Midterm

PART A (30%)

Solve problems posted in CodeZinger.

PART B (70%)

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.

1. Create a table with all the features in the dataset with their expected data type.
2. Obtain the basic statistical analysis for all numerical features. (Pandas)
3. Obtain the basic statistical analysis for all categorical variables. (Pandas)
4. As you may notice the data contains Missing values, inconsistent information, missing documentation, possible duplication and mixed text and numeric data. Clean this dataset minimizing the number of deleted records (RegEx). How many records could you clean? Save the cleaned dataset in CSV format and read it in a new dataframe.
5. Obtain all histograms for the numerical and categorical variables.