Preparing the Police:
Predicting Resistance to Police Activity Before Officers Arrive

Dieter Erben, Arushi Himatsingka, Ademola Oladosu, Brina Seidel

DS 1001 Final Project
December 2018

1. BUSINESS UNDERSTANDING

Policing practices are currently one of the most relevant social issues in the United States, where institutional racism and police violence have led to protests and movements like Black Lives Matter. Research in this area has been focused on police use of force [1], where police actions are scrutinized, and the solutions are focused on supervising these actions. We take a different approach by evaluating the likelihood of suspect resistance.

Excessive use of force often occurs in cases where police officers are not properly prepared to respond to resistance in a proportionate, appropriate manner. There is 1) research on the "relative amount of force used by the police compared to that used by suspects"[2] and 2) research on the response to resistance in officer shooting situations[3]. In the latter study in Houston, officers who were trained in "critical incident" situations were less likely to use their guns in such situations, as "an officer given information about a CIT [Critical Incident Training] incident before arriving at the scene was 82 percent less likely to shoot than other officers."[4] Therefore, the focus of our project is identifying these situations similar to the "CIT" ones mentioned above: situations where police officers can be informed about a potential suspect resistance. This creates direct value in terms of fewer lawsuits, improving public and media opinion, better allocation of police resources, and successful policing in general.

The process of data mining can be used to 1) collect potential indicators of suspect resistance that are available to the police before dispatch and 2) generate a classification model with a probability of resistance based on the specific circumstances. With these predictions, a police department could send officers who are trained in these situations or at least aware of the potential of suspect resistance. Our goal was therefore to build such a model for the Austin Police.

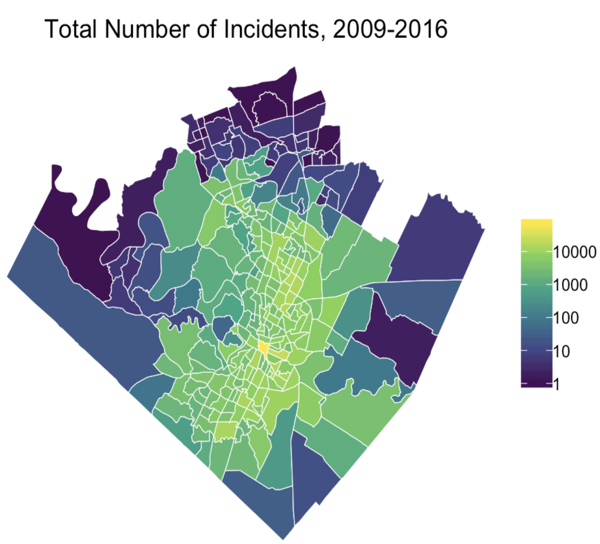
2. DATA UNDERSTANDING

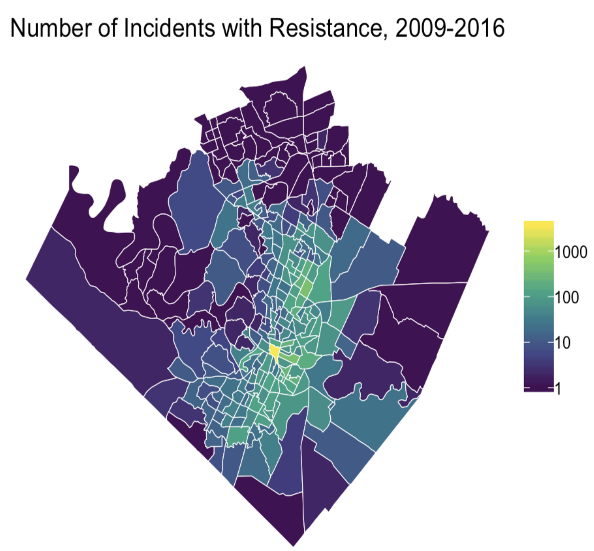
For the scope of this project, we decided to focus on one city and chose Austin, TX because of its proximity to Houston and because its data is publicly available. However, our process can be generalized to other cities depending on the data collected by their police department.

We chose two datasets from the Austin Police Department. The first one is the “Annual Crime Data”[5]. This data set contained crime report data in Austin from 2004 to 2018. However, we limited this dataset to the years 2009 to 2016 as we only had data for those years in the second dataset (below). Every data instance recorded a crime incident in Austin with a unique Incident Number and details like Location, Time, Description of Crime, etc.

The second one is the “Response to Resistance Dataset”[6]. It contains records of resistance to arrest with information like time, address, severity of the resistance, usage of weapon, injury, etc. Each record represents a different police report about an incident where the officer reported resistance. However, it is important to note that these are reports of resistance *by police officers*. There could be cases were there where resistance occurred but was not reported by the officer. Such cases introduce selection bias into our model and could perhaps result in an artificially inflated recall rate. A sample of both datasets, showing a subset of available features, can be found in the Appendix.

The maps below display the spatial distribution of both crimes and resistance. As the scales clearly show, resistance is a relatively rare event: just 1.3% of incidents involve resistance.





Since these two datasets focus on the nature of the incident, we decided to supplement the datasets with other variables we hypothesized could play a role in our predictions. We added Austin weather data[7], national holiday information[8], US census demographic data[9], and American Community Survey[10] data on employment, income and poverty. With these added features, our analysis can be more comprehensive.

3. DATA PREPARATION

3.1. Merging the datasets

We found that the crime incident numbers in the Crime Report Data matched primary key values in the Resistance Dataset. We confirmed this by ensuring that the date, time, location and description of crimes with the same primary key number in both datasets were the same. However, in the resistance data, while each primary key represented a unique incident, each instance of the data was a single officer's report of that incident. This means that there were often multiple records per primary key. In cases where the officers' reports conflicted, we used the highest level of resistance reported by any officer.

3.2. Cleaning the data

After merging the datasets, and including the Austin weather, US census and national holiday data, we performed a few more data cleaning operations:

- Dropping redundant variables: certain variables were not useful for modeling such as Internal Police Protocol Numbers, Reporting Date etc. and we decided to drop them.
- Handling features with missing values: some variables like Address had a very large number of missing values so we decided to remove these variables. Since we have a large number of features, we do not think the accuracy of our results will be majorly affected by removing these variables.
- Handling rows with missing values: We dropped rows with missing data because complete cases were needed to run any models. While the cases that we dropped were slightly less likely to have resistance than the cases we kept, we dropped less than 1% of our total dataset so it is unlikely that we introduced much bias.
- Type casting variables: many variables such as date and time of the incident were recorded as strings and we reformatted these into the correct data type.

