

Hackathon 2021

03/2021

Reconnaissance d'images

L'objectif de ce sujet est d'être en mesure d'identifier une catégorie de produit à partir de photos prises par l'utilisateur

Accueil > Rayons > Boissons > Eaux > **Eaux gazeuses** > Eau gazeuse minérale naturelle BADOIT



Potentiel client / catégorie

L'objectif de ce sujet est d'être en mesure d'identifier les prochains produits que le client va acheter, ainsi que son niveau de chiffre d'affaires sur les catégories de produits, à partir des achats qu'il a réalisés sur l'année précédente



SKU KEY	SKU	Y/N ?
1000001	BRSA LIMONADE PET 1.5L	
1000002	BRSA LIMONADE AROMATISEE	
1000003	BRSA LIMONADE 50CL	

100 - SOFTS DRINKS : ?€
101 - BIERES ET CIDRES : ?€
102 - EAUX : ?€
103 - VINS : ?€

Environnement

Support technique

- Individuellement pour chaque équipe :
 - VM avec notebook Jupyter associé
 - Dataset BigQuery (pour écriture de tables SQL)
 - Bucket dans Cloud Storage (pour intégrer / écrire des fichiers non structurés)
- Pour l'ensemble des équipes :
 - Vues aux tables communs (tickets / magasins / produits...)
 - Bucket contenant l'ensemble des images



Données disponibles

Datasets

- Transactions sur l'année 2019
 - Contenu des paniers pour un échantillon d'environ 100k clients
- Référentiel produit
- Référentiel magasin

- Images produits associés à un code EAN (code barres)
 - Les images sont fournies pour un produit donné, avec différents angles de vue
- Metadata associés aux images
 - Informations relative au produit, à son appartenance aux hiérarchies produits Drive et magasin



Images produits

Des images sur les produits du quotidien (180GB) ainsi que leur metadata respectives vous seront fournies

- Table BigQuery associée : `hackathon-2021-305208:common_referential.metadata_images`
- Répertoire des images sur Google Cloud Storage : `"gs://hackathon-2021-images/images"`

Format du nom du fichier : *(image_{barcode}_{photo_angle}_{random_seq}).jpeg (or .png)*

image_5449000052179_1_12b8bbc0.jpeg



Angles d'image :

Avant	: 1, 11, 31,
Gauche	: 2, 14, 15, 32
Droite	: 3, 23, 24, 25, 35
Dessus	: 4, 6, 17, 18, 19, 26, 27, 28, 33, 36
Arrière	: 5, 20, 21, 22, 34
¾	: 7, 12, 13, 16
Autre	: 8, 9, 10, 11, 29, 30



Metadata

Column	Description	Example
Barcode	Actual barcode that we see on the product (Unique)	3228021950051 (13 digits)
item_dec	Item description	Fromage Petit Brie PRESIDENT
brand_name	Brand name of the product	PRESIDENT
nature	Nature that was defined in the referential to identify it's nature	fromage
packaging	Description of the packaging	la boite de 500 g
arbonodes.is_primary_link	Sometimes the products is being listed into multiple categories Rayons on the website. This flag to say which one is primary	true or false
arbonodes.nodeid[1-4]	A unique identifier given to each rayon on the website	28127
arbonodes.level1	Description of the rayon displayed	Check next slide
structHyp	Different Nomenclature defined for hyper markets	



Arbonodes

L'arbonode est une manière de classer les produits et de les présenter sur le site carrefour.fr
Il est présent dans les métadatas dans BigQuery

Level 1

Level 2

Level 3

Level 4

Accueil > Rayons > Frais > Fromages à déguster > Camemberts, Coulommiers et Bries > Bries > Fromage Petit Brie PRESIDENT

