UCI Department of Mathematics

MSF AT UCI

Dublin Data Analysis

By Songhan, Jiaye, Luoning

Overview

01 Introduction

02 Data Preprocessing

03 Visualization

04 Modeling

05 Conclusion

06 Future Research

Objectives

1. Understand the User Portrait

2. Predict the Customer’s Stage in the Conversion Funnel

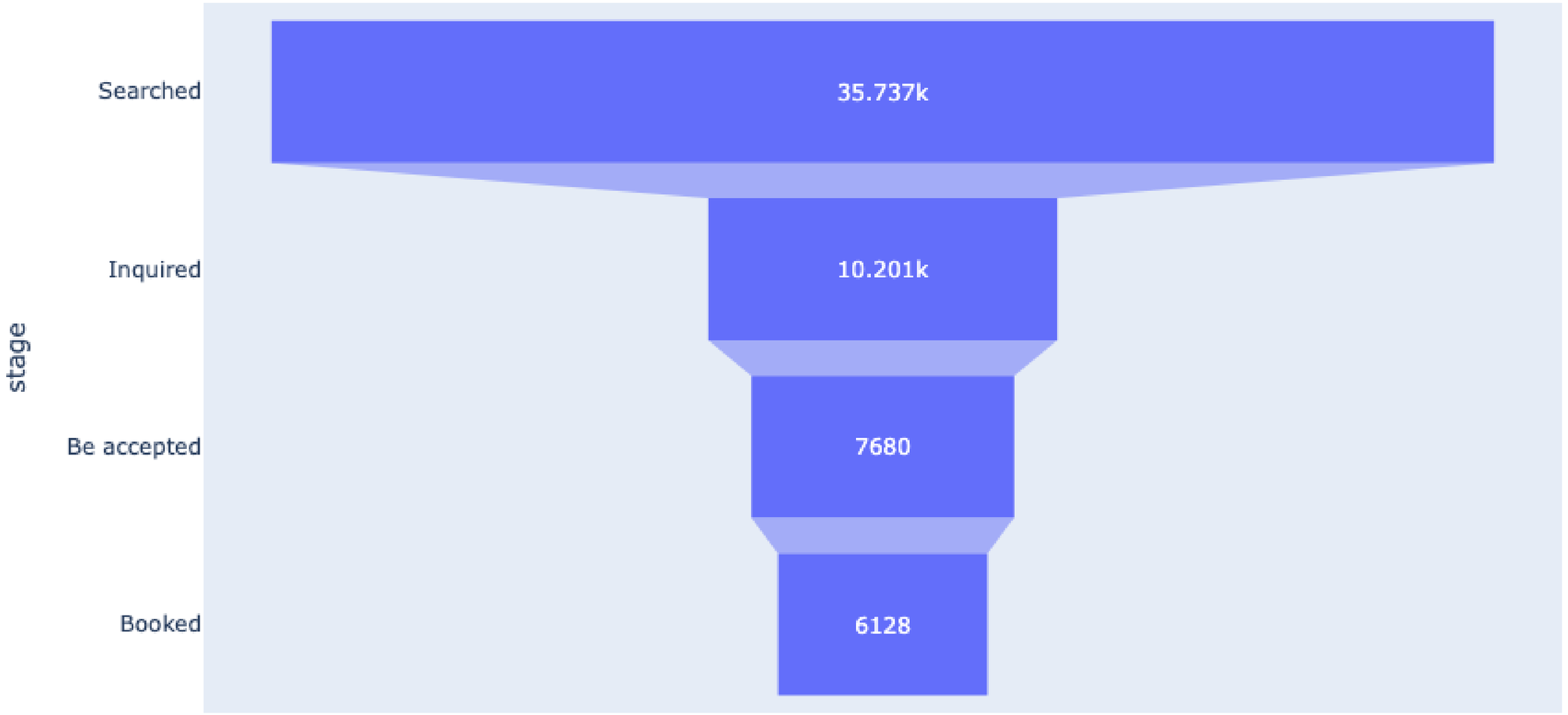
3. Model Training and Visualization

Data Overview

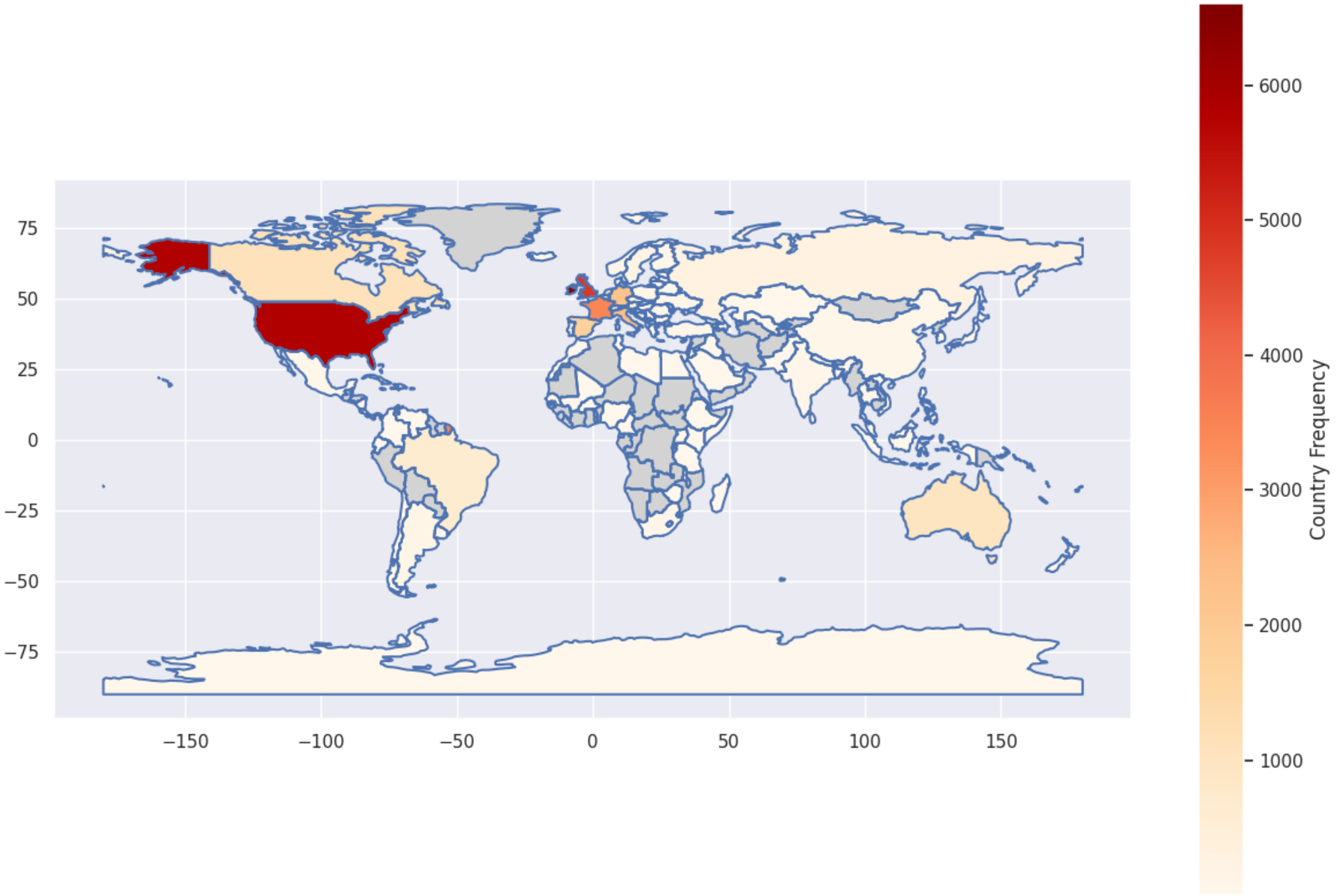
searches.tsv Ds	searches.tsv Id User	searches.tsv Ds Checkin	searches.tsv Ds Checkout	searches.tsv N Searches	searches.tsv N Nights	searche... N...	searche... N...	searches.tsv Origin...	searches.tsv Filter ...	searches.tsv Filter P...	searches.tsv Filter Ro...	searches.tsv Filter Neighborhoods
2014/10/1	0000af0a-6...	2014/10/9	2014/10/12	16	3	2	2	IE	0	67	,Entire home/...	*null*
2014/10/1	0000af0a-6...	2014/10/9	2014/10/19	3	10	1	2	IE	0	67	*null*	*null*
2014/10/1	000cd9d3-...	*null*	*null*	1	*null*	1	1	GB	*null*	*null*	*null*	*null*
2014/10/1	000cd9d3-...	2014/11/8	2014/11/10	11	2	1	2	GB	*null*	*null*	,Entire home/...	*null*
2014/10/1	001c04f0-5...	*null*	*null*	1	*null*	1	1	IE	*null*	*null*	*null*	*null*
2014/10/1	004e88da-...	2014/10/3	2014/10/5	7	2	5	5	SE	*null*	*null*	*null*	*null*
2014/10/1	00623353-...	2014/11/1	2014/11/9	6	8	1	1	IE	*null*	*null*	Entire home/...	*null*
2014/10/1	00623353-...	2015/3/30	2015/4/4	1	5	1	1	IE	*null*	*null*	Entire home/...	*null*