3.3. Preparing the Target Variable

For our target variable, we created a binary variable which measures whether there was any resistance to arrest. This is defined as a report submitted to the Austin Police Department describing any resistance. We assumed that the incidents which were recorded in the resistance dataset had faced resistance to arrest, whereas, those that were not recorded, had not faced any resistance.

3.4. Feature Engineering

For our features, we performed the following transformations:

- We separated our date time variables into separate columns for weekday, month, hour etc. To make features usable by the model, we transformed these categorical variables into dummy variables for each possible hour of day, day of week, etc.
- Similarly, to use the National Holiday Data and Weather Data in our models, we created dummy variables. For example, we made dummy variables for important US holidays like Thanksgiving and New Year's Day.
- We again transformed categorical variables like location type, zip code, and police department sector into dummy variables.
- Using the census data, we calculated the percent of individuals of each race and age group in each census tract based on the counts and the total population.
- Using the ACS data, we calculated the unemployment rate in the census tract as the number unemployed divided by the size of the labor force, and the poverty rate as the total number in poverty divided by the total population.

This left us with a total of 1,033,085 police incidents between 2009 and 2016. Because we created so many dummy variables, the number of features in our dataset jumped from about 40 to 315. Descriptive statistics for all features used can be found in the Appendix.

4. MODELING

4.1. Train and Test Split

After feature engineering was complete, we separated our data into a 20-80 split where 80% was be used for training and model selection, and 20% was used to test our final model's performance.

Our overall data had a base rate of 1.3%, meaning just 1.3% of incidents had any resistance to an arrest. We therefore needed to guard against the possibility of generating a training or testing set with too few resistance cases in order to reduce variance in our model's performance. We therefore preserved the percentage of samples for each class when splitting between the data by stratifying on the target variable when performing this initial 20-80 split.

4.2. Baseline Model

We hypothesized that a logistic regression on several key features would generate a useful baseline model. We selected features based on the simplest variables that are readily available to the police department without requiring further data mining. These included a dummy marking which police department sector the incident occurred in, as well as further dummies for the month, the day of the week, and the hour of the day. This represented a total of 22 features out of our total of 315. A logistic model was selected because of its interpretability and known computational efficiency. The model generated showed an AUC of 0.62, to which we can compare results obtained in subsequent models.

We proceed with using the AUC as our evaluation metric because of the precision and recall considerations for any model that is proposed for this government application. With AUC as an evaluation metric, it is possible to evaluate the false positive rate and true positive rate trade-off as we look into how to deploy the model.

4.3. Model Selection

Recognizing that generating more complex models would take considerably more time, we wanted to pare down our list of 315 features before trying additional models. We opted to do this by running a logistic regression with L1 (Lasso) regularization on a standardized copy of the data. This allows us to quickly discard features which had a 0 coefficient and proceed with a more meaningful set of features. The lasso regularization returned 192 non-zero features with non-zero coefficients and these formed the basis for the rest of our modeling. We also considered using principal component analysis for dimensionality reduction. However, transparency is important in any government application. We therefore wanted the informed public to be able to interpret our results, which would be challenging if we were using principal components instead of a subset of the original features.

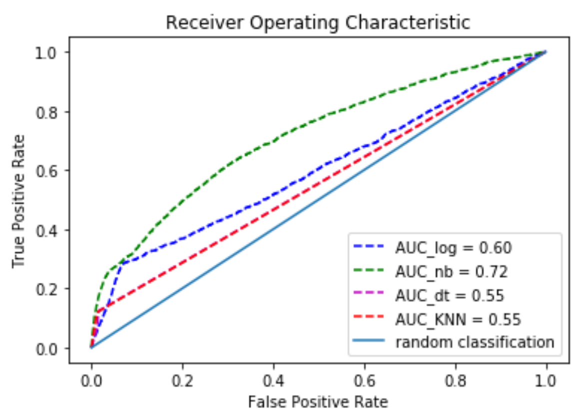
Based on this regression, the features that most predict resistance include the dummies marking the hours of 1am and 2am. The dummies marking streets, highways, and residences also had large coefficients, as did the dummy marking the 78701 zip code in downtown Austin.

Within each generated model below, we employed k-fold cross-validation for bias control, again stratifying by the target variable. By dividing our training data (80% selected from above) into 5 folds, we are able to run each model five times such that each observation acts as training data in 4 models and as test data in one model. While the smaller sample size increases the variance of our models, the use of multiple models reduces the bias. Below is a summary of the models we tried.

- Logistic Regression: The first generated model expanded the baseline model to develop a logistic regression model which included the 162 features that were selected after regularization. When this was done with an initial regularization corresponding to a C value of 10^{30}, we obtained an AUC of 0.60. We expect the results to degrade as C reduces and we increase regularization. When this is compared to the baseline model of 0.62, we see that adding more features yielded a lower AUC. This can be explained by the fact that adding non-informative features introduces noise and degrades the estimates of coefficients of useful features.
- Support Vector Machines: A logical next step was to try SVM classification. The general challenge with SVM modeling is computational expensiveness, which results from attempting to generate probability estimates with Platt scaling. Several attempts to generate probability estimates were unsuccessful even when the SVM classification was restricted to linear kernels using a small subset of the data. The same issue was encountered when the linearSVC package was used instead of the generalized SVC package available from sklearn. In general, we expect SVM and linear regression to perform similarly and prefer SVM methodologies when we seek to investigate a non-linear decision boundary. Given that we don't have prior information suggesting that a non-linear decision boundary applies, we decided to proceed with the logistic regression.
- Naive Bayes Classifier: The next attempted model was the Naive Bayes classifier, in which we assumed our features to be independent. Inspection of the data features suggests that this cannot be completely accurate, but this assumption may still provide a useful model. This model generated an AUC of 0.72, which represents a significant improvement over the baseline model.
- Decision Tree Classifier: We attempted a decision tree classifier and obtained an AUC of 0.55. We controlled complexity by first limiting the maximum allowable depth (high bias model) and then attempting the model with a higher limit (high variance model). The range of AUCs were 0.53-0.55 and the best score fell short of the AUC obtained from the logistic regression. Due to the poor performance of this model, random forests were not explored.

