BÁO CÁO: Tìm hiểu SweetViz và AutoViz

# 1. Giới thiệu

Trong lĩnh vực phân tích dữ liệu, việc thực hiện EDA (Exploratory Data Analysis) là bước quan trọng để hiểu rõ tập dữ liệu trước khi xây dựng mô hình. Hai công cụ phổ biến hỗ trợ EDA tự động là SweetViz và AutoViz. Báo cáo này sẽ tìm hiểu, hướng dẫn sử dụng và so sánh hai sản phẩm trên.

# 2. SweetViz

## 2.1 Giới thiệu

SweetViz là một thư viện Python được thiết kế để tạo các báo cáo trực quan EDA một cách nhanh chóng. Nó xuất báo cáo dưới dạng file HTML với các biểu đồ, thống kê mô tả, và so sánh dữ liệu.

## 2.2 Cài đặt

Cài đặt bằng lệnh:

pip install sweetviz

## 2.3 Cách sử dụng

Ví dụ sử dụng SweetViz trên tập dữ liệu Marketing Campaign:  
```python  
import sweetviz as sv  
import pandas as pd  
  
df = pd.read\_csv('marketing\_campaign.csv')  
report = sv.analyze(df)  
report.show\_html('sweetviz\_report.html')  
```  
Sau khi chạy, báo cáo EDA trực quan được xuất ra file HTML và có thể mở bằng trình duyệt.

# 3. AutoViz

## 3.1 Giới thiệu

AutoViz là một thư viện Python giúp tự động hóa việc trực quan hóa dữ liệu. Nó tạo nhiều loại biểu đồ để phân tích mối quan hệ giữa các biến, phân phối dữ liệu, và các đặc tính quan trọng khác.

## 3.2 Cài đặt

Cài đặt bằng lệnh:

pip install autoviz

## 3.3 Cách sử dụng

Ví dụ sử dụng AutoViz:  
```python  
from autoviz.AutoViz\_Class import AutoViz\_Class  
import pandas as pd  
  
df = pd.read\_csv('marketing\_campaign.csv')  
AV = AutoViz\_Class()  
report = AV.AutoViz(  
 filename='',  
 dfte=df,  
 depVar='Purchase\_Amount'  
)  
```  
AutoViz sẽ trực tiếp hiển thị các biểu đồ trong notebook (Jupyter Notebook).

# 4. So sánh SweetViz và AutoViz

|  |  |  |
| --- | --- | --- |
| Tiêu chí | SweetViz | AutoViz |
| Mục tiêu | Tạo báo cáo EDA trực quan, tập trung vào thống kê mô tả và so sánh dữ liệu | Tự động trực quan hóa dữ liệu với nhiều loại biểu đồ |
| Đầu ra | File HTML (báo cáo đầy đủ) | Hiển thị trực tiếp trong notebook |
| Dễ sử dụng | Rất dễ, chỉ vài dòng code | Dễ sử dụng nhưng nhiều tham số tuỳ chỉnh |
| Tính năng nổi bật | So sánh tập dữ liệu, trực quan chi tiết phân phối | Nhiều loại biểu đồ đa dạng, trực quan hoá nhanh chóng |
| Ứng dụng | Phân tích sơ bộ và báo cáo cho người không chuyên | Khám phá dữ liệu chi tiết trong quá trình phân tích, nghiên cứu |

# 5. Kết luận

SweetViz và AutoViz đều là các công cụ EDA mạnh mẽ. SweetViz phù hợp để tạo báo cáo nhanh và chia sẻ với người khác, trong khi AutoViz phù hợp khi làm việc trong môi trường Notebook, cần nhiều biểu đồ đa dạng. Tuỳ theo mục tiêu phân tích, người dùng có thể chọn công cụ thích hợp.