



Formation d'un groupe de **Statistique appliquée**

L'ENSAE organise chaque année un enseignement de Statistique Appliquée qui prend la forme d'un travail par groupe d'élèves. Cet enseignement complète les cours théoriques de statistique, d'économétrie, de sondages, de séries temporelles et d'introduction au *machine learning* de deuxième année. Il permet aux élèves de se familiariser avec les différentes étapes de la modélisation statistique (élaboration d'un modèle, estimation et test) et les appliquer sur des données réelles.

Ce projet est encadré par un (ou deux) animateur(s). Celui-ci propose un sujet en lien avec son occupation professionnelle et fournit les données au groupe retenu. Il a ensuite pour rôle de coordonner et de guider le travail en les rencontrant régulièrement dans l'année. En conséquence, il doit avoir une expertise statistique. Il assiste à la soutenance et participe à la notation finale.

L'essentiel du travail devra porter sur la modélisation statistique et la mise en oeuvre des méthodes vues en cours en deuxième année (régressions linéaires et logistiques, modèles usuels de séries temporelles, forêts aléatoires, etc.). Les concepts enseignés en troisième année (modèles de durée, méthodes non paramétriques, bootstrap, etc.) pourront également être abordés si besoin, avec mesure.

La production finale est constituée d'un compte-rendu d'étape (mi-février 2025), du mémoire et de la note de synthèse (début mai 2025) ainsi que d'une soutenance (entre le mi-mai et la mi-juin). Cela permet aux élèves d'apprendre à présenter un travail technique de façon claire et concise.

Objectifs pédagogiques

A travers le choix d'un animateur évoluant dans les milieux statistiques, économiques, financiers ou encore médicaux, l'ENSAE entend former les élèves sur les points suivants :

- Poser une question scientifique adaptée à une base de données réelles mise à disposition ;
- Mener une démarche scientifique collective (groupes de 3 ou 4 élèves) et guidée par un encadrant, qui inclut une revue de littérature, une description statistique des

- données et qui utilise les techniques quantitatives adaptées (statistiques et économétriques) pour répondre à la question posée ;
- Savoir synthétiser cette démarche à l'aide d'un rapport de 20 pages adressé à des spécialistes et d'une note de synthèse accessible à des non spécialistes.

Ce projet est *appliqué* dans le sens où le travail est effectué sur des données réelles, apportées par l'animateur, ou à tout le moins sur des simulations. Il s'agit d'un enseignement d'application et non d'une étude en grandeur réelle ; les fichiers de données mis à disposition des élèves doivent donc être propres et de taille raisonnable.

Il convient d'insister sur la nature pédagogique de ce projet : l'animateur doit posséder une expertise permettant de répondre aux questions des étudiants portant sur les aspects techniques et méthodologiques, et est rémunéré à ce titre. Il ne s'agit en aucun cas d'une mission de conseil similaire aux prestations que peut fournir la junior entreprise de l'ENSAE par exemple.

Le niveau de confidentialité s'appliquant aux données et au rapport doit être défini entre l'animateur et les élèves. Ces derniers sont tenus à l'obligation de discrétion professionnelle et sont à ce titre responsables de la sécurité des fichiers qui leur sont confiés.

Organisation pratique

Si vous souhaitez proposer un sujet, vous devez remplir le *modèle de présentation du sujet*, fichier attaché au courriel, et remplir le formulaire suivant :

<https://projets-entreprise.ensae.fr/>

Un exemple de sujet se trouve en annexe du présent document.

Les élèves forment eux-mêmes leurs groupes de trois ou quatre et prennent directement contact avec les encadrants des sujets qui les intéressent. Si l'intérêt est réciproque, le groupe et l'encadrant décident de travailler ensemble et nous tiennent au courant par mail. Nous mettons directement à jour en ligne l'affectation, de sorte que l'encadrant n'est plus contacté par d'autres groupes.

Le calendrier du projet est le suivant :

- début octobre : annonce des sujets aux étudiants
- mi-novembre : affectation des derniers groupes
- mi-février 2025 : remise d'une note d'étape
- mi-mai 2025 : remise du rapport
- fin mai – début juin 2025 : soutenance

Le volume horaire de l'enseignement est le suivant : dix séances de suivi de deux heures (soit en moyenne une séance et demi par mois), de préférence dans les locaux de l'ENSAE. La rémunération brute (payée en août 2025) est de l'ordre de 1215€ pour un animateur, 729€ par animateur s'ils sont deux.

Pour toutes questions, vous pouvez vous adresser à stat_app@ensae.fr et nous vous répondrons dans les plus brefs délais.

MODELE DE PRESENTATION DU SUJET

Prévision de la consommation d'électricité de la SNCF

Nom de l'animateur	Jean Martin
Entreprise	SNCF
N° de téléphone	01 45 XX XX XX
courriel	Jean.martin@sncf.fr

PROBLÉMATIQUE

Les prévisions de consommation d'électricité de traction sont un des leviers d'efficience de l'achat d'énergie. Une bonne anticipation des besoins permet d'améliorer la performance achat ainsi que la visibilité en matière de stratégie de couverture à mettre en place sur les marchés financiers.

La consommation d'électricité haute tension découle de l'activité des différentes entités de l'entreprise (Transport voyageurs et transports de marchandises) mais aussi de facteurs exogènes majeurs tels que la température.

Le but de l'étude proposée est d'élaborer un modèle efficient de prévision de la consommation d'électricité à un niveau de détail relativement fin (au niveau horaire par exemple).

Les enjeux sont multiples mais sont avant tout financiers. Les montants en jeu se comptent en centaines de millions d'euros.

Les élèves devront maîtriser le logiciel Python ainsi que les logiciels bureautiques standards (LaTeX, Pack office).

DONNÉES UTILISÉES

Les données utilisées dans le cadre de cette études sont de natures diverses :

- Données de trafic des différentes activités (Train kilomètres, tonnes remorquées etc.) ;
- Données d'alimentation du réseau ferré national en électricité ;
- Données météorologiques.

BIBLIOGRAPHIE INDICATIVE

Traction Electrique – *JM Allenbach / P. Chabpas / M. Comte / R. Kaller* – Presses polytechniques et universitaires romandes

Consommation française d'électricité caracteristiques et methode de prevision - *RTE*