

# **Automação de um Sistema *Business Intelligence***

*João Viterbo Vieira*

**Relatório Final do Estágio/Projeto**

Orientador na FEUP: Professor Doutor A. Miguel Gomes



**Licenciatura em Engenharia e Gestão Industrial**

2023-10-20

## Resumo

O presente relatório descreve um projeto realizado na empresa Clever Advertising, especializada em *marketing* digital, com foco na automação de um sistema *Business Intelligence* (BI).

O foco deste projeto passou por analisar e resolver problemas, relacionados com a implementação de automatização dos fluxos de dados e realização de relatórios e *dashboards*, com a criação de indicadores e métricas de desempenho, para monitorização e avaliação de resultados de *marketing* digital.

A abordagem do problema começou com uma observação da empresa, ambiente de trabalho, relações e processos internos.

Devido ao crescimento exponencial da informação e dos diferentes tipos de recolha de dados da web e redes sociais, os processos inicialmente implementados na empresa tornaram-se pouco eficientes.

Depois de identificado o problema e possíveis potenciais melhorias, começou-se com o estudo e revisão bibliográfica das soluções e tecnologias aplicáveis, destacando a crescente tendência de automação nos negócios.

A criação de um *Business Intelligence* eficiente com a automatização do fluxo de dados através da funcionalidade do Google App Scripts (JavaScript), melhorou a eficiência operacional. Além disso, realizou-se uma organização e estruturação de dados, resultando na produção de indicadores e métricas de desempenho, essenciais para o acompanhamento do negócio.

Criou-se a geração automática de relatórios por meio do Google Sheets e *dashboards* no PowerBI, eliminando a necessidade de análise manual, proporcionando relatórios atualizados e em tempo útil, bem como, a automação de fluxo de dados, resultando um ganho de tempo útil.

Os trabalhos desenvolvidos geraram ganhos estimados diários, de 1h30m para os elementos adstritos ao processo ETL e 1h por cada relatório gerado.

Este projeto na Clever Advertising proporcionou uma valiosa oportunidade para investigar e aplicar conhecimentos teórico-práticos em ambiente de trabalho, contribuindo significativamente para o melhoramento e operacionalização do fluxo, gestão e análise de dados.

O projeto teve um impacto positivo na Clever Advertising permitindo a criação de um sistema *Business Intelligence* e outras estruturas para futuras aplicações.

## **Abstract**

This report describes a project carried out at Clever Advertising, a company specialized in digital marketing, focusing on the automation of a Business Intelligence (BI) system.

The aim of this project was to analyze and solve issues related to the implementation of data flow automation, as well as the creation of reports and dashboards, equipped with performance indicators and metrics for monitoring and evaluating digital marketing results.

The problem-solving approach began with an observation of the company, its work environment, internal relationships, and processes.

Due to the exponential growth of information and various types of data collection from the web and social networks, the processes initially implemented in the company became inefficient.

After identifying the problem and potential areas for improvement, a literature review was conducted to study applicable solutions and technologies, highlighting the growing trend of business automation.

The creation of an efficient Business Intelligence system through data flow automation using Google App Scripts (JavaScript) improved operational efficiency. Additionally, data organization and structuring were carried out, resulting in the production of essential performance indicators and metrics for business monitoring.

Automated report generation was achieved through Google Sheets and dashboards in PowerBI, eliminating the need for manual analysis and providing timely and updated reports, as well as automating the data flow, resulting in significant time savings.

The work carried out led to estimated daily time savings of 1 hour and 30 minutes for personnel involved in the ETL (Extract, Transform, Load) process and 1 hour per generated report.

This project at Clever Advertising provided a valuable opportunity to investigate and apply theoretical and practical knowledge in a work environment, contributing significantly to the improvement and operationalization of data flow, management, and analysis.

The project had a positive impact on Clever Advertising, enabling the creation of a Business Intelligence system and other structures for future applications.

## **Agradecimentos**

Queria agradecer, em primeiro lugar ao meu orientador Professor Doutor A. Miguel Gomes pela orientação e ao regente da cadeira o Professor Doutor Gonçalo Figueira pelo apoio, que me proporcionaram no decorrer do estágio.

De seguida, queria agradecer à Clever Advertising, nomeadamente ao Engenheiro Filipe Magalhães pela orientação e ajuda no trabalho e a todos os elementos da equipa “Bold” pela sua disponibilidade, colaboração e bom ambiente que criaram no decorrer do projeto.

# Índice de Conteúdos

|     |  |    |
|-----|--|----|
| 1   | Introdução .....   | 1  |
| 1.1 | Apresentação da empresa .....  | 1  |
| 1.2 | Contexto e motivação .....   | 2  |
| 1.3 | Objetivos .....  | 2  |
| 1.4 | Planeamento e execução do projeto .....  | 2  |
| 1.5 | Estrutura do relatório .....   | 3  |
| 2   | <i>Business Intelligence</i> e Gestão de fluxos de dados .....                   | 4  |
| 2.1 | Automação de um sistema <i>Business Intelligence</i> .....                       | 4  |
| 2.2 | ETL ( <i>Extraction Transformation and Loading</i> ) .....                       | 5  |
| 2.3 | <i>Web scraper</i> e <i>Web crawler</i> .....                                    | 6  |
| 2.4 | API ( <i>Application Programming Interface</i> ) .....                           | 7  |
| 2.5 | Google Sheets para sistemas de automação <i>Business Intelligence</i> .....      | 7  |
| 2.6 | Google app scripts para sistemas de automação <i>Business Intelligence</i> ..... | 8  |
| 2.7 | Exploração do <i>Machine Learning</i> e AI em <i>Business Intelligence</i> ..... | 8  |
| 3   | Caso de estudo .....   | 10 |
| 3.1 | Processo de gestão de fluxo de dados .....                                       | 10 |
| 3.2 | Relatórios automatizados e monitorização .....                                   | 11 |
| 4   | Desenvolvimento e aplicação da solução de <i>Business Intelligence</i> .....     | 12 |
| 4.1 | Apresentação e escolha de soluções .....   | 12 |
| 4.2 | Automação do fluxo de dados: Processo ETL .....                                  | 13 |
| 4.3 | Geração automática de relatórios via Google Sheets e <i>dashboard</i> .....      | 15 |
| 4.4 | Análise de Dados e produção de indicadores e métricas de desempenho .....        | 16 |
| 4.5 | Discussão .....  | 18 |
| 5   | Conclusão .....  | 19 |
| 5.1 | Principais contribuições .....   | 19 |
| 5.2 | Limitações do trabalho e perspetivas de trabalho futuro .....                    | 20 |
|     | Referências .....  | 21 |
|     | Anexos .....   | 22 |

## Índice de Figuras

|   |    |
|---|----|
| Figura 1 - Apresentação de soluções.....                                    | 12 |
| Figura 2 - Exemplo de um cabeçalho para anotação dos conteúdos gerados..... | 14 |
| Figura 3 - Fluxograma dos <i>scripts</i> .....                              | 15 |
| Figura 4 - <i>Dashboard</i> do PowerBI.....                                 | 16 |
| Figura 5 - Exemplo de <i>template</i> com indicadores e métricas .....      | 17 |

## **Siglas e acrónimos**

BI - *Business Intelligence*

API - *Application Programming Interface*

ETL - *Extract, Transform and Load*

ROI - *Return On Investment*

AI - *Artificial Intelligence*

ML - *Machine Learning*

HTML - *Hypertext Markup Language*

KPI - *Key Performance Indicator*

# 1 Introdução

O projeto realizado na empresa Clever Advertising, especializada em *marketing* digital, uma área de trabalho dinâmica devido à evolução tecnológica e ao crescimento exponencial da informação proveniente de diversas fontes, como *web* e redes sociais, propiciou um ambiente em contexto de trabalho, para estudar e propor soluções exequíveis de automação de um sistema *Business Intelligence*.

A observação da empresa e do seu ambiente de trabalho, da interação das relações e processos internos, permitiu o estudo das atividades e compreender, que a recolha de dados da *web* e redes sociais, com os processos inicialmente implementados na empresa se tornaram pouco eficientes e impraticáveis a médio prazo.

O objetivo principal deste projeto foi analisar e resolver problemas relacionados com a implementação da automatização de um sistema *Business Intelligence*, o que implicou a extração, o tratamento e armazenamento de dados, para a realização sistemática de relatórios, para a monitorização e para a avaliação do processo.

O projeto em Engenharia e Gestão Industrial (PEGI) proporcionou um contacto com a realidade profissional, incrementando a proatividade, para aquisição de novas competências técnicas e não-técnicas (*soft-skills*), aprendizagem e aplicação de novas tecnologias, durante um período de seis semanas.

## 1.1 Apresentação da empresa

A Clever Advertising é uma empresa de *marketing* digital com um historial de 15 anos de atividade na indústria. O principal foco da empresa reside em aumentar a aquisição de clientes para os seus parceiros através de diversas estratégias e canais digitais. A Clever Advertising utiliza um modelo de negócios baseado em afiliação para ajudar os seus clientes a conquistar uma maior quota de mercado e a estabelecer uma posição sólida no setor.

Com um alcance global, a empresa presta serviços a mais de 50 operadores em cinco continentes. As áreas de especialização da Clever Advertising abrangem aspetos cruciais do *marketing* digital, incluindo *Media Buying*, *Search Engine Optimization* (SEO), publicidade por pagamento por clique (PPC) e *influencer marketing*. Ao longo da sua atividade, a Clever Advertising tem um compromisso constante em impulsionar a aquisição de clientes e expandir a presença digital dos seus parceiros a nível internacional.

A divisão de operação denominada “Bold” está mais direcionada para o *influencer marketing*, através das redes sociais como Youtube, Facebook, Instagram, X (antigo Twitter), Twitch, Telegram, Discord, etc...



