



Développement et déploiement d'un outil de management de la brand
safety dans DV360 avec Google Apps script et Google Sheet &
Prédictions du nombre d'impressions d'une campagne
GROUPM

Mamadou DIOUF

Master 1 Mathématiques et Interactions
Août 2022

Remerciements

Je tiens à remercier toutes les personnes qui ont contribué à la réussite de mon stage.

Premièrement l'entreprise GroupM qui m'a fait confiance pour ce poste et l'ensemble du groupe Programmatique Buying Unit(PBU). Notamment les membres du Programmatique Intelligence qui sont Antoine Beillan, Louis Dreyfus et Aidan Demello.

Ensuite, les personnes qui ont contribué à la réussite des projets plus particulièrement celui sur les automatisations : le Programmatique Intelligence, Alexandre Chanthavong PBU, Russell Müller GroupM UK et Philipp Brunner Xaxis Allemagne.

Enfin nos très chers enseignants de la primaire jusqu'à maintenant et spécialement ceux du master mathématiques et interactions de l'Université d'Evry - Paris Saclay.

Résumé

Avec l'explosion du numérique et ses nombres calculables de sites, d'applications etc. Certains d'entre eux sont inappropriés pour faire de la publicité de certains marques tandis d'autres sont très convoités pour publier des contenus. Dans le cadre de stage, nous développons un outil d'automatisation qui sert alimenter les listes des sites, applications ou mots clés d'exclusions et d'inclusions dans Display & vidéo 360 une plateforme de Google, pour gérer la sécurité des marques (Brand safety). Il y a aussi une deuxième partie sur la prédiction du nombre d'impressions d'une campagne.

Table des matières

1	Contexte du stage	7
1.1	Présentation de l'entreprise	7
1.1.1	Présentation générale de l'entreprise	7
1.1.2	Présentation du service	7
1.1.3	Missions du service	7
1.1.4	Objectif du stage	8
2	Automatisation des processus en brand safety	9
2.1	Brand Safety	9
2.1.1	Brand safety en programmation	9
2.1.2	Le contenu est roi	9
2.1.3	Meilleures pratiques	10
2.2	Présentations des outils	10
2.2.1	Google Cloud Platform	10
2.2.2	BigQuery	11
2.2.3	Google drive	11
2.2.4	Display & Video 360	11
2.2.5	Google Apps script	11
2.2.6	Google sheets	11
2.3	Définitions	12
2.3.1	Web services	12
2.3.2	URI	12
2.3.3	API	13
2.3.4	REST	13
2.4	Les méthodes HTTP	14
2.5	Apps script project	14
2.5.1	Fichiers manifestes	14
2.5.2	Champs d'applications des autorisations (Scopes)	15
2.5.3	Vérification du client OAuth	16
2.5.4	Créer et gérer des déploiements	16
2.5.5	Projets Google Cloud	17
2.5.6	Bibliothèques	19
2.6	Applications	21
2.6.1	Configuration du Gsheet	22
2.6.2	Pré-développement de l'apps script	26
2.6.3	Développement des fonctions de bases avec les API DV360	27
2.6.4	Développement des fonctions du Google sheet	31
2.6.5	Erreurs rencontrées et leurs solutions	32
2.6.6	Déploiement de l'outil	34
3	Apprentissage statistique	36
3.1	Apprentissage supervisé	36
3.1.1	Méthodes de régression	37
3.1.2	Régression linéaire	37
3.1.3	Gradient boosting	39
3.1.4	Gradient tree boosting	41
3.1.5	Régularisation	42

3.2	Applications	43
3.2.1	Présentation des données	43
3.2.2	Transformations des données	44
3.2.3	Construction de modèle de régression	44
Bibliographie		47

Chapitre 1

Contexte du stage

1.1 Présentation de l'entreprise

1.1.1 Présentation générale de l'entreprise

En 1971 est créée l'entreprise Wire & Plastic Products, un fabricant de paniers en fils de fer et en plastique. Elle est renommée WPP Group en 1985.

Dans la première décennie des années 2000, elle devient la plus importante agence de communication en terme chiffre d'affaires. Aujourd'hui, c'est le plus important réseau d'agences de publicité et de communication mondial qui emploie environ 107 000 personnes dans 3 000 bureaux à travers 112 pays. Parmi ses agences on compte GroupM.

GroupM est la première société d'investissement dans les médias au monde avec pour mission de créer une nouvelle ère de médias où la publicité fonctionne mieux pour les gens.

Responsable de plus de 60 milliards de dollars d'investissements annuels dans les médias, selon COMvergence, l'entreprise innove, se différencie et génère une valeur durable pour ses clients, où qu'ils exercent leurs activités. Le portefeuille de GroupM comprend les agences Mindshare, Wavemaker, EssenceMediacom et mSix & Partners, ainsi que Choreograph (Data & Technology), GroupM Nexus (Cross-Channel Performance & Activation) et GroupM Investment.

En France, les 1000 collaborateurs, répartis dans les 5 agences et structures transversales du groupe, accompagnent les clients dans leur stratégie d'investissement média, jusqu'à l'achat d'espaces publicitaires, mais aussi dans leur stratégie de croissance et de différenciation auprès de publics de plus en plus exigeants.

Notre portefeuille de clients nous donne l'opportunité de développer des expertises sectorielles très fortes et de déployer depuis Paris de grands hubs internationaux sur le luxe ou l'automobile par exemple.

Notre credo d'excellence marketing nous amène à développer des technologies digitales ultra-performantes, grâce à des investissements importants réalisés par le groupe et en étroite collaboration avec tous les acteurs majeurs du marché : Google, Facebook, Amazon, mais aussi les grands groupes de presse ou de télévision.

Pour nos collaborateurs, c'est l'assurance d'être confrontés à des problématiques complexes, qui nécessitent le travail en commun de nombreuses expertises, et de progresser chaque jour.

1.1.2 Présentation du service

Le programmatique est un modèle d'achat d'espace média automatisé, impression par impression, offrant la possibilité de cibler précisément des utilisateurs afin de leur adresser le bon message, au bon moment, au prix juste tout en permettant un contrôle complet sur la diffusion.

Au sein de GroupM France, le programmatique est opéré par la structure PBU, Programmatique Buying Unit, qui est composé de deux groupes celui des médias traders ou accounts managers (AM) qui gère le côté trading et celui de Programmatique intelligence (PI) qui assure les côtés technologie et data.

1.1.3 Missions du service

Le média trader est en charge de l'achat et de la gestion globale d'espaces publicitaires en temps réel (Real Time Bidding) pour un portefeuille d'annonceurs (des marques ou des grands groupes comme Mars, LVMH, Subway etc).

L'équipe de programmation intelligence a pour mission de gérer la base de données générée par ces campagnes pour créer des outils d'analyse de données et d'automatisation de processus mais aussi des modèles d'apprentissage statistique afin de faciliter le quotidien du média trader.

1.1.4 Objectif du stage

L'objectif de stage est de créer un outil d'automatisation pour automatiser la création et l'alimentation des canaux au niveau du plateforme Google Display & Video 360, qui est utilisée pour acheter et diffuser de la publicité programmatique, pour les listes d'exclusions ou inclusions des sites et mots clés. Mais aussi une exploration des données de campagnes pour trouver une perspective de machine learning.

Chapitre 2

Automatisation des processus en brand safety

2.1 Brand Safety

GroupM définit la brand safety ou sécurité de la marque comme tout risque auquel un annonceur peut être confronté dans la chaîne d’approvisionnement numérique.

Il y a trois catégories principales de risques : financière, réputationnelle et juridique.

Dans la catégorie financière nous avons la visibilité, fraude, démo et third party tracking (suivi tiers) (collectage de données), dans la réputationnelle : environnement de contenu, expérience utilisateur et dans la juridique : confidentialité des consommateurs, anti-piratage, termes et conditions.

2.1.1 Brand safety en programmation

Pour le Brand Safety dans le programmation nous avons les risques financière et réputationnelle. La brand safety contextuelle en programmation n’a pas pris autant de chaleur que la brand safety sur les plateformes sociales. Néanmoins, les problèmes sont les mêmes : piratage en ligne, discours de haine, désinformation en ligne, terrorisme et contenus obscènes.

Contrairement aux plates-formes sociales, le programmation offre davantage de tactiques d’atténuation, principalement par le biais d’une vérification par un tiers. Malgré ce fait, l’utilisation de la vérification par un tiers et du blocage des pré-enchères reste facultative pour de nombreux annonceurs, ce qui peut entraîner des contiguités inappropriées sur les sites Web à longue traîne généralement achetés via des bourses d’annonces et des réseaux publicitaires.

2.1.2 Le contenu est roi

La sécurité de la marque entraîne une plus grande demande de contenu de qualité et premium ou l’achat de médias dans des environnements contrôlés. Les marchés de confiance (Trust Market Place TMP) sont déjà à la mode en Europe, et nous verrons cette tendance culminer dans les 6 à 12 prochains mois à travers le monde. L’inventaire proposé via TMPS sera, par défaut, vérifié et approuvé par un tiers, et contextuellement sûr ainsi que conforme au contrat.

Outre l’inventaire de qualité disponible via TMPS, le marché voit déjà des éditeurs former des conglomérats et vendre de la qualité liée à des données de première partie. Ces coentreprises d’éditeurs sont susceptibles d’attirer davantage de dépenses publicitaires car elles offrent un contenu journalistique de haute qualité mélangé à des audiences appropriées. Le développement et la certification indépendante de technologies capables d’effectuer les mesures appropriées nécessaires pour faire confiance à ces médias seront d’une grande importance. À leur tour, les entreprises de mesure traditionnelles opérant dans le domaine de la mesure de la sécurité des marques sur le Web / les applications devront intensifier leur jeu en ce qui concerne le déploiement d’un apprentissage automatique plus efficace qui sera capable de digérer et de bloquer efficacement le contenu linguistiquement nuancé, commencer à analyser le visuel et l’audio dans contenu vidéo, et aller plus loin au-delà des métadonnées et de l’analyse du positionnement publicitaire.

2.1.3 Meilleures pratiques

Chaque annonceur doit définir le type de contenu adapté à sa marque. Selon leurs préférences et leur appréciation, différentes tactiques sont recommandées. Voici les meilleures pratiques tactiques de GroupM pour éviter une contiguïté inappropriée de la sécurité des marques dans le programmatique :

Commerce direct

Lorsque nous négocions des médias, nous préférons acheter des médias programmatiques directement auprès de propriétaires de médias de haute qualité et de confiance, et éviter les réseaux publicitaires et les échanges publicitaires non transparents.

S'il est impossible d'acheter directement des médias, nous exigeons l'utilisation de listes d'inclusion, de listes d'exclusion et d'une technologie indépendante vérifiant si le contenu des pages et des applications est sûr.

La sécurité de la marque entraîne une plus grande demande de contenu de qualité et premium ou l'achat de médias dans des environnements contrôlés.

Protéger contractuellement

GroupM vise à convenir de conditions contractuelles de sécurité de la marque qui protègent les intérêts de nos clients. Ces conditions peuvent stipuler des procédures de retrait et le type de contenu adapté à la contiguïté des marques.

Déployer la technologie de vérification 3D-PARTY disponible

GroupM utilise la technologie de deux manières : pour informer la liste d'exclusion globale de GroupM (plus de 300 000 applications et sites Web) et les listes d'inclusion ou d'exclusion personnalisées par le client, et pour limiter ou bloquer l'endroit où les publicités apparaissent au point de livraison.

Établir des processus opérationnels claires

Nous avons défini des processus opérationnels pour surveiller ou contrôler l'inventaire des propriétaires de médias afin de s'assurer qu'il respecte nos propres normes de sécurité ou celles de nos clients.

Travailler ensemble (avec l'industrie)

La coopération de l'industrie est d'une importance incroyable. Les organismes industriels du monde entier établissent des programmes et accréditent souvent (à la suite d'un audit indépendant) les entreprises pour leur conformité aux lignes directrices sur les meilleures pratiques établies par l'industrie. Nous collaborons avec plusieurs initiatives commerciales, notamment le Joint Industry Committee for Web Standards (JICWEBS) au Royaume-Uni et en Irlande, le Trustworthy Accountability Group (TAG), le Media Rating Council (MRC) et le China Media Assessment Council (CMAC) pour aider à développer des normes industrielles et les meilleures pratiques. L'équipe de sécurité de la marque GroupM siège au conseil d'administration de ces institutions et est certifiée à la fois par JICWEBS et TAG. D'autres initiatives notables sont Digital Ad Trust en France, l'IAB Quality Index en Italie et la Digital Trust Initiative en Allemagne.

Fournir de l'éducation

Éduquer le personnel, les clients et le public sur les risques et sur la manière de les atténuer.

2.2 Présentations des outils

2.2.1 Google Cloud Platform

Google Cloud Platform (GCP) est une plateforme de cloud computing fournie par Google, proposant un hébergement sur la même infrastructure que celle que Google utilise en interne pour des produits tels que son moteur de recherche¹. Cloud Platform fournit aux développeurs des produits permettant de construire une gamme de programmes allant de simples sites web à des applications complexes.

Google Cloud Platform fait partie d'un ensemble de solutions pour les entreprises appelé Google Cloud, et fournit des services modulaires basés sur le cloud, tels que le stockage d'informations, le calcul, des applications de traduction et de prévision, etc.

Elle est composée d'une famille de produits, chacun comportant une interface web, un outil de lignes de commande, et une interface de programmation applicative REST((representational state transfer) comme BigQuery un entrepôt de données à très grande échelle.

2.2.2 BigQuery

BigQuery est un service web RESTful intégré à Google Cloud qui permet l'analyse interactive massive de grands ensembles de données en collaboration avec l'espace de stockage Google. C'est un logiciel en tant que service (SaaS) qui peut être utilisé en complément de MapReduce.

BigQuery est également un entrepôt de données d'entreprise de Google, en mode sans serveur, donc sans infrastructure à gérer. Les requêtes peuvent être écrites en SQL legacy ou en SQL standard. Cet outil google permet d'analyser les données situées dans un entrepôt logique. Un entrepôt logique contient des datasets (bases de données). Les datasets contiennent des tables ou des vues. Le service est gratuit jusqu'à 1 To de données analysées par mois et 10 Go de données stockées.

2.2.3 Google drive

Google Drive ou Google Disque au Canada est un service de stockage et de partage de fichiers dans le cloud lancé par la société Google. Google Drive, qui regroupe Google Docs, Sheets, Slides et Drawings, est une suite bureautique permettant de modifier des documents, des feuilles de calcul, des présentations, des dessins, des formulaires, etc.

Les utilisateurs peuvent rechercher les fichiers partagés publiquement sur Google Drive par l'entremise de moteurs de recherche Web. Il sert à synchroniser, partager et modifier les données entre plusieurs ordinateurs et/ou utilisateurs.

2.2.4 Display & Video 360

Display & Video 360 est un outil de Google Market Platform unique pour la planification des campagnes, la conception et la gestion des créations, l'organisation et l'exploitation des données d'audience, la recherche et l'achat d'inventaires, ainsi que la mesure et l'optimisation des campagnes.

Display & Video 360 vous permet d'effectuer les opérations suivantes :

Collaborer entre équipes : partagez les données et les informations dont vous disposez sur les campagnes avec les équipes multimédia, publicitaire, en charge des données et des données analytiques en vue d'une collaboration interdisciplinaire.

Améliorer la transparence et le contrôle : découvrez précisément comment votre budget est dépensé ainsi que l'emplacement exact de diffusion de vos annonces, et ce, pour l'ensemble de vos campagnes.

Obtenir des informations plus détaillées, plus rapidement : la fonctionnalité de veille intégrée du produit permet d'automatiser des tâches telles que la définition et l'optimisation des enchères, ainsi que de présenter automatiquement les informations et recommandations qui vous permettront de prendre rapidement les mesures nécessaires en fonction de vos objectifs.

