Module Data Marketing

Session 1

Introduction

Parcours scolaire & professionnel







MSc Data Science & Organizational Behavior 2016 -. 2020

Data Engineer PwC Luxembourg 06.2020 - 03.2022 **Data Engineer** Société Générale 03.2022 - Aujourd'hui

Session 1 - Syllabus

Comprendre le Big Data

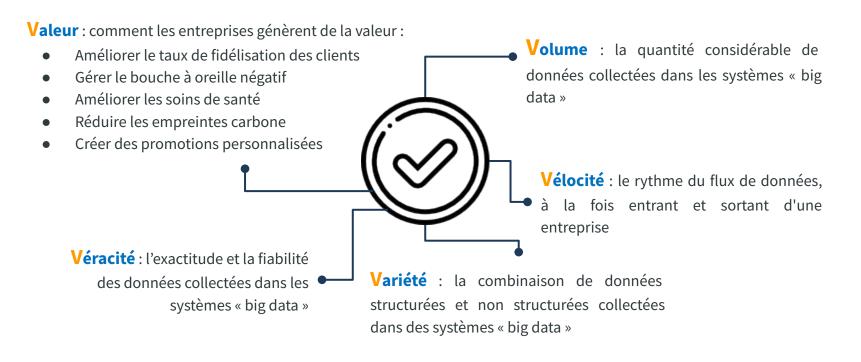
Donner de la valeur aux données

Préparer la certification Google Analytics

Comprendre le BIG DATA

Définition et enjeux du Big Data

Big Data : processus de capture, de fusion et d'analyse d'ensembles de données volumineux et variés pour comprendre les pratiques commerciales actuelles et rechercher de nouvelles opportunités pour améliorer les performances futures.



Types et format de données

Données structurées : chiffres ou textes stockés et organisées dans une structure de colonnes et de lignes.

→ Excel - ID dans une base de données - notation/avis

Données semi-structurées : données non formatées de manière conventionnelle et non référencé

→ Email - Tweet - Dossier

Données non structurées : données stockant de l'information non organisé, difficile à extraire et de toute autre forme difficile à organiser dans un format structuré.

→ Posts réseaux sociaux - vidéo - image - fichier son

Format de stockage des données : façon dont est représenté **un** type de données

→ Exemple :

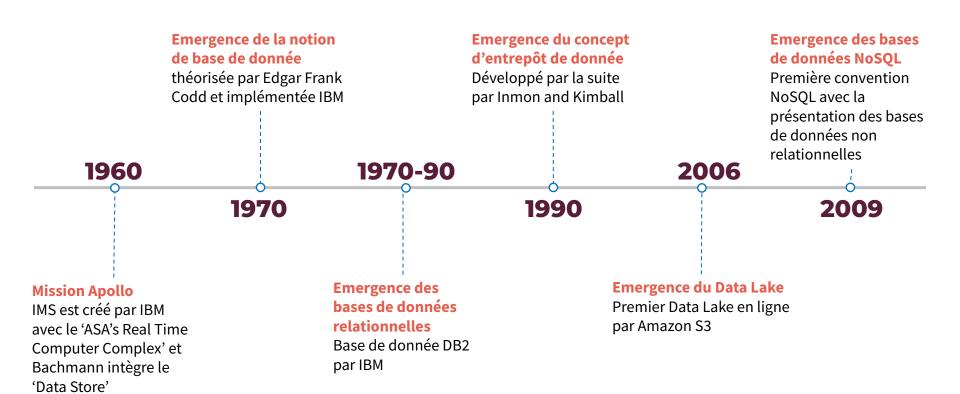
- Numérique Entier : 45 Décimal: 12.5
- Chaîne de caractère : data
- Chaîne de caractère unicode : donnée
- Booléen: Vrai
- Date et Heure: 01/01/2022 12:01:34

Le format de donnée est important pour la création des bases de données. Le format spécifique d'une colonne permet d'effectuer des opérations spécifiques

```
SELECT
symbol, name, exchange, assetType, ipoDate, delistingDate
FROM
default.ticker_listing
WHERE
symbol IN ('AAPL', 'GOOGL', 'AMZN', 'MSFT')
```

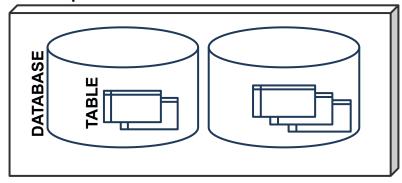
Stockage de la donnée

Histoire des bases de données



Mode de stockage analytique

DATA WAREHOUSE 'Entrepôt de donnée'

