**DATN-2376 6001 1 Analytical Tools and Met**

**The Pandas essentials for data analysis**

**Objectives**

**Applied**

1. Create a DataFrame by reading data from a file or by using the DataFrame constructor.
2. Save a DataFrame to disk as a pickle file, and restore the DataFrame by reading the pickle file.
3. Examine the data in a DataFrame by displaying the data and its attributes.
4. Examine the data in a DataFrame by using the info(), nunique(), and describe() methods.
5. Access columns, rows, or a subset of columns and rows by using some combination of dot notation, brackets, the query() method, and the loc[] or iloc[] accessor.
6. Use Pandas methods to get statistics for the columns of a DataFrame.
7. Use Python to do calculations on the data in the columns of a DataFrame.
8. Use the Pandas or Python replace() method to replace the data in a DataFrame or Series object.
9. Use Pandas methods to do the following DataFrame operations:  
   Sort the rows  
   Set and reset an index  
   Pivot the data  
   Melt the data  
   Group and aggregate the data  
   Plot the data based on the index that has been set

**Analyze the ramen data**

**The data file**

**ramen-ratings.csv**

**Tasks**

1. Read the data from the CSV file into a DataFrame.
2. Display the first five rows of data.
3. Display the last five rows of data.
4. Display statistical information for the numeric columns using the describe() method.
5. Display the number of unique values for each column.
6. Display only rows where the country is Vietnam.
7. Display only the Brand and Style columns.
8. Display only the Country column.
9. Display the data after it has been sorted by the Stars column from high values to low values.
10. In the Country column, replace “USA” with “United States”. Make sure this change is saved in the DataFrame, and then display the first five rows to be sure the change was made correctly.

**Questions**

1. How many countries are represented in the data?
2. Which three countries have the highest average rating?
3. Which three countries have the lowest average rating?
4. Which three countries have the most unique brands, and how many brands do each of these countries have?

**How to complete and submit**

* For each chapter project, create one Jupyter Notebook.
* When creating the filenames for your Notebooks, please use the convention specified by your instructor. Otherwise, name the Notebook *first\_last\_proj\_X-X* where *first\_last* specifies your first and last names and *X-X* specifies the project number. (example: if your name is Charlie Brown, then the notebook name for this assignment will be: *charlie\_brown\_proj\_2-1*
* Start each Notebook by importing any modules, classes, or functions required by the project.
* Use one Code cell to store the code that accomplishes each task and question.
* Use comments to describe your thought process and to document any code that’s hard to understand.
* Use Markdown cells to create headings that organize your Notebooks.
* Add a section titled 'Questions' at the end of your notebook with your answers to each question. Include the question text and your response. Your response should use complete sentences with correct grammar and spelling.
* Submit the notebook to this assignment's file dropbox.

**Knowledge**

1. Distinguish between a Series and a DataFrame.
2. Describe these attributes of a DataFrame: values, index, columns, shape

**Explain**

1. The first step to write the code

We have 2 phần mền là ANACONDA: [Anaconda.com/downloads](https://www.anaconda.com/download/)

và python

nhiệm vụ của hai phần mền này là để khi ta làm một bài tập về viết code hoặc phân tích chạy dữ liệu thì ta sẽ kéo tập tệp đó và hai phần mền này và chạy code những file bài tập này chỉ có thể mở trên những phần mền có thể đọc hiểu được nó như là ANACONDA và PYTHON .

hiện tại trong lớp DATN-2376 ta đang xài phần mền viết code là ANACONDA

2.syntax cú pháp : vd mình muốn gọi tên file để chạy ra thì sẽ dùng một kí tự bất kì như URL hay polls giống như thấy đang đặt trên bài mẫu để gọi tên tập file chạy ra

vd : đây là kí tự để gọi tên file

Graphical user interface, text, application

Description automatically generated

* url=pd.read\_csv : là mình đặt 1 kí từ bất kí để gọi tên tập file đó ra chạy chương trình
* mortality.csv" là tên file được nêu trên task

sau đó dựa theo video của thầy được lưu trong youtube coi và thực hành yêu cầu