

Recsys 2019 Challenge

# **trivago Session-based Hotel Recommendation Data Description**

2020.04.23 13기 조 상연



ACM RecSys Challenge 2019

# 1. Data Overview

Train.csv

1. 15.93M / 730K users / 910K Sessions
2. 2018.11.01 ~ 2018.11.06

	count	unique	top	freq
user_id	15932992	730803	6JWWFFNUMY6Y	6230
session_id	15932992	910683	3167404ed3197	3522
action_type	15932992	10	interaction item image	11860750
reference	15932992	400277	interaction sort button	235027
platform	15932992	55	BR	2634304
city	15932992	34752	London, United Kingdom	326255
device	15932992	3	mobile	7643538
current_filters	1153112	61980	Sort by Price	159376
impressions	1586586	1059891	1668573	59
prices	1586586	1066775	26	72

Test.csv

1. 3.78M / 250K users
2. 2018.11.07 ~ 2018.11.08

	count	unique	top	freq
user_id	3782335	250852	66CA5CUMNYZQ	3254
session_id	3782335	291381	1cf1d5f47e2e4	2815
action_type	3782335	10	interaction item image	2578455
reference	3520604	184878	interaction sort button	69156
platform	3782335	55	BR	564722
city	3782335	22382	London, United Kingdom	75012
device	3782335	3	desktop	1831157
current_filters	342844	21955	Sort by Price	53436
impressions	528779	352788	103526 5479186 106174 103473 103580 8742148 22...	53
prices	528779	355101	208 123 107 218 205 143 177 132 182 158 131 17...	53

item\_metadata.csv

df_item	
properties	
item_id	
5101	Satellite TV Golf Course Airport Shuttle Cosme...
5416	Satellite TV Cosmetic Mirror Safe (Hotel) Tele...
5834	Satellite TV Cosmetic Mirror Safe (Hotel) Tele...
5910	Satellite TV Sailing Cosmetic Mirror Telephone...
6066	Satellite TV Sailing Diving Cosmetic Mirror Sa...
...	...
11199990	House / Apartment
11236100	House / Apartment
11260526	House / Apartment
11263712	House / Apartment
11275316	Hotel

927142 rows x 1 columns



# 1. Data Overview

step	action_type	reference	platform	city	device	current_filters	impressions	prices	dt
1	search for destination	Púchov, Slovakia	CL	Púchov, Slovakia	desktop	Nan	Nan	Nan	2018-11-01 21:21:36
2	search for destination	Púchov, Slovakia	CL	Púchov, Slovakia	desktop	Nan	Nan	Nan	2018-11-01 21:21:38
3	filter selection	5 Star	CL	Púchov, Slovakia	desktop	5 Star	Nan	Nan	2018-11-01 21:22:42
4	filter selection	4 Star	CL	Púchov, Slovakia	desktop	5 Star 4 Star	Nan	Nan	2018-11-01 21:22:49
5	search for destination	Pucón, Chile	CL	Pucón, Chile	desktop	5 Star 4 Star	Nan	Nan	2018-11-01 21:23:09
6	clickout item	327111	CL	Pucón, Chile	desktop	5 Star 4 Star	327111 2217612 492491 505751 5101488 2061220 3... 62 118 209 116 124 54 169 57 140 102 78 107 72...		2018-11-01 21:28:20
7	search for destination	Valdivia, Chile	CL	Valdivia, Chile	desktop	Nan	Nan	Nan	2018-11-01 22:05:50
8	filter selection	5 Star	CL	Valdivia, Chile	desktop	5 Star	Nan	Nan	2018-11-01 22:06:23
9	filter selection	4 Star	CL	Valdivia, Chile	desktop	5 Star 4 Star	Nan	Nan	2018-11-01 22:06:30
10	filter selection	5 Star	CL	Valdivia, Chile	desktop	4 Star	Nan	Nan	2018-11-01 22:07:33
11	filter selection	3 Star	CL	Valdivia, Chile	desktop	4 Star 3 Star	Nan	Nan	2018-11-01 22:07:37

## 2. Data Analysis: Action Type

• **search for item**: user searches for an accommodation.

The reference value for this action is the item id.

• **search for destination**: user searches for a destination.

The reference value for this action is the name of the destination.

• **search for poi**: user searches for a point of interest (POI).

The reference value for this action is the name of the POI.

• **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.

• **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.

• **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.

