

Compétences clés

- Autonomie
- Polyvalence
- Formateur R et Python
- Veille technologique

Objectif

Après 7 années comme consultant Senior, je veux m'investir dans des projets de plus long terme, au sein d'une équipe qui pourra s'appuyer sur mon expérience.

Expérience professionnelle

Intégralité des missions : <https://www.linkedin.com/in/sebastienfoulle>

2021, débutée fin août – Mission dans une enseigne de Grands Magasins, équipe Insights : datascientist

- Etudes ponctuelles de CA, panier moyen, portefeuille actif/encarté, ... avec Big Query
 - Dashboard Data Studio sur le portefeuille de détenteurs de la carte de paiement
- Environnement technique** : Big Query, Google (Mail, Drive, Slides, Sheets, GCP), Python, Jira, Confluence

2021, 20 demi-journées – Hackathon sur la Google Cloud Platform pour un client dans l'Energie

- Choix de 3 machines méga-mémoire GCP qui ont tenu la charge (5G de données, 60 équipes de trois).
 - Support technique des équipes : gestion des environnements Python, commits et pushes sur Gitlab, ...
- Environnement technique** : GCP, JupyterLab, Python, Rstudio, R, Gitlab, Slack

2021, 3 mois – Mission dans l'Energie, Départ^t Sourcing Offre Electricité : industrialiser les reportings Excel

- Création d'un schéma, de 12 tables SQL et de 16 scripts R qui produisent automatiquement 2 des reportings. On économise une semaine de travail à chaque enchère (au moins 6 par an).
 - Création d'un rapport Power BI qui permet de suivre l'évolution des prévisions mensuelles des 3 dernières années. Trois présentations à l'équipe, retours très positifs des participants et des managers.
- Environnement technique** : PowerBI Report Server, Gitlab, Rstudio, Oracle, PostgreSQL, PgAdmin

2018 à 2020, 26 mois – Mission dans le Jeu en Ligne, BU Sport : datascientist

- Automatisation de l'alimentation des tables SAS : traitements lancés à 5h, le datamart est utilisable dès lundi 9h. Le client satisfait a prolongé ma mission de 3 mois à 2 ans et a recruté 2 collègues consultants.
- 9 vues SQL créées pour accélérer la conception des rapports Power BI : plus de 20 publiés en 9 mois.
- Lutte contre la fraude : rapport Power BI pour analyser des fraudes passées et en déduire des seuils d'alerte. En moyenne 20 alertes par jour lancées en quasi-temps réel, retours très positifs des traders.
- Segmentation mixte des nouveaux joueurs sous Python : 11 joueurs du segment « rentables et très gros miseurs » ont été remontés à la responsable du programme « VIP ».
- Détection des potentiels VIP par arbre de régression avec *scikit-learn* + seuil. Le lift vaut 140, la moyenne de 7 candidats par semaine est traitable par l'équipe VIP. Présentation au N+4 avant mise en production
- Mise en forme HTML des reportings journaliers : passage d'une simple phrase contenant 1 ou 2 chiffres à des mails structurés, contenant 4 à 5 tableaux de KPI et des graphiques.

Environnement technique : SQL Teradata, Python, Power BI Pro, SAS Guide, macros Excel, Trello

2017, 4 mois – Mission dans l'Assurance, Direction Connaissance Client : scorer la conversion prospect/client

- Calculs distribués avec le package R *sparklyr*. REX de 11 pages sur les bonnes pratiques Spark.
- Scoring sur 700 champs par forêt aléatoire. Calcul de l'importance des variables puis réduction aux 40 champs les plus importants pour accélérer les calculs, sans perte de performance du score.

Environnement technique : Cluster Spark, Rstudio, R

2017, 30 ½-journées – Mission dans l'Assurance Vie, Direction Connaissance Client : analyse de parcours Web

- Audit de la qualité des données de Google Analytics Premium : 2 erreurs d'alimentation des bases ont multiplié par 5 (resp. 80) les volumétries de 2 journées, problème vite corrigé une fois remonté.
- Identification par expressions régulières en SQL de 19 actions et 22 produits dans les URL.
- Mise en production d'un dashboard interactif réalisé avec le package R *flexdashboard*.

