# Design and develop a PySpark based system for basic data exploration.

Input - CSV (Strictly comma separated)
Output - Structured output (JSON) containing the exploration data.

1. Calculate missing value count for each column.

2. A column can be categorized into

    - **Categorical**
    - **Discrete**
    - **Continuous**
    - **Text**

3. Create rules based on which you can categorize the columns (after sampling the data stored in each column) in above mentioned classes.

    Based on classes, please compute:-

    For Categorical and Discrete
    - Extract distinct values & frequency

    For Continuous
    - Extract min and max value
    - Bins with frequency(Histogram)

    For text
    - Extract word count information

The Input CSV is just for reference. The spark code should be generic and work across a spectrum of CSV files.

For categorizing the variables, please come up with some heuristics and use that to divide the columns.

The assessment of this exercise will be determined by factors such as the modularity of the script, utilization of PySpark's built-in functions, error handling techniques, and other relevant aspects.