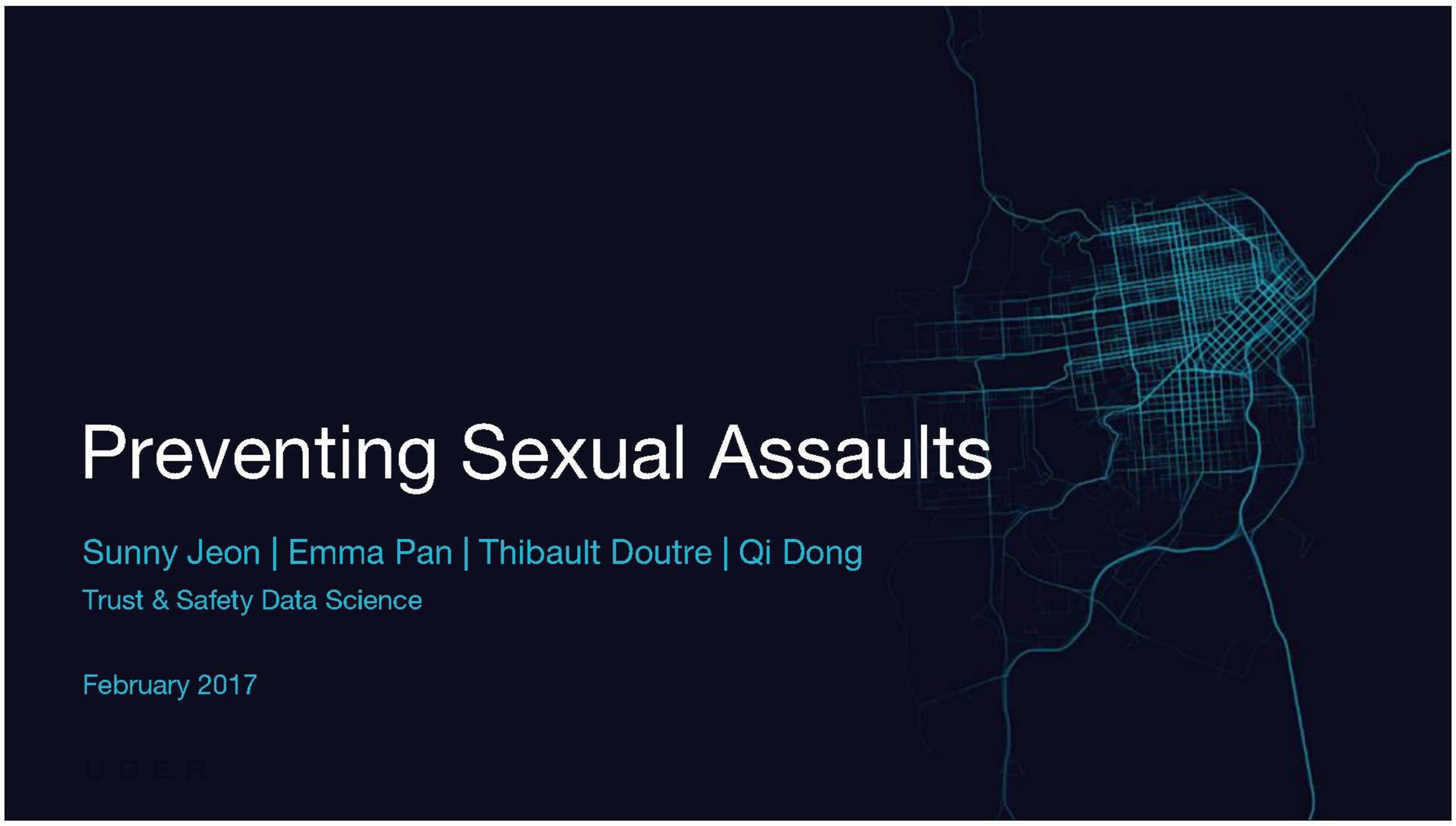


TO CHANGE THE BACKGROUND PHOTO:
– Go to 'View > Master'
– Right-click on the image and select 'Replace image...'
– Right-click on the photo, go to 'Order > Send to Back'

PHOTOS SHOULD BE AT LEAST 1700 x 866px FOR OPTIMAL RESOLUTION

WHEN THE PRESENTATION TITLE EXTENDS TO TWO OR MORE LINES:
– Select the background blur by clicking to the left of the presentation title
– Resize the background blur to ensure sufficient contrast with the background photo.

Contents

Summary of Findings

- Sexual assaults have patterns and precursors. Some correlates:
  - Male offenders and female victims.
  - Bars + late night + weekends.
  - Offenders have previous safety incidents, lower ratings, and higher Bouncer risk score.

- At **trip-level**, baseline models can predict sexual assaults with the following performance (at <0.50% trigger rate):
  - Heuristics: recall 3.61% / trigger rate 0.49% / precision 0.009%
  - Baseline model: recall 8.64% / trigger rate 0.49% / precision 0.019%

- Results demonstrate viability of quick rules to block risky matches while building out full production grade ML model.
  - Phase I: Naive rules + experiments to quickly prototype, set baseline, and iterate.
  - Phase II: Deploy model and evaluate (compared to naive rules / heuristics).
  - Phase III: Intelligent actioning -- blocking at high risk times/locations, down-ranking, triggering safety features (Dolby, SOS, Share).

