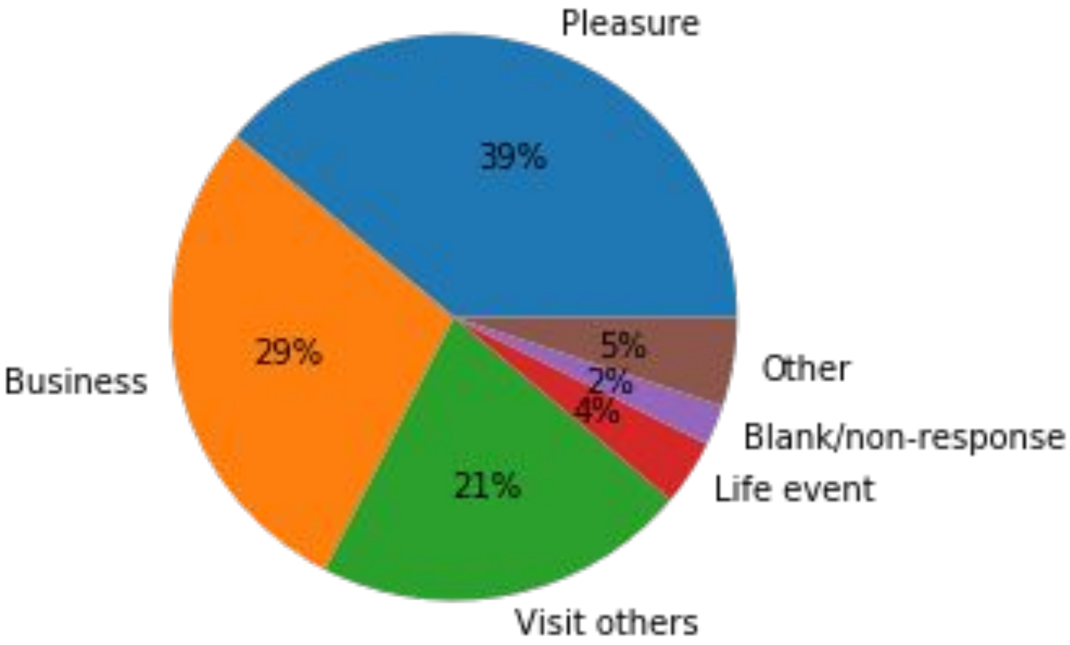
Modeling the Characteristics of Airport Travelers

