CA2: Static Data Visualization with Python

This assignment is worth 40% of the total marks that are available for this module.

This is an individual assignment. (estimated workload (outside class time): 10 hours.)

Requirements

You are required to use the dataset contained within the file "**us-names-by-decade.csv",**

which contains the following features:

- Gender - Gender of Individual – (M) Male, (F)Female
- Name – First Name of Individual
- Decade – 10 Year Period – 1990 = 1990-1999
- Count – Number of Individuals that were given the name detailed in the Name feature in the decade indicated in the Decade feature eg "F", "Olivia", "2010", "69799" = The number of Females named Olivia in the years 2010 -2019 was 69799

and then perform the following analysis:

1. You are then required to explain what you plan on doing with the data. E.g., Why did you choose the specific visualizations, etc **This must be detailed in the Mark-up of the Jupyter Notebook and include the rational for your choice**.

2. Generate a plot that details the top 5 number of Names for each of the Decades.

3. Plot a graph depicting the distribution of the Names that are Female in decade 1980.

4. Find out and visualize which decade had the MOST names.

5. Observe and visualize the average number of names per decade.

   No additional output will be graded.

   You must complete ALL data exploration PROGRAMMATICALLY and not using any other tool than python.

   NOTE This is an Individual Assessment and there are multiple ways to complete it

Marking Scheme (on a sliding scale)

Rationale	**Logical, Concise** Explanation of Your Plan for working with the data (Min 300 Words)	15
	Logical, Concise Explanation of Your Choice of each Visualization (Min 50 Words per visualization)	15
Python Solution in Jupyter Notebook	Data read and handling for csv	5
	Clear, working, code process	10
	Correct incorporation of Python Libraries (pandas, matplotlib, seacrest, etc)	5
	Solution to Requirement 2	10
	Solution to Requirement 3	10
	Solution to Requirement 4	10
	Solution to Requirement 5	10
	Inclusion of correct Title, axis title, annotation etc, suitable for the visualization chosen	10
	TOTAL MARKS:	100

Deadline

This assignment is due at 23:55 on **Wednesday 10th March 2020**.

Late submission will be graded as per CCT policy. Failure to upload your assignment will result in a 0% grade for this assessment.

Required Files

You are required to upload to Moodle:

1. **a Jupyter Notebook File file**, yourName_DV_CA2.ipynb.