2.2.5 Google Apps script

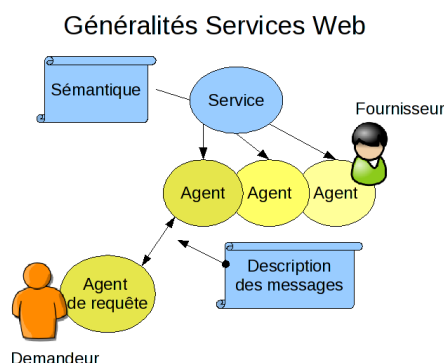
Google Apps Script est un langage de programmation propre à Google. Il permet de réaliser des scripts à l'intérieur de certains services de Google, dont Google Apps1. Ainsi il est possible de modifier le comportement du logiciel tableur, Google Spreadsheets, à l'instar de VBA pour Microsoft Excel.

Il est très proche du JavaScript. Si ce dernier est habituellement utilisé pour modifier des pages web, Google Apps Script accède aux données et comportements des applications Google telles que Display & Video 360, Google Docs, Sheets ou Drive.

2.2.6 Google sheets

Google Sheets est un tableur en ligne qui permet de créer et de mettre en forme des feuilles de calcul, et de les modifier en collaboration avec d'autres personnes.

Apps Script inclut des API spéciales pour vous permettre de créer, lire et modifier Google Sheets par programmation. Apps Script peut interagir avec Google Sheets de deux manières : n'importe quel script peut créer ou modifier une feuille de calcul si l'utilisateur du script dispose des autorisations appropriées pour la feuille de calcul, et un script peut également être lié à une feuille de calcul, ce qui donne au script des capacités spéciales pour modifier l'interface utilisateur ou répondre lorsque la feuille de calcul est ouverte. Pour créer un script lié, sélectionnez Extensions > Apps Script dans Google Sheets.



Le service de feuille de calcul traite Google Sheets comme une grille, fonctionnant avec des tableaux à deux dimensions. Pour récupérer les données de la feuille de calcul, vous devez accéder à la feuille de calcul dans laquelle les données sont stockées, obtenir la plage de la feuille de calcul qui contient les données, puis obtenir les valeurs des cellules. Apps Script facilite l'accès aux données en lisant les données structurées dans la feuille de calcul et en créant des objets JavaScript pour celles-ci.

2.3 Définitions

2.3.1 Web services

Un web service est un protocole d'interface informatique de la famille des technologies web permettant la communication et l'échange de données entre applications et systèmes hétérogènes dans des environnements distribués. Il s'agit donc d'un ensemble de fonctionnalités exposées sur internet ou sur un intranet, par et pour des applications ou machines, sans intervention humaine, de manière synchrone ou asynchrone.

Le protocole de communication est défini dans le cadre de la norme SOAP (Simple Object Access Protocol), protocole d'échange d'information structurée dans l'implémentation de services web bâti sur XML, dans la signature du service exposé (Web Services Description Language, WSDL). Actuellement, le protocole de transport est essentiellement protocole de contrôle de transmissions (TCP) (via HTTP). Le concept a été précisé et mis en œuvre dans le cadre de Web Services Activity2, au W3C (World Wide Web Consortium), particulièrement avec le protocole SOAP. Associé avec les échanges de données informatisés (EDI), le consortium ebXML l'a utilisé pour automatiser des échanges entre entreprises. Cependant le concept s'enrichit avec l'approfondissement des notions de ressource et d'état, dans le cadre du modèle REST, et l'approfondissement de la notion de service, avec le modèle architecture orientée services.

Un service web, avec ses agents, fournisseur et utilisateur Très grandes généralités sur un service web.

Dans sa présentation la plus générale, un service web se concrétise par un agent, réalisé selon une technologie informatique précise par un fournisseur du service. Un demandeur, à l'aide d'un agent de requête, utilise ce service. Fournisseur et demandeur partagent une même sémantique du service web, tandis qu'agent et agent de requête partagent une même description du service pour coordonner les messages qu'ils échangent.

Il existe plusieurs technologies derrière le terme services web. Parmi elles, les services web de type representational state transfer (REST) exposent entièrement ces fonctionnalités comme un ensemble de ressources identifiables par un URI et accessibles par la syntaxe et la sémantique du protocole HTTP. Les services web de type REST sont donc basés sur l'architecture du web et ses standards de base : HTTP et URI

2.3.2 URI

Un URI, de l'anglais Uniform Resource Identifier, soit littéralement identifiant uniforme de ressource, est une courte chaîne de caractères identifiant une ressource sur un réseau (par exemple une ressource Web) physique ou abstraite, et dont la syntaxe respecte une norme d'Internet mise en place pour le World Wide Web.

Un URI peut être de type « locator » ou « name » ou les deux.

Un Uniform Resource Locator (URL) est un URI qui, outre le fait qu'il identifie une ressource sur un réseau, fournit les moyens d'agir sur une ressource ou d'obtenir une représentation de la ressource en décrivant son mode d'accès primaire ou « emplacement » réseau. Par exemple, l'URL `http://www.wikipedia.org/` est un URI qui identifie une ressource (page d'accueil Wikipédia) et implique qu'une représentation de cette ressource (une page HTML en caractères encodés) peut être obtenue via le protocole HTTP depuis un réseau hôte appelé `www.wikipedia.org`.

Un Uniform Resource Name (URN) est un URI qui identifie une ressource par son nom dans un espace de noms. Un URN peut être employé pour parler d'une ressource sans que cela préjuge de son emplacement ou de la manière de la référencer. Par exemple, l'URN `urn:isbn:0-395-36341-1` est un URI qui, étant un numéro de l'International Standard Book Number (ISBN), permet de faire référence à un livre, mais ne suggère ni où, ni comment en obtenir une copie réelle.

2.3.3 API

En informatique, une interface de programmation d'application ou interface de programmation applicative (souvent désignée par le terme API pour Application Programming Interface) est un ensemble normalisé de classes, de méthodes, de fonctions et de constantes qui sert de façade par laquelle un logiciel offre des services à d'autres logiciels. Elle est offerte par une bibliothèque logicielle ou un service web, le plus souvent accompagnée d'une description qui spécifie comment des programmes consommateurs peuvent se servir des fonctionnalités du programme fournisseur.

On parle d'API à partir du moment où une entité informatique cherche à agir avec ou sur un système tiers, et que cette interaction se fait de manière normalisée en respectant les contraintes d'accès définies par le système tiers. On dit alors que le système tiers « expose une API ».

À ce titre, des interactions aussi diverses que la signature d'une fonction, une URL, un RPC (appel de procédure à distance) ... sont parfois considérés comme des API (ou micro-API) à part entière.

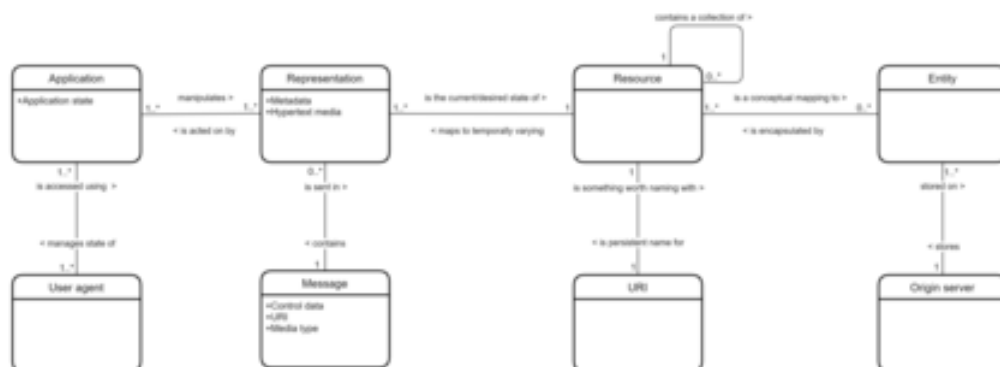
Dans l'industrie contemporaine du logiciel, les applications informatiques se servent de nombreuses interfaces de programmation, car la programmation réutilise des briques de fonctionnalités fournies par des logiciels tiers. Cette construction par assemblage nécessite pour le programmeur de connaître la manière d'interagir avec les autres logiciels, qui dépend de leur interface de programmation. Le programmeur n'a pas besoin de connaître les détails de la logique interne du logiciel tiers, et celle-ci n'est pas nécessairement documentée par le fournisseur. Seule l'API est réellement nécessaire pour utiliser le système tiers en question.

Des logiciels tels que les systèmes d'exploitation, les systèmes de gestion de base de données, les langages de programmation, ou les serveurs d'applications comportent une ou plusieurs interfaces de programmation.

2.3.4 REST

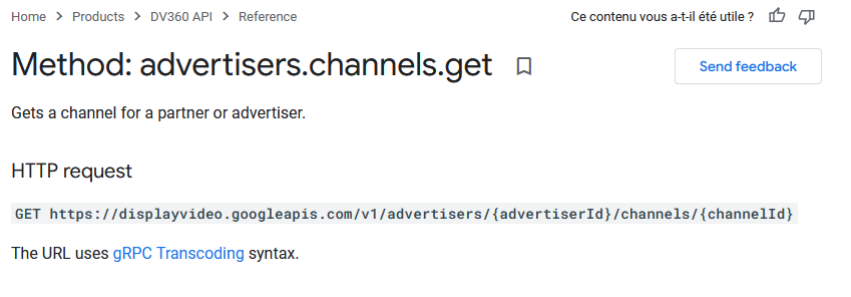
REST (representational state transfer) est un style d'architecture logicielle définissant un ensemble de contraintes à utiliser pour créer des services web. Les services web conformes au style d'architecture REST, aussi appelés services web RESTful, établissent une interopérabilité entre les ordinateurs sur Internet. Les services web REST permettent aux systèmes effectuant des requêtes de manipuler des ressources web via leurs représentations textuelles à travers un ensemble d'opérations uniformes et prédéfinies sans état. D'autres types de services web tels que les services web SOAP exposent leurs propres ensembles d'opérations arbitraires.

Les ressources web ont été définies pour la première fois sur le World Wide Web comme des documents ou des fichiers identifiés par leur URL. Cependant, elles ont aujourd'hui une définition beaucoup plus générique et abstraite qui inclut toute chose ou entité pouvant être identifiée, nommée, adressée ou gérée d'une façon quelconque sur le web. Dans un service web REST, les requêtes effectuées sur l'URI d'une ressource produisent une réponse dont le corps est formaté en HTML, XML, JSON ou un autre format. La réponse peut confirmer que la ressource stockée a été altérée et elle peut fournir des liens hypertextes vers d'autres ressources ou collection de ressources liées. Lorsque le protocole HTTP est utilisé, comme c'est souvent le cas, les méthodes HTTP disponibles sont GET, HEAD, POST, PUT, PATCH, DELETE, CONNECT, OPTIONS et TRACE.



2.4 Les méthodes HTTP

Dans le protocole HTTP, une méthode est une commande spécifiant un type de requête, c'est-à-dire qu'elle demande au serveur d'effectuer une action. En général l'action concerne une ressource identifiée par l'URL qui suit le nom de la méthode. Voici un exemple de méthode GET au niveau de DV360 :



Il existent plusieurs méthodes, celles que nous avons utilisées durant cet stage sont :

GET

La méthode HTTP GET, la plus utilisée, elle sert à récupérer les URI des ressources membres de la ressource collection ou membre dans le corps de la réponse sous forme HTML, XML ou JSON.

POST

La méthode HTTP POST crée une ressource membre dans la ressource collection ou membre en utilisant les instructions du corps de la requête. L'URI de la ressource membre créée est attribué automatiquement et retourné dans le champ d'en-tête Location de la réponse.

PATCH

Elle met à jour toutes les représentations des ressources membres de la ressource collection en utilisant les instructions du corps de la requête, ou crée éventuellement la ressource collection si elle n'existe pas.

De même s'il s'agit d'une ressource membre, elle met à jour toutes les représentations de la ressource membre, ou crée éventuellement la ressource membre si elle n'existe pas, en utilisant les instructions du corps de la requête

2.5 Apps script project

Un projet de script représente un ensemble de fichiers et de ressources dans Google Apps Script, parfois appelé simplement "un script". Un projet de script comporte un ou plusieurs fichiers de script qui peuvent être soit des fichiers de code (ayant une .gs extension), soit des fichiers HTML (une .html extension). Vous pouvez également inclure JavaScript et CSS dans les fichiers HTML.

L'éditeur de script a toujours un et un seul projet ouvert à un moment donné. Vous pouvez ouvrir plusieurs projets dans plusieurs fenêtres ou onglets de navigateur.

Nous pouvons créer un projet Apps Script à partir de Google Drive, Docs, Sheets, clasp command line tool, Slides ou Forms.

Pour la création avec un google sheet voici les étapes : Ouvrir un Sheets spreadsheet, Cliquer sur **Extensions** puis sur **Apps script** et Renommer le script éditeur en cliquant sur **Untitled project**

2.5.1 Fichiers manifestes

Un fichier manifeste de projet Apps Script est un fichier JSON spécial qui spécifie des informations de projet de base dont Apps Script a besoin pour exécuter le script avec succès.

Apps Script crée et met à jour automatiquement le manifeste du projet lorsque vous créez votre projet de script et apportez des modifications dans l'éditeur Apps Script. Dans la plupart des cas, vous n'avez jamais besoin d'afficher ou de modifier directement le manifeste ; cependant, dans certaines situations, cela peut être bénéfique ou nécessaire.

La structure du fichier manifeste et ses champs JSON sont décrits dans le guide de référence de la structure du manifeste .

2.5.2 Champs d'applications des autorisations (Scopes)

Les utilisateurs doivent autoriser les projets de script qui accèdent à leurs données ou agissent en leur nom. Lorsqu'un utilisateur exécute un script nécessitant une autorisation pour la première fois, l'interface utilisateur présente une invite pour démarrer le flux d'autorisation.

Au cours de ce flux, l'interface utilisateur indique à l'utilisateur ce que le script souhaite obtenir l'autorisation de faire. Par exemple, un script peut vouloir l'autorisation de lire les messages électroniques de l'utilisateur ou de créer des événements dans son calendrier. Le projet de script définit ces autorisations individuelles en tant que portées OAuth .

Pour la plupart des scripts, Apps Script détecte automatiquement les champs d'application dont vous avez besoin ; vous pouvez afficher les portées utilisées par un script à tout moment. Vous pouvez également définir des étendues explicitement dans votre manifeste à l' aide de chaînes d'URL. La définition explicite des étendues est parfois requise pour certaines applications telles que les modules complémentaires , car les applications publiées doivent toujours utiliser les étendues les plus étroites possibles.

Au cours du flux d'autorisation, Apps Script présente à l'utilisateur des descriptions lisibles par l'utilisateur des champs d'application requis. Par exemple, si votre script a besoin d'un accès en lecture seule à vos feuilles de calcul, le manifeste peut avoir la portée `https://www.googleapis.com/auth/spreadsheets.readonly`. Au cours du flux d'autorisation, un script avec cette portée demande à l'utilisateur d'autoriser cette application à "Afficher vos feuilles de calcul Google".

Certaines portées en incluent d'autres. Par exemple, lorsqu'il est autorisé, l'étendue `https://www.googleapis.com/auth/spreadsheets` autorise l'accès en lecture et en écriture aux feuilles de calcul.

Ajout de champs

Apps Script détermine automatiquement les étendues dont un script a besoin en analysant son code pour les appels de fonction qui en ont besoin. Pour la plupart des scripts, cela suffit et vous fait gagner du temps, mais pour les modules complémentaires publiés et les applications Web, vous devez exercer un contrôle plus direct sur les étendues.

Apps Script attribue parfois automatiquement aux projets des champs d'application très permissifs. Cela peut signifier que votre script demande à l'utilisateur plus que ce dont il a besoin, ce qui est une mauvaise pratique. Pour les scripts publiés, vous devez remplacer les portées larges par un ensemble plus limité qui couvre les besoins du script et rien de plus.

