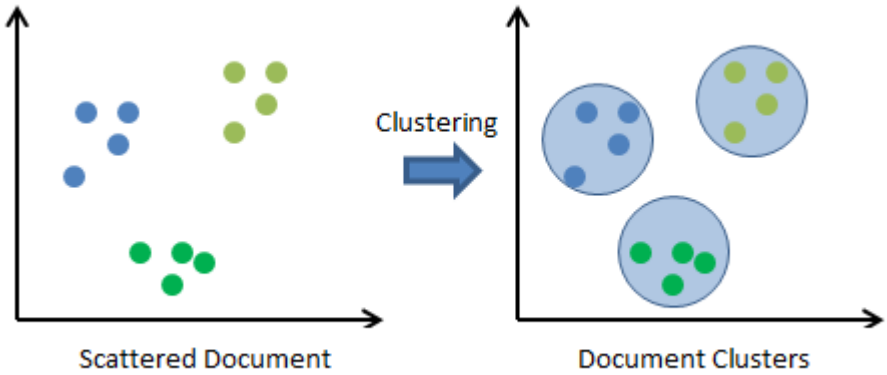
# 10 Interesting Use Cases for the K-means Algorithm

Below is the list of ten interesting use cases for k-means.

## 1. Document Classification

Cluster documents in multiple categories based on tags, topics, and the content of the document. This is a very standard classification problem and k-means is a highly suitable algorithm for this purpose. The initial processing of the documents is needed to represent each document as a vector and uses term frequency to identify commonly used terms that help classify the document. The document vectors are then clustered to help identify similarity in document groups.

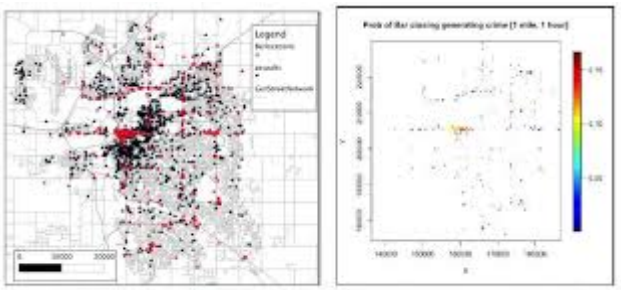



## 2. Delivery Store Optimization

Optimize the process of good delivery using truck drones by using a combination of k-means to find the optimal number of launch locations and a genetic algorithm to solve the truck route as a traveling salesman problem.


## 3. Identifying Crime Localities

With data related to crimes available in specific localities in a city, the category of crime, the area of the crime, and the association between the two can give quality insight into crime-prone areas within a city or a locality.



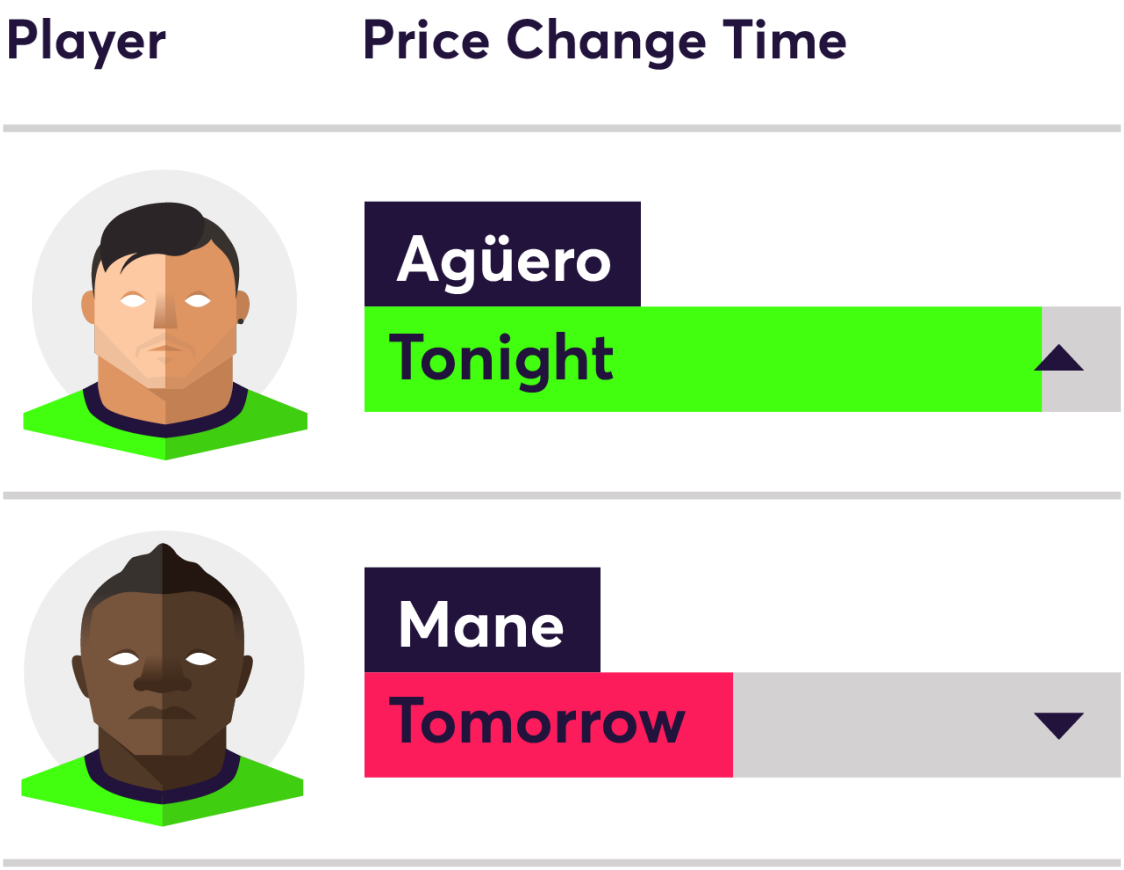
(a) Bar locations and Assault incidents in 2007 (b) Probability of Bars generating crime after bar closing

