

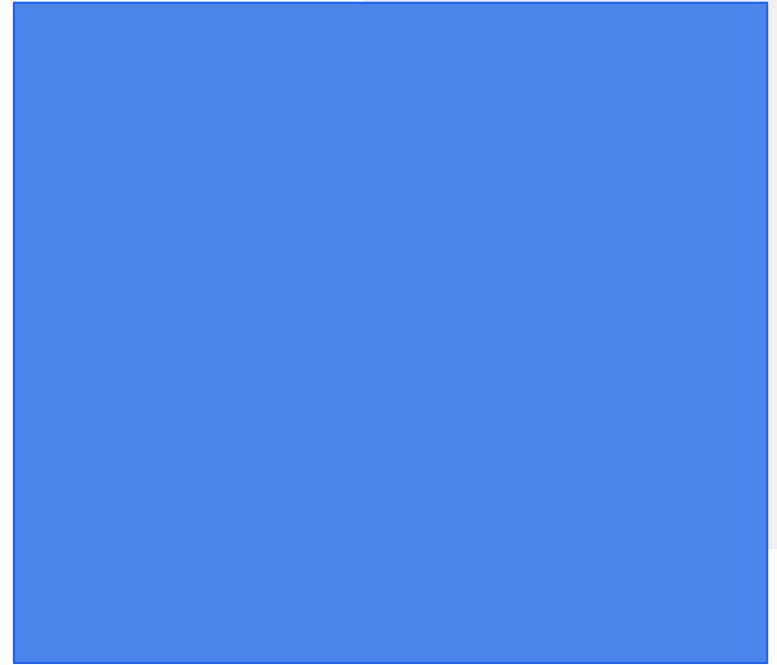


Big Data Analytics

Dr Sirintra Vaiwsri | Email: sirintra.v@itm.kmutnb.ac.th



Graph Analytics



Graph Analytics Implementation

- Import Libraries:
 - SparkSession
 - GraphFrame from graphframes
 - desc, col, and lit from pyspark.sql.functions

Graph Analytics Implementation

- Create SparkSession
- Read data from airline_routes.csv
- Show the Dataframe
- Create the vertice Dataframe using withColumnRenamed() and use the source_airport as the id column
- Create the edge Dataframe using withColumnRenamed() and use the source_airport as the src and destination_airport as the dst



Graph Analytics Implementation

- Show vertice Dataframe
- Show edge Dataframe
- Create graphframe using the created vertice and edge
- Show the number of vertices using `graph.vertices.count()`
- Show the number of edges using `graph.edges.count()`