Vous pouvez définir explicitement les étendues utilisées par votre projet de script en modifiant son fichier manifeste . Le champ manifeste **oauthScopes** est un tableau de toutes les portées utilisées par le projet. Pour définir les champs d'application de votre projet, procédez comme suit :

1. Ouvrez le projet de script.
2. À gauche, cliquez sur **Paramètres du projet** ⚙️.
3. Cochez la **case Afficher le fichier manifeste "appsscript.json" dans l'éditeur**.
4. À gauche, cliquez sur **Éditeur** <>.
5. À gauche, cliquez sur le `appsscript.json` fichier.
6. Localisez le champ de niveau supérieur intitulé `oauthScopes`. S'il n'est pas présent, vous pouvez l'ajouter.
7. Le `oauthScopes` champ spécifie un tableau de chaînes. Pour définir les étendues utilisées par votre projet, remplacez le contenu de ce tableau par les étendues que vous souhaitez qu'il utilise. Par exemple:

```
{
  ...
  "oauthScopes": [
    "https://www.googleapis.com/auth/spreadsheets.readonly",
    "https://www.googleapis.com/auth/userinfo.email"
  ],
  ...
}
```

8. En haut, cliquez sur **Enregistrer** 💾.

2.5.3 Vérification du client OAuth

Les clients Google OAuth qui demandent certains champs d'application OAuth sensibles sont soumis à la validation de Google.

Si vous ne validez pas le client OAuth de votre projet de script, un écran d'application non validée s'affiche pour les utilisateurs externes à votre domaine lorsqu'ils tentent d'autoriser votre script. Un flux d'autorisation non validé permet à ces utilisateurs d'autoriser des applications non validées et de les utiliser, mais uniquement après avoir confirmé qu'ils ont compris les risques. Le nombre total d'utilisateurs d'application non validés est également plafonné.

La validation n'est pas requise pour les projets Apps Script dont le propriétaire et les utilisateurs appartiennent au même domaine ou client Google Workspace.

2.5.4 Créer et gérer des déploiements

Un déploiement de projet Apps Script est une version du script mise à disposition pour une utilisation en tant qu'application Web, module complémentaire ou exécutable d'API. En créant et en gérant des déploiements, vous pouvez itérer sur votre code, suivre vos modifications et contrôler la version de code exacte à laquelle vos utilisateurs ont accès.

Il existe deux types de déploiement :

Déploiements principaux , qui sont toujours synchronisés avec le code du projet actuel.

Déploiements versionnés , qui sont connectés à une version de projet spécifique.

Déploiements de tête

Un déploiement principal est le code de projet actuel. Lorsque vous créez un projet Apps Script, vous créez automatiquement un déploiement principal pour ce projet.

Le déploiement principal est toujours synchronisé avec le code le plus récemment enregistré. Par exemple, si vous créez un déploiement versionné, puis modifiez votre code, le déploiement principal reflète ces modifications, tandis que le déploiement versionné reste intact.

Utilisez les déploiements principaux pour tester le code. N'utilisez pas les déploiements principaux pour un usage public.

Déploiements versionnés

Un déploiement versionné met à disposition une version spécifique du code du projet. Cela permet à vos utilisateurs de continuer à utiliser une version fonctionnelle pendant que vous apportez des modifications et des améliorations au code.

Lorsque votre application est publiée pour une utilisation publique, utilisez toujours un déploiement versionné. Vous pouvez avoir plusieurs déploiements versionnés actifs en même temps.

Créer un déploiement versionné

Si vous souhaitez déployer votre script en tant que module complémentaire ou exécutable d'API, vous devez d'abord faire passer l'association Google Cloud Project de votre Apps Script du projet par défaut à un projet standard .

Pour créer un déploiement avec version, procédez comme suit :

Nouvel éditeur Éditeur hérité

1. Ouvrez le projet Apps Script.
2. En haut à droite, cliquez sur **Déployer** > > **Nouveau déploiement** .
3. À côté de **Sélectionner le type**, cliquez sur **Activer les types de déploiement** ⚙️.
4. Sélectionnez le type de déploiement que vous souhaitez déployer.
5. Entrez les informations sur votre déploiement et cliquez sur **Déployer** .

Chaque nouveau déploiement peut être partagé en tant que bibliothèque. Si vous partagez le script en tant que bibliothèque, la description du déploiement est visible pour les utilisateurs de la bibliothèque.

2.5.5 Projets Google Cloud

Chaque projet Apps Script utilise Google Cloud pour gérer les autorisations, les services avancés et d'autres détails. Pour configurer et gérer ces paramètres, chaque projet Apps Script est associé à un projet Google Cloud . Votre projet de script peut utiliser un projet par défaut créé automatiquement par Apps Script ou un projet standard que vous créez vous-même. En général, les projets par défaut conviennent aux scripts simples ou courants, mais vous devez utiliser un projet standard pour toute application complexe, de qualité commerciale ou destinée à la publication.

Vous pouvez passer d'un projet par défaut à un projet standard à tout moment, mais vous ne pouvez pas revenir en arrière pour utiliser un projet par défaut. Il est préférable de sélectionner le projet Google Cloud que votre script utilise au début du développement. Changer plus tard peut entraîner des complications, comme obliger vos utilisateurs à autoriser à nouveau.

Projets Google Cloud par défaut

Lorsque vous créez un projet Apps Script, Apps Script crée un projet Google Cloud par défaut qui fonctionne en arrière-plan.

Pour la plupart des scripts, vous n'avez jamais besoin de voir ou d'ajuster ce projet par défaut. Apps Script gère les interactions nécessaires avec Google Cloud. Par exemple, si vous activez un service avancé dans l'éditeur Apps Script, Apps Script active le service avancé dans le projet Google Cloud par défaut lorsque vous enregistrez le projet de script.

Pour certains scripts, vous devez interagir avec la console Google Cloud. Dans ces cas, votre script doit plutôt utiliser un projet Google Cloud standard. Par exemple, pour afficher les journaux Google Cloud dans la console Google Cloud, votre script doit utiliser un projet standard.

Projets Google Cloud standards

Les projets Google Cloud par défaut constituent la meilleure option pour la plupart des projets de script, sauf si vous devez configurer manuellement le projet. Dans ces situations, vous devez changer votre projet de script pour utiliser un projet standard .

Les sections suivantes décrivent quand Apps Script requiert un projet standard, les propriétés de ces projets et les tâches courantes effectuées avec eux. Vous ne pouvez effectuer les tâches ci-dessous qu'avec des projets standard.

Vous devez utiliser un projet standard dans les situations suivantes :

- Pour publier votre projet de script en tant que module complémentaire sur Google Workspace Marketplace.
- Pour vérifier le client OAuth de votre projet de script.
- Lorsqu'une application doit exécuter des fonctions dans votre projet de script à l'aide de la méthode de l'API Apps Scripts `scripts.run`.
- Pour afficher les journaux Google Cloud de votre projet de script dans la console Google Cloud. La console fournit davantage d'outils pour filtrer et afficher les journaux, et peut être plus utile que la vue simplifiée fournie par le tableau de bord Apps Script.
- Pour afficher les rapports d'erreurs de votre projet de script à l'aide du rapport d'erreurs.
- Pour créer une boîte de dialogue d'ouverture de fichier.
- Lorsque vous avez autrement besoin d'un contrôle manuel sur les paramètres Google Cloud du projet.

Propriétés standards du projet Google Cloud

Les projets standard ont les propriétés suivantes :

- Vous pouvez accéder à tous les paramètres Google Cloud du projet directement depuis la console Google Cloud. Cela vous permet d'activer les API, d'ajuster les identifiants d'autorisation et de configurer d'autres détails.
- Lorsque vous supprimez un projet de script ou que vous le basculez pour utiliser un autre projet standard, le projet standard d'origine reste et peut être réutilisé.
- Lorsque vous activez un service avancé dans un projet de script, vous devez activer manuellement l'API correspondante dans le projet standard.
- Plusieurs projets de script et d'autres applications peuvent partager le même projet standard. Si vous avez l'intention de publier un projet de script sur Google Workspace Marketplace en tant que module complémentaire, il doit avoir son propre projet standard. Les applications publiées ne peuvent pas partager de projets Google Cloud avec d'autres applications.
- Si vous souhaitez exécuter des fonctions dans un projet de script à partir d'une autre application à l'aide de la méthode de l'API Apps Scripts `scripts.run`, le projet de script et l'application appelante doivent partager le même projet standard.
- Lorsque Apps Script demande à un utilisateur d'autoriser un script qui utilise un projet standard, le nom du projet Google Cloud est utilisé pour identifier le script (et non le nom du projet de script). Pour cette raison, assurez-vous de définir un nom de projet Google Cloud approprié.

Accéder à un projet Google Cloud standard

Pour accéder au projet standard associé à votre projet de script, procédez comme suit :

Nouvel éditeur **Éditeur hérité**

1. Ouvrez le projet Apps Script.
2. À gauche, cliquez sur **Paramètres du projet** ⚙️.
3. Sous **Projet Google Cloud Platform (GCP)**, cliquez sur le numéro du projet.

Activer une API dans un projet Google Cloud standard

Une application Apps Script a souvent besoin d'accéder à une autre API Google. Pour cela, vous devez activer l'API dans le projet Google Cloud correspondant. Activez une API en procédant comme suit :

1. Ouvrez le projet Google Cloud.
2. Cliquez sur **Menu > API & Services**.

3. Cliquez sur **Activer les API et les services** .
4. Dans la zone de recherche, entrez l'API que vous souhaitez activer et appuyez sur **Entrée** .
5. Cliquez sur l'API dans les résultats de la recherche, puis cliquez sur **Activer** pour activer l'API pour ce projet Google Cloud.

Vous serez peut-être invité à accepter les conditions d'utilisation des API Google ou de Google Cloud . Lisez attentivement les conditions d'utilisation avant de les accepter.

Selon l'application, vous devrez peut-être également configurer l'API en la sélectionnant dans le tableau de bord **API et services** .

2.5.6 Bibliothèques

Une bibliothèque est un projet de script dont les fonctions peuvent être réutilisées dans d'autres scripts.

Pour inclure une bibliothèque dans votre projet, vous devez y avoir au moins accès au niveau de la vue. Si vous n'êtes pas l'auteur de la bibliothèque que vous souhaitez inclure, contactez-le et demandez-lui de vous accorder l'accès.

Vous avez besoin de l'ID de script de la bibliothèque que vous souhaitez inclure. Une fois que vous avez accès à la bibliothèque, vous pouvez trouver l'ID de script sur la page **Paramètres du projet** .

Créer et partager une bibliothèque

Pour utiliser et partager votre projet de script en tant que bibliothèque, procédez comme suit :

1. Créez un déploiement avec versions gérées de votre script.
2. Partagez au moins l'accès au niveau de la vue avec tous les utilisateurs potentiels de la bibliothèque.
3. Attribuez à ces utilisateurs l'ID de script, disponible sur la page **Paramètres du projet**

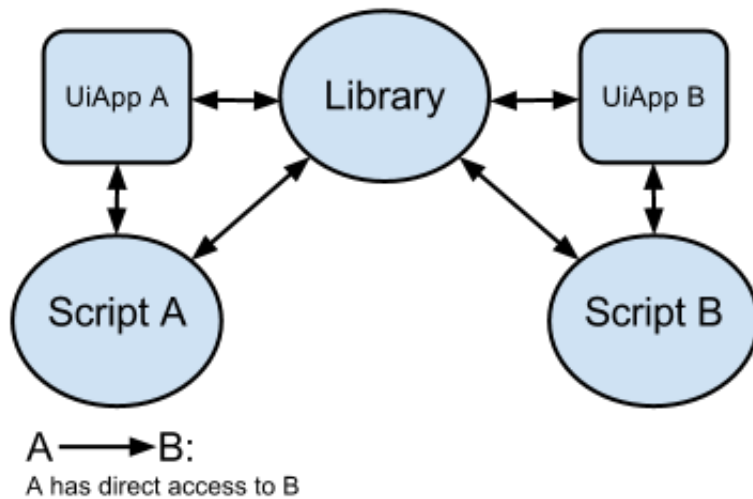
Bonnes pratiques

Voici quelques consignes à suivre lors de l'écriture d'une bibliothèque :

1. Choisissez un nom explicite pour votre projet, car il sera utilisé comme identifiant par défaut lorsque votre bibliothèque sera incluse par d'autres utilisateurs.
2. Si vous souhaitez que les utilisateurs de votre bibliothèque n'aient pas accès à une ou plusieurs méthodes de votre script, vous pouvez terminer le nom de la méthode par un trait de soulignement. Exemple : **myPrivateMethod_()**.
3. Seules les propriétés globales énumérables sont visibles pour les utilisateurs de la bibliothèque. Cela inclut les déclarations de fonction, les variables créées en dehors d'une fonction avec `var` et les propriétés définies explicitement sur l'objet global. Par exemple, **Object.defineProperty()** avec `enumerable` défini sur `false` crée un symbole que vous pouvez utiliser dans votre bibliothèque, mais vos utilisateurs ne peuvent pas y accéder.
4. Si vous souhaitez que les utilisateurs de votre bibliothèque utilisent la saisie semi-automatique de l'éditeur de scripts et la documentation générée automatiquement, vous devez disposer d'une documentation de type JSDoc pour toutes vos fonctions.

Portée des ressources

Il existe deux types de ressources lorsque vous utilisez des bibliothèques : partagées et non partagées. Une ressource partagée signifie que la bibliothèque et le script d'inclusion ont un accès intégré à la même instance de la ressource. Le schéma suivant illustre une ressource partagée utilisant l'exemple de propriétés utilisateur :

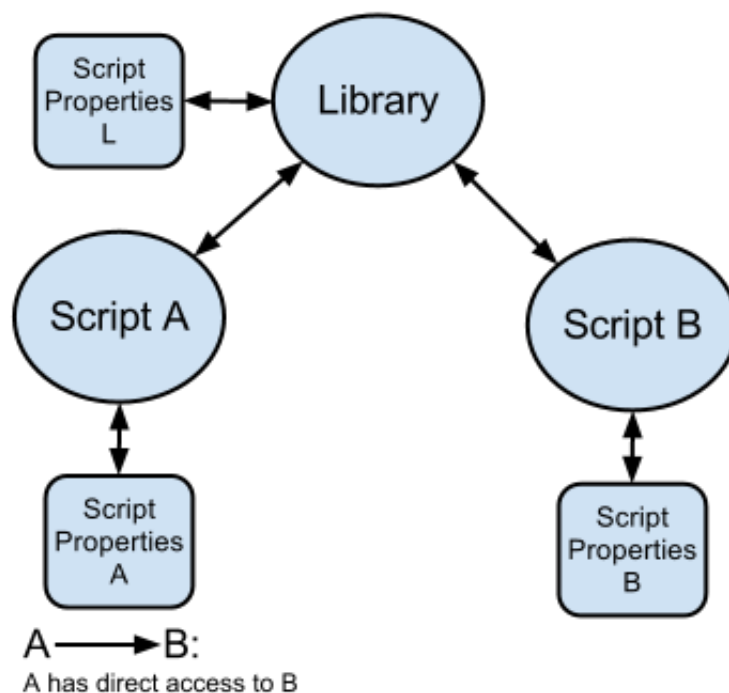


Une ressource non partagée signifie que la bibliothèque et le script inclus n'ont un accès intégré qu'à leur instance de la ressource. Cependant, une bibliothèque peut fournir l'accès à ses ressources non partagées en utilisant des fonctions explicites. Voici un exemple de fonction que vous pouvez inclure dans votre bibliothèque pour exposer ses propriétés de script :

```

function getLibraryProperty(key) {
  return ScriptProperties.getProperty(key);
}
  
```

Le schéma suivant illustre une ressource non partagée utilisant l'exemple de propriétés de script :



Ce tableau présente les ressources partagées et non partagées à titre d'information :

Ressource	Partagé*	Non partagé**	Remarques
Verrouiller		✓	La même instance est visible par tous les scripts, y compris lorsqu'ils sont créés dans la bibliothèque.
Propriétés du script		✓	La même instance est visible par tous les scripts, y compris lorsqu'ils sont créés dans la bibliothèque.
Cache		✓	La même instance est visible par tous les scripts, y compris lorsqu'ils sont créés dans la bibliothèque.
Triggers		✓	Les déclencheurs simples créés dans la bibliothèque ne sont pas déclenchés par le script "Inclure".
Application de script	✓		
UI	✓		
Propriétés utilisateur	✓		
Enregistreur et transcription d'exécution	✓		
Sites, feuilles de calcul et autres conteneurs	✓		Un appel à <code>getActive()</code> renvoie le conteneur du script "include".
MailApp et GmailApp	✓		
<p>* Cela signifie que la bibliothèque ne dispose pas de sa propre instance de la fonctionnalité/ressource, mais utilise celle créée par le script qui l'a appelée.</p> <p>** Cela signifie que la bibliothèque possède sa propre instance de la ressource/fonctionnalité, et que tous les scripts qui l'utilisent partagent la même instance.</p>			

2.6 Applications

Dans les meilleures pratiques en brand safety, sécurité de la marque, nous avons évoqué le commerce direct avec des listes d'exclusions et d'inclusions. En général, nous avons des listes d'exclusions de sites, d'applications ou mots clés très large. Dans notre cas nous avons une liste d'exclusions de sites avec plus 650 milles urls. Pour les ajoutés à un seul partenaire ou annonceur au niveau de DV360, qui est ici notre technologie de vérification de thirty party, ça prend deux voir trois heures.

