

Bài tập khoa học dữ liệu 1

1. Mở một google colab với tên theo cấu trúc: "problem1_ho_va_ten_lop". Và import các thư viện sau.

```
import numpy as np
from matplotlib import pyplot as plt
import pandas as pd
```

2. Upload file "Social_Network_Ads.csv" lên google colab.

3. Tạo DataFrame từ file trên

```
# Enter your code here
```

3. Tìm và in ra số khách hàng mua hàng (Purchased) và không mua

```
# Enter your code here
```

4. Vẽ pie plot biểu diễn tỉ lệ khách hàng mua hàng và không mua hàng

```
# Enter your code here
```

5. Thực hiện các tính toán các đặc trưng về tuổi và thu nhập của khách hàng để phân biệt khách hàng mua hàng và không mua?

```
# Enter your code here
```

6. Vẽ đồ thị biểu diễn trung vị (median) của thu nhập, giá trị trung bình của thu nhập, thu nhập nhỏ nhất, và giá trị lớn nhất theo tuổi của khách hàng

```
# Enter your code here
```

7. Vẽ histogram của thu nhập, thu nhập của khách hàng nữ, thu nhập của khách hàng nam. Cho nhận xét?

```
# Enter your code and your comments here
```

8. Tính hệ số tương quan của thu nhập và tuổi

```
# Enter your code here
```

9. Upload và tạo DataFrame có tên là df để lưu dữ liệu từ file "Salary_Data.csv"

```
# Enter your code here
```

10. Tính hệ số tương quan của các trường thông tin trong df. cho nhận xét

```
# Enter your code here
```

✓ 0s completed at 10:19 AM

