



Structured Query Language

Première année de Master en Statistique pour l'évaluation et la prévision

Pellerin Francois

Rezi Sabashvili (rezi.sabashvili@etudiant.univ-reims.fr)

Table des matières

1	Introduction et base de données	2
2	Classification pour les analyses	4
2.1	Caractéristiques générales	4
2.1.1	Identifiants de l'utilisateur et de la visite :	4
2.1.2	Informations sur l'appareil :	5
2.1.3	Informations géographiques :	5
2.2	Caractéristiques de l'indicateur d'engagement	5
2.2.1	Détails sur les actions de l'utilisateur :	5
2.2.2	Détails sur le temps et les interactions :	5
2.2.3	Informations sur l'action spécifique :	5
2.3	Caractéristiques temporelles	5
2.3.1	Horodatage général :	5
2.3.2	Première action :	6
2.3.3	Durée de la visite :	6
2.3.4	Horodatage des actions :	6
2.3.5	Historique des visites :	6
2.3.6	Heure locale :	6
2.4	Caractéristiques d'activités	6
2.4.1	Type d'action et URL :	6
2.4.2	Durée de l'activité :	6
2.4.3	Détails supplémentaires de la page :	7
2.4.4	Interactions et événements :	7
2.4.5	Produits et transactions :	7
2.5	Indicateurs à calculer : régularité, investissement, curiosité	7
2.5.1	L'indicateur de régularité	7
2.5.2	L'indicateur d'investissement	7
2.5.3	L'indicateur de curiosité	7

1 Introduction et base de données

Notre base de données est issue du site d'Éducapacité. Cette base de données anonymisée comporte 129 colonnes distinctes (sans l'itération des colonnes "*actionDetails n*"). Les colonnes uniques sont les suivantes :

- **idVisit** : Identifiant unique de la visite sur le site.
- **visitorId** : Identifiant unique de l'utilisateur qui effectue la visite.
- **fingerprint** : Empreinte unique associée à l'appareil de l'utilisateur.

Spécifiques à chaque action ou événement effectué durant la visite (n allant de 0 à 217) :

- **type (actionDetails n)** : Type d'action effectuée (par exemple, "action" indique une action réalisée par un utilisateur).
- **url (actionDetails n)** : URL de la page visitée par l'utilisateur.
- **pageTitle (actionDetails n)** : Le titre de la page visitée.
- **pageIdAction (actionDetails n)** : L'ID associé à l'action sur la page.
- **idpageview (actionDetails n)** : Identifiant unique de la page vue pendant cette visite.
- **serverTimePretty (actionDetails n)** : Date et heure de l'action, formatée de manière lisible.
- **pageId (actionDetails n)** : Identifiant unique de la page visitée.
- **timeSpent (actionDetails n)** : Le temps passé sur cette page pour cette action, en secondes (exemple : "2").
- **timeSpentPretty (actionDetails n)** : Le temps passé sur la page sous un format plus lisible (exemple : "11s").
- **pageviewPosition (actionDetails n)** : Position de la page vue dans l'ordre chronologique des actions de la visite.
- **title (actionDetails n)** : Titre de la page (semblable à pageTitle).
- **subtitle (actionDetails n)** : Un sous-titre pour la page, s'il existe.
- **icon (actionDetails n)** : Un icône représentant l'action ou la page (quasiment toujours vide).
- **iconSVG (actionDetails n)** : L'URL du fichier SVG de l'icône, représentant l'action ou la page.
- **timestamp (actionDetails n)** : Le moment exact de l'action, exprimé en timestamp Unix (nombre de secondes depuis le 1er janvier 1970).

Des attributs généraux concernant l'utilisateur et la visite dans son ensemble (à partir d'ici ce n'est pas pour n allant de 0 à 217):

