**สถานการณ์ (Scenario)**

ฝ่ายวิชาการต้องการวิเคราะห์ผลการเรียนของนักศึกษาในวิชาพื้นฐานวิศวกรรมศาสตร์ เพื่อหาแนวทางในการพัฒนาการเรียนการสอน คุณในฐานะวิศวกร ได้รับชุดข้อมูลผลการเรียนของนักศึกษาและต้องทำการจัดการข้อมูลให้พร้อมสำหรับการวิเคราะห์และสรุปผลเบื้องต้น

ชุดข้อมูลประกอบด้วยคอลัมน์ต่างๆ ดังนี้:

* StudentID: รหัสนักศึกษา
* FirstName: ชื่อจริง
* LastName: นามสกุล
* Major: สาขาวิชา
* MidtermScore: คะแนนสอบกลางภาค (เต็ม 100)
* FinalScore: คะแนนสอบปลายภาค (เต็ม 100)
* ProjectScore: คะแนนโครงงาน (เต็ม 50)
* Attendance: เปอร์เซ็นต์การเข้าเรียน

**ชุดข้อมูลสำหรับใบงาน (Dataset)**

ให้นักศึกษาสร้างตัวแปรสตริงใน Python เพื่อเก็บข้อมูล CSV ด้านล่างนี้ สำหรับใช้ในการโหลดเข้าสู่ DataFrame

student\_data = """StudentID,FirstName,LastName,Major,MidtermScore,FinalScore,ProjectScore,Attendance

6401001,Somchai,Jaidee,Computer Engineering,85,90,45,95

6401002,Somsri,Rakdee,Computer Engineering,76,,48,88

6401003,Mana,Petch,Civil Engineering,65,72,38,92

6401004,Chujai,Klaew,Electrical Engineering,92,95,50,98

6401005,Piti,Yindee,Computer Engineering,88,85,42,85

6401006,Malee,Boonma,Civil Engineering,,65,35,75

6401007,Weera,Chokdee,Electrical Engineering,78,82,40,90

6401008,Jinda,Suksai,Computer Engineering,95,98,49,99

6401009,Arun,Rungrueng,Industrial Engineering,55,60,30,68

6401010,Naree,Ngamdee,Civil Engineering,72,78,41,94

6401011,David,Smith,Computer Engineering,98,100,50,100

6401012,Emily,Jones,Electrical Engineering,68,75,39.0,80

6401013,Michael,Brown,Civil Engineering,81,80,44,91

6401014,Jessica,Davis,Computer Engineering,,92,47,93

6401015,Chris,Wilson,Industrial Engineering,62,70,33.0,65

"""

**ภาระงาน (10 Tasks)**

**1. การโหลดและตรวจสอบข้อมูลเบื้องต้น**

* โหลดข้อมูลจากตัวแปร student\_data เข้าสู่ DataFrame ชื่อ df
* แสดงข้อมูล 10 แถวแรกเพื่อทำความเข้าใจโครงสร้าง
* ใช้คำสั่ง .info() เพื่อตรวจสอบชนิดข้อมูลและหาค่าว่าง
* **คำสั่งที่ต้องใช้:** pd.read\_csv(), io.StringIO, df.head(), df.info()
* **คำถาม:** จากผลลัพธ์ของ .info() คอลัมน์ใดมีข้อมูลขาดหาย (Missing Values) และมีชนิดข้อมูล (Dtype) เป็นอะไร?

**2. การจัดการข้อมูลที่ขาดหาย (Missing Data)**

* เติมค่าที่ขาดหายในคอลัมน์ MidtermScore และ FinalScore ด้วย **ค่ามัธยฐาน (median)** ของแต่ละคอลัมน์
* ตรวจสอบอีกครั้งเพื่อให้แน่ใจว่าไม่มีข้อมูลขาดหายเหลืออยู่
* **คำสั่งที่ต้องใช้:** df['Column'].fillna(), df['Column'].median(), df.isnull().sum()

**3. การแปลงชนิดข้อมูล (Data Type Conversion)**

* ตรวจสอบชนิดข้อมูลของคอลัมน์ ProjectScore และหากยังไม่ใช่จำนวนเต็ม (integer) ให้ทำการแปลงชนิดข้อมูลของคอลัมน์นี้เป็น int
* **คำสั่งที่ต้องใช้:** df.dtypes, df['Column'].astype()

**4. การสร้างคอลัมน์ใหม่จากข้อมูลเดิม**

* สร้างคอลัมน์ใหม่ชื่อ FullName โดยการนำข้อมูลจากคอลัมน์ FirstName และ LastName มาต่อกัน (คั่นด้วยเว้นวรรค)
* **คำสั่งที่ต้องใช้:** การดำเนินการกับ Series โดยตรง (เช่น df['FirstName'] + ' ' + df['LastName'])

**5. การคำนวณคะแนนรวม**

* สร้างคอลัมน์ใหม่ชื่อ TotalScore ซึ่งเป็นผลรวมของ MidtermScore, FinalScore, และ ProjectScore
* **คำสั่งที่ต้องใช้:** การดำเนินการทางคณิตศาสตร์กับคอลัมน์

**6. การคำนวณและตัดเกรด**

* เขียนฟังก์ชัน Python ชื่อ calculate\_grade เพื่อคำนวณเกรดจาก TotalScore
* สร้างคอลัมน์ใหม่ชื่อ FinalGrade โดยใช้ฟังก์ชัน calculate\_grade
* **คำสั่งที่ต้องใช้:** def, if/elif/else, df['Column'].apply()

**7. การคัดกรองข้อมูลตามเงื่อนไขที่ซับซ้อน**

* คัดกรองและแสดงรายชื่อนักศึกษา **สาขาวิศวกรรมคอมพิวเตอร์ (Computer Engineering)** ที่ได้เกรด 'A'
* คัดกรองและแสดงรายชื่อนักศึกษาทั้งหมดที่มีเปอร์เซ็นต์การเข้าเรียน (Attendance) **น้อยกว่า 70%**
* **คำสั่งที่ต้องใช้:** Boolean Indexing (เช่น df[(condition1) & (condition2)])

**8. การเรียงลำดับข้อมูล**

* เรียงลำดับข้อมูลนักศึกษาทั้ง DataFrame ตาม TotalScore จาก **มากไปน้อย** และแสดง 5 อันดับแรก
* **คำสั่งที่ต้องใช้:** df.sort\_values(), df.head()

**9. การจัดกลุ่มและสรุปผลข้อมูล (Grouping and Aggregation)**

* จัดกลุ่มข้อมูลตามสาขาวิชา (Major) และคำนวณหา **คะแนนรวมเฉลี่ย** และ **คะแนนสอบปลายภาคเฉลี่ย** ของแต่ละสาขา
* **คำสั่งที่ต้องใช้:** df.groupby(), .agg()
* **คำถาม:** จากผลการวิเคราะห์ สาขาวิชาใดมีคะแนนรวมเฉลี่ยสูงสุด?

**10. การเตรียมและการบันทึกรายงานสรุป**

* สร้าง DataFrame ใหม่ชื่อ report\_df ที่มีเฉพาะคอลัมน์ StudentID, FullName, Major, TotalScore, และ FinalGrade
* บันทึก report\_df ที่ได้ลงในไฟล์ CSV ชื่อ student\_grade\_report.csv โดยไม่บันทึก index
* **คำสั่งที่ต้องใช้:** การเลือกหลายคอลัมน์ (เช่น df[['Col1', ...]]), df.to\_csv()