

SF Crime and Smoothies

Brayden Cleary
CDS, Red Wagon Smoothies Inc.



kaggle



Crime Dataset

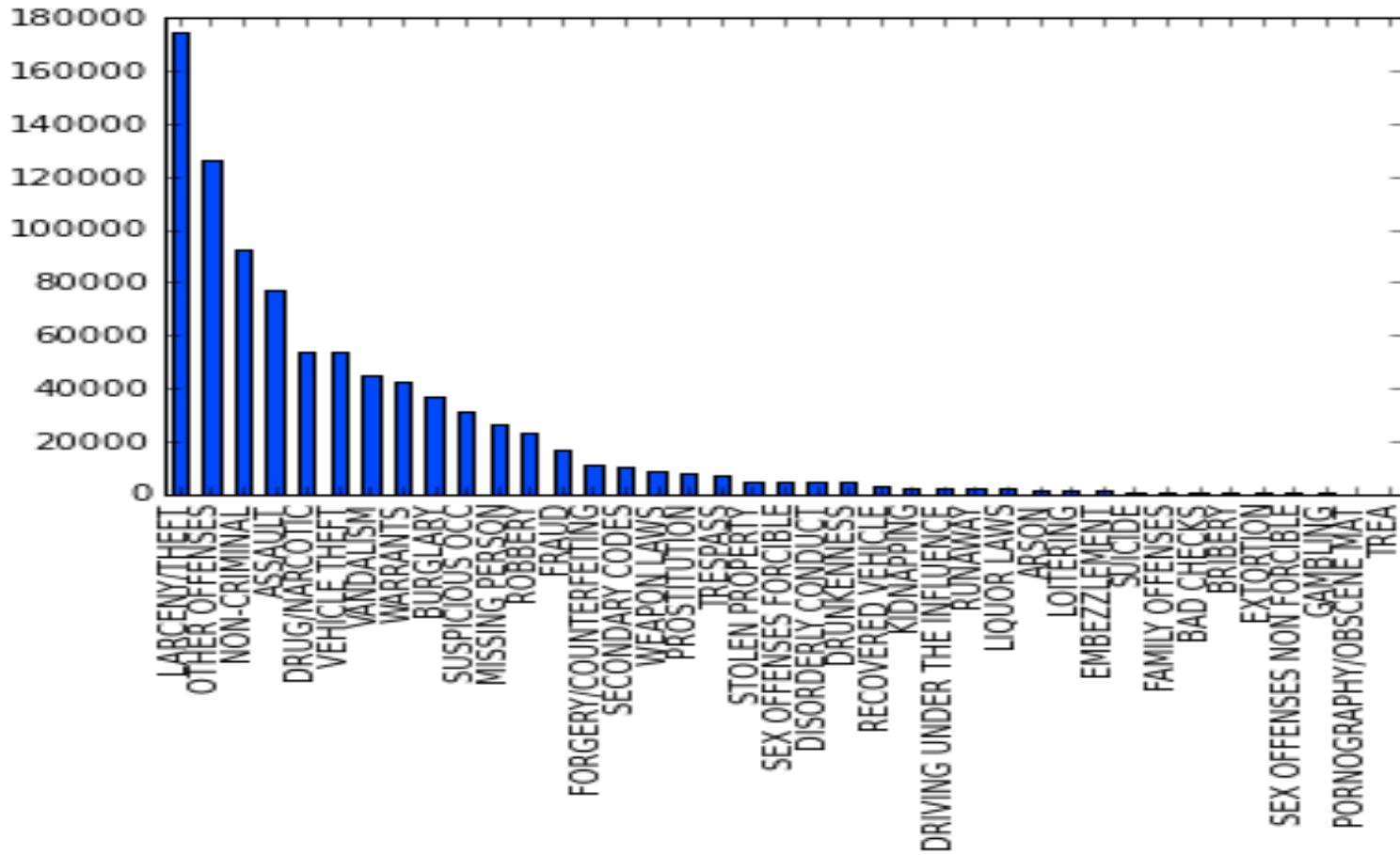
Response

Not in test data 😢

Not in test data 😢

Dates	Category	Descript	DayOfWeek	PdDistrict	Resolution	Address	X	Y
5/13/15 23:53	WARRANTS	WARRANT ARREST	Wednesday	NORTHERN	ARREST, BOOKED	OAK ST / LAGUNA ST	-122.426	37.7746
5/13/15 23:53	OTHER OFFENSES	TRAFFIC VIOLATION ARREST	Wednesday	NORTHERN	ARREST, BOOKED	OAK ST / LAGUNA ST	-122.426	37.7746
5/13/15 23:33	OTHER OFFENSES	TRAFFIC VIOLATION ARREST	Wednesday	NORTHERN	ARREST, BOOKED	VANNESS AV / GREEN	-122.424	37.80041
5/13/15 23:30	LARCENY/THEFT	GRAND THEFT FROM LOCKED	Wednesday	NORTHERN	NONE	1500 Block of LOMBA	-122.427	37.80087
5/13/15 23:30	LARCENY/THEFT	GRAND THEFT FROM LOCKED	Wednesday	PARK	NONE	100 Block of BRODERI	-122.439	37.77154
5/13/15 23:30	LARCENY/THEFT	GRAND THEFT FROM UNLOCK	Wednesday	INGLESIDE	NONE	0 Block of TEDDY AV	-122.403	37.71343
5/13/15 23:30	VEHICLE THEFT	STOLEN AUTOMOBILE	Wednesday	INGLESIDE	NONE	AVALON AV / PERU A	-122.423	37.72514
5/13/15 23:30	VEHICLE THEFT	STOLEN AUTOMOBILE	Wednesday	BAYVIEW	NONE	KIRKWOOD AV / DON	-122.371	37.72756
5/13/15 23:00	LARCENY/THEFT	GRAND THEFT FROM LOCKED	Wednesday	RICHMOND	NONE	600 Block of 47TH AV	-122.508	37.7766
5/13/15 23:00	LARCENY/THEFT	GRAND THEFT FROM LOCKED	Wednesday	CENTRAL	NONE	JEFFERSON ST / LEAVI	-122.419	37.8078
5/13/15 22:58	LARCENY/THEFT	PETTY THEFT FROM LOCKED A	Wednesday	CENTRAL	NONE	JEFFERSON ST / LEAVI	-122.419	37.8078
5/13/15 22:30	OTHER OFFENSES	MISCELLANEOUS INVESTIGATI	Wednesday	TARAVAL	NONE	0 Block of ESCOLTA W	-122.488	37.73767
5/13/15 22:30	VANDALISM	MALICIOUS MISCHIEF, VANDA	Wednesday	TENDERLOIN	NONE	TURK ST / JONES ST	-122.412	37.783
5/13/15 22:06	LARCENY/THEFT	GRAND THEFT FROM LOCKED	Wednesday	NORTHERN	NONE	FILLMORE ST / GEARY	-122.433	37.78435
5/13/15 22:00	NON-CRIMINAL	FOUND PROPERTY	Wednesday	BAYVIEW	NONE	200 Block of WILLIAM	-122.398	37.72993

The dreaded long tail



Feature Engineering

