

Statistique & Numérique en entreprise

À la découverte des métiers de la donnée



Ketsia Guichard

ketsia.guichard@gmail.com

Ingénieure en statistiques - Ensai 2015'

Data Scientist (Meetic, Vuitton, Leboncoin)

Enseignante en statistiques

Explosion des données disponibles



Le buzz autour des métiers de la donnée

Analytics And Data Science

Data Scientist: The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured
data. by Thomas H. Davenport and DJ Patil

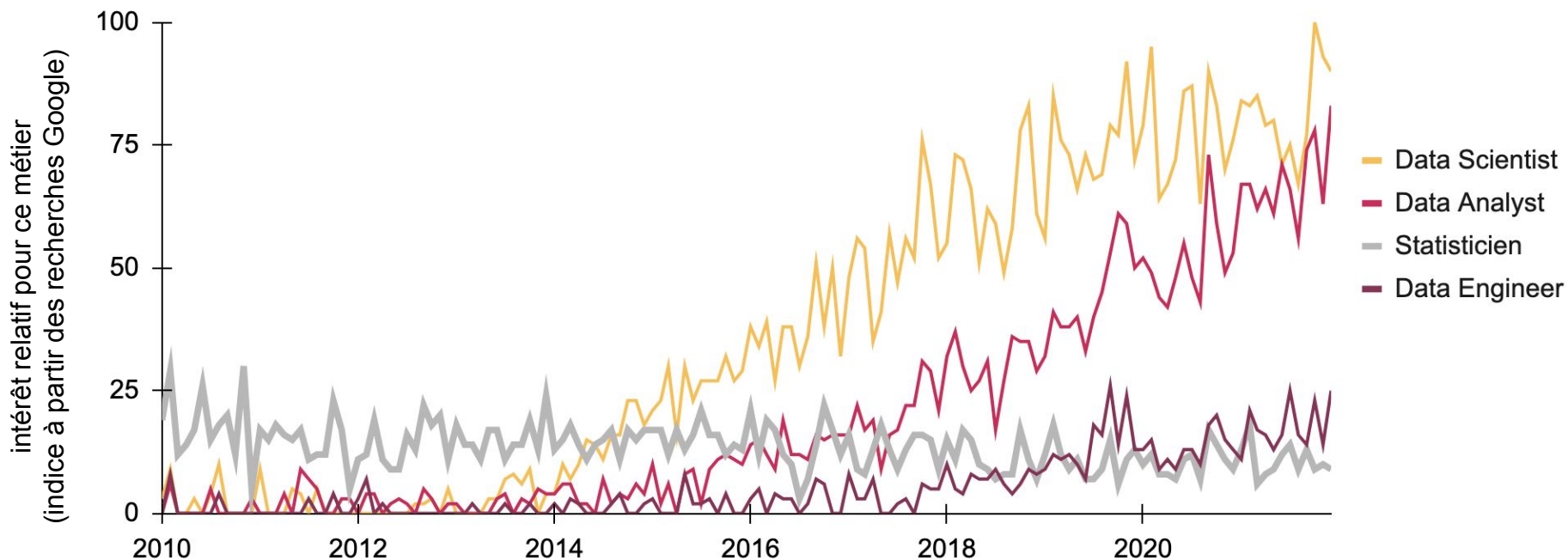
From the Magazine (October 2012)



Le buzz autour des métiers de la donnée

Explosion des métiers "Data"

Evolution des recherches Google liées aux métiers de la donnée (2010-2021)



Le buzz autour des métiers de la donnée



Stacby -
Data ninja



Stacby -
Data Wizard



Stacby -
Global Data Steward

Objectifs

- **clarifier** l'utilisation des données et des statistiques ;
 - Comment nos données personnelles sont-elles récoltées ?
 - Quelles méthodes statistiques sont employées ?
 - Quels sont les différents profils qui collaborent sur ce sujet ?
- faire connaître et **rendre accessible** les métiers liés à la donnée.

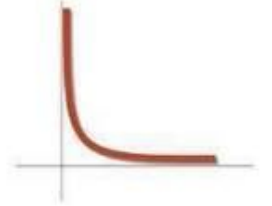
Au programme

1. **Cas pratique** d'utilisation de nos données et des statistiques
2. Récapitulatif des **métiers de la donnée**
3. Focus sur la **place des femmes** dans ces métiers

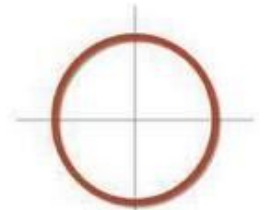
Cas pratique : les mathématiques de l'amour

