

## LAB3. THỰC HÀNH VỚI DATAFRAME NÂNG CAO

### 1. Load data từ PC của sinh viên.

- Cho sẵn 2 data. Sinh viên đọc 2 data này.

```
data1 = pd.read_excel('F:/Dạy Học/DH-Mở/HK1_NH2023_2024/Bài thực hành xác suất thống kê/Data/Điểm sinh viên/Điểm giữa kỳ.xlsx')
print(data1.head())
```

- Đọc 5 dòng đầu tiên của mỗi Data.

### 2. Gọi data2 là dữ liệu chứa điểm quá trình của sinh viên. Yêu cầu đổi tên các cột chữ ngày tháng năm thành tên: “Buổi 1”, “Buổi 2”, ...

```
tencot = ['Buổi '+str(i) for i in range(1,10)]
cot = list(data2.columns[0:4]) +tencot
data2.columns = cot
```

### 3. Ghép 2 Daframe này thành 1 dataframe theo Mã số sinh viên.

```
# Ghép DataFrame theo cột 'Student_ID'
merged_df = pd.merge(df1, df2, on='Student_ID', how='inner')
```

### 4. Xóa các cột thông tin bị trùng: cột họ và tên

- Dùng # Xóa cột 'B' và 'C' bằng phương thức drop

```
df = df.drop(['B', 'C'], axis=1)
```

- HOẶC dùng toán tử del dataframe['họ']

### 5. Fill NaN (not a number)

```
DataCuoi = ghep_df.fillna(0)
```

### 6. Đếm số buổi vắng tính điểm chuyên cần

```
#Đếm số buổi vắng tính điểm chuyên cần
# Tạo cột mới 'DiemChuyenCan' bằng cách đếm số chữ 'v' trong 9 cột cuối
DataCuoi['SoBuoiVang'] = DataCuoi.iloc[:, -9:].apply(lambda row:
    row.str.count('v')).sum(axis=1)
DataCuoi['DiemChuyenCan'] = 10-DataCuoi['SoBuoiVang']+2
for i in range(len(DataCuoi['DiemChuyenCan'])):
    if DataCuoi['DiemChuyenCan'][i]>=10:
        DataCuoi['DiemChuyenCan'][i] =10
```

### 7. Tính điểm bài tập bằng tổng tất cả các giá trị của các cột

```
def sum_integers(row):
    # Loại bỏ 'v' và chuyển đổi giá trị còn lại thành số nguyên
```

```

    numbers = [int(cell) for cell in row[4:] if isinstance(cell, (int)) and cell !=
'v']
    #Lọc ra các giá trị từ row[5:] sao cho giá trị đó là số nguyên và không phải là
'v'.
    return sum(numbers)

# Tạo cột mới 'DiemBaiTap' bằng cách áp dụng hàm cho từng hàng
DataCuoi['DiemBaiTap'] = DataCuoi.apply(sum_integers, axis=1)

```

8. tính điểm Quá trình bằng: Kiemtra\*30% + Điểm bài tập \*10% + Điểm danh\*10

9. Vẽ biểu đồ phân bố của điểm giữa kỳ

```

# Vẽ biểu đồ phân bố
plt.hist(DataCuoi['DiemGiuaKy'], bins=20, color='blue', edgecolor='black')

# Đặt tên cho biểu đồ và trục
plt.title('Phân Bố Điểm Giữa kỳ')
plt.xlabel('Điểm Điểm')
plt.ylabel('Số Lượng Sinh Viên')

# Hiển thị biểu đồ
plt.show()

```

10. Save dataframe thành file Excel

```

# Lưu DataFrame thành tệp Excel
DataCuoi.to_excel('Bảng điểm tổng hợp.xlsx', index=False)

```

File có tên là 'Bảng điểm tổng hợp.xlsx' sẽ được lưu ở thư mục làm việc hiện hành.