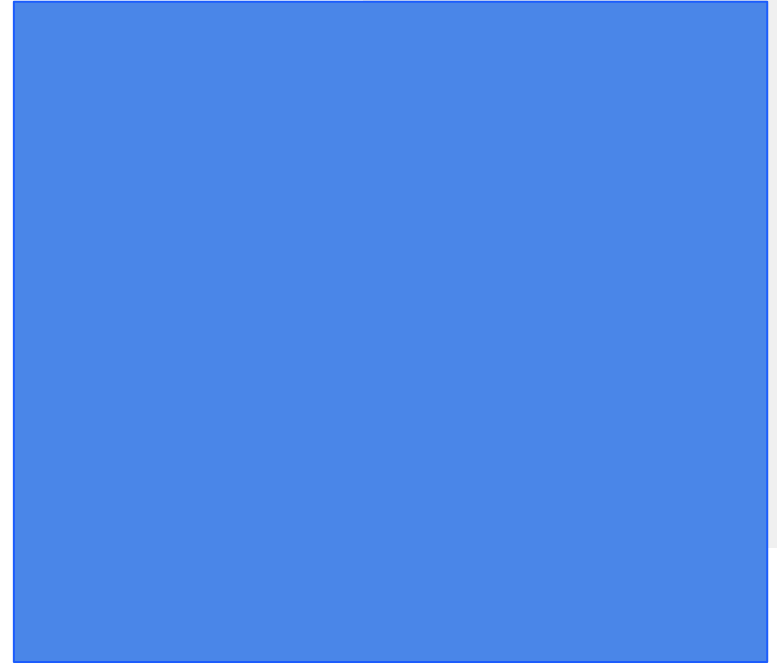


Graph Analytics Implementation



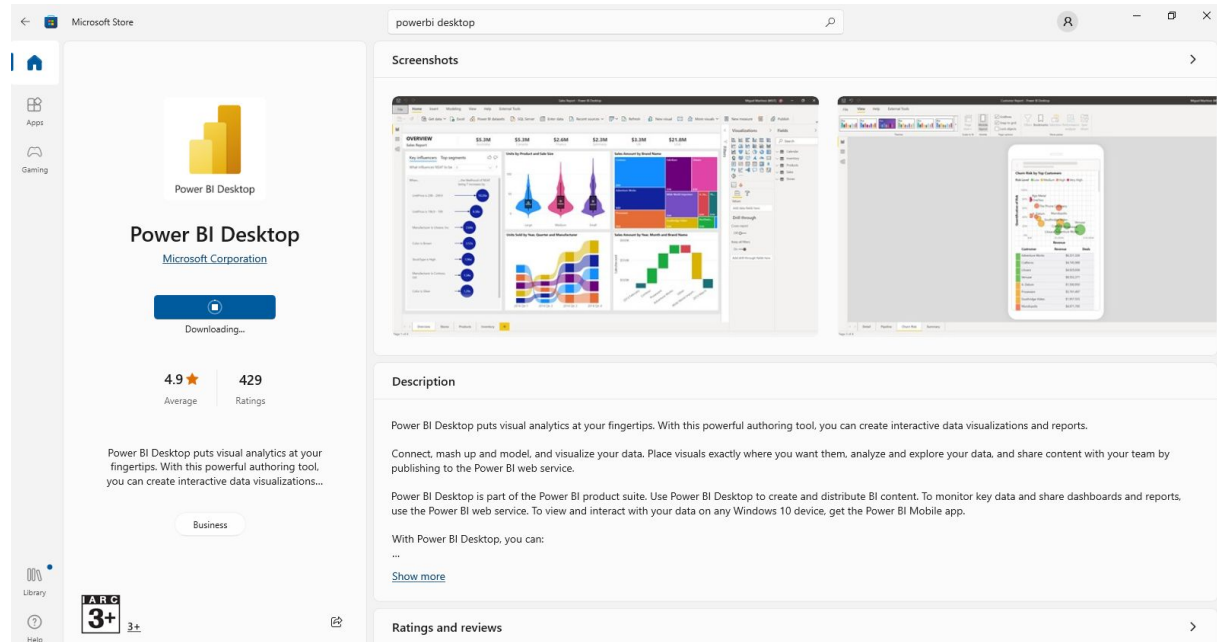
- Group the edges based on the src and dst where the count must be more than 5 and ordered in descending order, also add the column source_color with the value `lit('#3358FF')` and the column destination_color with the value `lit('#FF3F33')`
- Show data
- Write data into the CSV file using the overwrite mode and set the header to True