Melanie Malinas

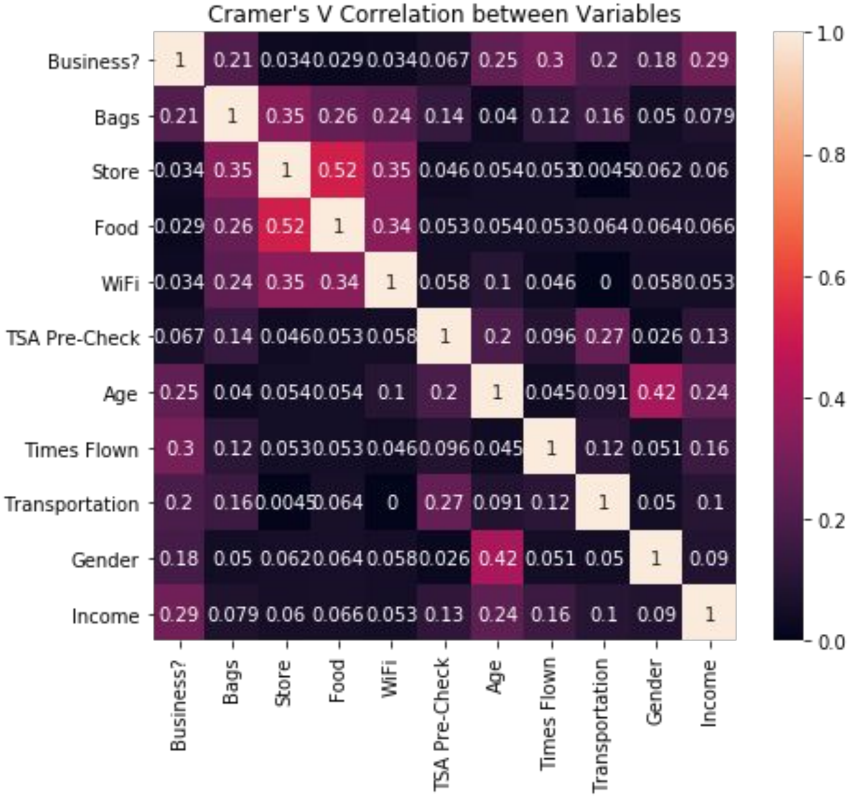
Springboard Data Science Career Track

Capstone 1

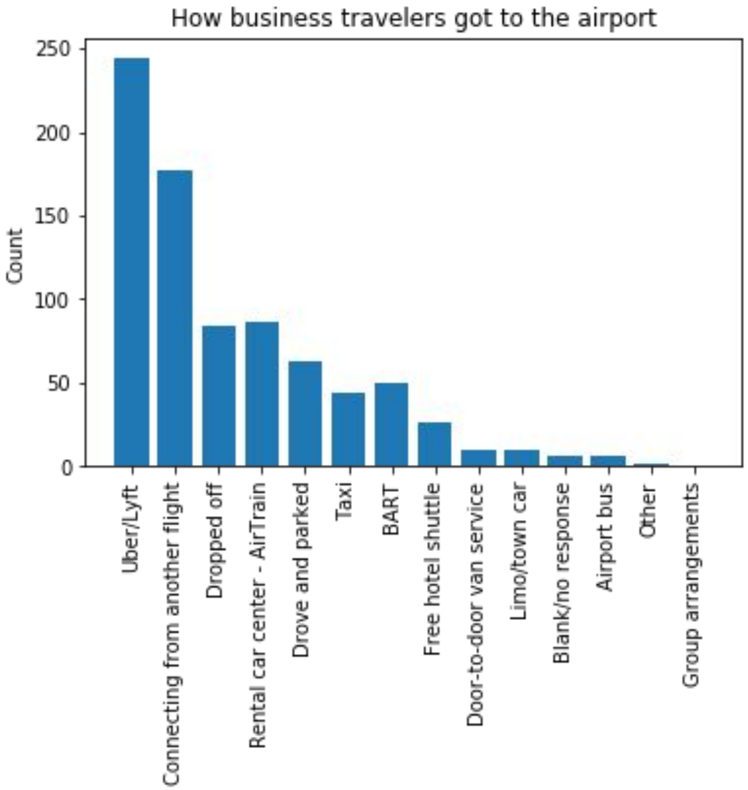
Types of SFO travelers



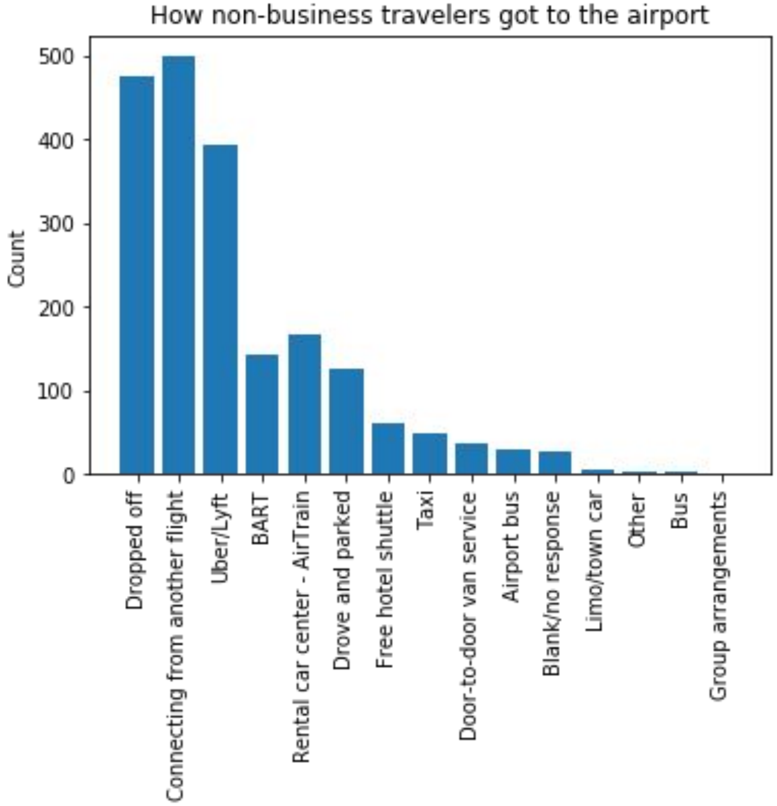
Categorical correlations between variables



