Projet de mise en place d’une plateforme d’estimation des écoulements.

# Objectif du projet

Nous souhaitons mettre en place une plateforme de développement des modèles d’écoulements des dépôts à vue. Cette plateforme sera mise à disposition des filiales.

# Livrables du projet

Une application web qui contiendra les fonctionnalités suivantes :

* La possibilité de créer son propre modèle en fonction des données téléchargées ;
* Un rapport de modélisation que la filiale va devoir rédiger en temps même temps qu’elle développe son modèle.

# Composition de la plateforme

Commençons par rappeler les grandes étapes de développement d’un modèle en général.

* Importation des données
* Data préparation
* Estimation
* Diagnostic du modèle (vérification des hypothèses)
* Performance et stabilité du modèle

## Panel 1 : Importation des données

L’utilisateur doit préparer les données en amont dans un fichier Excel dont le format est défini comme suit :

*Ici mettre le template du fichier excel en PJ*

Une fois les données préparées, l’importation peut débuter. L’utilisateur cliquera sur le bouton importer les données pour simplement importer et commencer le travail.

## Panel 2 : Analyse exploratoire

L’utilisateur a ensuite la possibilité d’effectuer une analyse exploratoire sur les données brutes mais aussi sur les données nettoyées.

Fonctionnalités :

Insérer un widget sur la sélection du type d’analyse : nous proposons ici essentiellement 3 types d’analyse :

* Scatter plot (nuage de points): ce graphe permet d’identifier une quelconque relation qui pourrait exister entre deux variables. Il est conseillé d’utiliser ce graphe lorsque les deux variables sont quantitatives.
* Line plot : ce graphe a le même objectif que le nuage de points. Il s’agit d’une courbe représentée par deux variables (abscisses et ordonnées).
* Boxplot : ce graphe est généralement utilisé pour vérifier la distribution et surtout la présence des points aberrants ou outliers.
* Barplot
* Histogram : ce graphe pêrmet également d’observer la distribution d’une variable.

Insérer un widget sur la sélection des variables à analyser : si l’utilisateur choisit :

* Boxplot : seule une variable apparait dans la sélection
* Scatter plot : deux variables apparaissent
* Line plot : deux variables apparaissent
* Histogram : une variable apparait
* Barplot une variable apparait

## Panel 3 : Préparation des données

Cette section est relative aux traitements appliqués sur les données importées. Il s’agit dans un premier temps de fournir une vue sur la qualité des données importées puis de nettoyer ces données pour l’étape d’estimation.

### Qualité des données

Dans cette section/étape, l’application offre la possibilité de connaitre :

* Le nombre de lignes,
* Le nombre de colonnes,
* Le nombre de données manquantes par variables,
* Des statistiques telles que le minimum, la moyenne, le quartile 25, la médiane, le quartile 75, le maximum

### Nettoyage des données

Dans cette section, il s’agit de proposer des solutions de nettoyage des données selon un certain nombre de scénarios :

* L’imputation : si le nombre d’observations pour une variable est inférieure à 5%.
* La suppression : l’utilisateur a le choix de ne pas considérer la variable pour des raisons qu’il devra justifier.

L’utilisateur pourra s’appuyer sur l’analyse de la qualité des données.

## Panel 3 : Processus de sélection des variables à intégrer dans le modèle

Dans cette section, il s’agit essentiellement de définir le processus de sélection des variables. Dans l’étape précédente, nous avons sélectionné les variables en termes de qualité (nombre d’observations manquantes). Ici, nous allons choisir statistiquement les variables à faire entrer dans le modèle. Nous nous appuierons principalement sur l’analyse de la corrélation entre les variables.

## Panel 4 : Estimation du modèle

Dans cette section, l’utilisateur a le choix sur un certain nombre d’approches.

* L’estimateur de Kaplan Meier ;
* L’estimateur de ...
* L’approche de Selvaggio ...

## Panel 5 : Validation des résultats

Dans cette section, la plateforme offre la possibilité à l’utilisateur de valider les résultats de l’estimation :

Sens économique des variables