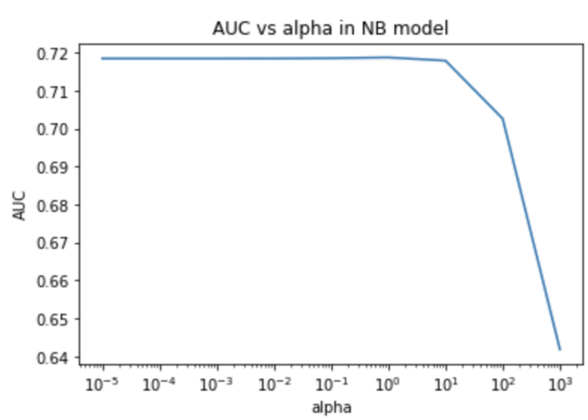
- <u>K-Nearest Neighbors Classifier</u>: A similar result of 0.55 was obtained for the best AUC using the K-nearest neighbors' classifier (K = 1000). The overfitted high-variance model (K = 1) showed the lowest AUC, which increased with K, until it started to flatten out at AUC = 0.55. Computational expensiveness prevented trying higher K values but we expect the AUC to marginally increase as we increase K (low variance model) before it starts to decrease. It is not expected that the marginally increasing AUC will match or exceed the performance of the logistic regression or Naive Bayes model.

Despite the inherent assumptions of the Naive Bayes, we proceed with this model because its AUC was considerably higher than that of the other models. Additionally, the shape of the ROC shows that this model strictly dominates the other models, with its true positive rate being higher than the other models for any given false positive rate. This can be seen below in the ROC curves for each generated model.



4.4. Model Tuning

After selecting the Naive Bayes model, we tuned the model by exploring various alpha parameters to introduce Laplace smoothing. While changing alpha had little effect for values less than 1, values greater than 1 sharply degraded the performance of our model.



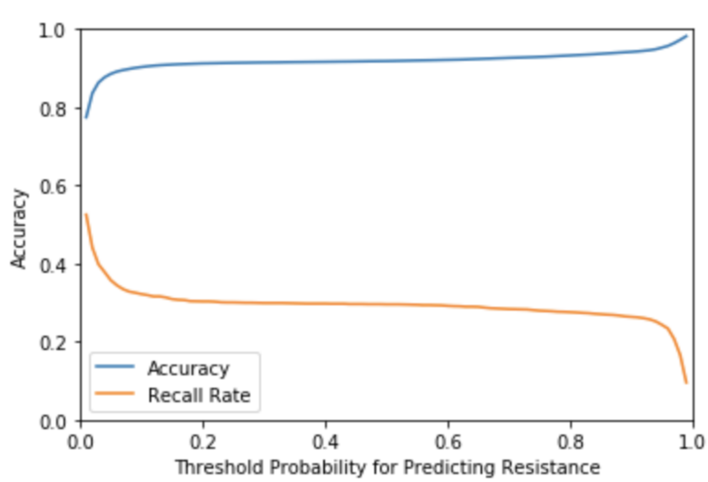
We therefore selected alpha=1 as our final hyper parameter because all alpha values less than or equal to 1 produced essentially identical results.

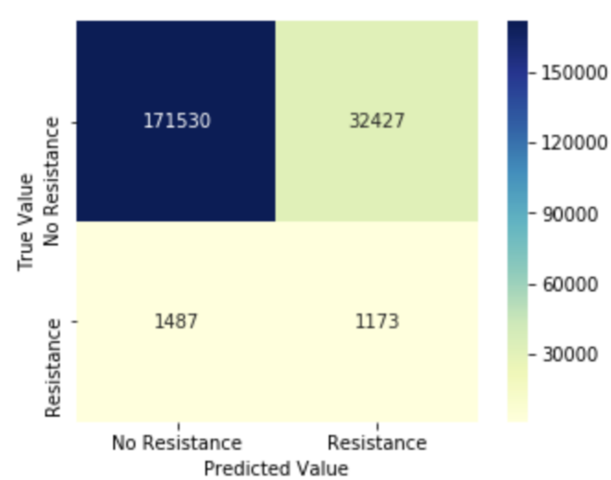
5. EVALUATION

We trained our final version of the Naive Bayes model on the full 80% of the training data and used this model to predict classes we obtain our final AUC by predicting on the 20% hold-out data that was withheld from our initial split. The AUC was 0.72.

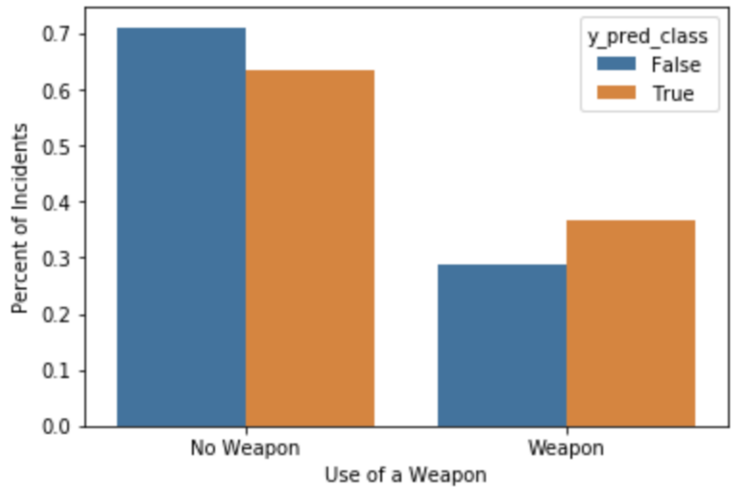
However, the problem remains of choosing a threshold for the predicted probability of resistance above which police officers are alerted to the risk. Deciding on this threshold required us to consider the different risks inherent to false positives and false negatives in the policing context. False positives would cause the dispatcher to send an officer into a safe situation thinking that the risk of resistance is high. This is unideal because it might result in a heightened emotional state that could lead to excessive use of force by an untrained officer. On the other hand, a false negative - telling an officer that there is little risk of resistance before encountering a suspect who resists arrest - could be very dangerous indeed. We therefore wanted to choose a threshold that maximizes the recall rate without sacrificing too much overall accuracy.