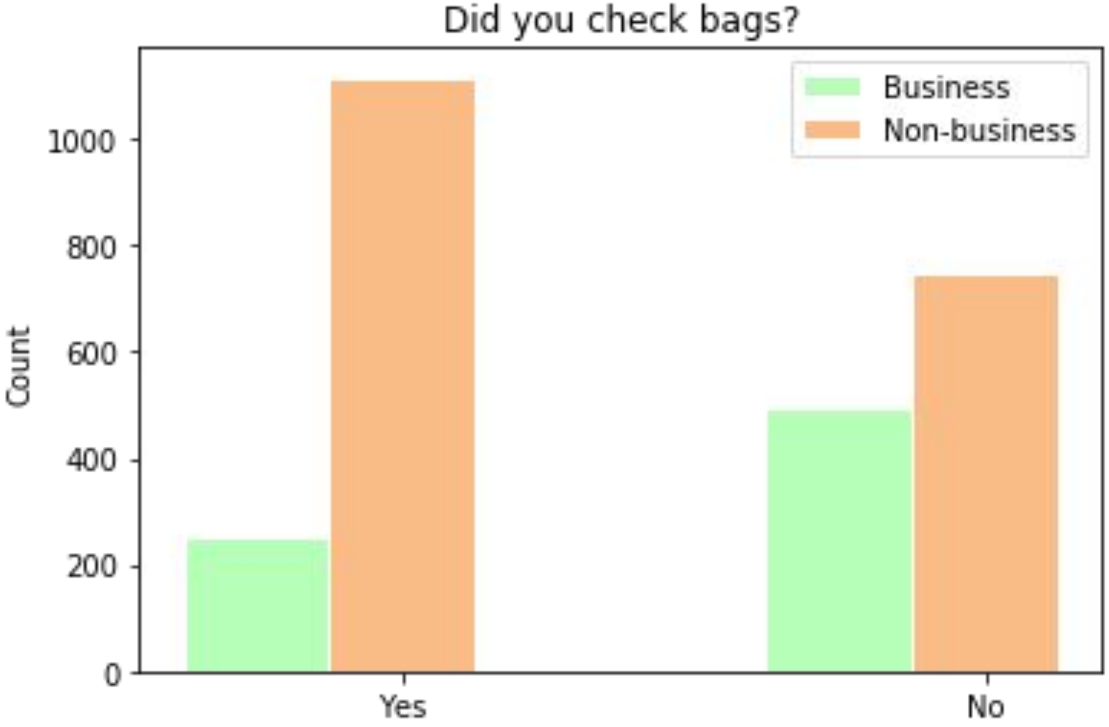
How business travelers got to the airport



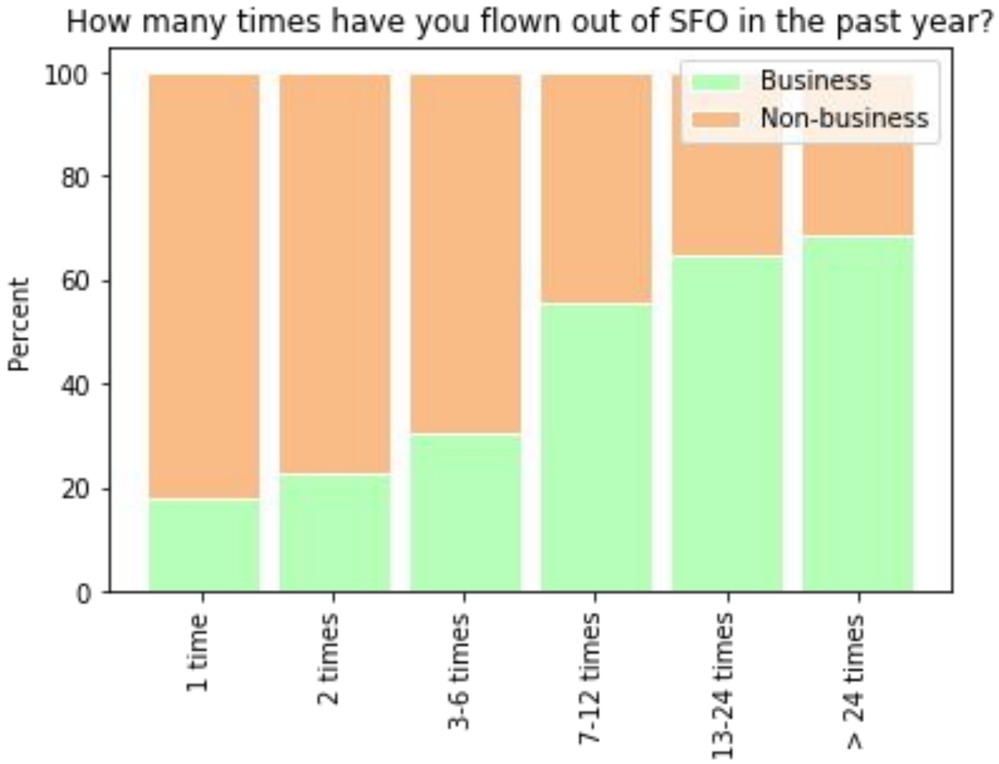
How non-business travelers got to the airport