contacts.tsv Id Guest	contacts.tsv Id Host	contacts.tsv Id Listing	contacts.tsv Ts Contact At	contacts.tsv Ts Reply At	contacts.tsv Ts Accepted At	contacts.tsv Ts Booking At	contacts.tsv Ds Ch...	contacts.tsv Ds Checkout (C...	contacts.tsv N Guests	contacts.tsv N Messages
00197051-c6c...	46aa3897-9c...	fb5ed09a-984...	2014/11/4 09:10:03	2014/11/4 09:4...	2014/11/4 09:45:50	2014/11/4 12:20:46	2014/11/27	2014/11/30	1	10
0027538e-aa9...	6bbb88ca-d...	d3871da6-801...	2014/10/10 12:02:50	2014/10/10 15:...	*null*	*null*	2014/10/17	2014/10/19	2	3
0027538e-aa9...	8772bc85-a9...	0d9b5583-80...	2014/10/10 15:23:53	*null*	*null*	*null*	2014/10/17	2014/10/19	2	2
0027538e-aa9...	ac162061-55...	ec68e0af-b0f...	2014/10/10 15:22:26	2014/10/10 15:...	2014/10/10 15:24:26	2014/10/10 15:52:42	2014/10/17	2014/10/19	2	14
006b68ec-937...	16a2ccae-60...	f375e1c9-3d5...	2014/10/4 09:15:47	2014/10/4 14:4...	*null*	*null*	2014/10/4	2014/10/5	3	6
006b68ec-937...	42bae547-e5...	9d8e30e9-dc...	2014/10/2 00:21:08	*null*	*null*	*null*	2014/10/4	2014/10/5	3	4
006c93e2-199...	ee55912a-88...	71bca3ca-cbbf...	2014/10/2 04:09:28	2014/10/2 06:5...	*null*	*null*	2014/11/5	2014/11/7	2	2
007a3626-1c7...	d5bf9afd-957...	d8a2dfe9-3ba...	2014/12/29 18:39:31	2014/12/29 23:...	*null*	*null*	2015/2/13	2015/2/16	2	2

