AIRKIM

THE AIRBNB ASSISTANT

Team KIM powered Airbnb chatbot

Aug 2022

AIRKIM

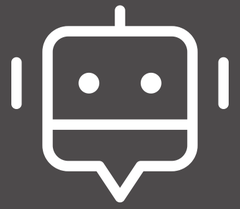
ABOUT

Our goal was to build a user-friendly, intuitive chatbot to help users make Airbnb vacation selections. We wanted to enhance the user experience by making the the text features on Airbnb listings filterable as part of the content recommendation process.

Data was sourced for all Airbnb listings in Hong Kong available from 14 Jun 2022 to 13 Jun 2023

NATURAL LANGUAGE PROCESSING

To deliver an intuitive user search experience, we used NLP methods to process text features of >5,600 Airbnb listings to be machine analyzable



RASA CHATBOT

Results delievered to user via a front-end chatbot , built on Rasa Open Source. NLU methods used to extract structured information from user messages.

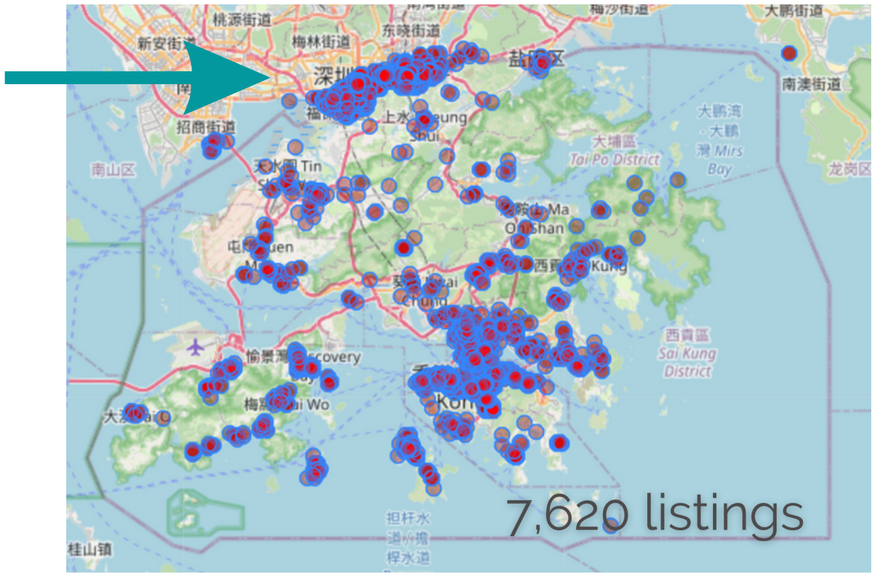
FLASK FRAMEWORK

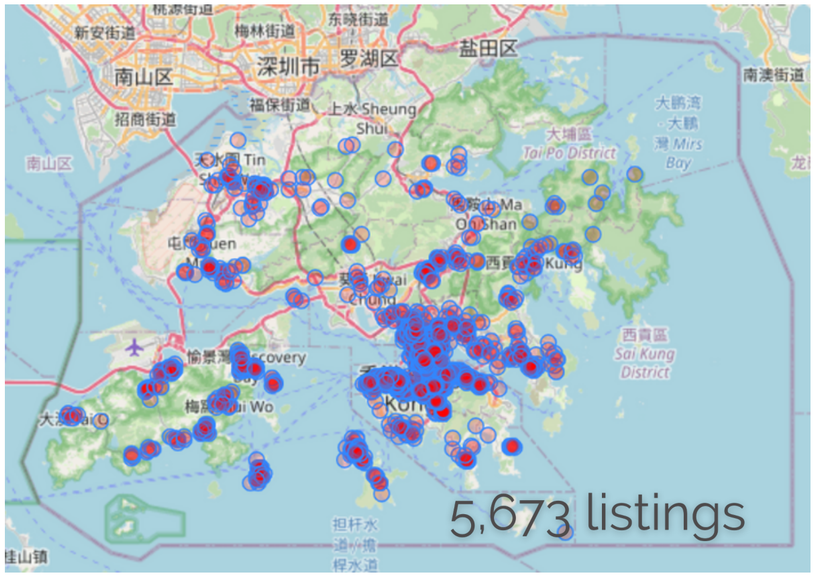
All functions integrated with a Flask framework, utilizing HTML, Javascript and Python. User experience enhanced using Bootstrap and user logins stored to SQL.