ALL YOU NEED IS

$$y = \frac{1}{x}$$



$$x^2 + y^2 = 9$$



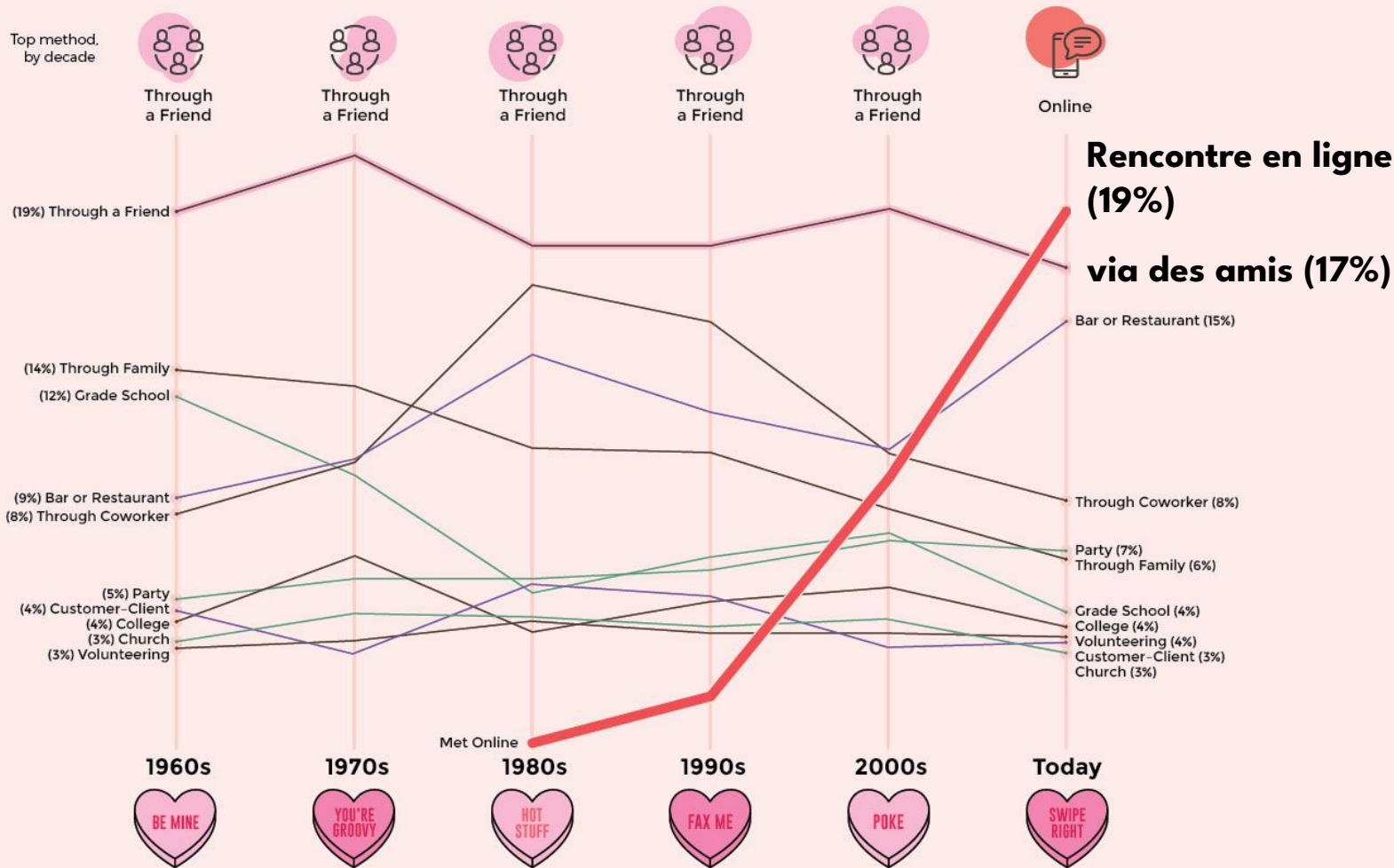
$$y = |-2x|$$



$$x = -3|\sin y|$$



Les sites de rencontres



Aperçu d'un site de rencontre

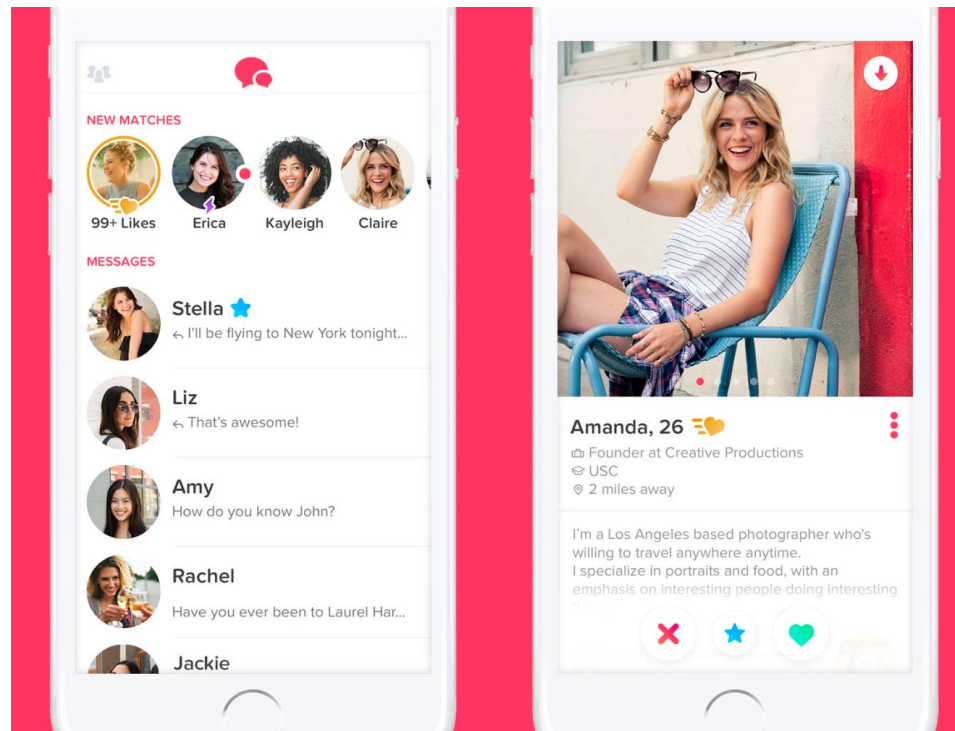
RÉGLAGES DE LA DÉCOUVERTE DE PROFILS

Distance maximale 21 km



Je recherche Tout le monde >

Tranche d'âge 18 - 32



Exemple de données

Profil des utilisateurs

RÉGLAGES DE LA DÉCOUVERTE DE PROFILS

Distance maximale 21 km

Je recherche Tout le monde >

Tranche d'âge 18 - 32

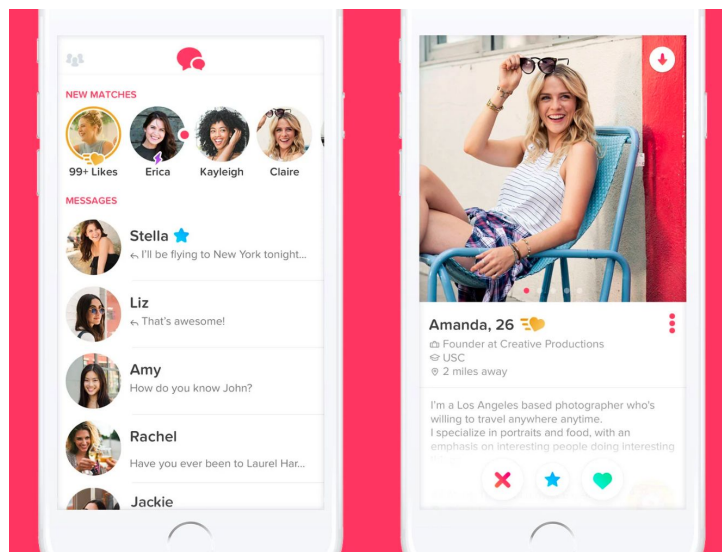
User	Genre	Âge	Cherche	Localisation	Photos	Distance max (km)	Age Min	Age Max
Ketsia	Femme	30	Homme Femme	Bruz Rennes Paris	4	21	18	32

+

ancienneté, fréquence des connexions, type de téléphone, mode d'acquisition, temps passé sur l'application, sommes dépensées, abonnement...

