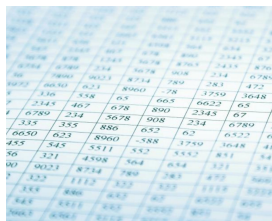


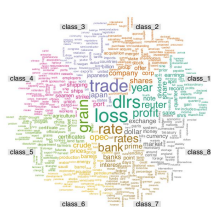
L'exploitation des données

La diversité des données

Données quantitatives



Données qualitatives



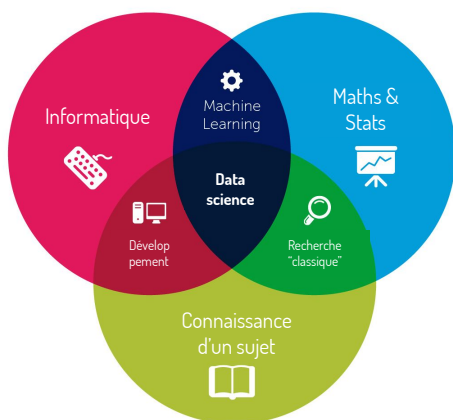
Données capturées



Données "crowdsourcées"



Le métier de data scientist



Copyright © 2014 by Steven Geringer Raleigh, NC.
Permission is granted to use, distribute, or modify this image,
provided that this copyright notice remains intact.

La science des données (en anglais *data science*) est une discipline qui s'appuie sur des outils mathématiques, de statistiques, d'informatique et mobilise des compétences spécifiques à une thématique.

Le premier objectif du « *data scientist* » est de produire des méthodes (automatisées, autant que possible) de tri et d'analyse de données de masse et de sources afin d'en extraire des informations utiles ou potentiellement utiles à des fins de **prise de décision, d'analyse ou de prédiction**.

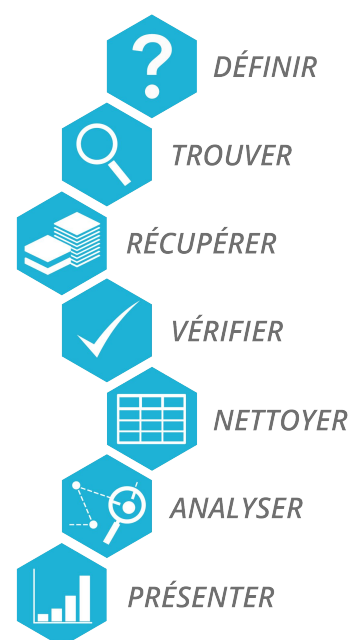
Le traitement des données

Selon ce modèle développé par School of Data, tout travail sur les données débute par la définition d'un problème.

Ensuite, il faut trouver les données, les récupérer et les vérifier pour identifier de potentiels problèmes.

Le nettoyage des données est la partie la plus difficile. On estime que **80% du temps d'un data scientist** est dédié au nettoyage et à la transformation des données.

Enfin, les données sont analysées et mise en forme grâce à la **data visualisation**.





DÉFINIR



TROUVER



RÉCUPÉRER



VÉRIFIER



NETTOYER



ANALYSER



PRÉSENTER