AIRKIM

DATA PREPROCESSING

During EDA, we found a quarter of the listings were in China; We geolocated these addresses and removed them from the final dataset





We also encountered about 20% of the data being non-English, which we had to translate

Original Text

我們的房源坐落於香港尖沙咀的中心地帶，在這裡，住客們可輕鬆前往市區內各大旅遊、購物、餐飲地點。 特色服務設施包括免費高速wifi，24小時安保，24小時前台，行李存放服務等 ...

Translated Text

Our property is located in the heart of Tsim Sha Tsui, Hong Kong, where guests can easily access the major tourist, shopping and dining spots in the city ...

AIRKIM

TOPIC EXTRACTION

We innovated upon Airbnb's native search by making owner-submitted text fields part of our search feature. To do so, we first used Natural Language Processing methods to prepare and embed the text, and then performed dimensionality reduction using Linear Discriminant Analysis to organize the unlabeled data into topics. This allowed the processed data to be compared for similarity.

Text is processed and given word embeddings

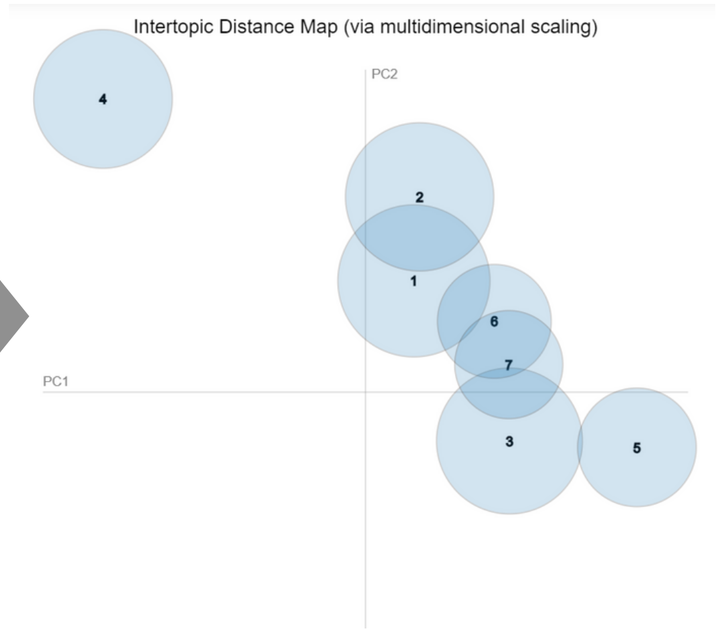
LDA applied to assign categorical groupings to text

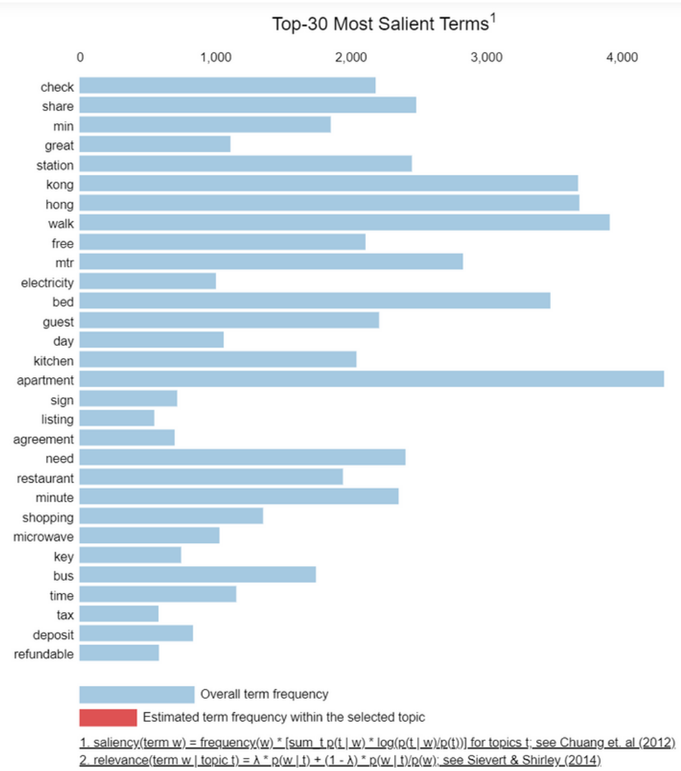
Topical groupings determined by keywords in text

Listing ID 17891 Description

Gorgeous and spacious loft, in the best location. Also available for Commercial Shoots and Longer term rentals ... Large & light-filled, designer open floor plan (800 sqf), fully furnished, with ultra-modern, open kitchen , available for short & long term. Includes: - full kitchen (stove top, oven) - fully equipped workstation with wi-fi & printer - balcony & large windows - queen size bed - walk-in wardrobe. ..

