Data science

test case

Included with the test case are the following three files

- **Reservations.csv**
- **Connections.csv**
- **Sizes.csv**

The following list explains the contents of each file

- **Reservations:** This dataset contains all the reservations made through the Mapiq app during the 90-day trial period (see part 1 for more info). A reservation is a booking made by an employee to indicate that they want to reserve a seat at the office on a particular date.
- **Connections:** This dataset contains information on all connections that have been made through Mapiq since the release of the product in 2021. Similar to connections on Linkedin, a connection in Mapiq is a link between two employees that allows them to follow each other's activity.
- **Sizes:** This dataset contains information on the total number of employees for each Mapiq customer.

Part 1

After Covid hit in 2019, many employees were required to work from home instead of going to the office. For many companies, this meant a lot of empty buildings and a lot of space management was required. To combat this problem, Mapiq developed a feature in their application that allows employees to reserve a place at the office. Employees could only come to the office if they had a reservation for that particular day.

This also meant that Mapiq customers could now limit the number of reservations being made on any given day and easily adjust the maximum capacity of the office to ensure that all social distancing rules were being followed and therefore ensuring their employee's safety.

After releasing the reservation feature, the data was collected for 90 days which was referred to as the trial period. After the trial period, the product team had some questions regarding the usage of the product. They sent the following list of questions to the data science team.

1. Which Mapiq customer has the most employees?
2. Before the reservation feature was released, how many connections had already been made for each customer individually? (One connection refers to a pair of users being connected with each other)
3. How many new connections were made during the trial period per customer?

4. Which employee(s) made the most reservations during the trial period? How many reservations did they make? (show the results in a table)
5. How many active employees were working for each Mapiq customer during the trial period? (An active employee is an employee that made at least one reservation during the trial period) (show the results as a graph)
6. What is the percentage of active employees compared to the overall number of employees for each Mapiq customer?

Please hand in all the answers to the above questions in a written format and include any code snippets, graphs and tables used to support your answers.

Part 2

After the trial period ended, more covid restrictions had already been lifted, but now Mapiq customers were struggling with getting their employees back to the office.

When brainstorming about how Mapiq could help our customers get more of their employees back to the office (without forcing them to do so), one question arose. The question was

"Do the number of connections of an employee influence his decision on whether to go to the office or not?"

This question was brought to you as a part of the data science team, and you were asked to present your findings a few days later. Using the data provided, perform an analysis of your choice and present your findings in two or three slides, including any tables or graphs you used to support your argument.