L'idée est d'utiliser les API de DV360 sur apps script (comme back end) pour créer des API que nous appellerons dans google sheet (comme front end) afin manager les listes d'exclusions et d'inclusions sur DV360.

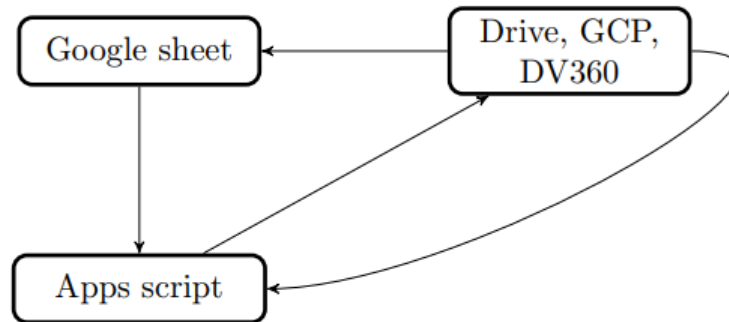


Diagramme d'utilisation des plate-formes.

2.6.1 Configuration du Gsheet

La première étape consiste à créer un spreadsheet et configurer l'ensemble des sheets, table et les boutons qui nous permettrons d'appeler les fonctions de l'appscript que nous créerons avec **Extensions**. Et grâce aux méthodes GET, POST et PATCH, ces fonctions pourront récupérer, modifier et ajouter des éléments aux niveaux des sheets mais aussi de DV360 et Drive.

Dans notre configuration, nous avons sept sheets :

GUIDELINES

Dans ce sheet, nous présentons de manière générale les sheets, expliquons comment récupérer l'identification(id) d'un fichier drive. Nous devons y ajouter les identifications des fichiers docs qui sont utilisées pour la plupart des fonctions, ce sont les fichiers des urls d'exclusions, d'inclusions et des mots clés d'exclusions, mais aussi les partenaires qui nous voudrions omettre :

PBU-FR-BRAND SAFETY

File Edit View Insert Format Data Tools Extensions Help Last edit was seconds ago

100% 123 Cabin 11

groupm AUTOMATED LIST MANAGEMENT

The Automated List Management Tool allows you to easily manage

- Channels (URLS) at both partner & advertiser levels
- Negative Keyword Lists (KEYWORDS) only for advertiser level

GLOSSARY

- KEYWORDS** Sheet where you can perform the following actions: Update/Add/Delete on **Keyword Negative Lists**
- URLS** Sheet where you can perform the following actions: Update/Add/Delete on **Channel Urls lists**
- KEYWORD IDS** List all keyword negative lists **by advertiser**
- URL IDS** List all channel urls lists (exclusion & inclusion) at both **partner & advertiser level**
- HISTORIC KEYWORDS** Find historic changes made for **keywords** management
- HISTORIC URLS** Find historic changes made for **urls** management

ID FINDER

Please, copy paste the url of your file to extract the file ID

ID FINDER (Paste below)	FILE ID
https://drive.google.com/filed/TYnEcY9vkhCmh8jAXa9zIo-Feu5OVL_Anews?usp=sharing	TYnEcY9vkhCmh8jAXa9zIo-Feu5OVL_

RESOURCES

Please, paste special (VALUES ONLY) the file ID retrieved from the ID Finder (H25)

URL EXCLUSION SOURCE ID	URL INCLUSION SOURCE ID	KEYWORD SOURCE ID
1NedjISuRu5SPibvBwJwXB45SV0hmUh	1EKbtGFpWhtCufqLEpdWOS3TYT9wNlc	TYnEcY9vkhCmh8jAXa9zIo-Feu5OVL_

Add 1000 more rows at bottom.

GUIDELINES KEYWORDS URLS KEYWORDS_IDS URL_IDS HISTORIC URLS HISTORIC KEYWORDS

Add the Partner Names you are not interested in to the list below

SKIPPED PARTNERS

- GroupM France
- NesopOgilvy France
- Mindshare FR
- KR Wavemaker
- Maxus FR
- LVMH
- AXA_BE_WPP
- RedPulse Media
- MediaCom FR
- Ford - Headlight - FR
- MEC FR
- PSA Global - Citroen
- PSA Global - DS
- PSA Global - Peugeot
- Ultra France
- The & Partners FR
- NesopOgilvy France - Nespresso
- Bayer EMEA DV360 FR
- TP - MAIF - CCN - DV360 - FR
- BAYER FRANCE
- AXA_IE_WPP
- AXA_ES_WPP
- AXA_CH_WPP
- AXA_ID_WPP
- AXA_MX_WPP
- AXA_TH_WPP
- AXA_PH_WPP
- AXA_FR_WPP
- Carrefour Media

GUIDELINES

KEYWORDS, KEYWORDS_IDS & HISTORIC KEYWORDS

The sheet **KEYWORDS** contient toutes tâches que nous voudrions effectuer avec les mots clés, création ou mises à jour des chaînes ou listes, ajouter ou supprimer des mots clés dans les chaînes, renommer les chaînes au niveau de DV360 et la mise à jour du fichier source des mots clés au niveau de Drive. Chacune de ces actions est effectuer grâce un bouton qui appelle la fonction qui lui est assigné dans apps script.

Pour alimenter les tables des identifications des partenaires, annonceurs et de leurs chaînes, nous utilisons le sheet **KEYWORDS_IDS** d'où sont stockées les identifications.

PBU-FR-BRAND SAFETY

File Edit View Insert Format Data Tools Extensions Help

Last edit was 26 minutes ago

100%

€ % 0 123 Cabin

11

B I A

	A	B	C	D	E	F	G	H	I	
1	Disclaimer: Please do not touch this tab									
2	You can exclude some partners from the listing by adding the Partner name in the tab "SKIPPED PARTNERS" in the GUIDELINES sheet (Col N)									
3	Status	Partner ID	Partner	Advertiser	Advertiser ID	NegativeKeyWordsList ID 1	NegativeKeyWordsList ID 2			
4	3696733	GRUPE MONDELEZ France	Autres Biscuits Lu	3763945	9153341873	9153341873	9144877794			
5	3696733	GRUPE MONDELEZ France	Belin	4149034	9147975874	9147975874	9153341675			
6	3696733	GRUPE MONDELEZ France	Belvita	3763932	9153342452	9153342452	9153342719			
7	3696733	GRUPE MONDELEZ France	Collection	3763944	9153342602	9153342602	9153343172			
8	3696733	GRUPE MONDELEZ France	Cookie de Lu	3763938	9153343241	9153343241	9147976426			
9	3696733	GRUPE MONDELEZ France	Côte d'Or	3763934	9147976465	9147976465	9144890190			
10	3696733	GRUPE MONDELEZ France	Cracotte	3763933	9147977320	9147977320	9153343637			
11	3696733	GRUPE MONDELEZ France	FR_Autres Biscuits Lu_HFSS	5674544	9153344339	9153344339	9153344408			
12	3696733	GRUPE MONDELEZ France	FR_Belin_HFSS	5674545	9147977491	9147977491	9144890454			
13	3696733	GRUPE MONDELEZ France	FR_Belvita_HFSS	5674546	9144892350	9144892350	9147977551			
14	3696733	GRUPE MONDELEZ France	FR_Collection_HFSS	5674547	9144892380	9144892380	9153344948			
15	3696733	GRUPE MONDELEZ France	FR_Cookie de Lu_HFSS	5674548	9144421041	9144421041	9147493879			
16	3696733	GRUPE MONDELEZ France	FR_Côte d'Or_HFSS	5674549	9144421128	9144421128	9144421134			
17	3696733	GRUPE MONDELEZ France	FR_Cracotte_HFSS	5674550	9147494137	9147494137	9144421344			
18	3696733	GRUPE MONDELEZ France	FR_Granola_HFSS	5674551	9144421353	9144421353	9144421356			
19	3696733	GRUPE MONDELEZ France	FR_Grany_HFSS	5743933	9152837468	9152837468	9147494458			
20	3696733	GRUPE MONDELEZ France	FR_Heudebert_HFSS	5767434	9152837837	9152837837	9144424728			
21	3696733	GRUPE MONDELEZ France	FR_Hollywood_HFSS	5743935	9144421392	9144421392	9144424836			
22	3696733	GRUPE MONDELEZ France	FR_LU CORE_HFSS	5675237	9152837930	9152837930	9147497695			
23	3696733	GRUPE MONDELEZ France	FR_Lu sans Gluten_HFSS	5674543	9152838068	9152838068	9147497977			
24	3696733	GRUPE MONDELEZ France	FR_Lulu_HFSS	5674542	9152842715	9152842715	9152842721			
25	3696733	GRUPE MONDELEZ France	FR_Mikado_HFSS	5730432	9147501301	9147501301	9144425685			
26	3696733	GRUPE MONDELEZ France	FR_Milka Biscuits_HFSS	5675235	9152842841	9152842841	9152842844			
27	3696733	GRUPE MONDELEZ France	FR_Milka_HFSS	5767435	9147501694	9147501694	9144425427			
28	3696733	GRUPE MONDELEZ France	FR_Napolitain_HFSS	5743931	9144425928	9144425928	9152843024			
29	3696733	GRUPE MONDELEZ France	FR_OREO_HFSS	5785831	9152842925	9152842925	9152843090			
30	3696733	GRUPE MONDELEZ France	FR_Patamlika_HFSS	5730431	9144425355	9144425355	9152843156			
31	3696733	GRUPE MONDELEZ France	FR_Pepito_HFSS	5743934	9152843288	9152843288	9144426234			
32	3696733	GRUPE MONDELEZ France	FR_Petit Ecolier_HFSS	5769531	9147502714	9147502714	9144426624			
33	3696733	GRUPE MONDELEZ France	FR_Philadelphia_HFSS	5769532	9147502192	9147502192	9152845058			
34	3696733	GRUPE MONDELEZ France	FR_Prince_HFSS	5675236	9147501850	9147501850	9147504799			
35	3696733	GRUPE MONDELEZ France	FR_TF FILIERES_HFSS	511309376	9144427416	9144427416	9144427347			
36	3696733	GRUPE MONDELEZ France	FR_Toblerone_HFSS	599730609	9152845700	9152845700	9144427482			
37	3696733	GRUPE MONDELEZ France	FR_Tuc_HFSS	5774334	9152845808	9152845808	9152845865			
38	3696733	GRUPE MONDELEZ France	Granola	3763935	9152845826	9152845826	9144429240			
39	3696733	GRUPE MONDELEZ France	Grany	3763936	9152848289	9152848289	9152845739			
40	3696733	GRUPE MONDELEZ France	Heubebert	3763937	9144429132	9144429132	9152850035			
41	3696733	GRUPE MONDELEZ France	Hollywood	3771635	9152849936	9152849936	9144429429			
42	3696733	GRUPE MONDELEZ France	LU CORE	4534331	9144429594	9144429594	9144429513			
43	3696733	GRUPE MONDELEZ France	Lu sans Gluten	3763939	9144429981	9144429981	9152850218			
44	3696733	GRUPE MONDELEZ France	Lulu	3763943	9152850326	9152850326	9152850266			
45	3696733	GRUPE MONDELEZ France	Mikado	3763946	9147509338	9147509338	9144429552			
46	3696733	GRUPE MONDELEZ France	Milka	3763947	9144430128	9144430128	9147510997			
47	3696733	GRUPE MONDELEZ France	Milka Biscuits	3763948	9147512572	9147512572	9147512578			
48	3696733	GRUPE MONDELEZ France	Napolitain	3763942	9144430308	9144430308	9152852129			
49	3696733	GRUPE MONDELEZ France	Oreo	3771634	9144430623	9144430623	9152852351			
50	3696733	GRUPE MONDELEZ France	Patamlika	3763949	9144430839	9144430839	9147507718			
51	3696733	GRUPE MONDELEZ France	Pepito	3763940	9144430728	9144430728	9144430887			

+

GUIDELINES

KEYWORDS

URLS

KEYWORDS_IDS

URL_IDS

HISTORIC URLS

HISTORIC KEYWORDS

KEYWORDS_IDS

KEYWORDS

Et l'historique des actions effectuées est ajouté dans le sheet **HISTORIC KEYWORDS**.

URLS, URL_IDS & HISTORIC URLS

Le sheet **URLS** contient toutes tâches que nous voudrions effectuer avec les urls, création ou mises à jour des chaînes ou listes, ajouter ou supprimer des mots clés dans les chaînes, renommer les chaînes au niveau de DV360 et la mise à jour du fichier source des urls d'exclusions ou d'inclusion au niveau de Drive. Chacune de ces actions est effectuer grâce un bouton qui appelle la fonction qui lui est assigné dans apps script.

Pour alimenter les tables des identifications des partenaires, annonceurs et de leurs chaînes, nous utilisons le sheet **URL_IDS** d'où sont stockées les identifications.

PBU- FR-BRAND SAFETY

File Edit View Insert Format Data Tools Extensions Help

Last edit was 2 hours ago

100%

€

0

123

Calc

11

URL_IDS

[illegible]

URLS

Et l'historique des actions effectuées est stocké dans le sheet **HISTORIC URLS**.

2.6.2 Pré-développement de l'apps script

Après cette configuration, nous allons ouvrir l'extension apps script du spreadsheet pour le développement des API et fonctions.

Services champs d'applications des autorisations

Avant tout, nous commençons par ajouter les autorisations dans le **appsscript.json** après l'avoir affiché comme dans **2.5.2**.

Voici l'ensemble des services Google Drive, BigQuery et Google sheets, et autorisations que nous aurons besoin :

```
1  {
2    "timeZone": "Europe/Paris",
3    "dependencies": {
4      "enabledAdvancedServices": [
5        {
6          "userSymbol": "Drive",
7          "serviceId": "drive",
8          "version": "v2"
9        },
10       {
11         "userSymbol": "BigQuery",
12         "serviceId": "bigquery",
13         "version": "v2"
14       },
15       {
16         "userSymbol": "Sheets",
17         "serviceId": "sheets",
18         "version": "v4"
19       }
20     ],
21     "libraries": [
22       {
23         "userSymbol": "GASRetry",
24         "libraryId": "1EugqzsWbWssuYne0wNZvNgeXqFfD8dPRMZM3PnpaV3SsiW5qUANNxhXL",
25         "version": "10"
26       },
27       {
28         "userSymbol": "OAuth2",
29         "libraryId": "1B7FSrk5Zi6L1rSxxTDgDEUsPzIukDsi4KGuTMorsTQHhGBzBkMun4i0F",
30         "version": "41",
31         "developmentMode": true
32       }
33     ]
34   },
35   "exceptionLogging": "STACKDRIVER",
36   "oauthScopes": [
37     "https://www.googleapis.com/auth/script.external_request",
38     "https://www.googleapis.com/auth/display-video",
39     "https://www.googleapis.com/auth/spreadsheets",
40     "https://www.googleapis.com/auth/bigquery",
41     "https://www.googleapis.com/auth/bigquery.insertdata",
42     "https://www.googleapis.com/auth/DoubleClickBidManager",
43     "https://www.googleapis.com/auth/drive",
44     "https://www.googleapis.com/auth/documents",
45     "https://www.googleapis.com/auth/script.scriptapp",
46     "https://www.googleapis.com/auth/userinfo.email",
47     "https://www.googleapis.com/auth/drive.file"
48   ],
49   "runtimeVersion": "V8"
50 }
```

Niveaux d'accès	Descriptions
https://www.googleapis.com/auth/script.external_request	Communiquer avec d'autres applications ou d'accéder à d'autres ressources sur le Web en récupérant des URL.
https://www.googleapis.com/auth/script.scriptapp	Créer des déclencheurs de script et de contrôler la publication du script en tant que service
https://www.googleapis.com/auth/display-video	Créer, consulter, modifier et supprimer définitivement vos entités et rapports Display & Video 360
https://www.googleapis.com/auth/spreadsheets	Consulter, modifier, créer et supprimer toutes vos feuilles de calcul Google Sheets
https://www.googleapis.com/auth/bigquery	Consulter et gérer vos données dans Google BigQuery et afficher l'adresse e-mail de votre compte Google
https://www.googleapis.com/auth/bigquery.insertdata	Insérer des données dans Google BigQuery
https://www.googleapis.com/auth/doubleclickbidmanager	Affichez et gérez vos rapports dans DoubleClick Bid Manager
https://www.googleapis.com/auth/drive	Voir, modifier, créer vos fichiers dans Google Drive, et tous les supprimer
https://www.googleapis.com/auth/drive.file	Consulter, modifier, créer et supprimer uniquement les fichiers Google Drive que vous utilisez avec cette application
https://www.googleapis.com/auth/documents	Consulter, modifier, créer et supprimer tous vos documents Google Docs
https://www.googleapis.com/auth/userinfo.email	Afficher votre adresse e-mail

Projet Google Cloud

Ensuite il nous faut ajouter l'identification(id) du projet cloud et activer les API des services Drive, BigQuery et Sheets dans le projet comme montrer dans **2.5.2**.