## 1.2 Contexto e motivação

O projeto realizado na Clever Advertising, proporcionou um ambiente real, para a aplicar e testar conhecimentos teórico-práticos para resolução de problemas em contexto de trabalho, bem como adquirir *skills* e competências não técnicas.

O ambiente de *marketing* digital está a evoluir muito rapidamente nos seus diversos aspetos, tendo sido motivação o estudo de novas tecnologias e novos conhecimentos, o que permitiu entrar em contacto com a área de publicidade digital nas suas diversas vertentes.

Outra motivação para realizar o projeto na Clever Advertising foram os desafios apresentados e sentidos para investigar, documentar, analisar e avaliar as atividades, com foco na automação do fluxo, gestão e análise de dados, para melhorar a eficiência operacional face à crescente complexidade da informação e proporcionar uma tomada decisão com base em informação estruturada.

Sendo a área do *marketing* digital uma área preferencial de aplicação da Inteligência Artificial aparecia uma oportunidade de estudar e ganhar conhecimentos para perceber e aplicar estes conceitos.

## 1.3 Objetivos

O projeto realizado teve como objetivos, uma oportunidade para a familiarização com a realidade profissional, através da integração e da realização duma solução viável e funcional para a empresa.

Os objetivos principais incluíram: a identificação de problemas existentes na automação de procedimentos internos; a investigação e análise de estudos relevantes; o desenvolvimento de ferramentas computacionais para implementar um sistema de *Business Intelligence*. Este sistema tinha como objetivo aumentar a eficiência operacional no processo ETL e na geração automatizada de relatórios, para a monitorização e avaliação de processos.

O projeto também pretendia investigar e desenvolver indicadores e métricas de desempenho críticos, para uma melhor e eficiente avaliação de resultados. E ainda se pretendia explorar e estudar o potencial de integração de tecnologias emergentes, como Inteligência Artificial (AI).

## 1.4 Planeamento e execução do projeto

O trabalho a desenvolver no projeto foi definido com a empresa, de acordo com as especificações do PEGI, cuja duração seria de 162 horas (até 6 semanas) com o *embedding* na estrutura da Clever Advertising, não podendo perturbar ou interromper a produção da empresa.

Durante as seis semanas do projeto, as duas primeiras foram dedicadas ao conhecimento da empresa e suas estruturas, rotinas e negócios. Seguidamente, ao estudo da automação do *Business Intelligence*, processos de ETL, geração de

relatórios, produção de indicadores, métricas de desempenho e o *state of the art* da Inteligência artificial.

Nas duas semanas seguintes, depois de propostas e decididas as soluções, já em ambiente de trabalho, foi realizada a programação em ambiente Python e JavaScript (App Scripts) para o processo ETL e desenvolvido em Google Sheets e PowerBI, a produção de relatórios e *dashboards*.

Na semana seguinte foi realizada a interação e instalação das soluções com a equipa “Bold”, de forma a responder a todos os seus requisitos, sendo dada formação aos elementos que iriam utilizar as ferramentas desenvolvidas e ainda, foi elaborado um pequeno manual de utilização.

Na última semana foi implementada a solução com uso intensivo e teste de todos os componentes. Estas atividades correram conforme projetado e planeado, não tendo ocorrido nenhuma intercorrência que alterasse o plano inicial, tendo-se cumprido os prazos previstos. Tendo a interface com a empresa e os seus trabalhadores, na mudança das rotinas, ocorrido sem problemas e de forma tranquila.

A introdução e alteração de rotinas na equipa “Bold” foi realizada sem qualquer tipo de perdas tempo, dados ou produtividade, resultando numa melhoria da eficiência operacional.

## 1.5 Estrutura do relatório

O relatório inicia-se com uma introdução, estabelecendo o contexto, os objetivos do projeto e a apresentação da empresa.

A segunda secção é dedicada ao estudo e exploração de aspetos técnicos em *Business Intelligence* e gestão e análise de fluxos de dados, abrangendo tópicos como, automação e ferramentas utilizadas.

A terceira secção oferece um caso de estudo para demonstrar a aplicação prática das abordagens investigadas e discutidas.

A quarta secção foca-se no desenvolvimento e implementação de uma solução de *Business Intelligence*, incluindo a análise de dados, avaliação de desempenho e discussão.

O relatório encerra-se com uma conclusão, que destaca as principais contribuições, limitações e sugere perspetivas para trabalhos futuros.

## 2 *Business Intelligence* e Gestão de fluxos de dados

A informação é um recurso essencial para a tomada de decisões em organizações e instituições científicas, sendo fundamental a capacidade de extrair dados relevantes, analisar, comparar e estruturar uma base de conhecimento, para uma resolução mais adequada. Em seguida, serão expostas as áreas de interesse que foram abordadas no projeto e que se reportam, nomeadamente a automação de um sistema *Business Intelligence*, o processo ETL, *web scraper* e *web crawler*, APIs, Google Sheets, App Scripts e exploração de ML e AI em *Business Intelligence*.

### 2.1 Automação de um sistema *Business Intelligence*

No contexto empresarial, a automação de *Business Intelligence* (BI) emerge como uma ferramenta crucial para a eficiência operacional. Esta automação envolve o uso de tecnologias sofisticadas para automatizar rotinas e processos complexos ligados aos sistemas de BI. Estes processos abrangem desde a recolha e integração de dados, até à sua subsequente análise e apresentação (Chen, Chiang e Storey, 2012).

O objetivo primordial da automação em BI é melhorar as operações, minimizar a necessidade de intervenção manual e fornecer *insights* de negócio de forma mais rápida e precisa.

Entre os componentes-chave da automação em BI, destaca-se a Extração, Transformação e Carregamento (ETL- *Extract, Transform and Load*). Este processo automatiza a recolha de dados de múltiplas fontes, a sua transformação e armazenamento. A análise de dados é crucial, onde os algoritmos são aplicados de forma automática para extrair informações auxiliares. A geração de relatórios e *dashboards* e a configuração de alertas automáticos para condições específicas nos dados, são outras funcionalidades que complementam este sistema. A manutenção de dados, que inclui tarefas como limpeza e verificação de integridade, também é automatizada (Wilkinson, et al., 2010).

No que diz respeito às vantagens da automação em BI, ela oferece uma eficiência operacional significativa. A automação reduz o tempo e o esforço necessários para a análise de dados e minimiza a margem de erro humano. Apresenta desafios, como o custo de implementação, manutenção e a necessidade de *expertise* multidisciplinar (Chen, Chiang e Storey, 2012).

No panorama tecnológico atual, observa-se uma integração crescente de tecnologias como *Machine Learning* e inteligência artificial nos sistemas de automação BI, permitindo análises mais avançadas e personalizadas.

De salientar que a automação BI não elimina a necessidade de intervenção humana, principalmente no planeamento, na interpretação contextual e na tomada de decisões estratégicas. A automação em *Business Intelligence* é uma área em constante evolução, cujas práticas e tecnologias subjacentes, continuam a ser objeto de investigação e desenvolvimento intensivos (Microsoft, 2023).

## 2.2 ETL (*Extraction Transformation and Loading*)

O ETL (Extração, Transformação e Carregamento) é um processo utilizado nos sistemas de informação para a integração de dados. Esta metodologia ganha especial relevância nos contextos de *Big Data* e base de dados, onde a qualidade e a consistência das informações são essenciais para as análises subsequentes (Vassiliadis, 2002).

Com o aumento exponencial da utilização de bases de dados, o ETL consolidou-se como o método de eleição para o tratamento e integração de dados. O ETL serve de alicerce às atividades de análise de dados e *Machine Learning*. Mediante um conjunto de regras de negócio, o ETL assegura a limpeza e organização dos dados, de forma a satisfazer requisitos específicos para a elaboração de relatórios recorrentes (Biplob e Haque, 2022). Contudo, a sua aplicabilidade estende-se a análises mais sofisticadas, que podem melhorar processos internos e experiências do utilizador.

O processo de ETL consiste em três etapas fundamentais:

- A etapa de *Extraction* (Extração) envolve a recolha de dados a partir de múltiplas fontes, que podem incluir bases de dados relacionais, ficheiros de texto, APIs e fluxos de dados em tempo real. O objetivo é realizar esta extração minimizando o impacto nos sistemas de origem e mantendo a integridade dos dados.
- A *Trasnformation* (Transformação) representa a fase onde os dados extraídos são trabalhados, validados e convertidos para um formato analisável. Operações comuns englobam a limpeza de dados, eliminação de duplicados, conversão de tipos de dados e aplicação de regras de negócio.
- O *Loading* (Carregamento) é a última etapa em que os dados transformados são inseridos, normalmente numa base de dados. Esta etapa deve ser melhorada para assegurar eficiência e minimizar impactos, no desempenho do sistema de destino.

As soluções de ETL melhoram a qualidade dos dados ao limpar as informações antes de as transferir para um novo repositório. Atualmente, há uma multiplicidade de ferramentas ETL (*open source* e em nuvem), que oferecem funcionalidades como automação completa, interfaces visuais intuitivas, suporte para o tratamento de dados complexos e conformidade com regulamentações de segurança (Ferreira, et al. 2010) (IBM, 2023).

A integração de dados pode ainda ser facilitada através de APIs, utilizando a Integração de Aplicações Empresariais, oferecendo uma solução mais flexível e escalável, particularmente em ambientes *web* (IBM, 2023).

O ETL é um componente crítico nos sistemas de informação modernos, desempenhando um papel vital na integração eficaz de dados heterogêneos. A seleção de ferramentas e técnicas de ETL deve ser meticulosamente ponderada, levando em consideração tanto os requisitos técnicos como os de negócio.

## 2.3 Web scraper e Web crawler

O **web scraper** é uma ferramenta ou programa de software desenvolvido para extrair informações de *websites*, cuja função é aceder a páginas específicas e recolher dados estruturados ou não estruturados. Este método é usado quando os sites não oferecem uma maneira formal de aceder às suas informações. Uma vez recolhidos, estes dados podem ser armazenados, analisados ou utilizados para diversos fins, como investigação, análise de mercado ou desenvolvimento de aplicações (Penã, 2012).