- **goalConversions** : Indique le nombre conversions atteints par l'utilisateur pendant sa visite (toujours 0).
- **siteCurrency** : La devise utilisée sur le site (toujours "EUR").
- **siteCurrencySymbol** : Le symbole de la devise (toujours "€").
- **serverDate** : La date sur le serveur au moment de la visite (format : jour/mois/année).
- **visitServerHour** : L'heure sur le serveur au moment de la visite.
- **lastActionTimestamp** : Le timestamp du dernier événement ou action effectué sur le site par l'utilisateur.
- **lastActionDateTime** : La date et l'heure exactes de la dernière action effectué par l'utilisateur.
- **siteName** : Le nom du site web visité (ici "Éducapacités").
- **serverTimestamp** : Un timestamp correspondant à un moment précis sur le serveur.
- **firstActionTimestamp** : Le timestamp de la première action effectuée sur le site par l'utilisateur.
- **serverTimePretty** : Heure sous un format lisible, correspondant à l'heure serveur.
- **serverDatePretty** : Date sous un format lisible, correspondant à la date serveur.
- **serverDatePrettyFirstAction** : La date lisible de la première action effectuée.
- **serverTimePrettyFirstAction** : L'heure lisible de la première action effectuée.
- **userId** : L'identifiant unique de l'utilisateur (les emails avant l'anonymisation).
- **visitorType** : Le type de visiteur ("new" pour un nouveau visiteur, "returningCustomer" pour un visiteur de retour).
- **visitorTypeIcon** : L'icône associée au type de visiteur (par exemple, un retour avec l'icône "returningVisitor").
- **visitConverted** : Indicateur qui indique si la visite a abouti à une conversion (par exemple, un achat ou une inscription).
- **visitConvertedIcon** : Icône associée à l'état de conversion de la visite.
- **visitCount** : Le nombre de visites effectuées par l'utilisateur.

- **visitEcommerceStatus** : Statut du commerce électronique de la visite (ex. si une conversion a eu lieu ou non).
- **visitEcommerceStatusIcon** : Icône associée au statut du commerce électronique.
- **daysSinceFirstVisit** : Nombre de jours écoulés depuis la première visite de l'utilisateur.
- **secondsSinceFirstVisit** : Nombre de secondes écoulées depuis la première visite de l'utilisateur.
- **daysSinceLastEcommerceOrder** : Nombre de jours écoulés depuis la dernière commande effectuée par l'utilisateur.
- **secondsSinceLastEcommerceOrder** : Nombre de secondes écoulées depuis la dernière commande effectuée.
- **visitDuration** : Durée totale de la visite de l'utilisateur.
- **visitDurationPretty** : Durée de la visite sous un format lisible.
- **searches** : Le nombre de recherches effectuées pendant la visite.
- **actions** : Le nombre d'actions réalisées pendant la visite (comme les clics, interactions avec les éléments de la page).
- **interactions** : Le nombre total d'interactions effectuées pendant la visite, souvent identique à la colonne "actions".
- **referrerType** : Type de référent qui a conduit à la visite (direct, moteur de recherche, réseau social, etc.).
- **referrerTypeName** : Nom du type de référent.
- **referrerName** : Le nom du site référent (par exemple, Google, Facebook).
- **referrerKeyword** : Le mot-clé utilisé dans le moteur de recherche (s'il y en a un).
- **referrerKeywordPosition** : Position du mot-clé dans les résultats de recherche.
- **referrerUrl** : L'URL du site référent.
- **referrerSearchEngineUrl** : URL du moteur de recherche référent.
- **referrerSearchEngineIcon** : Icône du moteur de recherche référent.
- **referrerSocialNetworkUrl** : URL du réseau social référent.
- **referrerSocialNetworkIcon** : Icône du réseau social référent.
- **languageCode** : Code de langue utilisé pendant la visite (ex. "fr-fr" pour le français).
- **language** : Langue de l'utilisateur pendant la visite.
- **deviceType** : Type d'appareil utilisé (par exemple, "Bureau" ou "Mobile").
- **deviceTypeIcon** : Icône associée au type d'appareil.
- **deviceBrand** : Marque de l'appareil utilisé (par exemple, Apple, Samsung).
- **deviceModel** : Modèle spécifique de l'appareil.
- **operatingSystem** : Système d'exploitation de l'appareil (par exemple, "Windows", "Mac").
- **operatingSystemName** : Nom du système d'exploitation.
- **operatingSystemIcon** : Icône associée au système d'exploitation.
- **operatingSystemCode** : Code correspondant au système d'exploitation.
- **operatingSystemVersion** : Version du système d'exploitation.
- **browserFamily** : Famille du navigateur utilisé (ex. Blink, Gecko).
- **browserFamilyDescription** : Description de la famille du navigateur.
- **browser** : Le navigateur utilisé (par exemple, Chrome, Firefox).
- **browserName** : Nom du navigateur utilisé.
- **browserIcon** : Icône associée au navigateur.
- **browserCode** : Code du navigateur.
- **browserVersion** : Version du navigateur.
- **totalEcommerceRevenue** : Revenus totaux du commerce électronique générés par la visite.
- **totalEcommerceConversions** : Nombre total de conversions dans le commerce électronique durant la visite.
- **totalEcommerceItems** : Nombre total d'articles achetés lors de la visite.
- **totalAbandonedCartsRevenue** : Revenus totaux générés par les paniers abandonnés.
- **totalAbandonedCarts** : Nombre total de paniers abandonnés.
- **totalAbandonedCartsItems** : Nombre total d'articles dans les paniers abandonnés.
- **events** : Nombre d'événements spéciaux générés pendant la visite (ex. des actions liées à des événements spécifiques).
- **continent** : Le continent où l'utilisateur se trouve pendant sa visite.
- **continentCode** : Code du continent (ex. "eur" pour l'Europe).
- **country** : Le pays de l'utilisateur.