PowerBI Desktop



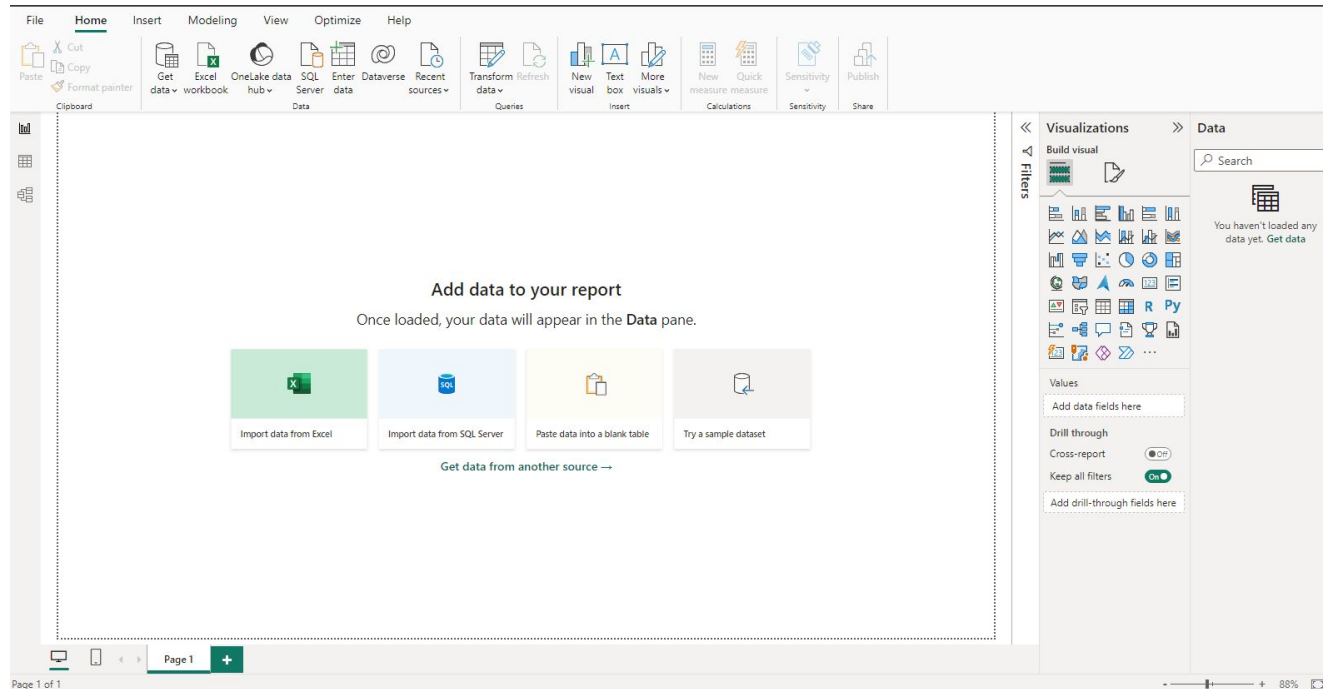
PowerBI Desktop

- PowerBI is one of the Microsoft platforms.
- It is used to visualise data.
- Download PowerBI Desktop from Microsoft Store