A operação de um **web scraper** inicia-se com o pedido ao servidor do *website* de onde se deseja extrair dados. Após receber a resposta do servidor, geralmente em formato HTML, o *scraper* analisa o conteúdo da página para identificar e extrair as informações desejadas. Para isso, o programa pode usar expressões regulares, árvores de sintaxe ou outras técnicas de processamento de texto.

É importante notar que o **web scraping** pode levantar questões éticas e legais. Nem todos os *websites* permitem a extração de seus dados e muitos, têm termos de serviço que proibem expressamente essa prática. Adicionalmente, o acesso excessivo a um *website*, por um *scraper* pode sobrecarregar o servidor, afetando a sua performance. Por isso, é essencial que os utilizadores de **web scrapers** sejam respeitosos e conscientes ao extrair informações da *web*, garantindo que não violem direitos de propriedade ou outros regulamentos (Krotov e Silva, 2018).

O **web scraping** é frequentemente realizado em ambiente Python, este método beneficia do uso de bibliotecas específicas como “*BeautifulSoup*” para análise de HTML e “*Requests*” para requisições HTTP. Adicionalmente, bibliotecas como “*Pandas*” podem ser empregues para manipulação e análise de dados, enquanto “*Selenium*” é muitas vezes utilizado para interagir com páginas *web* que requerem ações do utilizador, como preenchimento de formulários. O **web scraping** é vulnerável a alterações na estrutura do site alvo, tornando o código potencialmente obsoleto (Mitchell, 2015).

Um **web crawler**, frequentemente referido como “*spider*” ou “*robot*”, é um programa de software concebido para percorrer a *World Wide Web* de forma sistemática e automática, com o fim de recolher informações sobre os *websites*. A sua função é ler o conteúdo das páginas da *web* e criar entradas para um índice de busca, permitindo que motores, como o Google ou o Bing, forneçam resultados rápidos e precisos para as consultas dos utilizadores.

A operação de um **web crawler** começa com uma lista de endereços *web*, conhecidos como “*seeds*” (sementes). O *crawler* visita estes endereços e identifica todos os *hyperlinks* na página, adicionando-os à sua lista de páginas a serem visitadas. Este processo é repetido continuamente, permitindo ao *crawler* descobrir e indexar novas páginas à medida que surgem na *web* (Kausar, Dhaka e Singh, 2013).

Ao visitar um site, o **web crawler** não só recolhe informações sobre o conteúdo da página, mas também pode verificar meta-informações, como meta tags e cabeçalhos de resposta HTTP, para determinar a relevância e a frequência de atualização da página. É fundamental mencionar que a prática de *crawling* também tem implicações ética e legais. Muitos *websites* utilizam o arquivo “*robots.txt*” para especificar quais partes do site podem ser acedidos por *crawlers* e quais devem ser evitadas. Respeitar estas diretrizes é essencial para garantir que o *crawling* seja realizado de uma forma ética e sem sobrecarregar os servidores dos *websites*. Deve-se notar que uma abordagem eficaz de **web scraping** muitas vezes, incorpora elementos de *web*

*crawling*, especialmente quando se navega em sites de grande dimensão e complexidade.

## **2.4 API (*Application Programming Interface*)**

Uma API, (*Application Programming Interface*), é um conjunto de regras e especificações que permite a diferentes softwares comunicarem entre si. Funciona como uma ponte, possibilitando que aplicações distintas interajam e partilhem dados de forma estruturada, eficiente e segura. Este método é, muitas vezes, preferível devido à sua maior fiabilidade e conformidade com as políticas de uso de dados. No entanto, as APIs também apresentam limitações, como taxas de pedidos e uma dependência do fornecedor. Através de uma API, um software pode requisitar informações ou executar ações noutro software, sem necessidade de entender todo o seu funcionamento interno (Ofoeda, Boateng e Effah, 2019).

A utilização de APIs tornou-se essencial no mundo tecnológico atual, especialmente com o crescimento da computação em nuvem e das aplicações móveis. Há APIs que são disponibilizadas por empresas ou organizações e permitem, que desenvolvedores externos criem aplicações que interajam com os seus serviços, enquanto outras APIs são utilizadas internamente em empresas e organizações, para integrar os sistemas próprios.

É relevante destacar que, para garantir a segurança e a integridade dos dados, as APIs geralmente possuem mecanismos de autenticação e autorização, assegurando que apenas entidades autorizadas possam aceder e interagir com elas.

## **2.5 Google Sheets para sistemas de automação *Business Intelligence***

O Google Sheets é uma aplicação de folha de cálculo online que integra o lote de aplicações de escritório do Google Workspace. Tal como, outros programas semelhantes, como o Microsoft Excel, o Google Sheets permite a criação, edição e armazenamento de folhas de cálculo. A aplicação suporta fórmulas, gráficos e tabelas dinâmicas, permitindo uma ampla gama de funcionalidades.

No contexto do *Business Intelligence* (BI), o Google Sheets apresenta diversas vantagens. Primeiro, a sua base em nuvem facilita a acessibilidade e a colaboração em tempo útil. Com uma ligação à Internet, é possível aceder aos dados de qualquer dispositivo e múltiplos utilizadores, podem editar a mesma folha de cálculo simultaneamente. Esta característica torna o trabalho em equipa mais eficiente e colaborativo (Sierra, 2019).

Em termos de custo-benefício, o Google Sheets é uma opção atractiva, especialmente para pequenas e médias empresas. A aplicação oferece uma versão gratuita com uma variedade substancial de recursos, tornando-a uma alternativa económica a soluções de BI mais dispendiosas.

Além disso, o Google Sheets integra-se de forma fluida com outras ferramentas do Google Workspace, bem como, com uma série de plataformas e serviços externos, como o Google Analytics e o Google Ads. Esta integração é facilitada através do

Google Apps Script e de várias extensões disponíveis, permitindo uma maior flexibilidade e personalização (Google, 2023).

Finalmente, como o serviço é baseado em nuvem, todas as atualizações e adições de dados são reflectidas em tempo real, um aspecto crucial para a tomada de decisões de BI em tempo adequado e útil.

## 2.6 Google app scripts para sistemas de automação *Business Intelligence*

O Google Apps Script é uma linguagem de programação baseada em JavaScript que permite a automação e integração de serviços, dentro do ecossistema do Google Workspace. Lançada em 2009, esta ferramenta oferece uma série de funcionalidades que podem ser particularmente úteis para a automação de tarefas em sistemas de BI (Google, 2023).

A capacidade de integrar facilmente com outros serviços do Google, como Google Sheets e Google Data Studio, torna o Google Apps Script uma escolha atrativa para sistemas de BI que já utilizam estas plataformas. A execução em nuvem elimina a necessidade de infraestrutura dedicada, reduzindo assim os custos operacionais. A linguagem oferece uma grande flexibilidade na manipulação de dados, permitindo a criação de soluções personalizadas e a integração com outras APIs do Google e de terceiros, o que é crucial para a manipulação e análise de dados de um sistema BI. O Google Apps Script pode executar funções com base em acionadores temporais ou eventos específicos, uma característica altamente desejável em ambientes de BI, onde a atualização rápida é necessária (Google, 2023) (Sierra, 2019).

O Google Apps Script oferece uma série de vantagens para a automação de sistemas de BI, incluindo facilidade de integração, baixo custo e flexibilidade. No entanto, as suas limitações em termos de quotas de uso e capacidades de depuração (6 minutos de tempo máximo de execução por *script*) e armazenamento limitado, devem ser cuidadosamente consideradas tendo em conta o tamanho dos projetos (Google, 2023).

## 2.7 Exploração do *Machine Learning* e AI em *Business Intelligence*

Nos últimos anos, a integração de Inteligência Artificial (IA) e *Machine Learning* (ML) tem revolucionado o campo de *Business Intelligence* (BI). A AI, através do uso de *Machine Learning* (ML) e *Deep Learning* estão a permitir uma transição de sistemas reativos que oferecem *insights* baseados em dados históricos para sistemas proativos, capazes de prever tendências e tomar decisões autónomas. Em termos práticos, esta convergência tem inúmeras aplicações, desde a previsão e análise de sentimento do cliente até à operacionalização da *supply chain* (Boddu, et al., 2022).

No âmbito do *Digital Marketing*, estas tecnologias têm sido aplicadas em várias vertentes como *Customer Segmentation*, *Content Personalization* e *Sentiment Analysis*. As técnicas de ML e *Deep Learning* podem ser aplicadas para melhorar as campanhas de *marketing*, melhorar a Taxa de conversão (*Conversion Rate*) e aumentar o ROI (Nair e Gupta, 2021).

O *Machine Learning* ML é um subcampo da AI que foca na construção de algoritmos que permitem que as máquinas aprendam a partir dos dados. Modelos de ML como *Decision Trees* e *Support Vector Machines* têm sido usados para *Market Segmentation* e *Consumer Behavior Prediction* (S.Ullal, Hawaldar e Nadeem, 2022).

Na vertente da análise de dados, algoritmos de classificação podem ser empregues para categorizar tipos distintos de dados, como despesas num relatório financeiro. Além disso, pode-se utilizar algoritmos de séries temporais para prever tendências futuras, com base em dados históricos, útil para áreas de gestão. Também se destacam os algoritmos de deteção de anomalias, que identificam dados atípicos que podem ser resultado de erros de medição ou outras inconsistências. Na produção de indicadores e métricas de desempenho, a Inteligência Artificial pode desempenhar um papel fundamental na avaliação automática de KPIs (Indicadores-Chave de Desempenho), gerando alertas ou recomendações para ações corretivas ou de melhoria (S.Ullal, Hawaldar e Nadeem, 2022) (Boddu, et al., 2022).