Topic, Relevance:
2, 0.54316676
3, 0.41077125
7, 0.0376049





AIRKIM

RASA CHATBOT

Natural Language Understanding

Allows RASA to understand user input

```
- intent: room_listing
  examples: |
    - [1](room_id)
    - [20](room_id)

- intent: get_started
  examples: |
    - I need assistance
    - Help Me
    - Started
    - get started
    - i want to book an airbnb
    - get me airbnb

- intent: return_listing
  examples: |
```

Domain

Defining environment within which the chatbot performs its tasks

```
utter_goodbye:
- text: "Thank you for chatting with me today. Have a nice day. Goodbye."
- text: "Thanks for stopping by, I hope to hear from you again!"

utter_welcome:
- text: "Hello! I am AIRKIM, an airbnb search assistant! How can I help?"
  buttons:
  - title: "Get room recomendation"
    payload: "/room_suggestion"

utter_submit:
- text: "All done!"

utter_slots_values:
- text: "I am going to search airbnb using the following parameters:\n
```

Stories

Managing dialogue interactions with user

```
stories:

- story: happy path
  steps:
  - intent: greet
  - action: action_get_suggestion

- story: getting started
  steps:
  - intent: get_started
  - action: utter_welcome
```

RASA

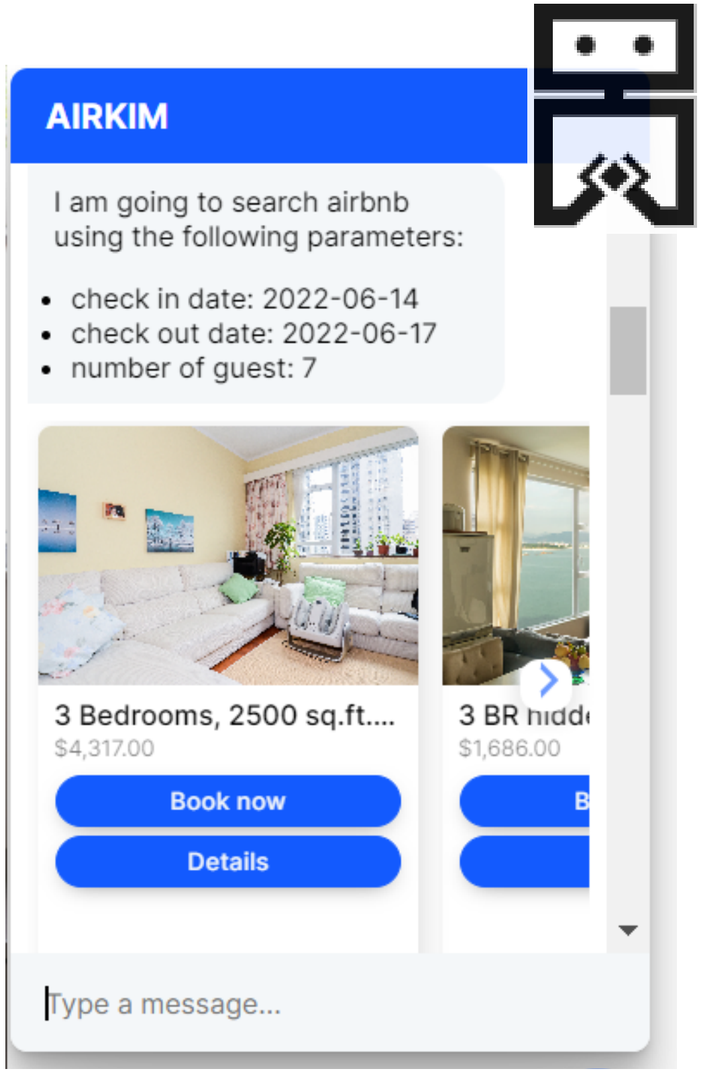
AIRKIM

RASA CHATBOT

Custom Search Actions

With the chatbot setup, we built-in actions allowing users to search our database for Airbnb listings and to finetune their search through content-based filtering

