LESSON PLAN

LESSON #7

Aim: Grouping and rearranging data part 1

Objective: After a lesson on reading and selection of data, students will be able to use Python to perform the same procedures on two other datasets.

Do Now: Why is data science important? Statistics to make you think! [0 – 3.5 minutes]

https://www.youtube.com/watch?v=mGEm3oT32BA

Standards: 9-12. CT.2 Computational Thinking, Data Analysis, and Visualization

9-12. CT.3 Computational Thinking, Data Analysis, and Visualization

9-12. CT.10 Computational Thinking, Algorithms, and Programming

9-12. CT.7 Computational Thinking, Algorithms, and Programming

9-12. DL.5 Digital Literacy, Digital Use

Mini-Lesson:

1. Grouping data
   a. Grouping according to a criterion and aggregating
2. Rearranging data
   a. Transforming the arrangement of data, redistributing the indexes and columns for better manipulation of the data. Example: specifying which columns will be he indexes, the new values and the new columns
   b. Using the new index to specify specific rows
3. Conclusion and summary
   Discussion:
   - What are some of the problems or challenges you encountered?
   - How did you resolve them?
   - What did you learn from this lesson?
   - Do you have any lingering questions on today's lesson or data science in general?

CODE

Grouping according to a criterion and aggregating

```
group = edu[["GEO", "Value"]].groupby('GEO').mean()
group.head()
```

	Value
GEO	
Austria	5.618333
Belgium	6.189091
Bulgaria	4.093333
Cyprus	7.023333
Czech Republic	4.16833

Rearranging Data

Transforming the arrangement of data, redistributing the indexes and columns for better manipulation of the data. Example: specifying which columns will be he indexes, the new values and the new columns

```
filtered_data = edu[edu["TIME"] > 2005]
pivedu = pd.pivot_table(filtered_data, values = 'Value',
                        index = ['GEO'],
                        columns = ['TIME'])
pivedu.head()
```

TIME	2006	2007	2008	2009	2010	2011
GEO						
Austria	5.40	5.33	5.47	5.98	5.91	5.80
Belgium	5.98	6.00	6.43	6.57	6.58	6.55
Bulgaria	4.04	3.88	4.44	4.58	4.10	3.82
Cyprus	7.02	6.95	7.45	7.98	7.92	7.87
Czech Republic	4.42	4.05	3.92	4.36	4.25	4.51

Using the new index to specify specific rows

```
pivedu.ix[['Spain','Portugal'], [2006,2011]]
```

TIME	2006	2011
GEO		
Spain	4.26	4.82
Portugal	5.07	5.27