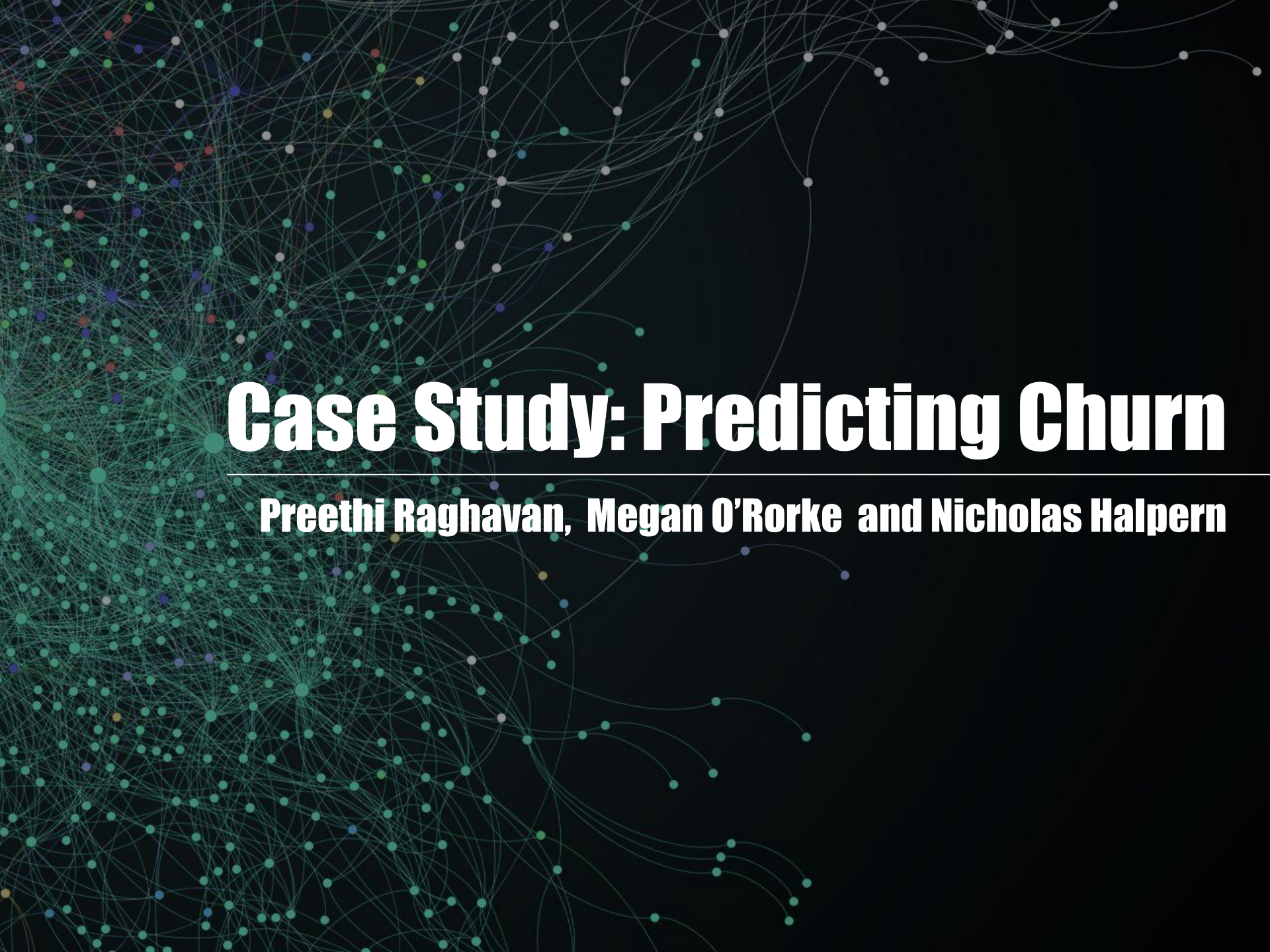


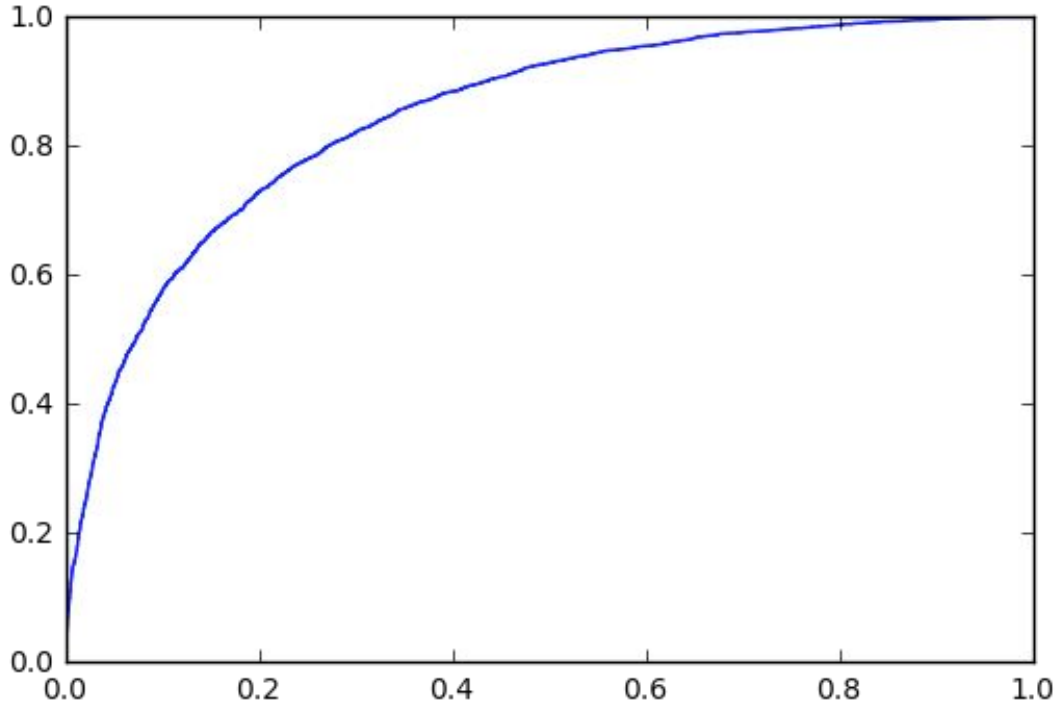
Case Study: Predicting Churn

Preethi Raghavan, Megan O'Rorke and Nicholas Halpern

Goal

Predict rider retention for rideshare company X.

Result



accuracy 0.77

precision 0.718787158145

recall 0.641039236479

Methods

Target:

Rode w/i last 30 days

Best model:

Random Forest on polynomial data of degree 2

Possible alternatives:

Logistic, Adaboost, KNN

Performance Metrics