Para concretizar a integração com o Google Sheets, diversas estratégias podem ser adotadas. O Google Apps Script permite uma integração direta com modelos de *Machine Learning*, possibilitando que dados sejam importados para uma folha de cálculo, analisados e, posteriormente, apresentados de forma compreensível, bibliotecas como “*TensorFlow*” e “*BigQuery ML*” podem ser integradas para análises mais complexas (Google Cloud, 2023).

A implementação dessas tecnologias também pode ser complexa e requerer um investimento inicial significativo. É recomendável que se proceda com uma validação rigorosa dos dados, um sistema de monitorização contínuo e, possivelmente, testes para avaliar a eficácia das métricas e indicadores gerados (Nair e Gupta, 2021).



### 3 Caso de estudo

Na empresa de publicidade Clever Advertising, no departamento dedicado a *influencer marketing* e denominado “Bold”, foi observado que toda a equipa estava envolvida diariamente, em tarefas de atualização manual de diversas folhas de cálculo, com várias tabelas de dados. Se inicialmente, essa abordagem era viável, devido ao rápido crescimento de informação e volume de negócios, essa abordagem tornou-se pouco eficiente e desatualizada, não permitindo uma gestão eficaz de informação e tempo.

Observou-se a inexistência de uma produção sistemática de relatórios para análise de dados, o que poderia condicionar o controlo e acompanhamento da eficácia dos resultados dos conteúdos e campanhas produzidos pelos *influencers*. Os relatórios elaborados anteriormente pela equipa eram demorados, por necessitarem da extração de dados dos *websites* das empresas clientes para o período desejado e posterior análise. A inexistência da produção atempada dos relatórios podia influenciar a possibilidade de negociar contratos em tempo útil com os *influencers*, mediante os objetivos cumpridos ou permitir contratos mais favoráveis à empresa.

Há uma grande diversidade de dados no *influencer marketing*, pois a eficácia da publicidade nas redes sociais tem vindo a aumentar. Por exemplo, só para o caso dos *influencers* (indivíduos que têm presença nas redes sociais) e que têm milhares/milhões de seguidores com um tipo de comunicação e comportamentos muito diversos, o tratamento da informação que resulta dos conteúdos publicitários por eles publicados nas redes sociais é muito diversificado e complexo.

Grande parte dos dados em questão e que necessitavam da automatização, resultavam das informações sobre os resultados dos conteúdos publicitadas pelos *influencers*, em forma de *post*, vídeo e *livestream*.

Sendo assim, foram estudadas as áreas de funcionamento da equipa “Bold” como o processo de gestão de fluxo de dados e a produção de relatórios.

#### 3.1 Processo de gestão de fluxo de dados

A equipa “Bold” realizava diariamente, o processo de aceder às plataformas *web* das empresas clientes publicitadas, que forneciam os dados sobre o resultado do *marketing* digital, para gerar os resultados dos dias anteriores e, em seguida, copia os dados para uma folha de Excel. Posteriormente, procediam ao tratamento como, eliminação de colunas não necessárias para estes registos e formatação dos dados, como por exemplo, alterar a data do formato “mm/dd /yyyy” para “dd/mm/yyyy” antes de os copiarem para o Google Sheets. No Google Sheets ainda realizavam outras formatações de valores, como por exemplo, um identificador referente ao ID do *influencer* com a data dos resultados.

Como este processo era realizado em vários *websites* e cada *website* distribuía os dados dos *influencers* por país, havia dezenas de folhas de cálculo para atualizar diariamente.

Pela observação da pouca eficácia da gestão deste fluxo de dados e da produtividade dos trabalhadores, relacionados com a atualização de bases de dados de *marketing*

digital para análise, controlo e armazenamento, foi proposto um estudo do problema e a apresentação de possíveis soluções.

### 3.2 Relatórios automatizados e monitorização

O *marketing* digital, especialmente no contexto de colaborações com *influencers*, requer uma monitorização rigorosa para avaliar o impacto e a eficácia das campanhas. No entanto, observou-se que a equipa carecia de um sistema automatizado e eficiente para a elaboração de relatórios e análise de dados, o que poderia ter implicações negativas, tanto a curto como a longo prazo.

A ausência de uma estrutura organizada para a elaboração de relatórios e análise de dados em tempo útil, representava uma lacuna significativa na eficácia operacional da equipa da “Bold”, particularmente no contexto de parcerias com *influencers* através de campanhas e geração de conteúdos nas redes sociais. Esta deficiência não só comprometia a capacidade de avaliar o retorno sobre o investimento (ROI) de forma eficaz, mas também limitava as oportunidades de negociação de contratos mais vantajosos.

Foi observado que os relatórios elaborados pela equipa eram extremamente demorados, pois a extração de dados dos *websites* das empresas clientes era um processo moroso e ineficiente, frequentemente limitado a intervalos de três meses, o que para maiores períodos de análise, se tornava ainda mais complicado a extração da informação, o que implicava maiores atrasos na elaboração de relatórios. Não existindo um sistema automatizado para gerar relatórios, resultava maiores esforços manuais redundantes e ineficientes por parte da equipa. A falta de dados atualizados e de relatórios em tempo útil, prejudicava as negociações de contratos, especialmente naquelas em que o tempo é um fator crítico.

Dada a concorrência no setor, a informação correta, pertinente em tempo útil e a tempo das renegociações e negociações, é fundamental para uma decisão e definição de valores dos contratos.

Foi definido criar uma solução para a implementação de um sistema de relatórios atualizados, mais eficiente e automatizado. É necessária uma monitorização, pois não é apenas uma necessidade operacional, mas uma estratégia crítica, para melhorar a eficácia do *marketing* digital e as negociações de contratos.

## 4 Desenvolvimento e aplicação da solução de *Business Intelligence*

A apresentação de soluções à empresa foi um processo de diálogo e discussão das tecnologias disponíveis, de avaliação de timings e capacidade de execução em tempo útil, tendo em consideração os critérios como, o tempo de implementação do projeto, a minimização dos custos, os recursos disponíveis, as questões legais e as tecnologias atuais. O desenvolvimento e aplicação da solução de *Business Intelligence* passou por: apresentação e escolha de soluções, automação de fluxos de dados: processo ETL, geração automática de relatórios via Google Sheets e *dashboard*, análise de dados e produção de indicadores e métricas de desempenho, e respetiva discussão que se descrevem a seguir.

### 4.1 Apresentação e escolha de soluções

A criação da automação de um sistema *Business Intelligence* passou pelo estudo e apresentação de 2 etapas (Ver Figura 1 - Apresentação de soluções):

- O processo ETL;
- O processo de geração automática de relatórios e análise de dados.

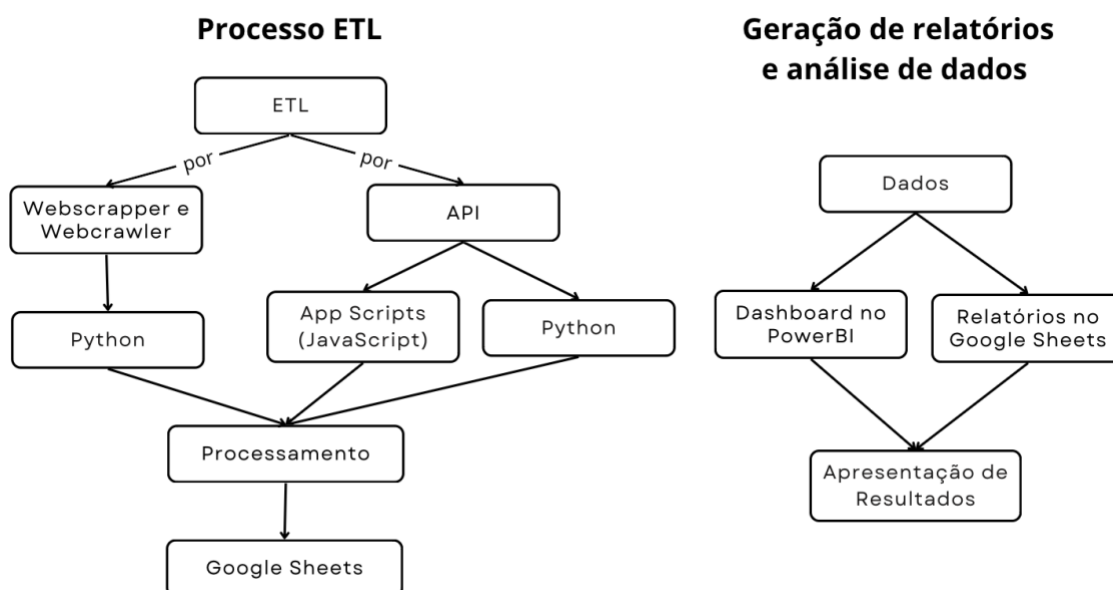


Figura 1 - Apresentação de soluções

Para o desenvolvimento do processo ETL foram estudadas e apresentadas soluções viáveis para melhorar a eficiência operacional da equipa. Em primeiro lugar, foi necessário optar por qual opção usar para proceder com o processo ETL:

- 1ª Opção - a implementação de um *web scraper* e *web crawler* utilizando Python;
- 2ª Opção - a extração de dados via APIs, por Python ou App Scripts (JavaScript).

Cada opção possui as suas vantagens e desvantagens, que foram discutidas e avaliadas com a empresa, conforme os critérios já estabelecidos.

A decisão de optar pela utilização de APIs para o processo ETL derivou da sua abordagem direta, estruturada e consistente na extração de dados, ao contrário do *web scraper* e *web crawler* que além da sua complexidade, podem necessitar de constantes atualizações. As APIs, frequentemente disponibilizadas por empresas e plataformas, permitem um acesso controlado e confiável dos dados, apresentando menos riscos legais e potencialmente menores custos associados. Dado a duração do estágio e a necessidade de uma solução eficiente, a escolha de APIs alinhou-se totalmente com os objetivos do projeto, maximizando a eficiência e minimizando os desafios.

