

The background of the slide is a blurred image of a document with a pen resting on it. A green circuit-like pattern with lines and circles is overlaid on the left side of the image.

# ONLINE SHOPPERS PURCHASING INTENTION DATASET

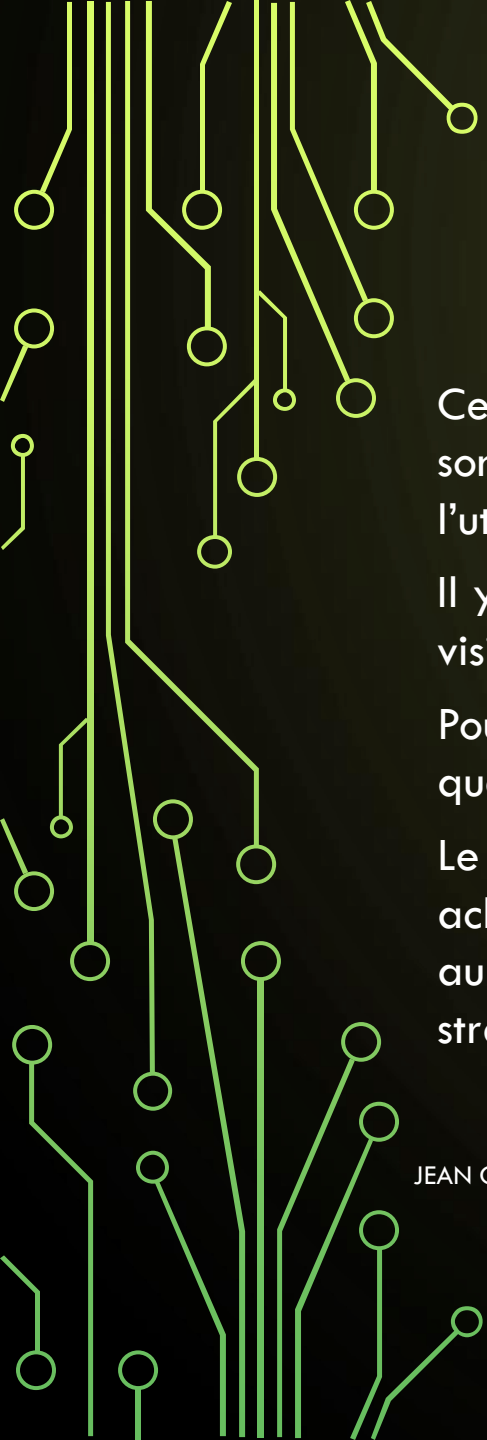
ESILV - PYTHON FOR DATA ANALYSIS - DEVOIR 2021

JEAN COUTURON - IBO2

A decorative graphic on the left side of the slide, consisting of a network of thin, light green lines and small circles, resembling a circuit board or a neural network diagram.

# PRÉSENTATION DU DATASET

JEAN COUTURON - IBO2



Ce dataset contient les informations de 12 330 sessions sur un site de vente en ligne. Les informations sont par exemple, le temps passé par un utilisateur sur une page du site, avec quel appareil l'utilisateur est allé sur le site, quel jour de l'année etc.

Il y a aussi une colonne indiquant si l'utilisateur a effectué un achat sur le site ou pas à la suite de sa visite. Cette colonne est de type boolean, soit true soit false. Ce sera notre colonne « TARGET ».