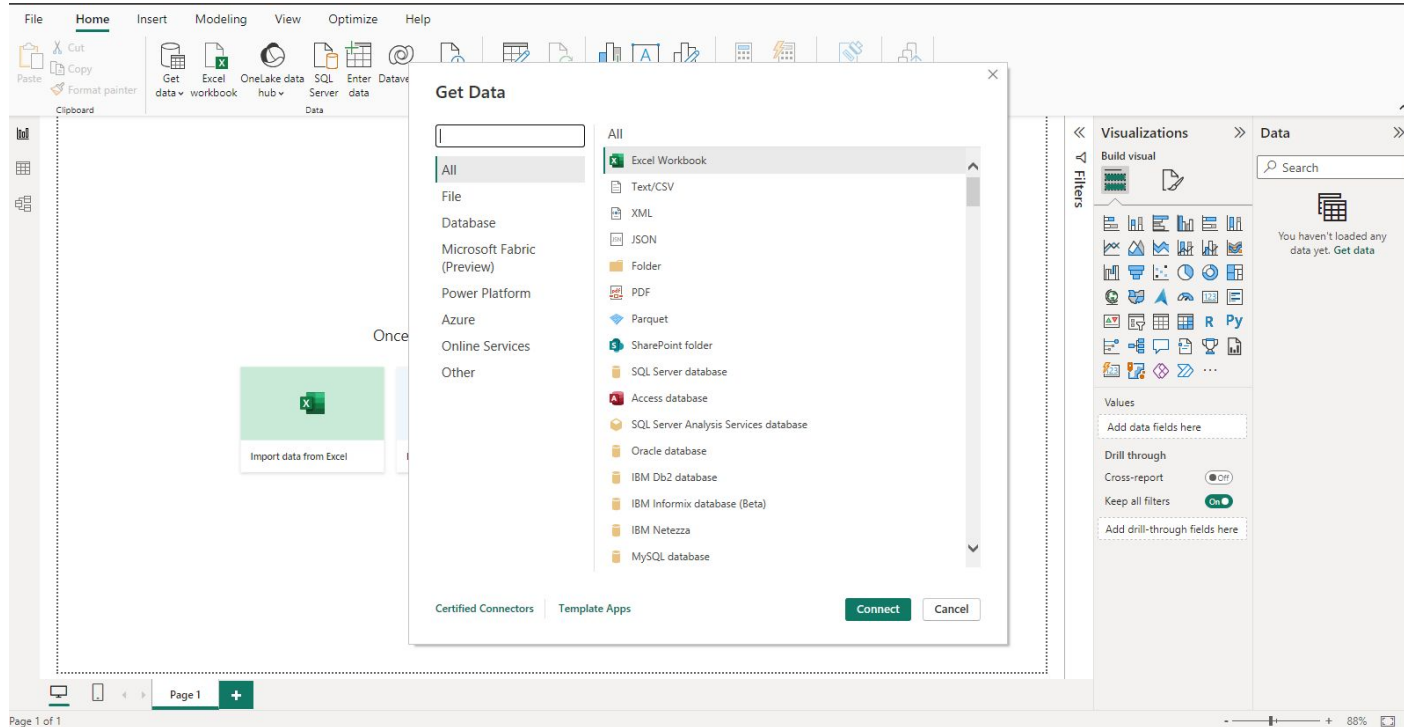
PowerBI Desktop

- Once the installation is finished, open PowerBI Desktop



PowerBI Desktop

- Get Data is used to import data into PowerBI



PowerBI Desktop

- Upload data to PowerBI, for example, airline_routes.csv file

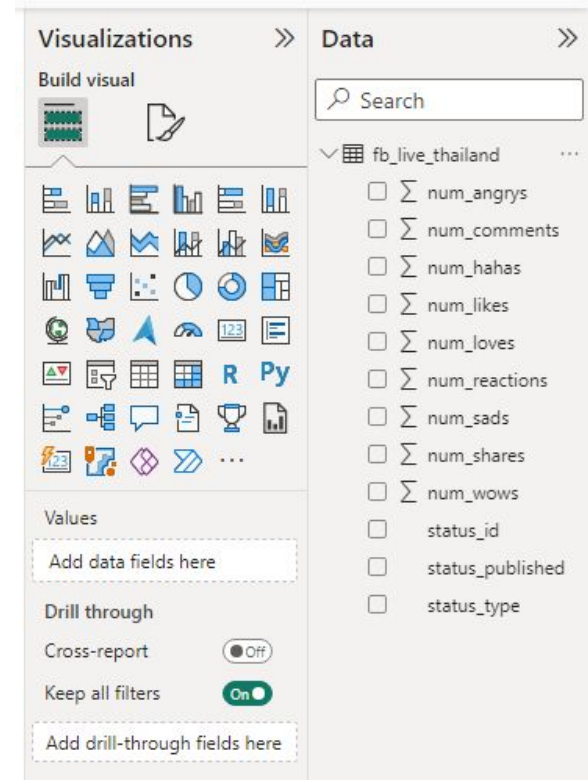
The screenshot shows the PowerBI Desktop interface. The 'Get Data' dialog box is open, displaying a list of files. The selected file is 'part-00000-b3489f9c-bd2a-4b90-ba89-e6fb884dde68-c000.csv'. The 'File Origin' is '874: Thai (Windows)', the 'Delimiter' is 'Comma', and 'Data Type Detection' is 'Based on first 200 rows'. A preview of the CSV data is shown below the file list.

| src | dst | count | source_color | destination_color |
|-----|-----|-------|--------------|-------------------|
| ORD | ATL | 20 | #3358FF | #FF3F33 |
| ATL | ORD | 19 | #3358FF | #FF3F33 |
| HKT | BKK | 13 | #3358FF | #FF3F33 |
| ORD | MSY | 13 | #3358FF | #FF3F33 |
| JFK | LHR | 12 | #3358FF | #FF3F33 |
| CAN | HGH | 12 | #3358FF | #FF3F33 |
| MIA | ATL | 12 | #3358FF | #FF3F33 |
| LHR | JFK | 12 | #3358FF | #FF3F33 |
| ATL | MIA | 12 | #3358FF | #FF3F33 |
| HKG | BKK | 12 | #3358FF | #FF3F33 |
| DOH | BAH | 12 | #3358FF | #FF3F33 |
| AUH | MCT | 12 | #3358FF | #FF3F33 |
| BKK | HKG | 12 | #3358FF | #FF3F33 |
| KGL | EBB | 11 | #3358FF | #FF3F33 |
| MSY | JFK | 11 | #3358FF | #FF3F33 |
| JFK | CDG | 11 | #3358FF | #FF3F33 |
| ATL | DEN | 11 | #3358FF | #FF3F33 |
| JFK | MSY | 11 | #3358FF | #FF3F33 |
| LAX | LHR | 11 | #3358FF | #FF3F33 |
| LHR | LAX | 11 | #3358FF | #FF3F33 |