## 4. Customer Segmentation

Clustering helps marketers improve their customer base, work on target areas, and segment customers based on purchase history, interests, or activity monitoring. How telecom providers can cluster pre-paid customers to identify patterns in terms of money spent in recharging, sending SMS, and browsing the internet. The classification would help the company target specific clusters of customers for specific campaigns.

## 5. Fantasy League Stat Anlaysis

Analyzing player stats has always been a critical element of the sporting world, and with increasing competition, machine learning has a critical role to play here. As an interesting exercise, if you would like to create a fantasy draft team and like to identify similar players based on player stats, k-means can be a useful option.

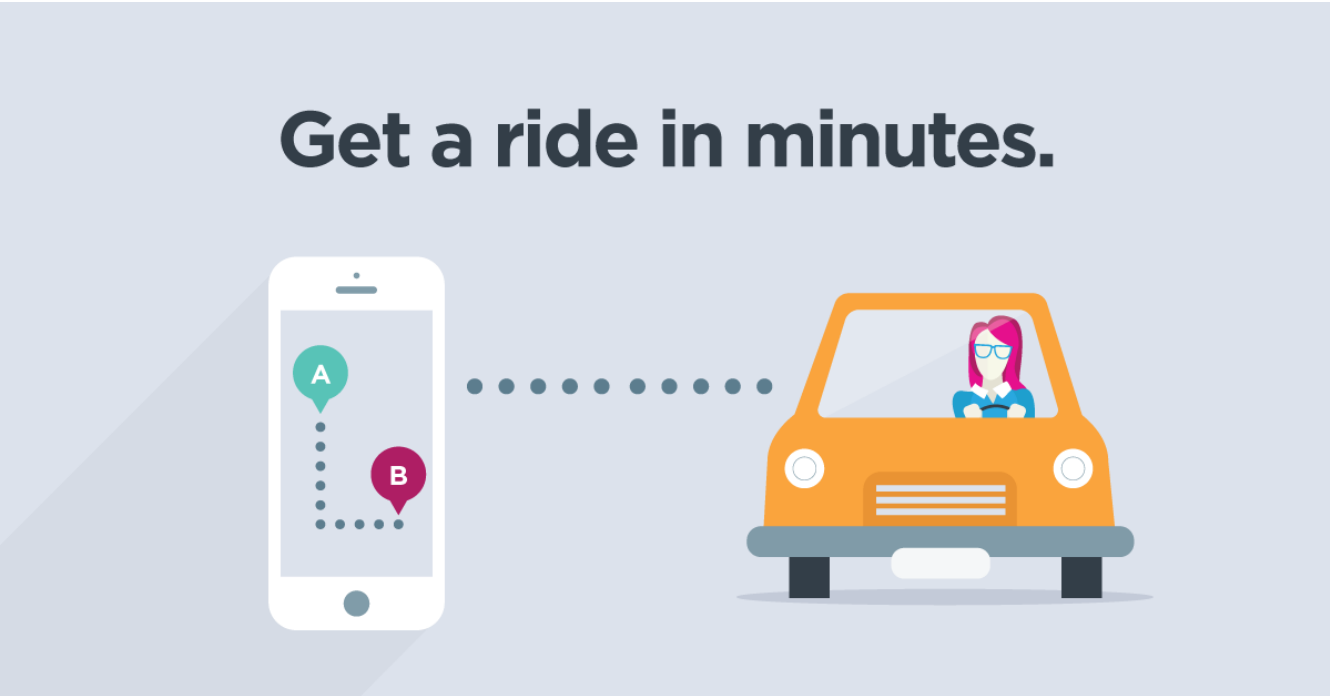



