

AirBnB

DETERMINE TRAVEL DESTINATIONS

By: Florian Philippi 05/17

Predicting travel destinations



AirBnB data sets

~250k users

2010 - 2014

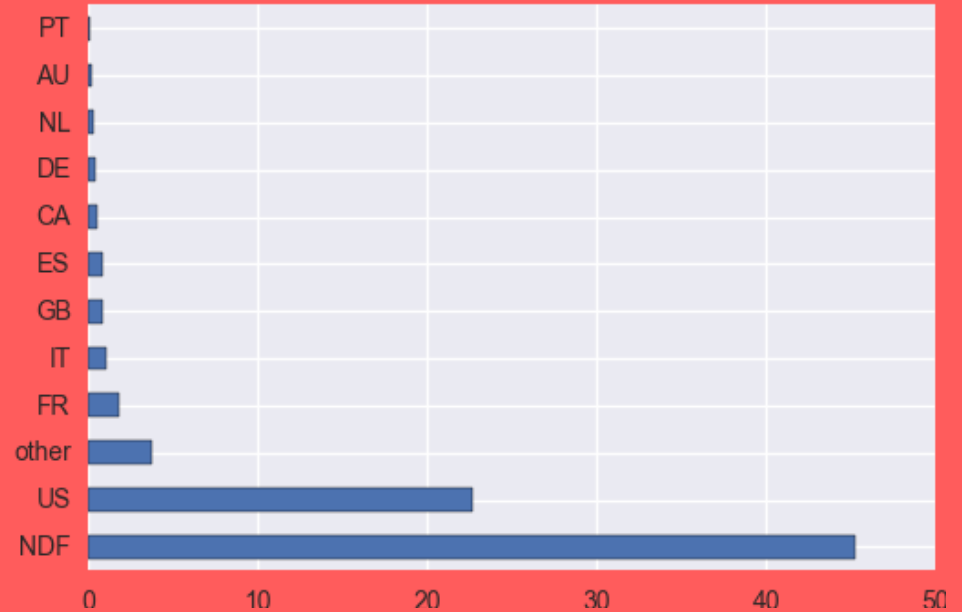
- Main data set with 250k users with demographics, date account created, first booking, etc.
- Web session records for users in 2014 tracking each interaction on the platform
- 12 different country destinations

Data & model

Y = country_destination

12 classes

24 features



NDF = No booking

- Age
- Gender
- Site user came from to sign up
- Sign up method
- Number of user actions on platform
- Total duration on platform

- Day/Month/Year account was created
- Average of seconds elapsed between actions
- Device type

