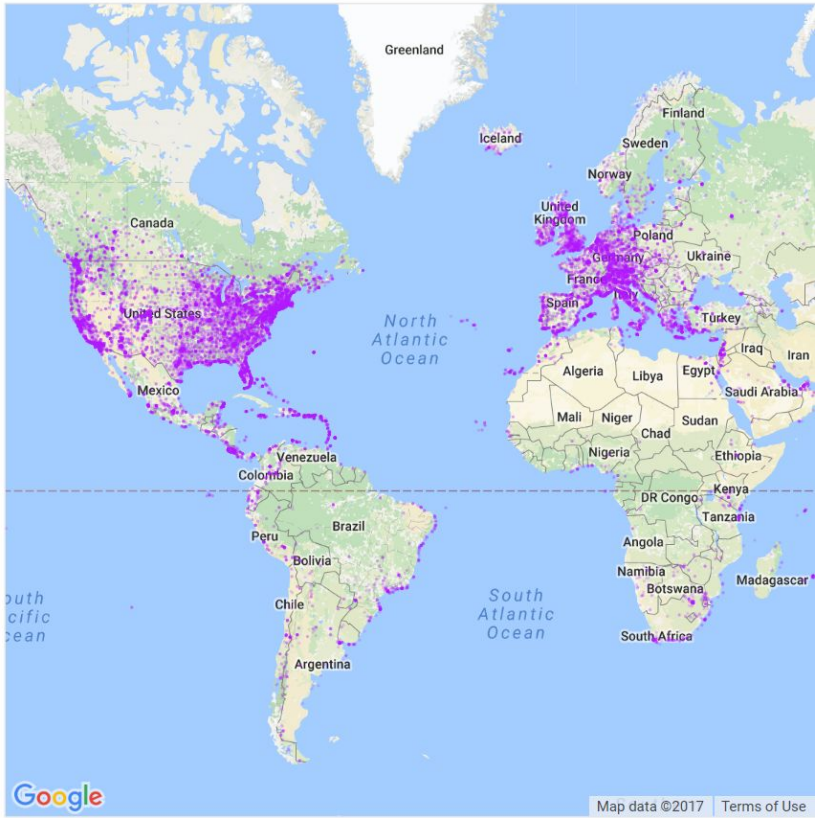
Reverse Engineering Data Secrets Time Delay Booking Predictions

Pandas
University of Massachusetts Amherst

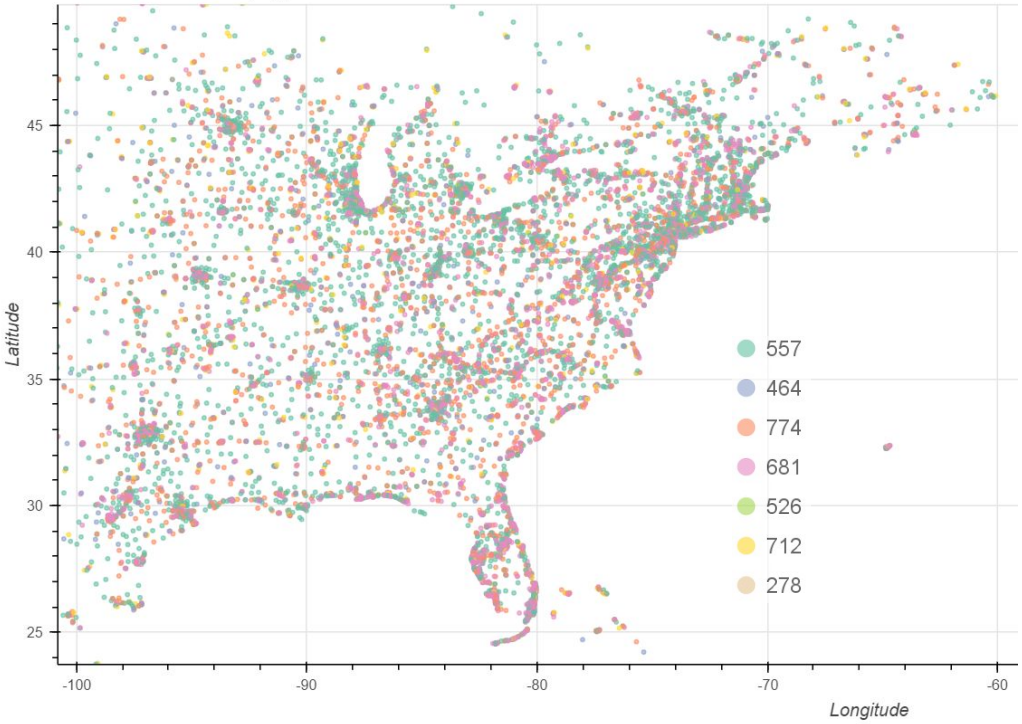
Ryan Cox, Kristina Yamkovoy, Anubhab Haldar, Hanfei Zhang