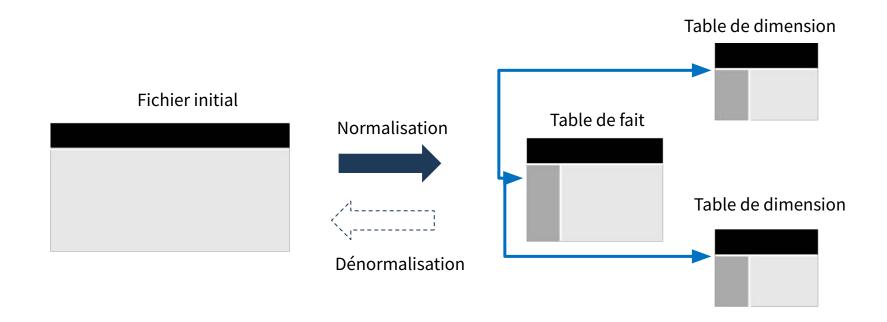


Donnée structurée

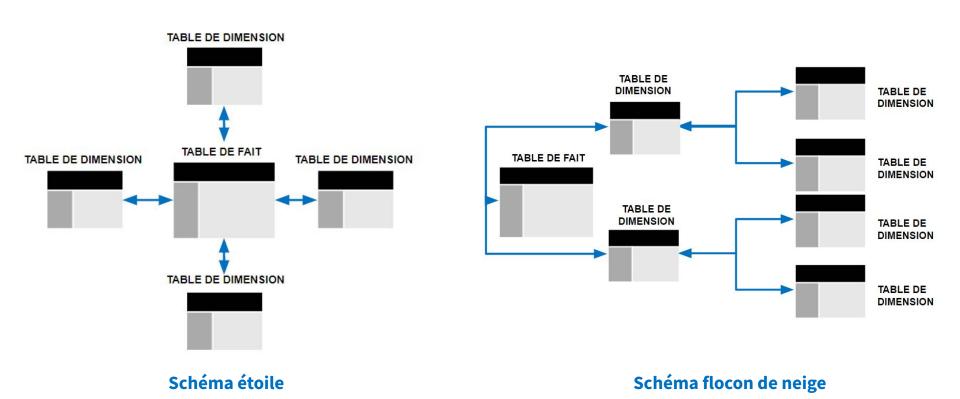


Donnée non-structurée

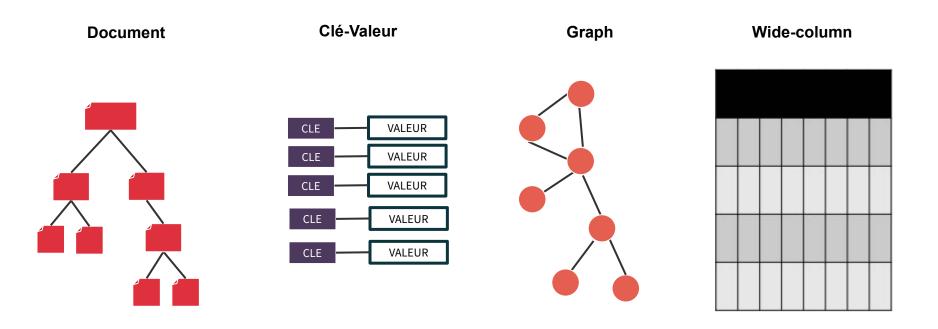
Base de donnée relationnelle



Optimiser le stockage de la donnée



Base de données NOSQL



Donner de la valeur aux données

Sources de données



→ Ex: récolter Nom, prénom, adresse mail du client



Données secondaires : les données ont déjà été collectées, souvent à d'autres fins ou par une autre organisation.

→ Ex: achat de fichier, données publiques.

Open Data : données accessibles librement et produites par une entreprise privée mais surtout par une administration/établissement/collectivité publique.

Data brokers (ou courtiers en données en français) : collectent elles-mêmes des données ou les achètent à d'autres entreprises et agrègent des informations, légalement ou non, avec des données provenant d'autres sources.

Qualité de la donnée

Exhaustivité



Existence de la donnée :

Présence obligatoire d'une valeur pour la donnée

Exhaustivité de périmètre

Exhaustivité des données pour un ensemble de données

Exactitude



Véracité:

Conformité de la donnée par rapport à la documentation de référence en interne

Conformité aux référentiels :

Conformité des valeurs de données de référentiels par rapport à la liste des valeurs du référentiel

Cohérence



Cohérence de données intra groupe de données :

Au sein d'un même groupe de données, cohérence d'une donnée par rapport à d'autres données

Cohérence de données inter groupe de données :

Entre deux groupes de données distincts, cohérence des données entre elles

Plausibilité:

Détection d'une valeur aberrante dans les données

Unicité



Unicité:

Conformité aux règles d'identification unique de la donnée quand celle-ci doit être unique

Ponctualité



Actualité:

Conformité de la dernière mise à jour de la donnée

Ponctualité:

Conformité entre la date de remise et la date de soumission effective des données

Digital Analytics

Data analytics: science qui consiste à analyser des données brutes afin de mieux interpréter des informations. Le processus d'analyse est le plus souvent automatisé et via des logiciels pour collecter et analyser la donnée

PREDICTIVE

Ce qui va probablement se produire à court terme.

