Data Analytics in Python
COMP 2800
Spring 2024

# Assignment 4

Complete each of the following exercises by writing the Python code required to get the given output. Put all of your code into a single Python file to submit it.

1. **For this exercise we will use a dataset on the course website:**

   http://faculty.otterbein.edu/dstucki/comp2800/billboard.csv

   **You can either download it and open it using a filename and path, or you can open it directly from the URL.**

   **This dataset presents information and rankings for 317 song tracks in 80 columns. The first several columns are** year**,** artist**,** track**,** time**,** genre**,** dateEntered**, and** datePeaked**. These first columns are intuitive descriptions of song tracks. The** dateEntered **column shows the date that the songs entered the hot 100 list. The** datePeaked **column shows the date that the songs reached their highest ranking on the hot 100 list. The rest of the 76 columns are song rankings at the end of each week from "w1" to "w76".**

2. **Write an instruction that reads this dataset into a pandas DataFrame.**

3. **Write a single line of code that computes how many null values are in each column. Then write code that drops all the columns that are completely null. Use that resulting DataFrame for the remainder of the exercises.**

4. **Write code to compute which song(s) was/were on the hot 100 list for the most number of weeks.**

5. **Run the following code and observe the result:**

   ```
   df.query('artist == "Houston, Whitney"')
   ```

   **Then write another method of getting the same result using Boolean masking.**

6. **Write code that shows just the artist, track, and datePeaked for any songs that contain the word love in their title (case-insensitive).**

7. **Write code that identifies the average time on the hot 100 list of the songs in the dataset.**

8. **Email your file as an attachment to prof. Stucki. Change the extension to .txt before attaching it.**

**Due: Monday, 2/21/2024 by 11:59pm**