

Rapport de Projet Big data

Introduction

Notre projet vise à développer une solution d'analyse et de prédiction des interactions en combat entre les Pokémon en utilisant des techniques de Big Data.

L'objectif principal est de comprendre comment les types et leurs avantages ou désavantages influencent les probabilités de victoire.

Les données proviennent de deux sources principales :

- Statistiques des Pokémon : Détails sur les attributs/statistiques de base de chaque Pokémon.
- Matchups de types : Relations entre les types de Pokémon (avantages/désavantages).

Ces informations ont été combinées et nettoyées pour créer un pipeline de données complet, prêt à l'analyse et à la modélisation.

Objectifs

- Identifier les types de Pokémon ayant un avantage stratégique en combat.
- Analyser les interactions entre les types pour proposer des recommandations stratégiques.

Sources des données

- [Pokémon Database](#) : Données sur les relations entre types (avantages et désavantages).
- [Pokémon Complete](#) : Informations détaillées sur les statistiques de base de chaque Pokémon.

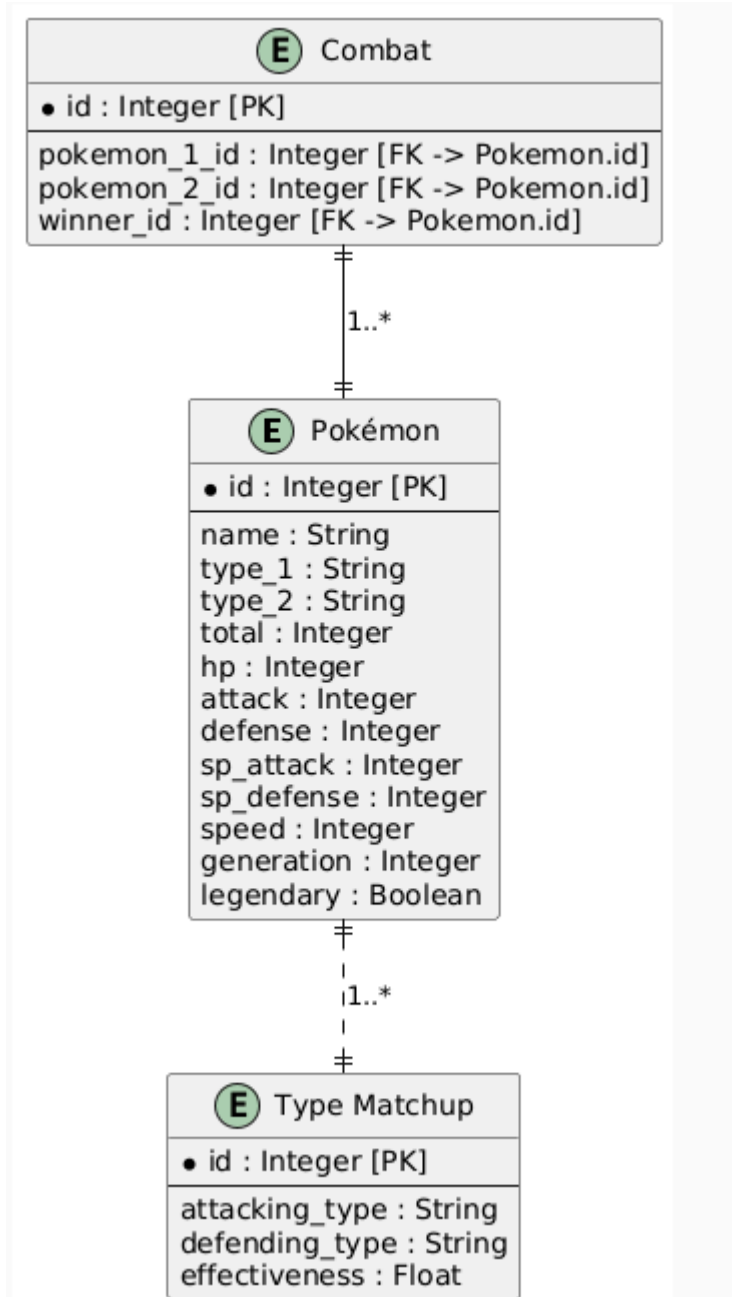
Dataset des Pokémon : Contient les informations sur les statistiques de base de chaque Pokémon.

