LESSON 5 : PYTHON VISUALIZATION

**- Deadline : *Giảng viên / mentor set deadline và điền vào mục này trước khi gửi cho học viên***

**- Phương thức nộp bài : *Nêu rõ học viên sẽ gửi bài tập trực tiếp cho mentors hay gửi vào 1 folder đã được tạo sẵn. Nếu là folder tạo sẵn, vui lòng đính kèm link nộp bài vào đây !***

# Mục đích

*Bài tập này sẽ giúp các bạn củng cố kiến thức đã học, sau khi hoàn thiện các bài tập, các bạn có có thể đạt được :*

**Kỹ năng :**

* Áp dụng Visualize Data với Python. Đây là bước trực quan hóa dữ liệu sau khi EDA, Prepare Data.

**Kiến thức :**

Các định nghĩa, khái niệm sẽ được củng cố thông qua phần bài tập này, bao gồm :

* Thư viện Matplotlib
* Thư viện Seaborn

# Bài tập

# Bài tập 1:

Tips dataset là bản ghi về số tiền tip mà khách hàng đưa trong một nhà hàng trong khoảng hai tháng rưỡi vào đầu những năm 1990. Nó bao gồm 6 cột như là total\_bill, tip, sex, smoker, day, time, size.

Bạn có thể download Tips dataset ở đây. [**Link**](https://drive.google.com/file/d/1F1GK8H-YK_y-5n57YmfK2-p7aB6Ien9L/view?usp=drive_link)

| **Column name** | **Description** |
| --- | --- |
| total\_bill | This column records the total bill amount of the transaction, which includes the cost of the meal and any additional charges. |
| tip | The "tip" column contains the amount of money given by the customer as a tip for the service received. |
| sex | This column indicates the gender of the customer |
| smoker | The "smoker" column represents whether the customer is a smoker or a non-smoker |
| day | The "day" column denotes the day of the week when the transaction occurred. |
| time | This column specifies whether the transaction took place during lunchtime or dinnertime |
| size | The "size" column indicates the size of the dining party (number of people) |

**Trực quan hóa dữ liệu bằng Python trả lời câu hỏi?**

1. Vẽ một biểu đồ phân tán (scatter chart) bằng thư viện matplotlib trong Python để hiển thị mối quan hệ giữa các ngày và số tiền tip.
2. Vẽ một biểu đồ cột (bar chart) bằng thư viện matplotlib trong Python để hiển thị mối quan hệ giữa các ngày và số tiền tip.
3. Vẽ một biểu đồ phân phối tần suất (histogram) bằng thư viện matplotlib trong Python để hiển thị phân phối giá trị hóa đơn “total\_bill”.
4. Vẽ một biểu đồ phân tán (scatter chart) bằng thư viện seaborn trong Python để hiển thị mối quan hệ giữa các ngày và số tiền tip, và sử dụng màu sắc để phân biệt giới tính của khách hàng.
5. Vẽ một biểu đồ đường (line chart) bằng thư viện seaborn trong Python để hiển thị mối quan hệ giữa số tiền tip và size.

=> Chúng ta có thể kết luận được gì từ các biểu đồ?

# Bài tập 2:

Customer Satisfaction là bài toán ở hầu hết các công ty Product, trong ngành dịch vụ giáo dục, như các trường đại học và ôn thi, sự hài lòng của học viên là yếu tố then chốt quyết định sự phát triển và duy trì hoạt động của doanh nghiệp. Hiểu rõ mức độ hài lòng của học viên và phụ huynh giúp các trung tâm xác định được những khía cạnh cần cải thiện, từ đó nâng cao trải nghiệm học tập và đem lại kết quả tốt hơn.

Một trường đại học A ý thức được việc trên, nên trong quá trình học tập, trường luôn chủ động thu thập các ý kiến phản hồi của học viên.

Dữ liệu [link](https://docs.google.com/spreadsheets/d/1eZ5RKB65JplPA73qHvkcxiqIQISsr4lv/edit?usp=drive_link&ouid=106613426743944653993&rtpof=true&sd=true) là kết quả của các biểu mẫu dữ liệu ghi nhận ý kiến phản hồi của sinh viên. Bạn hãy sử dụng python để xử lý lại bộ dữ liệu trên và sau đó trực quan hóa chúng thành các biểu đồ phân tích về trải nghiệm học của học viên.

Gợi ý:

* Hãy kiểm tra dữ liệu và xử lý tại cột: Mã lớp, chuẩn hóa lại dữ liệu theo chuẩn 1NF tại các cột feedback chi tiết của các yếu tố đánh giá, lấy điểm tại các yếu tố đánh giá (điểm là ký tự số đầu tiên trong các feedback đó)