MSCI:3200
Homework 3
Due: Friday, 4/3 @ 11:59 PM
Total Points: 50

Specifications
- For this homework, you need to submit three files:
  - Cleaned and processed Excel (.xlsx) file
  - APEX-encoded SQL script (.sql) with your commands and queries to answer each question. (**Note:** use the "Export" option under SQL Scripts.) Use SQL comments (--) to mark question number in your script.
  - Word or PDF file with screenshots of the code and output for each question
- Your file names should contain your last name and the assignment name (e.g., "lastname_homework3.xlsx", "lastname_homework3.sql", "lastname_homework3.docx")
- All files should be attached in one submission:
  - If you are using a cloud service (e.g., Google Drive, OneDrive), to save your files, you may need to download the files and submit the local copies (stored on your computer) to ICON
  - If you need to re-submit your assignment, make sure to include all files

---

Download the "nobel.xlsx" file from ICON. This file contains 54 rows tracking information about female winners of the Nobel Prize (https://www.nobelprize.org/). This data is sourced from Wikipedia (https://en.wikipedia.org/wiki/List_of_female_Nobel_laureates). The data contains the following fields:

Field	Type	Description
Year	Numeric	Year of Nobel Prize award
LaureateID	Text	Unique identifier for each winner
Laureate	Text	Winner's name (Title, FirstName, MiddleName, LastName)
Birth	Date	Winner's birthdate
Death	Date	Date of winner's death
Country	Text	Winner's home country(ies)
Category	Text	Award category
Rationale	Text	Reason for award
Shared	Numeric	1 if award is shared, 0 if award is individual

1. (13 points) Clean and normalize the data. Submit your processed Excel file with a sheet for each relation/table. You do not need to submit screenshots of the Excel data.
   a. The raw data contains a multivalued field (Country) that must be resolved before loading the data into APEX
   b. The raw data contains a composite field (Laureate) that must be resolved before loading the data into APEX. **Hint:** Every laureate has a first name. Title, middle name, and last name are optional.
   c. Break up the raw data into 3 tables. LAUREATE records information about each woman who has won a Nobel Prize, including ID, name (title, first, middle, last), birthdate, and death date. COUNTRY tracks the ID and country for each laureate. AWARD contains information about the award itself (year, recipient ID,

category, reasoning, and whether the prize is shared). Assume that the same person cannot receive multiple prizes in the same year.

2. (23.5 points) Import the data into APEX.
   a. Write a CREATE TABLE SQL statement for each table. Select appropriate data types, sizes, and constraints for each field. Don't forget to add primary and foreign key constraints. Provide screenshots of your SQL statements in your Word/PDF file. **Note:** Do not use the Table Design wizard or Data Workshop to generate the CREATE TABLE code.
   b. Use the Data Workshop to move your data into APEX. You can use the copy and paste or file upload method. Provide screenshots of the first few rows of your populated tables from the Object Browser.

3. (13.5 points) Use the database you've created to answer the following questions with **simple queries** (each query uses data from only one table). Provide screenshots of your SQL code and results tables in your Word/PDF file.
   a. Which countries have had the most female winners? For the top six countries with the most female laureates, return the country name and total winners. Make sure to rename any new, calculated fields.
   b. Many Nobel Prize winners have advanced degrees and titles like Doctor or Professor, but which winners do not have any titles? Return their IDs, full names (formatted as one field), and birthdates. Make sure to rename any new, calculated fields.
   c. Some Nobel Prizes are awarded to individuals, and some may be shared among multiple people. For all instances where a woman won the prize individually in a category other than Literature or Peace, return the year, category, and reasoning for the award.