Exemple de données

Actions et échanges entre utilisateurs



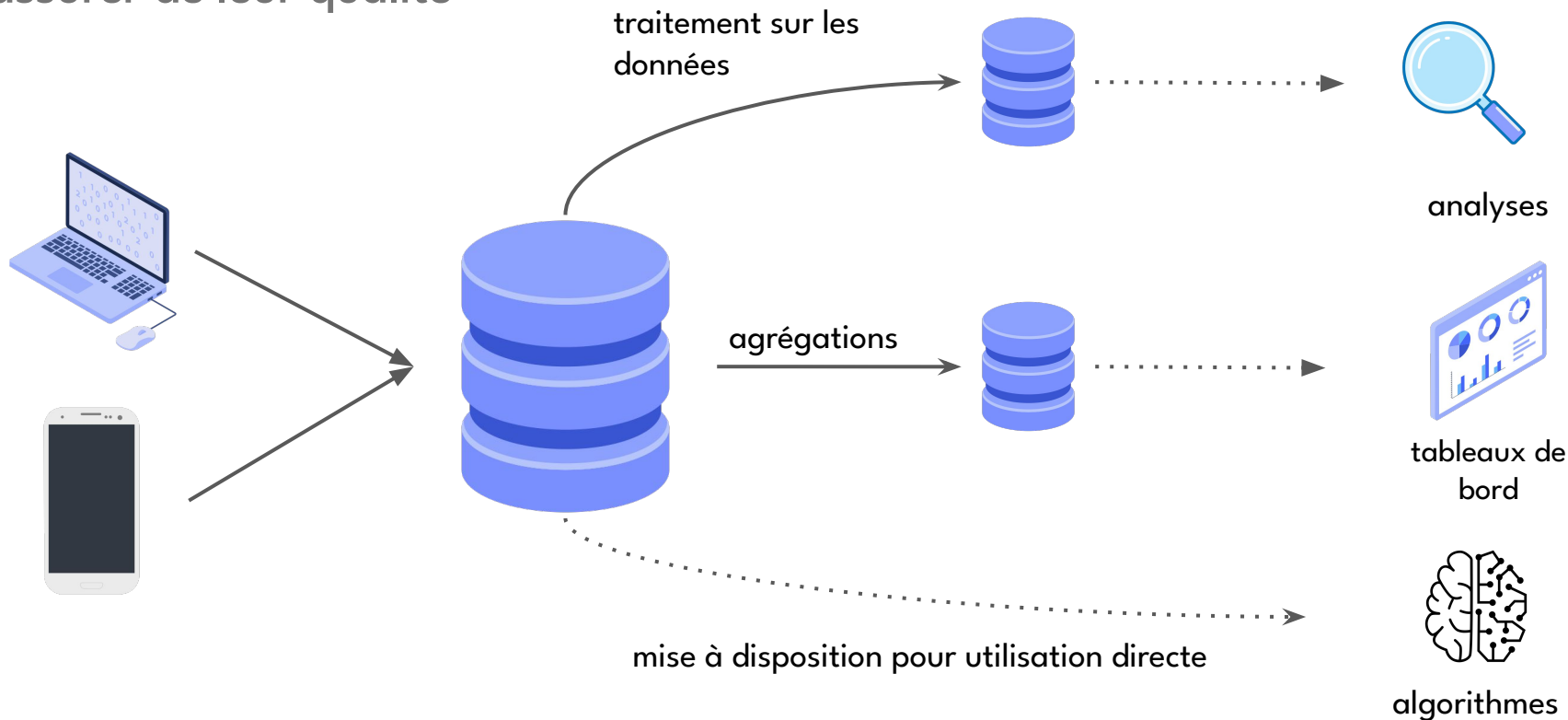
User	Profil	Choix	Match	Conver sation	A initié le dialogue	Nombre de messages
Ketsia	David	Nope	Non	-	-	-
Ketsia	Marc	Nope	Non	-	-	-
Ketsia	Amanda	Like	Oui	Non	-	-
Ketsia	Stella	Like	Oui	Oui	Non	2
Ketsia	Jacky	Like	Oui	Oui	Oui	30

+

attractivité du profil (taux de like), taux de réponse, temps de réponse, longueur des messages envoyés...