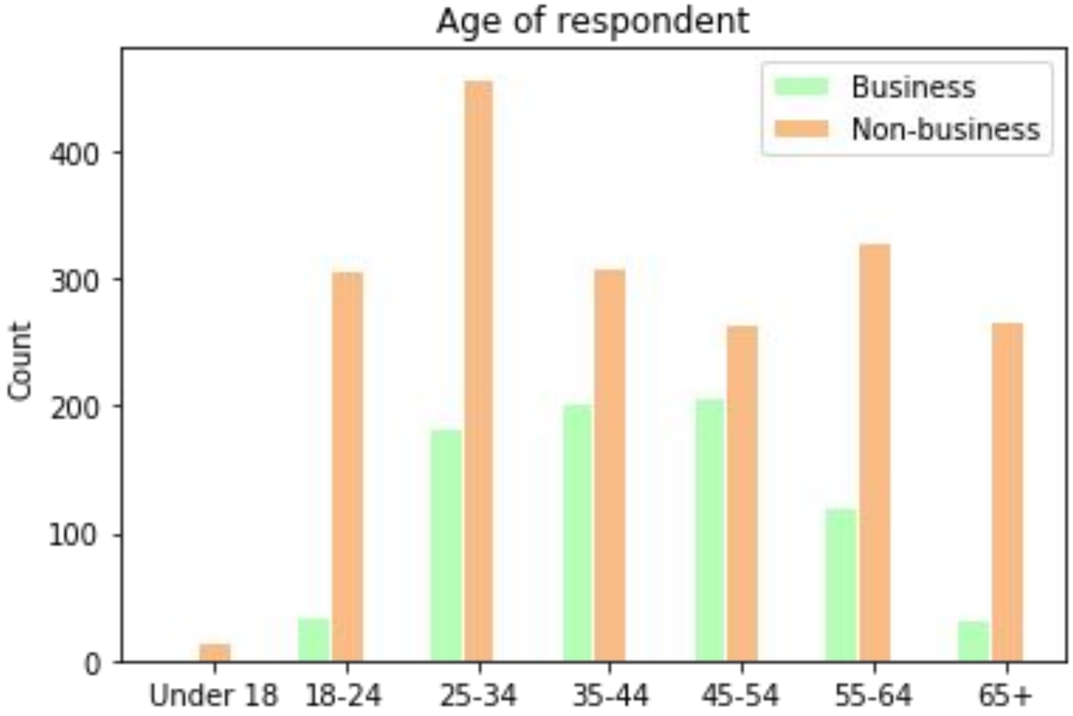
Checking bags



Times flown out of SFO

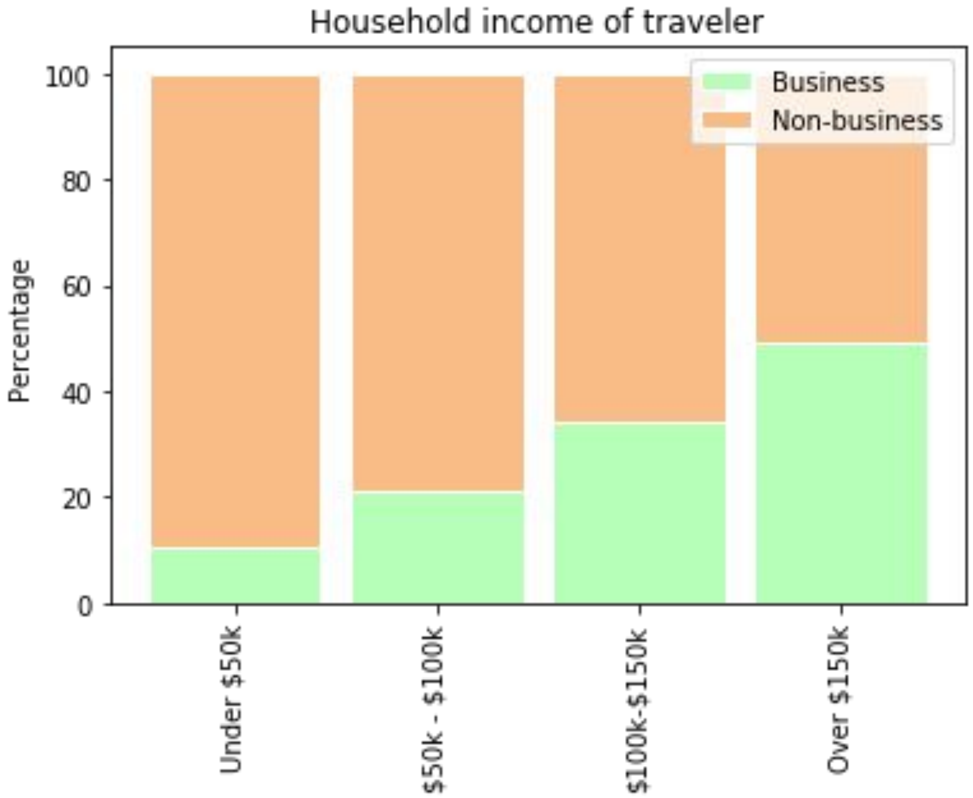


Age of traveler



Gender of traveler

Household income of traveler



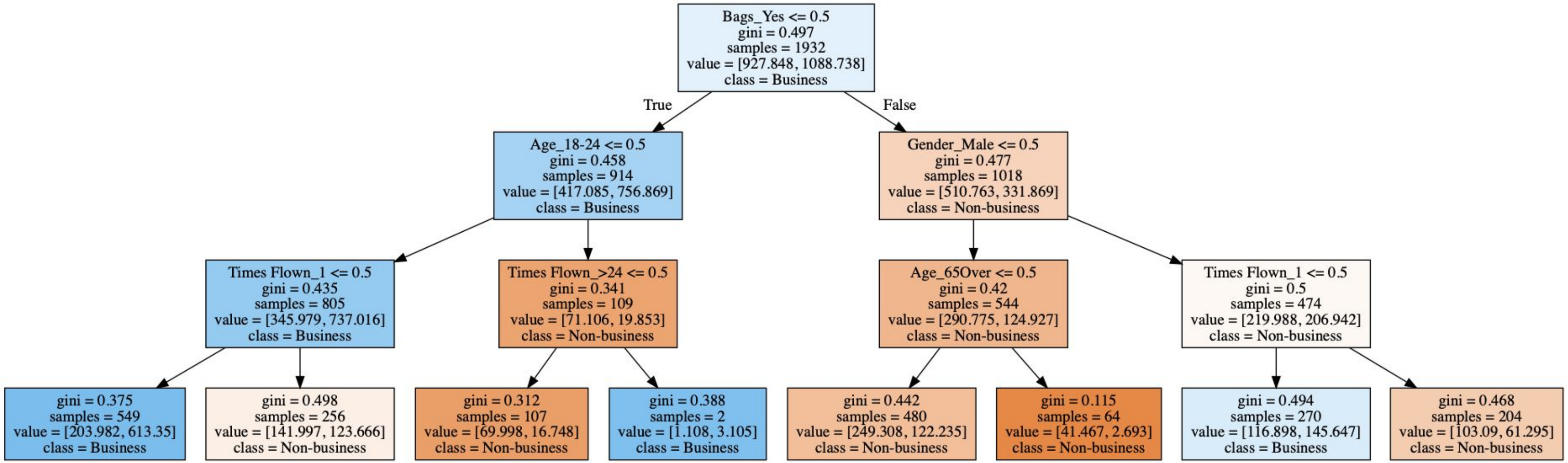
Business case

- Airline marketing a travel rewards credit card
- Giving deal when checking in at the desk of the airport
  - Features to use: Bags, Age, Gender, Times Flown
- Optimizing F1 score
  - Striking a balance between precision and recall

Machine learning methods

Method	Precision	Recall	F1 Score
Logistic regression - no class balance	0.54	0.63	**0.581**
Logistic regression - class balance	0.55	0.61	**0.578**
Random forests - no class balance	0.59	0.52	**0.550**
Random forests - class balance	0.61	0.51	**0.553**
SVM - no class balance	0.66	0.42	**0.511**
SVM - class balance	0.49	0.64	**0.553**
Logistic regression - upsampling	0.58	0.60	**0.591**

Decision tree visualization



Conclusions

- It is possible to predict whether someone is a business traveler based only on their gender, age, number of times flown out of SFO, and whether they checked a bag.
- Surprising that logistic regression was the best model and random forests did not do as well
- Usefulness of upsampling
- Assumptions of this project:
  - We did not have info on whether someone used a business credit card to book flight or whether they used a business-related email
- Work may be useful for distinguishing non-obvious business travelers from non-business travelers