Bibliothèques

Enfin ajouter les bibliothèques qui seront utilisées dans les fonctions ou pour les vérifications des autorisations. Dans notre cas, nous utiliserons les bibliothèques :

- **OAuth 2.0** : est un protocole qui permet à un utilisateur d'accorder à un site Web tiers ou à une application l'accès aux ressources protégées de l'utilisateur, sans nécessairement révéler ses informations d'identification à long terme ni même son identité. OAuth introduit une couche d'autorisation et sépare le rôle du client de celui du propriétaire de la ressource. Dans OAuth, le client demande l'accès aux ressources contrôlées par le propriétaire de la ressource et hébergées par le serveur de ressources et reçoit un ensemble d'informations d'identification différent de celui du propriétaire de la ressource. Au lieu d'utiliser les informations d'identification du propriétaire de la ressource pour accéder aux ressources protégées, le client obtient un jeton d'accès, une chaîne indiquant une portée, une durée de vie et d'autres attributs d'accès spécifiques. Les jetons d'accès sont délivrés aux clients tiers par un serveur d'autorisation avec l'approbation du propriétaire de la ressource. Ensuite, le client utilise le jeton d'accès pour accéder aux ressources protégées hébergées par le serveur de ressources.
- **GASRetry** : Invoque une fonction, effectuant jusqu'à 5 tentatives avec une temporisation exponentielle. Réessaie avec des retards d'environ 1, 2, 4, 8 puis 16 secondes pour un total d'environ 32 secondes avant d'abandonner et de relancer la dernière erreur.

2.6.3 Développement des fonctions de bases avec les API DV360

Nous créons des scripts pour les fonctions qui permettront de créer ou changer les noms des canaux pour des partenaires et/ou annonceurs, de récupérer leurs identifications, remplacer leurs contenus, d'ajouter ou supprimer du contenu grâce aux méthodes et requêtes HTTP de DV360. Ces fonctions seront appelée dans les sheets.

Méthodes HTTP : GET, POST et PATCH

L'ensemble de nos actions nécessite que les méthodes GET, POST et PATCH, donc nous avons défini une fonction qui prends comme argument la méthode, l'url de la requête HTTP et le corps de la requête et effectuer

l'action de la méthode. Pour cela nous avons besoin de la classe **UrlFetchApp** d'Apps Script. Ce service permet aux scripts de communiquer avec d'autres applications ou d'accéder à d'autres ressources sur le Web en récupérant des URL. Un script peut utiliser le service URL Fetch pour émettre des requêtes HTTP et HTTPS et recevoir des réponses. Le service URL Fetch utilise l'infrastructure réseau de Google à des fins d'efficacité et d'évolutivité.

Ce service nécessite la `https://www.googleapis.com/auth/script.external_request` portée. Dans la plupart des cas, Apps Script détecte et inclut automatiquement les étendues dont un script a besoin, mais si vous définissez explicitement vos étendues, vous devez ajouter manuellement cette étendue à utiliser `UrlFetchApp`. Dans cette classe nous utiliserons sa méthode **fetch(url,params)** :

Paramètres

Nom	Type	La description
<code>url</code>	<code>String</code>	URL à récupérer.
<code>params</code>	<code>Object</code>	L'objet JavaScript facultatif spécifiant les paramètres avancés tels que définis ci-dessous.

Paramètres avancés

Nom	Type	La description
<code>contentType</code>	<code>String</code>	le type de contenu (par défaut, 'application/x-www-form-urlencoded'). Un autre exemple de type de contenu est 'application/xml'; jeu de caractères=utf-8'.
<code>headers</code>	<code>Object</code>	un mappage clé/valeur JavaScript des en-têtes HTTP pour la requête
<code>method</code>	<code>String</code>	la méthode HTTP de la requête : <code>get</code> , <code>delete</code> , <code>patch</code> , <code>post</code> , ou <code>put</code> . La valeur par défaut est <code>get</code> .
<code>payload</code>	<code>String</code>	la charge utile (c'est-à-dire le corps POST) de la requête. Certaines méthodes HTTP (par exemple, GET) n'acceptent pas de charge utile. Il peut s'agir d'une chaîne, d'un tableau d'octets, d'un blob ou d'un objet JavaScript. Un objet JavaScript est interprété comme une correspondance entre les noms de champs de formulaire et les valeurs, où les valeurs peuvent être des chaînes ou des blobs.
<code>useIntranet</code>	<code>Boolean</code>	Obsolète. Cela demande à fetch de résoudre l'URL spécifiée dans l'intranet lié à votre domaine via (obsolète) SDC
<code>validateHttpsCertificates</code>	<code>Boolean</code>	Si <code>false</code> la récupération ignore les certificats non valides pour les requêtes HTTPS. La valeur par défaut est <code>true</code> .
<code>followRedirects</code>	<code>Boolean</code>	Si <code>false</code> la récupération ne suit pas automatiquement les redirections HTTP ; il renvoie la réponse HTTP d'origine. La valeur par défaut est <code>true</code> .
<code>muteHttpExceptions</code>	<code>Boolean</code>	Si <code>true</code> la récupération ne lève pas d'exception si le code de réponse indique un échec et renvoie à la place le fichier HttpResponse . La valeur par défaut est <code>false</code> .
<code>escaping</code>	<code>Boolean</code>	Si <code>false</code> les caractères réservés dans l'URL ne sont pas échappés. La valeur par défaut est <code>true</code> .

UrlFetchApp.fetch paramètres

Mais qu'avec les paramètres avancés, **headers**, **method** et **muteHttpExceptions** pour la méthode **GET**, nous y ajoutons **accept**, **contentType** et **payload** pour les autres méthodes. Dans headers, nous mettons le jeton d'accès OAuth 2.0 pour l'utilisateur effectif récupérer par la méthode **getOAuthToken** de la classe **ScriptApp** et nous le corps de la requête en JSON string.

```

1  /** DV360 API FUNCTION */
2  /**
3   * @param {string} method : the method GET, POST or PATCH
4   * @param {string} url : Api's url in dv360
5   * @param {object} body : the body that we want to POST or PATCH
6   */
7  function dv360_get_post_patch(method,url,body={}){
8
9      const token = ScriptApp.getOAuthToken();
10     const headers = {
11         Authorization: 'Bearer ' + token,
12     };
13     let options = {
14         headers: headers,
15         method: method,
16         muteHttpExceptions: true
17     };
18     if (method !== "GET") {
19         Object.assign(options, {
20             accept: "application/json",
21             contentType: "application/json",
22             payload: JSON.stringify(body),
23         });
24     }
25
26     const response = UrlFetchApp.fetch(url, options);
27     // Logger.log(response)
28     if (response.getResponseCode() !== 200) {
29         throw new Error(response.getContentText());
30     }
31
32     if (method === "GET") {
33         return response
34     }
35 }
36 }
37

```

DV360 méthodes GET, POST et PATCH

Si la réponse est différent de 200, c'est à dire OK, nous renvoyons l'erreur. Sinon, si la méthode est GET elle retourne la réponse.

Récupérations des identifications des partenaires et annonceurs

Avec la méthode **GET** et l'url de la requête HTTP de DV360 en y ajoutant un rangement suivant le "diplay-Name", nous récupérons la réponse en JSON string que nous convertissons en dictionnaire afin de récupérer les identifiants et les noms des partenaires. Nous faisons de même pour annonceurs mais avec un url qui nécessite l'identifiant du partenaire concerné, les annonceurs sont classés dans les partenaires.

- <https://displayvideo.googleapis.com/v1/partners> pour le partenaire et
- <https://displayvideo.googleapis.com/v1/advertisers?partnerId> pour les annonceurs

Création des canaux dans un partenaire ou annonceurs

Pour la création des canaux, nous utilisons la méthode **POST**, le corps de la requête est un dictionnaire qui a pour clé "**displayName**" et pour valeur le nom du canal. La requête HTTP est :

Method: partners.channels.create

[Send feedback](#)

Creates a new channel. Returns the newly created channel if successful.

HTTP request

POST <https://displayvideo.googleapis.com/v1/partners/{partnerId}/channels>

The URL uses [gRPC Transcoding](#) syntax.

Path parameters

Parameters	
partnerId	string (int64 format) The ID of the partner that owns the created channel.

Method : *partners.channels.create* DV360

OU Method: advertisers.channels.create

[Send feedback](#)

Creates a new channel. Returns the newly created channel if successful.

HTTP request

POST <https://displayvideo.googleapis.com/v1/advertisers/{advertiserId}/channels>

The URL uses [gRPC Transcoding](#) syntax.

Path parameters

Parameters	
advertiserId	string (int64 format) The ID of the advertiser that owns the created channel.

Method : *advertisers.channels.create* DV360

Nous avons créer une fonction qui prend comme argument une id, un nom, le nombre canaux que nous voulons et type de l'id soit partners ou advertisers et créer les canaux dans le partenaire ou annonceur.

Récupérer les identifiants des canaux

Avec la méthode **GET** et l'url de la requête HTTP de DV360 en y ajoutant , nous récupérerons la réponse en JSON string que nous convertirons en dictionnaire afin de récupérer les identifiants des canaux.

Requête HTTP :

- <https://displayvideo.googleapis.com/v1/partners/partnerId/channels> pour le partenaire et
- <https://displayvideo.googleapis.com/v1/advertisers/advertiserId/channels> pour les annonceurs

Création des sites des canaux

La requête HTTP de **bulkEdit** avec la méthode POST, nous permettent de créer des sites c'est à dire d'ajouter des urls dans les canaux. Pour cela nous aurons d'un corps de requête qui est un dictionnaire de clé **createdSites** et de valeur une liste de dictionnaire de clé **urlOAppId** avec comme valeur l'url. Cependant l'url doit en UTF-8 (Universal Character Set Transformation Format - 8 bits) .

Requête HTTP :

- <https://displayvideo.googleapis.com/v1/partners/partnerId/channels/channelId/sites:bulkEdit> pour le partenaire et
- <https://displayvideo.googleapis.com/v1/advertisers/advertiserId/channels/channelId/sites:bulkEdit> pour les annonceurs

Remplacer les sites d'un canal

La requête **replace**, elle est à peu près comme la création mais elle remplace ce qui est dans le canal par le corps de la requête même si le canal ou le corps est vide. Le corps est un dictionnaire de clé **newSites** et de valeur une liste de dictionnaire de clé name avec comme valeur nom et **urlOAppId** avec comme valeur l'url.

Requête HTTP :

- `https://displayvideo.googleapis.com/v1/partners/partnerId/channels/channelId/sites` :replace pour le partenaire et
- `https://displayvideo.googleapis.com/v1/advertisers/advertiserId/channels/channelId/sites` :replace pour les annonceurs

Supprimer des sites d'un canal

Elle est comme la création à différence près le corps qui est un dictionnaire de clé **deletedSites** et de valeur une liste d'urls.

Changer le nom d'un canal

Pour celle là, nous utilisons la méthode **PATCH** et le corps de la requête est un dictionnaire de clé **displayme** et de valeur le nouveau nom :

Requête HTTP :

- `https://displayvideo.googleapis.com/v1/partners/partnerId/channels/channelId?updateMask=diplayName` pour le partenaire et
- `https://displayvideo.googleapis.com/v1/advertisers/advertiserId/channels/channelId?updateMask=diplayName` pour les annonceurs

Remarque

Chaque corps d'une requête est crée par fonction à part et elles sont appelées dans méthodes, pour les méthodes **POST** et **PATCH**

2.6.4 Développement des fonctions du Google sheet

Comme les fonctionnements des sheets **URLS** et **KEYWORDS** sont les mêmes, nous développerons sur celui d'**URLS**. Nous avons sept boutons chacun associé à une fonction apps script :

Nous utiliserons les méthodes de la classe **SpreadsheetApp** pour accéder et modifier un sheet. Ces méthodes utilisées sont :

Méthodes	Descriptions
<code>appendRow(rowContents)</code>	Ajoute une ligne au bas de la région de données actuelle de la feuille avec comme contenu le "rowContents"
<code>getRange(a1Notation)</code>	Range Renvoie la plage spécifiée au format A1 ou R1C1.
<code>getActiveSheet()</code>	Récupère la feuille active dans une feuille de calcul.
<code>getLastRow()</code>	Renvoie la position de la dernière ligne comportant du contenu.
<code>getValue()</code>	Renvoie la valeur de la cellule en haut à gauche dans la plage.
<code>getValues()</code>	Renvoie la grille rectangulaire des valeurs de cette plage.
<code>getSheetByName(name)</code>	Affiche une feuille portant le nom indiqué.
<code>clearContent()</code>	Efface le contenu de la plage, sans modifier la mise en forme.
<code>setValue(value)</code>	Définit la valeur de la plage.
<code>setValues(values)</code>	Définit une grille rectangulaire de valeurs (doit correspondre aux dimensions de cette plage).

Et les méthodes de la classe **DriveApp** qui permet aux scripts de créer, de rechercher et de modifier des fichiers et des dossiers dans Google Drive. Les méthodes utilisées :

<code>getFileById(id)</code>	Récupère le fichier avec l'ID donné.
<code>setContent(content)</code>	Ce paramètre permet d'écraser le contenu du fichier par un élément de remplacement donné.

Lister les identifiants des partenaires ou annonceurs

Après avoir rempli les informations demandées dans **GUIDELINES** qui sont nécessaire pour les fonctions de cet sheet à voir les identifications des fichiers des urls d'exclusions et d'inclusions et les noms des partenaires que nous voudrions omettre, il faut lister les identifiants des canaux. Cette fonction récupère les identifiants des partenaires (non omis) depuis DV360, puis ceux de leurs canaux, suivant un nom, s'ils sont créés et les liste dans le sheet **URL_IDS**. Ensuite il suffit de sélectionner le ou les partenaires ou annonceurs et de cliquer sur l'action.

Remplacer ou remplir les canaux

Cette fonction récupère les identifiants des partenaires ou annonceurs sélectionnées et ceux de leurs canaux (s'ils ne sont pas nous les créons) pour remplacer leurs contenus par les éléments de la liste d'exclusions ou inclusions dans **Drive**. Dans notre cas, nous avons une liste d'exclusions de près 620 milles sites et une liste d'inclusions de 1200 sites, et comme chaque canal peut contenir que 65 milles sites. Donc nous créons dans dix canaux et dix sous listes d'exclusions et un canal d'inclusion.

Renommer les canaux

Elle récupère les identifiants des partenaires ou annonceurs sélectionnées et ceux de leurs canaux et les renomme.