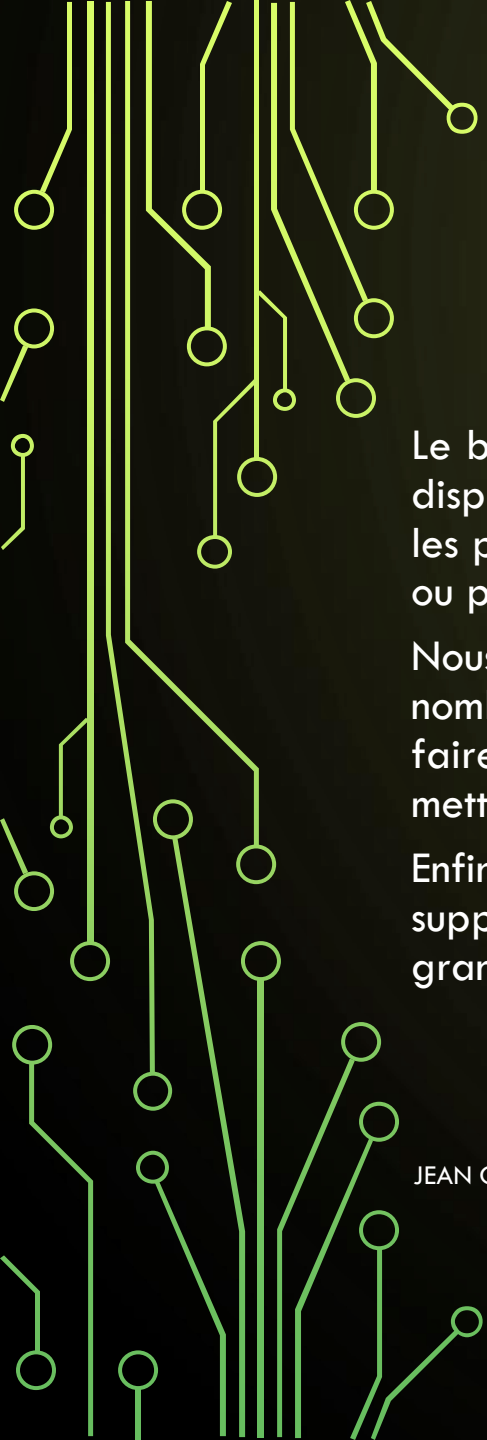
Pour finir chaque session correspond à un utilisateur différent. De plus les données n'ont été prélevées que sur une année.

Le but de mon étude va donc être de déterminer si un utilisateur lors d'une session va effectué un achat ou non et quels seront les informations utiles pour déterminer cela. Elle pourra donc permettre au propriétaire du site de savoir à l'avance si un visiteur va acheter ou non et d'adapter ses stratégies en fonction des informations récupérées.

A decorative graphic on the left side of the slide, consisting of a network of thin, light green lines and small circles, resembling a circuit board or a neural network diagram.

# RÉFLEXIONS SUR LE DATASET

JEAN COUTURON - IBO2

A decorative graphic on the left side of the slide, consisting of a network of thin, light blue lines and small circles, resembling a circuit board or a stylized tree structure.

Le but va donc être de déterminer sous quelle condition un utilisateur va faire un achat sur le site. Comme nous disposons de la date à laquelle un achat sera effectué, l'un des éléments qu'il sera intéressant de regarder est si les périodes de fête comme Noël, la Saint-Valentin ou encore la fête des mères seront des moments dans l'année où plus d'achat seront effectués, auront une influence sur notre « TARGET ».


Nous disposons aussi du nombre de pages visitées lors d'une session et du type de ces pages. Regarder si le nombre de pages visitées a un impact sur l'achat ou non mais aussi quel type de page son le plus susceptible de faire acheter un visiteur peut nous renseigner sur différentes stratégies commerciales que ce site souhaiterait mettre en place.

Enfin nous disposons du temps passé par un visiteur sur chaque page et à quel moment il quitte le site. On suppose que plus un utilisateur passe de temps sur le site plus la probabilité qu'il passe une commande est grande et inversement, nous pourrions donc confirmer cela.

A decorative graphic on the left side of the slide, consisting of a network of thin, light green lines and small circles, resembling a circuit board or a stylized tree structure.

# LIMITE DU DATASET

JEAN COUTURON - IBO2



Les limites de ce dataset sont le fait qu'il ne se limite qu'à une seule session par utilisateur. Or, la fidélité des utilisateurs au site a souvent une grande influence sur l'achat ou non d'un article par un utilisateur.

