

Analyse des Pals

Non, ce ne sont pas des Pokémons, mais des Pals !

Eidolon Parlu



Palworld (ou Eidolon Parlu), qui a rapidement gagné en popularité dans le monde entier début janvier 2024, est **un jeu de crafting et de survie multijoueur en monde ouvert**, du développeur japonais Pocket Pair.

Avec son gameplay innovant, ce jeu a attiré un grand nombre de joueurs et de fans, attirant l'attention sur les Pals, ces créatures fantastiques peuplant le monde virtuel du jeu. **Vous apprivoisez et collectionnez ces mystérieuses créatures afin de construire, se battre, cultiver et travailler dans des usines afin de développer votre campement !**

N'étant pas insensible vous-même à ce jeu, vous avez consacré des heures interminables à ce jeu ! Votre nature d'analyste de données aussi important



que votre nature d'amateur de jeux vidéo, vous souhaitez réaliser une étude de vos chers Pals. **L'étude des caractéristiques et du comportement de ces créatures peut se montrer utile pour développer une expérience de jeu plus attrayante.**



Analysez les tous

Vous récupérez un dataset obtenu après une extraction des données du jeu.

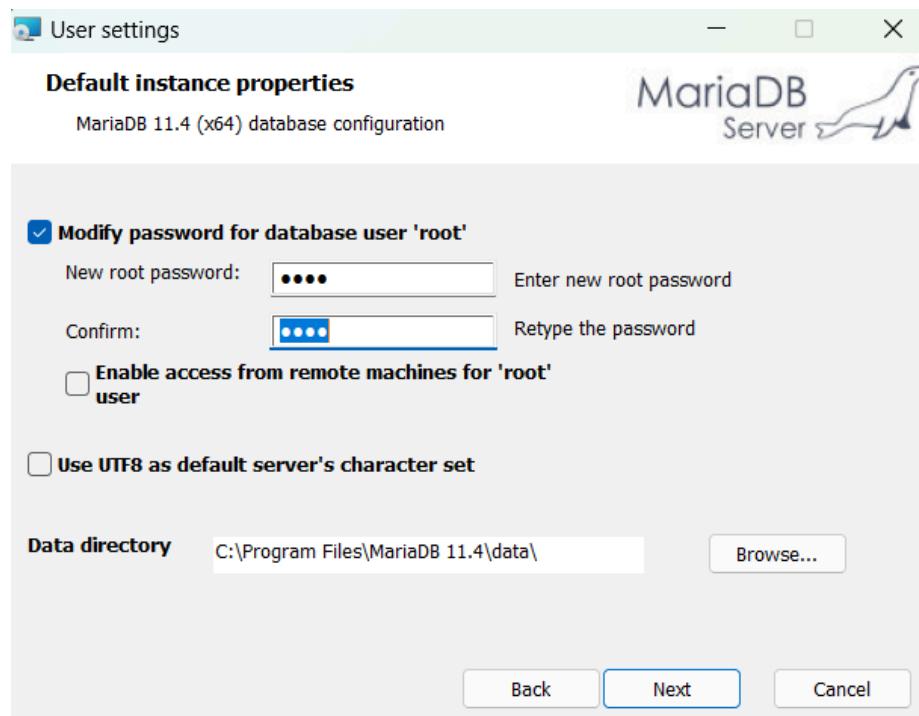
Vous avez donc accès à **des informations précises et en temps réel des Pals** telles que les attributs, les comportements, les compétences, la nature, la force, ... etc. Le jeu de données se trouvant [ici](#) et son descriptif [ici](#).

1. Vous commencez par **récupérer** le dataset d'intérêt et à le parcourir afin de vous familiariser avec. **Notez tout ce qui semble pertinent à vos yeux** (le nombre de tables, de colonnes, les données qualitatives, les



données quantitatives, les données intéressantes, données mal formatées, données erronées, ..., etc).

2. **Vous installez MariaDB (ou MySQL)** sur votre machine afin de pouvoir créer et manipuler votre base de données. Depuis l'interface d'installation, **vous paramétrez l'utilisateur root** et **vous définissez un mot de passe sécurisé** qu'il ne faudra surtout pas oublier. Le reste des paramètres d'installation peuvent être laissés par défaut. L'installation de MariaDB, inclut HeidiSQL, un outil d'administration de base de données possédant un éditeur SQL et un constructeur de requête.



3. Une fois l'installation terminée, vous ouvrez un Command Prompt mariaDB sur votre machine et vous connectez votre utilisateur root. Vous créez une base de données appelée **palworld_database**. Afin de vous assurer que votre base de données a bien été créée, affichez les



bases de données existantes. Précisez depuis votre terminal que vous allez utiliser cette base de données.

```
Command Prompt (MariaDB 1) + ^

Setting environment for MariaDB 11.4 (x64)

C:\Windows\System32>mysql -u root -p
Enter password: *****
Welcome to the MariaDB monitor. Commands end with ; or \g.
Your MariaDB connection id is 3
Server version: 11.4.2-MariaDB mariadb.org binary distribution

Copyright (c) 2000, 2018, Oracle, MariaDB Corporation Ab and others.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

MariaDB [(none)]> CREATE DATABASE palworld_database;
Query OK, 1 row affected (0.003 sec)

MariaDB [(none)]>
MariaDB [(none)]> USE palworld_database;
Database changed
MariaDB [palworld_database]> SHOW DATABASES;
+-----+
| Database |
+-----+
| information_schema |
| mysql |
| palworld_database |
| performance_schema |
| sys |
+-----+
5 rows in set (0.032 sec)
```

