Using Text Analysis to Uncover the Motivators of Solo Travel

Dhruval Bhatt [email: dhruval@uchicago.edu]

Masters of Arts, Computational Social Science, University of Chicago

THE UNIVERSITY OF CHICAGO

Background and Objective

Solo travel leads the list of popular trends in travel yet little is known about the true motivations of why someone travels alone for leisure. Small scale qualitative analysis conducted by Constanza Bianchi[1] suggests top motivators include: personal factors such as independence or bravery, destination factors such as sites or landscapes and human interaction with locals and service providers. We seek to use computational methods on online accommodation reviews to uncover solo traveler motivators using experience satisfaction as the bridge.

The three main questions of this study are:
- Can online hostel reviews reveal patterns in expectation?
- Is there a distinction in travel expectation between solo travelers and other travelers?
- Do solo travelers value human interaction more than, less than or same as other traveler types?

Data

The source of data is Trip Advisor, one of the largest travel review sites that lists traveler type along with rating, date of review and text for reviews. A dataset of over 500 reviews from 3 hostels in San Francisco area, a top tourist destination, is collected. The data is limited to hostels as it has a large subset of solo travelers.

The graphs below summarize the distribution of traveler types in each dataset.
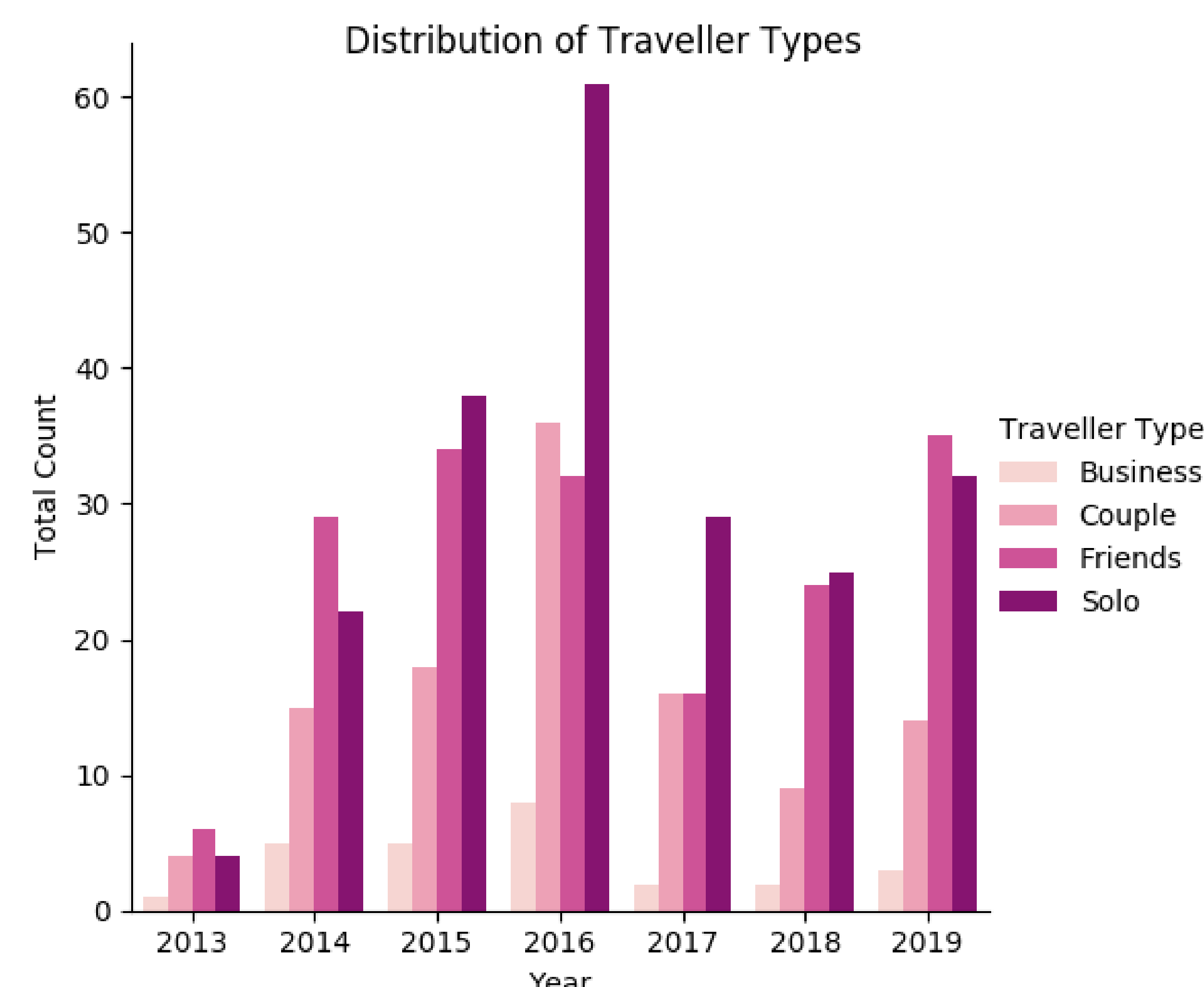
The distribution of solo versus non solo is roughly even