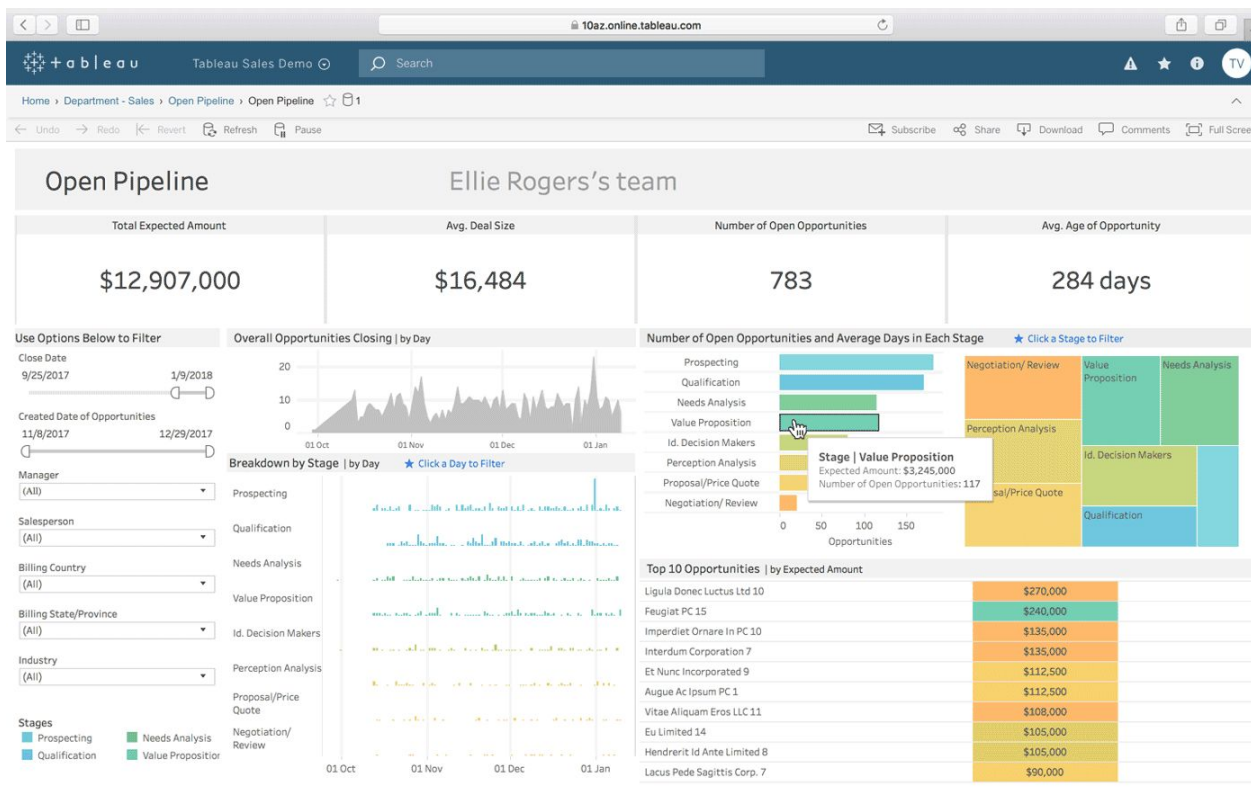
Stocker les données

Et s'assurer de leur qualité



Tirer de la valeur des données

Suivi business



Tirer de la valeur des données

Les algorithmes, au support des différents métiers de l'entreprise

- **Marketing** : segmentation des utilisateurs → *statistiques descriptives*
- **Finance** : prédiction de chiffres d'affaires → *séries temporelles*
- **Relation client** : offres spéciales d'abonnement → *score d'appétence*
- **Produit** : recommandation de profils → *algorithmes de recommandation*
- **Modération** : aide à la modération

Tirer de la valeur des données

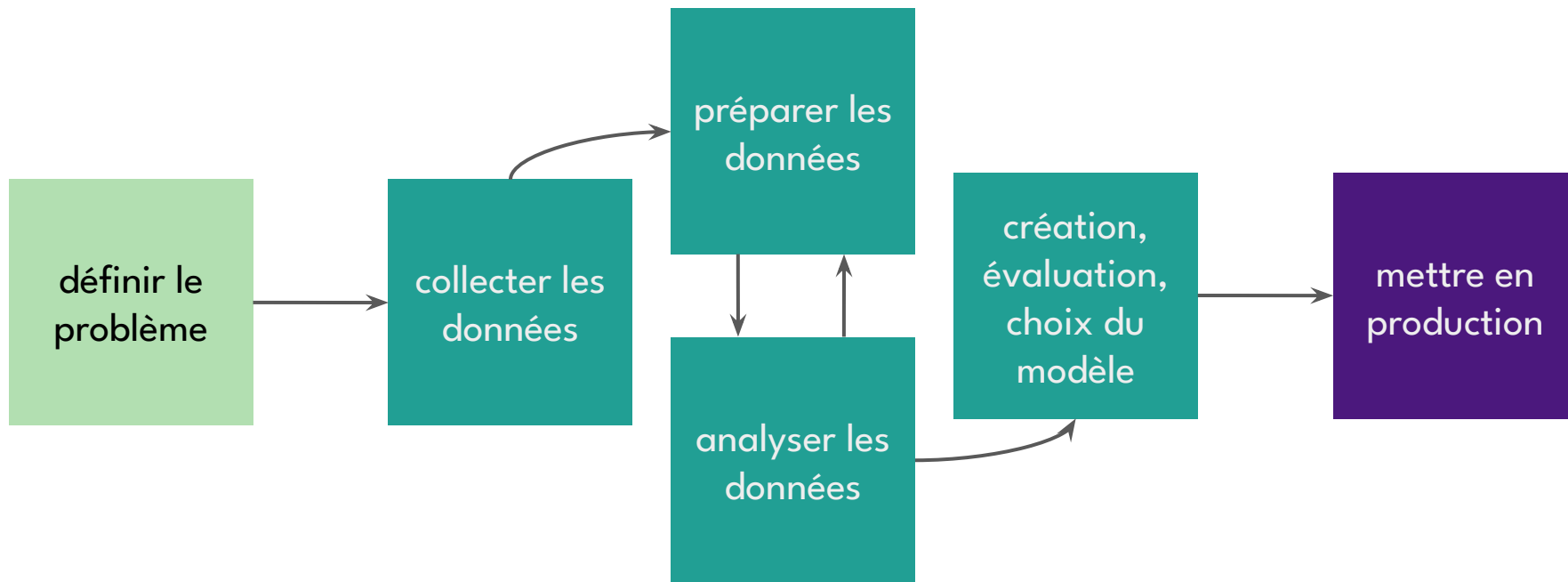
Aider la modération

Les différents rôles de la modération :

- repérer les faux profils → *score de risque*
- modérer les messages envoyés → *aide à la définition des mots clés*
- appliquer les règles du site, y compris **l'envoi de photos dénudées** (particulièrement quand elles sont non consenties)

Tirer de la valeur des données

Processus de création d'un algorithme d'apprentissage automatique



besoin

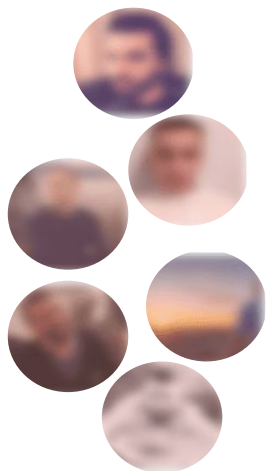
modélisation

automatisation

Repérer automatiquement les images à censurer

Deep Learning & Analyses d'images

Constitution des jeux de données (apprentissage/test)