- **countryCode** : Le code du pays (ex. “fr” pour la France).
- **countryFlag** : Icône du drapeau du pays.
- **region** : Région où l'utilisateur se trouve.
- **regionCode** : Code de la région.
- **city** : Ville où l'utilisateur se trouve.
- **location** : Localisation géographique de l'utilisateur (pays, région, ville).
- **latitude** : Latitude géographique de l'utilisateur.
- **longitude** : Longitude géographique de l'utilisateur.
- **visitLocalTime** : Heure locale de l'utilisateur pendant la visite.
- **visitLocalHour** : Heure locale de l'utilisateur, uniquement l'heure.
- **daysSinceLastVisit** : Nombre de jours écoulés depuis la dernière visite de l'utilisateur.
- **secondsSinceLastVisit** : Nombre de secondes écoulées depuis la dernière visite de l'utilisateur.
- **resolution** : Résolution d'écran de l'utilisateur.
- **plugins** : Liste des plugins utilisés par l'utilisateur pendant la visite (par exemple PDF).
- **pluginIcon (pluginsIcons n)** : Icône du premier plugin utilisé.
- **pluginName (pluginsIcons n)** : Nom du premier plugin utilisé.
- **pluginIcon (pluginsIcons n)** : Icône du deuxième plugin utilisé.
- **pluginName (pluginsIcons n)** : Nom du deuxième plugin utilisé.
- **pageLoadTime (actionDetails n)** : Temps de chargement de la page.
- **pageLoadTimeMilliseconds (actionDetails n)** : Temps de chargement de la page en millisecondes.

Actions et transactions d'achats sur le site :

- **productViewName (actionDetails n)** : Nom du produit consulté par l'utilisateur pendant la visite.
- **productViewPrice** : Prix du produit lorsqu'il est consulté par l'utilisateur sur le site, avant toute action d'achat
- **productViewSku (actionDetails n)** : Code SKU du produit consulté. Il s'agit d'un identifiant unique pour chaque produit
- **orderId (actionDetails n)** : Identifiant unique de la commande. Attribué au moment où un achat est effectué.
- **revenue (actionDetails n)** : Le revenu généré par la transaction.
- **revenueSubTotal (actionDetails n)** : Le sous-total de la commande avant l'ajout de taxes, frais d'expédition ou remises.
- **revenueTax (actionDetails n)** : Montant des taxes appliquées sur la commande.
- **revenueShipping (actionDetails n)** : Montant des frais d'expédition associés à la commande.
- **revenueDiscount (actionDetails n)** : Montant des remises appliquées à la commande.
- **items (actionDetails n)** : Nombre d'articles inclus dans la commande.

Colonnes associées à des détails sur les articles commandés :

- **itemSKU (itemDetails 0)** : ode SKU spécifique à l'article acheté.
- **itemName (itemDetails 0)** : Nom de l'article acheté.
- **itemCategory (itemDetails 0)** : Catégorie à laquelle appartient l'article.
- **price (itemDetails 0)** : Prix unitaire de l'article acheté.
- **quantity (itemDetails 0)** : Quantité de l'article achetée dans la commande.

2 Classification pour les analyses

Voici une classification des colonnes pertinentes en fonction de chaque objectif :

2.1 Caractéristiques générales

2.1.1 Identifiants de l'utilisateur et de la visite :

Identifier de manière unique l'utilisateur et la visite qu'il a effectuée sur le site.

- **visitorId** : Identifiant unique de l'utilisateur.
- **idVisit** : Identifiant unique de la visite sur le site.