The screenshot shows the trivago search interface with the following details:

- Search Criteria:** 서울 (Seoul), 체크인 2020.05.07 (목), 체크아웃 2020.05.08 (금), 객실 더블룸.
- Sort Order:** 1박당 요금 (1 night price) - 670,000원+.
- Filters:** 숙박 시설 모든 유형, 고객 평점 모두, 호텔 위치 서울, 추가 필터 선택.
- Message:** 공지 사항: 코로나19(COVID-19)로 인해 여행 계획이 영향을 받을 수 있습니다. 자세한 정보는 [여기](#)에서 확인해 주세요. 건강과 안전에 유의하시기 바랍니다.
- Results:**
  - 신라스테이 역삼**: ★★★★☆ 호텔, 서울, 8.3 아주 좋음 (후기 772개).  
Image: A night photograph of a modern, multi-story hotel building with illuminated windows.  
Details: 추천 요금 무료 취소 · 호텔에서 결제 95,472 원 (선택).
  - 라마다 앙코르 서울 마곡**: 호텔, 서울, 8.1 아주 좋음 (후기 143개).  
Image: An interior view of a bright hotel room with a double bed and large windows.  
Details: 추천 요금 무료 취소 65,106 원 (선택).
  - Orakai Insadong Suites**: 주택/아파트, 서울, 8.8 최고 좋음 (후기 2405개).  
Image: An interior view of a spacious apartment room with a king-size bed and a balcony.  
Details: 추천 요금 무료 취소 · 무료 조식 116,185 원 (선택).

## 2. Data Analysis: Action Type

- **search for item:** user searches for an accommodation.

The reference value for this action is **the item id**.

- **search for destination:** user searches for a destination.

The reference value for this action is **the name of the destination**.

- **search for poi:** user searches for a point of interest (POI).

The reference value for this action is **the name of the POI**.

- **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.

- **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.

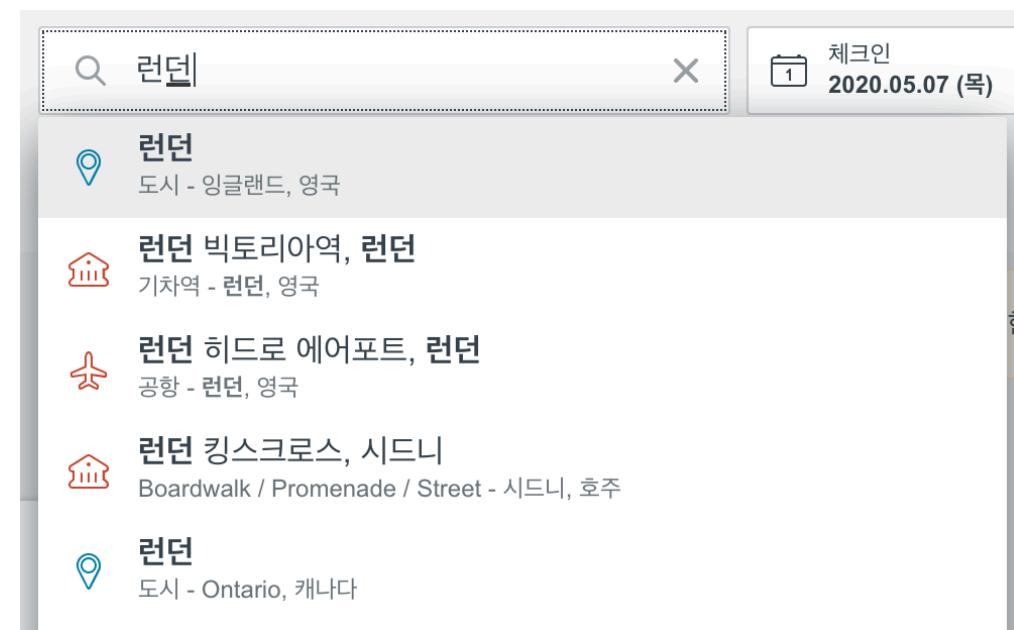
- **interaction item image:** user interacts with an image of an item.

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.

- **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.



## 2. Data Analysis: Action Type

- **search for item:** user searches for an accommodation.

The reference value for this action is the item id.

- **search for destination:** user searches for a destination.

The reference value for this action is the name of the destination.

- **search for poi:** user searches for a point of interest (POI).

The reference value for this action is the name of the POI.

- **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**.

- **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.

