**Case Study Data Visualization and Descriptive Statistics**

**Yêu cầu**:

- Sử dụng Python và trình bày với Markdown (tương tự bài tập Data Wrangling)

**Nội dung:**

Bạn là Data Analyst của công ty X. File dữ liệu **Home\_Values.xlsx** chứa các giá trị median của các biến sau: giá nhà (**Home Value**), thu nhập hộ gia đình (Household income - **HH Inc**), thu nhập trên đầu người(Income per capita - **Per Cap Inc**) và tỷ lệ phần trăm số nhà có chủ sở hữu (percent of homes that are owner occupied - **Pct Owner Occ** ) cho mỗi bang trong nước Mỹ (bao gồm cả District of Columbia). Trước khi đi vào những phần tích chuyên sâu hơn, công ty của bạn muốn hiểu rõ 4 biến số này (ví dụ: central tendency, variability, hình dạng của phần phối dữ liệu - shape of the distribution, các pattern về mối quan hệ giữa các biến số). Để giúp công ty hiểu rõ hơn về dữ liệu, bạn được yêu cầu thực hiện các bước phân tích sau:

1. Tính toán mean, median, range và standard deviation của mỗi biến. Tổng hợp và trình bày kết quả vào 1 bảng.
2. Tạo 1 frequency histogram cho mỗi biến để xác định hình dạng của phần phối dữ liệu. Lưu ý mỗi chart cần có tiêu đề và các trục của biểu đồ cần được ghi chú rõ ràng.
3. Tạo biểu đồ boxplots cho mỗi biến. Lưu ý mỗi chart cần có tiêu đề và các trục của biểu đồ cần được ghi chú rõ ràng.
4. Tạo biểu đồ điểm (scatterplots) cho mỗi biến trong tương quan với 1 biến khác (gợi ý: bạn sẽ cần phải vẽ 6 scatterplots). Lưu ý mỗi chart cần có tiêu đề và các trục của biểu đồ cần được ghi chú rõ ràng.
5. Tính toán correlation coefficient của mỗi biến với các biến còn lại. Trình bày thành 1 bảng.

Trình bày (dưới dạng markdown trong Python) các nhận định của bạn từ các bước phân tích 1-5 bên trên (đưa cả các bảng, chart đã chuẩn bị ở các bước trên vào để người đọc tiện theo dõi), các nhận định cần bao gồm:

1. Trình bày bảng dữ liệu có được ở bước 1 thể hiện gì về mức độ tập trung và sự phần phối dữ liệu của từng biến (the central tendency, variability and distribution of each variable).
2. Trình bày các frequency histograms ở bước 2 và các boxplots ở bước 3 hỗ trợ và giúp làm rõ các nhận được ở bước 1 như thế nào. Từ bước 1 và bước 2, hãy tìm ra các outliers đối với từng biến.
3. Trình bày các biểu đồ scatterplots ở bước 4 và correlation coefficients ở bước 5 thể hiện gì về mối quan hệ giữa các cặp biến (Ví dụ: các biến có quan hệ với nhau không?, các mối quan hệ đó có vẻ là tuyến tính (linear) hay phi tuyến tính (nonlinear)?, các mối quan hệ đó là tương quan đồng biến hay nghịch biến (positive or negative)?.
4. Trong các biến HH Inc , Per Cap Inc và Pct Owner Occ thì biến nào có mối quan hệ mạnh nhất với Home Value?