ATTORNEY-CLIENT PRIVILEGED

1 | Sexual Assault Definitions & Trends

DO NOT CITE OR CIRCULATE

Definitions (JIRA)

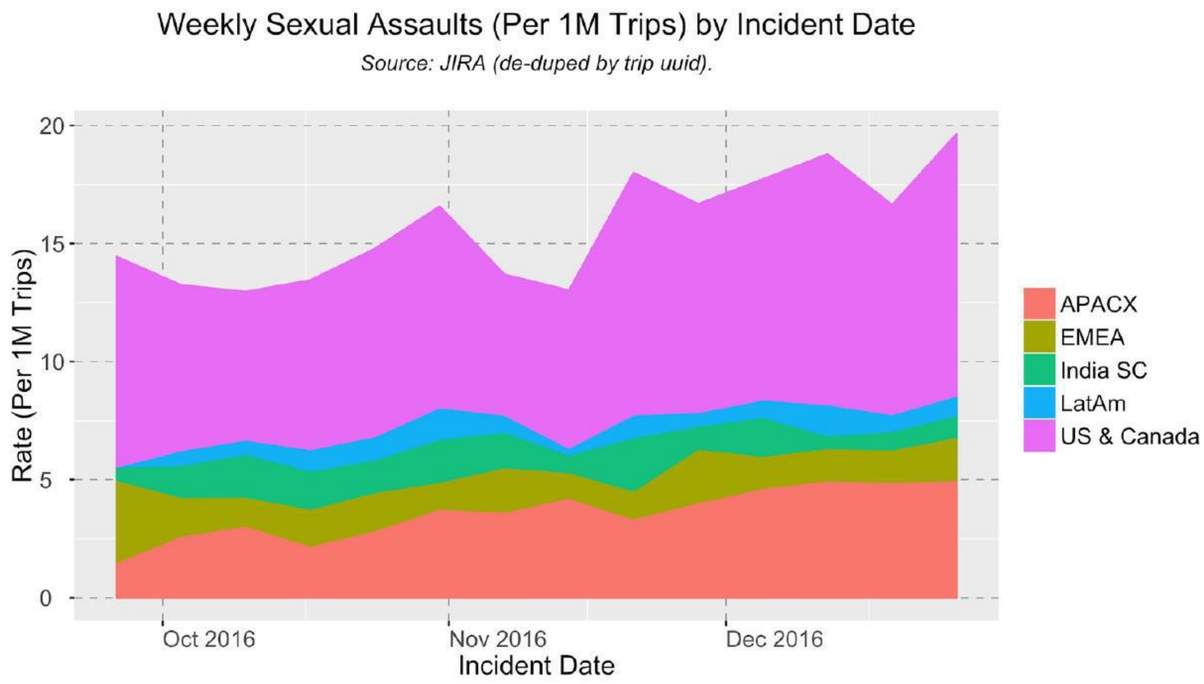
Sexual Misconduct & Assaults (US Cities, Sept 1 - Jan 9, 2017)

Focus of this deck

Category	Incident Type	Freq	Perc
Sexual Assault	**Sexual Misconduct -> Non-Consensual Touching**	**2683**	**65.7%**
Misconduct	Inappropriate Contact -> Physical Stalking	458	11.21%
Misconduct	Sexual Misconduct -> Sexually Inappropriate Remarks or Conversation	399	9.77%
Misconduct	Sexual Misconduct -> Indecent Exposure	232	5.68%
Misconduct	Sexual Misconduct -> Masturbation	97	2.38%
Sexual Assault / Rape	**Sexual Misconduct -> Non-Consensual Intercourse**	**85**	**2.08%**
Misconduct	Sexual Misconduct -> Explicit Gesture	70	1.71%
Misconduct	Inappropriate Contact -> Text Messages/Phone Calls After Trip	51	1.25%
Misconduct	Sexual Misconduct -> Staring or Leering	9	0.22%
		4084	**100%**

Notes: Categories and frequencies based on JIRA data for US cities only. Trip uuid used for de-duping and linking to city, so incidents without trip uuids are not included. All offender types.

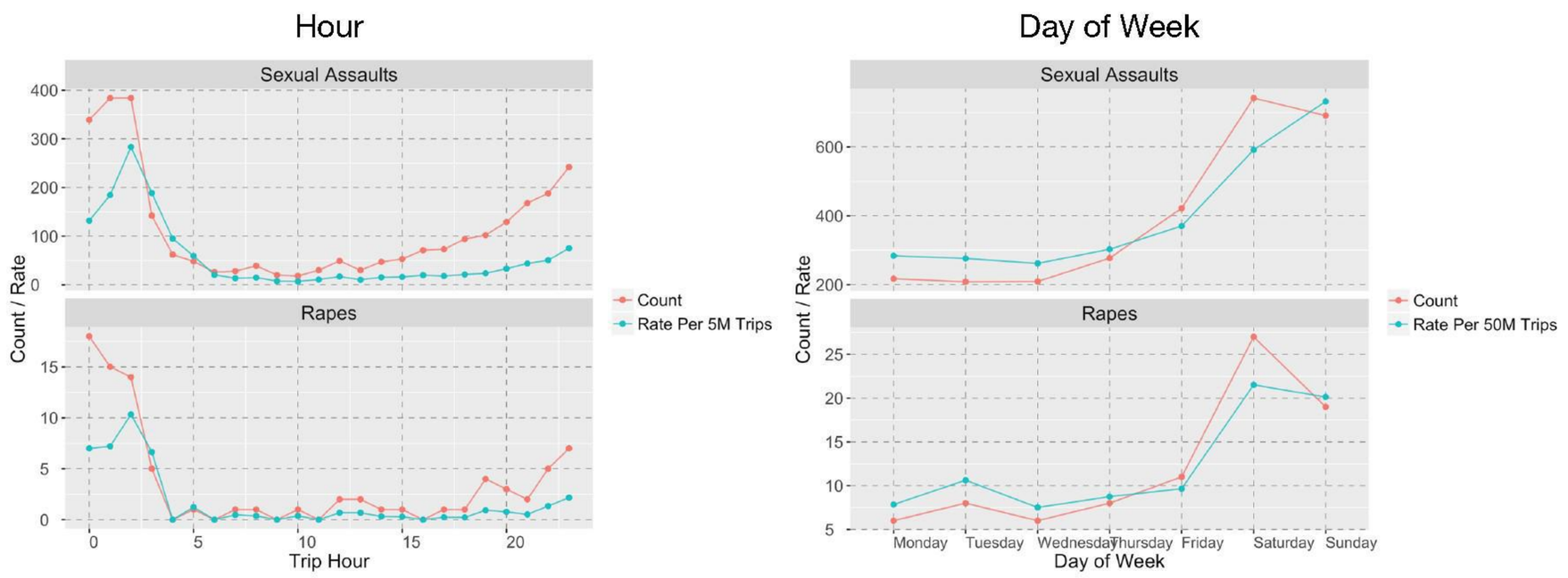
Sexual Assaults Over Time



Notes: Stacked area graph of region-level incident rates. Sexual assaults defined as "non-consensual intercourse" and "non-consensual touching" JIRA categories. Trip uuid used for de-duping and linking to city, so incidents without trip uuids are not included. All offender types.