Dataset des Matchups : Fournit les relations entre les types de Pokémon, indiquant les avantages et les désavantages.

Ces données ont été pré-traitées et intégrées dans notre pipeline.

Modèle Conceptuel de Données

Un modèle en étoile a été utilisé pour structurer les données dans la zone Gold :



Entité **Pokémon** :

- Contient les caractéristiques principales des Pokémon (statistiques, types, génération, etc.).
- Les champs `type_1` et `type_2` permettent de définir les types primaires et secondaires.

Entité **TypeMatchup** :

- Définit les relations entre les types avec le champ **effectiveness** (coefficient indiquant l'avantage ou le désavantage).

Entité **Combat** :

- Enregistre les combats entre deux Pokémon et le vainqueur.
- Les clés étrangères **pokemon_1_id**, **pokemon_2_id**, et **winner_id** pointent vers l'entité **Pokemon**.

Relations :

- Une relation entre **Pokemon** et **TypeMatchup** (un Pokémon peut être associé à plusieurs types).
- Une relation entre **Combat** et **Pokemon** (chaque combat concerne deux Pokémon, avec un vainqueur).

Architecture et Traitements

Les données ont été organisées dans une architecture de type Lakehouse :

Zone Bronze : Contient les données brutes sous forme de fichiers CSV.

Zone Silver : Données nettoyées, avec suppression des doublons et des entrées manquantes.

Zone Gold : Données transformées et préparées pour l'analyse, stockées dans un modèle optimisé.

Nettoyage et Transformation

Les étapes suivantes ont été suivies :

1. Suppression des doublons et des valeurs manquantes.
2. Normalisation des colonnes :
 - Uniformisation des noms.
 - Association des colonnes **Number** (Pokémon Complete) et **pokedex_number** (Pokémon Database).
3. Standardisation des formats de données.

Analyse et Résultats

Statistiques Descriptives

Les statistiques descriptives et analyses des deux fichiers fournissent des informations clés :

Premier fichier (types et faiblesses)

- **Attaque (attack)** : Moyenne de 81, avec un maximum de 185, indiquant des Pokémon très puissants offensivement.
- **Défense (defense)** : Moyenne de 77 avec une valeur maximale de 230.
- **Vitesse (speed)** : Moyenne de 68, avec des variations importantes (5 à 160).

Faiblesses et résistances

- **Types les plus faibles** : against_normal (0.82), against_poison (0.88).
- **Types les plus résistants** : against_ice (1.38), against_rock (1.38).

Deuxième fichier (caractéristiques étendues)

- **Base Stat Total** : Moyenne de 620 pour les Pokémon légendaires.
- **Catch Rate** : Très faible (moyenne de 23.8).
- **Poids moyen** : 206.5 kg, avec un maximum de 999.9 kg.

Prédictions et Insights Stratégiques

- Les types comme rock ou ice sont stratégiquement avantageux.
- Les Pokémon avec des statistiques élevées (attaque et défense) ont une meilleure probabilité de victoire.

Visualisations

Les résultats ont été présentés sous forme de :

1. **Carte thermique** : Montrant les avantages et désavantages entre types.
2. **Histogrammes** : Comparaison des statistiques moyennes par type.
3. **Courbes** : Taux de victoire moyen par type.

Conclusion et Perspectives

Ce projet met en évidence les interactions complexes entre les types Pokémon et leurs impacts sur les résultats des combats. Les travaux futurs pourraient inclure :

- L'intégration de données supplémentaires, comme les capacités spéciales.
- L'analyse des combinaisons multi-types pour des prédictions plus fines.