# Trivago 2019 Dataset

Problem-definition

The data provided for this task consists of a training and test set, and metadata for accommodations (items). The training set contains user actions up to a specified time (split date). It can be used to build models of user interactions and specifies the type of action that has been performed (filter usage, search refinements, item interactions, item searches, item click-outs) as well as information about impressed items and prices at the time of a click-out.

The recommendations should be provided for a test set that contains information about sessions after the split date but is missing the information about the accommodations that have been clicked in the last part of the sessions. The required output is a list of maximum 25 items for each click-out ordered by preferences for the specific user. The higher the actually clicked item appears on the list the higher the score.

The following schematic illustrates the problem setting and the separation of the data into training and test sets.
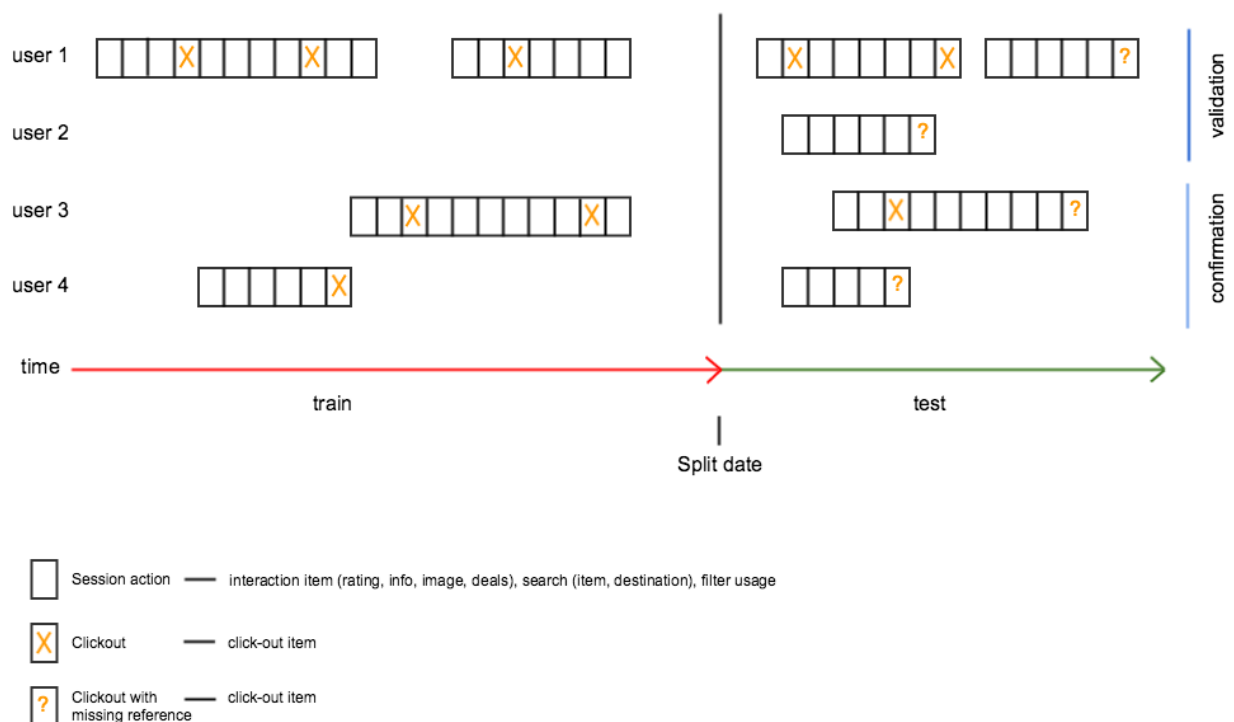
Evaluation

We use the [mean reciprocal rank](#) as metric.

```
Example:
     query 1:
```

- `impressions = [100, 101, 102, 103, 104, 105]`
- `clicked_item_id = 102`
- `results = [101, 103, 104, 102, 105, 100]`
- `reciprocal rank = 0.25`

```
     query 2:
```

- `impression = [101, 103, 104, 100, 105]`
- `clicked_item_id = 105`
- `results = [103, 105, 101, 100, 104]`
- `reciprocal rank = 0.5`
- `mrr = (0.25 + 0.5) / 2 = 0.375`

Data
Session actions (train.csv and test.csv)

- `user_id`: identifier of the user
- `session_id`: identifier of each session
- `timestamp`: UNIX timestamp for the time of the interaction
- `step`: step in the sequence of actions within the session
- `action_type`: identifier of the action that has been taken by the user.
    - **clickout item**: user makes a click-out on the item and gets forwarded to a partner website. The reference value for this action is the item_id. Other items that were displayed to the user and their associated prices are listed under the 'impressions' and 'prices' column for this action.
    - **interaction item rating**: user interacts with a rating or review of an item.
      The reference value for this action is the item id.
    - **interaction item info**: user interacts with item information.
      The reference value for this action is the item id.
    - **interaction item image**: user interacts with an image of an item.
      The reference value for this action is the item id.
    - **interaction item deals**: user clicks on the view more deals button.
      The reference value for this action is the item id.
    - **change of sort order**: user changes the sort order.
      The reference value for this action is the sort order description.
    - **filter selection**: user selects a filter.
      The reference value for this action is the filter description.