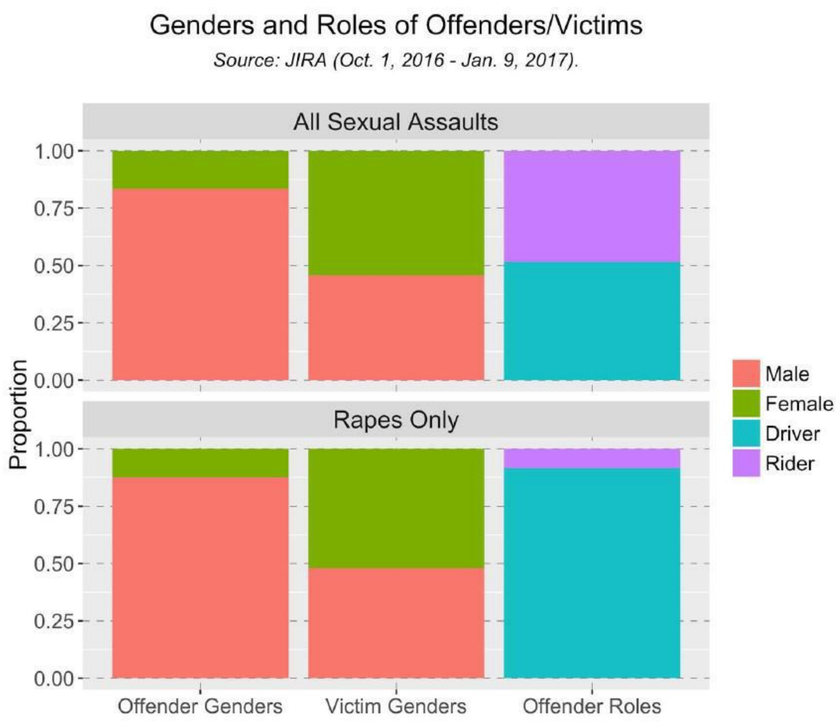
2 | Trip-level Correlates

Sexual Assaults by Hour and Day of Week



Notes: Sexual assaults defined as “non-consensual intercourse” and “non-consensual touching” JIRA categories. Trip uuid used for de-duping and linking to city, so incidents without trip uuids are not included. All offender types. US data from September 1 - January 9, 2017.

Genders and Roles of Offenders & Victims



Incidents between Different Genders

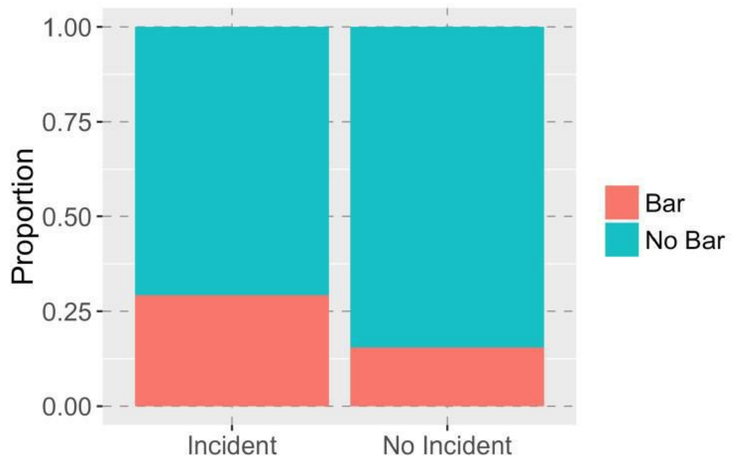
	Trips with Incident	Trips without Incident
All Sexual Assaults (n = 2237)		
Proportion of Trips with Driver-Rider from **Different Genders**	0.56	0.40
	Chi-squared test statistic = 136.39 ($p < 0.001$)	
Rapes Only (n = 65)		
Proportion of Trips with Driver-Rider from **Different Genders**	0.59	0.40
	Chi-squared test statistic = 5.9 ($p = 0.015$)	

Notes: Gender of users classified using R package gender, which draws from US census data. Offenders and victims identified using JIRA data (Incident Offender field). Trip uuid used for de-duping and linking to city, so incidents without trip uuids are not included. US incidents and trips only.

Bar Pick-ups and Sexual Assaults

Proportion of Bar Pick-ups

Source: JIRA (All Offenders, Oct. 1, 2016 - Jan. 9, 2017).



Notes: A pickup is defined as a "bar pick-up" if there is a bar within 50 meters of the request lat/lng (as indicated by Foursquare data).