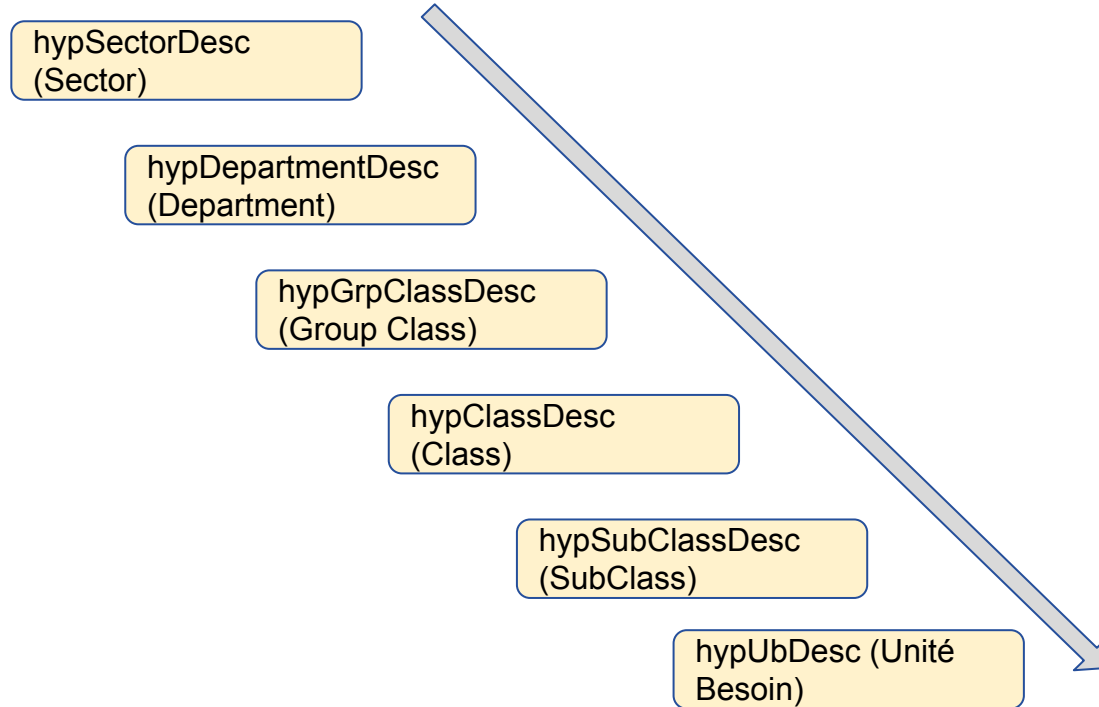


Fromag
PRESIDE
la boîte de 5

4,85€
9.70 € / Kilogram

Nomenclature produit

Tous les produits vendus en magasin ne sont pas présents dans l'assortiment e-commerce.
La nomenclature "classique" de Carrefour permet de retrouver l'ensemble des produits
Ces données sont présentes dans les metadata dans BigQuery & dans la table bv_barcode



Transactions

Column	Description	Example
id_household	Customer ID	1026684021
format_cmd	OFFLINE (stores) / DRIVE / CLCV (home delivery)	CLCV
datekey	Day of the transaction	2019-03-06
storeEan	Store ID	3020180261296
trxid	Transaction ID	2019-03-06-1073-79418825
ean	Product barcode	3041091099914
itemkey	Product internal ID	5858996
amount	Amount paid by the customer (for the given product)	41.8
quantity	Quantity bought by the customer	6.0
promoFlag	1 if there was promotion for this product	1
promoAmount	Amount of promotion (for the given product)	20.9



Stores

Column	Example
stoEan	3021520203228
stoDisplayDesc	Market Chevigny-Saint-Sauveur
address1	ROUTE DE DIJON
address2	
postcode	21800
town	CHEVIGNY ST SAUVEUR
storeSalSize	2450
drivePickingFlag	1
gasStationFlag	1
chaintypeopkey	SPM
chaintypeopdesc	SUPERMARCHES
chaintypekey	SUP
chaintypedesc	SUPER



Products

Column	Example
barcode	5000112611861
itemKey	6545729
itemDesc	PET 1.75L COCA
mainBarcode	5000112611861
brandTypeKey	N
brandTypeDesc	National brand
brandKey	COCA COLA
brandDesc	COCA COLA
dispoKey	P
dispoDesc	PERMANENT
prdUnitDesc	UVC NORMAL
prdModKey	X
prdModDesc	N/A

Column	Example
hypSectorKey	1
hypSectorDesc	PGC
hypDepartmentKey	10
hypDepartmentDesc	BOISSONS
hypGrpClassKey	100
hypGrpClassDesc	SOFTS DRINKS
hypClassKey	1000
hypClassDesc	GAZEIFIES
hypSubClassKey	10004
hypSubClassDesc	COLAS
hypUbKey	1000402
hypUbDesc	BRSA COLA STD PET 2L



Questions?