The dialog box also includes a 'Load' button, a 'Transform Data' button, and a 'Cancel' button. A note at the bottom states: 'The data in the preview has been truncated due to size limits.'

PowerBI Desktop

- If the import succeeds, you will see your data on the right.
- By default, the column with numbers will be summarised.



PowerBI Desktop

- Get more visuals by clicking ...
- Sign in using your KMUTNB email

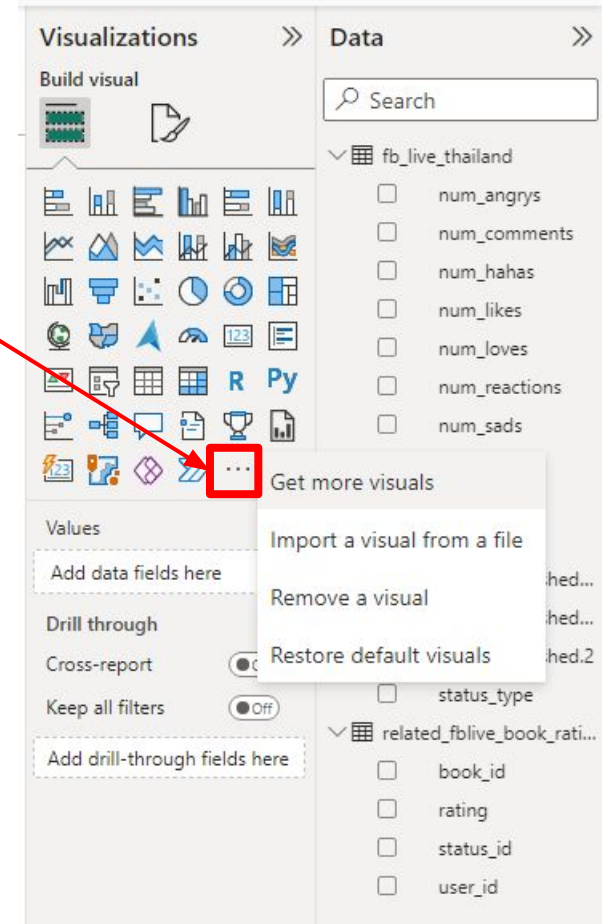
Enter your email address

Power BI Desktop and the Power BI service work better together. Sign in to enhance your collaboration and access organizational content.

Email:

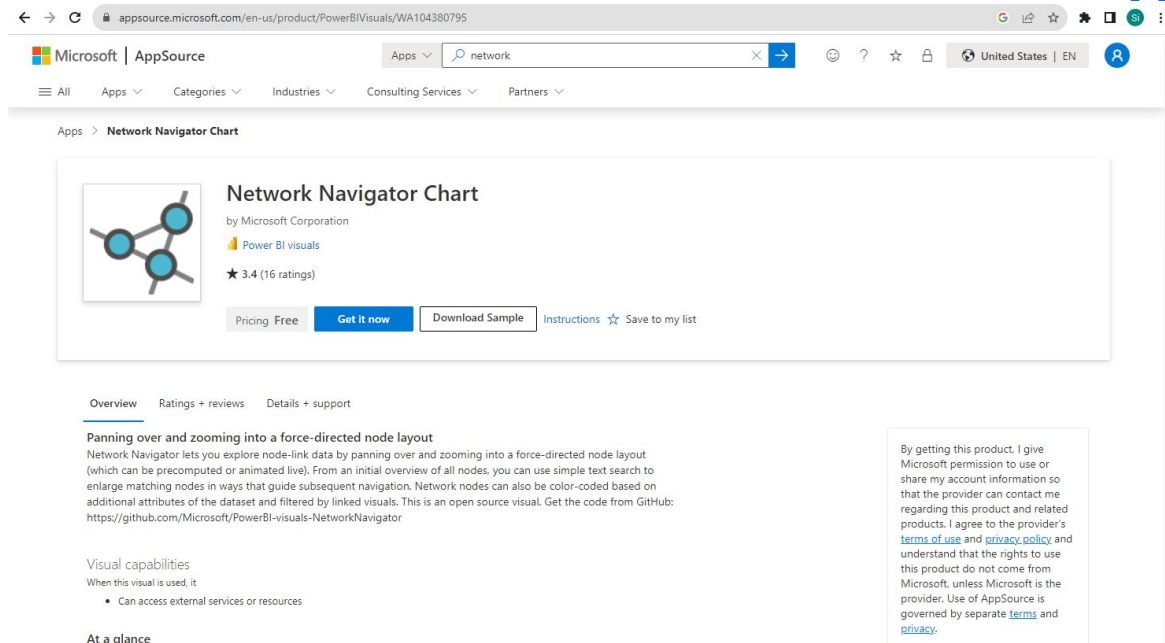
Continue

Cancel



PowerBI Desktop

- Download a new visual from Microsoft AppSource.
- For example, search for Network Navigator Chart