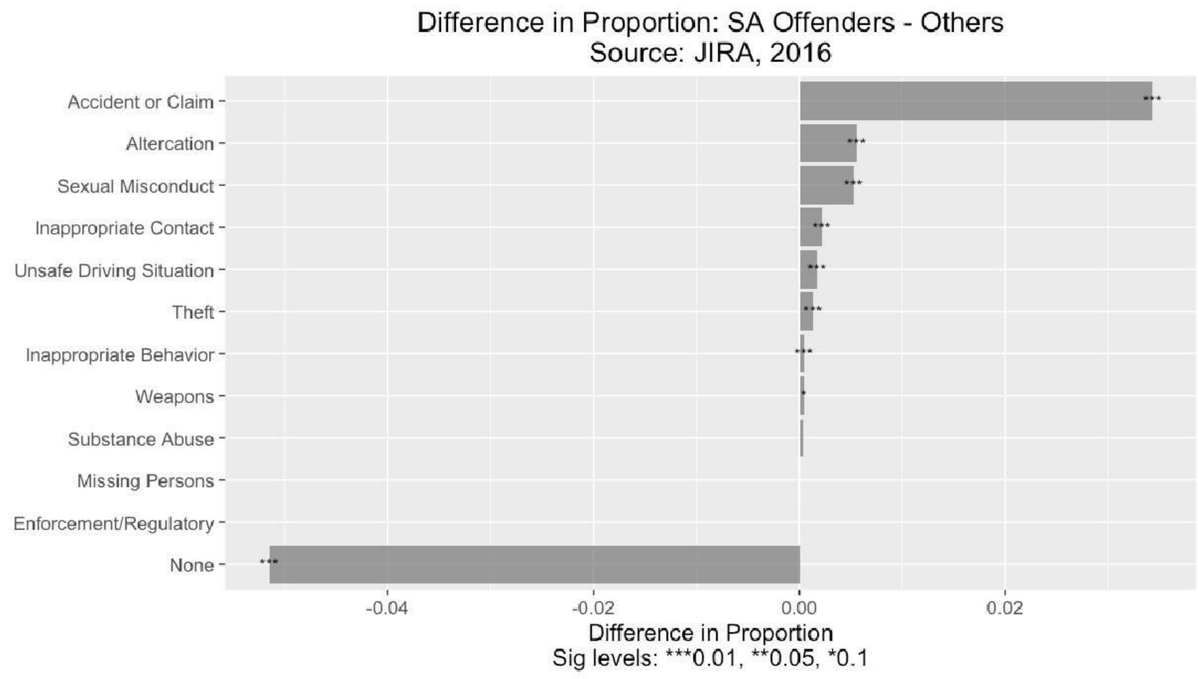
Bars within 50 Meters of Request Lat/Lng

	Trips with Sexual Assault	Trips without Sexual Assault
Mean Number of Bars within 50 Meters	0.55	0.28
Difference in Means	t-statistic = 11.45 ($p < 0.001$)	
Proportion of Trips with >0 Bars	0.29	0.16
Difference in Proportions	chi-squared test statistic = 322.59 ($p < 0.001$)	

3 | Driver-level Correlates

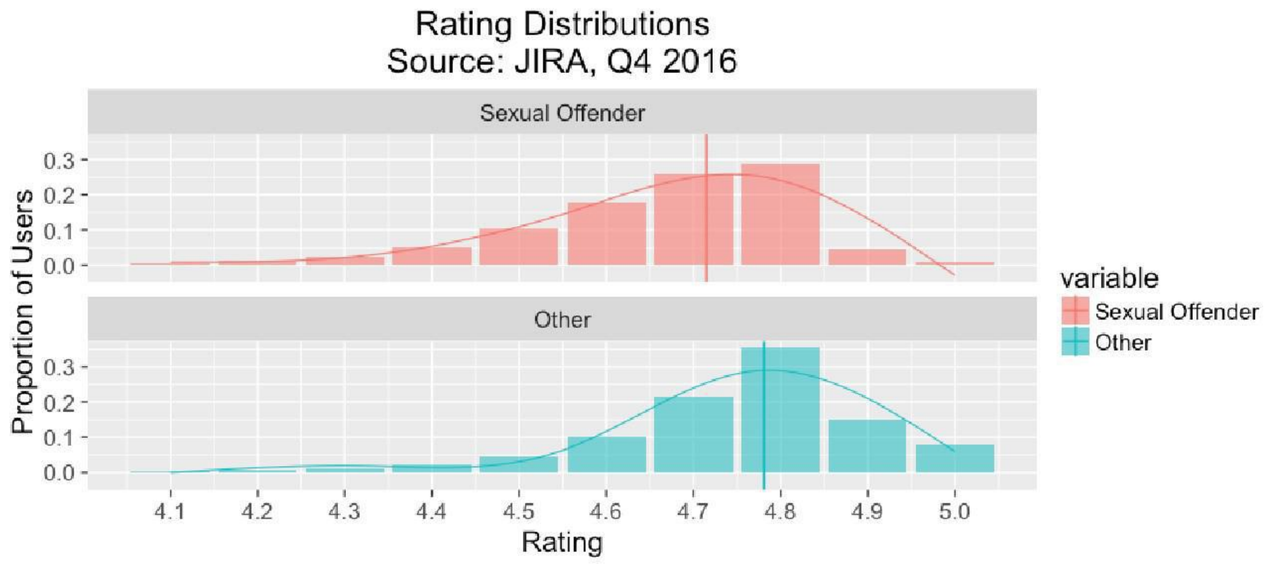
- Incident History
- Ratings: Overall and by Opposite Gender
- Age
- Language

Driver Offenders Have Previous Incidents



Notes: For each issue type, the bar represents the difference between **1**- Proportion of drivers who committed an incident associated with this issue type among drivers who have committed a sexual assault 2- Proportion of drivers who committed an incident associated with this issue type among drivers who didn't commit a sexual assault. For sexual assault offenders, only data prior to the first sexual assault in the US, in 2016 are considered. Proportion tests are computed for each issue type.

Driver Offenders Receive Lower Ratings



	Average for Sexual Offenders	Average for All Others	P-value (T-test)	Relative Difference
Ratings	4.715	4.781	<0.001	**-1.34%**