« Que va-t-il se passer?»

Qu'est-il arrivé aux ventes la dernière fois que nous avons eu un été chaud ? Combien de modèles météorologiques prédisent un été chaud cette année ?

PRESCRIPTIVE

Suggère un plan d'action.

« Que dois-je faire ? »

Nous devrions ajouter un quart de soir à la brasserie et louer un réservoir supplémentaire pour augmenter la production si la probabilité d'un été chaud est mesurée comme la moyenne de ces cinq modèles météorologiques et que la moyenne est supérieure à 58 %

DESCRIPTIVE

Décrit ce qui s'est passé sur une période de temps donnée.

« Que s'est-il passé ? »

Le nombre de vues a-t-il augmenté ? Les ventes sont-elles plus fortes ce mois-ci que le mois dernier ?

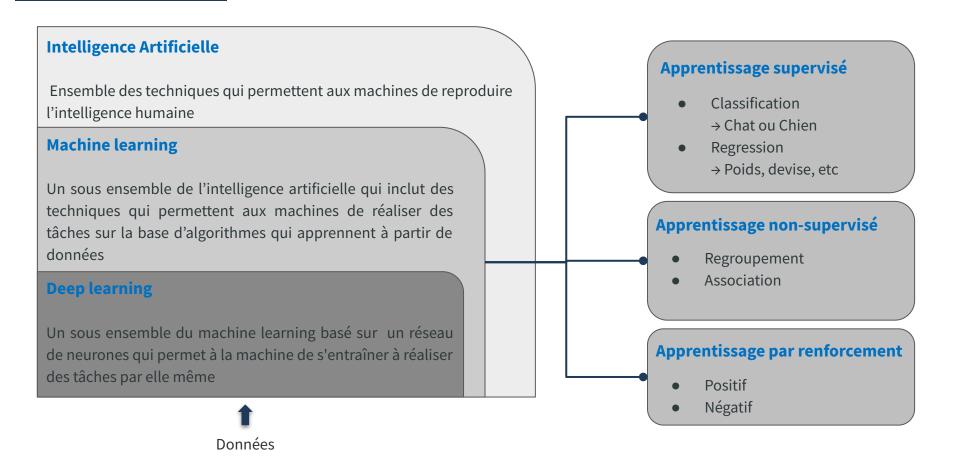
DIAGNOSTIQUE

Raison pour laquelle quelque chose s'est produit. Cela implique des entrées de données plus diverses et un peu d'hypothèses.

« Pourquoi cela s'est-il produit ? »

La météo a-t-elle affecté les ventes de bière ? Cette dernière campagne marketing a-t-elle eu un impact sur les ventes ?

Données : base de l'intelligence artificielle



Grands modèles de langage

Les grands modèles de langage (ou LLM de l'anglais Large Language Model) sont des programmes entraînés sur un ensembles de données linguistiques massives pour identifier les caractéristiques et les relations entre des éléments de données similaires sans intervention humaine (apprentissage automatique de la structure du langage humain.)







- **Robots conversationnels et assistants virtuels**: chatbots clients pour assistance, suivi des prises de contact via un site web ou assistant personnel (réponses questions ouvertes).
- Génération de code et débogage : extraits de code, identification et correction d'erreur de code, et complément des programmes sur instructions.
- Analyse du sentiment d'un texte : automatisation de la compréhension de la satisfaction des clients.
- Classification et regroupement de textes : catégorisation et tri de grands volumes de données pour identifier des thèmes et des tendances
- Traduction de documents et des pages web
- **Synthèse et paraphrase** : résumés d'articles, de publications, d'appels, de réunions, avec mise en avant des points importants.
- **Génération de contenu** : rédaction de texte, synopsis ou du nouveau contenu pouvant servir d'ébauche ou d'amorce.

CAS D'USAGES

Règlement Général sur la Protection des Données (RGPD)

Règlement Général sur la Protection des Données (RGPD) :

en anglais « General Data Protection Regulation » ou GDPR, encadre le traitement des données personnelles sur le territoire de l'Union européenne.