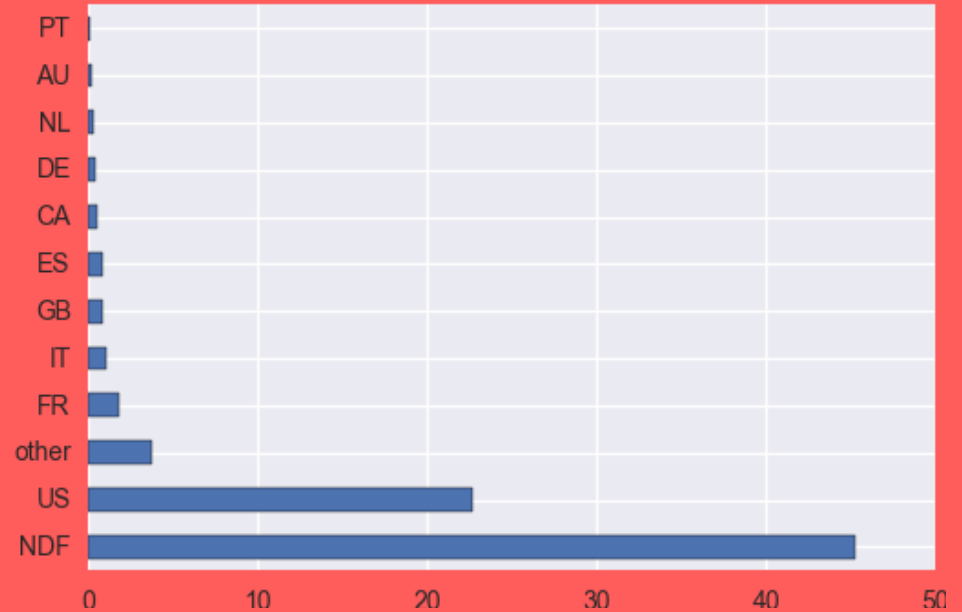
Problem: Class imbalance

sklearn.imblearn

Y = Booking / No Booking

Future works

- Over-/Undersampling
- Try different thresholds



	precision	recall	f1-score	support
AU	0.03	0.01	0.01	162
CA	0.01	0.00	0.00	393
DE	0.01	0.00	0.00	334
ES	0.01	0.00	0.00	677
FR	0.04	0.01	0.02	1563
GB	0.02	0.00	0.01	691
IT	0.02	0.01	0.01	856
NDF	0.68	0.81	0.74	37578
NL	0.00	0.00	0.00	226
PT	0.00	0.00	0.00	61
US	0.45	0.41	0.43	18518
other	0.07	0.02	0.03	3026
avg / total	0.53	0.60	0.56	64085

Model with binary classes

Y = Booking / No booking

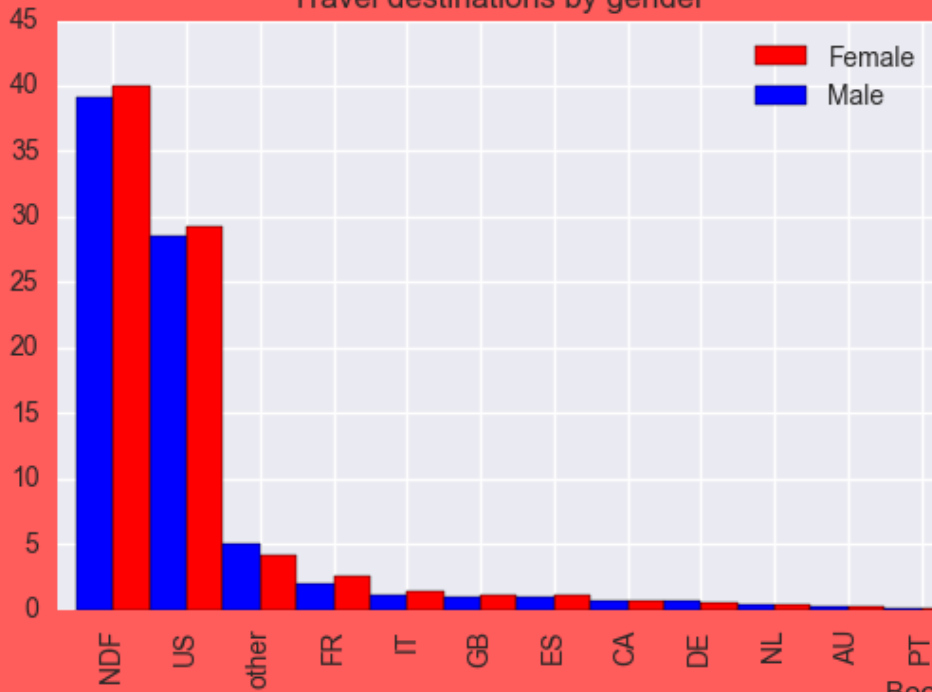
- Logistic Regression
- Random Forest
- Gradient Boost
- XGB

	pre	rec	spe	f1	geo	iba	sup
0	0.73	0.82	0.56	0.77	0.71	0.50	37426
1	0.69	0.56	0.82	0.62	0.71	0.50	26659
avg / total	0.71	0.71	0.67	0.71	0.71	0.50	64085