In absence of a cost info/loss functions we chose accuracy, precision and recall.

Recommendations

Surge % has a bigger impact than # times surged

→ **Try fewer but higher surges**

iPhone churn less

→ **Try improving android app experience**

People who take a luxury car churn less

→ **Try promos for luxury car rides**

People who ride more in 30 days churn less

→ **Try promos, ads for 1st 30 days**

Avg rating BY > OF driver:

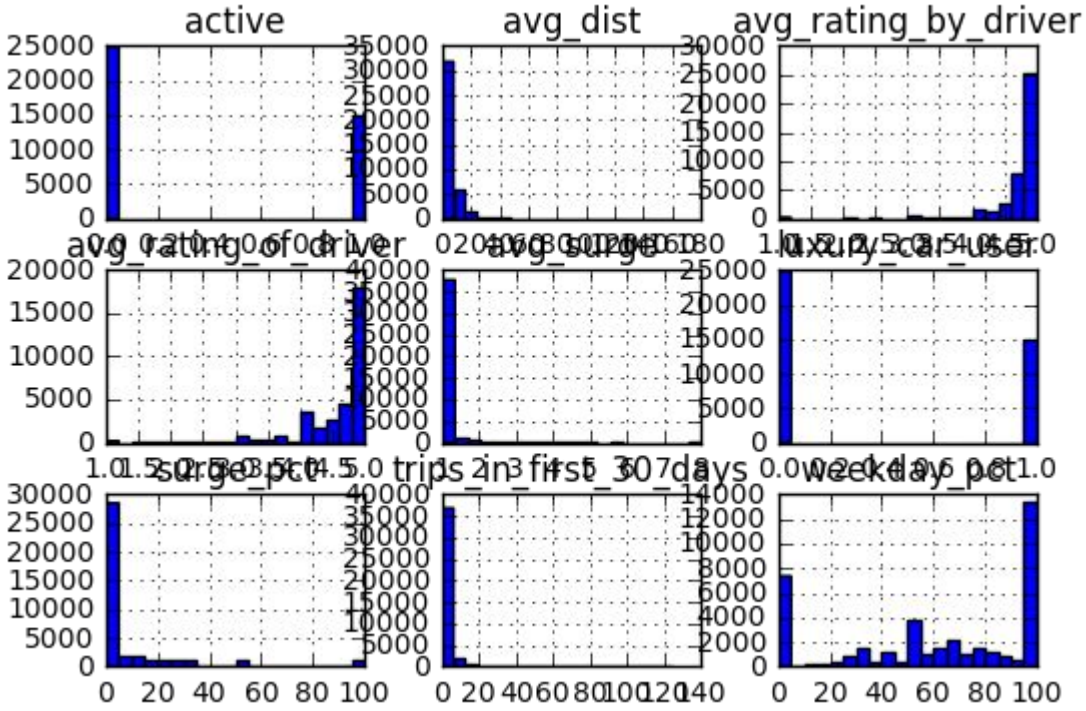
Need more info. Maybe a rider w/ low score is assigned worse drivers?