Objectifs:

- Adapter le contexte juridique aux évolutions technologiques de la société
- Renforce le contrôle des citoyens sur l'utilisation de leurs données
- Harmoniser les règle en europe pour les entreprises afin de développer le numérique basé sur la confiance des utilisateurs



Qui est concerné:

Tout organisme, publique ou privé quels que soient sa taille, son pays d'implantation et son activité, qui traite des données personnelles ciblant des résidants européens ou étable sur le territoire de l'Union européenne.

Protection des données

Données personnelles:

Toute information se rapportant à une personne physique, identifiée ou identifiable, de manière direct ou indirect, à partir d'une seule donnée ou du croisement d'un ensemble de données.



Traitement de données personnelles :

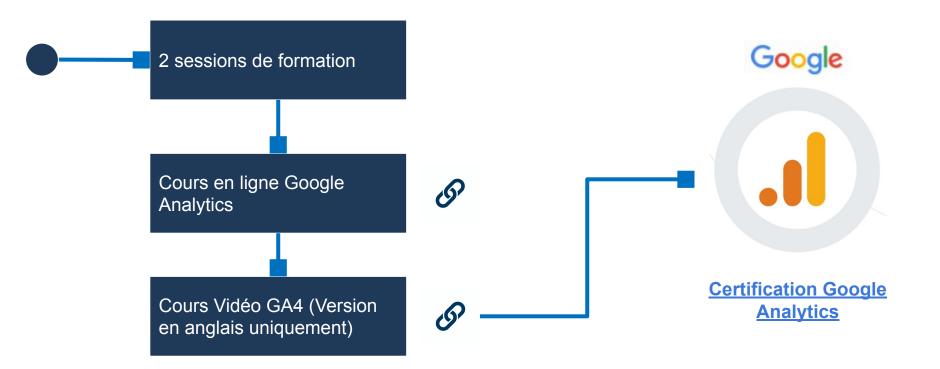
Opération ou ensemble d'opérations, informatisé ou physique, portant sur des données personnelles, quel que soit le procédé utilisé. Il doit avoir un objectif et une finalité légal et légitime vis à vis de l'activité professionnelle.



Les personnes morales ne sont pas concernées par la RGPD (coordonnée d'une entreprise, mail et téléphone standard).

Introduction à la certification Google Analytics

Certification Google Analytics



Information sur l'examen

Comment obtenir la certification?

- Obtenir un score avec un minimum de 80% de bonnes réponses
- Ne pas quitter la fenêtre de l'évaluation avant la fin

Contenu de la certification?

- 50 questions portant sur Google Analytics
- Impossible de revenir à la question précédente
- Répondre aux questions dans le temps imparti
- En cas d'échec, vous devez patienter 24 heures avant de repasser cette évaluation.

Durée de l'évaluation?

• 75 minutes pour 50 questions

Type de question?

QCM uniquement

Répondez à toutes les questions!

Introduction à Google Analytics

Comprendre le fonctionnement de Google Analytics et ses avantages

Définition du Digital Analytics

OBJECTIFS

Digital Analytics

Le digital analytics est le processus qui consiste à collecter, mesurer et analyser les données provenant de sources numériques telles que des sites Web ou des applications mobiles.



Analyser le comportement des utilisateurs



Améliorer la prise de décision



Évaluer l'efficacité de vos campagnes marketing



Améliorer la performance commerciale

Cycle de vie du client

ENTONNOIR MARKETING

ACQUISITION

Faire connaître votre produit/service et susciter l'intérêt des utilisateurs.

ENGAGEMENT

L'ensemble des interactions des utilisateurs avec votre entreprise.

MONETIZATION

Un utilisateur devient un client et effectue un achat.

RETENTION

Inciter un client à devenir un client récurrent

Identifier les canaux m clients sur un site Web/Ap

Identifier les canaux marketing qui attirent les clients sur un site Web/Application

OBJECTIFS

Déterminer les contenus avec le plus d'interaction

Savoir combien d'utilisateurs deviennent des clients

Connaître la fidélité d'un client sur le site Web/l'application de l'entreprise

ANALYTI ITAL, U

Impact de la promotion sur un site web

Sites référents : sites compilant des avis sur la boutique et les produits pour effectuer des achats sur le site web d'origine

Recherche : afficher le site web dans les résultats de différents moteurs de recherche

Réseaux sociaux: être présent sur plusieurs réseaux sociaux et fait la promotion des produits pour rediriger les clients vers le site web principal

Campagne par e-mail: communiquer sur les produits et promotions via email.