- **idpageview (actionDetails n)** : Identifiant unique de la page vue pendant la visite.

2.1.2 Informations sur l'appareil :

Détails sur l'appareil utilisé par l'utilisateur pour effectuer la visite.

- **fingerprint** : Empreinte unique associée à l'appareil.
- **deviceType** : Type d'appareil utilisé (par exemple, mobile, desktop, tablette).
- **deviceBrand** : Marque de l'appareil utilisé (par exemple, Apple, Samsung).
- **deviceModel** : Modèle de l'appareil.
- **operatingSystem** : Système d'exploitation de l'appareil (par exemple, Windows, macOS, Android).
- **browser** : Navigateur utilisé par l'utilisateur (par exemple, Chrome, Firefox, Safari).

2.1.3 Informations géographiques :

Informations sur la localisation géographique de l'utilisateur pendant la visite.

- **continent** : Le continent de l'utilisateur.
- **country** : Le pays de l'utilisateur.
- **region** : Région géographique de l'utilisateur (par exemple, province, département).
- **city** : Ville de l'utilisateur.
- **location** : Localisation géographique de l'utilisateur (par exemple, coordonnées GPS ou zone approximative).

2.2 Caractéristiques de l'indicateur d'engagement

2.2.1 Détails sur les actions de l'utilisateur :

Informations sur le type et le moment des actions effectuées par l'utilisateur.

- **type (actionDetails n)** : Type d'action effectuée par l'utilisateur (par exemple, "click", "scroll").
- **serverTimePretty (actionDetails n)** : Date et heure de l'action de manière lisible.

2.2.2 Détails sur le temps et les interactions :

Le temps passé par l'utilisateur sur les pages et le nombre d'interactions réalisées.

- **timeSpent (actionDetails n)** : Temps passé sur chaque page ou action (en secondes).
- **actions** : Nombre d'actions réalisées durant la visite (clics, interactions avec les éléments de la page).
- **interactions** : Nombre total d'interactions effectuées durant la visite, souvent identique à "actions".

2.2.3 Informations sur l'action spécifique :

L'action spécifique réalisée sur la page.

- **pageIdAction (actionDetails n)** : Identifiant lié à l'action sur la page (pour suivre l'action particulière dans l'ensemble de la visite).

2.3 Caractéristiques temporelles

2.3.1 Horodatage général :

Liées à l'heure et la date de la visite de l'utilisateur, vision globale du moment où l'interaction a lieu.

- **serverDate** : La date sur le serveur au moment de la visite (jour/mois/année).
- **visitServerHour** : L'heure sur le serveur au moment de la visite.
- **serverTimePretty** : Heure sous un format lisible correspondant à l'heure serveur.
- **serverDatePretty** : Date sous un format lisible correspondant à la date serveur.

2.3.2 Première action :

Première interaction de l'utilisateur avec la plateforme (comprendre l'engagement initial ?)

- **serverDatePrettyFirstAction** : La date de la première action effectuée sous un format lisible.
- **serverTimePrettyFirstAction** : L'heure de la première action effectuée sous un format lisible.
- **firstActionTimestamp** : Timestamp de la première action effectuée sur le site.

2.3.3 Durée de la visite :

Mmesurer la durée totale de la visite et d'autres aspects temporels relatifs à la session de l'utilisateur.

- **visitDuration** : Durée totale de la visite (en secondes ou format lisible).
- **pageLoadTime (actionDetails n)** : Temps de chargement de la page (en secondes).
- **pageLoadTimeMilliseconds (actionDetails n)** : Temps de chargement de la page en millisecondes.

2.3.4 Horodatage des actions :

Associées à l'heure précise des actions effectuées durant la visite.

- **timestamp (actionDetails n)** : Moment exact de l'action en timestamp Unix.
- **lastActionTimestamp** : Timestamp du dernier événement ou action effectué sur le site.

2.3.5 Historique des visites :

Visites précédentes, étudier les interactions dans un contexte plus large.

- **daysSinceFirstVisit** : Nombre de jours écoulés depuis la première visite de l'utilisateur.
- **secondsSinceFirstVisit** : Nombre de secondes écoulées depuis la première visite.
- **daysSinceLastVisit** : Nombre de jours écoulés depuis la dernière visite.
- **secondsSinceLastVisit** : Nombre de secondes écoulées depuis la dernière visite.

2.3.6 Heure locale :

Heure à laquelle l'utilisateur interagit avec la plateforme selon son propre fuseau horaire.

