**Assignment 6 – Pandas**

### General Instructions:

The assignment can be performed in pairs or singles.

**Late submission** will lead to a reduction in your grade – 5 points per day.

Assignment will written in a **Jupyter Notebook** (.IPYNB file), which will be uploaded to **GitHub**.

In the course’s Moodle you should go to the assignment activity and upload a text file, named:

hw-<id>.txt, where <id> is your ID number. The file should contain the URL of your repository, for example: [https://github.com/israel\_israeli/*DS\_Intro\_HW\_*](https://github.com/israel_israeli/DS_Intro_HW_5)*6*

### Tasks

1. To the dataframe created in Assignment 5 (Scraping from LonleyPlanet.com), add a new column called “num\_of\_words” that counts the number of words in the description of the city.
2. Add a new column called “No\_punct\_description” that contains the text of the description column after removing all new line characters (\n) and all punctuation marks (e.g., “,” , “.”, “:”, “-“, “;”, … )
3. Add a new column called “has\_restaurants” that contains the number of times the word “restaurant” appears in the description.
4. Add a new column called “has\_museums” that contains the number of times the word “museum” appears in the description.
5. Add a new column called “has\_beaches” that contains the number of times the words “beach” or “ocean” or “sea” appears in the description.
6. In a new cell of the notebook: Calculate the average number of words in the description column
7. In a new cell of the notebook: find the city with the longest description column – that is, the highest number of words.
8. Draw a histogram of the column “num\_of\_words” (how many destinations have different levels of num\_of\_words.
9. Draw a scatter plot that shows the relationship between number of beaches and number of restaurants. The plot should include titles.
10. Draw a scatter plot that shows the relationship between number of museums and the number of beaches. The plot should include titles.