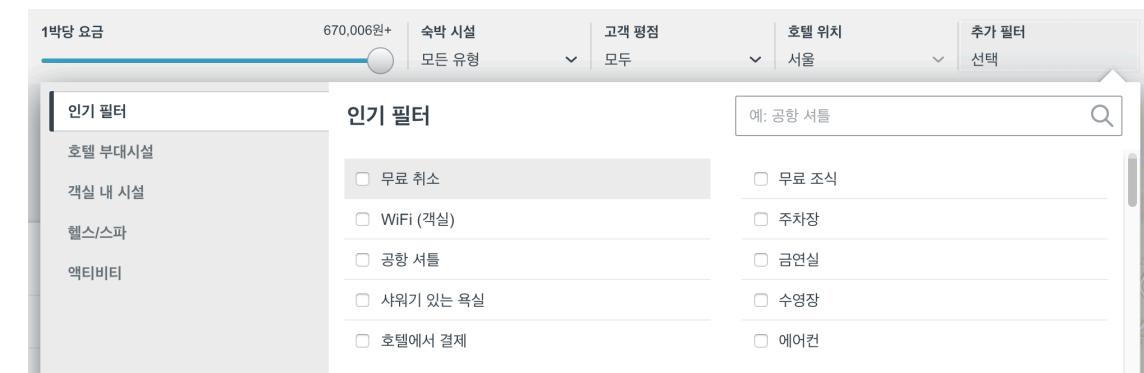
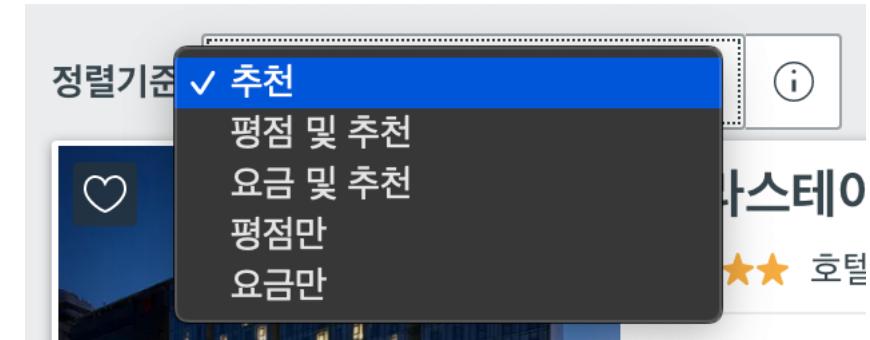
- **interaction item image:** user interacts with an image of an item.

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.

- **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.



## 2. Data Analysis: Action Type

• **search for item:** user searches for an accommodation.

The reference value for this action is the item id.

• **search for destination:** user searches for a destination.

The reference value for this action is the name of the destination.

• **search for poi:** user searches for a point of interest (POI).

The reference value for this action is the name of the POI.

• **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.

• **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.

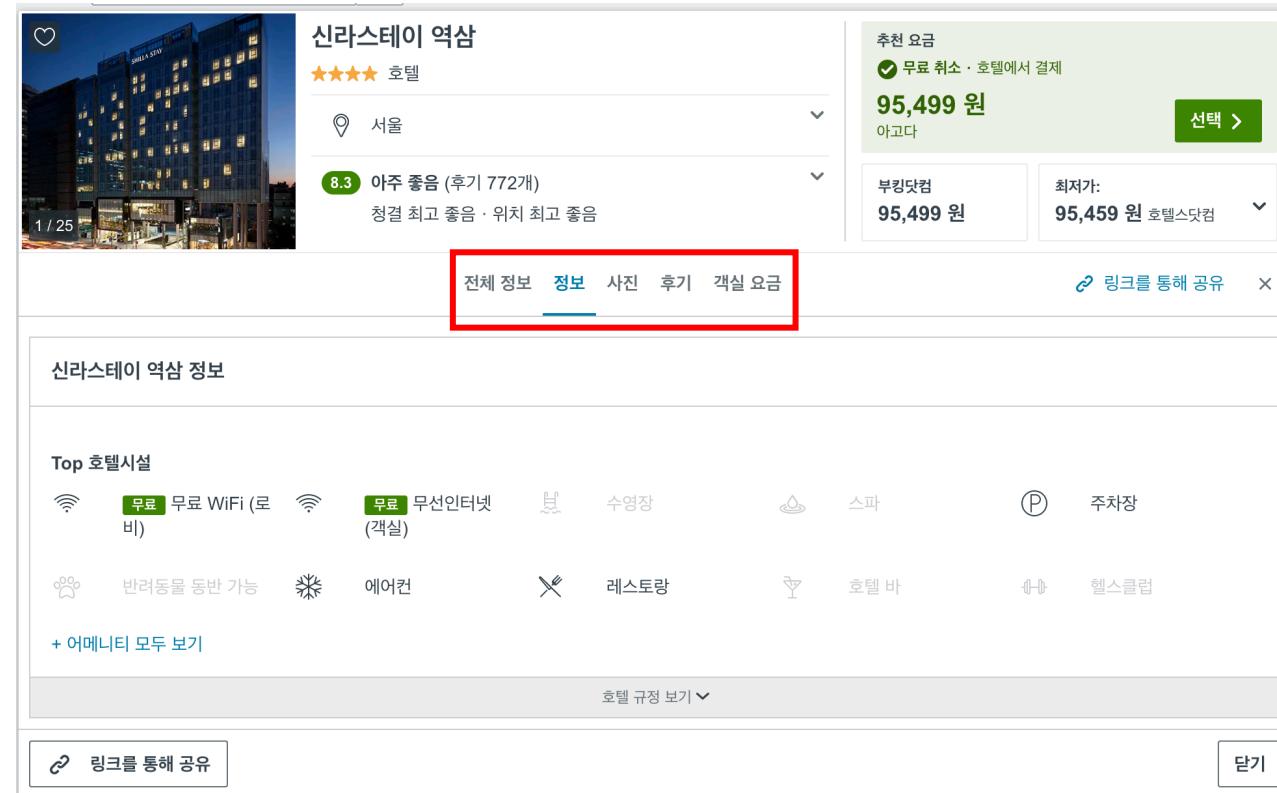
• **interaction item image:** user interacts with an image of an item.