Nous ne savons pas non plus à quel type de site vente en ligne nous avons à faire. Ainsi, si c'est un site comme amazon les périodes de fêtes peuvent avoir une grande influence sur la « TARGET » mais si c'est un site de vente d'outillage pour professionnel, nous pouvons supposé que l'impact sera tout de suite beaucoup moins grand.

Nous n'avons pas non plus le détail des commandes. Par exemple, il pourrait être avantageux pour le propriétaire du site de concentrer sa communication sur les utilisateurs passant de grosses commandes qui rapportent plus d'argent que sur les utilisateurs qui ne vont faire qu'une petite commande. Toutes les commandes ne se valent pas.

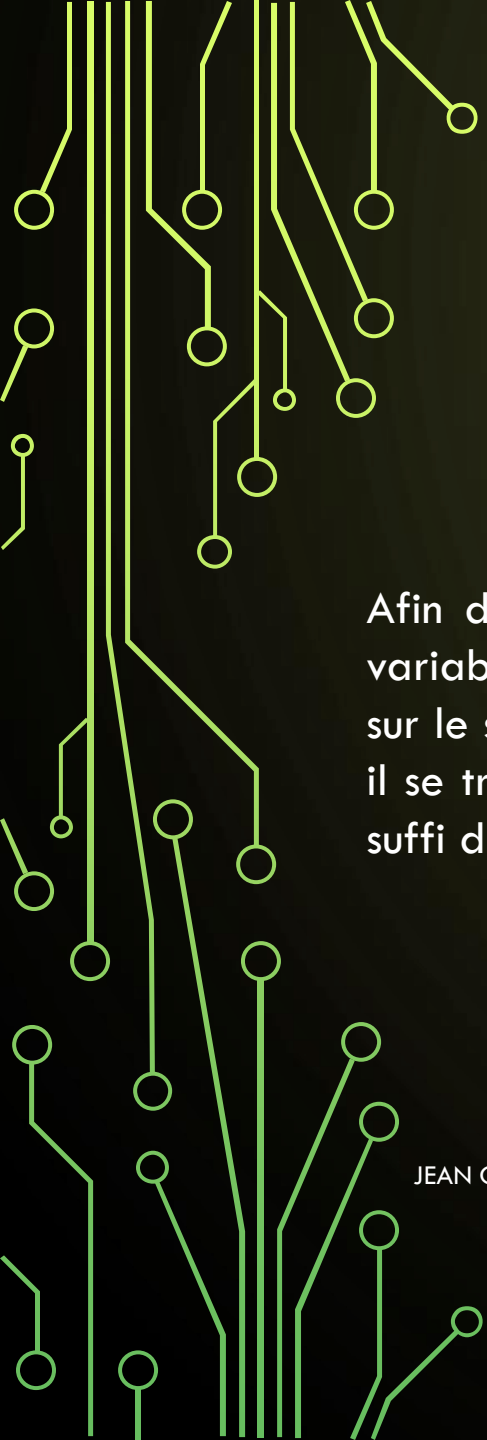


A decorative graphic on the left side of the slide, consisting of a network of light green lines and small circles, resembling a circuit board or a stylized tree structure.

# CRÉATION DE VARIABLE

JEAN COUTURON - IBO2



A decorative graphic on the left side of the slide, consisting of a network of thin, light blue lines and small circles, resembling a circuit board or a stylized tree structure. The lines and circles are arranged in a way that suggests a flow or a path, with some lines ending in small circles.

Afin de pouvoir déterminer plus précisément si un achat va être effectué ou non, j'ai décidé de créer la variable du temps total passé par un utilisateur sur le site. J'ai supposé que plus un utilisateur passait de temps sur le site plus la probabilité qui fasse un achat était grande et inversement quelque soit la page sur laquelle il se trouvait. Le temps passé sur chaque type de page du site était déjà présent dans le dataset, il a donc suffi de les additionner pour créer cette variable.