

CS 327E Milestone 3 due Sunday, 02/16.

1. Create an ERD of the staging tables in your `dataset1`:
 - Each relation in the diagram should contain the relation's field names and associated data types.
 - If a field is a primary key, it should be marked as such in the diagram.
 - If a field is a foreign key, it should be marked as such via a relationship with the primary key that it refers to.
 - If a table is wide and contains 30+ fields, only list the fields that represent keys on this table (i.e. PK, FK) as well as those which you consider informative and plan to query.
 - Draw all valid relationships between the staging tables (i.e. 1:1, 1:m, m:1, m:n), remembering that a table can also have a self-relationship.
 - Save the queries you used to discover the keys and relationships between tables in a Jupyter notebook named `<source>_erd_staging_queries.ipynb` where `<source>` is the source of your `dataset1`.
 - Name the ERD file `<source>_erd_staging.pdf` where `<source>` is the source of your `dataset1`.
2. Practice writing SQL queries with joins:
 - Create a new Jupyter notebook named `<source>_joins.ipynb` where `<source>` is the source of your `dataset1`.
 - Write **6** queries on `dataset1` with at least 1 join per query.
 - At least **2/6** queries must use an outer join.
 - All queries must use a `WHERE` clause or `ORDER BY` clause or both.
 - Show the output of each query in the notebook. If a query produces `> 12` rows, add a `LIMIT` clause to the query.
 - Add a short Markdown comment above each SQL statement to describe its function.

CS 327E Milestone 3 Rubric

Due Date: 02/16/19

<p>Create an ERD of dataset1 and save it as <code><source>_erd_staging.pdf</code>. Create a notebook with the queries you used to discover the primary keys and foreign keys saved as <code><source>_erd_staging_queries.ipynb</code>.</p> <p>The diagram should contain entities for all the tables, with each table named <code><dataset>.<table name></code>.</p> <p>-40 <code>./<source>_erd_staging.pdf</code> or <code>./<source>_erd_staging_queries.ipynb</code> not found in repository</p> <ul style="list-style-type: none"> -5 each incorrect relationship, up to -20 -10 each primary key/foreign key missing -10 missing data types or fields of tables -10 missing queries 	40
<p>Create a notebook with 6 queries involving joins. 2 queries should contain outer joins. Each SQL query should have a comment preceding it, describing its function.</p> <p>-60 <code>./<source>_joins.ipynb</code> not found in repository</p> <ul style="list-style-type: none"> -10 each missing queries with joins, up to -60 -5 each missing or incorrect comment, up to -30 -10 each missing outer join, up to -20 -10 each missing <code>where</code> clause or <code>order by</code> clause, up to -30 	60
<p><code>submission.json</code> submitted into Canvas. Your project will not be graded without this submission. The file should have the following schema:</p> <pre>{ "commit-id": "your most recent commit ID from Github", "project-id": "your project ID from GCP" }</pre> <p>Example:</p> <pre>{ "commit-id": "dab96492ac7d906368ac9c7a17cb0dbd670923d9", "project-id": "some-project-id" }</pre>	Required
<p>Total Credit:</p>	100