4. **Vous créez les différentes tables** dans lesquelles vous importez les 6 tables de données récupérées précédemment :
- La table combat-attribute,
 - Une table job-skill,
 - Une table hidden-attribute,
 - La table refresh-area,
 - Une table ordinary-boss-attribute,
 - Une table tower-boss-attribute.



5. A l'aide de vos notes de la première étape, **vous nettoyez les différentes tables de votre base de données** et vous les rendez exploitables pour une analyse. Traitez les valeurs manquantes, normalisez les formats et corrigez les erreurs. **Vous modifiez vos tables afin qu'il n'y ait pas de redondances.**

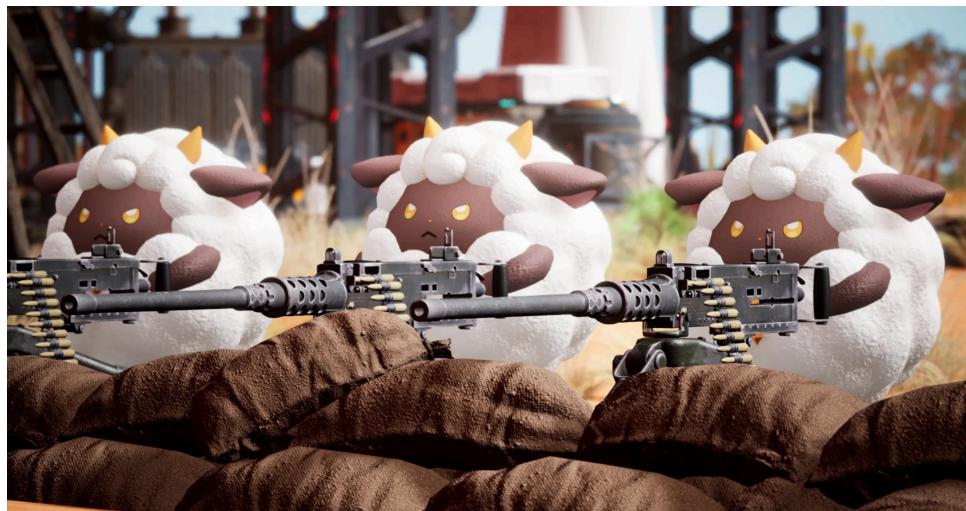


6. **Vous réalisez une analyse exploratoire de votre base de données afin d'identifier les informations contenues dans l'ensemble des données de Palworld**, permettant une optimisation plus efficace de votre stratégie de jeu. Vous utiliserez **SQL** pour le traitement de vos données et **Python** pour la création de visuels pertinents.

- Quelle est la distribution de **la taille des Pals** ?
- Quelle est la distribution de **la catégorie des Pals** ?
- Quelle est la distribution **des points de vie des Pals** ?



- d. Quelle est la distribution de **la rareté des Pals** ?
- e. Quelle est la distribution de **la consommation alimentaire des Pals** ?
- f. Quels sont les **Pals pouvant offrir des produits utiles** à votre campement grâce au ranch (laine, oeuf, lait, ...) ?
- g. Quelle est **la distribution de la puissance de combat parmi les Pals** ? Fournissez une liste des 10 Pals les plus puissants (selon l'attaque, la défense,...).
- h. Quelles sont les **corrélations entre les différents attributs de combat** ?
- i. Comment **la rareté d'un Pal** affecte-t-elle les valeurs de ses attributs de base ?
- j. Quelle est **la rareté moyenne des Pals** ayant la puissance d'attaque la plus élevée ?
- k. **La taille des Pals** affecte-t-elle leur performance au combat ?





- I. **Les Pals les plus rapides** sont-ils généralement **plus efficaces au combat** ?
- m. **Comment former l'équipe la plus équilibrée** en combinant des Pals aux attributs complémentaires ? Proposez une équipe équilibrée de 5 Pals.
- n. Quelles sont **les compétences de travail** les **plus répandues** chez les Pals ?
- o. Quelles sont **les compétences de travail** les **moins répandues** chez les Pals ?
- p. Combien de Pals conviennent au **travail de nuit** ?
- q. Quelles sont les **caractéristiques communes des Pals** qui conviennent au **travail de nuit** ?
- r. Quelle est la **rareté moyenne des Pals** possédant **le plus de compétences** ?
- s. Quels sont **les Pals qui ont la vitesse de travail la plus élevée** ?
- t. Quels sont **les Pals dont la probabilité de capture est la plus élevée** ? Comment cette information peut-elle être utilisée pour **élaborer une stratégie de capture** dans le jeu ?





u. Quel **Ordinary Boss ou Tower Boss** a le **score** d'attributs de **combat** combinés **le plus élevé** ?

v. Quelle est la **répartition des niveaux d'apparition** des Pals ?

w. Quelle est la **répartition des zones d'apparition** ?

x. De nombreuses autres observations peuvent être réalisées à l'aide des données et ne se limitent pas à la liste ci-dessus.

