

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

PROJET : Analyse des performances de campagnes e-mails marketing

Ce projet est à rendre le **dimanche 8 mars** avant minuit. Il est à faire en **trinôme**. 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 séquentiellement de haut en bas pour assurer la reproductibilité des résultats obtenus. Le rapport pdf incluera 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 : Une entreprise de e-commerce souhaite analyser la performance de ses campagnes d'e-mails marketing afin d'améliorer l'efficacité de ses actions. Vous disposez d'un jeu de données synthétique (*email_campaigns.csv*) contenant les caractéristiques de plus de 10000 campagnes envoyées à différents clients entre 2023 et 2025. L'objectif du projet est d'effectuer une analyse descriptive complète des campagnes, d'explorer les données et de visualiser les variables afin de mettre en évidence les leviers de performance, et de formuler des recommandations pour optimiser la stratégie d'e-mail marketing.

Les différentes variables du jeu de données sont définies ci-dessous :

- **CampaignID** : Identifiant unique de la campagne
- **SendDate** : Date d'envoi de la campagne
- **SendHour** : Heure d'envoi de la campagne (0 – 23)
- **AudienceSize** : Taille de la cible (nombre d'adresses)
- **Subject** : Sujet de l'e-mail
- **TopKeywords** : 6 mots-clés principaux extraits du contenu de l'e-mail (TF-IDF simplifié)
- **OpenRate** : Taux d'ouverture du mail (en %)
- **ClickRate** : Taux de clic (en %)
- **ConversionRate** : Taux de conversion (en %)
- **DeviceType** : Type d'appareil utilisé (Mobile, Desktop, Tablet)
- **CustomerSegment** : Segment client (New, Returning, VIP, Inactive, Loyal)

- **CampaignType** : Type de campagne (Promotion, Newsletter, Reminder, Seasonal, Re-engagement)

1 Présentation et description des données

2 Analyse univariée pour chaque variable quantitative/qualitative

3 Analyse bivariée

- Relation entre **SendHour** et **OpenRate**.
- Relation entre **AudienceSize** et **OpenRate**.
- Influence du **DeviceType** sur le **ClickRate** ou **ConversionRate**.
- Différences de performances selon le **CustomerSegment**.
- Comparaison des types de campagne (**CampaignType**) : quelles catégories performent le mieux ?

4 Analyse des mots-clés

- Faire une analyse de fréquence des mots apparaissant dans la colonne **TopKeywords**.
- Identifier les mots les plus utilisés dans les campagnes les plus performantes (avec **OpenRate** élevé).
- Comparer les mots-clés dominants par type de campagne.

[Visualiser les résultats de façon ludique et parlante pour une audience non technique.](#)

5 Analyse croisée : contenu et performance

- Combiner vos analyses numériques et textuelles pour résumer les campagnes avec les meilleurs taux d'ouverture et de conversion.

6 Conclusion générale du projet

Rédiger une conclusion claire et opérationnelle :

- Quels sont les principaux facteurs de succès d'une campagne e-mail ?
- À quelle(s) heure(s) ou à quel(s) type(s) de clients envoyer ?
- Quels types de messages ou mots-clés fonctionnent le mieux ?
- Quelles pistes d'amélioration proposeriez-vous pour les futures campagnes ?