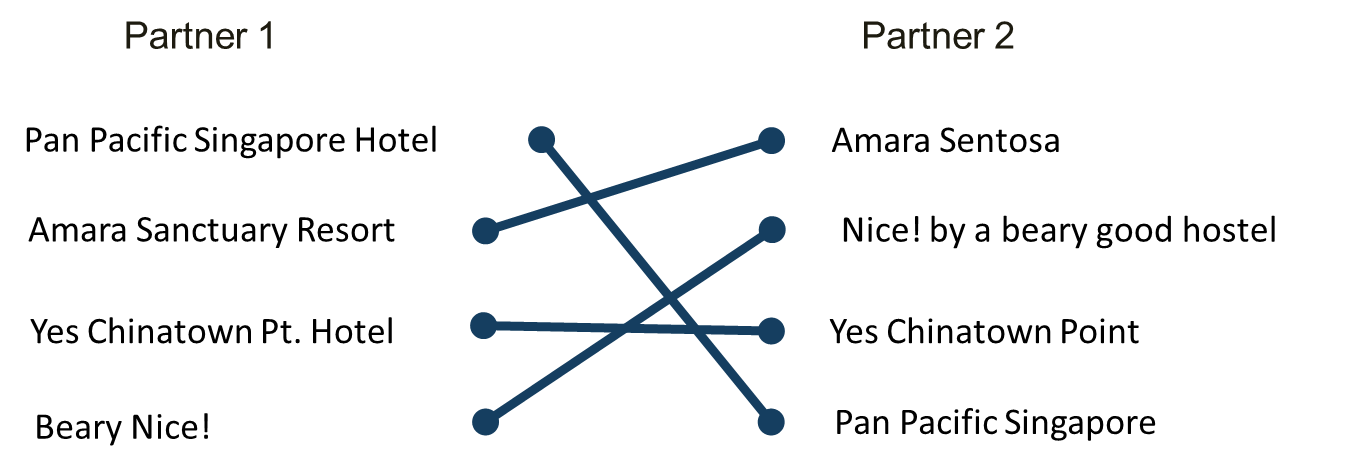
**Agoda mapping challenge**

**Objective:** As Agoda is one of the largest OTAs (Online travel agents) in the world, it gets many partners to sell their inventory through the Agoda platform. In order to make the experience smooth, there is a need to map between different partner inventories. Since this is a high scale problem, the aim is to design and implement an algorithm that automatically finds matches between two inventory lists of hotels. The objective is to create the maximum possible matches, given that the level of accuracy in the match is high enough.

Example:



**Input data sets:**

mappinghotelsdataset.xlsx contains three sheets:

* Partner1 hotels inventory list
* Partner2 hotels inventory list
* Examples of matches

**Notes:**

* Each partner contains metadata: key, hotel\_name, country\_code, city\_name, star\_rating, hotel\_address and postal code. There is no obligation to use them, but they might help you in the approach to the solution.
* For this challenge in particular you can assume that most of the hotels (>99%) in Partner1 list have a corresponding hotel in the Partner2 list.
* The accuracy levels you want to achieve should be very high (>98%), as we don’t want to send customers to the wrong hotels ☺. Is better not to map a hotel, than to map it incorrectly. Make sure the result you got make sense and that they comply the numbers above.
* Take into consideration that the data is noisy, meaning that even if two hotels have the same exact name, they can relate to different hotels in different cities.
* Matches can only be made within the same country. (hotels with different country codes won’t match)
* Use utf-8 for character encoding of special characters (there might be some names in other languages)

**Outputs:**

* **The outputs should be composed of 3 parts:**
  1. **Mappings csv file**
  2. **Code**
  3. **Slides summarizing approach/solution/results**

1. **Mappings CSV file**

* Provide the output in the following format:
  + Zip file containing a directory with your name and last name without space.
    - **[NAME]\_[LASTNAME]**
  + Include a file inside the directory named **mappings.csv** including the following
    - Csv file containing the keys of the respective matches suggested by the algorithm
    - **Filename**: mappings.csv

|  |  |
| --- | --- |
| P1.key | P2.key |
| 074BF1CC1F1C150E080EBB9855D23EAC | CBEF956F35D16548C939056575C7E0C7 |
| 103756D573E5A0C80ED374C9637DB142 | 603122E5379E7354B532B876BB149C8D |
| … |  |

1. **Code**

* Code of the algorithm
* [BONUS] executable for running on a new set. (the executable should only take as argument an xlsx file with the same format as mappinghotelsdataset\_v2.xlsx)

1. **Slides**

* Between 1-3 slides describing the approach to the solution. Basic parts of the solution and results.
* **Filename**: results.[ppt/pdf]