Vous continuez donc l'analyse exploratoire avec des recherches et une étude complémentaire.





Palstream

```
MyApp.py
```

```
import streamlit as st
import pandas as pd

st.write("""
# My first app
Hello *world!*
""")

df = pd.read_csv("my_data.csv")
st.line_chart(df)
```

My App • Streamlit

My first app

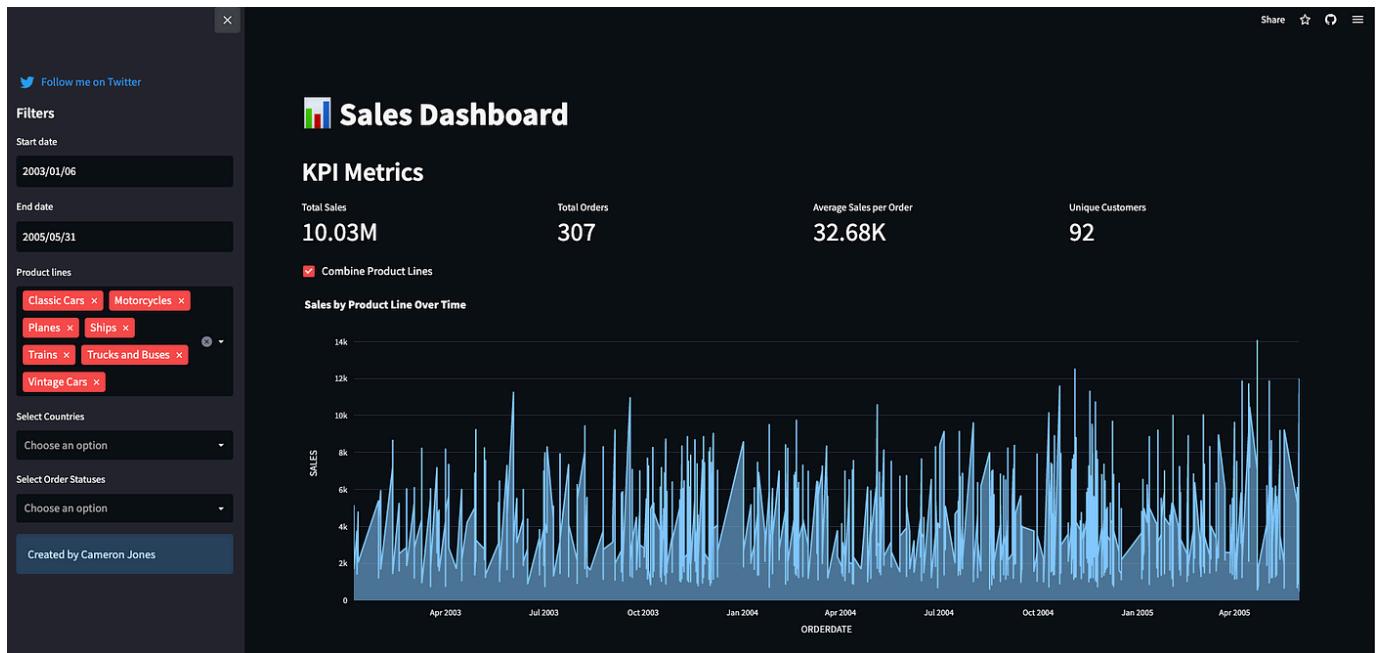
Hello world!

A line chart showing fluctuating data points.

Streamlit est une bibliothèque Python open source qui permet de créer rapidement et facilement des applications web interactives pour la science des données et l'apprentissage automatique. C'est un outil très populaire parmi les data scientists et les ingénieurs en machine learning, car il simplifie le processus de partage de visualisations, de modèles d'apprentissage et d'analyses avec des utilisateurs non techniques.

Vous utilisez Streamlit pour la création d'un outil d'analyse des données du jeu Palworld pour l'optimisation des stratégies de combat et la main d'œuvre de production des campements.

Vous ne gardez que les graphiques et les visuels les plus pertinents à la problématique.



Compétences visées

→ Analyse de données

Rendu

L'évaluation de ce projet se fera sur deux aspects :

1. Une présentation explicative de votre travail sous forme de diapositives.
2. Un repository github public nommé **pals-analysis**, contenant les éléments suivants :
 - a. Un **notebook jupyter** relatant votre analyse de données et la réflexion derrière votre travail,
 - b. Le script **app.py** de votre application Streamlit,
 - c. Un fichier **README.md** présentant le contexte du projet, les données utilisées, votre analyse et une conclusion sur votre travail.



Base de connaissances

- [Palworld - Fandom : Wiki](#)
- [Streamlit - A faster way to build and share data apps : documentation](#)
- [Building a dashboard in Python using Streamlit](#)
- [A Guide To Getting Data Visualization Right](#)
- [Data Visualization Reference Guides](#)