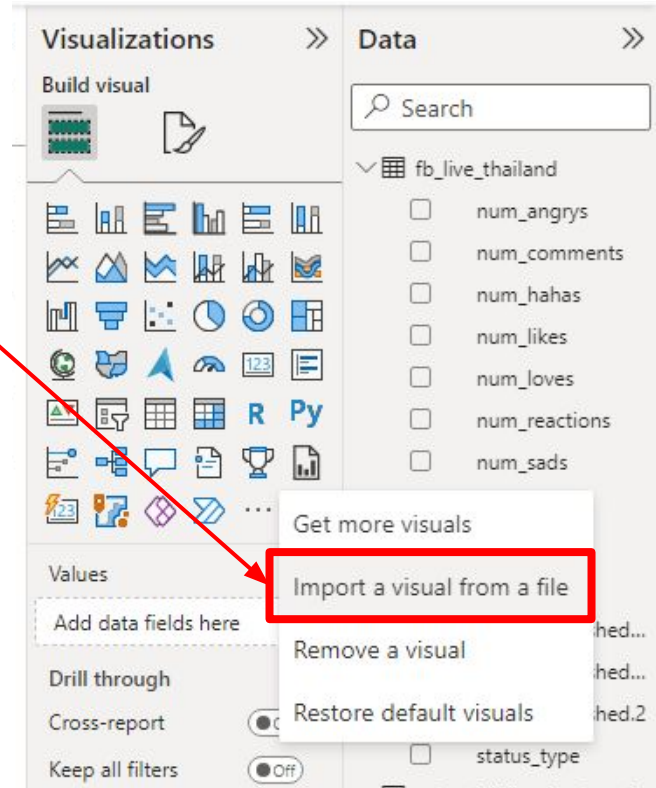
PowerBI Desktop

- Import the downloaded file to PowerBI

Import successful

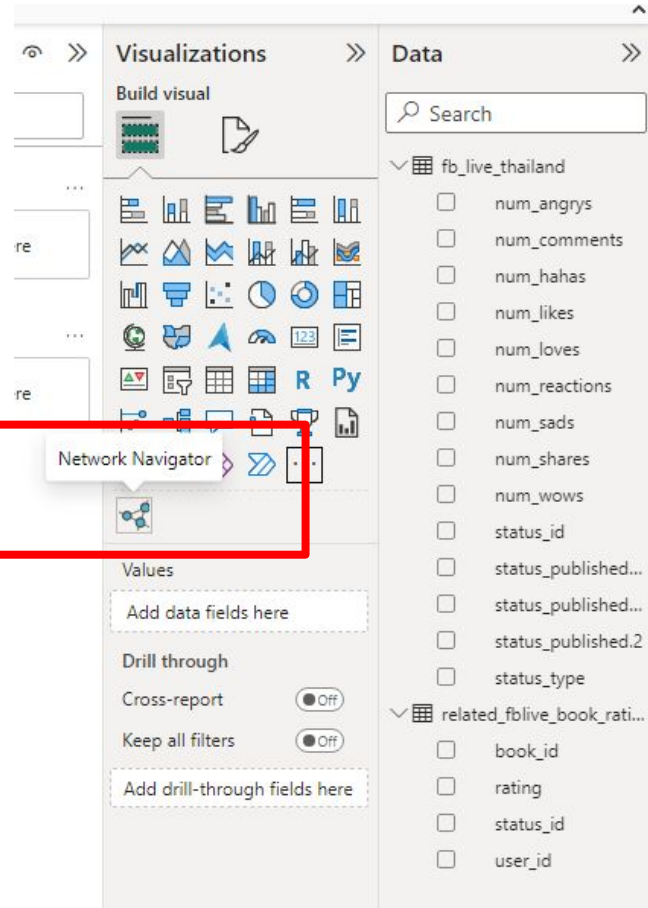
The visual was successfully imported into this report.

OK



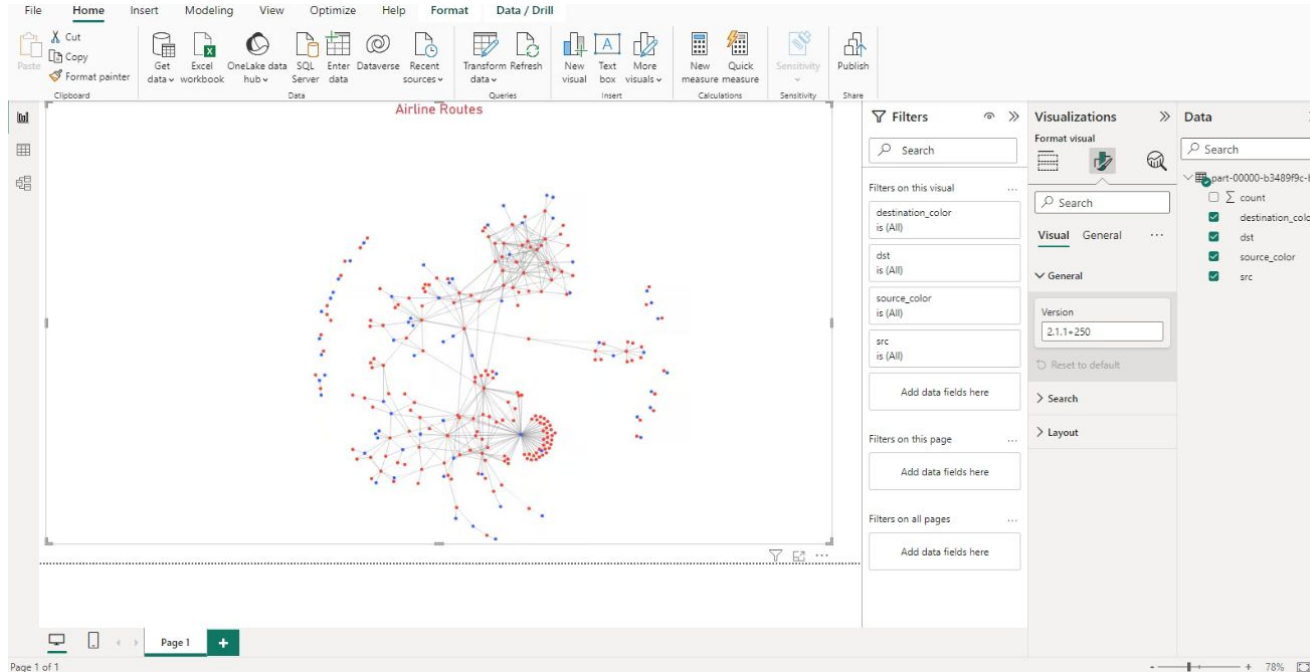
PowerBI Desktop

- New visual imported



PowerBI Desktop

- Generate graph



Assignment (2 points)

- Please implement the graph analytics following slides 3-6.
- Please use the file you created to visualise the graph in PowerBI.