**DataFrame ด้วย PySpark และ Pandas**

A screen shot of a computer

Description automatically generated

A black background with green text

Description automatically generated





--------------------------------------------------------------------------------------------------

A black screen with white text

Description automatically generated

A screenshot of a computer screen

Description automatically generated

------------------------------------------------------------------------------------------

A computer screen with text

Description automatically generated

A screen shot of a computer program

Description automatically generated

A screen shot of a computer code

Description automatically generated

--------------------------------------------------------------------------------------------------

A screen shot of a black background

Description automatically generated

A computer screen with text

Description automatically generated

**Graph Analytics PowerBI**

A computer screen shot of text

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Output CSV**

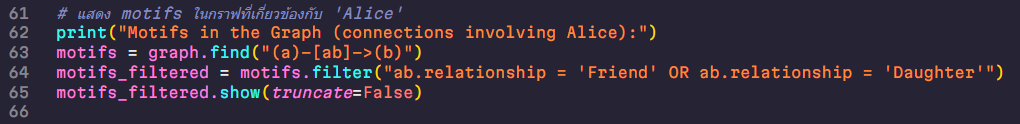
**Graph Analytics**

**A computer screen shot of a program code

Description automatically generated**

**A computer code on a dark background

Description automatically generated**

****

**A computer screen shot of text

Description automatically generated**

A screenshot of a computer

Description automatically generated**A screen shot of a computer

Description automatically generated**

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

คำสั่ง groupBy("src", "dst").count() ใช้ในการนับจำนวนการเชื่อมต่อระหว่างสนามบินต้นทาง

และปลายทาง

มีการใช้คำสั่ง .filter(col("count") > 5) เพื่อกรองเส้นทางที่มีการบินซ้ำมากกว่า 5 ครั้ง

คำสั่ง .orderBy(desc("count")) ทำการเรียงลำดับจากมากไปน้อย

ส่วนสี (source\_color, destination\_color) ถูกเพิ่มเข้าไปใน DataFrame เพื่อใช้ในงานแสดงผล เช่น การวิเคราะห์กราฟขั้นสูง หรือการ visualizing กราฟ.

ASC เรียงจากน้อยไปหามาก

DESC เรียงจากมากไปหาน้อย

.orderBY เรียง

.filter(col("distance") > 10)  # ตัวอย่างกรองระยะทางมากกว่า 10 กิโลเมตร

.filter(col("trip\_date").between('2023-01-01', '2023-12-31'))  # ตัวอย่างกรองการเดินทางในปี 2023

.filter(col("src") != col("dst"))  # ตัวอย่างกรองเส้นทางที่มีจุดเริ่มต้นและปลายทางไม่ซ้ำกัน

hub\_stations = graph.edges.groupBy("src").count().orderBy(desc("count"))

# ตัวอย่างหาสถานีที่มีการออกเดินทางมากที่สุด

.filter(col("trip\_type") == "leisure")  # ตัวอย่างกรองเฉพาะการเดินทางเพื่อพักผ่อน

grouped\_edges = graph.edges.groupBy("src", "dst").count()

.filter(col("count") > 5)

.filter(col("src") != col("dst"))  # กรองเฉพาะเส้นทางที่ไม่มีการย้อนกลับ

.filter(col("distance") > 10)  # กรองเฉพาะเส้นทางที่มีระยะทางมากกว่า 10 กิโลเมตร

.orderBy(desc("count"))

.withColumn("source\_color", lit("#3358FF"))

.withColumn("destination\_color", lit("#FF3F33"))

หาสถานีที่มีการเดินทางออกจากมากที่สุด

popular\_start\_stations = graph.edges.groupBy("src").count().orderBy(desc("count"))

หาสถานีที่มีการเดินทางเข้ามามากที่สุด

popular\_end\_stations = graph.edges.groupBy("dst").count().orderBy(desc("count"))

popular\_start\_stations.show()

popular\_end\_stations.show()

popular\_routes = graph.edges.groupBy("src","dst").count().orderBy(desc("count"))

หาสถานีที่มีเส้นทางเชื่อมโยงกับหลายสถานีที่สุด

station\_connectivity = graph.degrees.orderBy(desc("degree"))

--------------------------------------------------------------------------------------------------

**การ Join Table แบบ Pandas Output เอาเป็นไฟล์ csv**

A computer screen shot of text

Description automatically generated

**ผลลัพธ์ index=False**

A screenshot of a table

Description automatically generated

**ผลลัพธ์ index=True**

**A screenshot of a table

Description automatically generated**

**การ Join Table แบบ PySpark ไม่มี Output csv**

A screen shot of a computer program

Description automatically generated

**PySpark จัดการ Datframe โดยใช้ภาษา SQL**

A computer screen shot of a program code

Description automatically generated

**ตัวอย่างการใช้ต่าง ๆ**

A screen shot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Fillter PySpark**

A screen shot of a computer screen

Description automatically generated

A screen shot of a computer code

Description automatically generated

A computer screen with white text

Description automatically generated

A screen shot of a computer

Description automatically generated

A screen shot of a computer code

Description automatically generated

**จาก Code Text Analytics**

A screenshot of a computer program

Description automatically generated

A screen shot of a computer code

Description automatically generated

**ถ้าหากค่ามันว่างใน dataset**

A screenshot of a computer program

Description automatically generated

A screen shot of a computer program

Description automatically generated

A computer screen with text

Description automatically generated