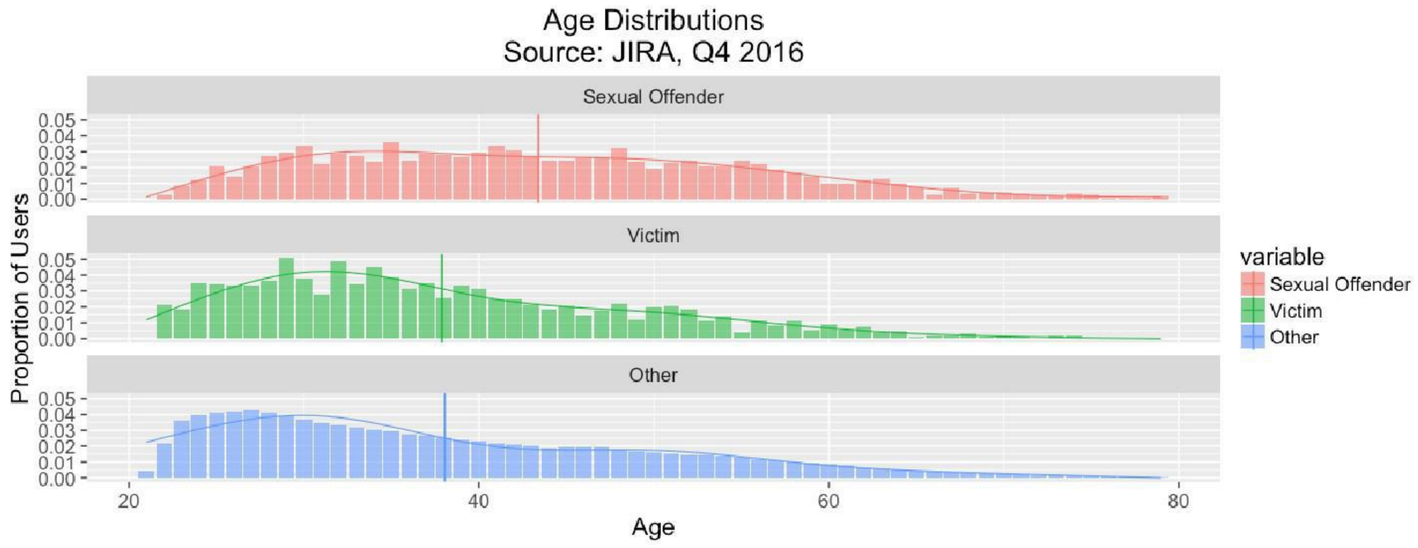
Notes: Others refer to all US drivers who got at least one rating and haven't committed a sexual assault in Q4, 2016. Only drivers with at least 5 rated trips during this time period are considered.

Driver Ratings From Opposite Gender

	Average Rating from Different Gender	Average Rating from Same Gender	P-value (T-test)	Relative Difference
Offenders	4.715	4.777	<0.001	**-1.30%**
Non Offenders	4.781	4.816	<0.001	**-0.73%**

Notes: Others refer to a random sample of all US drivers who got at least one rating and haven't committed a sexual assault in Q4, 2016. Only drivers with at least 5 rated trips are considered.

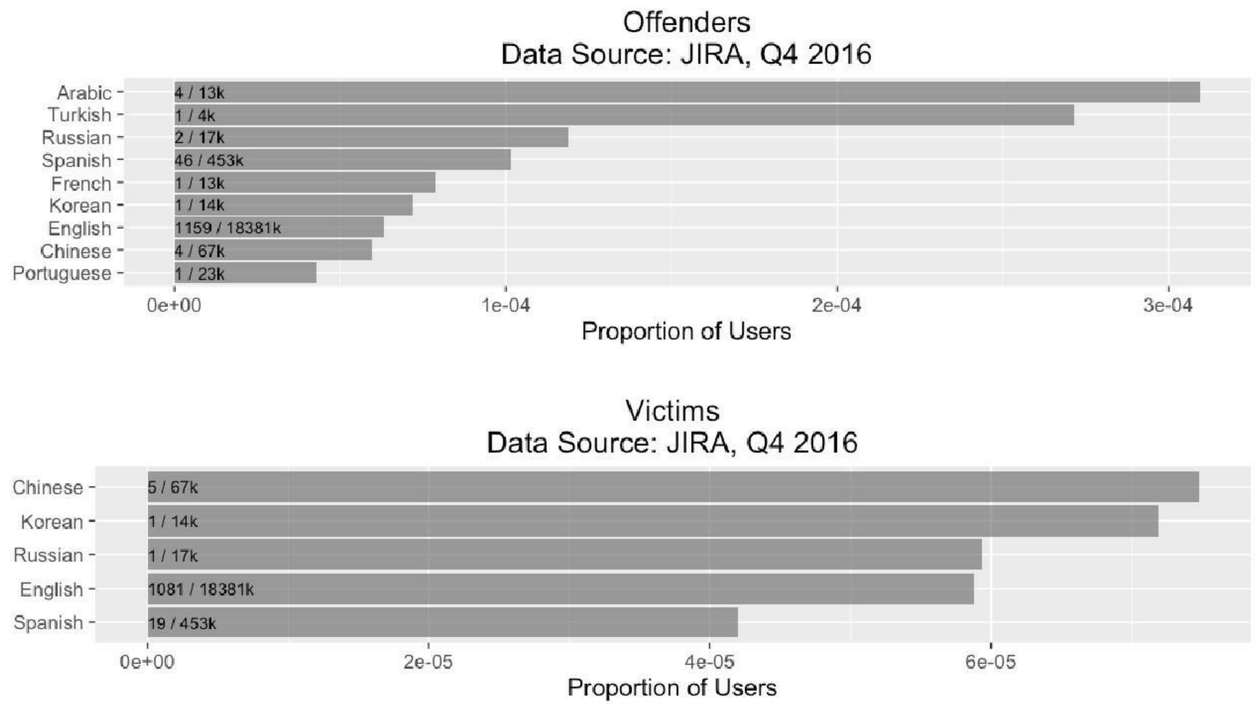
Driver Offenders are Older than Victims



	Average Age	Average Age of All Others	P-value (T-test)	Relative Difference
Offenders	43.4	38.06	<0.001	**+14%**
Victims	37.89	38.06	0.6252	-0.45%

Notes: Others refer to all US drivers who got at least one rating and haven't committed a sexual assault in Q4, 2016.

Driver Language (Note: tiny sample size here)



Notes: For each language, the bar represents the proportion of US drivers with this phone language who have committed a sexual assault in Q4 2016.

