

Données de géolocasition pour savoir le nombre de personnes de tel pays visitant tel musée, vendre les données à la chambre de commerce.

2 types de projets : Data Science et Infra SI

1. Projet Data Science

- Marketer, campagne etc (définit le besoin et valide la pertinence des résultats)

- Développeur qui se charge de toute la partie Data Lake

- Data Scientist : construit le smodèles analytiques (remet en forme les data, donc plus le développeur a bien fait son taff plus le travail du Data Scientist est facilité)

- Data Analyst : 2 types :

- le MOA (se rapproche du métier et comprennent et ont la maîtrise du métier, il rédigent des spécifications et testent le produit)

- l’analyste BI : compétences en data vis (data visualisation) on fait parler les odnnées au travers d’images, un graphe etc

Compétences en SQL et data vis

Le Data Scientist : itératif, est-ce que le modèle est bon ? sinon il recommence en rajoutant des nouvelles variables etc

Comment il voit si le modèle est bon ou pas ?

En prenant les mêmes types de données pour les tests et l’apprentissage (si 10% de femmes en test alors que 90% en apprentissage pas bon)

ETL : extract transform load

ELT : extract load transform