The left plot below shows that, at low thresholds, the false negative rate falls much faster than the accuracy rate rises, suggesting that we should pick a low threshold to maximize the recall rate without sacrificing much accuracy. We therefore selected a threshold of 0.02, at which point the recall rate is 44% and the accuracy is still relatively high at 84%. The resulting confusion matrix is also shown below.





While our model does not retrieve all cases in which there was resistance, it does disproportionately retrieve the more dangerous cases. When comparing incidents in which our model correctly identified resistance to incidents with resistance that were missed, we find that those which we identified were more likely to involve the use of a weapon. 37% of our true positive cases involved the use of a weapon, compared to 29% of false negatives and a base rate of 32%.



This model will provide the Austin Police Department with an improved ability to predict resistance to an arrest scenario and either have the department deploy trained officers in this scenario or at least allow police officers to arrive at the scene with better awareness of the environment.

6. DEPLOYMENT

In an actual production system, one of the issues we would encounter would be promptness of the recommendation to the officer. As the Austin Police Department receives a call about a crime, our model would need to generate the probability in a timely manner so that the officers can react accordingly and make decisions. One way to handle this would be to integrate the relevant features into the existing police dispatch system, which would then calculate the probability of encountering resistance to an arrest. Based on this output, the Austin Police Department can decide if they need to send a special team to handle a situation with the potential of resistance.

Luckily, the time-consuming part of our model is fitting the data to the model, so an officer could run the model and get a probability right after submitting the form. To maintain the model updated with more recent data and ensure there is no concept drift, the model could be trained and fitted at the end of every year, when the annual crime dataset is generated.

For the "Response to Resistance" dataset, the records are all reported by police officers. There could be reporting bias here, as some records could be influenced by the officer's perception of the event. An officer could make a mistake when writing the report, or even recording the event in an inaccurate way, either by not reporting resistance or changing the facts of the incident in favor of the police department.

We also have to be careful with census data regarding race and ACS data regarding income and economic status. It is important to notice that the model is a representation of reality, which can discriminate towards people of certain races and socioeconomic status. As certain areas are policed more often, it is more likely that a crime is reported in these areas. This societal bias can lead to a disparate impact where certain members of the population are being affected by these predictions in a disproportionate way. In order to address this issue, we have to consider individual fairness when deploying the model, meaning that two similar individuals should receive similar outcomes [11]. This means that an officer should treat two cases with the same probability in the same way, which is an attitude that can be refined through police training.

An additional risk stems from the fact that our model is not optimized to reduce false positives, in which resistance was predicted for incidents that will not have suspect resistance. This bias (from low precision) could firstly lead to a situation where police officers are anticipating resistances and are more likely to use excessive force because of this. Even if we assume that the deployed policemen are well-trained to avoid this tendency, there still exists the issue of the cost of training more officers or overusing some specialized police resources.

7. FUTURE CONSIDERATIONS

In this project, our goal was to predict the probability of resistance to an arrest. However, APD could be interested in differentiating between different types of resistance. For example, is this resistance seen with a suspect trying to run away, or is it a violent resistance where a gun is involved, and some shots were fired? These two very different kinds of resistance could lead to different action plans for APD. Tuning the model to highlight more severe situations would be a natural next step in this study.

LINK TO GITHUB REPOSITORY:

https://github.com/brinaseidel/Response_to_Resistance

BIBLIOGRAPHY

[1] Geoffrey Alpert and Roger Dunham, Force Factor: Measuring Police Use of Force Relative to Suspect Resistance, National Criminal Justice Reference Service.

[2] Geoffrey Alpert, et al., Police use of force: examining the relationship between calls for service and the balance of police force and suspect resistance, Journal of Criminal Justice

[3] Anthony L. Colucci, et al., Houston Police Department Officer-Involved Shooting, Northwestern University

[4] *Ibid.*

[5] https://data.austintexas.gov/Public-Safety/Crime-Reports-beta-/fdj4-gpfu

[6] https://data.austintexas.gov/Public-Safety/R2R-2010/q5ym-htjz

[7] https://www.kaggle.com/grubenm/austin-weather

[8] https://www.kaggle.com/gsnehaa21/federal-holidays-usa-19662020

[9] https://www.census.gov/data/developers/data-sets/decennial-census.html

[10] https://www.census.gov/data/developers/data-sets/acs-5year.html

[11] Julia Stoyanovich, *Responsible Data Science*, New York University, delivered November 7 2018

CONTRIBUTIONS

Dieter Erben Vasconcelos (dev241):

- Researched policing literature and previous studies on predictive policing
- Explored stratification methods to ensure balanced classes
- Wrote down business understanding and implications of model deployment

Arushi Himatsingka (ah3243):

- Cleaned variables for merging the datasets
- Conducted feature engineering for the models
- Drafted report structure and wrote out the data preparation and some modeling sections

Ademola Oladosu (ao1584):

- Obtained and cleaned holiday & weather information
- Ran SVM, KNN and NB models and set up hyperparameter optimization for these models
- Provided discussion on model selection / evaluation

Brina Seidel (bs3743):

- Pulled census and ACS data from API
- Set up stratified sampling for k-fold cross validation
- Ran L1 norm regression and some initial models
- Produced figures showing performance of the final model

APPENDIX

Crimes Dataset

	inc_number	date_time	high_off_desc	address	zip_code	apd_sector	cen_tract	loc_t ype
0	20145043640	10/01/2014 08:19:00 PM	VIOL OF PROTECTIVE ORDER	1708 WHELESS LN	78723.0	ID	21.12	RESIDENCE / HOME
1	20115047788	10/11/2011 06:00:00 AM	RUNAWAY CHILD	7403 LANGSTON DR	78723.0	ID	21.13	SCHOOLS / COLLEGES
3	20112011254	07/20/2011 05:05:00 PM	VIOL CITY ORDINANCE - OTHER	2000 BLOCK LOU NEFF RD	78746.0	DA	19.11	GOVERNMENT / PUBLIC BUILDING
6	20145045267	10/11/2014 05:00:00 PM	BURGLARY OF VEHICLE	E 7TH ST / TRINITY ST	78701.0	GE	11	PARKING LOTS / GARAGE
7	2013141039	01/14/2013 03:16:00 PM	AGG ROBBERY/DEADLY WEAPON	5431 N IH 35 SVRD NB	78723.0	ID	21.05	COMMERCIAL / OFFICE BUILDING