Popular Destination Searches

Search Destinations



Search Destinations by Type ID



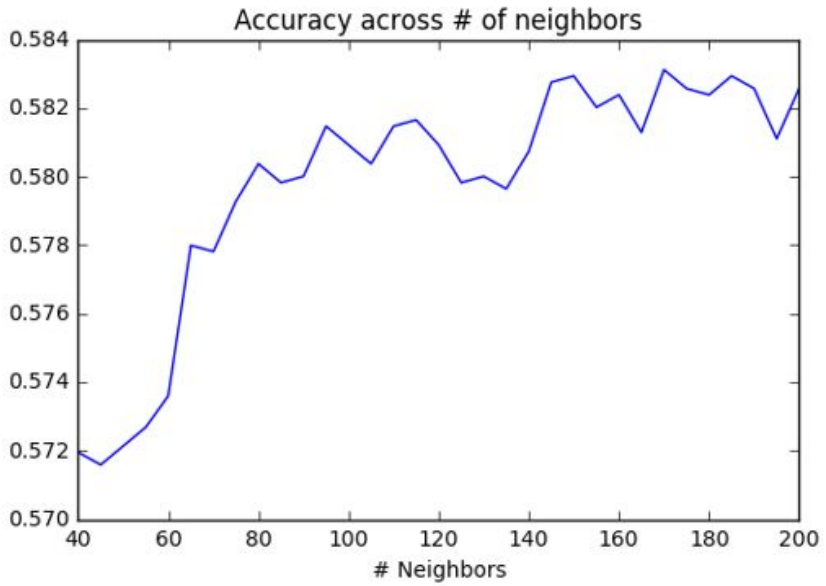
Reverse Engineering: srch_destination_type_ids

```
# create x and y vectors for training
y = destinations['srch_destination_type_id']
x = destinations.select_dtypes(['float64'])
```

```
# set up lda object and fit our training sets
lda = da.LinearDiscriminantAnalysis(n_components=7)
x_fit = lda.fit_transform(x, y)
```

```
# separate into training and test sets
train = int(len(x_fit[:, 0])*0.85)
x_train = x_fit[:train, :]
x_test = x_fit[train:, :]

y_train = y[:train]
y_test = y[train:]
```




```
knn.score(x_test, y_test)
```

```
0.58037348956426216
```

```
Out[56]: Index(['popular_activity_snorkeling', 'popular_activity_windsurfing',
                'popular_activity_ecotourism', 'popular_activity_indoorskiing',
                'popular_cultural_entertainment'],
               dtype='object')
```

Time-Delay Booking Predictions

Hypothesis: There should be a way to predict whether a user will book or not, based on their check-in date, website visit date, stay duration, and potentially even hotel rating.

Data and classifier: date_time, srch_ci, srch_co, orig_destination_distance, prop_starrating fields, all in a KNN classifier.