Dojo - Week 2

Some Basics

Create pandas dataframes from the inputs you find here.

Example 1

- Display the first 6 rows
- Display the last 3 rows
- Display 4 random rows
- Rename the column “score” to “Total Score”
- Rename one element in the index
- Transform the “attempts” column to string
- Set the “name” column as the index
- Reset the index
- Add two more rows (with made-up information)
- Display only the rows where qualify = yes
- Display only the rows where the score is bigger than 11

- Count the yes and no in the qualify column
- Display the sum of all attempts
- Sort the dataframe by name
- Sort the dataframe by number of attempts
- Display the rows where the scores are missing

Example 2

- Display summary statistics; which of the calculated outputs does actually make sense for this dataset?
- Display the total number of missing values in the dataframe
- Display the number of missing values in each column
- Create new dataframes where you dealt with the missing values:
    1. Delete all rows with missing values
    2. Replaces them with Zeros
    3. Replace them with the Median of the column
    4. Replace them with the following value
       Check if this deals with all NaNs
    5. Replace them with the previous value
       Check if this deals with all NaNs
    6. Decide for each column individually which of the methods is suited best

Combining Dataframes

You can find the data here.

- Create two dataframes from student_data1 and student_data2
- Combine them by the columns (add the second as new columns to the first)
- Combine them by the rows (add the rows of the second to the first) and save it as student_data
- Add the new row "Susi Lustigs" to this student_data dataframe
- Join the exam_id to the student_data on the common column
  Which type of join would you use?

The Movies

Create a dataframe out of this dataset here.

- Display the size (number of columns and rows)
- Display the columns of the DataFrame
- Get the details of the third movie of the DataFrame
- Display every second movie
- How many unique genres are there?
- get the details of the movie with title 'Grumpier Old Men'.
- Create a smaller dataframe with only the first 4 columns and the 10th to 15th rows
- Create a smaller dataframe with only every second column and the last 20 rows
- Explain .iloc with an example and 1-2 sentences of explanation.
- Explain .loc with an example and 1-2 sentences of explanation.
- Sort the dataframe by release date
- Only display the movies released after 1995-01-01.
- Display the mean and median of the runtime of all movies
- get those movies whose revenue more than 2 million and spent less than 1 million.
- get the longest runtime and shortest runtime
- display the movies (title, number of votes) that received specified number of votes.

Alphabet Inc.

Here you can find stock price data of the google mother corporation Alphabet.

Description of the columns:

- Date: It gives the date of which stocks details are given.
- Open: It gives the opening price of stock on that date.
- High: It gives the highest price to which the stock ascened on that day.
- Low: It gives the highest price to which the stock plummeted on that day.

- Close: It gives the closing price of stock on that date.
- Volume: It gives the amount of stock traded on that date.
- Adjusted Close: An adjusted closing price is a stock's closing price on any given day of trading that has been amended to include any distributions and corporate actions that occurred at any time prior to the next day's open.

1. Write a Pandas program to create a simple bar plot of the trading volume of Alphabet Inc. stock between two specific dates.
2. Write a Pandas program to create a simple line plot of the opening, closing stock prices of Alphabet Inc. between two specific dates.
3. Do the same using a.) Matplotlib and b.) seaborn. Change three elements of the plots respectively (like titles, font sizes, colors….)
4. Write a Pandas program to create a plot of adjusted closing prices, thirty days and forty days simple moving average of Alphabet Inc. between two specific dates. (You can use the `.rolling(window=30).mean()` method to do that.)
5. Write a Pandas program to create a histogram to visualize daily return distribution of Alphabet Inc. stock price between two specific dates. (You can use the `df.pct_change(periods=1)` method to do that.