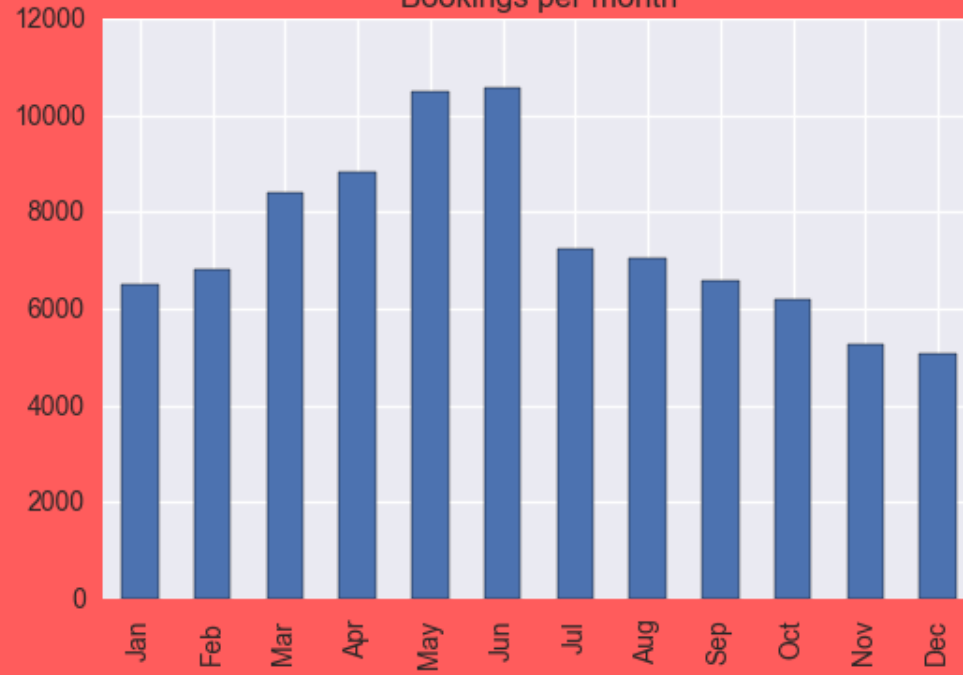
Web App

Appendix

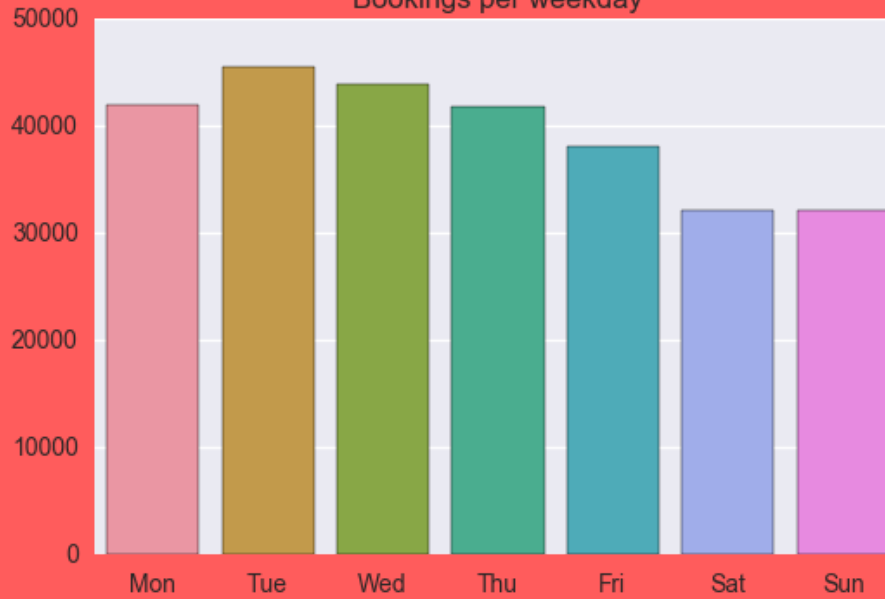