Vérifier un url dans une liste

Elle récupère les urls sur la table **Check Urls** afin de vérifier s'ils sont dans la liste des urls du fichier.

Ajouter des nouveaux urls pour un partenaire ou annonceur

Elle ajoute les nouveaux urls qui sont placés dans la colonne **Add Urls** dans le dernier canal du partenaire ou annonceur.

Supprimer des urls pour un partenaire ou annonceur

Elle supprime les urls qui sont placés dans la colonne **Delete Urls** dans le dernier canal du partenaire ou annonceur.

Mise à jour du fichier des urls

Elle ajoute les nouveaux urls qui sont placés dans la colonne **Add Urls** et supprime les urls qui sont placés dans la colonne **Delete Urls** dans le fichier concerné inclusion ou exclusion au niveau de **Drive**.

2.6.5 Erreurs rencontrées et leurs solutions

Les principales erreurs rencontrées sont des erreurs d'API et de maximum de temps d'exécution d'une fonction :

Erreurs d'API

Code	RPC	Action recommandée
400	INVALID_ARGUMENT	Il y a un problème avec votre demande. Passez en revue le champ de message dans la réponse d'erreur et modifiez votre demande en conséquence.
401	UNAUTHENTICATED	Votre demande n'a pas pu être correctement authentifiée. Vérifiez que vous incluez des informations d'identification OAuth valides dans votre demande. Pour plus d'informations à ce sujet, veuillez consulter notre guide de démarrage .
403	PERMISSION_DENIED	Vous n'avez pas les autorisations nécessaires pour compléter votre demande. Vérifiez que votre compte utilisateur dispose des autorisations appropriées pour le partenaire ou l'annonceur concerné.
404	NOT_FOUND	La ressource que vous tentez de récupérer est introuvable. Vérifiez que l'ID que vous utilisez est correct et correspond au type de ressource récupéré.
409	ABORTED	La ressource que vous tentez de modifier est en cours de modification ailleurs. Attendez quelques secondes, puis réessayez la demande.
429	RESOURCE_EXHAUSTED	<p>Vous avez dépassé votre limite de débit ou votre quota de demandes quotidiennes. Examinez votre utilisation de l'API à l'aide de la console d'API Google et modifiez votre flux de travail pour qu'il fonctionne dans les limites de quota existantes. Pour plus d'informations, veuillez consulter notre page sur l'optimisation des quotas.</p> <p>Si vos opérations ne peuvent pas être effectuées dans le cadre de votre quota quotidien, vous pouvez demander un quota supplémentaire.</p>
500	INTERNAL	L'API a rencontré une erreur interne. Attendez quelques secondes, puis réessayez la demande. Si l'erreur persiste, contactez l'assistance .
504	DEADLINE_EXCEEDED	L'API a pris trop de temps pour traiter la demande. Attendez quelques secondes, puis réessayez la demande. Si l'erreur persiste, contactez l'assistance .

Temps d'exécution maximum

Les quotas sont définis à des niveaux différents pour les utilisateurs de comptes personnels (comme gmail.com) ou les comptes de l'édition offerte de G Suite (obsolète) et les comptes Google Workspace. Nous avons un quota de trente minutes d'exécution pour un script. Et pour nos fonctions de mise à jour si nous avons sélectionnés plus partenaires, nous dépassons ce temps d'exécution.

Pour contourner cette erreur nous stockons les propriétés de la fonction et crée un **trigger** qui une fois lancée utilisera ces propriétés.

Pour les propriétés, nous utilisons les méthodes de la classe **Properties** :

Méthodes

Méthode	Type renvoyé	Brève description
<code>deleteAllProperties()</code>	Properties	Supprime toutes les propriétés du magasin Properties actuel.
<code>deleteProperty(key)</code>	Properties	Supprime la propriété avec la clé donnée dans le magasin Properties actuel.
<code>getKeys()</code>	String[]	Récupère toutes les clés dans le magasin Properties actuel.
<code>getProperties()</code>	Object	Récupère une copie de toutes les paires clé/valeur du datastore Properties actuel.
<code>getProperty(key)</code>	String	Récupère la valeur associée à la clé donnée dans le magasin Properties actuel ou null s'il n'existe aucune clé de ce type.
<code>setProperty(properties)</code>	Properties	Définit toutes les paires clé/valeur de l'objet donné dans le magasin Properties actuel.
<code>setProperty(properties, deleteAllOthers)</code>	Properties	Définit toutes les paires clé/valeur de l'objet donné dans le datastore Properties actuel, et éventuellement toutes les autres propriétés du magasin.
<code>setProperty(key, value)</code>	Properties	Définit la paire clé-valeur indiquée dans le magasin Properties actuel.

Pour créer et supprimer des déclencheurs de manière automatisée à l'aide du service de script. Nous utilisons la méthode **ScriptApp.newTrigger**, exemple :

```
/**
 * Creates two time-driven triggers.
 * @see https://developers.google.com/apps-script/guides/triggers/installable#time-driven_triggers
 */
function createTimeDrivenTriggers() {
  // Trigger every 6 hours.
  ScriptApp.newTrigger('myFunction')
    .timeBased()
    .everyHours(6)
    .create();
  // Trigger every Monday at 09:00.
  ScriptApp.newTrigger('myFunction')
    .timeBased()
    .onWeekDay(ScriptApp.WeekDay.MONDAY)
    .atHour(9)
    .create();
}
```

2.6.6 Déploiement de l'outil

Avant tout, il faut déployer la version de l'apps script en bibliothèque comme montrer dans **2.5.4**. Ensuite il faut partager le google sheet du projet aux personnes dont nous souhaitons déployer l'outil en ajoutant leurs adresses mail. Nous les partageons ce projet en **lecteur**.

PBU-FR-BRAND SAFETY

File Edit View Insert Format Data Tools Extensions Help Last edit was 2 days ago

100% € % .0 .00 123 Cabin 24 B I A

CHANNEL URLS EXCLUSION/INCLUSION LIST MANAGEMENT

This sheet is to update your management list in DV360 and drive.
You can add, delete or check items in your channel urls or list and update channel urls lists names.

Select your type of management: **Exclusion**

UPDATE

Functions	Description
Run List channels Urls by Partner/Advertiser	List all Channel Urls Lists from GroupM EMEA if it exist (DV360) ex: AUTOMATIC_EXCLUSION-LIST_082022_N'01
Run Replace channels Urls values	Replace all Channel Urls Lists from GroupM EMEA if it exists. (DV360) Otherwise, it will create the Channel Urls Lists.
Run Update channel Urls names	Update the Channel Urls Lists names with current the current date (DV360) ex: AUTOMATIC_EXCLUSION-LIST_082022_N'01

Add urls like "name.domain"
 google.com OK
 www.google.com NOT OK
 https://www.google.com NOT OK

BULK
☒ Please use "Check urls" function to check if urls are
 You can add in bulk urls if you are 100% sure there
 The function will fail otherwise.

Level All Partner All Advertiser All

Timestamp	Status	ID DSP	Partner	Advertiser	Select
2022-08-17 18:20	REPLACE OK	3696733	GROUPE MONDELEZ France	GROUPE MONDELEZ France	<input type="checkbox"/>
		3763945	GROUPE MONDELEZ France	Autres Biscuits Lu	<input type="checkbox"/>
		4149034	GROUPE MONDELEZ France	Belin	<input type="checkbox"/>
		3763932	GROUPE MONDELEZ France	Belvita	<input type="checkbox"/>
		3763944	GROUPE MONDELEZ France	Collection	<input type="checkbox"/>

LIST_KEYWORDS NegativeKeywordLists GUIDELINES KEYWORDS URLS KEYWORDS_IDS URL_IDS HISTORIC URLS HISTORIC KEYWORDS

Share "PBU-FR-BRAND SAFETY"

Russell Miller X

☒ Notify people

Message

Viewer

☒ Viewer

Commenter

Editor

Copy link Cancel Send

Enfin, nous créons un nouveau spreadsheet en copiant le nôtre puis nous configurons son apps script en ajoutant notre bibliothèque, celle de **GASRetry** et celle des autorisations **OAuth2**, l'identifiant de notre projet apps script et récréer nos fonctions pour les boutons en utilisant la bibliothèque.

Chapitre 3

Apprentissage statistique

L'apprentissage statistique, ou apprentissage automatique, de l'anglais machine learning, est un champ d'étude de l'intelligence artificielle qui se fonde sur des approches mathématiques et statistiques pour donner aux ordinateurs la capacité d'« apprendre » à partir de données, c'est-à-dire d'améliorer leurs performances à résoudre des tâches sans être explicitement programmés pour chacune. Plus largement, il concerne la conception, l'analyse, l'optimisation, le développement et l'implémentation de telles méthodes.

L'apprentissage automatique comporte généralement deux phases. La première consiste à estimer un modèle à partir de données, appelées observations, qui sont disponibles et en nombre fini, lors de la phase de conception du système. L'estimation du modèle consiste à résoudre une tâche pratique, telle que traduire un discours, estimer une densité de probabilité, reconnaître la présence d'un chat dans une photographie ou participer à la conduite d'un véhicule autonome. Cette phase dite « d'apprentissage » ou « d'entraînement » est généralement réalisée préalablement à l'utilisation pratique du modèle. La seconde phase correspond à la mise en production : le modèle étant déterminé, de nouvelles données peuvent alors être soumises afin d'obtenir le résultat correspondant à la tâche souhaitée. En pratique, certains systèmes peuvent poursuivre leur apprentissage une fois en production, pour peu qu'ils aient un moyen d'obtenir un retour sur la qualité des résultats produits.

Selon les informations disponibles durant la phase d'apprentissage, l'apprentissage est qualifié de différentes manières. Si les données sont étiquetées (c'est-à-dire que la réponse à la tâche est connue pour ces données), il s'agit d'un apprentissage supervisé. On parle de classification ou de classement si les étiquettes sont discrètes, ou de régression si elles sont continues. Si le modèle est appris de manière incrémentale en fonction d'une récompense reçue par le programme pour chacune des actions entreprises, on parle d'apprentissage par renforcement. Dans le cas le plus général, sans étiquette, on cherche à déterminer la structure sous-jacente des données (qui peuvent être une densité de probabilité) et il s'agit alors d'apprentissage non supervisé. L'apprentissage automatique peut être appliqué à différents types de données, tels des graphes, des arbres, des courbes, ou plus simplement des vecteurs de caractéristiques, qui peuvent être des variables qualitatives ou quantitatives continues ou discrètes.

Nous pouvons avoir une analyse statistique qui correspond trop précisément à une collection particulière d'un ensemble de données, appelée sur-apprentissage, ou sur-ajustement (en anglais « overfitting »). Ainsi, cette analyse peut ne pas correspondre à des données supplémentaires ou ne pas prévoir de manière fiable les observations futures. Un modèle sur-ajusté est un modèle statistique qui contient plus de paramètres que ne peuvent le justifier les données.

Dans le cadre de notre stage nous avons étudié des cas d'apprentissage supervisé plus précisément des méthodes de régressions.

3.1 Apprentissage supervisé

Nous pouvons représenter l'ensemble des données étiquetées appelé aussi base de données d'apprentissage, ou ensemble d'apprentissage, comme un ensemble de couples entrée-sortie $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$, où \mathbf{x} représente l'entrée, \mathbf{y} le sortie et N le nombre de données d'apprentissage.

Chaque entrée x_i est un vecteur de réels de dimension finie \mathbf{d} , elles sont aussi appelées features(caractéristiques) et peuvent être très complexes comme une image, une série temporelle, graphe.

Les sorties peuvent aussi être très variées mais en général elles sont nominales ou catégoriques $y_i \in \{1, 2, \dots, C\}$, il s'agit d'un problème de classification ou des variables réelles, un problème de régression.

3.1.1 Méthodes de régression

La régression recouvre plusieurs méthodes d'analyse statistique permettant d'approcher une variable à partir d'autres qui lui sont corrélées. On distingue essentiellement deux cas selon la nature de la variable expliquée, représentée ici par une variable aléatoire \mathbf{y} . Les variables explicatives seront notées \mathbf{x} . Si certaines de ces colonnes sont qualitatives, il est parfois judicieux de vectoriser leurs modalités en distinguant une modalité de référence représentée par le vecteur nul, et en représentant les autres modalités par les vecteurs de base d'un espace euclidien. Sous certaines conditions, on peut aussi quantifier les modalités de ces variables.

Nous pouvons formaliser le problème comme celui d'une fonction d'approximation. Nous supposons que $y = f(x)$, f une fonction inconnue, et le but d'estimer cette fonction suivant les données d'entraînement. Puis de prédire \mathbf{y} par $\hat{\mathbf{y}} = \hat{\mathbf{f}}(\mathbf{x})$, \hat{f} est l'estimation de f . Le principal objectif est de généraliser la prédiction pour des nouvelles données d'entrées.

Notons la probabilité de la distribution des possibles sorties sachant un vecteur d'entrée \mathbf{x} et l'ensemble d'apprentissage \mathcal{D} par $p(y|x, \mathcal{D})$. Nous pouvons prédire \mathbf{y} par la sortie qui a la meilleure probabilité dans l'ensemble des sorties :

$$\hat{y} = \hat{f} = \operatorname{argmax}_c p(y = c | \mathbf{y}, \mathcal{D}) \quad (3.1)$$

Le principal problème réside sur l'estimation de la loi de distributions des la probabilité.

3.1.2 Régression linéaire

Elle est la modèle de régression la plus utilisé. Dans cette modèle, nous supposons que la sortie ou le label est une fonction linéaire des entrées :

$$y(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + \epsilon = \sum_{j=1}^d w_j x_j + \epsilon \quad (3.2)$$

où \mathbf{w}^T représente le produit scalaire entre le vecteur d'entrée \mathbf{x} et les poids du modèle \mathbf{w} , et ϵ l'erreur entre notre prédiction linéaire et la vraie variable expliquée.

Nous admettons souvent que ϵ suit une distribution normale ou gaussienne notée $\epsilon \sim \mathcal{N}(\mu, \sigma^2)$, avec μ comme moyenne et σ^2 comme variance. Nous pouvons donc écrire le modèle comme suit :

$$p(y|\mathbf{x}, \theta) = \mathcal{N}(y|\mu(\mathbf{x}), \sigma^2(\mathbf{x})) \quad (3.3)$$

Cette formule nous permet de dire que le modèle est une densité de probabilité conditionnelle. Dans le cas simple, nous supposons que μ est une fonction linéaire de \mathbf{x} et que la variance est fixe, $\mu = \mathbf{w}^T \mathbf{x}$ et $\sigma^2(x) = \sigma^2$, et $\theta = (\mathbf{w}, \sigma^2)$ sont les paramètres du modèle.

Par exemple, si l'entrée est de dimension 1, nous avons $\mu = w_0 + w_1 x = \mathbf{w}^T \mathbf{x}$, où w_0 l'intercepte et w_1 la pente et le vecteur $\mathbf{x} = (1, x)$.

La régression linéaire peut être faite pour modéliser des relations non linéaires en remplaçant \mathbf{x} par certains fonction non linéaire des entrées $\phi(\mathbf{x})$. Nous aurons alors

$$p(y|\mathbf{x}, \theta) = \mathcal{N}(y|\mathbf{w}^T \phi(\mathbf{x}), \sigma^2(\mathbf{x})) \quad (3.4)$$

Maximum de vraisemblance

Une manière courante d'estimer les paramètres d'un modèle statistique consiste à calculer l'estimation du maximum de vraisemblance (moindres carrés), en anglais maximum likelihood estimation (MLE), qui est défini comme :

$$\hat{\theta} = \arg \max_{\theta} \log p(\mathcal{D}|\theta) \quad (3.5)$$

Il est courant de supposer que les exemples de formation sont indépendants et distribués de manière identique, communément abrégé en iid. Cela signifie que nous pouvons écrire la log-vraisemblance comme suit :

$$l(\theta) = \log p(\mathcal{D}|\theta) = \sum_{i=1}^N \log p(y_i|\mathbf{x}_i, \theta) \quad (3.6)$$

Au lieu de maximiser la log-vraisemblance, nous pouvons de manière équivalente minimiser la log-vraisemblance négative, en anglais negative log likelihood (NLL) :

$$NLL(\theta) = - \sum_{i=1}^N \log p(y_i|\mathbf{x}_i, \theta) \quad (3.7)$$

La formulation NLL est parfois plus pratique, car de nombreux logiciels d'optimisation sont conçus pour trouver les minima des fonctions, plutôt que les maxima.