Resistance Dataset

	rin	prim_key	date	time	address	area_command	nature_of_contact	reason_desc	r2r_level
0	19654	200910720	01/01/2009 12:00:00 AM		200 W 4TH ST	GE	VIEWED OFFENSE	NECESSARY TO EFFECT ARREST / DETENTION	3.0
1	19656	200910883	01/01/2009 12:00:00 AM	0330	6309 BURNS ST	ID	DISPATCHED CALL	NECESSARY TO DEFEND REPORTING OFFICER	2.0
2	19734	200910883	01/01/2009 12:00:00 AM		6309 BURNS ST	ID	DISPATCHED CALL	NECESSARY TO EFFECT ARREST / DETENTION	1.0
3	19735	200911936	01/01/2009 12:00:00 AM	1640	6710 ARROYO SECO	ID	DISPATCHED CALL	OTHER (DOCUMENT IN SUPPLEMENT)	3.0
4	19736	200911936	01/01/2009 12:00:00 AM	1635	6710 ARROYO SECO	ID	DISPATCHED CALL	OTHER (DOCUMENT IN SUPPLEMENT)	3.0

Descriptive Statistics

Variable Name	Mean	Minimum	25%	50%	75%	Maximum
Target1	0.013	0.000	0.000	0.000	0.000	1.000
temp_max	82.039	26.000	72.000	84.000	94.000	112.000
temp_avg	71.320	22.000	61.000	73.000	84.000	96.000
temp_min	60.096	17.000	49.000	63.000	73.000	82.000
dew_max	60.734	0.000	52.000	65.000	72.000	80.000
dew_avg	55.479	0.000	45.000	60.000	68.000	76.000
dew_min	49.122	-2.000	36.000	54.000	63.000	75.000
hum_max	85.711	0.000	82.000	88.000	93.000	100.000
hum_min	40.847	0.000	28.000	39.000	51.000	93.000
wind_max	13.346	0.000	10.000	13.000	15.000	31.000
wind_min	0.850	0.000	0.000	0.000	0.000	9.000
pres_max	30.018	0.000	29.970	30.060	30.190	30.840
pres_min	29.838	0.000	29.810	29.890	30.010	30.660
prec_avg	0.094	0.000	0.000	0.000	0.000	7.040
new_year	0.003	0.000	0.000	0.000	0.000	1.000
mlk_day	0.002	0.000	0.000	0.000	0.000	1.000
wash_bday	0.002	0.000	0.000	0.000	0.000	1.000
mem_day	0.003	0.000	0.000	0.000	0.000	1.000
ind_day	0.003	0.000	0.000	0.000	0.000	1.000
labor_day	0.003	0.000	0.000	0.000	0.000	1.000
col_day	0.003	0.000	0.000	0.000	0.000	1.000
vet_day	0.003	0.000	0.000	0.000	0.000	1.000
thanksgiving	0.002	0.000	0.000	0.000	0.000	1.000
christmas	0.002	0.000	0.000	0.000	0.000	1.000
month1	0.083	0.000	0.000	0.000	0.000	1.000
month2	0.075	0.000	0.000	0.000	0.000	1.000
month3	0.087	0.000	0.000	0.000	0.000	1.000
month4	0.086	0.000	0.000	0.000	0.000	1.000
month5	0.089	0.000	0.000	0.000	0.000	1.000
month6	0.085	0.000	0.000	0.000	0.000	1.000
month7	0.088	0.000	0.000	0.000	0.000	1.000
month8	0.087	0.000	0.000	0.000	0.000	1.000
month9	0.082	0.000	0.000	0.000	0.000	1.000
month10	0.084	0.000	0.000	0.000	0.000	1.000
month11	0.077	0.000	0.000	0.000	0.000	1.000
month12	0.078	0.000	0.000	0.000	0.000	1.000
day0	0.139	0.000	0.000	0.000	0.000	1.000
day1	0.138	0.000	0.000	0.000	0.000	1.000
day2	0.139	0.000	0.000	0.000	0.000	1.000
day3	0.140	0.000	0.000	0.000	0.000	1.000
day4	0.153	0.000	0.000	0.000	0.000	1.000
day5	0.150	0.000	0.000	0.000	0.000	1.000
day6	0.140	0.000	0.000	0.000	0.000	1.000
hour0	0.061	0.000	0.000	0.000	0.000	1.000
hour1	0.043	0.000	0.000	0.000	0.000	1.000
hour2	0.043	0.000	0.000	0.000	0.000	1.000
hour3	0.028	0.000	0.000	0.000	0.000	1.000
hour4	0.017	0.000	0.000	0.000	0.000	1.000
hour5	0.012	0.000	0.000	0.000	0.000	1.000
hour6	0.014	0.000	0.000	0.000	0.000	1.000
hour7	0.022	0.000	0.000	0.000	0.000	1.000
hour8	0.029	0.000	0.000	0.000	0.000	1.000
hour9	0.030	0.000	0.000	0.000	0.000	1.000
hour10	0.036	0.000	0.000	0.000	0.000	1.000
hour11	0.035	0.000	0.000	0.000	0.000	1.000
hour12	0.066	0.000	0.000	0.000	0.000	1.000
hour13	0.037	0.000	0.000	0.000	0.000	1.000
hour14	0.040	0.000	0.000	0.000	0.000	1.000
hour15	0.043	0.000	0.000	0.000	0.000	1.000
hour16	0.046	0.000	0.000	0.000	0.000	1.000
hour17	0.053	0.000	0.000	0.000	0.000	1.000
hour18	0.057	0.000	0.000	0.000	0.000	1.000
hour19	0.053	0.000	0.000	0.000	0.000	1.000
hour20	0.055	0.000	0.000	0.000	0.000	1.000
hour21	0.058	0.000	0.000	0.000	0.000	1.000
hour22	0.064	0.000	0.000	0.000	0.000	1.000
hour23	0.057	0.000	0.000	0.000	0.000	1.000
P0030001	4,963.288	567.000	3,616.00	4,996.00	6,161.00	16,495.00