Annonces display payantes : diffuser des publicités sur ses produits auprès des clients lors de leurs navigation web

Annonces payantes sur le réseau de recherche : diffuser des publicités auprès des utilisateurs qui recherchent des produits sur des moteurs de recherche

Audience: ensemble des personnes qui visitent un site web

- se mesure via différents indicateurs (nombre de visiteurs uniques, le nombre de pages vues, le nombre de visites, la durée moyenne des visites)
- segmenter les utilisateurs par dimension, métrique et événement en fonction de votre entreprise.

Événement: interaction mesurable d'un utilisateur avec un site Web ou une application.

Événement clé: interaction importante et intermédiaire entre une évènement et une conversion.

Conversion: activités des utilisateurs considérés comme une réussite pour l'entreprise (achat d'un produit, téléchargement d'un document, inscription à un évènement).

Les conversions constituent un sous-ensemble des événements qui sont les événements les plus intéressants pour une entreprise.

Présentation Google Analytics

Evolution de Google Analytics



Google Analytics Classic

Rachat Urchin par Google et développement d'un outil d'analyse statistique



Google Analytics 4 (GA4)

Refonte de la collecte et du traitement des données

1995

2013



Création Urchin

Recueil de données statistiques Web par Urchin Software Corporation



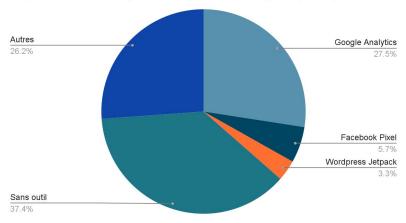
Google Analytics Universal (GA3)

Refonte et nouvelles fonctionnalités



Part de marché Google Analytics et revenu Google 2023





Usage Statistics and Market Share of Traffic Analysis Tools for Websites, April 2024

Financial Results

Revenues

The following table presents revenues by type (in millions):

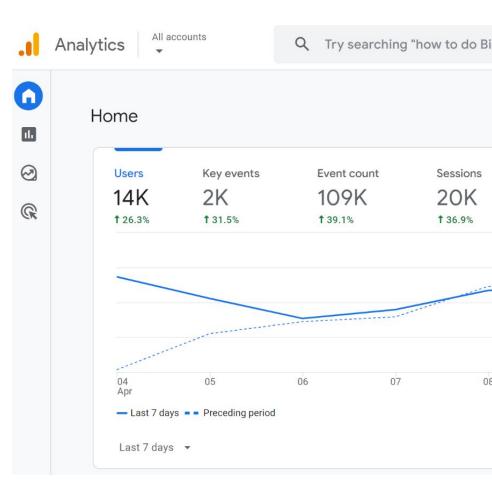
	Year Ended December 31,			
	2022		2023	
Google Search & other	\$ 162,450	\$	175,033	
YouTube ads	29,243		31,510	
Google Network	32,780		31,312	
Google advertising	224,473		237,855	
Google subscriptions, platforms, and devices	29,055		34,688	
Google Services total	253,528	Me.	272,543	
Google Cloud	26,280		33,088	
Other Bets	1,068		1,527	
Hedging gains (losses)	1,960		236	
Total revenues	\$ 282,836	\$	307,394	

Alphabet Inc. 2023 Annual Report

Google Analytics



Google Analytics est une plate-forme de collecte de données afin d'analyser l'**audience** de sites Web et d'applications. L'outil permet de créer des rapports d'activités Web, d'analyse de parcours client et de performance marketing.



Avantages Google Analytics



- Mesurer les interactions tout au long du parcours client
- Optimiser les campagnes marketing grâce à l'intégration d'outils publicitaires et d'édition Google
- Anticiper le comportement des clients grâce au Machine Learning de Google
- Faciliter l'analyse de donnée et favoriser la collaboration sur les rapports d'analyses
- Exporter vos conversions Google Analytics vers Google Ads afin de créer des rapports sur les annonces et de définir des enchères.

Ecosystème Google Analytics

Search Console

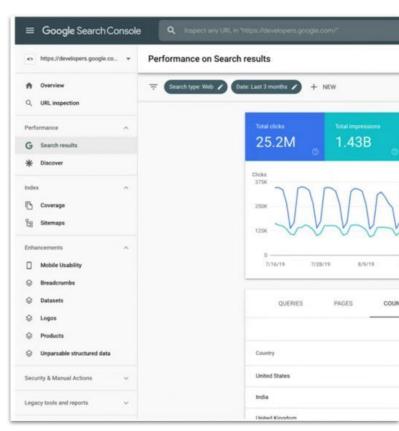


Search Console

Search Console est un outil Google d'optimisation de référencement et de performance pour un site Web.

