

Crime, What's Next?

Springboard Data Science Career Track - Capstone 1
John Peterson - February Cohort





Overview

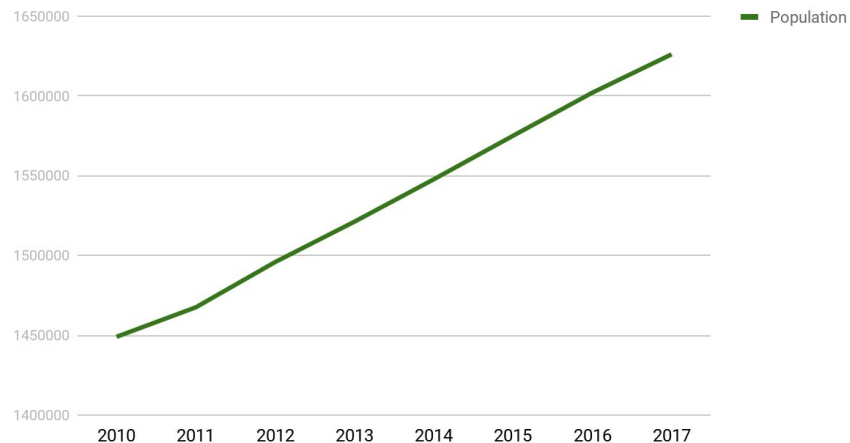
- Background
- Project Goal
- Data Wrangling
- Exploratory Data Analysis (EDA)
- Feature Engineering
- Machine Learning
- Conclusion and Next Steps



Background

- Phoenix 5th largest city in US in 2017
- Since 2010 there has been steady population growth
- With an increase in population there is the possibility for an increase in crime

City of Phoenix Population Growth



**Can we predict future
types of crime using
machine learning to aid
law enforcement?**





Project Goal

- Use historical crime data from the City of Phoenix to predict future crimes
- Explore the data and create features to better train the model
- Use different machine learning algorithms to get the most accurate prediction
- Determine the best algorithm and who can benefit from this

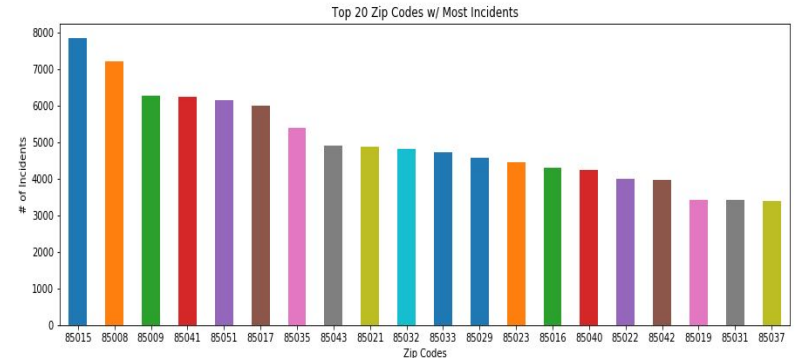
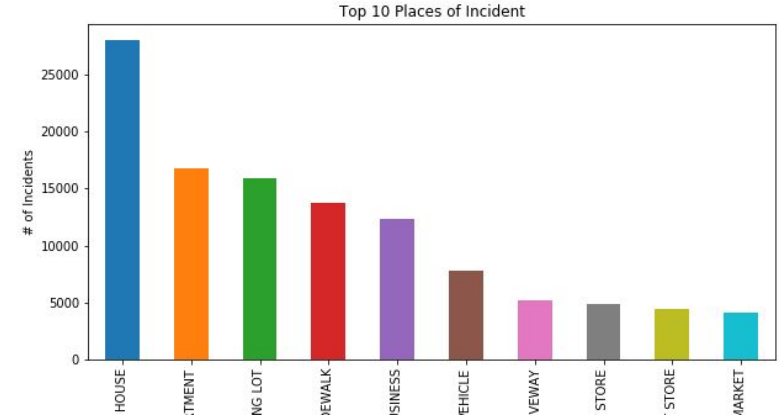


Data Wrangling

- Phoenix Crime Data
 - Type of Crime
 - Zip Code
 - 100 Block Address
 - Date Occurred On
 - Date Occurred To
 - Premise Type
- Categorical Type Data
- Filled in missing or NA values
- Removed rows
- Cleaned up naming conventions inside variable classes
- Converted date fields to datetime and brokedown datetime into new columns
- Created hour groups and season groups