## 6. Insurance Fraud Detection

Machine learning has a critical role to play in fraud detection and has numerous applications in automobile, healthcare, and insurance fraud detection. Utilizing past historical data on fraudulent claims, it is possible to isolate new claims based on its proximity to clusters that indicate fraudulent patterns. Since insurance fraud can potentially have a multi-million dollar impact on a company, the ability to detect frauds is crucial.

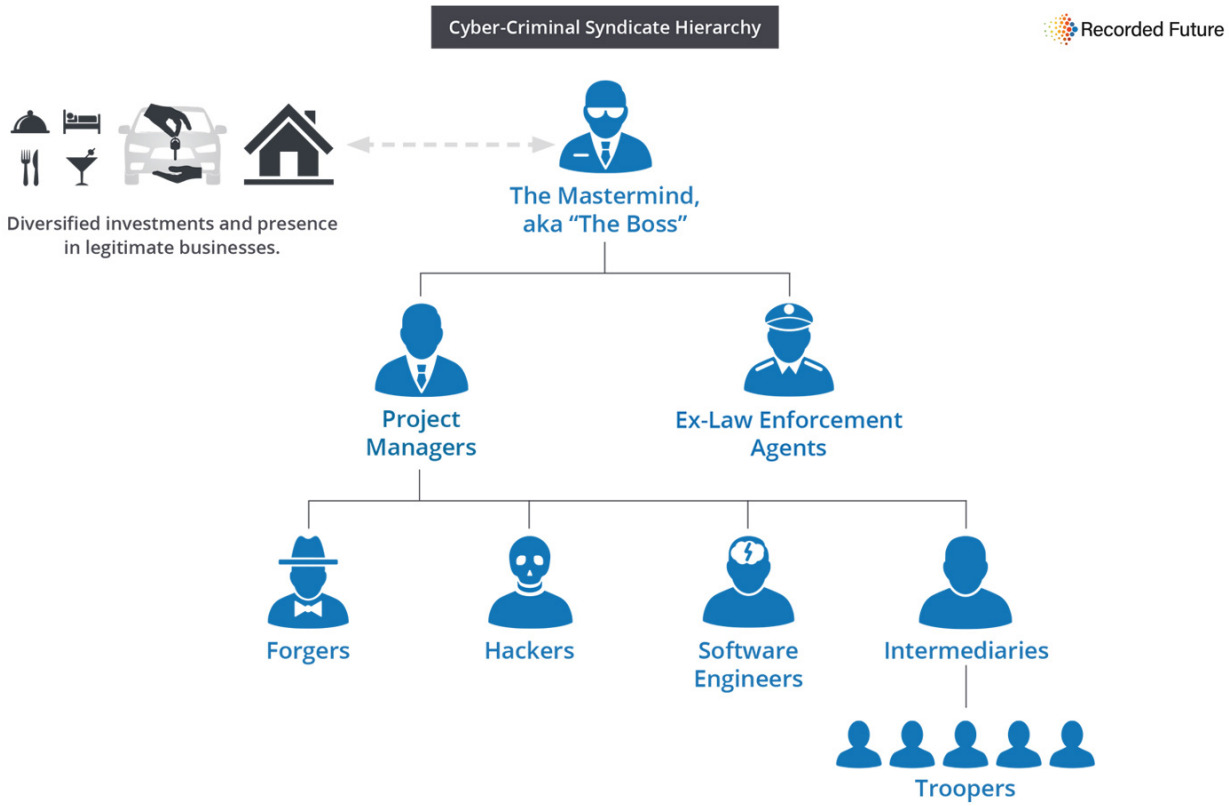

## 7. Rideshare Data Analysis

The publicly available Uber ride information dataset provides a large amount of valuable data around traffic, transit time, peak pickup localities, and more. Analyzing this data is useful not just in the context of Uber but also in providing insight into urban traffic patterns and helping us plan for the cities of the future.



