

Database **E-Commerce Brazilian Project**

Equipe: Marceline, Silvain, Temirlan, Clara

Sommaire

01

Exploration

Données Kaggle Brazilian
E-commerce

03

Préparation

Et nettoyage des données

02

Importation

Des données sur le serveur
MySQL via Jupyter Notebook

04

Visualisation

Création des tableaux de bord
avec Power BI

Project Management

The screenshot displays a web browser window showing a GitHub repository page for 'dataIA-2021 / Clara_Marceline_Silvain_Tim'. The page is set to the 'Projects' tab, which features a Kanban board titled 'Brazilian E-Commerce Equipe 2'. The board is organized into three columns: 'To do' (12 items), 'In progress' (1 item), and 'Done' (1 item). The 'To do' column contains tasks such as 'Nettoyage des données', 'script python', 'connexion à la base de données', 'Création du dictionnaire de données', and 'observation des clefs primaires et secondaires'. The 'In progress' column contains a welcome message and a checklist for creating a new project. The 'Done' column contains a task 'Prise en connaissance de la structure des Données'. The interface includes a search bar, navigation links for Code, Issues, Pull requests, Actions, Projects, Wiki, Security, Insights, and Settings, and a taskbar at the bottom with various application icons and system information.

dataIA-2021 / Clara_Marceline_Silvain_Tim Public

<> Code Issues 1 Pull requests Actions Projects 1 Wiki Security Insights Settings

Brazilian E-Commerce Equipe 2
Updated 9 days ago

Filter cards + Add cards Fullscreen Menu

12 To do

- Nettoyage des données
Added by Marceline37
- script python
Added by Marceline37
- connexion à la base de données
Added by Marceline37
- Création du dictionnaire de données
Added by Marceline37
- observation des clefs primaires et secondaires

1 In progress

- Welcome to GitHub Projects
We're so excited that you've decided to create a new project! Now that you're here, let's make sure you know how to get the most out of GitHub Projects.
- ☒ Create a new project
- ☒ Give your project a name
- ☐ Press the ? key to see available keyboard shortcuts
- ☐ Add a new column
- ☐ Drag and drop this card to the new column
- ☐ Search for and add issues or PRs to your project
- ☐ Manage automation on columns

1 Done

- Prise en connaissance de la structure des Données
Added by Silvain.

+ Add column

PBI-translations-re...zip PBI-translations-re...zip PBI-translations-re...zip Confusion_Matrix...pdf e_commerce.pbix

Tout afficher

Taper ici pour rechercher 5°C Nuageux 1409 19/01/2022

01

Exploration

Données Kaggle Brazilian E-commerce

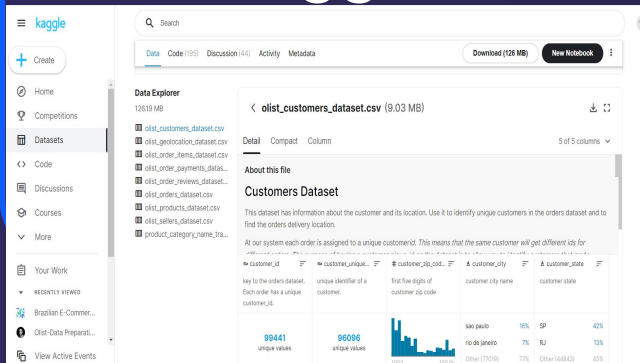
Exploration

On a exploré les données avec python et on a observé les metadata sur Kaggle.

What are we working on?



Kaggle



The screenshot shows the Kaggle Data Explorer interface for the 'olist_customers_dataset.csv' file (9.03 MB). The interface includes a sidebar with navigation options like Home, Competitions, Datasets, Code, Discussions, Courses, and More. The main area displays the dataset's details, including a description: 'This dataset has information about the customer and its location. Use it to identify unique customers in the orders dataset and to find the orders delivery location.' Below the description, there is a table with columns: customer_id, customer_unique_id, customer_zip_code, customer_city, and customer_state. The table shows the first few rows of data, including customer IDs, unique identifiers, zip codes, city names, and state abbreviations. A histogram is also visible, showing the distribution of unique values for the customer_id column.