P0130001	31.113	19.600	27.800	30.600	34.300	50.700
P0130002	31.039	19.700	27.900	30.600	33.600	49.100
P0130003	31.147	19.500	27.500	30.700	33.500	51.900
P0030002_pct	0.649	0.281	0.525	0.656	0.789	0.961
P0030003_pct	0.099	0.003	0.045	0.078	0.114	0.528
P0030004_pct	0.010	0.000	0.007	0.009	0.014	0.029
P0030005_pct	0.048	0.001	0.013	0.030	0.056	0.274
P0030006_pct	0.001	0.000	0.000	0.001	0.001	0.004
P0030007_pct	0.159	0.003	0.039	0.139	0.259	0.394
P0030008_pct	0.034	0.013	0.028	0.035	0.040	0.065
P0040003_pct	0.410	0.049	0.160	0.400	0.641	0.845
P0200002_pct	0.282	0.008	0.194	0.283	0.372	0.645
P0250002_pct	0.121	0.005	0.080	0.110	0.156	0.420
P0190007_pct	0.312	0.013	0.201	0.303	0.403	0.881
B19113_001E	76,983.82	8,606.0	39,500.0	61,543.0	90,366.0	250,001.0
B17001_001E_pct	0.983	0.080	0.987	0.996	1.000	1.000
C18120_003E_pct	0.947	0.828	0.931	0.950	0.968	0.997
apd_sector1	0.000	0.000	0.000	0.000	0.000	0.000
apd_sector1124	0.000	0.000	0.000	0.000	0.000	0.000
apd_sector2	0.000	0.000	0.000	0.000	0.000	0.000
apd_sector4	0.000	0.000	0.000	0.000	0.000	0.000
apd_sector8	0.000	0.000	0.000	0.000	0.000	0.000
apd_sector83	0.000	0.000	0.000	0.000	0.000	0.000
apd_sector88	0.002	0.000	0.000	0.000	0.000	1.000
apd_sector99	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorA1	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorAD	0.094	0.000	0.000	0.000	0.000	1.000
apd_sectorADAM	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorAP	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorBA	0.106	0.000	0.000	0.000	0.000	1.000
apd_sectorBAKR	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorC1	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorCH	0.099	0.000	0.000	0.000	0.000	1.000
apd_sectorDA	0.129	0.000	0.000	0.000	0.000	1.000
apd_sectorDAVD	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorED	0.135	0.000	0.000	0.000	0.000	1.000
apd_sectorFR	0.126	0.000	0.000	0.000	0.000	1.000
apd_sectorG	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorGE	0.088	0.000	0.000	0.000	0.000	1.000
apd_sectorGRGE	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorH	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorHE	0.112	0.000	0.000	0.000	0.000	1.000
apd_sectorHENRY	0.000	0.000	0.000	0.000	0.000	1.000
apd_sectorHR	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorI	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorID	0.109	0.000	0.000	0.000	0.000	1.000
apd_sectorIDA	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorRF	0.000	0.000	0.000	0.000	0.000	0.000
zip0	0.000	0.000	0.000	0.000	0.000	0.000
zip76574	0.000	0.000	0.000	0.000	0.000	0.000
zip78610	0.000	0.000	0.000	0.000	0.000	1.000
zip78612	0.000	0.000	0.000	0.000	0.000	0.000
zip78613	0.004	0.000	0.000	0.000	0.000	1.000
zip78617	0.007	0.000	0.000	0.000	0.000	1.000
zip78619	0.000	0.000	0.000	0.000	0.000	0.000
zip78621	0.000	0.000	0.000	0.000	0.000	0.000
zip78626	0.000	0.000	0.000	0.000	0.000	0.000
zip78640	0.000	0.000	0.000	0.000	0.000	0.000
zip78641	0.000	0.000	0.000	0.000	0.000	1.000
zip78645	0.000	0.000	0.000	0.000	0.000	1.000
zip78652	0.000	0.000	0.000	0.000	0.000	1.000
zip78653	0.001	0.000	0.000	0.000	0.000	1.000
zip78660	0.003	0.000	0.000	0.000	0.000	1.000
zip78664	0.000	0.000	0.000	0.000	0.000	1.000
zip78669	0.000	0.000	0.000	0.000	0.000	0.000
zip78681	0.000	0.000	0.000	0.000	0.000	1.000
zip78701	0.075	0.000	0.000	0.000	0.000	1.000
zip78702	0.057	0.000	0.000	0.000	0.000	1.000
zip78703	0.016	0.000	0.000	0.000	0.000	1.000
zip78704	0.067	0.000	0.000	0.000	0.000	1.000
zip78705	0.025	0.000	0.000	0.000	0.000	1.000
zip78712	0.000	0.000	0.000	0.000	0.000	1.000