- Dummy columns for day of week, police district, and is_weekend
- time_of_day_bucket

zip	CensusAgeMedianAge	CensusAge18To19	CensusHispanicMexican	CensusRacePercentOtherOnly	CensusHouseholdsPercentSingleMoms
94108	43.8	206	476	2	1.1
94102	40.9	644	3974	9.2	2.2
94109	37.6	995	3779	4.3	1.2
94103	39.3	372	3100	8.9	2.7
94102	40.9	644	3974	9.2	2.2
94112	39.8	1986	9594	14.5	4.7
94102	40.9	644	3974	9.2	2.2
94110	36.2	1027	13356	17.4	4.4
94103	39.3	372	3100	8.9	2.7
94117	33.8	1561	2104	2.9	1.8
94132	32.9	3193	1707	5.1	3.4
94102	40.9	644	3974	9.2	2.2
94123	35	97	601	1.2	1
94111	51.4	24	76	1.3	0.6
94103	39.3	372	3100	8.9	2.7
94134	39.5	1066	3956	11.7	8.1
94111	51.4	24	76	1.3	
94117	33.8	1561	2104	2.9	
94124	34.6	1107	3902	15.1	
94122	39.1	952	1633	2	

Crime Models

- Baseline accuracy: 19.93%

Model	Accuracy	Params
Knn	20.02%	n_neighbors = 27
Decision Tree	23.45%	max_depth = 4
Naïve Bayes	17%	features used = 32
Random Forest	12%	n_estimators = 301



Kaggle Submission

268	↓24	BenB	2.66353	4	Wed, 12 Aug 2015 01:05:09 (-14d)
269	↓24	citrucel	2.66353	2	Tue, 30 Jun 2015 18:17:13 (-1.1h)
270	↓24	S. J. Stanley	2.66780	9	Tue, 11 Aug 2015 01:59:27 (-25.5h)
271	new	Brayden Cleary	2.66956	1	Mon, 24 Aug 2015 00:55:01
272	↓25	woodwill	2.67171	1	Tue, 30 Jun 2015 23:20:32
273	↓25	rhoslug	2.67235	5	Wed, 08 Jul 2015 06:44:21 (-25.2h)
274	↓25	G.M.S.	2.67274	2	Fri, 24 Jul 2015 02:55:15 (-30.4d)

Predicting Optimal Smoothie Selling Location



Red Wagon Smoothies



Smoothie Data

- Crimes that help me selling smoothies:
 - Arson, drunkenness, drug/narcotic, gambling, and loitering
- Crimes that hurt my selling smoothies:
 - Vehicle theft, recovered vehicle, stolen property, robbery, and DUI
- Reduced dataset to 147,862 rows 

Smoothie Model

- Baseline accuracy: 57.03%

Model	Accuracy	Specificity	Sensitivity
Decision Tree	72%	70%	75%
Knn	68%	60%	72%
Random Forest	67%	65%	68%
Normalized Knn	66%	59%	68%
Logistic Regression	62%	60%	63%
SVM	64%	60%	65%
Naïve Bayes	58%	57%	62%

<http://www.smoothiecrime.herokuapp.com>

