Technical Report:

We downloaded files from Kaggle the patient file and patient\_route file. Next, we manually created database in PgAdmin Named: sk\_cov19\_db. We then created two tables manually in PgAdmin, the tables were named: patient and patient\_route we used the schema.sql to create the tables.

We ran the corona\_virus\_etl.ipynb on Jupyter notebook. The code does the following:

1. Read files into data frame (Extract)
2. Transformed data frame by removing columns (Transform)
3. Loaded data frame into corresponding tables in the database: sk\_cov19\_db (load)

Then, we performed the following analysis:

* Visual:

We created a bar graph that showed the number of patients organized and grouped by ages in the following buckets: 1-10, 10-25, 25-50, 50-100. (via Jupyter notebook)

* SQL Analysis:

We created a SQL file: sql\_analysis.sql that contains SQL statements to perform the following:

* + Total count of patients
  + Count of patients for each state (released, isolated, or deceased)
  + Number of contacts per patient sorted on descending order of number of contacts
  + Details of routes for some of the patients from previous query