zip78717	0.006	0.000	0.000	0.000	0.000	1.000
zip78719	0.000	0.000	0.000	0.000	0.000	1.000
zip78721	0.021	0.000	0.000	0.000	0.000	1.000
zip78722	0.009	0.000	0.000	0.000	0.000	1.000
zip78723	0.054	0.000	0.000	0.000	0.000	1.000
zip78724	0.017	0.000	0.000	0.000	0.000	1.000
zip78725	0.002	0.000	0.000	0.000	0.000	1.000
zip78726	0.005	0.000	0.000	0.000	0.000	1.000
zip78727	0.015	0.000	0.000	0.000	0.000	1.000
zip78728	0.000	0.000	0.000	0.000	0.000	1.000
zip78729	0.012	0.000	0.000	0.000	0.000	1.000
zip78730	0.001	0.000	0.000	0.000	0.000	1.000
zip78731	0.012	0.000	0.000	0.000	0.000	1.000
zip78732	0.000	0.000	0.000	0.000	0.000	1.000
zip78733	0.000	0.000	0.000	0.000	0.000	1.000
zip78734	0.000	0.000	0.000	0.000	0.000	1.000
zip78735	0.007	0.000	0.000	0.000	0.000	1.000
zip78736	0.002	0.000	0.000	0.000	0.000	1.000
zip78737	0.000	0.000	0.000	0.000	0.000	1.000
zip78738	0.000	0.000	0.000	0.000	0.000	1.000
zip78739	0.003	0.000	0.000	0.000	0.000	1.000
zip78741	0.092	0.000	0.000	0.000	0.000	1.000
zip78742	0.001	0.000	0.000	0.000	0.000	1.000
zip78744	0.057	0.000	0.000	0.000	0.000	1.000
zip78745	0.065	0.000	0.000	0.000	0.000	1.000
zip78746	0.013	0.000	0.000	0.000	0.000	1.000
zip78747	0.007	0.000	0.000	0.000	0.000	1.000
zip78748	0.032	0.000	0.000	0.000	0.000	1.000
zip78749	0.018	0.000	0.000	0.000	0.000	1.000
zip78750	0.012	0.000	0.000	0.000	0.000	1.000
zip78751	0.029	0.000	0.000	0.000	0.000	1.000
zip78752	0.035	0.000	0.000	0.000	0.000	1.000
zip78753	0.080	0.000	0.000	0.000	0.000	1.000
zip78754	0.011	0.000	0.000	0.000	0.000	1.000
zip78756	0.007	0.000	0.000	0.000	0.000	1.000
zip78757	0.025	0.000	0.000	0.000	0.000	1.000
zip78758	0.069	0.000	0.000	0.000	0.000	1.000
zip78759	0.028	0.000	0.000	0.000	0.000	1.000
apd_sector1	0.000	0.000	0.000	0.000	0.000	0.000
apd_sector1124	0.000	0.000	0.000	0.000	0.000	0.000
apd_sector2	0.000	0.000	0.000	0.000	0.000	0.000
apd_sector4	0.000	0.000	0.000	0.000	0.000	0.000
apd_sector8	0.000	0.000	0.000	0.000	0.000	0.000
apd_sector83	0.000	0.000	0.000	0.000	0.000	0.000
apd_sector88	0.002	0.000	0.000	0.000	0.000	1.000
apd_sector99	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorA1	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorAD	0.094	0.000	0.000	0.000	0.000	1.000
apd_sectorADAM	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorAP	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorBA	0.106	0.000	0.000	0.000	0.000	1.000
apd_sectorBAKR	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorC1	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorCH	0.099	0.000	0.000	0.000	0.000	1.000
apd_sectorDA	0.129	0.000	0.000	0.000	0.000	1.000
apd_sectorDAVD	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorED	0.135	0.000	0.000	0.000	0.000	1.000
apd_sectorFR	0.126	0.000	0.000	0.000	0.000	1.000
apd_sectorG	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorGE	0.088	0.000	0.000	0.000	0.000	1.000
apd_sectorGRGE	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorH	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorHE	0.112	0.000	0.000	0.000	0.000	1.000
apd_sectorHENRY	0.000	0.000	0.000	0.000	0.000	1.000
apd_sectorHR	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorI	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorID	0.109	0.000	0.000	0.000	0.000	1.000
apd_sectorIDA	0.000	0.000	0.000	0.000	0.000	0.000
apd_sectorRF	0.000	0.000	0.000	0.000	0.000	0.000
zip0	0.000	0.000	0.000	0.000	0.000	0.000
zip76574	0.000	0.000	0.000	0.000	0.000	0.000
zip78610	0.000	0.000	0.000	0.000	0.000	1.000

zip78612	0.000	0.000	0.000	0.000	0.000	0.000
zip78613	0.004	0.000	0.000	0.000	0.000	1.000
zip78617	0.007	0.000	0.000	0.000	0.000	1.000
zip78619	0.000	0.000	0.000	0.000	0.000	0.000
zip78621	0.000	0.000	0.000	0.000	0.000	0.000
zip78626	0.000	0.000	0.000	0.000	0.000	0.000
zip78640	0.000	0.000	0.000	0.000	0.000	0.000
zip78641	0.000	0.000	0.000	0.000	0.000	1.000
zip78645	0.000	0.000	0.000	0.000	0.000	1.000
zip78652	0.000	0.000	0.000	0.000	0.000	1.000
zip78653	0.001	0.000	0.000	0.000	0.000	1.000
zip78660	0.003	0.000	0.000	0.000	0.000	1.000
zip78664	0.000	0.000	0.000	0.000	0.000	1.000
zip78669	0.000	0.000	0.000	0.000	0.000	0.000
zip78681	0.000	0.000	0.000	0.000	0.000	1.000
zip78701	0.075	0.000	0.000	0.000	0.000	1.000
zip78702	0.057	0.000	0.000	0.000	0.000	1.000
zip78703	0.016	0.000	0.000	0.000	0.000	1.000
zip78704	0.067	0.000	0.000	0.000	0.000	1.000
zip78705	0.025	0.000	0.000	0.000	0.000	1.000
zip78712	0.000	0.000	0.000	0.000	0.000	1.000
zip78717	0.006	0.000	0.000	0.000	0.000	1.000
zip78719	0.000	0.000	0.000	0.000	0.000	1.000
zip78721	0.021	0.000	0.000	0.000	0.000	1.000
zip78722	0.009	0.000	0.000	0.000	0.000	1.000
zip78723	0.054	0.000	0.000	0.000	0.000	1.000
zip78724	0.017	0.000	0.000	0.000	0.000	1.000
zip78725	0.002	0.000	0.000	0.000	0.000	1.000
zip78726	0.005	0.000	0.000	0.000	0.000	1.000
zip78727	0.015	0.000	0.000	0.000	0.000	1.000
zip78728	0.000	0.000	0.000	0.000	0.000	1.000
zip78729	0.012	0.000	0.000	0.000	0.000	1.000
zip78730	0.001	0.000	0.000	0.000	0.000	1.000
zip78731	0.012	0.000	0.000	0.000	0.000	1.000
zip78732	0.000	0.000	0.000	0.000	0.000	1.000
zip78733	0.000	0.000	0.000	0.000	0.000	1.000
zip78734	0.000	0.000	0.000	0.000	0.000	1.000
zip78735	0.007	0.000	0.000	0.000	0.000	1.000
zip78736	0.002	0.000	0.000	0.000	0.000	1.000
zip78737	0.000	0.000	0.000	0.000	0.000	1.000
zip78738	0.000	0.000	0.000	0.000	0.000	1.000
zip78739	0.003	0.000	0.000	0.000	0.000	1.000
zip78741	0.092	0.000	0.000	0.000	0.000	1.000
zip78742	0.001	0.000	0.000	0.000	0.000	1.000
zip78744	0.057	0.000	0.000	0.000	0.000	1.000
zip78745	0.065	0.000	0.000	0.000	0.000	1.000
zip78746	0.013	0.000	0.000	0.000	0.000	1.000
zip78747	0.007	0.000	0.000	0.000	0.000	1.000
zip78748	0.032	0.000	0.000	0.000	0.000	1.000
zip78749	0.018	0.000	0.000	0.000	0.000	1.000
zip78750	0.012	0.000	0.000	0.000	0.000	1.000
zip78751	0.029	0.000	0.000	0.000	0.000	1.000
zip78752	0.035	0.000	0.000	0.000	0.000	1.000
zip78753	0.080	0.000	0.000	0.000	0.000	1.000
zip78754	0.011	0.000	0.000	0.000	0.000	1.000
zip78756	0.007	0.000	0.000	0.000	0.000	1.000
zip78757	0.025	0.000	0.000	0.000	0.000	1.000
zip78758	0.069	0.000	0.000	0.000	0.000	1.000
zip78759	0.028	0.000	0.000	0.000	0.000	1.000
loc_typeABANDONED/CONDEMNED STRUCTURE	0.000	0.000	0.000	0.000	0.000	1.000
loc_typeAMUSEMENT PARK	0.000	0.000	0.000	0.000	0.000	1.000
loc_typeARENA / STADIUM / FAIRGROUNDS / COLISEUM	0.000	0.000	0.000	0.000	0.000	1.000
loc_typeATM SEPARATE FROM BANK	0.000	0.000	0.000	0.000	0.000	1.000
loc_typeAUTO DEALERSHIP NEW / USED	0.000	0.000	0.000	0.000	0.000	1.000
loc_typeBANKS / SAVINGS & LOAN	0.003	0.000	0.000	0.000	0.000	1.000
loc_typeBAR / NIGHT CLUB	0.011	0.000	0.000	0.000	0.000	1.000
loc_typeCAMP / CAMPGROUND	0.000	0.000	0.000	0.000	0.000	1.000