Photos de profil



Photos (légitimes) échangées
entre utilisateurs



Photos à censurer (dick pics)
annotées par la modération

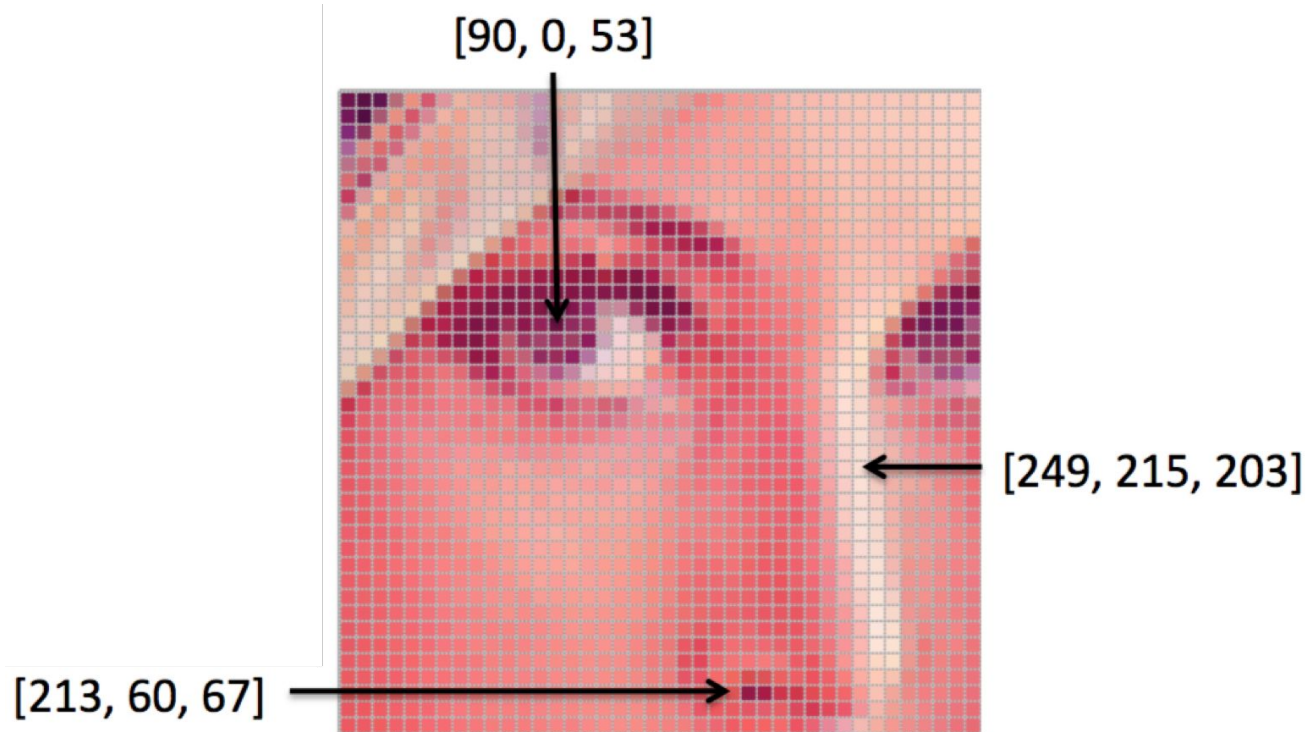
Repérer automatiquement les images à censurer

Mais... une image c'est des données ?

R = 255
V = 0
B = 0

R = 0
V = 255
B = 0

R = 0
V = 255
B = 0



Repérer automatiquement les images à censurer

Deep Learning & Analyses d'images



Jeu d'apprentissage

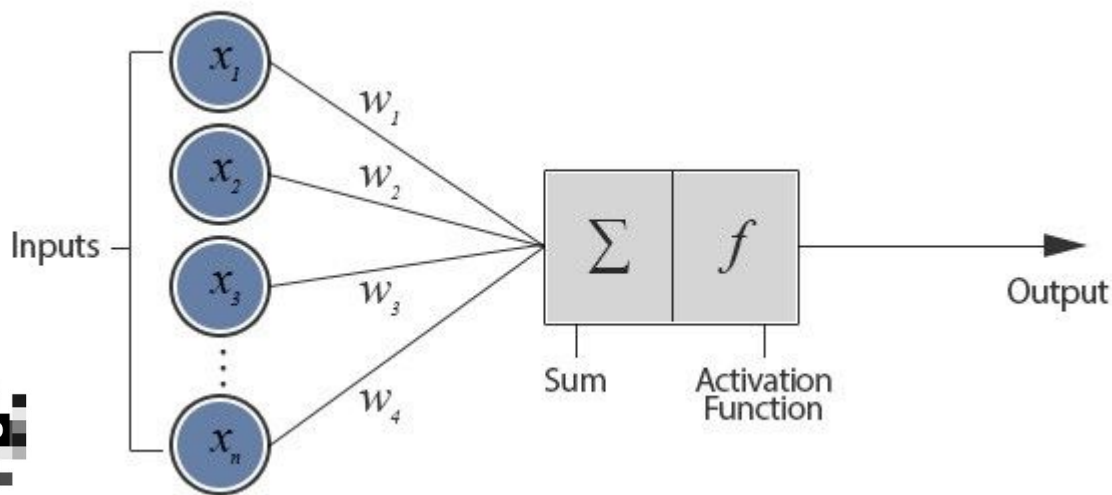
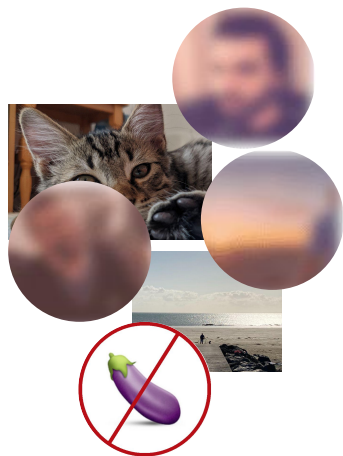


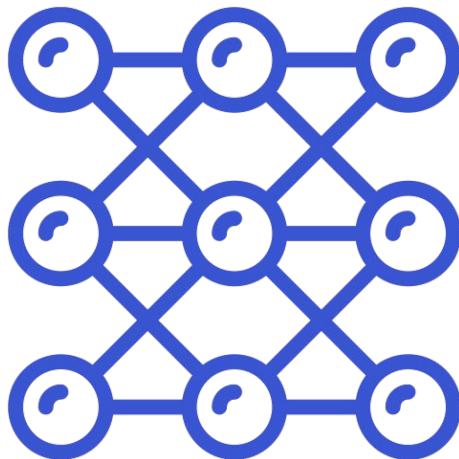
Image à
modérer
ou non
(score)

Repérer automatiquement les images à censurer

Evaluation des résultats



Jeu de test
(nouvelles données)



Evaluation :

- taux de biens classés (*accuracy*)
- taux d'images effectivement à censurer parmi celles prédit comme telles (*précision*)
- taux d'images prédits comme à censurer parmi toutes celles effectivement à censurer (*rappel*)
- ...