Bouncer Scores

Average Bouncer Scores

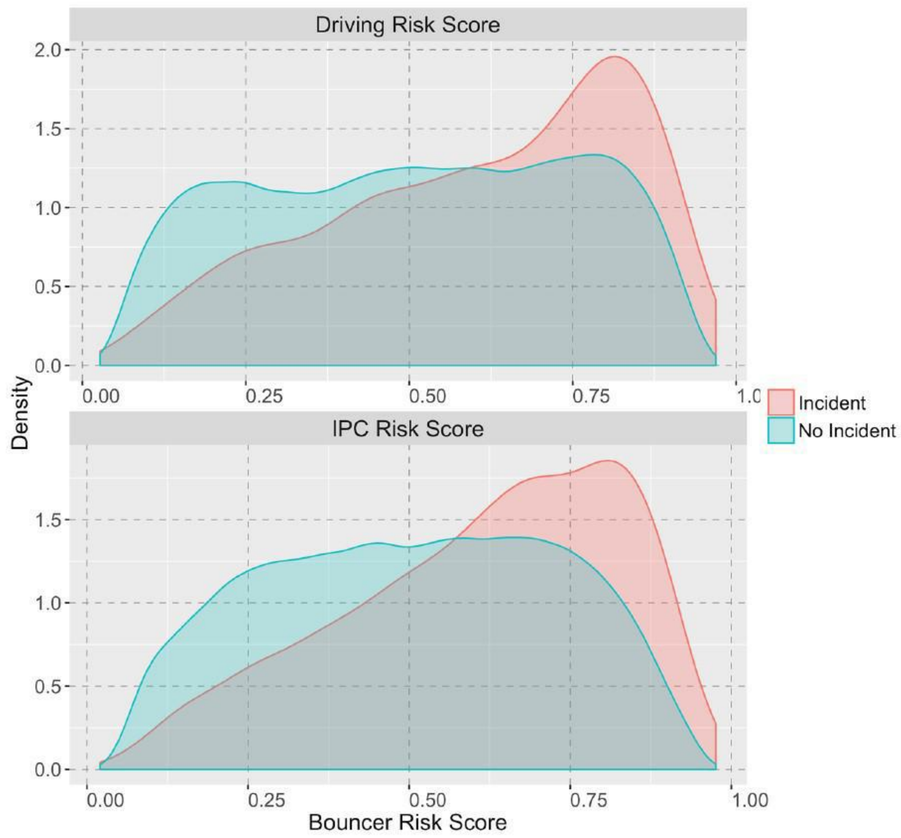
	Sexual Assault Offender	Non-Offender	Difference in Means (p-value from t-test)
Driving Risk Score	0.60	0.50	0.10 ($p < 0.001$)
IPC Risk Score	0.61	0.50	0.11 ($p < 0.001$)

"Creepy Driver" Feedback from Riders

	Trips with Sexual Assault	Trips without Sexual Assault
Proportion of Drivers with "Creepy Driver" Feedback from Riders	0.017	0.007
	Chi-squared test statistic = 14.084 ($p < 0.001$)	

Bouncer Scores for Drivers With/Without Sexual Assaults

Source: JIRA (Driver Offenders, Non-Consensual Touching + Intercourse, Oct. 1, 2016 - Jan. 9, 2017).

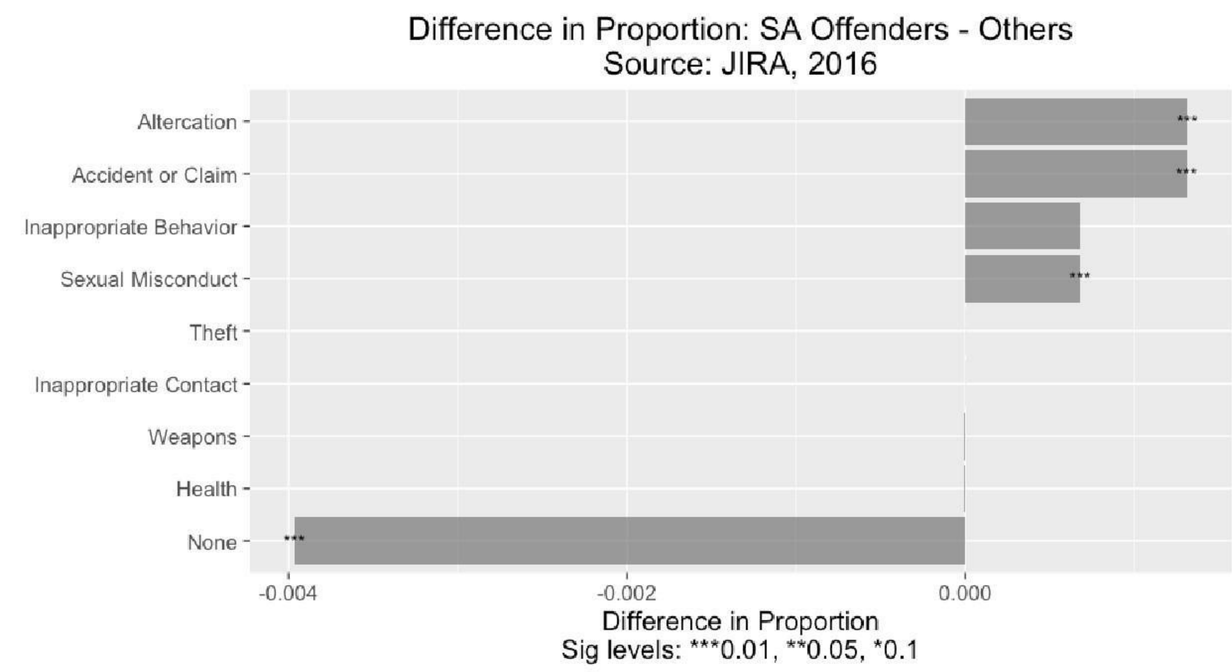


* All statistics represent pre-incident statistics (e.g., Bouncer score *before* sexual assault, number of creepy driver comments *before* sexual assault).