loc_typeCHURCH / SYNAGOGUE / TEMPLE / MOSQUE	0.002	0.000	0.000	0.000	0.000	1.000
loc_typeCOMMERCIAL / OFFICE BUILDING	0.073	0.000	0.000	0.000	0.000	1.000
loc_typeCOMMUNITY CENTER	0.000	0.000	0.000	0.000	0.000	1.000
loc_typeCONSTRUCTION SITE	0.003	0.000	0.000	0.000	0.000	1.000
loc_typeCONVENIENCE STORE	0.012	0.000	0.000	0.000	0.000	1.000
loc_typeDAYCARE FACILITY	0.000	0.000	0.000	0.000	0.000	1.000
loc_typeDEPARTMENT / DISCOUNT STORE	0.020	0.000	0.000	0.000	0.000	1.000
loc_typeDOCK / WHARF / FREIGHT / MODAL TERMINAL	0.000	0.000	0.000	0.000	0.000	1.000
loc_typeDRUG STORE / DR. OFFICE / HOSPITAL	0.013	0.000	0.000	0.000	0.000	1.000
loc_typeFARM FACILITY	0.000	0.000	0.000	0.000	0.000	1.000
loc_typeFIELD / WOODS	0.003	0.000	0.000	0.000	0.000	1.000
loc_typeGAMBLING FACILITY / CASINO / RACE TRACK	0.000	0.000	0.000	0.000	0.000	1.000
loc_typeGAS / SERVICE STATIONS	0.007	0.000	0.000	0.000	0.000	1.000
loc_typeGOVERNMENT / PUBLIC BUILDING	0.012	0.000	0.000	0.000	0.000	1.000
loc_typeGROCERY / SUPERMARKET	0.016	0.000	0.000	0.000	0.000	1.000
loc_typeHOTEL / MOTEL / ETC.	0.015	0.000	0.000	0.000	0.000	1.000
loc_typeINDUSTRIAL SITE	0.000	0.000	0.000	0.000	0.000	1.000
loc_typeJAIL / PRISON	0.000	0.000	0.000	0.000	0.000	1.000
loc_typeLAKE / WATERWAY	0.001	0.000	0.000	0.000	0.000	1.000
loc_typeLIQUOR STORE	0.001	0.000	0.000	0.000	0.000	1.000
loc_typeMILITARY INSTALLATION	0.000	0.000	0.000	0.000	0.000	1.000
loc_typeOTHER / UNKNOWN	0.029	0.000	0.000	0.000	0.000	1.000
loc_typePARK / PLAYGROUND	0.004	0.000	0.000	0.000	0.000	1.000
loc_typePARKING LOTS / GARAGE	0.112	0.000	0.000	0.000	0.000	1.000
loc_typeRENTAL STORAGE FACILITY	0.001	0.000	0.000	0.000	0.000	1.000
loc_typeRESIDENCE / HOME	0.404	0.000	0.000	0.000	1.000	1.000
loc_typeREST AREA	0.000	0.000	0.000	0.000	0.000	1.000
loc_typeRESTAURANTS	0.014	0.000	0.000	0.000	0.000	1.000
loc_typeSCHOOL - UNIVERSITY	0.000	0.000	0.000	0.000	0.000	1.000
loc_typeSCHOOL - ELEMENTARY / SECONDARY	0.000	0.000	0.000	0.000	0.000	1.000
loc_typeSCHOOLS / COLLEGES	0.003	0.000	0.000	0.000	0.000	1.000
loc_typeSHELTER-MISSION / HOMELESS	0.000	0.000	0.000	0.000	0.000	1.000
loc_typeSHOPPING MALL	0.001	0.000	0.000	0.000	0.000	1.000
loc_typeSPECIALTY STORE (TV FUR ETC.)	0.005	0.000	0.000	0.000	0.000	1.000
loc_typeSTREETS / HWY / ROAD / ALLEY	0.228	0.000	0.000	0.000	0.000	1.000
loc_typeTRANSPORTATION (AIR / BUS / TRAIN - TERMINALS)	0.003	0.000	0.000	0.000	0.000	1.000
loc_typeTRIBAL LANDS	0.000	0.000	0.000	0.000	0.000	1.000