Posteriormente, o App Scripts (JavaScript) para automação da extração dos dados via API emergiu como a opção mais viável. A sua integração nativa com o ecossistema Google, incluindo o Google Sheets, já em utilização pela equipa, foi um fator determinante, o que permitiu uma implementação mais rápida e eficiente para o estudo e desenvolvimento. Salienta-se ainda, como os *scripts* são executados nos servidores da Google, evitou à empresa custos adicionais na criação de infraestruturas.

É importante referir que foram desenvolvidos algoritmos em Python no editor de código Visual Studio Code da Microsoft de *web scraper*, *web crawler* e extração via API, com a utilização de bibliotecas como “*BeautifulSoup*”, “*Requests*”, “*Pandas*” e “*Selenium*”, etc... Mas como a solução escolhida foi a App Scripts não houve necessidade de aplicação dos algoritmos, no entanto poderão ser usados, caso seja necessário existir a transição para o Python, pois já incorporam o tratamento dos dados e a estrutura dos *websites* das empresas clientes. Foi estudado ainda comandos básicos de Git para boas práticas no rastreio do desenvolvimento destes algoritmos.

Em seguida, apresentou-se à empresa, a organização e geração automatizada de relatórios por Google Sheets com o auxílio da App Scripts para abordar as necessidades imediatas relacionadas ao sistema de *Business Intelligence* (BI). A empresa concordou com a proposta em gerar relatórios por Google Sheets com o auxílio também da App Scripts.

Para além da geração de relatórios durante o desenvolvimento do projeto foi apresentada como ferramenta adicional para interpretação da informação, um *dashboard* para a visualização de resultados com o software da Microsoft, PowerBI.

Foi estudada, uma exploração de ferramentas de *Machine Learning* como o “*TensorFlow*” e “*BigQuery ML*”, pela boa integração no ecossistema Google e facilidade de utilização para realizar análises preditivas, no entanto a empresa não optou por uma abordagem deste género.

## 4.2 Automação do fluxo de dados: Processo ETL

Na criação e desenvolvimento do processo de ETL foi necessário acompanhamento com a equipa da “Bold” para compreensão do armazenamento dos dados, suas utilidades e também o seu processamento. Cada folha de cálculo tinha um armazenamento e tratamento de dados diferentes e cada API das empresas clientes, fornecia os dados com diferentes formatos o que implicou desenvolver *scripts* diferentes, para diferentes folhas de cálculo (em anexo - Programa 1 e Programa 2). Nalguns casos, adaptou-se a extração e o armazenamento e tratamento de dados para a formatação da API, noutros, adaptou-se a importação de dados para o

armazenamento e processamento, já realizados pela equipa no Google Sheets. Todos os *scripts* executavam numa média de três minutos, por prevenção foi criado um algoritmo de “*check up*” (em anexo) que irá verificar e atualizar os dados, caso exista um erro por *time out* ou um erro importação na solicitação (“*request*”) à API.

Os *scripts* criados para a automação do fluxo de dados têm um ativador que ocorre diariamente, com o objetivo de substituir o trabalho de rotina da equipa “Bold”, com a função de retirar e tratar das informações referentes até ao antepenúltimo dia, que antecedia a data corrente.

Para o desenvolvimento dos *scripts* em App Scripts foi necessário realizar o estudo de programação em JavaScript.

Foi escolhido o formato de dados a ser obtido através da API. As opções disponíveis eram CSV (*Comma-Separated Values*) ou XML (*eXtensible Markup Language*). Escolheu-se o formato CSV, uma decisão fundamentada na facilidade de integração deste formato com a plataforma Google Sheets.

O *script* define variáveis essenciais, como o ID da folha de cálculo do Google Sheets, data corrente (dd/mm/yyyy) e (dd/mm/yyyy - 2 dias).

Em seguida, o *script* formatava as datas nos formatos devidos e realizava as ações relacionadas com a gestão da folha de cálculo, como a duplicação mensal de um *template*, para registo das campanhas dos *influencers* com os resultados gerados.

A função “*duplicarFolhaEInserirData*” é ativada todos os meses no seu primeiro dia e irá duplicar um *template* (Ver Figura 2) já criado no Google Sheets, modificando o *template* para o mês em questão. É necessário referir que este *template* apenas funcionará para controlo e se necessário para registo dos conteúdos gerados pelos *influencers* por parte da equipa “Bold”.





| OLD ID | NEW ID | IG | YT | T | STREAMER | 01/08/2023  |   |     |     |
|--------|--------|----|----|---|----------|---|---|-----|-----|
|        |        |    |    |   |          | terça-feira   |   |     |     |
|        |        |    |    |   |          |  |  |     |     |
|        |        |    |    |   |          |  |  | REG | FTD |

Figura 2 - Exemplo de um cabeçalho para anotação dos conteúdos gerados

Seguidamente, realizou-se a construção do URL da API que é usada para extrair os dados necessários. A chave da API (denominada “*apiKey*” no código) é uma informação sensível e confidencial, que permite o acesso à API externa. Para garantir a segurança, a chave é armazenada de forma encriptada e é recuperada quando necessário, garantindo uma camada extra de segurança. Posteriormente, o URL da API é formatado para incluir esta “*apiKey*” e outros parâmetros necessários para a consulta. Um dos parâmetros mais cruciais nesta URL é o intervalo de datas para o qual se deseja obter os dados. A função “*getLatestDateAboveLastRow*” determina a última data para a qual já se tem dados, para que ao fazer uma nova consulta à API, se garanta a continuidade e a integridade dos dados ao longo do tempo. O *script* realiza uma solicitação (*HTTP GET*) à API e obter os dados no formato CSV, esses dados são processados usando a biblioteca “*CSVParser*” ficando uma estrutura de dados utilizável. Depois de processar os dados CSV, são executadas algumas operações para limpar e estruturar esses dados, de acordo com os requisitos pedidos pela empresa. Em seguida, percorrem-se os dados e extraem-se informações específicas, para serem inseridas nas colunas apropriadas da folha de cálculo. O

*script* ainda cria na coluna A um identificador para cada *influencer* para os resultados diários, que é definido como o ID e insere os dados processados nas colunas apropriadas, da folha de cálculo.

O *script*, em resumo, extrai dados via API em formato CSV, processa e insere na folha de cálculo do Google Sheets, de acordo com a formação e estrutura desejada (Ver Figura 3 - Fluxograma dos *scripts*).

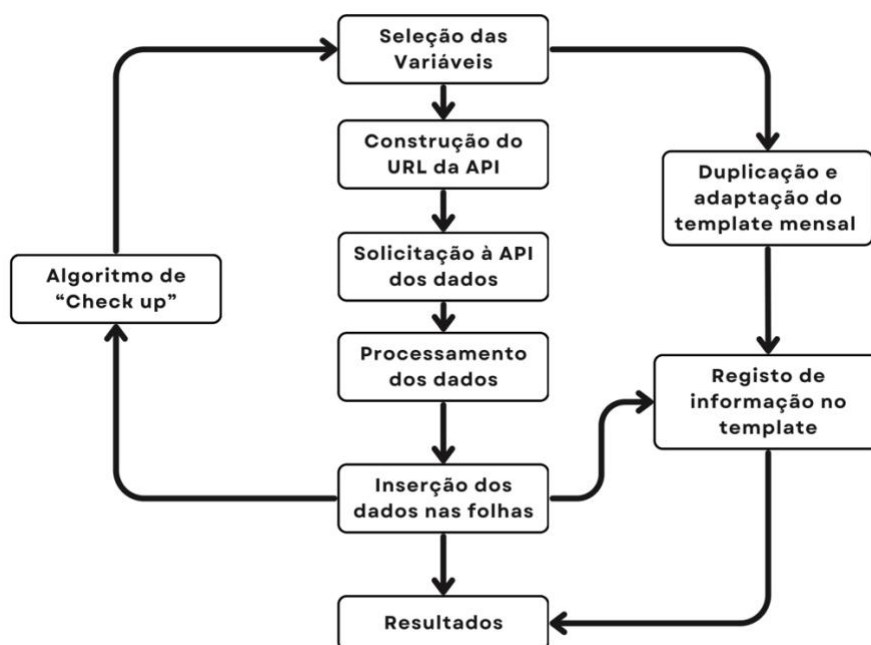


Figura 3 - Fluxograma dos *scripts*

Durante a semana de teste e aplicação, não ocorreram nenhuns problemas, anomalias nem pedidos de alteração à solução criada, satisfazendo plenamente, os pedidos da equipa “Bold” e da empresa. Os elementos da equipa “Bold” adstritos às tarefas automatizadas tiveram um ganho estimado de 1h30min por dia.

#### 4.3 Geração automática de relatórios via Google Sheets e *dashboard*

O processo de produção de relatórios automatizados levou à criação de vários *templates* capazes de analisar e apresentar indicadores numéricos e visuais dos resultados obtidos pelas campanhas e conteúdos gerados pelos *influencers*.

Durante o desenvolvimento da geração automática dos relatórios, existiu a necessidade de criar um *script* (em anexo) para importar dados de *influencers*, para realizar uma análise mais completa e personalizada, pois existiam períodos que não tinham sido armazenados pela equipa da “Bold”.

O novo *script* desenvolvido foi adaptado da extração de dados diários. A funcionalidade de extração de dados para um período pré-determinado foi retirada; em contrapartida, implementou-se um novo mecanismo, que concede ao utilizador a liberdade de selecionar o intervalo temporal para o *influencer* ou *influencers*, dos quais se pretendem realizar a análise. Salienta-se que a empresa só começou a

armazenar dados a partir de uma data recente. Em virtude disto, a extração de dados do *influencer* torna-se indispensável, caso o período ainda não esteja armazenado.