Références pour jouer avec les données de l'amour

Les mathématiques de l'amour, Hanna Fry, 2015

contient de nombreux exemples d'utilisations des mathématiques et des statistiques sur le thème de l'amour : calculer son nombre de partenaires potentiels, comment optimiser le placement des invités à un mariage, etc. Très agréable à lire, drôle et donnant beaucoup d'idées de petits contenus pédagogiques !

Jeux de données (Kaggle)

- [Speed Dating Experiment](#), données issues d'une expérience de psychologie sociale
- [OkCupid Profiles](#), 60 000 profils OKCupid anonymisés !

Jouer avec les modèles ou comprendre les algorithmes

- [Bumble Private Detector](#)
- [Finding Love on a First Date: Matching Algorithms in Online Dating](#), Harvard Data Science Review



Récapitulatif : Les métiers de la donnée

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21st century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing packages, e.g., R
- ☆ Databases: SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau



BUSINESS PROCESS OWNER

Il est le garant d'un processus et donc le propriétaire des données sous-jacentes.



CHIEF DATA OFFICER

Il est le leader de la transformation autour de la donnée. Il organise la chaîne de création de valeur par la donnée. Il définit les politiques applicables, le modèle de gouvernance des données et l'organisation cible. Il dirige les instances décisionnelles sur la Data.



DATA MANAGER

Il définit les données clé et est garant opérationnel de la qualité des données pour son processus.



DATA STEWARD

Il élabore le référentiel de données et s'assure que le catalogue des données soit complet et régulièrement enrichi.



CRÉER DE LA VALEUR
GRÂCE AUX DONNÉES



DATA SCIENTIST

Il exploite différentes disciplines scientifiques (mathématique, statistique, informatique) pour extraire des connaissances et des idées de nombreuses données (structurées ou non).



DATA ARCHITECT

Il identifie où sont les données ciblées. Il élabore la cartographie des flux de données (entre les différents systèmes informatiques).



DATA ENGINEER

Il collecte, stocke, structure et connecte les données sources aux applications qui seront utilisées en garantissant un fonctionnement fluide.

FOURNIR
DES ÉCLAIRAGES
AUX MÉTIERS/BUSINESS



DATA ANALYST

Il extrait, met en forme les données en utilisant des outils informatiques dédiés pour la réalisation d'outils décisionnels.

DISPOSER DE DONNÉES
ACCESSIBLES
ET EXPLOITABLES

BUSINESS PROCESS OWNER

Il est le garant d'un processus et donc le propriétaire des données sous-jacentes.



CHIEF DATA OFFICER

Il est le leader de la transformation autour de la donnée. Il organise la chaîne de création de valeur par la donnée. Il définit les politiques applicables, le modèle de gouvernance des données et l'organisation cible. Il dirige les instances décisionnelles sur la Data.



DATA MANAGER

Il définit les données clé et est garant opérationnel de la qualité des données pour son processus.



DATA STEWARD

Il élabore le référentiel de données et s'assure que le catalogue des données soit complet et régulièrement enrichi.



CRÉER DE LA VALEUR
GRÂCE AUX DONNÉES

DATA ARCHITECT

Il identifie où sont les données ciblées. Il élabore la cartographie des flux de données (entre les différents systèmes informatiques).



DISPOSER DE DONNÉES
ACCESSIBLES
ET EXPLOITABLES

DATA ENGINEER

Il collecte, stocke, structure et connecte les données sources aux applications qui seront utilisées en garantissant un fonctionnement fluide.



FOURNIR
DES ÉCLAIRAGES
AUX MÉTIERS/BUSINESS

DATA SCIENTIST

Il exploite différentes disciplines scientifiques (mathématique, statistique, informatique) pour extraire des connaissances et des idées de nombreuses données (structurées ou non).



DATA ANALYST

Il extrait, met en forme les données en utilisant des outils informatiques dédiés pour la réalisation d'outils décisionnels.



Data Analyst

Activités principales

- structurer les données
 - ◆ définir les règles de nettoyage des bases
 - ◆ maîtriser la qualité des données
- analyse et exploration de données
 - ◆ études
 - ◆ réaliser des tableaux de bord
 - ◆ visualisation de données
- communiquer avec les autres équipes
 - ◆ traduire les besoins de manière analytique
 - ◆ présenter les résultats

Compétences techniques

- requêtage de bases de données
- connaissance des tests et méthodes statistiques
- langages d'analyse de donnée (Python, R)
- outils de data visualisation (Tableau, PowerBI...)

Diplômes

- Bac+3 en statistique, traitement de l'information
- Bac+5 en statistiques, Big Data, parfois Marketing ou école de commerce

Data Scientist

Activités principales

- structurer les données
 - ◆ définir les règles de nettoyage des bases
 - ◆ maîtriser la qualité des données
- analyse et exploration de données
- élaborer les algorithmes
 - ◆ construire les données d'entraînement
 - ◆ créer et tester des algorithmes d'apprentissage automatique
 - ◆ amélioration continue des modèles
- industrialiser les algorithmes

Compétences techniques

- requêtage de bases de données
- langages d'analyse de donnée (Python, R)
- très bonnes connaissances en statistique
- expertise en méthodes d'apprentissage automatique (Machine Learning, Deep Learning) et environnements de développements associés (TensorFlow, PyTorch, Keras...)

