

Alternance

« Construction et traitement de bases de données volumineuses pour les modèles de coûts »

SNCF Réseau

Description du poste

Au sein de SNCF Réseau, vous êtes rattaché(e) à la Direction Clients et Marchés, dans le département Tarification et Valorisation.

L'alternance s'inscrit dans les travaux de modélisation des coûts du Gestionnaire d'Infrastructure ferroviaire français, avec un fort enjeu autour de la structuration, fiabilisation et exploitation de bases de données volumineuses.

Missions

La mission principale de l'alternant(e) sera de participer à la conception, l'amélioration et l'industrialisation d'un outil de construction et de traitement de bases de données massives, utilisé pour alimenter les estimations économétriques des modèles de coûts directement imputables au trafic ferroviaire (coûts marginaux) qui déterminent directement une partie des péages dont s'acquittent les entreprises ferroviaires (en accord avec le droit européen).

À ce titre, il/elle devra :

- Contribuer à la conception d'un outil de traitement et de structuration de données (en Python)
- Participer à la construction de bases de données consolidées et exploitables à partir de sources multiples et volumineuses
- Améliorer les processus existants de collecte, nettoyage et transformation des données
- Assurer la qualité, la cohérence des bases de données produites
- Collaborer avec les économistes du département pour adapter les données aux besoins des modèles

Une fois les bases construites, l'alternant(e) contribuera également à :

- La réalisation d'analyses quantitatives et économétriques
- L'exploitation des données sous Stata (prioritairement) ou R
- La production de résultats permettant d'alimenter les modèles de coûts marginaux liés au trafic

Activités principales

Encadré(e) par les économistes du département, l'alternant(e) devra :

- Développer et maintenir des scripts de traitement de données (Python principalement)
- Concevoir, si possible, des pipelines de données robustes et automatisés
- Manipuler des bases de données de grande taille (data engineering)
- Documenter les processus de construction des bases
- Participer à l'analyse statistique et économétrique des données produites
- Faire un suivi des avancées de la littérature sur l'estimation des coûts marginaux
- Contribuer à l'amélioration et à l'adaptation des modèles existants pour assurer leur bon fonctionnement, leur évolution et leur exploitation sur des différents sujets.

Aptitudes et compétences

Formation

- Formation en data science, data engineering, économétrie ou statistiques
- Niveau Master (M1/M2) ou école d'ingénieurs / ENSAE / ENS / équivalent

Compétences techniques

- Très bonne maîtrise de Python (indispensable)
- Bonne maîtrise de R
- Maîtrise de Stata fortement appréciée
- Expérience en traitement de données volumineuses
- Compétences en construction et gestion de bases de données
- Bonne maîtrise d'Excel

Qualités personnelles

- Rigueur et capacité d'analyse
- Autonomie et esprit d'initiative
- Capacité à travailler sur des problématiques complexes
- Intérêt pour l'économie des transports et la data appliquée

Localisation du poste

Siège – Saint-Denis (RER D – Stade de France Saint-Denis)

Contacts

Carlos Olarte Bácsares (carlos.olarte-bacares@reseau.sncf.fr), Jean-Charles Fourot (jean-charles.fourot@reseau.sncf.fr)