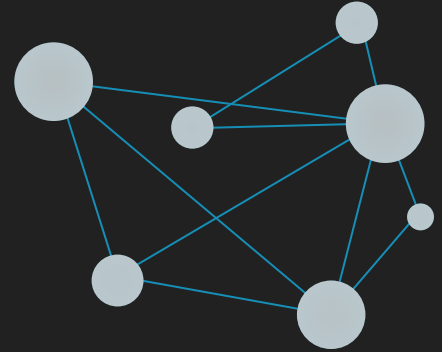




COLLECT DATA ON CRIME OUTCOMES

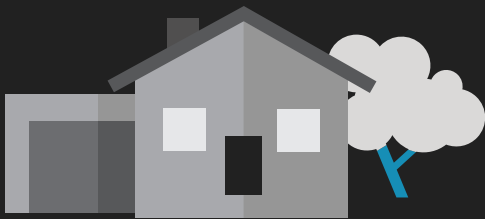
DATA ARE MESSY

Each source of crime data is inspected and analyzed to flag data-entry errors. To fix mistakes, we blend information from diverse sources to triangulate crime risk moving over time at the neighborhood scale.



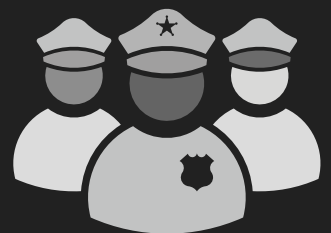
NOT EVERY CITY REPORTS CRIME IN THE SAME WAY

A bicycle theft is more likely to be reported in a small town than in a big city. Such differences complicate the comparison of crime rates in different places. To enable an apples-to-apples comparison, we combine data on crimes reported to police with national crime victimization surveys.



NOT EVERY POLICE AGENCY REPORTS CRIME IN THE SAME WAY

Crime data from police agencies can differ in surprising ways. For example, while most agencies report monthly, some report bi-monthly, or even less frequently. To estimate monthly crime risk in a consistent way over time, we use a wide variety of statistical methods to reconcile reporting differences across agencies.





IDENTIFY THE DETERMINANTS OF CRIME



CAUSES OF CRIME

Pinkerton leads a collaborative effort, combining the expertise and practical experience of law enforcement with leading criminology experts. The effort guides our selection of practical frameworks for determining the underlying causes of crime. An example is Routine Activities Theory, which recognizes that crimes occur when a motivated offender finds a suitable target in the absence of a capable guardian.

QUANTIFYING DETERMINANTS

To quantify the causes of crimes, we have built systems to collect a wide variety of continuously updated data that reflect important aspects of each concept. For example, a high percentage of crimes are committed by persons between the ages of 16-24, known as the age-crime curve. Thus, one dimension of our **Motivated Offenders** concept is captured by Census data on the age distribution within a community. Similarly, an important dimension of **Target Suitability** is captured by recent property value data as well as income and wealth data for residents. Formal **Guardianship** can be quantified using measures such as police or security force density. Informal guardianship can be quantified using variables related to housing tenure. This type of logic motivates our selection of hundreds of variables relevant for predicting crime.

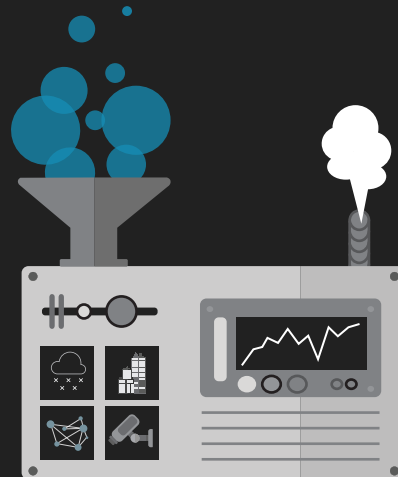




USE DATA SCIENCE TO PREDICT FUTURE CRIME

OUR ALGORITHMS

We use a wide variety of algorithms to ensure robust and accurate predictions. Time-series models detect seasonality and shifting trends, deep neural networks find complicated patterns among variables, and econometric models capture structural characteristics of place. By separately forecasting each crime type, in each county, we find models best suited to capture the unique dynamics of each community.



THE TOURNAMENT

With over 3100 counties and seven different crime types, our data scientists perform over 20 thousand forecasting problems every month. We pit our algorithms against one another in a double-elimination tournament. Winners are chosen based on the algorithm's ability to predict recent crime data that the algorithm has not yet seen.

INDEX ASSEMBLY

Our final predictions represent an **ensemble** (or averaging) of top-performing models. We also quantify the range of uncertainty surrounding every estimate. In the end, a single risk score summarizes total crime risk in each place, while separate indices summarize risk for property crime and violent crime. Each score reflects a community's risk relative to the national-average level.



THE PINKERTON CRIME INDEX

HOW WE BUILD OUR CRIME INDEX



STEP ONE



COLLECT DATA ON CRIME OUTCOMES

Our data collection systems continuously ingest the best available crime data from key sources, including FBI and local police. Data are cross-referenced to flag errors and establish a realistic view of crime in America.

IDENTIFY THE DETERMINANTS OF CRIME

Combining Pinkerton's grounded experience fighting crime with leading experts from criminology and policing, we have developed effective models to track the shifting determinants of crime risk over time.

STEP TWO



USE DATA SCIENCE TO PREDICT FUTURE CRIME

Our data scientists use a double elimination tournament to select the best-performing prediction algorithms, optimally honed for different crime categories and different community types.

STEP THREE



BUILD INDEX TO SUMMARIZE LOCAL RISK

The result is an index of overall crime risk at the neighborhood scale, along with separate measures for property and violent crime. Each index measures risk relative to the national-average level.

2.6x