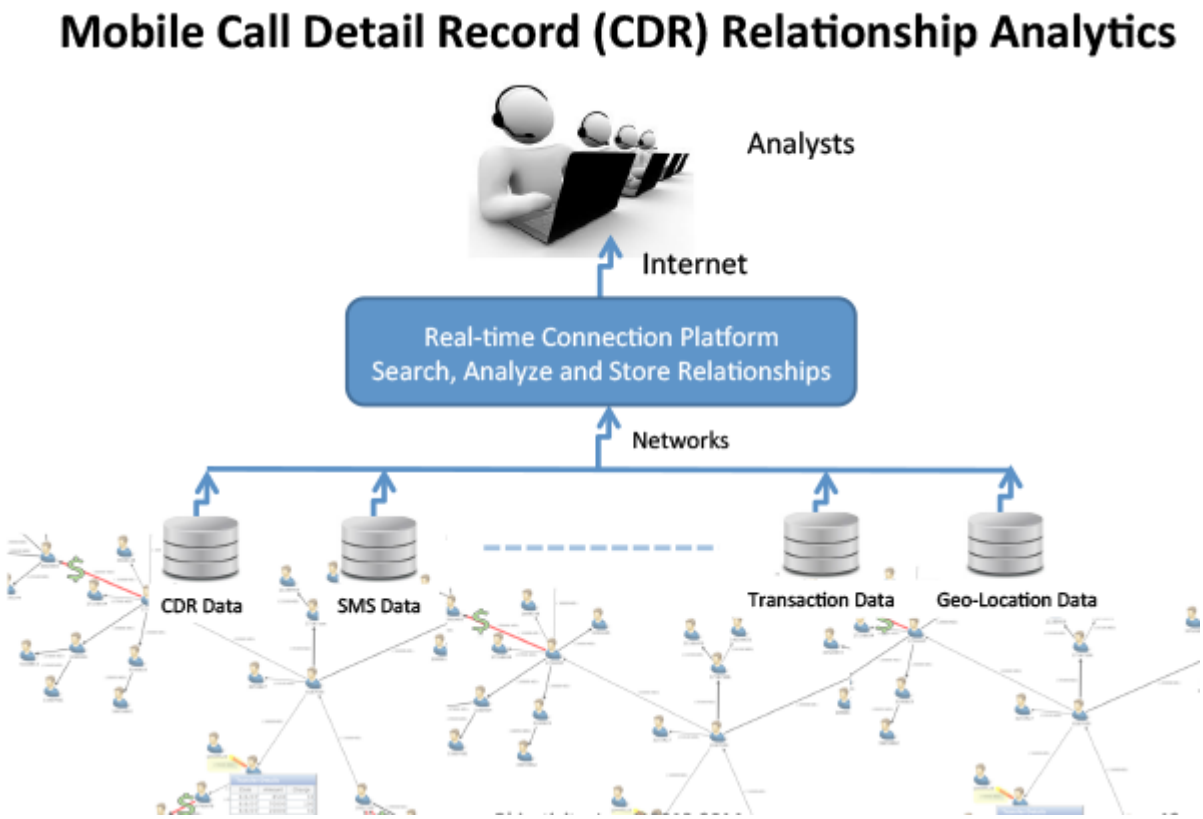

## 8. Cyber-Profiling Criminals

Cyber-profiling is the process of collecting data from individuals and groups to identify significant co-relations. The idea of cyber profiling is derived from criminal profiles, which provide information on the investigation division to classify the types of criminals who were at the crime scene.




## 9. Call record Detail Analysis

A call detail record (CDR) is the information captured by telecom companies during the call, SMS, and internet activity of a customer. This information provides greater insights about the customer's needs when used with customer demographics.



## 10. Automatic Clustering of IT Alerts

Large enterprise IT infrastructure technology components such as network, storage, or database generate large volumes of alert messages. Because alert messages potentially point to operational issues, they must be manually screened for prioritization for downstream processes. Clustering of data can provide insight into categories of alerts and mean time to repair, and help in failure predictions.