A) **Match riders scored low w/drivers w/high scores**

B) Pre-drop off hide passenger ratings from drivers

C) **Hide aggregate passenger rating from users, only show most recent ride score**

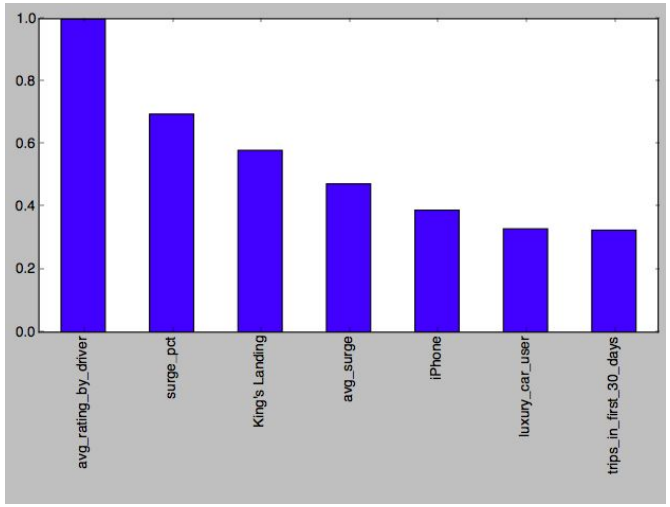
Visualizing the Data



Visualizing the Data

Feature Importance

	Astapor	Kings Landing	Winterfell
avg_dist	-0.041806	-0.048080	-0.042418
price_proxy	-0.002070	0.030611	0.012578
rating_diff	0.049979	-0.048833	0.006851
avg_rating_by_driver	-0.173132	-0.004830	-0.138625
avg_rating_of_driver	-0.223111	0.044003	-0.145477
avg_surge	-0.028523	-1.000000	-0.173950
trips_in_first_30_days	0.083707	0.040812	0.137561
surge_pct	0.000238	0.009726	0.003491
iPhone	1.000000	0.647150	1.000000
luxury_car_user	0.788027	0.634118	0.752056



Visualizing the Data

By City

city	Astapor		King's Landing		Winterfell	
active	0.0	1.0	0.0	1.0	0.0	1.0
avg_dist	5.781319	4.395897	6.765955	5.456761	6.409935	5.194314
avg_rating_by_driver	4.811858	4.764107	4.864733	4.842392	4.747069	4.697280
avg_rating_of_driver	4.638518	4.579752	4.674120	4.711130	4.564250	4.507584
avg_surge	1.087629	1.083137	1.071966	1.071580	1.066477	1.071306
surge_pct	10.293805	9.754758	8.969548	10.562390	7.284810	7.747342
trips_in_first_30_days	1.602656	3.016899	1.649621	2.922942	1.704655	3.776046
luxury_car_user	0.332657	0.565372	0.257671	0.506552	0.280603	0.473419
weekday_pct	60.204421	60.080522	62.820851	63.379210	60.434173	60.247082
one_ride	0.071783	0.000000	0.058067	0.000000	0.081166	0.000000
weekday	0.670283	0.656389	0.681623	0.683747	0.686268	0.669416
weekend	0.329717	0.343611	0.318377	0.316253	0.313732	0.330584
day_of_week	3.346345	3.408835	3.308809	3.304518	3.281845	3.266881
Astapor	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000
King's Landing	0.000000	0.000000	1.000000	1.000000	0.000000	0.000000
Winterfell	0.000000	0.000000	0.000000	0.000000	1.000000	1.000000
Android	0.374835	0.153276	0.382052	0.169372	0.387610	0.173388
iPhone	0.617459	0.840202	0.610030	0.824760	0.602369	0.819584

Future Insights

Data we'd like to analyze:

- # declined rides
- individual trip data per user including each ride's cost
- all rider's data for multiple yrs (currently have 6 months of aggregate data from Jan sign ups)
- a list of changes (with dates) for rider and passenger interfaces
- additional biz info for ex: cost benefit tradeoffs, profit curves

Q&A


Thank You!