Course 732 (100 points)

Project Instructions.

Due Date – Week 12.

Set-up environment: Using the marketplace locate one or more **free** data sets that interest you. (5 pts)

- Then Install (acquire) the data. This will usually require email verification. Data is then shared with your account. Deltas flow in automatically.
- After you've obtained a copy of the data set(s) you are to treat this data as a source system(s). **Assume the data is now in your first tier "raw" or "Ingested".** Data cannot be altered in this layer.
- Email the instructor your data set choice asap so that we can avoid overlaps. Everyone in the class needs a different data set. Pick one quickly.

Worksheet #1. Build a curation layer. (20 pts)

- Create a specific schema for this layer. Using SQL within a worksheet: Create tables and views using a naming convention that makes it clear. (SLV_ or CUR_)
- Tag your tables using the Semantic policy option. Give the tag = project for all new tables and views.
- Within your SQL: You are to enhance the data with additional fields based on logic you create (e.g. case statements or IFF). Add flags or Yes/No indicators. You can remove data if it's missing values or update null values based on logic you decide. Be creative. **At least 5 steps** are needed within the procedure to "curate" the data. By "steps" we mean changes to the original data.
- Each step within your curation process needs to be reproducible when the script is run by the instructor. This means all objects you create must have the "Create or Replace" syntax. This means that within your script for curation, you are creating tables and/or views to hold the results.

Worksheet #2. Create stored procedure (10)

- Create a stored procedure that performs at least 2 more data transformations. These must be different from the changes you're making in worksheet #1.
  Be sure to use "create or replace" for this procedure.

Worksheet #3. Build an aggregation layer. Put these objects in a different schema using a naming convention that distinguishes it from curation. Such as (GLD_ or AGG_) (20 pts)

- Step 1. Create a second SQL worksheet: **At least 4 different tables or views** must be created, **and 4 different types of aggregation** should be used. (Avg, sum, count, min, max etc.)
- Step 2. Create a materialized view that queries some of the data in step 1.

  NOTE: Get creative with your data. Act as though you are trying to provide your colleagues with multiple ways to look at your data. How would others want to view that data and what questions would be interesting to them. Remember to use "create or replace" so that the instructor can recreate your objects.

Worksheet #4. Create a function (10 pts)

- Create a table function to generate results from one of your aggregated tables or views. Be sure to use the "Create or Replace" clause.
- The output of the function should be in the form of a table.

Worksheet #5. Data Sharing (5 pts)

- Create an internal marketplace listing for your materialized view in worksheet #3

Worksheet #6 Task (10 pts)

- Create a task that runs every Sunday at 4 am. (Test it but then be sure to suspend it)
- This task should execute the stored procedure you created in worksheet #2

Introduction to tasks | Snowflake Documentation

Worksheet #7. Create a summary with the following information: (10 pts)

A. Provide the name and description of the data set you chose.
B. Explain in 1-2 sentences what your naming convention is and what your schemas are so the instructor can locate them quickly.
C. Briefly explain the logic/formulas used for any custom fields you create in your curation layer (This is a mini data catalog)

Create a simple dashboard. (15 pts)

- 4 tiles minimum. Be sure to label each tile with a business-friendly name and each axis in the charts should be understandable as well. In short – act as though a business team member is going to be looking at it. Share the dashboard with the instructor – use your animal's name as the prefix. Example "Cheetah_Project_Dashboard"

General Note:

Using comments in your code will help you maximize your points. Being thoughtful and creative will also help you maximize your grade.

List of deliverables:

1. **Pick data set – email instructor**
2. **Worksheet #1 – Curation layer**
3. **Worksheet #2 – Stored procedure**
4. **Worksheet #3 -- Aggregation**
5. **Worksheet #4 --Function**
6. **Worksheet #5 -- Sharing/Listing**
7. **Worksheet #6 -- Tasks**
8. **Worksheet #7 -- Summary (worksheet 5)**
9. **Dashboard (dashboard with 3 tiles or more)**
10. **Put all your files into a folder. Label the folder: 732Project_<lastname>**