Environnement technique : Hue, Hive, Impala, Rstudio, R

2015-2016, 8 mois – Mission dans l'Energie : migration de SAS vers R

script SAS qui s'exécute en 6 heures : les requêtes SQL sous R sont plus lentes, mais les packages *data.table* et *Rcpp* permettent d'accélérer les jointures et le script R tourne aussi en 6 heures.

Environnement technique : SAS, Rstudio, R

Depuis 2014 – AI&Data (conseil en marketing), AI&Data Academy : veille et animation d'équipe

- Formations créées et animées, seul ou en binôme, chez 5 clients et en interne.
 - ✓ En Python : modélisation des parcours clients par RNN avec *Keras*, NLP sur des tweets avec *spaCy*.
 - ✓ En R : calculs sur grosses volumétries avec *data.table*, faire un *ShinyDashboard*.
- Veille quotidienne Python et R sur 17 sites anglo-saxons, newsletter interne hebdomadaire.
- Cartographie des carreaux de l'INSEE en binôme : parallélisation pour passer de 48h à 1h de calculs, présentation aux rencontres R <https://r2016-toulouse.sciencesconf.org/106664.html>.
- Alimentation du Wiki de Redmine, rédaction de 6 articles du blog en 18 mois. Documentation html des bonnes pratiques : 13 pages sur le serveur Rstudio, 25 pages sur la programmation.
- Encadrement d'un stagiaire en 2016 : benchmark de ML, XGboost présenté à la BU Datascience.

Environnement technique : JupyterLab, Python, Rstudio, R, Rmarkdown, Shiny, Gitlab, Teams

2010 à 2014 – BCA Expertise (expertise automobile), BU Datamining : dataminer en interne

- Gestion du projet « acquisition outil BI » sur 8 mois. Délais et budget respectés. Présentation d'une maquette aux directeurs généraux avant déploiement de l'outil dans les 82 agences.
- Préviation des volumes mensuels de commandes. Elle a amené une révision des projections du 1er trimestre 2012 de la direction commerciale, puis un plan d'économie.
- Refonte des calculs d'objectifs. La performance des agences est estimée par les Plus Proches Voisins puis combinée à l'objectif annuel de l'entreprise fixé par la direction. La méthode est validée par le métier.

Environnement technique : R, macros Excel, solution BI « Reportive Studio »

2010, 4 mois – Pacifica : chargé d'études statistiques en CDD

- Production du reporting annuel d'activité de la direction commerciale.
- Refonte des macros Excel/Powerpoint : mise à jour passée de plusieurs jours à quelques minutes.

Environnement technique : Excel, Powerpoint, VBA

2001 à 2009 – Education Nationale : enseignant, chercheur

- Enseignement en mathématiques en classe préparatoire HEC.
- Travaux dirigés en statistiques et probabilités, algèbre, analyse (Universités de Lyon et de Caen).
- Cryptographie : publication sur la sécurité du vote par internet en collaboration avec un chercheur d'Orange Labs (« A practical and secure coercion-resistant scheme for internet voting »).

Environnement technique : LaTeX, Scilab

Formation

2021 Azure Certificat « Examen AZ-900 : Microsoft Azure Fundamentals »

GCP Certificat Coursera de la spécialisation de 5 cours « Machine Learning with TensorFlow on GCP »

2020 Dataiku certificats de manipulation de donnée et modélisation « Core designer », « ML practitioner »

Python 5 certificats Datascientest dont « Advanced ML with Scikit-learn »

2013 Cours du soir sur le Marketing Stratégique et le CRM, CNAM (Paris)

2009 Master 2 Pro « Statistiques, Informatique, Techniques Numériques », Lyon 1

2004 Doctorat de mathématiques, théorie des représentations des groupes, Lyon 1

2000 Agrégation de mathématiques, option probabilités et statistiques, ENS Cachan