Avantages:

- Recevoir une alerte et résoudre les problèmes sur votre site
- Optimiser le contenu Web
- Optimiser le classement dans les résultats de recherche
- Héberger votre contenu (sitemaps et URL) sur Google
- Déterminer comment la recherche Google accède à votre site
- Vérifier la bonne indexation de votre site Web



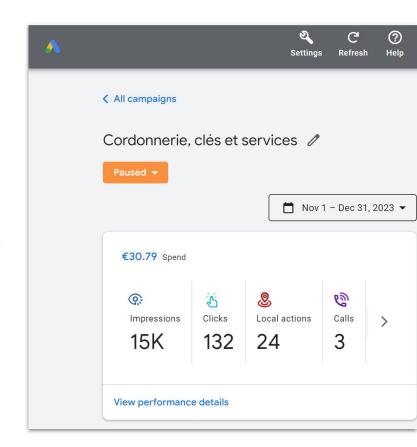
Google Ads



Google Ads est un programme publicitaire proposé par Google pour diffuser des campagnes publicitaires.

Avantages:

- Maximiser les prospects, les conversions et les ventes en lignes
- Améliorer la visibilité de l'entreprise
- Générer du trafic en magasin
- Améliorer le ROI et maîtriser votre budget



Looker Studio

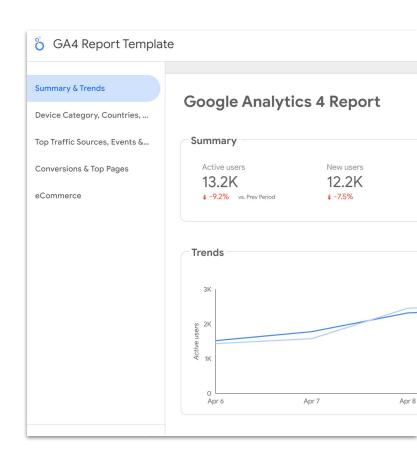


Looker Studio

Looker Studio est un outil Google de Tableau de bord et de visualisation de donnée.

Avantages:

- Créer des Tableaux de bord personnalisés
- Connecter de nombreuses sources de données
- Créer des modèles de rapports à partager



Structure Google Analytics

Gestion de balises d'un site web



Google Tag

Une balise est un petit extrait de code qui est placé sur toutes les pages de votre site Web afin de mesurer le comportement des utilisateurs ou les performances des annonces. Vous pouvez ajouter les balises directement dans le code de votre site Web.



Google Tag manager

Un système de gestion de balises permettant de piloter l'ensemble des Google Tag présent dans le site Web.



Google Tag Assistant

Un outil de débogage qui permet de vérifier la bonne implémentation des balises de votre site.

Gestion de balises d'une application





Kit de développement logiciel qui permet de collecter les données d'une application de la même manière qu'une balise Google le fait pour les sites Web.

On parle aussi de SDK (de l'anglais "software development kit").





Développement

Simplifie le développement d'application grâce à une infrastructure backend intégrée



Lancement et mesure

Sécurise le lancement et mesure la performance et la stabilité



Engagement

Augmente l'engagement utilisateur avec des analyses de données riches, A/B testing, et de messages de campagnes

Configuration des balises Google

Pour associer votre site Web ou une application à Google analytics, vous devez ajouter une balise Google ou le SDK Firebase pour créer un flux de donnée :

Guide Association Site Web

- Google Tag Manager pour ajouter une balise sans modifier le code d'un site web
- Ajouter manuellement une balise Google au site
- Indiquer un ID de mesure ou une balise Google dans l'outil de création de sites Web

Guide Association Application

- Depuis la console si l'application utilise déjà Firebase
- Créer un projet Firebase depuis Google Analytics



Créer un flux de données pour chaque plateforme d'application si celle-ci est disponible sous iOS et Android

Rapport Temps réel

Le rapport "Temps réel" collecte et affiche via Google Analytics l'activité en ligne des utilisateurs au cours des 30 dernières minutes.

Ce rapport vous permet aussi de vérifier la bonne installation d'une balise. A noter qu'il faut attendre jusqu'à 24h pour alimenter le rapport lors de l'installation de Google Analytics.



Tester le rapport en temps réel en vous connectant à votre site web

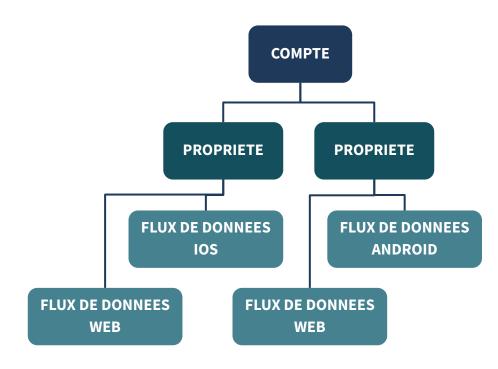
Architecture de Google Analytics



Un compte est unique avec des propriétés et des données liés à une entité.