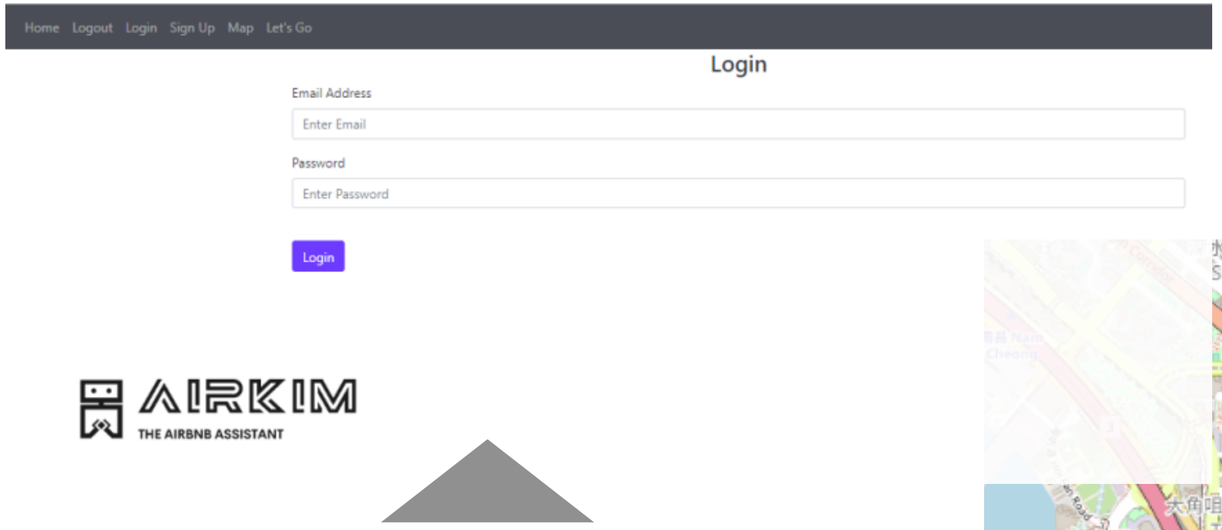
```
def searchListing(checkIn = '2022-06-14', checkOut = '2023-06-13', numGuests = 2):
    # convert dates to datetime and get length of stay
    checkIn, checkOut = pd.to_datetime(checkIn), pd.to_datetime(checkOut)
    stayLen = int((checkOut - checkIn) / np.timedelta64(1, 'D'))

    # return listings that match chosen period of stay
    dfStay = vacancyHk.query('minimum_nights >= @stayLen <= maximum_nights').query('dat
    dfList = dfStay.groupby('listing_id').date.count()
    availListing = [dfList.index[i] for i in range(len(dfList)) if dfList.values[i] ==

    # return listings that match number of guests, period of stay
    return [listing for listing in availListing
            if listingHk[listingHk.id == listing].accommodates.values[0] >= numGuests]
```



