24

`csp-layer.ipynb` has all required information and has implemented suggestions from design review	60
ERD diagram accurately depicts relations between the csp tables **-5** for each missing important link **-5** for each missing staging table **-10** ERD not aligned with data dictionary columns **-20** missing file	20
Data dictionary has all important information about csp tables **-2** for each missing column of staging table **-5** missing description column **-10** missing file	20

`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**