Travel destinations by gender

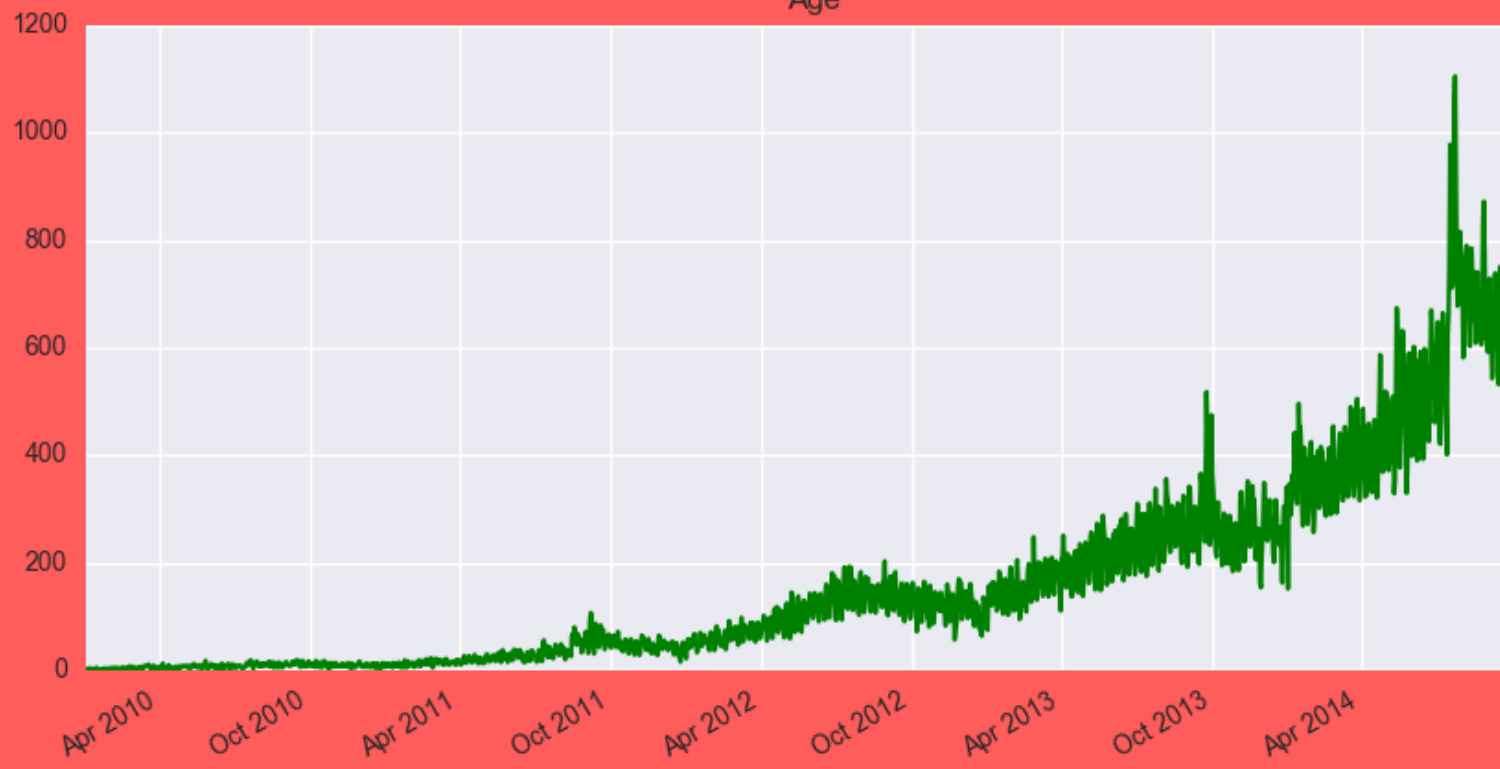
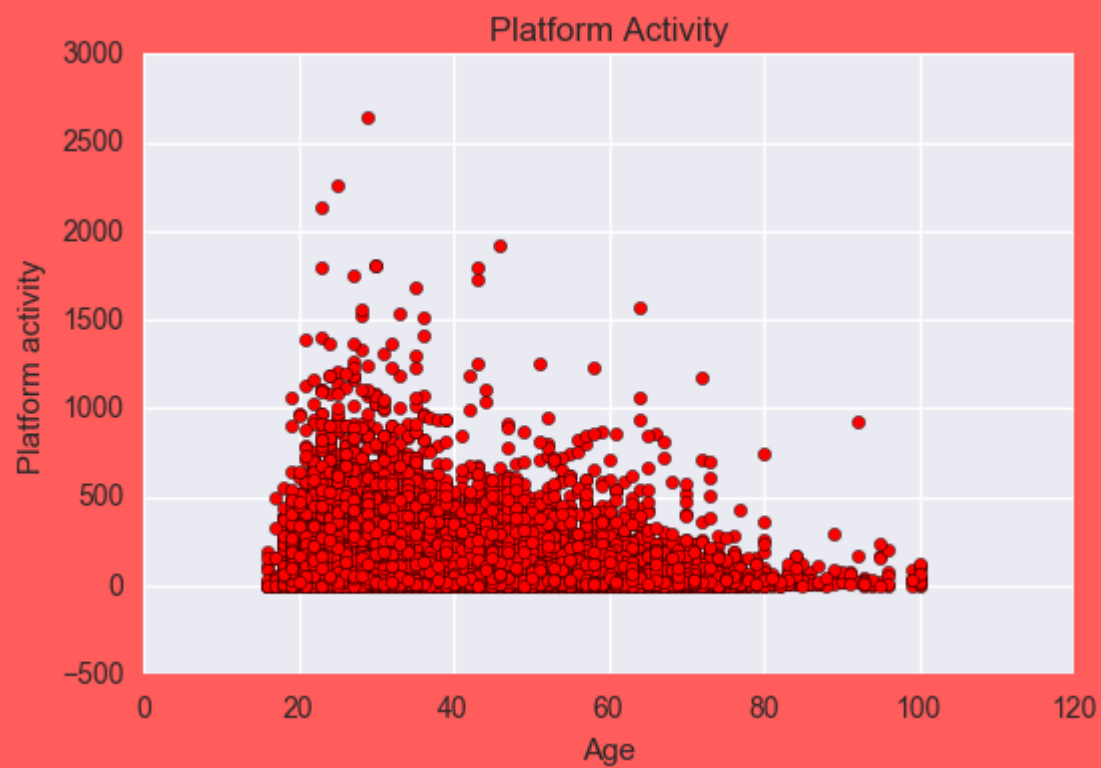