Une propriété est un conteneur de rapports basé sur les données collectées d'utilisateurs. Google Analytics permet des associations avec des produits Google à ce niveau.

Un flux de données est une source de donnée Web ou applicative. Ce niveau permet de contrôler les données collectées.



Exemple d'architecture

Gestion de l'accès au compte - Rôles utilisateurs













- contrôle total sur le compte Analytics.
- gérer les utilisateurs*
- accorder des autorisations complètes à n'importe quel utilisateur, y compris à eux-mêmes.
- contrôle total sur les paramètres du compte et de ses propriétés.
- ne peuvent pas gérer les utilisateurs.
- créer, modifier et supprimer des audiences, des conversions, des modèles d'attribution, des événements et des périodes d'analyse.
- créer, modifier et supprimer certains éléments de propriété, comme les explorations.
- collaborer sur des éléments partagés.
- voir les paramètres et les données
- modifier les données des rapports**
- voir les nouveaux rapports et les nouvelles collections dans la navigation de gauche
- ne peuvent pas modifier la navigation.
- aucun rôle pour cette propriété ou ce compte
- avoir un rôle pour une propriété ou un compte associés.

Gestion de l'accès au compte - Créer des restrictions



Pas de métriques sur les revenus

 Les métriques sur les revenus générés par les achats des clients ne sont pas disponibles.

Pas de métriques sur les coûts

 Les métriques sur les coûts de la publicité en ligne et des investissements marketing ne sont pas disponibles.

Segmentation de l'audience

Afin de mieux comprendre les segments de votre base d'utilisateurs il est possible de combiner les dimensions répondent aux questions "qui", "quoi" ou "où" tandis que les métriques répondent à la question "combien"

Dimension

Une dimension est un libellé au format texte qui représente vos données (nom de pays, type d'appareil utilisé).

Propriétés utilisateur

Des attributs sur l'identité des utilisateurs de votre application ou de votre site Web (emplacement géographique ou appareil utilisé, par exemple).

Métrique

Une métrique est une valeur numérique qui représente vos données. (total d'utilisateurs, d'événements et de conversions, et revenu).

Paramètres d'événement

Des données supplémentaires liées aux événements pour préciser l'action effectuée ou contextualiser l'événement (nom de la vidéo regardée ou durée de visionnage)

Technique d'analyse avancée

Conversions modélisées

Conversions modélisées

Prédire grâce au machine learning les conversions en ligne qui ne peuvent pas être observées directement. Le modèle recherche des tendances entre les conversions directement observées ou non

Pourquoi?

- Renforcement des contrôles sur la collecte de données
- Demande des utilisateurs de contrôle et de transparence de la collecte et de l'utilisation de leurs données
- Evolution de la réglementation dans le monde visant à mieux encadrer la collecte et l'utilisation des données

Analyser un sous-ensemble de conversions dissocié des annonces



- Attribuer les conversions avec précision sans identifier les utilisateurs
- Pas d'authentification autorisée des utilisateurs individuels
- Regroupement des données afin d'estimer la probabilité que des conversions se produisent

Optimiser les campagnes publicitaires et améliorer les enchères automatiques





Permet de recouvrer la perte d'observabilité et éviter la surestimation

Modélisation et gestion du changement

- Séparation de la modélisation plus générale afin de représenter les activités et les comportements des clients
- La validation par réservation, une bonne pratique de machine learning, garantit la précision des modèles Google.
- Les conversions modélisées sont comparées aux conversions observées qui ont été réservées, et les informations ainsi obtenues permettent d'affiner les modèles.
- Nous communiquons les changements qui peuvent avoir une incidence majeure sur les données.





Attribution

Modèles multicanaux

Attribution : action de répartir le crédit de chaque conversion entre les différents clics, annonces et autres facteurs impliqués sur le chemin emprunté par l'utilisateur avant la conversion.

Modèle d'attribution : peut être une règle, un ensemble de règles ou un algorithme basé sur les données qui détermine la façon dont le crédit des conversions est attribué aux points de contact sur les chemins de conversion



Dernier Clic

Attribue la conversion au dernier canal avec lequel l'utilisateur a interagi avant la conversion



Premier Clic

Attribue la conversion au premier canal avec lequel l'utilisateur a interagi avant la conversion



Dépréciation dans le temps

Attribue davantage de crédit aux points de contact qui sont chronologiquement les plus proches de la conversion



Linéaire

la conversion est réparti de manière égale entre tous les canaux



Basé sur la position

Attribue 40 % du crédit aux première et dernière interactions, et les 20 % restants sont répartis uniformément entre les interactions intermédiaires.