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.

• **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.



## 2. Data Analysis: Action Type

- **search for item:** user searches for an accommodation.

The reference value for this action is the item id.

- **search for destination:** user searches for a destination.

The reference value for this action is the name of the destination.

- **search for poi:** user searches for a point of interest (POI).

The reference value for this action is the name of the POI.

- **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.

- **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.

- **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.



trivago > >

agoda  
● ● ● ●

trivago에서 파격 특가 상품 확인

아고다 웹사이트로 이동 중

# 1. Data Overview

step	action_type	reference	platform	city	device	current_filters	impressions	prices	dt
1	search for destination	Púchov, Slovakia	CL	Púchov, Slovakia	desktop	Nan	Nan	Nan	2018-11-01 21:21:36
2	search for destination	Púchov, Slovakia	CL	Púchov, Slovakia	desktop	Nan	Nan	Nan	2018-11-01 21:21:38
3	filter selection	5 Star	CL	Púchov, Slovakia	desktop	5 Star	Nan	Nan	2018-11-01 21:22:42
4	filter selection	4 Star	CL	Púchov, Slovakia	desktop	5 Star 4 Star	Nan	Nan	2018-11-01 21:22:49
5	search for destination	Pucón, Chile	CL	Pucón, Chile	desktop	5 Star 4 Star	Nan	Nan	2018-11-01 21:23:09
6	clickout item	327111	CL	Pucón, Chile	desktop	5 Star 4 Star	327111 2217612 492491 505751 5101488 2061220 3... 62 118 209 116 124 54 169 57 140 102 78 107 72...		2018-11-01 21:28:20
7	search for destination	Valdivia, Chile	CL	Valdivia, Chile	desktop	Nan	Nan	Nan	2018-11-01 22:05:50
8	filter selection	5 Star	CL	Valdivia, Chile	desktop	5 Star	Nan	Nan	2018-11-01 22:06:23
9	filter selection	4 Star	CL	Valdivia, Chile	desktop	5 Star 4 Star	Nan	Nan	2018-11-01 22:06:30
10	filter selection	5 Star	CL	Valdivia, Chile	desktop	4 Star	Nan	Nan	2018-11-01 22:07:33
11	filter selection	3 Star	CL	Valdivia, Chile	desktop	4 Star 3 Star	Nan	Nan	2018-11-01 22:07:37

# 1. Data Overview

step	action_type	reference	platform	city	device	current_filters	impressions
1	search for destination	London, United Kingdom	UK	London, United Kingdom	desktop	Nan	Nan
2	search for poi	Covent Garden	UK	London, United Kingdom	desktop	Nan	Nan
3	search for poi	Oxford Street	UK	London, United Kingdom	desktop	Nan	Nan
4	interaction item image	1032816	UK	London, United Kingdom	desktop	Nan	Nan
5	interaction item image	1032816	UK	London, United Kingdom	desktop	Nan	Nan
6	clickout item	1032816	UK	London, United Kingdom	desktop	Nan 12693 46363 81657 18448 47687 152913 18417 927...	
7	clickout item	1032816	UK	London, United Kingdom	desktop	Nan 12693 46363 81657 18448 47687 152913 18417 927...	