4 | Rider-level Correlates

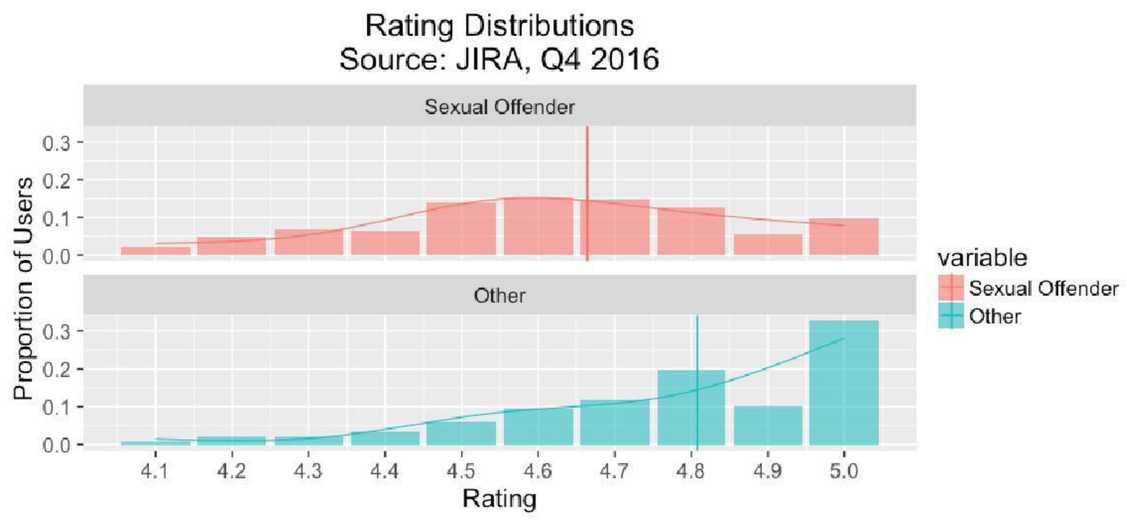
- Incident History
- Ratings: Overall and by Opposite Gender
- Language

Rider Offenders Have Previous Incidents



Notes: For each issue type, the bar represents the difference between 1- Proportion of riders who committed an incident associated with this issue type among riders who have committed a sexual assault 2- Proportion of riders who committed an incident associated with this issue type among riders who didn't commit a sexual assault. For sexual assault offenders, only data prior to the first sexual assault in the US, in 2016 are considered. Proportion tests are computed for each issue type.

Rider Offenders Receive Lower Ratings



	Average for Sexual Offenders	Average for All Others	P-value (T-test)	Relative Difference
Ratings	4.664	4.808	<0.001	**-3.0%**

Notes: Others refer to all US riders who got at least one rating and haven't committed a sexual assault in Q4, 2016. Only riders with at least 5 rated trips during this time period are considered.

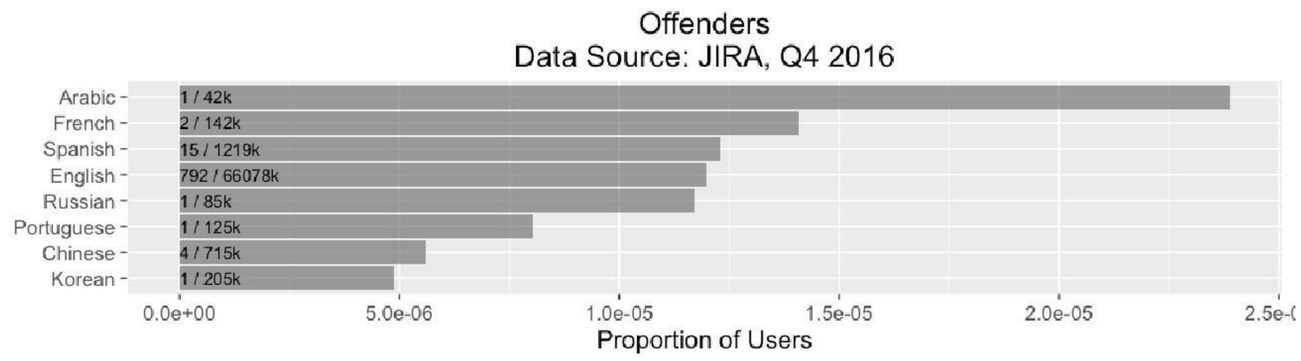
Rider Ratings From Opposite Gender

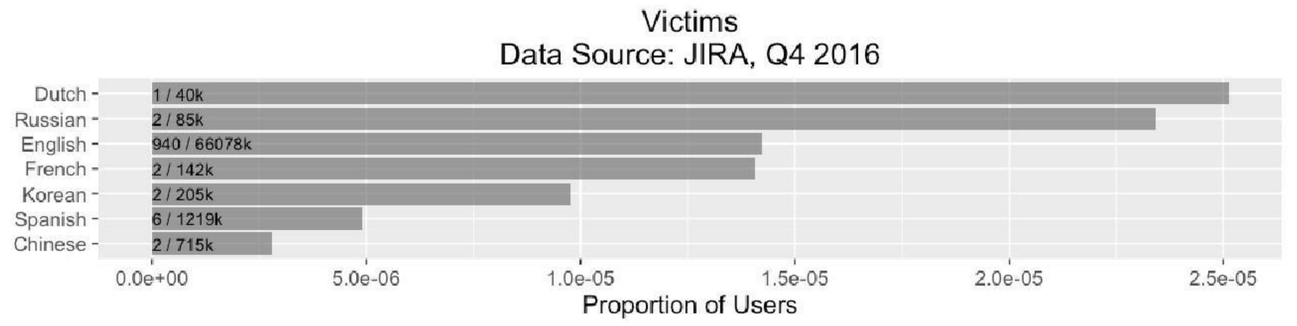
	Average Rating from Different Gender	Average Rating from Same Gender	P-value (T-test)	Relative Difference
Offenders	4.604	4.696	<0.001	**-2.0%**
Non Offenders	4.805	4.810	0.078	-0.12%

Notes: Others refer to a random sample of all US riders who got at least one rating and haven't committed a sexual assault in Q4, 2016. Only riders with at least 5 rated trips are considered.

