

*Intervenants* : Caroline Bérard ([caroline.berard@univ-rouen.fr](mailto:caroline.berard@univ-rouen.fr))  
Kaci Makour ([kacislemakour@gmail.com](mailto:kacislemakour@gmail.com))

## PROJET : Segmentation de clients d'une boutique en ligne

---

Ce projet est à rendre le **samedi 1er mars** avant minuit. Il est à faire en **trîôme**. Vous devrez déposer un **rapport en .pdf** (n'excédant pas 20 pages), ainsi que votre **script python commenté en .ipynb** sur l'espace de dépôt de la page de cours Moodle. Le code doit être exécutable dans l'ordre séquentiel de haut en bas afin de nous assurer des résultats obtenus. Le rapport pdf inclura une introduction du projet, une présentation des données, les différentes sections d'analyse et une conclusion. Pour chaque question, vous mettrez les graphiques obtenus, ainsi que vos **descriptions, commentaires, remarques et interprétations**.

---

**Objectif** : Le jeu de données concerne des informations clients d'une boutique en ligne. L'idée est d'explorer les données, d'analyser ce jeu de données de la manière la plus complète possible, et de visualiser les relations afin de segmenter les clients. L'objectif est de faire une étude descriptive, et en particulier de construire une typologie des clients de la boutique en ligne. Les différentes variables du jeu de données sont définies ci-dessous :

- **CustomerID** : Identifiant unique du client
- **Age** : Âge du client
- **Gender** : Sexe (Male, Female)
- **AnnualIncome** : Revenu annuel (en milliers d'euros)
- **SpendingScore** : Score (0-100) représentant la propension du client à dépenser
- **OnlinePurchases** : Nombre d'achats en ligne au cours des 12 derniers mois
- **WebsiteVisits** : Nombre de visites sur le site au cours des 12 derniers mois
- **CategoryPreference** : Catégorie préférée (Clothing, Electronics, Books, etc.)

Les données (fictives) sont stockées dans le fichier *customer\_segmentation.csv*.

## 1 Présentation et description des données

## 2 Analyse univariée pour chaque variable quantitative/qualitative

## 3 Analyse bivariée

- Étudier la relation entre le revenu annuel du client et sa propension à dépenser.
- Étudier la relation entre le nombre de visites du site web et le nombre d'achats en ligne.
- Étudier les différences d'aptitude à dépenser entre les sexes.

## 4 Analyse comportementale

- Identifier les caractéristiques des clients visitant très souvent le site web de la boutique.
- Analyser les catégories préférées par sexe et par âge.
- Classer les clients en 3 catégories selon leur aptitude à dépenser : Low, Medium, High et identifier leurs caractéristiques.

À ce stade, et avant la dernière section, vous pouvez faire une conclusion/interprétation de toutes vos analyses descriptives des données.

## 5 Clustering des clients

- Utiliser une méthode de classification non supervisée pour faire un clustering des patients ayant des caractéristiques similaires.
- Visualiser les clusters et donner une interprétation.

## 6 Conclusion générale du projet

Quelles conclusions en tirez-vous? Quel serait le plan d'action pour augmenter le nombre de ventes en ligne ?