Jupyter

```
Entrée [17]: datasheet = "/Data/olist_products_dataset.csv"
pro = pd.read_csv(datasheet);
pro.shape
pro.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 32951 entries, 0 to 32950
Data columns (total 9 columns):
 #   Column              Non-Null Count  Dtype
---  --
 0   product_id          32951 non-null  object
 1   product_category_name 32341 non-null  object
 2   product_name_length  32341 non-null  float64
 3   product_description_length 32341 non-null float64
 4   product_photos_qty   32341 non-null  float64
 5   product_weight_g      32949 non-null  float64
 6   product_length_cm     32949 non-null  float64
 7   product_height_cm     32949 non-null  float64
 8   product_width_cm      32949 non-null  float64
dtypes: float64(7), object(2)
memory usage: 2.3+ MB
```

```
Entrée [18]: pro.head()
Out[18]:
```

	product_id	product_category_name	product_name_length	product_description_length	product_photos_qty	product_weight_g	product
0	1e16e9f04b0c4f541ec08957ead17a5	perfumaria	40.0	287.0	1.0	225.0	
1	3aa071193dc10b67c0e050ead41aa2c5	artes	44.0	276.0	1.0	1000.0	
2	9f0b7f6c8d10374ed1c05e261975717f	esporte_lazer	46.0	260.0	1.0	154.0	
3	c6ff7bce19066a032b703ac239eb23d3	bebes	27.0	261.0	1.0	371.0	
4	8dc1a76027444849c219c19560b71	utilidades_domesticas	37.0	402.0	4.0	625.0	

02

Importation

Des données et création des tables sur My
SQL

```
import pandas as pd
import pymysql
from sqlalchemy import create_engine
import yaml
with open('votre répertoire/config.yml', 'r') as f:
    conf = yaml.safe_load(f)
my = conf['MYSQL']

# Connection MYSQL
str = f"mysql+pymysql://{my['user']}:{my['password']}@{my['host']}:{my['port']}/{my['database']}"
engine = create_engine(str, echo=False)

connection = engine.connect()

connection.execute('DROP TABLE IF EXISTS products;')

connection.execute('CREATE TABLE IF NOT EXISTS products(product_id varchar(60) PRIMARY KEY NOT NULL,
product_category_name text(70), product_name_lenght float(10), product_description_lenght float(20),
product_photos_qty float(5),product_weight_g float(20),product_length_cm float(10),product_height_cm
float(10),product_width_cm float(10)) DEFAULT CHARSET=utf8;')

# data set
products=pd.read_csv('https://raw.githubusercontent.com/VictorGuedes/Brazilian-E-Commerce-Public-Dataset-example
s/master/dataset/olist_products_dataset.csv')
products.to_sql('products', con=engine, if_exists='append', index=False)
```


03

Nettoyage

Des données et création dictionnaire des
tables

Nettoyage

On a effectué un nettoyage, on a utilisé OpenRefine
Le nettoyage que l'on a effectué concerné :

- Les fautes de frappe
- La clusterisation des commentaires

OpenRefine 127.0.0.1_bdd_brand_1642521727121

[Personal](#)

[Open](#)
[Export](#)
[Help](#)

Facet/Filter

[Undo](#)
[Redo](#)

99224 rows

Extensions

Wizards

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?

[Watch these screencasts](#)

Show as

rows

records

Show 5
10
25
50
100
500
1000
rows

[x](#)
[list](#)
[previous](#)

1
of 9623 pages

[next](#)
[last](#)

	All	review_id	order_id	review_score	review_comment_title	review_comment_message	review_creation_date	review_answer_creation_date
1	1	0001230c19e2c032059592c2a645	10a66f7b717f16384240870104	5			2018-03-20 00:00:00	2018-03-20 18:30:04
2	1	0001058100a90871784131a5c265	048554201caedc3c783a0e5593e	5			2018-06-22 00:00:00	2018-06-25 13:31:29
3	3	0001730c19e2c032059592c2a645	e38a5e730a303030303030303030	5	Enrega rapidat	A entrega foi super rápida e o vendedor é incrível (graças a isso mandei)	2018-04-25 14:30:36	
4	1	0001020741445467189b3030303030	04b7c559939a3030303030303030	5			2017-05-30 00:00:00	2017-06-01 23:28:55
5	1	0001400000000000000000000000	050607b7002045327030303030	5			2017-08-13 00:00:00	2017-08-16 19:58:53
6	1	0001030c19e2c032059592c2a645	8706a0e36142c19e2c032059592c2a645	5			2018-02-05 00:00:00	2018-02-06 20:51:37
7	1	000101787f74550f80303030303030	99910e0e595740e330003030303030	1			2018-05-24 00:00:00	2018-05-25 12:42:58
8	1	000106a5002102c032059592c2a645	8706a0e36142c19e2c032059592c2a645	5	SUPER RECOMENDO	GOSTO DA CAPINHA DE ESQUILAR, MAS COMO EU ESPERAVA	2018-07-01 00:00:00	2018-07-01 12:22:24
9	1	0001054773388c040e083030303030	a091a9e000000000000000000000	5	ótimo produto, prazo cumprido, seller atencioso multi-lingua		2017-09-01 00:00:00	2017-09-02 13:04:08
10	1	000105a500000000000000000000	930a264240e303030303030303030	5	Creador lo entregou muito antes do esperado. Como agilizou.		2017-11-25 00:00:00	2017-11-27 11:37:22

04

Visualisation

Création des tableaux de bord avec PowerBI

MERCI !



MERCI !

Marceline, Silvain, Tim et Clara