DO NOT CITE OR CIRCULATE

Rider Language (Note: tiny sample size here)





Notes: For each language, the bar represents the proportion of US riders with this phone language who have committed a sexual assault in Q4 2016.

5 | Baseline Model

Baseline Model

- Machine learning model for predicting driver-rider matches likely to lead to a safety incident.
- Trained to predict L3/L4 sexual assaults using US data (Oct 1 - Jan 9, 2017).
- 43 predictors used in baseline model:
  - Age ← *R&D Purposes Only*
  - Bouncer Scores
  - "Creepy Driver" Feedback
  - Gender ← *R&D Purposes Only*
  - Ratings
  - Safety Incident History
  - Spatial Features (e.g., number of bars near pick-up)
  - Trip History (e.g., trip count, tenure, percent of trips at night/weekends)

Thres, recall,precision,trigger,efficiency
row1: 0.192 0.081967213 6.086095e-05 4.915672e-03 16.674671
row2: 0.060 0.017883756 3.084191e-04 2.116406e-04 84.500595
row3: 0.192 0.09230769 6.639377e-06 4.915672e-03 18.778245
row4: 0.090 0.07692308 6.426746e-05 4.231924e-04 181.768571

Baseline Model - Performance

Prediction Rule	Precision	Recall	Trigger Rate	Efficiency (Recall / Trigger Rate)
All Sexual Assaults*				
Optimize recall with trigger rate < 0.5%	0.019%	8.64% (58/671)	0.49%	**17.6x**
Optimize Efficiency Ratio	0.1%	1.49%	0.01%	**112.6x**
Non-Consensual Intercourse Only*				
Optimize recall with trigger rate < 0.5%	8.83E-04%	12.3% (8/65)	0.49%	**25.0x**
Optimize Efficiency Ratio	0.005%	1.54%	0.01%	**131.8x**

Notes:

- Recall is based on balanced data; Trigger rate and precision adjusted according to the incident rate in raw unbalanced data
- Information unavailable at dispatching is excluded(e.g. trip status, trip duration/distance and whether pool is matched);
- Modeling approach is 1) using SMOTE to rebalance training data and 2) gradient boosting modeling (200 trees & interaction depth 4) with 5-repeated 5-fold cross validation. AUC on test data is 0.8. Decision tree and random forrest were tested but they gave worse performance than GBM did.

If removing gender and age, SA's recall is reduced to 7.1% and efficiency to 14.3;

Baseline Model - Variable Importance

All Sexual Assaults

	Variable Importance
hour	100.00
driver_ipc_count	34.48
driver_genderfemale	29.98
num_bars_50m	16.28
driver_safety_incident_count	15.54
num_bars_100m	12.88
eta	12.33
driver_age	11.79
client_weeks_active	9.05
driver_perc_night_trips	8.77
driver_l3l4_count	7.55
ipc_risk	5.47
client_gendermale	5.36
driver_num_ratings	5.27
driver_trip_count	5.21
client_perc_canceled	5.16
num_bars_500m	4.61
driver_weeks_active	4.10
driver_sum_rating	3.95
driver_avg_rating	3.87

	Variable Importance
num_restaurants_km2	3.80
client_perc_night_trips	3.75
gender_diff	3.61
sub_regionTri-State	3.38
client_safety_incident_count	3.05
client_num_ratings	2.97
client_avg_rating	2.84
num_colleges	2.84
num_bars_1km	2.72
num_restaurants	2.71
client_genderfemale	2.35
regionUS East	2.02

Notes: Variable importance measures a variable's ability to classify trips with sexual assault vs trips without sexual assault. The higher the value the stronger the ability.

Comparison to Heuristics / Naive Model

Naive Prediction Rule: Predict trip will end in sexual assault if:

- Trip takes place on Sat/Sun,
- Trip time between 12am-4am,
- Driver-rider different gender,
- Number of bars within 50m of request, point ≥ national average,
- Bouncer score is above threshold p^*.

Performance Metrics

Prediction Rule	Precision	Recall	Trigger Rate	Efficiency (Recall / Trigger Rate)
All Sexual Assaults				
Optimize recall with trigger rate < 0.5% (Bouncer p^* = 0.07)	0.009%	3.61% (82/2303)	0.49%	**7.37x**
Optimize Efficiency Ratio (Bouncer p^* = 0.9)	0.03%	0.26% (6/2303)	0.01%	**26.00x**
Non-Consensual Intercourse Only				
Optimize recall with trigger rate < 0.5% (Bouncer p^* = 0.29)	7.77 E-04%	10.77% (7/65)	0.49%	**21.98x**
Optimize Efficiency Ratio (Bouncer p^* = 0.96)	0.002%	1.54% (1/65)	0.03%	**51.33x**

Notes: Performance predicting JIRA sexual assaults occurring in US cities between Oct 1 - Jan 9, 2017 (all offenders). Trip uuid used for de-duping and linking to city, so incidents without trip uuids are not included. Trigger rate calculated on random sample of 125,000 US trips taking place during the same time period. Precision estimated by taking number of true positives and dividing by the trigger rate multiplied by the total number of trips occurring during sample time period. In Q4, there was an average of 12M P2P trips per week (QUERY). So a trigger rate of 0.01% = 1.2k trips per week, 0.50% = 60k trips per week.

6 | Next Steps

Next Steps

- New features to investigate:
  - NLP on rider feedback + complaints from opposite gender.
  - History of contact after trip.
  - Criminal violations (background check data).
  - Anomalies in cancellations (e.g., cancel till female rider matched).
  - Driver/rider profiles (e.g., percent of trips late night Fri/Sat => frequents bars).
  - Spatial features (e.g., bars, colleges, offices, demography, income, crime rate).

- Build, deploy, and test high-performing, production grade version of model.

- Intelligent interventions / rules / actioning.
  - Blocking during high risk times and locations,
  - Down-ranking,
  - Monitor and trigger safety features, like Dolby, SOS, Share Your Trip.