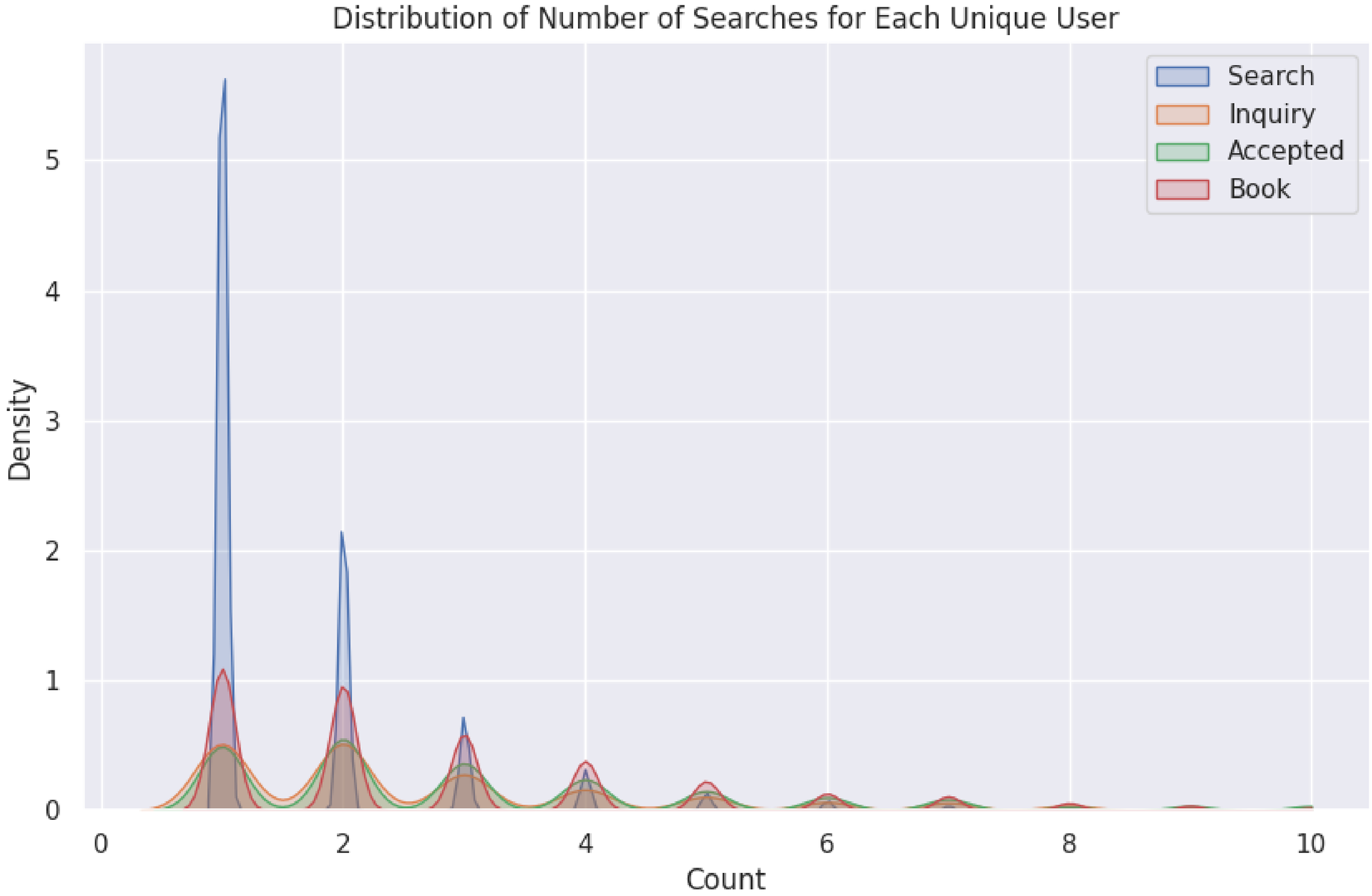
Introduction



Visualization

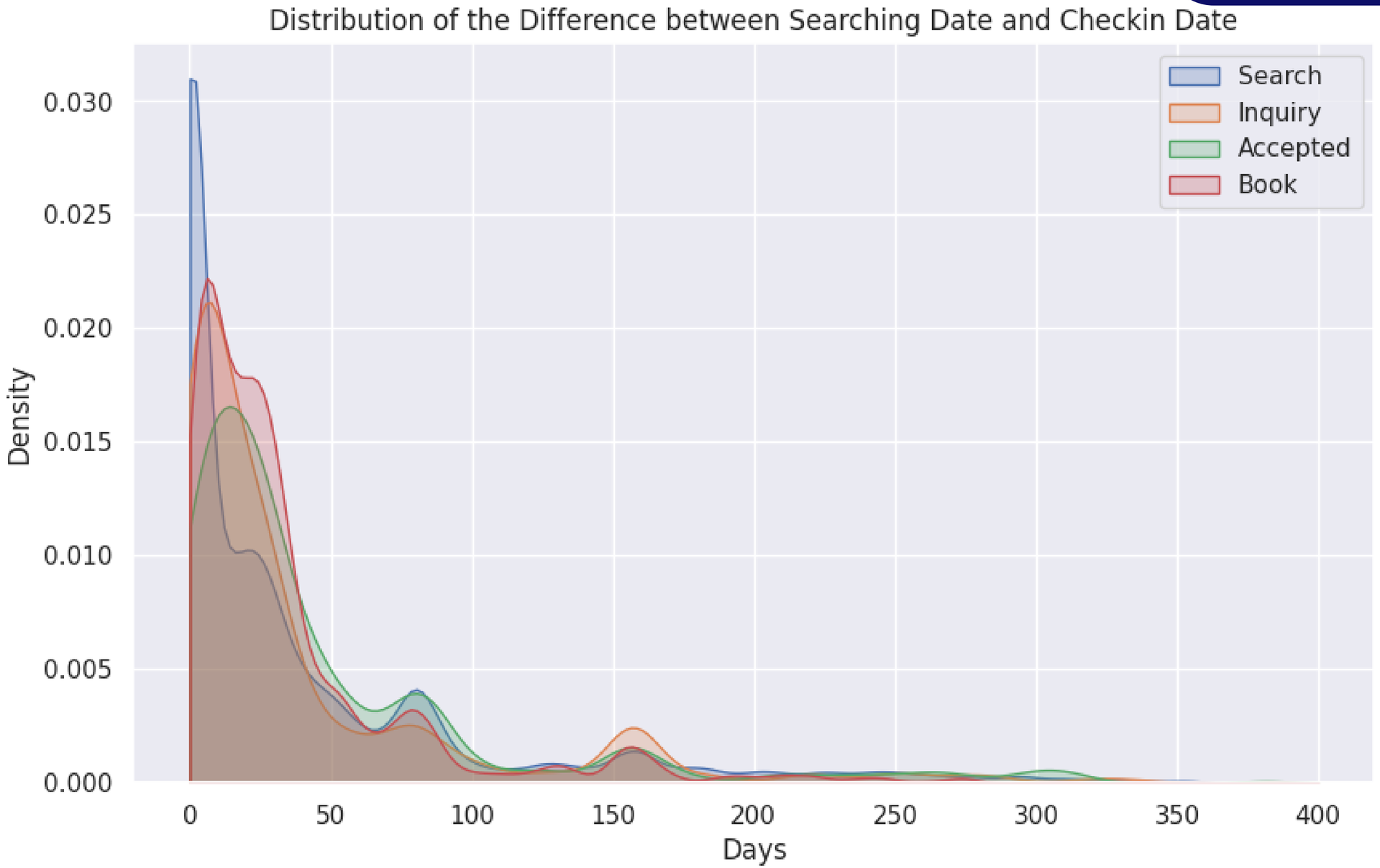


Visualization



More Searching Time Leads to Higher Conversion Rate

Visualization



Searchings made around 25 days before the check-in date has a better conversion rate

Data Preprocessing

01 Data Cleansing and Manipulation

1. NaN value
2. Outlier
3. Level of Data
4. Data Merge

02 Variable Choosing

1. Independent Variable
2. Dependent Variable

```
(['id_user', 'num_of_searches', 'n_night_mid', 'n_night_var',
  'n_guests_max_mid', 'n_guests_min_mid', 'filter_price_min_mid',
  'filter_price_max_mid', 'num_room_types', 'num_neighborhoods',
  'day_diff', 'type_order', 'AD', 'AE', 'AL', 'AM', 'AQ', 'AR', 'AT',
  'AU'],
```

18605 rows × 143 columns pd.Dataframe ↗

Model

XGBClassifier

- A Tree-based model
- High-performance
- Accuracy score for validation dataset: 0.8213