Exploratory Analysis
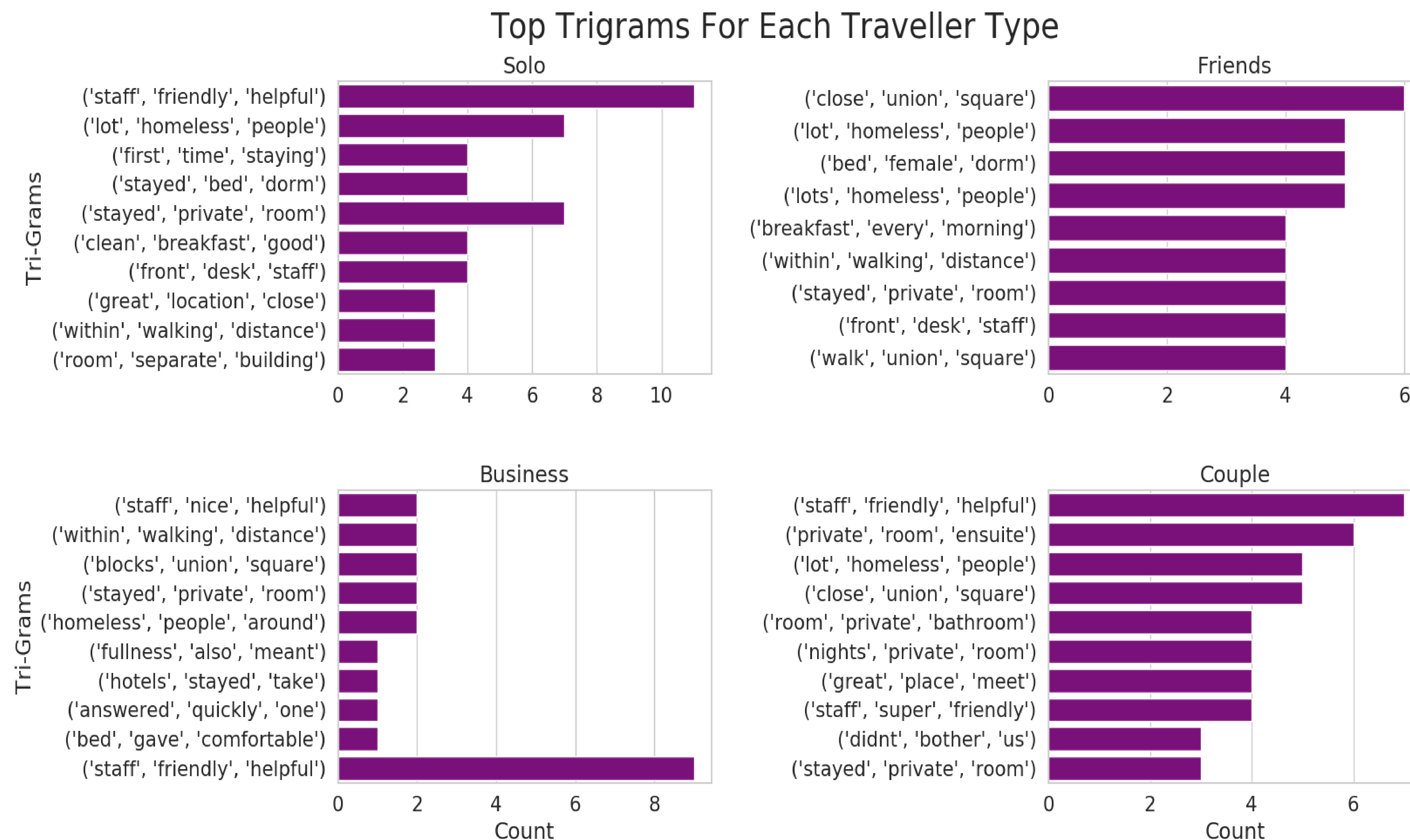
Different traveler types have different comments; Solo travelers value friendly staff and location

For high level insights of text from reviews, the top 10 trigrams for each of the category is plotted. This shows that solo travelers and friends are more similar than business and couple. Business travelers value comfort and promptness while couples value privacy.

Method Used
The text is processed to:
- Remove punctuation
- Convert to lower case
- Remove numbers
- Split sentences into words
- Remove stop words
- Remove domain specific words
- Create word pairs and count

Python Packages used:
NLTK, Collections, Seaborn

Predictive Model

To explore if solo travelers can be identified by using the reviews given, a simple predictive model is implemented.

Method Used:
- Use count tokenizer from sklearn to create bigrams
- Extract bigram count for the entire corpus of text
- Convert to dataframe and add column of traveler type
- Convert traveler type to numerical: 1 for solo and 0 for others
- Split the dataframe into train and test sets
- Train and test on a decision tree and random forest classifier

Python Packages Used: sklearn

Model Type	Test Accuracy	Time Taken
Decision Tree	62.1 %	41.4 s
Random Forest	61.9 %	7.3 s

Model accuracy to predict solo traveler based on reviews: ~60%

Conclusions

From the results of this analysis, it can be concluded:

- Using frequency count of words from review data reveals patterns in expectations of different travelers which can be used to infer motivation.
- There is a clear distinction between business, couple and friends and solo traveler expectations.
- Solo travelers highly value the friendly staff. This is similar to those travelling with friends. To distinguish these two clearly, a further focused study is required.

The major limitation of this study was the lack of a large corpus of reviews data. With a larger, more varied data, the results would be more interesting and generalizable.

References

[1] Bianchi, Constanza. "Solo Holiday Travellers: Motivators and Drivers of Satisfaction and Dissatisfaction." *International Journal of Tourism Research*, vol. 18, no. 2, 2015, pp. 197–208.