O *template* apresenta vários gráficos, que facilitam a visualização de resultados e a evolução do desempenho dos *influencers* ao longo do tempo, que mostram o número de registros, receitas, clientes, clientes que repetem a compra, cliques e interações ao longo do tempo, podendo oferecer uma melhor visualização e compreensão dos resultados gerados pelo *influencer*. Essas métricas visuais podem ser particularmente úteis, para avaliações em tempo real e ajustes de estratégia.

Para complementar os relatórios em Google Sheets e proporcionar uma melhor visualização de resultados, foi desenvolvido um PowerBI para melhor apresentação dos dados (Ver Figura 4 - *Dashboard* do PowerBI). No PowerBI é possível selecionar o intervalo de tempo pretendido para análise e visualização, sendo também possível realizar outras seleções, como visualizar o número de *clicks* e outros indicadores pedidos pela equipa da “Bold”. Os dados foram carregados manualmente pela transferência do Google Sheets para um Excel e depois exportados para o PowerBI.



Figura 4 - *Dashboard* do PowerBI

Durante a semana de teste e aplicação para a geração de relatórios, as soluções encontradas satisfizeram plenamente os pedidos da equipa “Bold” e da empresa. Os elementos da equipa “Bold” adstritos às tarefas automatizados tiveram um ganho estimado de 1h por relatório.

#### 4.4 Análise de Dados e produção de indicadores e métricas de desempenho

Para a análise de dados, foram criados diversos *templates* (Ver Figura 5 - Exemplo de *template* com indicadores e métricas) que incluem uma análise de dados pré-definida. Esses *templates* funcionam eficazmente porque os dados estão todos

armazenados de forma uniforme, permitindo a criação de *templates* para todas as configurações de armazenamento de dados.

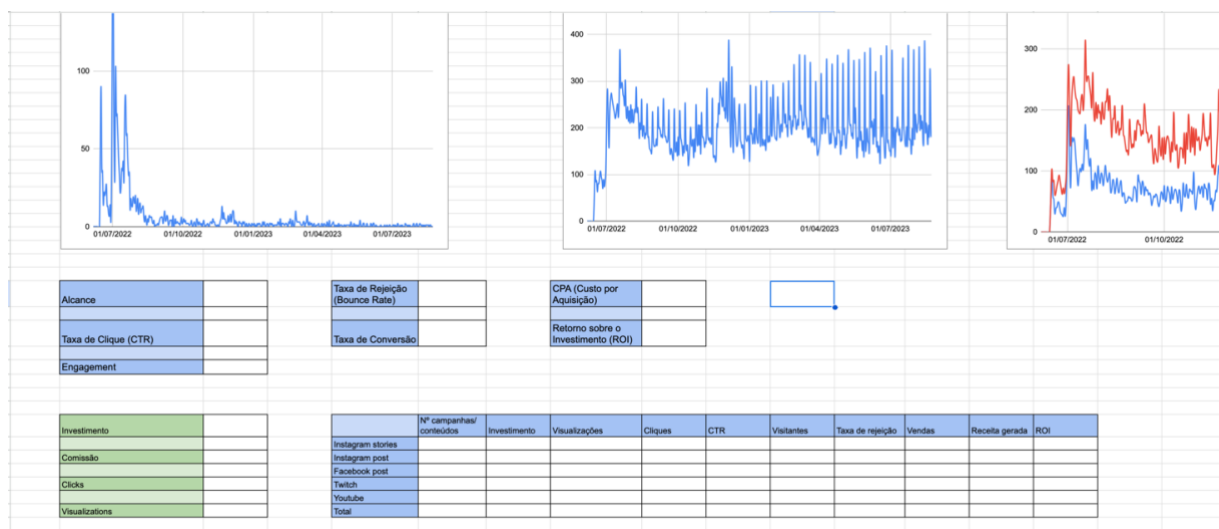


Figura 5 - Exemplo de *template* com indicadores e métricas

Para apresentar os relatórios foi necessário definir métricas e indicadores de desempenho (KPIs) apropriados para avaliar a eficácia dos *influencers*. As métricas e os indicadores de desempenho utilizados na geração de resultados para a automação BI foram os seguintes:

- **Alcance** - mede o tamanho da audiência do *influencer*. É calculado como o número total de visualizações.
- **Taxa de Clique (CTR)** - Percentagem de pessoas que clicam em um anúncio em relação ao número de visualizações do anúncio.
- **Engagement** - Esta métrica avalia o nível de interação da audiência com o conteúdo do *influencer*. Pode incluir vários indicadores como o número total de *likes*, partilhas e comentários. No contexto deste projeto, as métricas de cliques, interações e registos poderiam ser usadas para avaliar o *engagement*.
- **Taxa de Rejeição (Bounce Rate)** - Percentual de visitantes e que saem do site após visualizar apenas uma página. Uma taxa alta pode indicar que o conteúdo ou design da página não é relevante ou atraente.
- **Taxa de Conversão (Conversion Rate)** - A taxa de conversão mede a eficácia do *influencer* em motivar a audiência a tomar uma ação desejada, como fazer uma compra ou registrar-se. É calculada como o número de ações desejadas realizadas dividido pelo número total de interações.
- **CPA (Custo por Aquisição)** - O CPA é uma métrica que ajuda a entender o custo associado à aquisição de um cliente através das atividades promocionais do *influencer*. É calculado dividindo o custo total da campanha ou do *influencer* pelo número de aquisições (por exemplo, novos clientes, vendas, etc.).
- **Retorno sobre o Investimento (ROI)**- O ROI por *influencer* é uma métrica financeira que avalia a eficácia do investimento feito no *influencer* ou na campanha. É calculado usando a fórmula (1):

$$ROI = \left( \frac{(\text{Receitas obtidas} - \text{Custo do Investimento})}{\text{Custo do Investimento}} \right) * 100 \quad (1)$$



- **Receitas Geradas** - As Receitas Geradas referem-se ao total de receitas obtidas através das atividades promocionais do *influencer*.

#### 4.5 Discussão

O projecto destaca-se pela implementação de uma solução de *Business Intelligence* bem estruturada e eficiente. Ao optar pela utilização de App Scripts em JavaScript, conseguiu-se alinhar com ferramentas já utilizadas pela a equipa, como o Google Sheets. Esta escolha não só facilitou a integração, como também reduziu a curva de aprendizagem, permitindo uma implementação mais rápida e eficaz. Este alinhamento estratégico com as ferramentas existentes foi um factor crítico de sucesso.

A decisão de utilizar APIs, em detrimento de *web scraping*, para o processo de ETL foi estratégica. Além de minimizar os riscos legais, esta opção também garantiu uma maior fiabilidade e consistência na extração de dados.

O ganho de eficiência foi considerável, com a equipa a poupar estimadamente 1h30min por elemento adstrito ao processo ETL e cerca de 1h por relatório gerado. Estes números ilustram o impacto direto e mensurável na produtividade produzida pela solução, justificando plenamente o trabalho desenvolvido.

Ainda vale a pena referir trabalho efetuado, mas não utilizado, em Python para *web scraping*, *web crawling* e extração via API. Estes algoritmos representam um recurso para empresa, possibilitando flexibilidade para eventuais futuras soluções.

Outro aspecto interessante do projeto foi o estudo preliminar em Inteligência Artificial e *Machine Learning*, que demonstra um olhar atento às tecnologias emergentes. Esta investigação prévia em AI pode servir como uma base de trabalho para futuras iterações do projeto, permitindo análises mais complexas e *insights* mais profundos.

Em resumo, o projecto apresenta uma implementação bem-sucedida de uma solução de BI, com ganhos mensuráveis em eficiência e uma integração cuidada, com as ferramentas já implementadas pela equipa.

## 5 Conclusão

O projeto com a implementação de um sistema de *Business Intelligence* na empresa Clever Advertising, representou um avanço significativo na sua eficiência operacional, na gestão e análise de dados.

### 5.1 Principais contribuições

O projeto na Clever Advertising teve como principal contribuição o desenvolvimento de ferramentas computacionais para automação de um sistema *Business Intelligence* funcional, viável e eficiente.

Com as principais contribuições deste projeto, a equipa “Bold” obteve uma melhoria significativa na eficiência operacional, economizando aproximadamente 1h30min por dia e elemento em tarefas de extração, transformação e carregamento de dados, e cerca de 1h por relatório gerado. Esta economia de tempo foi em grande parte devida ao estudo, planeamento e escolha acertada das tecnologias.

Optou-se por utilizar APIs para o processo de extração de dados, uma decisão que não só minimizou os riscos legais, mas também garantiu uma maior fiabilidade e uma maior rapidez na obtenção dos dados. Esta escolha tecnológica também se mostrou estrategicamente alinhada com as ferramentas já em uso pela equipa. A implementação de App Scripts em JavaScript facilitou a integração com o Google Sheets, acelerando assim o processo de implementação e diminuindo a curva de aprendizagem para os membros da equipa.

No entanto, o projeto não se limitou a solucionar problemas do presente. Além de implementar melhorias imediatas, também estabeleceu uma base sólida para futuras inovações. Os algoritmos desenvolvidos em Python, embora não utilizados, acrescentaram uma possibilidade de flexibilidade, preparando a empresa para eventuais transições tecnológicas que possam surgir.

Não menos importante, foi o foco na monitorização e análise avançada dos dados. Foram desenvolvidos *templates* para a geração automática de relatórios e criados *dashboards* no PowerBI, que mostra a diversidade e amplitude das soluções desenvolvidas, permitindo uma melhor visualização e compreensão dos dados, um conhecimento crucial para a tomada de decisões informadas. Além disso, o projeto fez um estudo preliminar em Inteligência Artificial e *Machine Learning*, revelando interesse e preparação para adotar tecnologias mais avançadas, no futuro.