Diplômes

- Bac+5 en informatique, Data science ou statistique
- Bac+8 : doctorat en statistique ou informatique

Data Engineer

Ingénieur·e data/big data, Développeur·euse data

Activités principales

- acheminer la donnée
 - ◆ collecter la donnée
 - ◆ choisir les solutions de stockage
 - ◆ développer l'infrastructure de donnée

- mettre à disposition la donnée
 - ◆ automatiser le nettoyage des données
 - ◆ maintenir et documenter les bases

- industrialiser les algorithmes

Compétences techniques

- gestion de bases de données
- langages de programmation (Scala, Python, Java...)
- connaissances de base en statistiques et intelligence artificielle

Diplômes

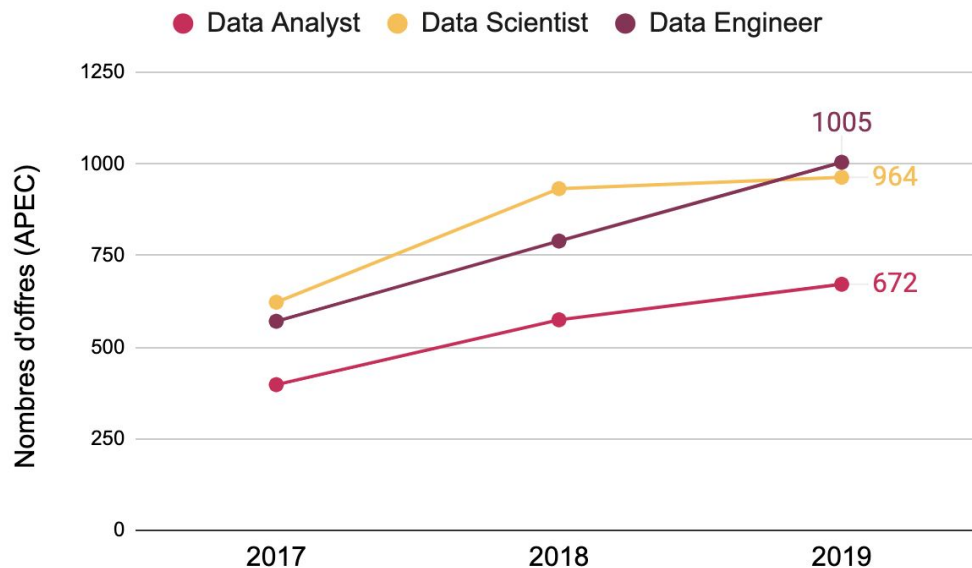
- Bac+5 en informatique, Data science ou statistique

- Bac+2 en statistique ou informatique avec une expérience dans le traitement des données

Tendances du marché

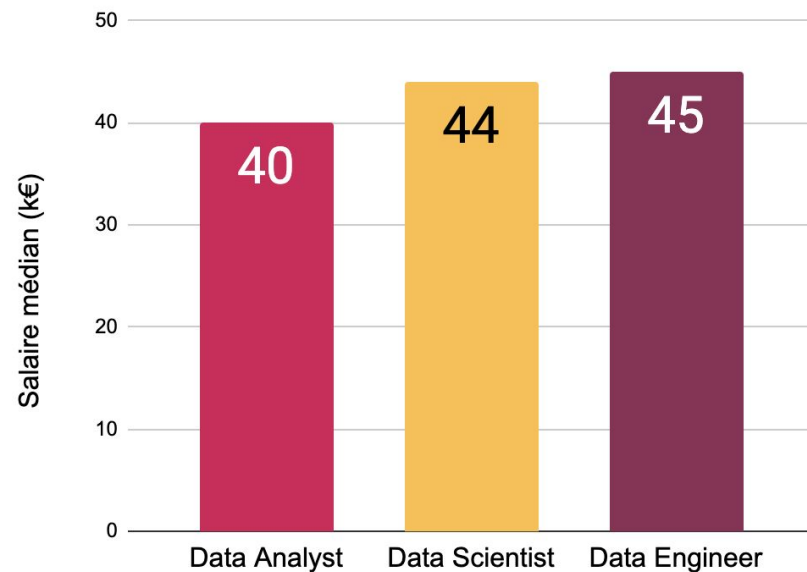
Evolution du nombre d'offres

Données APEC (2019)



Salaire médian (2019)

Données APEC



Ressources

- [Les métiers de la Data](#), APEC
- [BD Data, la Data en 10 métiers](#), Syntec Conseil
- [Le futur des métiers de la Data](#), Syntec Conseil
- [Le marché du travail numérique en France](#), Indeed

Focus :

La place des femmes dans les métiers de la donnée



Etat des lieux : les femmes dans la tech

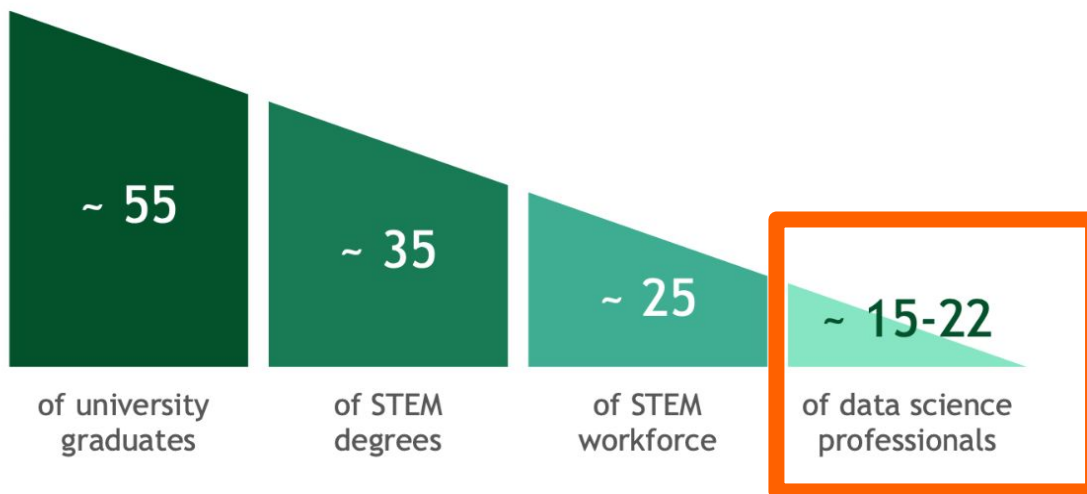
Selon le rapport DESI (Digital Economy and Society Index) de la Commission européenne,

- 19% des spécialistes en TIC* sont des femmes en 2022 (21% en France) ;
**TIC = Technologies de l'information et de la communication*
- Ce taux progresse, mais très lentement (+2.1pts en Europe, +3,3pts en France en 5 ans) ;
- Le programme “Path to the digital decade” a donc ajouté cette année comme objectif **la convergence des genres** pour les spécialistes en TIC.

Etat des lieux : les femmes dans la data

Share of women at each stage of the STEM talent funnel (in %)

Women make up ...