Appliquons maintenant la méthode de MLE au paramètre de régression linéaire. Insertion de la définition de la gaussienne dans ce qui précède, nous constatons que le log de vraisemblance est donné par :

$$l(\theta) = - \sum_{i=1}^N \log \left[\left(\frac{1}{2\pi\sigma^2} \right)^{\frac{1}{2}} \exp \left(-\frac{1}{2\sigma^2} (y_i - \mathbf{w}^T \mathbf{x}_i)^2 \right) \right] \quad (3.8)$$

$$= -\frac{1}{2\pi\sigma^2} SCR(\mathbf{w}) - \frac{N}{2} \log(2\pi\sigma^2) \quad (3.9)$$

Avec SCR la somme des carrées résiduelles définies par :

$$SCR(\mathbf{w}) = \sum_{i=1}^N (y_i - \mathbf{w}^T \mathbf{x}_i)^2 \quad (3.10)$$

Il peut aussi s'écrire comme le carré de la norme l_2 du vecteur des erreurs résiduelles :

$$SCR(\mathbf{w}) = \|\epsilon\|_2^2 = \sum_{i=1}^N \epsilon_i^2 \quad (3.11)$$

avec $\epsilon_i = (y_i - \mathbf{w}^T \mathbf{x}_i)$

Dérivation du MLE

Tout d'abord, nous réécrivons l'objectif sous une forme qui se prête mieux à la différenciation :

$$NLL(\mathbf{w}) = \frac{1}{2} (\mathbf{y} - \mathbf{X}\mathbf{w})^T (\mathbf{y} - \mathbf{X}\mathbf{w}) = \frac{1}{2} \mathbf{w}^T (\mathbf{X}^T \mathbf{X}) \mathbf{w} - \mathbf{w}^T (\mathbf{X}^T \mathbf{y}) \quad (3.12)$$

où

$$\mathbf{X}^T \mathbf{X} = \sum_{i=1}^N \begin{pmatrix} x_{i,1} & \dots & x_{i,1}x_{i,D} \\ & \ddots & \\ x_{i,D}x_{i,1} & \dots & x_{i,D}^2 \end{pmatrix} \quad (3.13)$$

est la matrice somme des carrés et

$$\mathbf{X}^T \mathbf{y} = \sum_{i=1}^N \mathbf{x}_i y_i. \quad (3.14)$$

En calculant le gradient de $NLL(\mathbf{w})$ nous avons :

$$g(\mathbf{w}) = [\mathbf{X}^T \mathbf{X} \mathbf{w} - \mathbf{X}^T \mathbf{y}] = \sum_{i=1}^N \mathbf{x}_i (\mathbf{w}^T x_i - y_i) \quad (3.15)$$

En égalant à zéro, nous obtenons

$$\mathbf{X}^T \mathbf{X} \mathbf{w} = \mathbf{X}^T \mathbf{y} \quad (3.16)$$

C'est ce qu'on appelle l'équation normale. La solution correspondante $\hat{\mathbf{w}}$ à ce système linéaire d'équations est appelée la solution des moindres carrés ordinaires, qui est donnée par :

$$\hat{\mathbf{w}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (3.17)$$

3.1.3 Gradient boosting

Le gradient boosting est une technique d'apprentissage automatique utilisée dans les tâches de régression et de classification, entre autres. Il donne un modèle de prédiction sous la forme d'un ensemble de modèles de prédiction faibles, qui sont typiquement des arbres de décision. Lorsqu'un arbre de décision est l'apprenant faible, l'algorithme résultant est appelé arbres à gradient renforcé. Un modèle d'arbres boostés par le gradient est construit par étapes comme dans d'autres méthodes de boosting, mais il généralise les autres méthodes en permettant l'optimisation d'un nombre arbitraire fonction de perte différentiable.

Gradient boosting

Comme d'autres méthodes de boosting, le gradient boosting combine des "apprenants" faibles en un seul apprenant fort de manière itérative. Il est plus facile à expliquer dans le cadre de la régression des moindres carrés, où le but est "d'enseigner" un modèle F prédire des valeurs de la forme $\hat{y} = F(x)$ en minimisant l'erreur quadratique moyenne $\frac{1}{n} \sum_i (\hat{y}_i - y_i)^2$, où je i index sur un ensemble d'entraînement de taille n des valeurs réelles de la variable de sortie y :

\hat{y}_i = la valeur prédite $F(x_i)$

y_i = la valeur observée

n = le nombre d'échantillons dans y

Considérons maintenant un algorithme de renforcement de gradient avec M étapes. A chaque étape m ($1 \leq m \leq M$) de renforcement du gradient, supposons un modèle imparfait F_m (pour les faibles m , ce modèle peut simplement retourner $\hat{y}_i = \bar{y}$, où l'échelle de droite est la moyenne de y). Pour améliorer F_m , notre algorithme devrait ajouter un nouvel estimateur, $h_m(x)$. Ainsi,

$$F_{m+1}(x_i) = F_m(x_i) + h_m(x_i) = y_i \quad (3.18)$$

ou équivalent,

$$h_m(x_i) = y_i - F_m(x_i). \quad (3.19)$$

Par conséquent, l'amplification du gradient conviendra h_m au résiduel $y_i - F_m(x_i)$. Comme dans d'autres variantes de boosting, chaque F_{m+1} tente de corriger les erreurs de son prédécesseur F_m . Une généralisation de cette idée aux fonctions de perte autres que l'erreur quadratique et aux problèmes de classification et de classement découle de l'observation que les résidus $h_m(x_i)$ pour un modèle donné sont proportionnels aux gradients négatifs

de la fonction de perte de l'erreur quadratique moyenne (en anglais mean squared error, MSE) (par rapport à $F(x_i)$) :

$$L_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (y_i - F(x_i))^2 \quad (3.20)$$

$$-\frac{\partial L_{\text{MSE}}}{\partial F(x_i)} = \frac{2}{n} (y_i - F(x_i)) = \frac{2}{n} h_m(x_i). \quad (3.21)$$

Ainsi, l'amplification de gradient pourrait être spécialisée dans un algorithme de descente de gradient, et sa généralisation implique de "brancher" une perte différente et son gradient.

Algorithme

Dans de nombreux problèmes d'apprentissage supervisé, il existe une variable de sortie y et un vecteur de variables d'entrée x , liés les uns aux autres avec une certaine distribution probabiliste. Le but est de trouver une fonction $\hat{F}(x)$ qui se rapproche le mieux de la variable de sortie à partir des valeurs des variables d'entrée. Ceci est formalisé en introduisant une fonction de perte $L(y, F(x))$ et en le minimisant dans l'attente :

$$\hat{F} = \arg \min_F \mathbb{E}_{x,y} [L(y, F(x))]. \quad (3.22)$$

La méthode de renforcement du gradient suppose un y à valeur réelle. Il cherche une approximation $\hat{F}(x)$ sous la forme d'une somme pondérée de M fonctions $h_m(x)$ d'une certaine classe \mathcal{H} , appelés apprenants de base (ou faibles) :

$$\hat{F}(x) = \sum_{m=1}^M \gamma_m h_m(x) + \text{const.} \quad (3.23)$$

On nous donne généralement un ensemble de formation $\{(x_1, y_1), \dots, (x_n, y_n)\}$ des valeurs d'échantillon connues de x et des valeurs correspondantes de y . Conformément au principe empirique de minimisation du risque, la méthode tente de trouver une approximation $\hat{F}(x)$ qui minimise la valeur moyenne de la fonction de perte sur l'ensemble d'apprentissage, c'est-à-dire minimise le risque empirique. Pour ce faire, il part d'un modèle constitué d'une fonction constante $F_0(x)$, et le développe progressivement de manière gourmande :

$$F_0(x) = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, \gamma) \quad (3.24)$$

$$F_m(x) = F_{m-1}(x) + \arg \min_{h_m \in \mathcal{H}} \left[\sum_{i=1}^n L(y_i, F_{m-1}(x_i) + h_m(x_i)) \right], \quad (3.25)$$

pour $m \geq 1$, où $h_m \in \mathcal{H}$ est une fonction d'apprentissage de base.

Malheureusement, choisir la meilleure fonction h_m à chaque étape pour une fonction de perte arbitraire L est un problème d'optimisation informatiquement infaisable en général. Par conséquent, nous limitons notre approche à une version simplifiée du problème.

L'idée est d'appliquer un pas de descente la plus raide à ce problème de minimisation (descente de gradient fonctionnelle).

L'idée de base derrière la descente la plus raide est de trouver un minimum local de la fonction de perte en itérant sur $F_{m-1}(x)$. En fait, la direction de descente maximale locale de la fonction de perte est le gradient négatif.

Par conséquent, déplacer une petite quantité γ telle que l'approximation linéaire reste valable :

$$F_m(x) = F_{m-1}(x) - \gamma \sum_{i=1}^n \nabla_{F_{m-1}} L(y_i, F_{m-1}(x_i)) \quad (3.26)$$

où $\gamma > 0$. Pour les petits γ , cela implique que $L(y_i, F_m(x_i)) \leq L(y_i, F_{m-1}(x_i))$.

De plus, nous pouvons optimiser γ en trouvant le γ valeur pour laquelle la fonction de perte a un minimum :

$$\gamma_m = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, F_m(x_i)) = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, F_{m-1}(x_i) - \gamma \nabla_{F_{m-1}} L(y_i, F_{m-1}(x_i))). \quad (3.27)$$

Si l'on considère le cas continu, c'est-à-dire où \mathcal{H} est l'ensemble des fonctions différentiables arbitraires sur \mathbb{R} ,

nous mettons à jour le modèle conformément aux équations suivantes

$$F_m(x) = F_{m-1}(x) - \gamma_m \sum_{i=1}^n \nabla_{F_{m-1}} L(y_i, F_{m-1}(x_i)) \quad (3.28)$$

où γ_m est la longueur du pas, définie comme $\gamma_m = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, F_{m-1}(x_i) - \gamma \nabla_{F_{m-1}} L(y_i, F_{m-1}(x_i)))$.

Dans le cas discret cependant, c'est-à-dire lorsque l'ensemble \mathcal{H} est finie, on choisit la fonction candidate h la plus proche du gradient de L pour laquelle le coefficient γ peut alors être calculé à l'aide d'une recherche linéaire sur les équations ci-dessus. Notez que cette approche est une heuristique et ne donne donc pas une solution exacte au problème donné, mais plutôt une approximation. En pseudo code, la méthode générique de boosting de gradient est :

Entrée : ensemble d'entraînement $\{(x_i, y_i)\}_{i=1}^n$, une fonction de perte différentiable $L(y, F(x))$, nombre d'itérations M .

Algorithme:

1. Initialiser le modèle avec une valeur constante :

$$F_0(x) = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, \gamma).$$

2. Pour $m = 1$ à M :

1. Calculer des *pseudo-résidus* :

$$r_{im} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)} \quad \text{for } i = 1, \dots, n.$$

2. Ajuster un apprenant de base (ou un apprenant faible, par exemple un arbre) fermé sous mise à l'échelle $h_m(x)$ aux pseudo-résidus, c'est-à-dire l'entraîner à l'aide de l'ensemble d'apprentissage $\{(x_i, r_{im})\}_{i=1}^n$.

3. Calculer le multiplicateur γ_m en résolvant le problème d'optimisation unidimensionnel suivant :

$$\gamma_m = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)).$$

4. Mettez à jour le modèle :

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x).$$

3. Production $F_M(x)$.

3.1.4 Gradient tree boosting

Le gradient boosting est généralement utilisée avec des arbres de décision (en particulier des arbres CART) d'une taille fixe en tant qu'apprenants de base. Pour ce cas particulier, Friedman propose une modification de la méthode de renforcement du gradient qui améliore la qualité d'ajustement de chaque apprenant de base.

L'augmentation de gradient générique à la m -ième étape correspondrait à un arbre de décision $h_m(x)$ aux pseudo-résidus. Laisser J_m être le nombre de ses feuilles. L'arborescence partitionne l'espace d'entrée en J_m régions disjointes $R_{1m}, \dots, R_{J_m m}$ et prédit une valeur constante dans chaque région. En utilisant la notation de l'indicateur, la sortie de $h_m(x)$ pour l'entrée x peut être écrit comme la somme :

$$h_m(x) = \sum_{j=1}^{J_m} b_{jm} \mathbf{1}_{R_{jm}}(x), \quad (3.29)$$

où b_{jm} est la valeur prédite dans la région R_{jm} .

Alors les coefficients b_{jm} sont multipliés par une certaine valeur γ_m , choisie à l'aide de la recherche linéaire de manière à minimiser la fonction de perte, et le modèle est mis à jour comme suit :

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x), \quad \gamma_m = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)). \quad (3.30)$$

Friedman propose de modifier cet algorithme afin qu'il choisisse une valeur optimale distincte γ_{jm} pour chacune des régions de l'arbre, au lieu d'un seul γ_m pour tout l'arbre. Il appelle l'algorithme modifié "TreeBoost". Les coefficients b_{jm} de la procédure d'ajustement d'arbre peut alors être simplement ignorée et la règle de mise à jour

du modèle devient :

$$F_m(x) = F_{m-1}(x) + \sum_{j=1}^{J_m} \gamma_{jm} \mathbf{1}_{R_{jm}}(x), \quad \gamma_{jm} = \arg \min_{\gamma} \sum_{x_i \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma). \quad (3.31)$$

J , le nombre de nœuds terminaux dans les arbres, est le paramètre de la méthode qui peut être ajusté pour un ensemble de données à portée de main. Il contrôle le niveau d'interaction maximal autorisé entre les variables du modèle. Avec $J = 2$ (souches de décision), aucune interaction entre les variables n'est autorisée. Avec $J = 3$ le modèle peut inclure les effets de l'interaction entre jusqu'à deux variables, et ainsi de suite.

Hastie et al. commentent que typiquement $4 \leq J \leq 8$ fonctionnent bien pour booster et les résultats sont assez insensibles au choix de J dans cette gamme, $J = 2$ est insuffisant pour de nombreuses applications, et $J > 10$ est peu susceptible d'être nécessaire.

3.1.5 Régularisation

Un ajustement trop serré de l'ensemble d'apprentissage peut entraîner une dégradation de la capacité de généralisation du modèle. Plusieurs techniques dites de régularisation réduisent cet effet de sur-ajustement en contraignant la procédure d'ajustement.

Un paramètre naturel de régularisation est le nombre d'itérations de renforcement de gradient M (c'est-à-dire le nombre d'arbres dans le modèle lorsque l'apprenant de base est un arbre de décision). L'augmentation de M réduit l'erreur sur l'ensemble d'apprentissage, mais une valeur trop élevée peut entraîner un sur-ajustement. Une valeur optimale de M est souvent sélectionnée en surveillant l'erreur de prédiction sur un ensemble de données de validation séparé. Outre le contrôle de M , plusieurs autres techniques de régularisation sont utilisées.

Un autre paramètre de régularisation est la profondeur des arbres. Plus cette valeur est élevée, plus le modèle est susceptible de sur-ajuster les données d'apprentissage.

Rétrécissement

Une partie importante de la méthode de gradient boosting est la régularisation par rétrécissement qui consiste à modifier la règle de mise à jour comme suit :

$$F_m(x) = F_{m-1}(x) + \nu \cdot \gamma_m h_m(x), \quad 0 < \nu \leq 1, \quad (3.32)$$

où paramètre ν s'appelle le "taux d'apprentissage".

Empiriquement, il a été constaté que l'utilisation de petits taux d'apprentissage (tels que $\nu < 0.1$) apporte des améliorations spectaculaires dans la capacité de généralisation des modèles par rapport à l'amplification de gradient sans rétrécissement ($\nu = 1$). Cependant, cela se fait au prix d'une augmentation du temps de calcul à la fois pendant l'apprentissage et l'interrogation : un taux d'apprentissage plus faible nécessite plus d'itérations.