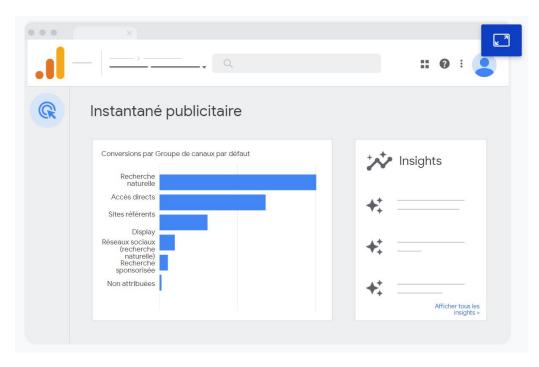
Basé sur les données

Répartit le crédit de la conversion en fonction des données observées pour chaque type de conversion. Calcule la contribution réelle de chaque clic à l'aide des données de votre compte.

L'espace de travail Publicité

L'espace de travail Publicité fournit des analyses supplémentaires sur les parcours utilisateur. Les rapports de cette section vous aident à mieux comprendre l'impact de des actions marketing sur différents canaux.

- Consultez rapidement vos performances publicitaires.
- Le rapport "Instantané publicitaire" vous permet de vérifier rapidement les métriques

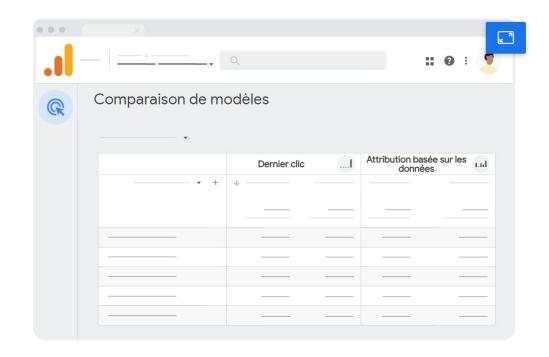


Comparaison de modèles

Comparaison de modèles indique aux annonceurs comment les différents modèles d'attribution attribuent le crédit aux canaux, sources, supports et campagnes.

- Dans le menu déroulant en haut à droite de la page, sélectionnez une plage de dates.
- Dans le menu déroulant en haut à gauche, sélectionnez une ou plusieurs conversions.

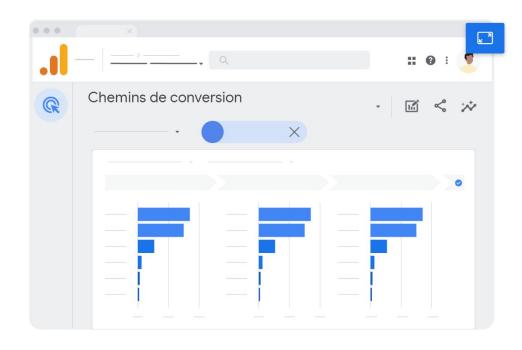
Par défaut, tous les événements de conversion sont sélectionnés et regroupés dans le rapport.



Rapport Chemins de conversion

Le rapport "**Chemins de conversion**" comporte deux sections :

- une visualisation des données et un tableau de données : permet de voir rapidement sur quels canaux la conversion démarre, se poursuit et se conclut.
- le tableau de données : affiche les chemins empruntés par les utilisateurs qui ont réalisé des conversions, ainsi que les métriques suivantes : "Conversions", "Revenus issus des achats", "Nombre de jours avant une conversion" et "Points de contact avant la conversion".





Comparez le modèle basé sur le dernier clic multicanal au modèle basé sur le premier clic multicanal afin d'identifier les campagnes sous-évaluées qui amènent les utilisateurs sur le chemin de conversion. Cette approche est utile si vous souhaitez attirer davantage de nouveaux clients sur un site Web ou dans une application.

Evaluation

Travail de groupe

- Par groupe de 4 personnes
- 10 min de présentation
- Ecrit : document (5 pages max)
- Oral: support visuel (powerpoint ou autres)

Situation : vous êtes consultant.e pour un cabinet, vous devez conseiller l'entreprise pour mettre en place une stratégie marketing.

Phase 1: comprendre l'entreprise

- 1.a. Présentation de l'entreprise
- 1.b. Comment collecter la donnée de l'entreprise : canaux de vente (magasin, site internet, caisse, etc.)

Phase 2 : mettre en place la stratégie :

- 2.a.Comment les données collectées peuvent aider l'entreprise à mieux performer ? (meilleur connaissance client, création d'un nouveau produit, etc.)
- 2.b.Comment l'intelligence artificielle peut aider l'entreprise à améliorer sa stratégie marketing?