

Compte-rendu du projet d'analyse de données: applications au jeu Stardle

Charles Azais de Vergeron et Octave Feuilland

Mai 2025

1 Introduction

1.1 Présentation du jeu Stardle

Le jeu Stardle est un jeu en ligne (stardle.net) basé sur le jeu Wordle / SUTO. Le but est de deviner chaque jour un vaisseau choisi au hasard parmi une liste d'environ 200 éléments.

Chaque vaisseau est défini par deux types de caractéristiques:

- Qualitatives
 - Fabricant (Manufacturer)
 - Type
 - Role
 - Année d'apparition (Release Date)
 - En Jeu ou Non (Status)
- Quantitatives
 - Nombre de personnes dans l'équipage (Crew)
 - Valeur en \$ (Price)
 - Valeurs en monnaie du jeu (Price In game)
 - Capacité en tant que cargo (Cargo Capacity)
 - Vitesse de croisière (SCM) et maximale (Max)
 - Longueur (Length), Largeur (Beam) et Hauteur (Height)

Voici un exemple de partie du jeu :

Nous avons 3 possibilités de résultats: Vrai, Faux ou Proche (uniquement pour les caractéristiques quantitatives).

1.2 Objectif du projet

L'objectif de ce projet est d'analyser les données du jeu Stardle afin de créer un algorithme capable de trouver un vaisseau pris au hasard dans la base de données en un minimum de tentatives et de déterminer une combinaison d'essais de vaisseaux qui permet de minimiser le nombre de tentatives.

Nous avons donc récupérer les données et recréer le jeu puis essayer de trouver un algorithme optimal pour déterminer le vaisseau le plus efficacement possible.

2 Récupération des données

La récupération des données est faite via le script Python CreateShipDatabase.py. Ce script utilise l'API officielle de Star Citizen Wiki pour récupérer les informations sur les vaisseaux.

2.1 Structure du script

Le script se décompose en plusieurs étapes principales :

1. Collecte des données

Nous utilisons l'API <https://api.star-citizen.wiki/> pour obtenir la liste complète des vaisseaux. Pour chaque vaisseau, le script récupère les données via l'API et extrait toutes les caractéristiques – dans la configuration "All" – et uniquement celles du jeu Stardle – dans la configuration "Stardle" –

Dans la suite du rapport, "Stardle" désigne la configuration du jeu et "All" la configuration complète.

2. Stockage des données

- Création de deux fichiers JSON :
 - `shipList.json` : liste des vaisseaux et leurs liens
 - `shipDB_All.json` : base de données complète des vaisseaux
 - `shipDB_Stardle.json` : base de données avec les catégories du jeu Stardle

3 Nettoyage et préparation des données

Pour ajouter de la complexité aux données d'origine, nous avons mélangés les deux bases de données. Stardle est la base du projet. Nous avons extrait de All les valeurs de `scm`, `max`, `length`, `beam`. Nous avons aussi rajouté une colonne "price in game" qui correspond à la valeur du vaisseau dans le jeu (via le site Erkul.com)

Pour la gestion des valeurs manquantes, nous avons étudié les dependances entre les colonnes. Nous avons remarqué que certaines colonnes étaient très corrélées entre elles. Par exemple, le prix \$ (sans valeurs manquantes) et celle du prix en jeu (avec beaucoup des valeurs manquantes). Nous avons tenté de remplr les valeurs manquantes en utilisant une regression lineaire et une regression quadratique mais ce ne fut pas concluant. Nous avons donc appliqué l'algorithme suivant:

Result: Colonne sans valeurs manquantes

```

groupby by status,type,manufacturer for vaisseaux in names do
| if mean(groupby by status,type,manufacturer) ≠ Nan and price_in_game ==
|   0 then
|   | price_in_game[vaisseaux] = mean(groupby by status,type,manufacturer)
| else
|   | price_in_game[vaisseaux] = mean
| end
end

```

4 Implementation du jeu

Nous avons recréé le jeu Stardle en Python : `stardle.py` Nous avons commencé par afficher l'intégralité des vaisseux du jeu pour éviter au joueur de se tromper lors de la saisie. Nous reproduisons la meme logique que le jeu original puis après nous avons à taper le nom du vaisseux. Ca affiche les résultats.

Voici ce que donne le début du jeu :

INSERT CODE

Il est possible de faire défiler les noms des 200 vaisseaux

Ensuite, tout en bas, il y a une barre de recherche où l'on écrit le nom du vaisseau que nous voulons proposer.

Voici le résultat que ça donne quand nous nous trompons de vaisseau :

```

Vous vous êtes trompés de vaisseau. Voici les résultats de chaque variables :
- pour la variable cargo_capacity ce n est pas la bonne réponse X
- pour la variable mass ce n est pas la bonne réponse X
- pour la variable crew ce n est pas la bonne réponse X
- pour la variable manufacturer ce n est pas la bonne réponse X
- pour la variable type c est la bonne réponse V
- pour la variable status ce n'est pas la bonne réponse X
- pour la variable role c est la bonne réponse V
- pour la variable price ce n est pas la bonne réponse X
- pour la variable release_date ce n est pas la bonne réponse X
- pour la variable price_ingame ce n est pas la bonne réponse X

```

```

-pour la variable scm ce n est pas la bonne réponse    X
-pour la variable max ce n est pas la bonne réponse    X
-pour la variable length ce n est pas la bonne réponse X
-pour la variable beam ce n est pas la bonne réponse   X
-pour la variable height ce n est pas la bonne réponse X

```

Nous pouvons y constater que chaque caractéristique du vaisseau est comparée à celle du vaisseau mystère, pour donnée des indications. Dans le cas d'un succès, on félicite le joueur pour son résultat comme ci-dessous :

```

Bravo, vous avez trouvez le bon vaisseau!
Félicitations! Vous avez trouvé le vaisseau!

```

ON TRAITE L ANNEE DE SORTIE ET LE NBRE D EQUIPAGE COMME
DU QUALITATIF PCQ C EST DES ENTIERS ET PAS LISSE DU TT MET-
TRE UN GRAPHIQUE QUI MONTRE QUE LE NOMBRE D EQUIPAGE
ET LA DATE DE SORTIE VARIE PEU SUR TT LES DONNEES

5 Algorithme de jeu

Dans un premier temps, nous avons étudié les répartitions des vaisseaux entre différents critères qualitatifs par des histogrammes et des graphes. Les histogrammes précisent la répartition des vaisseaux entre leurs rôles, fabricants.. alors que des graphes précise les relations entre eux.

PSEUDO CODE DE STARDLE AUTO ICI

5.1 Complexité

- Temporelle : $O(n \log n)$ où n est le nombre de vaisseaux
- Spatiale : $O(n)$ pour stocker les vaisseaux visités et les distances

5.2 Points clés

- Utilisation d'un système de distances entre vaisseaux
- Mémorisation des vaisseaux déjà testés
- Sélection intelligente du prochain vaisseau basée sur la distance
- Fallback sur une sélection aléatoire si aucune distance n'est disponible