Câu 1: Vui lòng tham khảo file Jupyter NoteBook đính kèm

**Người thực hiện:** Đoàn Trọng Tân

**Email:** [tandoannrc@gmail.com](mailto:tandoannrc@gmail.com)

**Vị trí ứng tuyển:** Business Intelligence

|  |  |
| --- | --- |
| **Xử lý, làm sạch dữ liệu** | |
| Bước 1: Import thư viện cần thiết | #1. Import Necessary libraries  import pandas as pd  import numpy as np  import matplotlib.pyplot as plt |
| Bước 2: Tạo handle kết nối Jupyter Notebook vs ExelFile | #2. Create handle to connect Python with Excel  handle\_excel = pd.ExcelFile('TestPython.xlsx')  data = handle\_excel.parse('InputData')  **Output:** |
| Bước 3: Chỉnh lại dữ liệu bị lỗi ở cột YearMonth | #2.1 Fix Errors in data  data['Year\_month'] = pd.Series(str(month).replace("204304","202304").replace("203303", "202303") for month in data['Year\_month'])  **Output:** |
| **Bảng 1: Vẽ thống kê dạng bảng theo Manager, Director với Division là Corporate** | |
| Bước 1:  Lọc ra Division Corporate | #3. CORPORATE DIVISION  #3.1 Filter Corporate Division  corperate\_data = data[data['Division'] == 'Corporate']  Output: |
| Bước 2: Pivot theo Corporate Division | #3.2 Create Pivot Table  corporate\_table =pd.pivot\_table(corperate\_data, values =['Revenue', 'Asset', 'Level\_number', 'Staff\_quantity'], index =['Director', 'Manager'],  columns =['Y6M'], aggfunc = np.sum)  **Output:** |
| Bước 3: Tính Subtotal (Giám đốc) và Total của Corporate | #3.3 Create Subtotal for Director and Corporate  data\_1 = pd.concat([  y.append(y.sum().rename((x, ('Total of ' +x ))))  for x, y in corporate\_table.groupby(level=0)  ]).append(corporate\_table.sum().rename(('Corporate', 'Total')))  **Output:** |
| Bước 4: Lọc ra Private Division và Pivot kết quả | #4. PRIVATE DIVISION  private\_data = data[data['Division'] == 'Private']  table\_2 = pd.pivot\_table(private\_data, values =['Revenue', 'Asset', 'Level\_number', 'Staff\_quantity'], index =['Director', 'Manager'],  columns =['Y6M'], aggfunc = np.sum)  **Output:** |
| Bước 5: Tính Subtotal và lấy dòng cuối Private Total | #4.1 Create Subtotal for Director and Private, then take the final row as Private Total  data\_2 = pd.concat([  y.append(y.sum().rename((x, ('Total of ' +x ))))  for x, y in table\_2.groupby(level=0)  ]).append(table\_2.sum().rename(('Private', 'Total')))  data\_2.iloc[[-1]] **#Note**  **Output:** |
| Bước 6:  Append dòng cuối của Private vào Corporate để có bảng Pivot tổng | #4.2 Get the Final Pivot Result of CORPORATE  df\_corporate = data\_1.append(data\_2.iloc[[-1]])  **Output:** |
| Bước 7: Lấy list cột cần tính toán và so sánh | #4.3 Get the column names of the Final Pivot Table  columns = list(set([item[0] for item in df\_corporate.columns]))  **Output:** |
| Bước 8: Tinh sự tang giảm (dòng vòng lặp for) | #5 FINAL RESULT OF CORPORATE DIVISION  for column in columns:  df\_corporate[column + ' tăng/giảm (%)'] = round((df\_corporate[column][['T3&4']]['T3&4'] - df\_corporate[column][['T1&2']]['T1&2']) / df\_corporate[column][['T1&2']]['T1&2'], 2) \* 100  df\_corporate.reindex(sorted(df\_corporate.columns), axis=1) #reorde columns  **Công thức:** [(Value T3&4 – Value T1&2)/ Value T1&2] \*100  **Output:** |
| **Bảng 2: Vẽ thống kê dạng bảng theo Manager, Director với Division là Private (option)** | |
| Cách thức giống như cách xử lý và tính toán Corporate, | #6 FINAL RESULT OF PRIVATE DIVISION  for column in columns:  df\_private[column + ' tăng/giảm (%)'] = round((df\_private[column][['T3&4']]['T3&4'] - df\_private[column][['T1&2']]['T1&2']) / df\_private[column][['T1&2']]['T1&2'], 2) \* 100  df\_private.reindex(sorted(df\_corporate.columns), axis=1)  **Output:** |
| **Nhận xét về Revenue** | |
| * Doanh thu của nhìn chung giảm 18% (Corporate giảm 21% và Private giảm 14%) trong giai đoạn tháng 3-4/2023 so với tháng 1-2/2023. * Xét Corporate Division: * Trong đội ngũ của Giám Đốc A, xét thấy Trưởng phòng 2 giảm đều ở Staff\_quantity, Level\_number, Asset và Revenue. Vậy có thể đặt giả định là 1 trong 2 nhân viên vừa nghỉ trong T3&4 của Trưởng phòng 2 đóng góp khá nhiều vào Performance của đội ngũ. * So sánh làm tăng độ tin cậy của độ tin cậy trên như sau: * Xét đội ngũ Giám Đốc B, trưởng phòng 3, có 2 nhân viên nghỉ trong giai đoạn T3&4, nhưng các chỉ số còn lại không giảm, thậm chí tang đáng kể ở Asset và Level\_number, vậy thì 1 trong 2 nhân viên đó nghỉ việc mang lại dấu hiệu tốt để tái cấu trúc đội ngũ. * Vậy Trưởng phòng 2 cần xem lại cách quản lý, để tìm hiểu lý do nghỉ việc của 1 trong 2 nhân viên đóng góp hiệu quả cho đội ngũ, để đề xuất Ban giám đốc có những chế độ, phương pháp tốt hơn để giữ nhân viên tốt. * Xét Private Division: * Đội ngũ Giám Đốc B, dù có tăng nhẹ ở phần doanh thu ở Corporate Division, tuy nhiên là ở Private giảm đáng kể, đơn cử Trưởng phòng 4 dù số lượng nhân viên tăng lên 9 người (180%), nhưng doanh thu giảm tận 74%), vậy thì Giám Đốc B cần xem xét xem trưởng phòng 4 có phù hợp cho phân khúc Private nữa hay không để có định hướng phù hợp. | |
| **Vẽ line chart thể hiện Revenue Performance của Manager qua các tháng** | |
| Bước 1:  Lấy ra các cột cần thiết để Pivot thông tin | #7 Pivot Chart  df\_chart = data[['Year\_month', 'Manager', 'Revenue']]  df\_chart\_pivot = pd.pivot\_table(df\_chart, values =['Revenue'], index =['Year\_month'],  columns =['Manager'], aggfunc = np.sum)  **Output:** |
| Bước 2:  Vẽ chart | #8 Line chart visualization  plot\_df = df\_chart\_pivot['Revenue']  ax = plot\_df.plot(grid=True, figsize=(15, 10), marker='o', ylabel = "Revenue", title = "Sales Performance of Managers by Month")  #Show data value here  for col in plot\_df.columns:  for idx, val in enumerate(plot\_df[col]):  ax.text(idx, val, str(val))  **Output:** |