- **visitLocalTime** : Heure locale de l'utilisateur pendant la visite.
- **visitLocalHour** : Heure locale de l'utilisateur, uniquement l'heure.

2.4 Caractéristiques d'activités

2.4.1 Type d'action et URL :

Type d'interaction de l'utilisateur avec la plateforme et à la page qu'il a visitée.

- **type (actionDetails n)** : Type d'action effectuée (par exemple, "action" indique une action réalisée par un utilisateur).
- **url (actionDetails n)** : URL de la page visitée par l'utilisateur.
- **pageTitle (actionDetails n)** : Le titre de la page visitée.
- **pageIdAction (actionDetails n)** : Identifiant associé à l'action sur la page.
- **idpageview (actionDetails n)** : Identifiant unique de la page vue pendant cette visite.

2.4.2 Durée de l'activité :

Temps passé sur une page ou une action spécifique pendant la visite.

- **timeSpent (actionDetails n)** : Temps passé sur cette page pour cette action, en secondes.
- **pageviewPosition (actionDetails n)** : Position de la page vue dans l'ordre chronologique des actions de la visite.

2.4.3 Détails supplémentaires de la page :

Informations supplémentaires concernant la page ou le produit que l'utilisateur a consulté.

- **title (actionDetails n)** : Titre de la page (semblable à pageTitle).
- **subtitle (actionDetails n)** : Un sous-titre pour la page, s'il existe.

2.4.4 Interactions et événements :

Actions spécifiques de l'utilisateur pendant sa visite, y compris les clics, interactions et événements spéciaux.

- **actions** : Nombre d'actions réalisées pendant la visite (clics, interactions avec les éléments de la page).
- **interactions** : Nombre total d'interactions effectuées pendant la visite, souvent identique à "actions".
- **events** : Nombre d'événements spéciaux générés pendant la visite (ex. des actions liées à des événements spécifiques).

2.4.5 Produits et transactions :

Interaction avec des produits spécifiques et les achats réalisés lors de la visite.

- **productViewName (actionDetails n)** : Nom du produit consulté par l'utilisateur.
- **productViewPrice** : Prix du produit consulté, avant toute action d'achat.
- **productViewSku (actionDetails n)** : Code SKU du produit consulté (identifiant unique pour chaque produit).
- **orderId (actionDetails n)** : Identifiant unique de la commande effectuée.
- **revenue** : Revenu généré par la transaction.
- **items (actionDetails n)** : Nombre d'articles achetés lors de la transaction.

2.5 Indicateurs à calculer : régularité, investissement, curiosité

Pour effectuer les calculs des indicateurs de régularité, d'investissement et de curiosité, nous pouvons utiliser les colonnes suivantes :

2.5.1 L'indicateur de régularité

- **serverDatePretty** : Date de la visite sous un format lisible, utile pour observer la régularité temporelle.
- **serverTimePretty** : Heure de la visite sous un format lisible.
- **pageviewPosition (actionDetails n)** : Position de la page vue, pour repérer la séquence des actions.
- **timestamp (actionDetails n)** : Moment exact de l'action en timestamp Unix.
- **daysSinceFirstVisit** : Nombre de jours depuis la première visite, utile pour analyser la régularité.
- **secondsSinceFirstVisit** : Nombre de secondes depuis la première visite.
- **daysSinceLastVisit** : Nombre de jours depuis la dernière visite.
- **secondsSinceLastVisit** : Nombre de secondes depuis la dernière visite.

2.5.2 L'indicateur d'investissement

- **actions** : Nombre d'actions réalisées.
- **visitDuration** : Durée totale de la visite de l'utilisateur.
- **searches** : Nombre de recherches effectuées pendant la visite.
- **totalEcommerceRevenue** : Revenu généré durant la visite, utile pour analyser l'investissement financier.
- **totalEcommerceConversions** : Nombre total de conversions (achats).
- **totalEcommerceItems** : Nombre d'articles achetés.

2.5.3 L'indicateur de curiosité

- **productViewName (actionDetails n)** : Nom du produit consulté, utile pour mesurer la variété des contenus.
- **productViewPrice** : Prix du produit consulté, utile pour observer les comportements d'achat.
- **productViewSku (actionDetails n)** : Code SKU pour identifier les produits consultés.

- **itemCategory (itemDetails 0)** : Catégorie des articles consultés, pour observer la diversité du contenu exploré.