The dataset I choose is the Flight Route Database from Kaggle:
https://www.kaggle.com/datasets/open-flights/flight-route-database

This data set details all routes counted in January 2012, include Airline name, airline id, Source airport 3-letter (IATA), Source airport ID , Destination airport 3-letter (IATA), Destination airport ID, Stops Number of stops on this flight, and Equipment 3-letter codes for plane type(s) generally used on this flight. For my analyze, I use the information of Source airport 3-letter (IATA), Source airport ID , Destination airport 3-letter (IATA), Destination airport ID.

I mainly use the ids of the airports as their names for analysis. For airports without ids, I transform the three letters of their names into 36-digit numbers to ensure that they are stored in the integral data type and are unique. After processing the data, I used a HashMap with each airport name as a key, and the airports they can reach as elements, forming a graph.

With this graph, I am try exploring how many connections are needed on average to reach another random airport from a random airport in today's world of flights, and what percentage of airports can be reached by taking off from a random airport and going through two transfer. I think through my studies, when people are flying on their own, if they find themselves with an above average number of connections between two not very niche airports, it's easy to see that there may be better options to get to their destination. At the same time, most people are reluctant to make too many connections, and two connections is the maximum number of connections that most people can tolerate. If an airline's flights can reach airports close to the total number of all flights within two connections, it means that the airline has a very wide coverage

To study these two problems, I am using the bfs algorithm, starting with a node to find all nodes that have an edge connection to him, and then continuing from those nodes. With this algorithm, it is possible to quickly find the number of places that can be reached within a certain distance, while finding the shortest path from one place to another.

According to my research, within two connections, people can reach an average of 200 different airports. Meanwhile, if a person wants to get from one identified airport to another, it takes an average of 4.18 connections to get there. And, there are about 2% of airports that may be too remote for them to reach from a plane. (4.18 transfers does not count this 2%).