College Readiness
Montgomery County, MD
Mesa, AZ

Station #3

The Bottom Line	City governments often don't have direct control over education. Focusing on data and analytics capacity is one way for cities to engage with education-related initiatives, especially as they may relate to other city services, such as public safety, parks and recreation, and neighborhood and community services. Cities should also consider applying this study's method in other areas where identifying risk is important to service provision.
Problem	Identifying an at-risk population for intervention before that population goes off track. In Mesa and Montgomery County, the problem is college matriculation. Some students don't graduate from high school and even those who do graduate often go to schools below their potential performance level. The goal is to use past data to identify which kids are at risk of not graduating or undermatching before they go off track and to provide interventions.
Leadership	School district leadership championed the project and ensured that technical staff had time and resources to work with the data and partners who completed modelling and analysis.
Staffing	District - From a few days to a few weeks of one person's time, depending on data readiness. Usually a Director of Data or IT related staff member. University - One or two researchers to create models and run analysis.
Data	Average of five years of historical, individual student data from two school districts, including student grades, attendance, and demographics.
Technology	PostgreSQL database for data and Python for analysis.
Methodology	Several standard machine learning models were applied to the student data to determine which were able to most accurately predict the top 5% of students at risk of going off track. Grades, attendance, and demographic information were used as predictors to determine whether a student could be "off track."
Results	Certain models ("Random Forests" being the most prominent) were better at predicting risk of going off track than others. The models also identified certain factors, such as absence rates, as better predictors than others.
Replication	Technology solutions are easy to replicate in this case, but data readiness is the key challenge. School districts (and cities) need to implement best practices in collecting, recording, and defining data to effectively replicate. Once data is ready, partnering with experts or acquiring machine learning expertise is the next step.
Learn More	Data Science for Social Good - dssg.uchicago.edu/projects