No que diz respeito à análise de dados, a definição e desenvolvimento de métricas e indicadores de desempenho, como o Retorno sobre o Investimento (ROI), a Taxa de Clique e a Taxa de Conversão, permitiram uma avaliação mais precisa da eficácia das campanhas e do desempenho dos *influencers* para a adoção de estratégias, para melhorar a eficácia do *marketing* digital e as negociações de contratos.

Em suma, o projeto resultou numa implementação de sucesso de um sistema *Business Intelligence* contribuindo significativamente para o melhoramento e operacionalização do fluxo, gestão e análise de dados, tendo sido implementado sem qualquer tipo de perdas de tempo, dados ou produtividade, resultando numa melhoria da eficiência operacional.

O projeto teve um impacto positivo na Clever Advertising permitindo a criação de um sistema *Business Intelligence* e outras estruturas para futuras aplicações.

## 5.2 Limitações do trabalho e perspectivas de trabalho futuro

O trabalho desenvolvido apresentou algumas limitações, nomeadamente ao nível do tempo disponível relativamente à complexidade associada das tarefas realizadas. A carga horária foi um fator limitante para aprofundar algumas áreas, especialmente considerando a complexidade de algumas tarefas e a necessidade de estudo prévio para a sua realização. Além disso, as opções da empresa para a minimização de custos e minimização de questões legais e de confidencialidade, condicionaram as soluções implementadas.

A empresa não enquadrava a exploração nem o estudo de Inteligência Artificial (AI) na proposta inicial, contudo, por iniciativa do autor do relatório, o estudo de AI e *Machine Learning* (ML) foi incluído nos objetivos do projeto. Este estudo trouxe um valor acrescido ao trabalho, tornando os produtos acabados mais propícios a uma abordagem futura que incorpore AI e ML.

No que diz respeito às perspectivas de trabalho futuro, é evidente o potencial para a implementação de AI e ML na automação de um *Business Intelligence* especialmente na análise de dados, tanto no Google Sheets como no PowerBI. As automações já implementadas no processo ETL via App Scripts, os algoritmos em Python para *web scraping* e *web crawling* e extração via API, oferecem uma base sólida para esta implementação. Adicionalmente, os templates com indicadores e métricas já produzidos podem ser adaptados para acomodar soluções de AI. Entre as soluções de AI que se poderiam aplicar incluem-se análises preditivas de *Customer Segmentation* e *Content Personalization* para melhorar as campanhas de *marketing* digital.

Deste modo, o trabalho desenvolvido e efetuado serve não apenas como uma solução para as necessidades imediatas, mas também como um alicerce para futuros desenvolvimentos.

## Referências

- Biplob, Badiuzzaman, e Mokammel Haque. 2022. "Development of an Efficient ETL Technique for Data Warehouses."
- Boddu, Raja Sarath Kumar, Ashwinkumar A. Santoki, Shopita Khurana, Poonam Vitthal Koli, and Ravi Rai. 2022. "An analysis to understand the role of machine learning, robotics and artificial intelligence in digital marketing."
- Chen, H., R. H. Chiang, and V. C. Storey. 2012. "Business Intelligence and Analytics: From Big Data to Big Impact." (MIS Quarterly.).
- Cloud Google. 2023. *Google Cloud*. Acedido em setembro de 2023. <https://cloud.google.com/bigquery/docs/bqml-introduction>.
- Ferreira, João, Miguel Miranda, António Abelha, and José Machado. 2010. "O Processo ETL em Sistemas Data Warehouse."
- Glez-Penã, Daniel, Analia Lourenco, Hugo Lopez-Fernandez, and Miguel Reboiro-Jato. 2013. "Web scraping technologies in an API world ."
- Google. 2023. Acedido em setembro de 2023. <https://developers.google.com/apps-script/reference>.
- IBM. 2023. Acedido em setembro de 2023. <https://www.ibm.com/topics/etl>.
- Kausar, Md. Abu, V. S. Dhaka, and Sanjeev Kumar Singh. 2013. "Web Crawler: A Review ." (International Journal of Computer Applications).
- Krotov, Vlad, and Leiser Silva. 2018. "Legality and Ethics of Web Scraping."
- Microsoft. 2023. Acedido em setembro de 2023. <https://powerbi.microsoft.com/pt-pt/what-is-business-intelligence/>.
- Mitchell, R. 2015. *Web Scraping with Python: A Comprehensive Guide*. . O'Reilly Media.
- Nair, Kiran, and Ruchi Gupta. 2021. "Application of AI technology in modern digital marketing environment." *World Journal of Entrepreneurship, Management and Sustainable Development*.
- Ofoeda, Joshua, Richard Boateng, and John Effah. 2019. *Application Programming Interface (API) Research: A Review of the Past to Inform the Future*.
- S.Ullal, Mithun, Iqbal Thonse Hawaldar, and Mohammed Nadeem. 2022. "The Role of Machine Learning in Digital Marketing." *Sage Journal*.
- Sierra, Cristóbal Leiva. 2019. "Modelo de Negocio basado en una metodología de conversión y desarrollo de sistemas basados en Excel VBA hacia tecnologías de Google Apps Scripts."
- Vassiliadis, P., A. Simitsis, and S. Skiadopoulos. 2002. "Modeling ETL activities as graphs."
- Wilkinson, Kevin, Alkis Simitsis, Malu Castellanos, and Umeshwar Dayal . 2010. *Leveraging Business Process Models for ETL Design*.

## Anexos

### Processo ETL- Programa 1

```
function updateSheetWithCSV() {  
    var spreadsheetId = 'xxxxxxxx';  
    var SheetName = 'REPORTS'; // Caso o Sheetname troque de nome é necessário que se troque para o novo  
  
    // Obtém a data de hoje  
    var today = new Date();  
    var yesterday = new Date(today);  
    yesterday.setDate(yesterday.getDate() - 2);  
  
    // Formata as datas no formato "yyyy/M/d"  
    var formattedToday = Utilities.formatDate(today, "GMT", "yyyy/M/d");  
    var formattedYesterday = Utilities.formatDate(yesterday, "GMT", "yyyy/M/d");  
  
    duplicarFolhaEInserirData('Template', yesterday, spreadsheetId);  
  
    Logger.log('Data inicio: ' + getLatestDateAboveLastRow());  
    Logger.log('Data fim: ' + formattedYesterday);  
  
    var apiKey = PropertiesService.getScriptProperties().getProperty('apiKey');  
    // Constrói a URL da API usando a API Key recuperada  
    var apiUrl = 'https://xxxxxxxxxxxxxxxxxxxx.com/api/affreporting.asp?key=' + apiKey  
    + '&reportname=Member%20Report%20-%20Detailed&reportformat=csv&reportmerchantid=0&reportstartdate='  
    + getLatestDateAboveLastRow() + '&reportenddate=' + formattedYesterday;  
  
    var response = UrlFetchApp.fetch(apiUrl);  
    var csvData = response.getContentText();  
  
    var Sheet = SpreadsheetApp.openById(spreadsheetId).getSheetByName(SheetName);  
  
    // Processa o CSV usando a biblioteca CSVParser
```

```

var csv = Utilities.newBlob(csvData).getDataAsString();
var parsedData = Utilities.parseCsv(csv);

parsedData.shift();
parsedData.shift();
parsedData.shift();
parsedData.shift();
parsedData.shift();
parsedData.shift();

parsedData.pop();

var numRows = parsedData.length;
var numColumns = parsedData[0].length;

var lastRow = Sheet.getRange("A" + Sheet.getLastRow()).getRow();
var valuesToFill = [];
var valuesToFillE = [];
var valuesToFillO = [];
var valuesToFillF = [];
var valuesToFillI = [];
var valuesToFillJ = [];
for (var i = 0; i < numRows; i++) {
    var dateValue = contarDiasDesde1900(parsedData[i][3]);
    valuesToFill.push([dateValue + "|" + parsedData[i][9]]);

    var formattedDate = convertToDDMMYYYY(parsedData[i][3]);
    valuesToFillE.push([formattedDate]);

    var reg = parsedData[i][15];
    valuesToFillO.push([reg]);

    var ftd = parsedData[i][20];
    valuesToFillF.push([ftd]);

```

```

    var id = parsedData[i][9];
    valuesToFillI.push([id]);

    var nameid = parsedData[i][10];
    valuesToFillJ.push([nameid]);
  }
  var rangeColumnA = Sheet.getRange(lastRow + 1, 1, numRows, 1);
  var rangeColumnE = Sheet.getRange(lastRow + 1, 5, numRows, 1); // Coluna E é a
coluna 5
  var rangeColumnO = Sheet.getRange(lastRow + 1, 15, numRows, 1); // Coluna O é a
coluna 15
  var rangeColumnF = Sheet.getRange(lastRow + 1, 20, numRows, 1); // Coluna M é a
coluna 20
  var rangeColumnI = Sheet.getRange(lastRow + 1, 9, numRows, 1); // Coluna I é a
coluna 9
  var rangeColumnJ = Sheet.getRange(lastRow + 1, 10, numRows, 1); // Coluna J é a
coluna 10

  rangeColumnA.setValues(valuesToFill);
  rangeColumnE.setValues(valuesToFillE);
  rangeColumnO.setValues(valuesToFillO);
  rangeColumnF.setValues(valuesToFillF);
  rangeColumnI.setValues(valuesToFillI);
  rangeColumnJ.setValues(valuesToFillJ);

  // Registra o número de linhas adicionadas no Logger
  Logger.log('Ultima linha: ' + lastRow+ " + " + numRows);
}

```

## Funções

```

function duplicarFolhaEInserirData(SheetName, date, spreadsheetId) {
  // Define the name of the new Sheet based on the date
  var nomeNovaFolha = getNomeMesProximoAno(date);

  var ss = SpreadsheetApp.getActiveSpreadSheet();

