

Testovací úloha pro kandidáty na pozici Datového Engineera v Publicis Groupe

# Vyhodnocení průběhu kampaně 1AHA810

Tato testovací úloha se zaměřuje na načtení, zpracování dat a jejich vizualizaci v Pythonu. Úloha představuje jednoduchou situaci, kterou můžete zažít v datově-analytickém oddělení.

Pomocí HTTP requestu a SAS tokenu vyextrahujte 2 datasety z Azure Blob Storage (URL je v hyperlincích názvů data setů):

* [**1AHA810\_plan.csv**](https://detask.blob.core.windows.net/de-task/1AHA810_plan.csv)
* [**1AHA810\_postbuy.csv**](https://detask.blob.core.windows.net/de-task/1AHA810_postbuy.csv)
* **SAS token:** *?sv=2022-11-02&ss=bfqt&srt=co&sp=r&se=2024-09-08T21:42:10Z&st=2024-07-08T13:42:10Z&spr=https&sig=GPNpEjOEf0bdd5l3en38h6zjhBgNFNsaoalVbKj7k%2FU%3D*

## 1AHA810\_plan.csv

Datový set reprezentuje *plán kampaně* tak, jak jej předkládáme klientovi. Tento plán je zároveň i objednávkou. Dataset má denní granularitu. Jednotlivé bannery a inzertní plochy na serverech se nakupují buď na *imprese*, nebo na *clicks*, nebo na *views*. Položky *other* obsahují další, blíže nespecifikované služby. Plochy, které nakupujeme na imprese plánujeme v čase na imprese, tzn. jsou obsaženy v poli Impressions\_DailyGoal a Impressions\_Client. To samé platí pro Clicks i pro Views (zhlédnutí videa).

|  |  |
| --- | --- |
| **Sloupec** | **Popis** |
| **Date** | Den, na který je naplánována marketingová investice |
| **JobNumber** | Identifikátor kampaně (jedné zakázky) |
| **Start** | Počáteční datum kampaně |
| **End** | Poslední oficiální datum kampaně |
| **Impressions\_DailyGoal** | Denní plán na počet zobrazení přes všechny bannery |
| **Impressions\_Client** | Naplánovaná, rozpočítaná cena na den jako cena za odvysílané bannery |
| **Clicks\_DailyGoal** | Denní plán nakoupených prokliků |
| **Clicks\_Client** | Denní rozpočet na prokliky |
| **Views\_DailyGoal** | Plán realizovaných zhlédnutí videa |
| **Views\_Client** | Denní rozpočet na zhlednutí videí |
| **Other\_DailyGoal** | Denní plán na jakékoliv další služby |
| **Other\_Client** | Denní rozpočet na jakékoliv další služby |

### 1AHA810\_postbuy.csv

Postbuy reprezentuje *reálné plnění* plánu po dnech. Postbuy získáváme z platformy Keboola, která načítá data ze všech marketingových platforem (AdWords, Facebook, Sklik, Adform atd.). Díky nastaveným datových pipelines pak získáváme na každý den počet impresí, prokliků, views atd. ze všech marketingových platforem, které jsme využili pro nákup reklamy.