Renforcement du gradient stochastique

Peu de temps après l'introduction de l'amplification de gradient, Friedman a proposé une modification mineure de l'algorithme, motivée par la méthode d'agrégation bootstrap ("bagging") de Breiman. Plus précisément, il a proposé qu'à chaque itération de l'algorithme, un apprenant de base soit ajusté sur un sous-échantillon de l'ensemble d'apprentissage tiré au sort sans remise. Friedman a observé une amélioration substantielle de la précision de l'amplification du gradient avec cette modification.

La taille du sous-échantillon est une fraction constante f de la taille de l'ensemble d'apprentissage. Lorsque $f = 1$, l'algorithme est déterministe et identique à celui décrit ci-dessus. Des valeurs plus petites de f introduisent du caractère aléatoire dans l'algorithme et aident à prévenir le surajustement, agissant comme une sorte de régularisation. L'algorithme devient également plus rapide, car les arbres de régression doivent être adaptés à des ensembles de données plus petits à chaque itération. Friedman a obtenu que $0.5 \leq f \leq 0.8$ conduit à de bons résultats pour les ensembles d'entraînement de petite et moyenne taille. Par conséquent, f est généralement défini sur 0,5, ce qui signifie qu'une moitié de l'ensemble de formation est utilisée pour construire chaque apprenant de base.

De plus, comme dans le bagging, le sous-échantillonnage permet de définir une erreur hors sac de l'amélioration des performances de prédiction en évaluant les prédictions sur les observations qui n'ont pas été utilisées dans la construction de l'apprenant de base suivant. Les estimations prêtes à l'emploi permettent d'éviter le besoin d'un ensemble de données de validation indépendant, mais sous-estiment souvent l'amélioration réelle des performances et le nombre optimal d'itérations.

Nombre d'observations dans les feuilles

Les implémentations de renforcement d'arbre de gradient utilisent souvent également la régularisation en limitant le nombre minimum d'observations dans les nœuds terminaux des arbres. Il est utilisé dans le processus de construction de l'arborescence en ignorant toutes les scissions qui conduisent à des nœuds contenant moins que ce nombre d'instances d'ensemble d'apprentissage.

L'imposition de cette limite permet de réduire la variance des prédictions au niveau des feuilles.

Pénaliser la complexité de l'arbre

Une autre technique de régularisation utile pour les arbres à gradient boosté consiste à pénaliser la complexité du modèle appris. La complexité du modèle peut être définie comme le nombre proportionnel de feuilles dans les arbres appris. L'optimisation conjointe de la perte et de la complexité du modèle correspond à un algorithme de post-élagage pour supprimer les branches qui ne parviennent pas à réduire la perte d'un seuil. D'autres types de régularisation tels qu'un ℓ_2 une pénalité sur les valeurs des feuilles peut également être ajoutée pour éviter le surajustement .

3.2 Applications

3.2.1 Présentation des données

Grâce à BigQuery, nous avons récupéré nos données de campagne et les connectées à un Google sheet pour les analyser et créer une nouvelle base données.

Nous avons créé une base qui contient le coût média, le nombre total d'impressions, la durée, le mois du commencement, le mois de la fin, le mode d'achat (buying mode), l'environnement, le catégorie de l'annonceur, le nombre total de cliques d'un élément de campagne (line item). Puis la téléchargée en csv pour l'importer dans Google colab afin de continuer notre étude et créer notre modèle.

Voici un exemple de données :

	Environment	Advertiser_Category	Line_item_type	Buying_Mode	SUM of Impressions	SUM of Clicks	SUM of Media_Cost	Durée	Start month	End month
line_item_id										
1.032303e+10	Connected TV	Banque/Assurance	YouTube & partners	Open Auction	2446	0	11.195281	3	1	1
NaN	Desktop	Banque/Assurance	YouTube & partners	Open Auction	263	0	1.387379	3	1	1
NaN	Mobile In-App	Banque/Assurance	YouTube & partners	Open Auction	1149	1	6.603395	3	1	1
NaN	Mobile WebM	Banque/Assurance	YouTube & partners	Open Auction	149	0	0.564388	3	1	1
1.032304e+10	Connected TV	Banque/Assurance	YouTube & partners	Open Auction	3343	0	14.405610	3	1	1
...
NaN	Mobile In-App	Automobile	Real-time bidding	Open Auction	4954	6	11.908246	22	9	9
NaN	Mobile WebM	Automobile	Real-time bidding	Open Auction	61301	57	119.407588	22	9	9
5.717970e+07	Desktop	Automobile	Real-time bidding	Open Auction	20230	12	42.385215	21	9	9
NaN	Mobile In-App	Automobile	Real-time bidding	Open Auction	4719	6	11.315065	21	9	9
NaN	Mobile WebM	Automobile	Real-time bidding	Open Auction	60784	54	120.130286	21	9	9

28140 rows × 10 columns

Description de nos données :

	Environment	Advertiser_Category	Line_item_type	Buying_Mode	SUM of Impressions	SUM of Clicks	SUM of Media_Cost	Durée	Start month	End month
count	28140	28140	28140	28140	2.814000e+04	28140.000000	28140.000000	28140.000000	28140.000000	28140.000000
unique	5	16	4	4	NaN	NaN	NaN	NaN	NaN	NaN
top	Desktop	Banque/Assurance	Real-time bidding	Open Auction	NaN	NaN	NaN	NaN	NaN	NaN
freq	8928	9164	11545	19131	NaN	NaN	NaN	NaN	NaN	NaN
mean	NaN	NaN	NaN	NaN	5.833945e+05	941.630135	1992.014087	39.280313	4.969225	4.969225
std	NaN	NaN	NaN	NaN	2.013497e+06	5111.412045	4951.349274	36.586322	2.955378	2.955378
min	NaN	NaN	NaN	NaN	1.000000e+00	0.000000	0.000003	1.000000	1.000000	1.000000
25%	NaN	NaN	NaN	NaN	3.065500e+03	0.000000	16.638375	15.000000	3.000000	3.000000
50%	NaN	NaN	NaN	NaN	5.917000e+04	5.000000	303.135593	28.000000	5.000000	5.000000
75%	NaN	NaN	NaN	NaN	3.950468e+05	188.000000	1733.038092	55.000000	7.000000	7.000000
max	NaN	NaN	NaN	NaN	6.553471e+07	324628.000000	156315.987100	353.000000	12.000000	12.000000

3.2.2 Transformations des données

Comme vous pouvez le constater, nous avons des variables qualitatives, le type de l'élément de campagne, l'environnement, la catégorie de l'annonceur et le mode d'achat dans nos données. Nous allons vectoriser chacun élément avec la méthode `get_dummies` de `pandas` :

```

# W = df2[['sum_of_impressions', 'sum_of_media_cost', 'duree', 'environment', 'line_item_type', 'buying_mode', 'advertiser_category']]
# W = pd.get_dummies(W, drop_first=True)

```

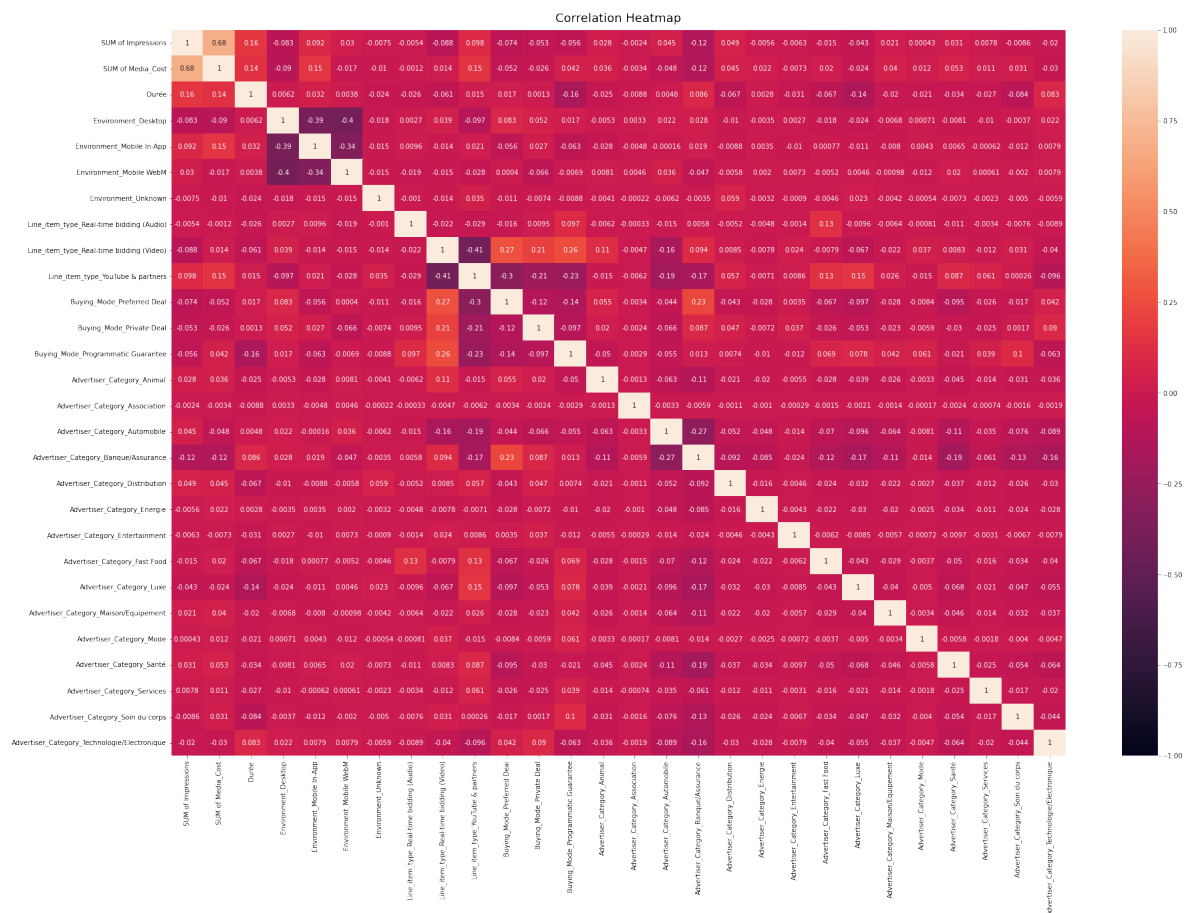
	sum_of_impressions	sum_of_media_cost	duree	environment_desktop	environment_mobile_in-app	environment_mobile_webm	environment_unknown	line_item_type_real-time bidding (audio)	line_item_type_real-time bidding (video)	line_item_type_youtube & partners	...	advertiser_category_energie	advertiser_category_entertainment	advertiser_category_fast food	advertiser_category_luxe	...
line_item_id																
1.032303e+10	2446	11.195281	3	0	0	0	0	0	0	1	...	0	0	0	0	0
Null	263	1.307379	3	1	0	0	0	0	0	1	...	0	0	0	0	0
Null	1149	6.603395	3	0	1	0	0	0	0	1	...	0	0	0	0	0
Null	149	0.564308	3	0	0	1	0	0	0	1	...	0	0	0	0	0
1.032304e+10	3343	14.405610	3	0	0	0	0	0	0	1	...	0	0	0	0	0
...
Null	4954	11.900246	22	0	1	0	0	0	0	0	...	0	0	0	0	0
Null	61301	119.407588	22	0	0	1	0	0	0	0	...	0	0	0	0	0
5.717970e+07	20280	42.385215	21	1	0	0	0	0	0	0	...	0	0	0	0	0
Null	4719	11.315065	21	0	1	0	0	0	0	0	...	0	0	0	0	0
Null	60764	120.130286	21	0	0	1	0	0	0	0	...	0	0	0	0	0

25140 rows x 28 columns

3.2.3 Construction de modèle de régression

Nous voulons prédire le nombre d'impressions d'un élément de campagne à partir des autres campagnes. Pour cela, nous calculons d'abord les corrélations.

Corrélations



Nous constatons que les impressions sont faiblement corrélées aux autres variables à part le coût média et la durée. Nous allons construire des modèles linéaire régression et gradient boosting pour les comparer et améliorer le meilleur.

Modèles de régressions

Nous importons les modèles **LinearRegression** depuis **sklearn.linear_model** et **GradientBoostingRegressor** depuis **sklearn.ensemble**. Nous avons une base de données de 28140 échantillons que nous allons en deux groupe 90% pour les données d'apprentissage et 10% pour les données de test . Les scores sont calculés avec R2 score ou MSE.

```
GradientBoostingRegressor()
Training time: 2.738s
Prediction time: 0.006s
Explained variance: 0.753100352340148
Mean absolute error: 313224.6845760442
score in train: 0.7444312476742341
R2 score : 0.752951785752194

LinearRegression()
Training time: 0.039s
Prediction time: 0.003s
Explained variance: 0.580558519175128
Mean absolute error: 485310.3313658662
score in train: 0.4887753771218155
R2 score : 0.5805277561632876
```

Nous constatons le gradient boosting a des scores plus élevés sur l'apprentissage et le test, il présente aussi le plus faible erreur moyenne absolue. Cependant sa variance, et les temps mis à l'apprentissage et à la prédiction sont plus élevés.

Nous avons aussi essayé de centré et réduire la matrice d'entrée mais nous avons obtenu les mêmes résultats.

Amélioration du modèle gradient boosting

Nous recherchons son meilleur modèle suivant ses hyperparamètres. Pour cela, nous utilisons **GridSearchCV** de **sklearn.model_selection** qui nous sélectionne le modèle, il utilise la cross-validation.

La sélection avec les hyperparamètres :

```
[19] parameters = {'loss':['squared_error', 'absolute_error', 'huber', 'quantile'], 'n_estimators' : [75,100],
                  'criterion' : ['friedman_mse', 'squared_error'], 'learning_rate':[0.1,0.2]}

grid = GridSearchCV(GradientBoostingRegressor(),parameters)
model = grid.fit(M_train,m_train)
print(model.best_params_,'\n')
print(model.best_estimator_,'\n')

{'criterion': 'squared_error', 'learning_rate': 0.2, 'loss': 'squared_error', 'n_estimators': 75}

GradientBoostingRegressor(criterion='squared_error', learning_rate=0.2,
                           n_estimators=75)
```

Il a comparé 32 modèles, et le meilleur a pour est celui avec les paramètres : *'criterion' : 'squared_error', 'learning_rate' : 0.2, 'loss' : 'squared_error' et 'n_estimators' : 75.*

Après avoir entraîné ce dernier avec nos données d'apprentissage, nous l'avons testé avec nos données de test nous obtenons :

```
[21] GBR = GradientBoostingRegressor(criterion='squared_error', learning_rate=0.2,n_estimators=75).fit(M_train,m_train)

[22] m_pred=GBR.predict(M_test)

print("\tExplained variance:", explained_variance_score(m_test, m_pred))
print("\tMean absolute error:", mean_absolute_error(m_test, m_pred))

print("\tR2 score :", r2_score(m_test, m_pred))

Explained variance: 0.7705358717431527
Mean absolute error: 305962.9211397919
R2 score : 0.7703969689907411
```

Nous constatons une amélioration de notre modèle avec une augmentation de notre score et une diminution de notre erreur moyenne absolue.

Conclusion

Dans le domaine de la publicité, la brand safety (sécurité de la marque) est primordiale. Avec cet outil d'automatisation, les traders vont créer leurs canaux d'exclusions et d'inclusions de manière automatique et en gagnant deux voir trois heures de temps sur ce travail.

Cet stage m'a permis de découvrir le domaine du marketing et d'apprendre de nouvelles compétences comme développement d'API avec apps cript, de créer un outil d'automatisation mais aussi de développer mes compétences mathématiques comme apprentissage statistique.

Références

Wikipedia : Brand safety, Services Web, apprentissage statistique, gradient boosting
Machine LearningA Probabilistic Perspective, Kevin P. Murphy
Cours Modèle linéaire, Marie Luce Taupin
[https ://developers.google.com/display-video/api/reference](https://developers.google.com/display-video/api/reference)
[https ://developers.google.com/apps-script/reference](https://developers.google.com/apps-script/reference)
Bibliothèque GASRetry Peter Herrmann Brand Safety GroupM, Edition 2019