```

```

// Check if a Sheet with the same name already exists
var existingSheet = ss.getSheetByName(nomeNovaFolha);
if (existingSheet) {
    Logger.log("Sheet with name '" + nomeNovaFolha + "' already exists. Skipping...");
    return;
}

var                folhaOrigem                =
SpreadSheetApp.openById(spreadSheetId).getSheetByName(SheetName);

var novaFolha = folhaOrigem.copyTo(ss);
novaFolha.setName(nomeNovaFolha);

var primeiroDiaProximoMes = new Date(date.getFullYear(), date.getMonth(), 1+1);
var    dataFormatada    =    Utilities.formatDate(primeiroDiaProximoMes,    'GMT',
'd/MM/yyyy');
novaFolha.getRange('M2').setValue(dataFormatada);
}

function getNomeMesProximoAno(date) {
    var proximoMes = new Date(date.getFullYear(), date.getMonth());
    var nomeMeses = ['JANEIRO', 'FEVEREIRO', 'MARÇO', 'ABRIL', 'MAIO', 'JUNHO', 'JULHO',
'AGOSTO', 'SETEMBRO', 'OUTUBRO', 'NOVEMBRO', 'DEZEMBRO'];
    var nomeMes = nomeMeses[proximoMes.getMonth()];
    return "CALENDÁRIO " + nomeMes + ' ' + proximoMes.getFullYear();
}

function contarDiasDesde1900(data) {
    var dataBase = new Date("1900-01-01");
    var dataFornecida = new Date(data);

    var diferencaEmMilissegundos = dataFornecida.getTime() - dataBase.getTime();

    // Converte a diferença em dias
    var diasDesde1900 = Math.floor(diferencaEmMilissegundos / (1000 * 60 * 60 * 24)+3);

```



```

    return diasDesde1900;
}

function convertToDDMMYYYY(dateString) {
    // Verifica se a string está no formato mm/dd/yyyy
    var dateRegex = /^(\d{1,2})\/(\d{1,2})\/(\d{4})$/;
    var match = dateString.match(dateRegex);

    if (!match) {
        throw new Error("A data fornecida não está no formato mm/dd/yyyy.");
    }

    var month = Number(match[1]);
    var day = Number(match[2]);
    var year = Number(match[3]);

    // Formata a data para dd/mm/yyyy
    var formattedDate = Utilities.formatString('%02d/%02d/%04d', day, month, year);

    return formattedDate;
}

function getLatestDateAboveLastRow() {
    var spreadsheetId = 'xxxxxxxx';
    var SheetName = 'REPORTS';
    var dateColumnIndex = 5; // Column E
    var valuesAboveLastRow = 500;

    var Sheet = SpreadsheetApp.openById(spreadsheetId).getSheetByName(SheetName);
    var lastRow = Sheet.getLastRow();

    var range = Sheet.getRange(lastRow - valuesAboveLastRow + 1, dateColumnIndex,
    valuesAboveLastRow, 1);

    var values = range.getValues();

    var latestDate = new Date(0);

```

```

for (var i = 0; i < valuesAboveLastRow; i++) {
    var cellValue = values[i][0];
    if (cellValue instanceof Date && cellValue > latestDate) {
        latestDate = cellValue;
    }
}

var nextDay = new Date(latestDate);
nextDay.setDate(nextDay.getDate() + 1);

var todayMinus2 = new Date();
todayMinus2.setDate(todayMinus2.getDate() - 2);

if (nextDay > todayMinus2) {
    return Utilities.formatDate(todayMinus2, "GMT", "yyyy/M/d");
} else {
    nextDay.setDate(nextDay.getDate() + 1);
    return Utilities.formatDate(nextDay, "GMT", "yyyy/M/d");
}
}

```

## Processo ETL - Programa 2

```

function updateSheetWithCSV() {
    var spreadsheetId = 'xxxxxxx';

    var SheetName = 'API-REPORTS'; // Caso o Sheetname troque de nome é necessário que
    se troque para o novo

    var today = new Date();
    var yesterday = new Date(today);
    yesterday.setDate(yesterday.getDate() - 2);

    var formattedToday = Utilities.formatDate(today, "GMT", "yyyy/M/d");
    var formattedYesterday = Utilities.formatDate(yesterday, "GMT", "yyyy/M/d");

    duplicarFolhaEInserirData('Template', yesterday, spreadsheetId);
}

```

```

Logger.log('Data inicio: ' + getLatestDateAboveLastRow());
Logger.log('Data fim: ' + formattedYesterday);

var apiKey = PropertiesService.getScriptProperties().getProperty('apiKey');

var apiURL = 'https://affiliates.xxxxxx.com/api/affreporting.asp?key=' + apiKey +
'&reportname=Member%20Report%20-
%20Detailed&reportformat=csv&reportmerchantid=0&reportstartdate='
getLatestDateAboveLastRow() + '&reportenddate=' + formattedYesterday;

var response = UrlFetchApp.fetch(apiURL);
var csvData = response.getContentText();

var Sheet = SpreadsheetApp.openById(spreadSheetId).getSheetByName(SheetName);

// Processa o CSV usando a biblioteca CSVParser
var csv = Utilities.newBlob(csvData).getDataAsString();
var parsedData = Utilities.parseCsv(csv);

// Remove as 5 primeiras linhas do array parsedData
parsedData.shift(); // Repete o método para remover as outras 4 linhas
parsedData.shift();
parsedData.shift();
parsedData.shift();
parsedData.shift();
parsedData.shift();

// Remove a última linha do array parsedData
parsedData.pop();

var numRows = parsedData.length;
var numColumns = parsedData[0].length;

var lastRow = Sheet.getRange("B" + Sheet.getLastRow()).getRow();

```

```

var range = Sheet.getRange(lastRow + 1, 2, numRows, numColumns);
range.setValues(parsedData);

var valuesToFill = [];
var valuesToFillE = [];

for (var i = 0; i < numRows; i++) {
    var dateValue = contarDiasDesde1900(parsedData[i][3]);
    valuesToFill.push([dateValue + "|" + parsedData[i][9]]);

    var formattedDate = convertToDDMMYYYY(parsedData[i][3]);
    valuesToFillE.push([formattedDate]);
}

var rangeColumnA = Sheet.getRange(lastRow + 1, 1, numRows, 1);
var rangeColumnE = Sheet.getRange(lastRow + 1, 5, numRows, 1); // Coluna E é a
coluna 5

rangeColumnA.setValues(valuesToFill);
rangeColumnE.setValues(valuesToFillE);

// Registra o número de linhas adicionadas no Logger
Logger.log('Ultima linha: ' + lastRow+ " + " + numRows);
}

```

### Algoritmo de “Check up”

```

function checkAndUpdateData() {
    var SheetName = "REPORTS"; // Substituir pelo nome da folha
    var Sheet = SpreadsheetApp.getActiveSpreadsheet().getSheetByName(SheetName);

    var lastRow = Sheet.getLastRow();
    var numRowsToCheck = 500;
    if (lastRow < numRowsToCheck) {
        numRowsToCheck = lastRow; // Se o número de linhas a verificar for maior do que
a quantidade total de linhas, ajustamos para a quantidade total de linhas.
    }
}

```

```

}

var todayMinus2 = new Date();
todayMinus2.setDate(todayMinus2.getDate() - 2);

var dataRange = Sheet.getRange(lastRow - numRowsToCheck + 1, 5, numRowsToCheck, 1);
var dataValues = dataRange.getValues();

var isDateInLast20Rows = false;
for (var i = 0; i < numRowsToCheck; i++) {
    var rowDate = dataValues[i][0];
    if (rowDate.toString() === todayMinus2.toString()) {
        isDateInLast20Rows = true;
        break;
    }
}

if (!isDateInLast20Rows) {
    updateSheetWithCSV();
}
}

```

## Script para a análise de dados

```

function updateSheetWithCSVManual() {
    var spreadsheetId = 'xxxxxxx';
    var SheetName = 'Influencer name'; //nome do influencer para identificar

    var data_inicio = "2023/8/2";
    var data_fim = "2023/8/24";
    var id = "14916"; // Id do influencer
    var apiKey = PropertiesService.getScriptProperties().getProperty('apiKey');
    var apiUrl = 'https://xxxxxxxxxxxxxxxx.com/api/affreporting.asp?key=' + apiKey +
    '&reportname=Member%20Report%20-%20Detailed&reportformat=csv&reportmerchantid=0&reportstartdate=' + data_inicio +
    '&reportenddate=' + data_fim;

    var response = UrlFetchApp.fetch(apiUrl);
    var csvData = response.getContentText();
}

```

```

var Sheet = SpreadsheetApp.openById(spreadSheetId).getSheetByName(SheetName);

duplicarFolha('Template', spreadSheetId);

// Processa o CSV usando a biblioteca CSVParser
var csv = Utilities.newBlob(csvData).getDataAsString();
var parsedData = Utilities.parseCsv(csv);

var dataToWrite = [];
var datesToFillE = [];

for (var i = 5; i < parsedData.length - 1; i++) { // Começa da sexta linha (índice
5) e evita a última linha
    var row = parsedData[i];

    if (row[9] === id) {
        var formattedDate = convertToDDMMYYYY(row[3]);
        dataToWrite.push(row);
        datesToFillE.push([formattedDate]);
    }
}

if (dataToWrite.length > 0) {
    var lastRow = Sheet.getLastRow() + 1;
    var rangeColumnsToWrite = Sheet.getRange(lastRow, 1, dataToWrite.length,
dataToWrite[0].length);
    rangeColumnsToWrite.setValues(dataToWrite);
    var rangeColumnE = Sheet.getRange(lastRow, 4, datesToFillE.length, 1);
    rangeColumnE.setValues(datesToFillE);
}

Logger.log('Data inicio: ' + data_inicio);
Logger.log('Data fim: ' + data_fim);
Logger.log('Linhas adicionadas: ' + dataToWrite.length);
}

```