

Health Data Science: Chapter 8 Questions

Karis Van Noord

1. How do you import the pandas library in Python?
 - a. To import the pandas library into Python, you must use the code `import pandas as pd`. This allows you to reference the panda library when you import the code “pd”.
2. How do you read a CSV file named `data.csv` into a pandas DataFrame?
 - a. To read a CSV file called `data.csv` into a pandas dataframe, you would input the code: `df = pd.read_csv('data.csv')`.
3. How do you display the first 5 rows of a DataFrame named `df`?
 - a. To display the first five rows of a dataframe named `df`, you would input the code: `print(df.head())`.
4. How do you find the number of rows and columns in a DataFrame `df`?
 - a. To find the number of both rows and columns in a dataframe called `df`, you would input the following rows of code:
 - i. `Num_rows, num_columns = df.shape`
 - ii. `print("Number of rows:", num_rows)`
 - iii. `print("Number of columns:", num_columns)`
5. How do you print the names of all the columns in a DataFrame `df`?
 - a. To print the names of all the columns in a dataframe called `df` you would input the following code:
 - i. `print(df.columns)`.
6. How do you get the summary information of the DataFrame `df` including data types and non-null values?
 - a. To get all the summary information of the dataframe `df` including data types and non-null values you would input the following code:
 - i. `print(df.info())`.
7. How do you select only the rows in DataFrame `df` where the column `A` is greater than 5?
 - a. To select only the rows in the dataframe where the column `A` is larger than five you would input the following code:
 - i. `selected_rows = df[df['A'] > 5]`.
8. How do you select the columns `A` and `B` from the DataFrame `df`?
 - a. To select only the columns `A` and `B` from the dataframe you would input the following code:
 - i. `Selected_columns = df[['A', 'B']]`.
9. How do you rename the column `A` to `Alpha` in DataFrame `df`?
 - a. To rename the column `A` to `Alpha` in the dataframe, you would input the following code:
 - i. `Df_rename = df.rename(columns={'A': 'Alpha'})`.

10. How do you replace all occurrences of the value 999 in DataFrame **df** with **nan**?
 - a. To replace all occurrences of the value 999 in Dataframe df with nan input the following lines of code:
 - i. Import numpy as np
 - ii. `df.replace(999, np.nan, inplace=True)`
11. How do you drop all rows in DataFrame **df** that contain missing values?
 - a. To drop all rows in the dataframe df that contain missing values input the following code:
 - i. `Df_without_missing_values = df.dropna()`.
12. How do you fill all missing values in DataFrame **df** with 0?
 - a. To fill all missing values in the dataframe with 0, input the following code:
 - i. `df.fillna(0, inplace=True)`.
13. How do you calculate the mean of column **A** in DataFrame **df**?
 - a. To calculate the mean of column A in the dataframe, input the following lines of code:
 - i. `mean_A = df['A'].mean()`
 - ii. `print("Mean of column A:", mean_A)`.
14. How do you create a new column **C** in DataFrame **df** that is the sum of columns **A** and **B**?
 - a. To create a new column C in the data frame that is the sum of columns A and B, input the following code:
 - i. `df['C'] = df['A'] + df['B']`.
15. How do you save the DataFrame **df** to a CSV file named **output.csv**? (Find the solution by searching the pandas documentation.)
 - a. To save the dataframe to a CSV file named output.csv, input the following code:
 - i. `df.to_csv('output.csv', index=False)`.