Voting Classifier

- Combination of 5 widely-used model

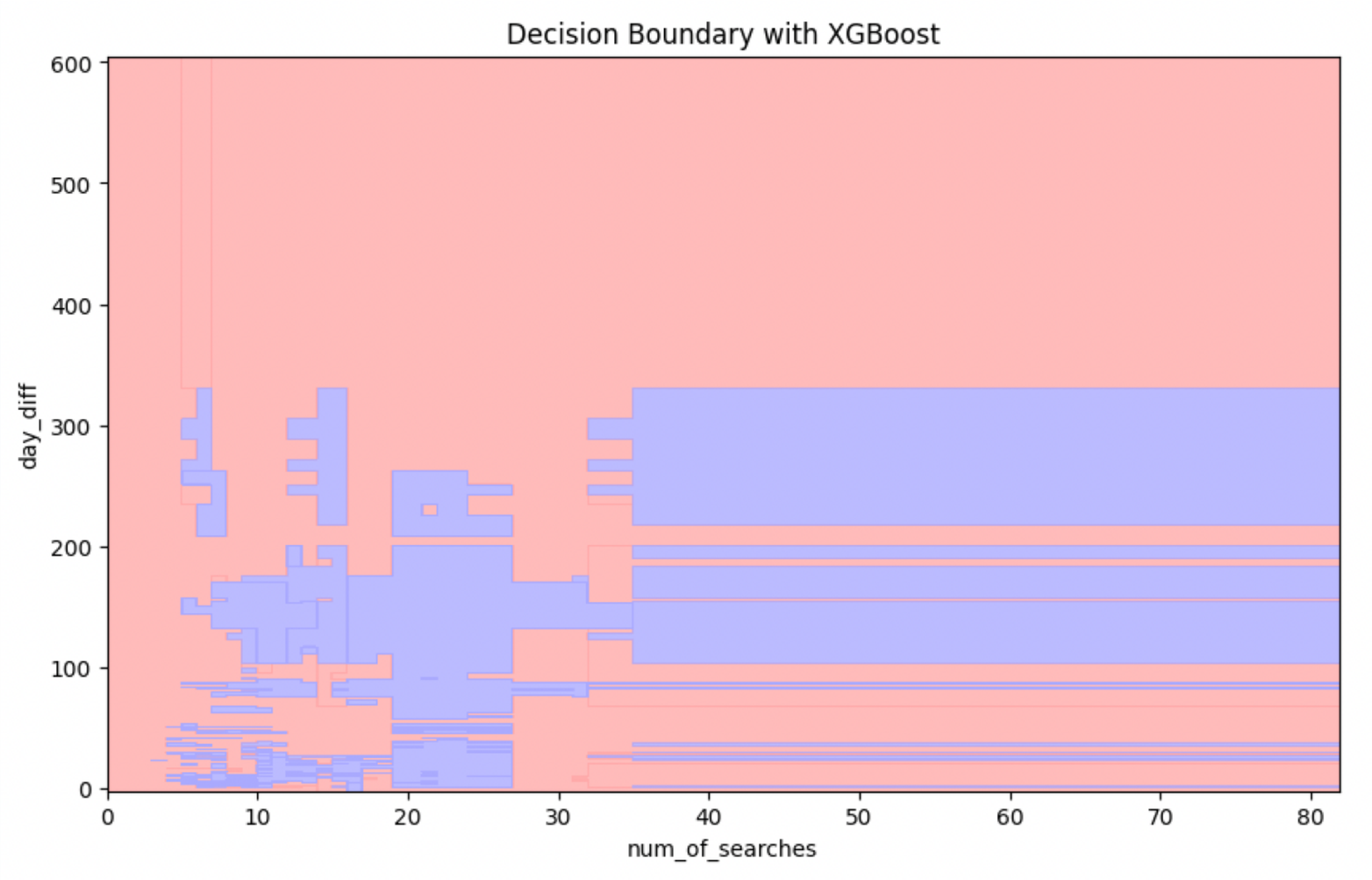
Hyperparameter Optimization

- Optuna
- Cross-validation score

Feature Importance

- Discover features that have a high impact on prediction

Decision Boundary



Conclusion

1. Marketing campaign should be performed around 25 days before the busy season for traveling

2. Encourage users to do more searching

3. Our model can predict the user's "drop out" stage, therefore we can make accurate action to avoid this.

Future Research

1. Expand data about the host (e.g. reply time)

2. Explore more Useful Features