Data Assignment: Thai Travel Thailand (TTT)

Thai Travel Thailand (TTT; fictional organization) has released an app to track the locations of its users (with permission) for the purpose of domestic travel promotion. As one of the coolest data consulting companies, we are hired to analyze the data. Given the input files, our first task is to transform individual user locations into trips defined as follows.

Input

Two flat files in CSV format as shown below. The data headers and formats should be self-explanatory. Dates are in `YYYY-MM-DD` form and hours are `hh:mm`.

user.csv

user_id	hometown
0001	Bangkok
0002	Bangkok
0003	Nonthaburi

transaction.csv

date	hour	user_id	province
2022-02-21	15:00	0001	Bangkok
2022-02-22	06:00	0001	Bangkok
2022-02-22	07:00	0001	Nakhon Pathom
2022-02-22	16:00	0001	Bangkok
2022-02-22	17:00	0001	Samut Prakan
2022-02-22	18:00	0001	Chiang Mai
2022-02-23	12:00	0001	Chiang Rai
2022-02-24	08:00	0001	Chiang Mai
2022-03-01	20:00	0001	Chiang Mai
2022-03-01	22:00	0001	Bangkok

Download full input files from the following Google Drive link: https://drive.google.com/drive/folders/1g1rEEdPfZhXvYT0KMSZSMCUSmbyF58yt?usp=sharing

Output

A flat file in CSV with each row as a **trip** as defined below. The following output has 3 trips derived from the input examples above.

trip_id	user_id	start_date	end_date	province_list
0001	0001	2022-02-22	2022-02-22	Nakhon Pathom
0002	0001	2022-02-22	2022-02-24	Samut Prakan, Chiang Mai, Chiang Rai
0003	0001	2022-03-01	2022-03-01	Chiang Mai

A trip is a series of provinces traveled to by one user. A trip can be on the same day or span many days. Their hometowns mark the start and the end of a trip. Each province appears at most once in output `province_list` and the list does not include its user's hometowns. `trip_id` is arbitrary.

After this task, we would like to answer the following questions:

1. What is the total number of trips?
2. How many provinces are there in the trip with the most number of provinces?
3. What are the most common province pairs that people travel to in the same trip?

4. How can we use the data to support our client’s business?

Deliverable

Complete all the steps and show your work. Submit the output file and your answers to the questions above along with your code.

- **Submission deadline:** details written in email
- **Submission email:** job+data@oho.chat
- **Tools**: SQL, Python, or other programming languages. Feel free to use any common libraries.
- **Submission attachments:**
  - For the questions, write your answers on any editors and export as **PDF.**
  - For the code, prepare **the codes or git repo** with instructions on how to run.