Bookings per month



Bookings per weekday





	precision	recall	f1-score	support
AU	0.03	0.01	0.01	162
CA	0.01	0.00	0.00	393
DE	0.01	0.00	0.00	334
ES	0.01	0.00	0.00	677
FR	0.04	0.01	0.02	1563
GB	0.02	0.00	0.01	691
IT	0.02	0.01	0.01	856
NDF	0.68	0.81	0.74	37578
NL	0.00	0.00	0.00	226
PT	0.00	0.00	0.00	61
US	0.45	0.41	0.43	18518
other	0.07	0.02	0.03	3026
avg / total	0.53	0.60	0.56	64085

```
[ (0.16905445, 'age'),
  (0.10028654, 'signup_flow'),
  (0.067335241, 'action_update_listing'),
  (0.061604585, 'action_booking_requests'),
  (0.055873927, 'signup_method_basic'),
  (0.054441262, 'gender'),
  (0.045845274, 'action_p5'),
  (0.041547276, 'first_device_type_other'),
  (0.038681947, 'account_year'),
  (0.035816617, 'action_search'),
  (0.032951288, 'sum'),
  (0.032951288, 'affiliate_channel_other'),
  (0.031518623, 'timestamp_month'),
  (0.028653296, 'max'),
  (0.028653296, 'first_device_type_Desktop'),
  (0.025787966, 'sumcount'),
  (0.021489972, 'timestamp_year'),
  (0.021489972, 'first_affiliate_tracked_untracked'),
  (0.021489972, 'count'),
  (0.020057306, 'action_view_trips'),
  (0.018624641, 'stddev'),
  (0.017191976, 'affiliate_channel_sem-non-brand'),
  (0.015759312, 'first_affiliate_tracked_omg'),
  (0.012893983, 'affiliate_provider_facebook')]
```

	pre	rec	spe	f1	geo	iba	sup
0	0.73	0.82	0.56	0.77	0.71	0.50	37426
1	0.69	0.56	0.82	0.62	0.71	0.50	26659
avg / total	0.71	0.71	0.67	0.71	0.71	0.50	64085