**Group**: Cindy ([cindy.reznikov@gmail.com](mailto:cindy.reznikov@gmail.com) 314/265-4315), James ([jamesa820@gmail.com](mailto:jamesa820@gmail.com) 314/971-7711), Robert ([rhanlon70@gmail.com](mailto:rhanlon70@gmail.com) / 314/303-7578)

**Topic:**  Predictive Analysis on Recidivism

**Data Source:** [**https://dev.socrata.com/foundry/data.iowa.gov/mw8r-vqy4**](https://dev.socrata.com/foundry/data.iowa.gov/mw8r-vqy4)(26,000 records)

Iowa Department of Corrections

**Project Plan:**

* Utilize Python Pandas and PySpark for data cleanup
* Utilize Scikit-Learn
* Utilize RFE for feature selection
* Machine Learning models used:
  + Logistic Regression
  + SVM
  + Naive Bayes
* Tableau for data visualization

**Features for Analysis:**

* Main Supervising District, Release Type, Race - Ethnicity, Age, Sex, Offense Classification, Offense Type, Offense Subtype, Return to Prison (y/n), Recidivism Type, New Offense Class, New Offense Type, New Offense Subtype, Target population

**Questions to answer:**

* Machine Learning Question: What is the chance of a released prisoner returning to prison?
  + Which features correspond to increased/decreased recidivism? In other words, does the type of crime committed tend to predict recidivism? Does gender play a role? Does age at release play a role?
* What is the difference in recidivism between the Supervising Districts? (Could imply differences in case management between districts -- could lessons in the district with least recidivism be taught in the others?)
* Ultimate goal: how do we reduce or eliminate recidivism?

**Visualizations:**

* Most common offense types for those who return to prison
* Male/Female comparison
* Race comparison
* District comparison
* Age comparison