AIRKIM

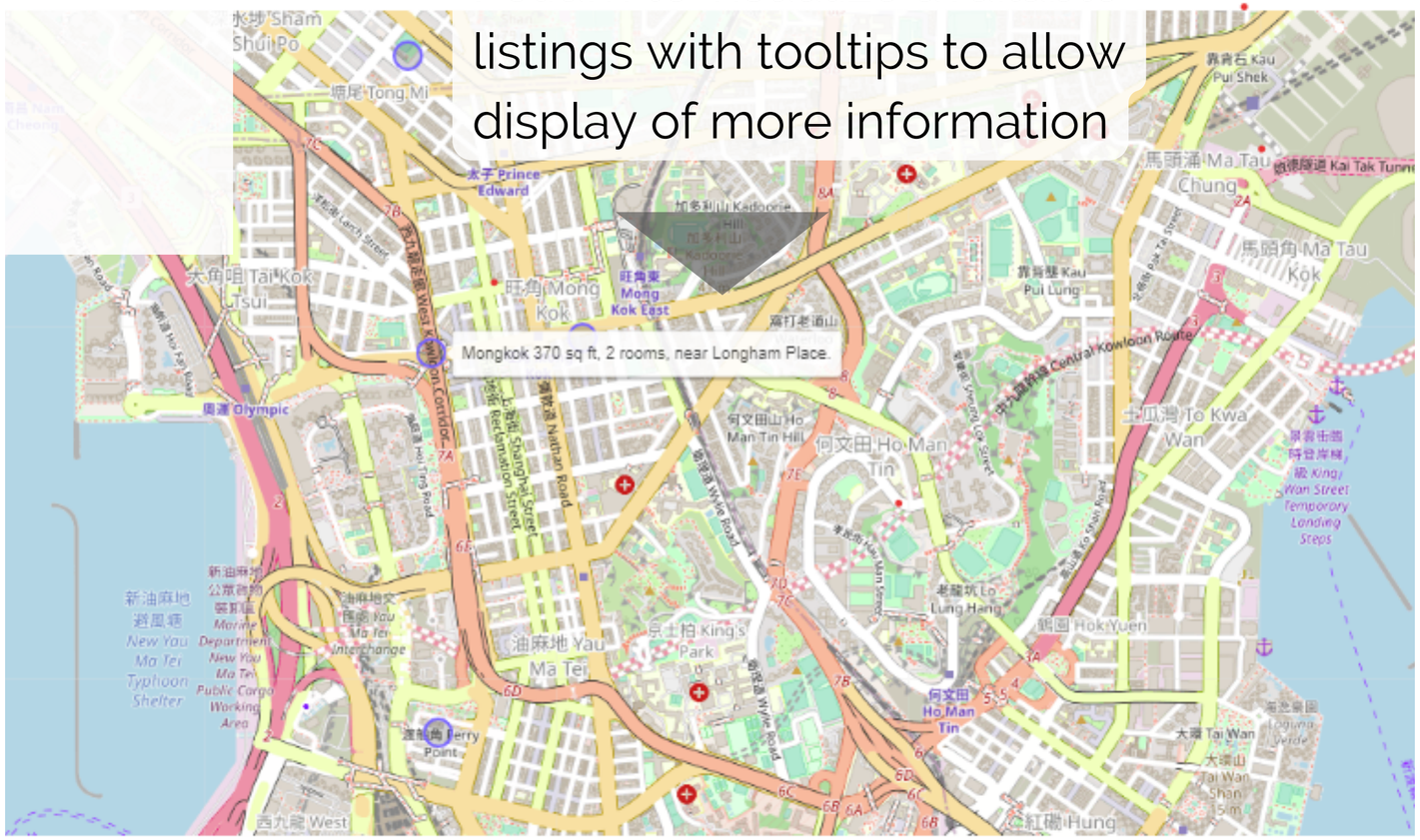
FLASK INTEGRATION

We built a webpage to host the AIRKIM chatbot and added features to improve the user experience



A user login function was introduced. Login information is stored on SQL with passwords hashed and subject to sha256 encoding to maximize website security

An interactive map feature was implemented using Folium to visualize available listings with tooltips to allow display of more information



AIRKIM

CHALLENGES

NLP

- Tried multiple methods to extract context from data but not all successful
- More sophisticated methods out of reach

RASA

- Not intuitive to install and use
- Time-consuming to incorporate functions into the architecture

FLASK / Website

- Learning HTML, Jinga and Python integration
- Storing and passing user input variables within a session was a challenge

AIRKIM

UPGRADES

NLP

- Finetune ability of model to segment text data to improve data/topic labelling
- Incorporate review data into recommendations; We had performed simple Aspect-based Sentiment Analysis on review data but this provided limited insights to improve our model

```
Example 65: great apartment and great <decoration:Positive Confidence:0.9990229606628418> . the <bathroom:Negative Confidenc
e:0.9984305500984192> is a bit too small but functional .

Example 48044: the <location:Positive Confidence:0.9998146891593933> is good , but the <sound of the air conditioner:Negative
Confidence:0.9998100399971008> is a bit loud at night , but it is acceptable
```

RASA

- Increase intent and entity training to improve natural language functions of the chatbot
- Data is "sandboxed" currently; User experience would be enhanced if data can be passed between website and chatbot

FLASK / Website

- Expand interactivity with Folium geo-map, ideally be able to pass chatbot search variables into map
- SQL: Storing user interactions with website to analyze user search behavior for future implementation of collaborative filtering recommendations