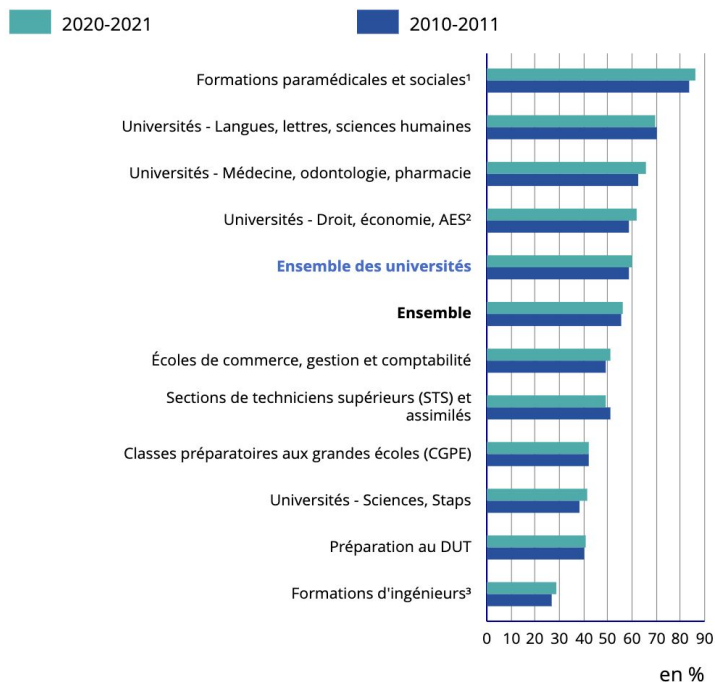


Quelques causes selon BCG :

- la data science a une image de secteur théorique, très “geek”;
- réputation de compétitivité ;
- manque d'exemple de réels impacts pour les entreprises et la société.

Etat des lieux : les femmes dans les études scientifiques

Figure 1 - Part des femmes dans les formations d'enseignement supérieur



L'INSEE constate une spécialisation des genres selon le type de cursus et les matières étudiées

“Au sein des disciplines scientifiques à l'université, seulement 23 % des étudiants en DUT dans les spécialités de production et d'informatique et 31 % des étudiants en sciences fondamentales et applications sont des femmes.

[...]

*Ces écarts se sont **peu réduits** durant la dernière décennie”*

L'orientation des lycéen·ne·s

Selon l'Observatoire sur la féminisation des métiers du numérique (Ipsos) les lycéennes ne se dirigent pas vers le numérique car :

- sous-estimation des compétences dans ces matières ;
- manque de représentation féminines ;
- crainte d'être désavantagées professionnellement dans ces voies ;
- parents moins encourageants vers ces voies.

94%

des lycéens pensent qu'il est important voire indispensable d'avoir un très bon niveau dans les matières scientifiques pour être admis et réussir dans une école d'informatique.

Or même lorsqu'elles ont plus de **14/20** de moyenne dans les matières scientifiques, les filles sont bien moins nombreuses que les garçons à penser avoir le niveau pour suivre une formation en école d'informatique (**43%** contre **78%**), et leurs parents sont du même avis.

37%

des lycéennes envisagent de s'orienter vers une école d'informatique ou une école d'ingénieur, contre **66%** des garçons. Pourtant **56%** des lycéennes sont intéressées par l'informatique / le numérique.

Pistes d'actions dès le secondaire

- Orientation dès la seconde pour éviter la disparition progressive des filles dans les matières scientifiques ;
- Mettre les filles au contact de l'informatique ;
- Permettre de rencontrer des femmes scientifiques servant de modèles d'identification.

Selon Marion Monnet, chercheuse à l'INED, *"les filles qui ont bénéficié d'intervention de ce type au lycée ont 30% de chance en plus d'intégrer une prépa scientifique"*.



Pistes d'actions dès le secondaire

*“**Une attitude soutenante** d'un point de vue social de la part de l'enseignant **n'est pas un élément particulièrement favorable** contrairement à un soutien davantage axé sur la relation académique (via les feed-back).*

*Se montrer attentif aux progrès et aux difficultés des élèves, **leur indiquer comment se corriger et s'améliorer** sont des comportements qui, s'ils sont perçus par les filles, **renforcent sensiblement la vision qu'elles ont de l'utilité des mathématiques et leur intérêt pour cette discipline.**”*

Doriane Jaegers, Dominique Lafontaine

Aspirer à une carrière mathématique :
quel rôle jouent le soutien et les attentes de l'enseignant chez les filles et les garçons ?

Ressources

Place des femmes dans la tech/la data

- [Rapports DESI \(Europe/France\)](#)
- [Women in Tech Statistics Show the Industry Has a Long Way to Go](#), BuiltIn
- [RESETTING TECH CULTURE : 5 strategies to keep women in tech](#), GirlsWhoCode x Accenture
- [What's Keeping Women out of Data Science?](#), BCG Gamma

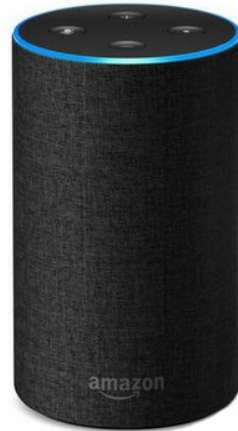
Part des femmes dans les études scientifiques

- [Femmes et hommes, l'égalité en question \(édition 2022\)](#), INSEE
- [La disparition des filles dans les études d'informatique : les conséquences d'un changement de représentation](#), Isabelle Collet, Carrefours de l'éducation 2004
- [Étudiantes en filières scientifiques : où en sommes-nous ?](#), L'Étudiant

Orientation dans le secondaire

- [Observatoire sur la féminisation des métiers du numérique](#), Ipsos x Epitech
- [Réforme du lycée : quand les filles désertent les maths, les prépas scientifiques en pâtissent](#), l'Étudiant
- [Do boys and girls differ in their attitudes towards school and learning?](#), PISA 2018
- [Les freins à l'accès des filles aux filières informatiques et numériques du lycée](#), Centre Hubertine Auclert

Les enjeux de la diversité



Les enjeux de la diversité

“develop the advanced technical skills of women and girls so they can steer the creation of frontier technologies alongside men”

UNESCO, rapport EQUALS

Merci !

Des questions ?