## 2. Data Analysis

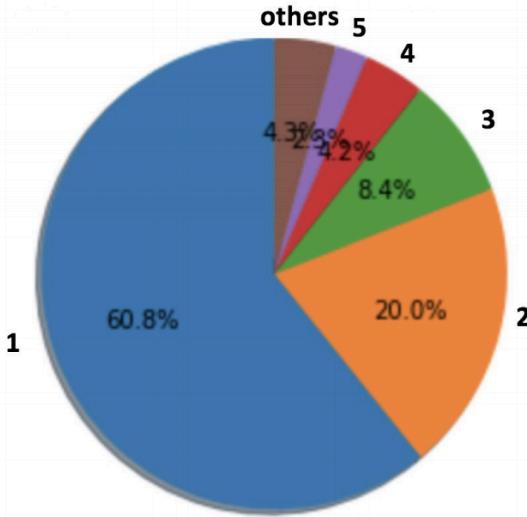


Figure 2: Clickout Number in a Session Distribution

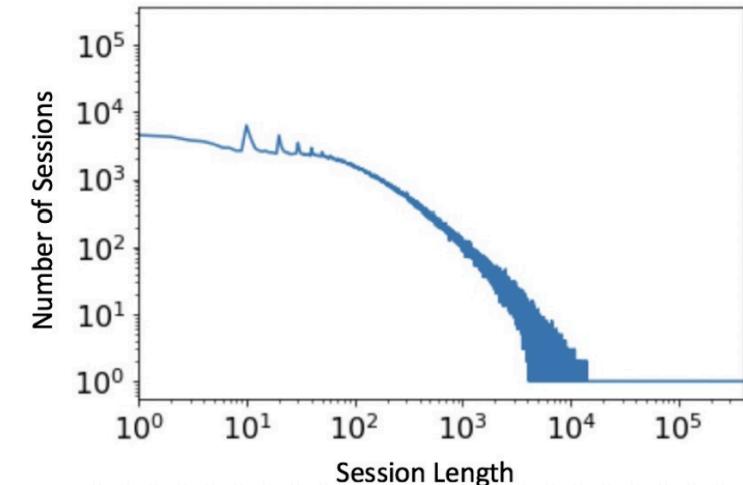
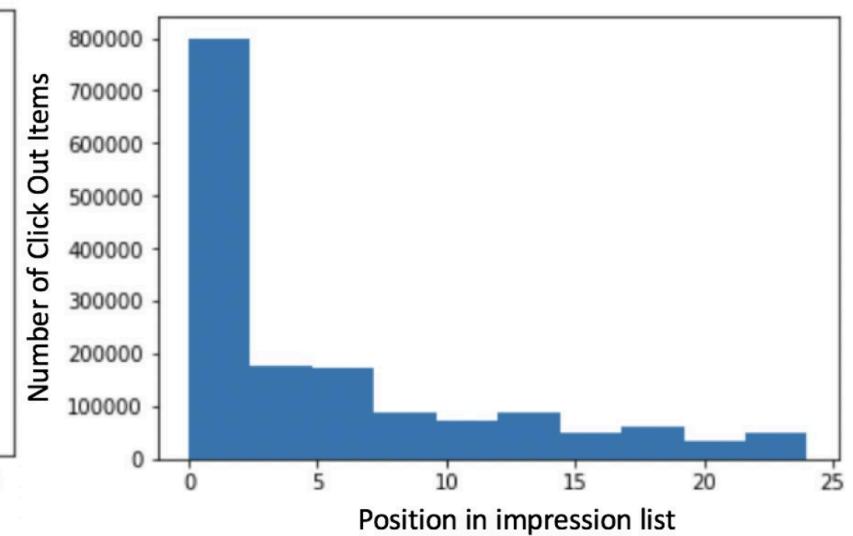
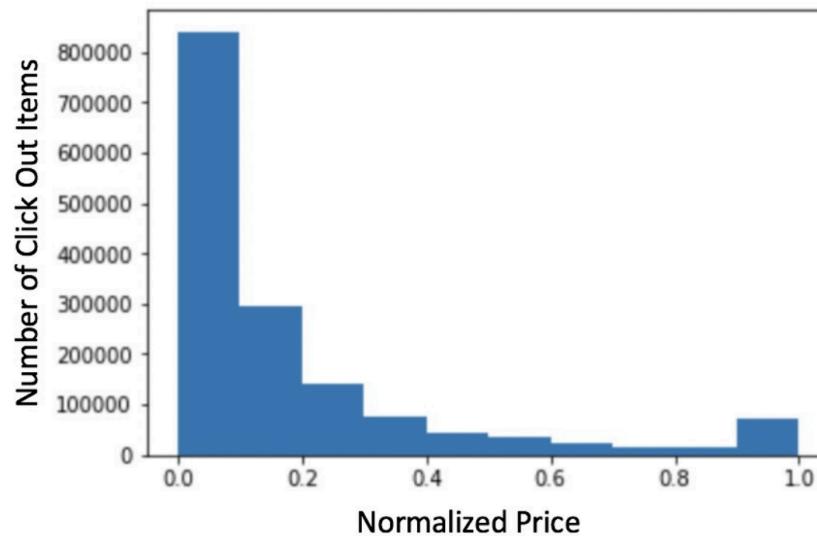
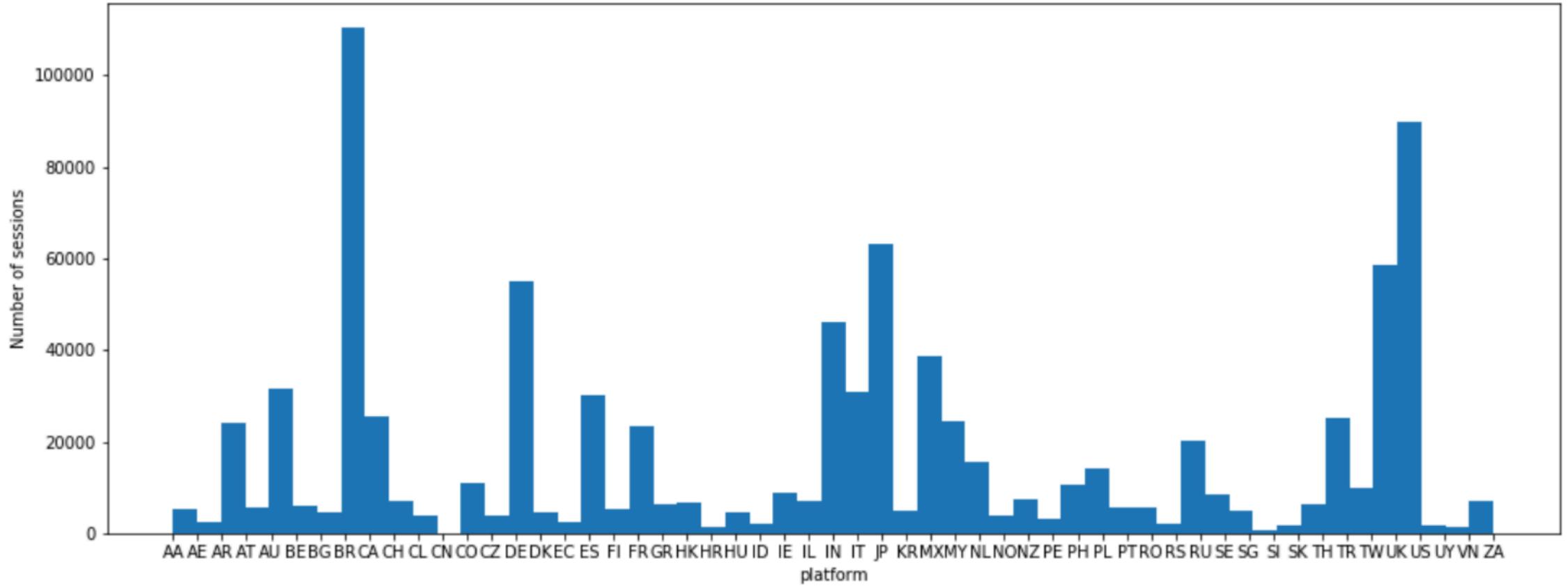


Figure 3: Session Length Distribution



## 2. Data Analysis



## 2. Data Analysis : Feature Engineering

### Session Context

- Item actions that precede clickout
- Non-item actions that precede clickout
- Number of session steps
- Time duration of session
- Device used

### Item Session Statistics

- Price and price rank in impression
- Rank position in impression
- Previous actions on this item in current session
- Item dwell time in current session
- Item properties metadata