|  |  |
| --- | --- |
| **Sloupec** | **Popis** |
| Id | Identifikátor řádku |
| Account | Klientský účet |
| BusinessUnit | Divize firmy, příp. identifikátor účtu klienta |
| Date | Den, kdy se reklama zobrazila |
| ReportGranularity | Označuje, že pracujeme s denními daty. Můžeme využívat i více granulární, tedy hodinová, desetiminutová, nebo minutová data, ale odpovídá tomu i zpracovávaný objem dat. |
| SalesChannel | Ve velké míře online, občas je součástí dat i nějaká offline reklama. |
| PlatformId | Identifikátor marketingové platformy (adform, adwords, sklik apod.) |
| PlatformCode | Název marketingové platformy (adform, adwords, sklik apod.) |
| ChannelId | Identifikátor kanálu |
| ChannelCode | Název kanálu |
| TagCodes | Nevyužíváno … |
| Site | Web, kam směřujeme traffic |
| AdSource | Zdroj návštěvnosti (bráno z UTM) |
| AdMedium | Médium (bráno z UTM) |
| AdCampaign | UTM\_Campaign (parametr předávaný v URL odkazu) |
| JobNumber | Identifikátor kampaně (zakázky) |
| LineItem | Číslo řádku v mediaplánu |
| Source | Zdroj návštěvnosti |
| Format | Identifikátor plochy (např. Idnes/ square160x160 Homepage) |
| Target | Cílení |
| Other1 | Další custom parametry, které si předáváme z odkazů na webu |
| Other2 | Další custom parametry, které si předáváme z odkazů na webu |
| MarketingInvestment | Investice do inzerce |
| Views | Počet shlédnutí videa |
| Impressions | Počet zobrazení bannerů |
| Clicks | Počet prokliků |
| Visits | Počet návštěv |
| Conversions | Počet konverzí |
| Deliveries | Počet dokončených objednávek (v případě ecommerce … ) |
| NetRevenue | Obrat vytvořený dokončenými objednávkami (v případě ecommerce … ) |
| GrossProfit | Profit (v případě ecommerce … ) |
| CurrencyCode | Kód měny |
| AppliedAttributionModel | Zvolený atribuční model |
| EtlLoader | Technický sloupec bez využití |
| CreatedAt | Datum vytvoření záznamu / řádky |
| ModifiedAt | Datum poslední modifikace řádku |

# Cíl úlohy

* Pro zpracování úkolu využijte programovací jazyk Python a můžete využít libovolný IDE/code editor.
* Vyextrahujte data z Azure blobu.
* Načtěte data z plánu a postbuy.
* Vyčistěte postbuy data (odstraňte irrelevantní data a záznamy s null hodnotami)
* Spočítejte, kolik procent z celkového rozpočtu je vyčerpáno.
* Zobrazte grafem porovnání plánované *(součet hodnot Impressions\_Client, Clicks\_Client a Views\_Client)* a realizované investice *(MarketingInvestment)* po dnech.
* Zobrazte grafem tyto 2 hodnoty kumulovaně. Tato vizualizace ukazuje, jestli naše celková investice k dnešnímu dni je v souladu s celkovou plánovanou investicí k dnešnímu dni.
* Zobrazte poměr mezi kumulovaným plánem a kumulovaným postbuy k datu poslední aktualizace (tzn. tam, kde máme poslední data z postbuye).
* Zkuste navrhnout, co dalšího zajímavého by se v datasetu dalo vizualizovat.
* Výstup prosím pošlete nejpozději 3 dny před osobním setkáním (vyexportovaný notebook s výstupy a dokumentací v .html formátu).

# Teoretické řešení automatizace reportu

Tato úloha je příkladem typického zadání, které DE tým dostává od kolegů i klientů. Ti potřebují mít připravená data v různých reportech, která jsou generována na denní bázi (a tedy aktuální k poslednímu dni).

Úkolem je navrhnout řešení, jak automatizovat stahování reportu, který je generován reklamní platformou (např. Adform) na denní bázi a zasílán jako .csv příloha emailem.

Cílem je bez manuálního zásahu report z emailu stáhnout, zpracovat a nakonec zapsat do cílové databáze. Předpokládejte, že pro úkol nemáte k dispozici fyzický server, a proto využijete služeb cloudových platforem (Keboola, Microsoft Azure, AWS, atp.).

Odpověď není nutné psát, postačí promyslet, jak byste k řešení přistupovali:

* Jak přistoupíte ke stahování reportu, který je generován jako příloha emailu
* Kde zpracujete data
* Jak a odkud zapíšete data do cílové databáze
* Jak celý proces automatizovat

Zkuste se zamyslet nad slabými a silnými stránkami různých řešeních tohoto problému, včetně Vámi navrženého.