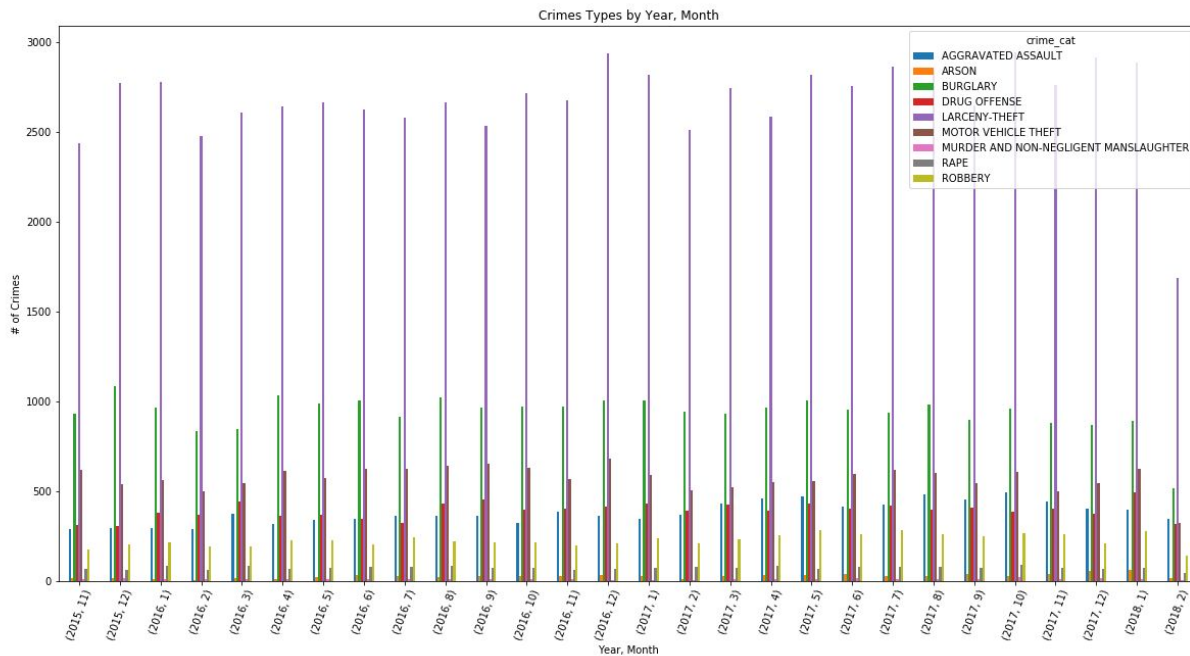
Exploratory Data Analysis

- Graphed data to find insights and correlations
 - # of reported incidents by year and month
 - Crime types by year and month
 - Top premises of incident
 - Top zip codes by incident
- Performed Market Basket Analysis to find associations between variables





Exploratory Data Analysis





Feature Engineering

- Created aggregation features
 - Zip Code and Crime Category
 - Crime Category, Hour Group and Zip Code
 - Crime Category, Premise Type and Zip Code
 - Crime Category, Season and Zip Code
 - Crime Category, Day of Week and Zip Code
- Features joined to training and test sets



Machine Learning

- 3 algorithms tested
 - Logistic Regression
 - Naive Bayes
 - Decision Tree
- Cross validation score performed for model selection
- Accuracy Score for model evaluation

	Logistic Regression	Naive Bayes	Decision Tree
CV Score Avg	0.6262	0.5971	0.6479
Accuracy Score	0.6619	0.6268	????

0.996

Decision Tree Accuracy Score



Conclusion and Next Steps

- Decision Tree accuracy score very high with possibility of overfitting
 - CV Scores for all the models did not indicate overfitting
 - Test data set worked best with Decision Tree algorithm
- Next Steps
 - Add demographic data to the dataset, provide more data to the model
 - Create additional features
 - Try different algorithms; Gradient Boosting, Random Forest
- Who Benefits?
 - City of Phoenix and Phoenix Police Dept. could use the model to focus patrols on particular areas, hire more officers or be proactive in areas where a crime type could occur



References

City of Phoenix Open Data. 2017. *Crime Data*. [ONLINE] Available at:
<https://phoenixopendata.com/dataset/crime-data>. [Accessed 27 February 2018].

Phoenix Population. (2018-06-03). Retrieved 2018-08-01, from
<http://worldpopulationreview.com/us-cities/phoenix/>