- o **search for item**: user searches for an accommodation.
    The reference value for this action is the item id.
  - o **search for destination**: user searches for a destination.
    The reference value for this action is the name of the destination.
  - o **search for poi**: user searches for a point of interest (POI).
    The reference value for this action is the name of the POI.
- `reference`: reference value of the action as described for the different action types
- `platform`: country platform that was used for the search, e.g. trivago.de (DE) or trivago.com (US)
- `city`: name of the current city of the search context
- `device`: device that was used for the search
- `current_filters`: list of pipe-separated filters that were active at the given timestamp
- `impressions`: list of pipe-separated items that were displayed to the user at the time of a click-out (see action_type = clickout_item)
- `prices`: list of pipe-separated prices of the items that were displayed to the user at the time of a click-out (see action_type = clickout_item)

   NOTE

- `1. The impressions and prices in the train.csv are sorted according to the way they were displayed.`
- `2. There is an additional action type "interaction item rating" in the train.csv and test.csv.`
- `3. The "current_filters" have more values for the "clickout item" action type in the train.csv and test.csv.`
-

Item metadata (item_metadata.csv)

- `item_id`: identifier of the accommodation as used in the reference values for item related action types, e.g. clickout_item and item interactions, and impression list
- `properties`: pipe-separated list of filters that are applicable for the given item

Results

The results should consist of a list of recommended hotels for each click-out that is missing in the test set.

Therefore the output file should be structured in the following way:

- `user_id`: identifier of the user that made the click-out
- `session_id`: identifier of the session in which the click-out was made
- `timestamp`: UNIX timestamp of the click-out
- `step`: step in the sequence of actions that marks the click-out
- `item_recommendations`: Space-separated list of items that the user will most likely have clicked

Explanation of user actions in a sample session

user_id	session_id	timestamp	step	action_type	reference	platform	city	device	current_filters	impressions	prices
93F7WGHBPO3A	569f5ea70df51	1541543231	1	search for destination	Barcelona, Spain	US	Barcelona, Spain	desktop			
93F7WGHBPO3A	569f5ea70df51	1541543269	2	filter selection	Focus on Distance	US	Barcelona, Spain	desktop	Focus on Distance		
93F7WGHBPO3A	569f5ea70df51	1541543269	3	search for poi	Port de Barcelona	US	Barcelona, Spain	desktop	Focus on Distance		
93F7WGHBPO3A	569f5ea70df51	1541543371	4	interaction item deals	40255	US	Barcelona, Spain	desktop			
93F7WGHBPO3A	569f5ea70df51	1541543425	5	clickout item	40255	US	Barcelona, Spain	desktop		6744|40181|40630|84610|2282416|1258693|974937|147509|128238|7998246|40255|3058538|1637385|40285|147502|921707|40849|6757|12770|893733|685091|147522|40708|860451|6819	162|91|218|190|176|365|272|159|139|240|136|5099|164|116|90|192|191|213|109|178|131|128|168|101|331
93F7WGHBPO3A	569f5ea70df51	1541543741	6	search for item	81770	US	Barcelona, Spain	desktop			
93F7WGHBPO3A	569f5ea70df51	1541543770	7	interaction item info	81770	US	Barcelona, Spain	desktop			
93F7WGHBPO3A	569f5ea70df51	1541543813	8	clickout item	81770	US	Barcelona, Spain	desktop		6832|40396|6621784|40197|6743|147488|40635|6177052|6742|1319782|40763|945255|83855|39937|1870125|1354432|6812|82400|40181|6834|81770|5056102|40797|923935|40284	347|245|199|65|359|233|227|270|294|625|208|174|121|217|226|616|293|166|91|198|274|272|123|130|131

In this session, a user from the US platform has used trivago on a desktop device. The actions in this session are the following:

1. (action type: search for destination, reference: Barcelona, Spain): User searches for Barcelona, Spain.
2. (action type: filter selection, reference: Focus on Distance): The 'focus on distance' filter is activated. At this point the current_filters column indicates that this is the only filter that is active.
3. (action type: search for poi, reference: Port de Barcelona): User searches for a point-of-interest (POI), the Port de Barcelona.
4. (action type: interaction item deals, reference: 40255): User viewed at the 'More Deals' button on item 40255. The 'focus on distance' filter is no longer activated.
5. (action type: clickout item, reference: 40225): The user clicks out on item 40225. The full list of displayed items and their associated prices can be seen in the 'impressions' and 'price' columns.
6. (action type: search for item, reference: 81770): User searches for item 81770.
7. (action type: interaction item info, reference: 81770): User interacts with the item information of item 81770.
8. (action type: clickout item, reference: 81770): User clicks out on item 81770. The full list of items and their associated prices can be seen in the 'impressions' and 'price' columns.