### Item Global Statistics

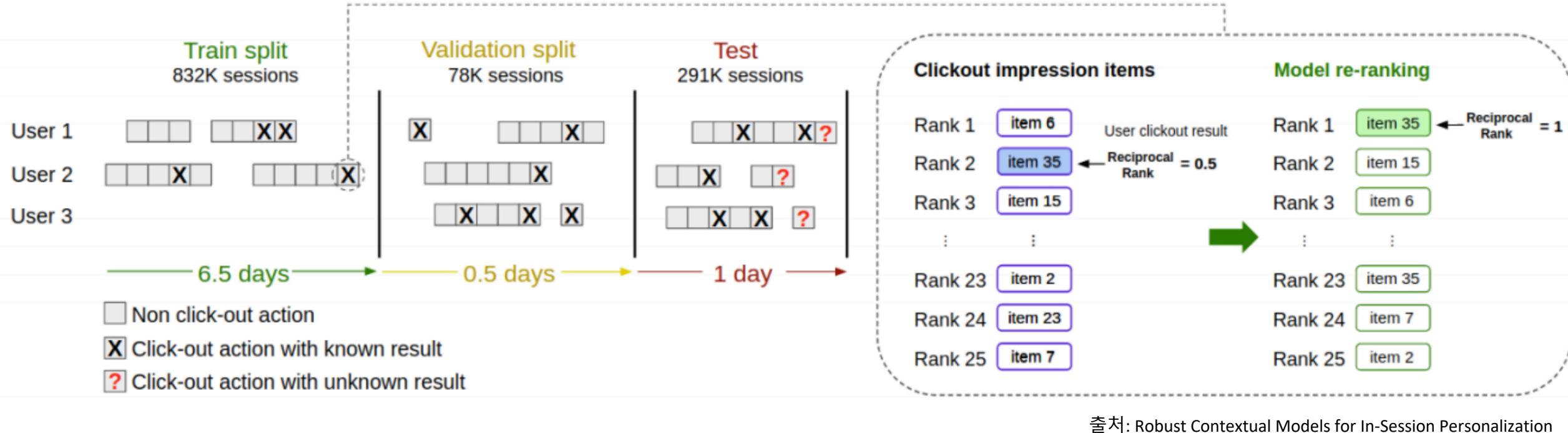
- Global item action counts across sessions
- Impression statistics across sessions such as average position, price and price rank
- Number of unique users that interacted with this item
- Difference between current session and global statistics

### Impression Items Statistics

- Price summary across impression items
- Metadata properties across impression items
- Item similarity within impression
- Global statistics summary across impression items



### 3. Problem Definition: Re-ordering



### 3. Problem Definition: Binary Classification

session_id		user_id	session_id	timestamp	step	action_type	reference	platform	city	device	current_filters	impressions
89643988fdbfb	3090782	000324D9BBUC	89643988fdbfb	1541593942	10	clickout item	NaN	BG	Budapest, Hungary	desktop	Car Park	923407 1729121 1050792 97171 353141 106315 218...
9de47d9a66494	1030763	0004Q49X39PY	9de47d9a66494	1541641157	1	clickout item	NaN	PH	Iloilo City, Philippines	mobile	NaN	2213014 3184842 10213134 4504242 4486372 38120...
beea5c27030cb	1030764	0004Q49X39PY	beea5c27030cb	1541561202	1	clickout item	NaN	PH	Iloilo City, Philippines	mobile	NaN	3812004 3505150 3202894 2292254 1984229 222789...
9617600e1ba7c	2997882	00071784XQ6B	9617600e1ba7c	1541630328	2	clickout item	NaN	IT	Turin, Italy	desktop	NaN	22721 3067559 16121 22727 22854 22819 22764 14...
2d0e2102ee0dc	77194	0008BO33KUQ0	2d0e2102ee0dc	1541636411	6	clickout item	NaN	BR	Canela, Brazil	mobile	NaN	507861 2176280 1669587 502066 1352530 4342348 ...

	user_id	session_id	timestamp	step	item_recommendations
0	000324D9BBUC	89643988fdbfb	1541593942	10	924795 106315 1033140 119494 101758 903037 105...
1	0004Q49X39PY	9de47d9a66494	1541641157	1	3505150 3812004 2227896 2292254 3184842 222702...
2	0004Q49X39PY	beea5c27030cb	1541561202	1	4476010 3505150 3812004 2227896 2292254 222702...
3	00071784XQ6B	9617600e1ba7c	1541630328	2	22854 3067559 22721 22713 16121 22772 22727 22...
4	0008BO33KUQ0	2d0e2102ee0dc	1541636411	6	9857656 5849628 655716 1352530 502066 1405084 ...



