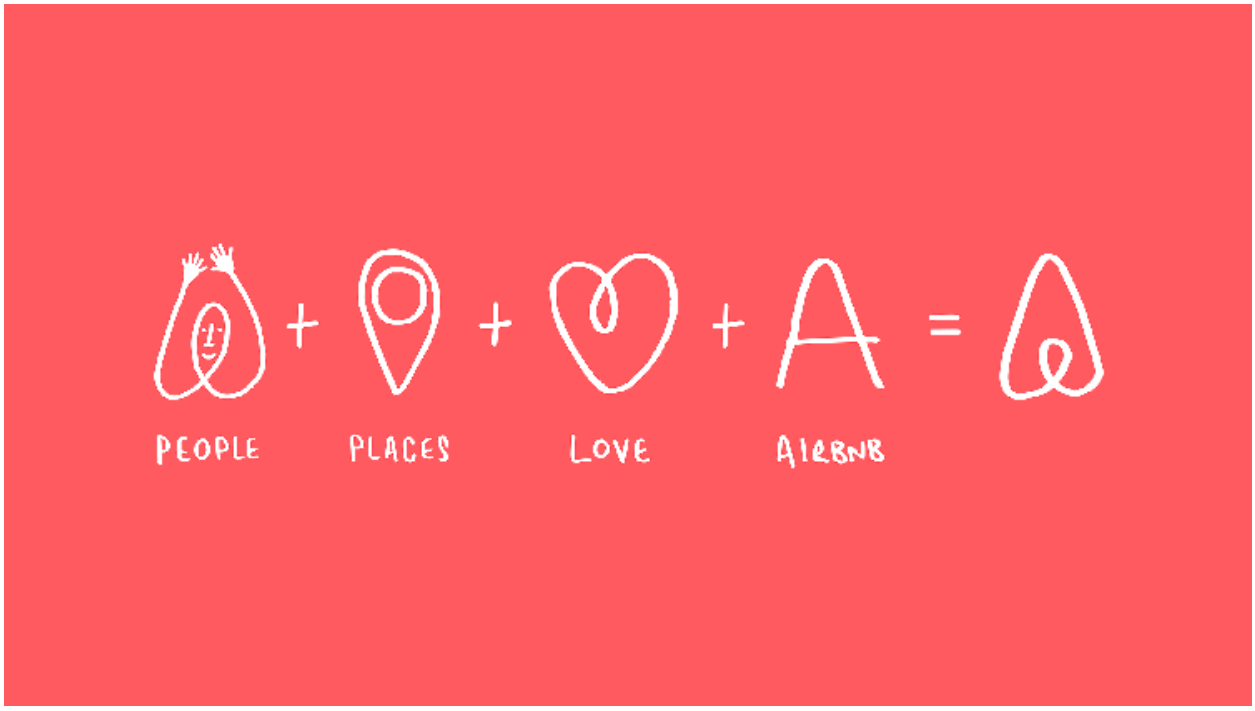

Airbnb Top Listings

Predicting the popularity of a Airbnb listing

06.02.2021

Valentina Rizzati
Data Scientist
Airbnb, Host Team

Opportunity

I am a Data Scientist for the Host team at Airbnb and I've recently been approached by a Product Manager who intends to enhance the set of Host tools with a dashboard visualizing, amongst other things, two categories of listings: top listing and not top listing. This is a binary classification problem that I intend to solve with a classification methodology (e.g. kNN or logistic regression).

As aligned with my business counterparts, the first prototype of this dashboard will be based on the Airbnb listings in NYC.

Impact Hypothesis

If hosts have more visibility on the expected performance of their listing, they will be able to optimize the listing's most critical features and work towards improving their performance and achieving the *Top Listing* status. The more listings and guests' experiences are optimized, the higher retention will be on the Airbnb platform.

Data

I will collect Airbnb NYC listings data from Inside Airbnb.

I will use the *Avail_365* (i.e. how many days will a specific listing be available for the next 365 days) as the discrete variable that will be used to identify the two classes in our binary problem. I will define *Top Listing* as a listing with *Avail_365* greater than the median of *Avail_365*.

Features like *host_response_time* and *property_type,* that are expected to impact the status of *Top Listing*, will be included in the model.

Solution Path

I am planning to follow the following solution path:

- Conduct data cleaning and initial EDA
- Create a new feature *top_listing* as a new column in the dataframe
- Run different types of classification models (e.g. kNN, Logistic Regression) and compare them through cross-validation
    - Interpret coefficients and identify most relevant features that a host should optimize for if their listing was classified as *Not Top Listing*

- Select optimal classification model
- Build Tableau Dashboard to visualize the results

Tools

To build the classification model I will use pandas, numpy, scikit-learn.

For diagnostic visualizations within the Jupyter Notebook I am planning to use seaborn and plotly.

Finally, to create the host-facing dashboard, I will use Tableau.

MVP Goal

As a MVP, I am planning to present a first version of the classification model and of the Tableau dashboard.