HW1 Pre LEC

1. To find if there is missing vlaues in your datasets, you can use the function df.isnull().sum(), this will show you which section has missing vlaues, for example, it will return this after running, name 2, age 3, income 4...

2. Observations: as my understanding, I think that every thing in you data set is a observation, like each row could be your observation, for example, if your data set is about student, then each row would represent a student's data

Variables: everything inside your dataset that you are measuring is variables, in the case of employee dataset, your variables could be name, age, income, work hours....

3. To summaries you data set, you can use the function df.describe(), This gives you summary statistics, for example like count, mean, min, max... for numeric columns in your dataset you can use something like age, and it could calculate the max, min... for you.

4.

df.shape: This function returns the total number of rows and columns in the dataset, regardless of the type of data in the columns (numeric or non-numeric).

For example, if your dataset has 5 rows and 3 columns, df.shape will return (5, 3).

df.describe(): This function only provides summary statistics for numeric columns by default (like Age in your dataset). It won't include non-numeric columns (like Name or City).

So if you have 3 columns, but only 1 of them is numeric, df.describe() will give you summaries for just that 1 numeric column, not all 3.

df.shape counts all rows, even if some are missing data.

df.describe()'s "count" column counts only the non-missing values.

5. Attribute (like df.shape): An attribute gives you information about the data, but you don't need to call it with parentheses. It's like a property of the dataset that's always available, like its size, column names, or data types.
Method (like df.describe()): A method does something to the data. It performs an action or calculation, and that's why you need to use parentheses. Methods allow you to get a summary, filter the data, or make changes to the dataset.

An attribute is a piece of information you can directly access about a dataset, while a method is a function that performs an action on the dataset, often needing parentheses to run it.