## 4. Plan

1. 데이터 전처리
2. Feature 엔지니어링
3. Baseline 모델 선정
4. LightGBM, XGBoost, CNN, LSTM, Transformer 등 모델 적용 및 벤치마크
5. 실시간 처리 inference 모델 개발

1. Item\_id 기반하여 실제 상품 정보 크롤링 - 일치 확인
2. 호텔 추천 웹사이트 제작
  1. 도시 및 impression 고정
  2. 여러 interaction log 수집 및 전송 서버 제작
  3. 수집한 정보 실시간 적재 및 전처리
  4. 라이브 모델 기반 추론 및 백엔드 서버 반영

# Appendix

**Table 3: Feature importance (top 20 features)**

Feature name	Description	Feature importance
session_start_ts	Time since session started	2431
price_vs_mean_price	The price of the hotel relative to the average price of all hotels in query	2147
clickout_step_rev	How many steps until the last step	2049
clickout_item_item_last_timestamp	Time since the last clickout clicked	1963
rank	Rank of the hotel	1927
clickout_prob_time_position_offset	Probability of a click given the time difference and position of the last click	1915
last_price_diff	What is the price difference between this hotel and the previous click	1913
mean_rank_counter_mean	Mean across historical mean rank for hotels on the impression list	1772
avg_price_similarity	What is the difference between the price and average prices of clicked hotels	1390
clickout_max_step	The maximum step	1218
mean_rank_counter_min	The minimum across historical mean rank for hotels on the impression list	1062
user_start_ts	Time since the first user event	~
clickout_item_ctr_rank_weighted	CTR of a hotel weighted by the rank	
clickout_item_ctr_corr	CTR of a hotel (impressions are corrected)	
clickout_counter_vs_interaction_counter_pure	Hotel's total number of clicks divided by the total number of interactions	
item_clickouts_intersection	How similar is an impression to the previous impression	
item_id	Item ID (original numerical value)	
clickout_item_uniq_prob	How unique are the hotels that are clicked by a user	
interaction_item_image_item_last_timestamp	Time since the last interaction with the hotel images	
price_pct_by_platform	What is a price of a hotel relative to the platform prices	
interact_item_uniq_prob	How unique are the hotels that a user interacted with	
clickout_item_ctr_corr_by_platform	CTR of the hotel (impressions are corrected) – grouped by the platform	

**Table 4: Hyper parameters of the models**

Boosting type	Objective function	Learning rate	Number of trees	Max leaves	Min data in leaf	Drop rate	Bagging fraction	Feature fraction	Max position	XGBoost dart mode	MRR	Blending coefficient
dart	NDCG	0.2	5000	64	5	0.015	0.8	0.7	20	False	69.068	0.51223
dart	NDCG	0.2	5000	64	5	0.03	0.8	0.7	20	False	69.026	0.76749
dart	NDCG	0.2	3200	64	5	0.1	1	1	20	False	69.018	0.37716
dart	NDCG	0.2	5000	64	5	0.03	0.8	0.7	20	False	69.011	0.47778
dart	NDCG	0.2	4800	32	5	0.1	1	1	20	False	69.010	0.33917
dart	NDCG	0.2	3200	64	5	0.1	1	1	20	False	69.006	0.25198
dart	NDCG	0.2	3200	64	5	0.1	1	1	20	True	69.002	0.58178
dart	NDCG	0.2	5000	64	5	0.015	0.8	0.7	20	False	69.001	0.79173
dart	NDCG	0.2	3200	64	5	0.1	1	1	25	False	68.995	0.00569
dart	NDCG	0.2	3200	64	3	0.1	1	1	25	False	68.985	0.00161
dart	MRR3	0.2	3200	64	5	0.1	1	1	20	False	68.970	-0.12063
dart	NDCG	0.2	1600	128	5	0.1	1	1	20	False	68.949	0.15376
dart	MRR3	0.2	1600	128	5	0.1	1	1	20	False	68.912	-0.30801
dart	NDCG	0.2	1600	128	5	0.1	1	1	20	False	68.894	0.13886
dart	NDCG	0.2	1000	256	5	0.03	1	0.5	20	False	68.860	-0.32652
dart	NDCG	0.2	1600	128	5	0.1	1	1	20	False	68.825	-0.10392
gbdt	NDCG	0.1	5000	62	5	-	1	1	20	False	68.783	0.17662
dart	NDCG	0.4	3200	64	5	-	1	1	25	False	68.782	0.40968
gbdt	NDCG	0.2	1600	62	20	-	1	1	20	False	68.773	0.32795
gbdt	NDCG	0.1	1600	62	20	-	1	1	20	False	68.768	-0.27577
gbdt	NDCG	0.1	5000	62	5	-	1	1	20	False	68.766	-0.10544
gbdt	Ranker3	0.1	1600	62	20	-	1	1	20	False	68.755	-0.46916