**Data Description:** The Metropolitan Museum of Art presents over 5,000 years of art from around the world for everyone to experience and enjoy. The Museum lives in three iconic sites in New York City—The Met Fifth Avenue, The Met Breuer, and The Met Cloisters. Millions of people also take part in The Met experience online.

Since it was founded in 1870, The Met has always aspired to be more than a treasury of rare and beautiful objects. Every day, art comes alive in the Museum's galleries and through its exhibitions and events, revealing both new ideas and unexpected connections across time and across cultures.

The Metropolitan Museum of Art provides select datasets of information on more than 470,000 artworks in its Collection for unrestricted commercial and noncommercial use.

**Critical Details and Instructions:**

i. Included with these directions should be a .csv file (**MetObjects\_Subset.csv**) that consists of only a small subset of objects (~17.3k) in the museum. You should use this file as a basis for all instructions that follow.

ii. You must use either a .ipynb notebook with separate cells per problem or a .py file with separate functions per problem in your submission.

iii. For problems 1-5, you can manipulate the data-frames/dictionaries as you see fit and using whatever functions/libraries you want.  **However, it is critically important that your end results for each problem match the provided variable name (ex: the result of problem 1 is called df\_init) so that they are accessible for grading.**

iv. With the exception of problem 1 (which is trivial) you should include a few comments in your code that make it clear what your thought process and/or code does to address each problem.

1. Load the .csv file into a pandas data-frame (DF) called **df\_init** with appropriate rows and columns.

2. Many columns of this data are missing *entirely* (i.e. no entries are present) or have a majority of missing values for each object entry. Use Python to determine which columns are missing for **at least 50%** of the provided objects and create a modified version of the data-frame from problem 1 without these columns called **df\_prob2**.  
Hint 1: remember that ***drop*** is a simple way to remove a column or columns.  
Hint 2: You may want to look into the pandas member function **isna**.

3. Find the 10 most common values for “Object Name” for the data-frame from #3. Filter out any rows with objects **not** among these 10 most common ones and store the result (a data-frame with common objects only) into a data-frame named **df\_prob3**.

4. Most objects are associated with a country. Compute the percentage of objects from df\_prob3 whose “Country” column value matches “United States”, “Mexico”, “Canada”, and “Other” (anything else, including NaN values). Store these results in a dictionary named **dict\_prob4**, where each key is an entry (“United States”, “Mexico”, “Canada”, or “Other”) and the paired value is the corresponding percentage. Your percentages should add up to 100.

5. Most of the objects in the dataset include a completion date (“Object End Date”). For this problem, you should create a dictionary **dict\_prob5** with ten keys corresponding to each of the decades in the 20th century (1900s, 1910s,…,1990s), with value pairs corresponding to the count (*not percentage*) of objects from df\_prob3 completed in the corresponding decade. (For example, if 5 objects were completed in the 1990s, the dictionary would include {…‘1990s’: 5}).

6. Provide plots (histograms/bar plots or line plots) for each of the dictionaries in problems 4 and 5, where each key is given on the x-axis, and the percentage or count is depicted on the y-axis.

You should upload your exam via the File Response dialogue through the Blackboard exam – but if you *cannot* do so, email it to me ASAP. Note that if you are submitting a .py file you are highly encouraged to include a README to explain what should be run to produce the required structures for problems 1-5 and graphs for problem 6.