

Market Analysis in Dublin

[Download Datasets](#)

This data project has been used as a take-home assignment in the recruitment process for the data science positions at Airbnb.

Assignment

A new city manager for Airbnb has started in Dublin and wants to better understand:

- what guests are searching for in Dublin,
- which inquiries hosts tend to accept.

Based on the findings the new city manager will try to boost the number and quality of hosts in Dublin to fit the demands from guests. The goal of this challenge is to **analyze, understand, visualize**, and **communicate** the demand / supply in the market. For example you may want to look at the breakdown of start date day of the week, or number of nights, or room type that is searched for, and how many hosts accepted the reservation. In particular, we are interested in:

- what the gaps are between guest demand and host supply that the new city manager could plug to increase the number of bookings in Dublin,
- what other data would be useful to have to deepen the analysis and understanding.

Data Description

There are 2 datasets

1. `searches.tsv` - Contains a row for each set of searches that a user does for Dublin
2. `contacts.tsv` - Contains a row for every time that an assigned visitor makes an inquiry for a stay in a listing in Dublin

`searches` dataset contains the following columns:

- `ds` - Date of the search
- `id_user` - Alphanumeric user_id
- `ds_checkin` - Date stamp of the check-in date of the search
- `ds_checkout` - Date stamp of the check-out date of the search
- `n_searches` - Number of searches in the search set
- `n_nights` - The number of nights the search was for
- `n_guests_min` - The minimum number of guests selected in a search set
- `n_guests_max` - The maximum number of guests selected in a search set
- `origin_country` - The country the search was from
- `filter_price_min` - The value of the lower bound of the price filter, if the user used it
- `filter_price_max` - The value of the upper bound of the price filter, if the user used it
- `filter_room_types` - The room types that the user filtered by, if the user used the room_types filter
- `filter_neighborhoods` - The neighborhoods types that the user filtered by, if the user used the neighborhoods filter

`contacts` dataset contains the following columns:

- `id_guest` - Alphanumeric user_id of the guest making the inquiry
- `id_host` - Alphanumeric user_id of the host of the listing to which the inquiry is made
- `id_listing` - Alphanumeric identifier for the listing to which the inquiry is made
- `ts_contact_at` - UTC timestamp of the moment the inquiry is made.
- `ts_reply_at` - UTC timestamp of the moment the host replies to the inquiry, if so
- `ts_accepted_at` - UTC timestamp of the moment the host accepts the inquiry, if so

- `ts_booking_at`` - UTC timestamp of the moment the booking is made, if so
- `ds_checkin`` - Date stamp of the check-in date of the inquiry
- `ds_checkout`` - Date stamp of the check-out date of the inquiry
- `n_guests`` - The number of guests the inquiry is for
- `n_messages`` - The total number of messages that were sent around this inquiry

Practicalities

Analyze the provided data and answer the questions to the best of your abilities. Include the relevant tables/graphs/visualization to explain what you have learnt about the market. Make sure that the solution reflects your entire thought process including the preparation of data - it is more important how the code is structured rather than just the final result or plot. You are expected to spend no more than 3-6 hours on this project.

