Comprehensive analysis of Pokemon data

Team Name: Data Wizards
Manasa Akula - S560998
Mounika Jakkula - S560445
Gayathri Nangineni - S559973
PoojaSri Ramineni - S559300

1   Project Idea

The project idea is to undertake an in-depth analysis of Pokemon data to acquire insight into numerous elements of Pokemon traits and gaming dynamics. This analysis may include studying evolutionary patterns, examining basic stats, investigating type variety, assessing ability impact, doing sentiment analysis on descriptions, investigating physical qualities, performing ranking analysis, analyzing generation trends, and so on. The purpose is to discover patterns, trends, and linkages in the Pokemon universe in order to better comprehend its evolution over time and the impact it has on gaming dynamics.

2   Technology Summary

Data Cleaning: Python (Pandas, NumPy)
PySpark: For handling large-scale data analysis. PySpark provides the ability to work with big data efficiently using the Spark framework.
Tableau: To visualize the results in a pictorial form.
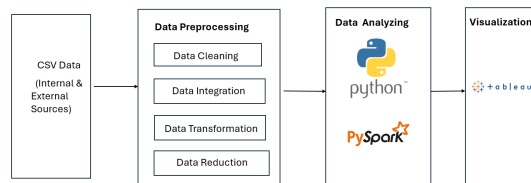
3   Architecture Diagram

Figure 1: Architecture Diagram

4 Architecture Summary

Architecture designed for managing various data sources, including CSV files.
Stages: data cleaning, integration, transformation, and reduction.
Data Cleaning: Identify and rectify inconsistencies, errors, and missing values.
Data Integration: Merge information from multiple sources into a unified dataset.
Data Transformation: Standardize formats, apply calculations, and create new features.
Data Reduction: Minimize dataset size without sacrificing analytical value.
Framework implemented in PySpark and Python for scalability and flexibility with large datasets.
PySpark's distributed computing capabilities ensure efficient processing.
Python allows for versatile manipulation tasks Visualizing processed data through Tableau for intuitive exploration and presentation.
Interconnected workflow ensures smooth data flow from ingestion to visualization.
Benefits: Improved data integrity, reduced processing time, enhanced decision-making capabilities.

5 Project Goals

The project goals encompass a range of analytical tasks aimed at understanding different facets of the Pokemon universe using data analysis techniques.
Goal 1: Conducting a statistical analysis of the base stats (hp, atk, def, spatk, spdef, speed, total) to identify the average, median, and variation of stats among different Pokemon generations or types, providing insights into the overall power balance in the Pokemon world.
Goal 2: Examining the physical attributes of Pokemon, such as height and weight, to understand the diversity and distribution of sizes among different species, and investigate whether there are any correlations between physical attributes and other characteristics like base stats or evolutionary stages.
Goal 3: Analyzing Pokemon rankings to understand the factors influencing a Pokemon's rank, such as its base stats, evolutionary stage, or abilities, and identify any patterns or trends in how Pokemon are ranked across different generations or types.
Goal 4: Analyzing trends in Pokemon characteristics (such as base stats, types, and abilities) across different generations to understand how the Pokemon universe has evolved over time and identify any significant changes or developments in gameplay mechanics.
Goal 5: Segmenting Pokemon based on weight classes (e.g., lightweight, middleweight, heavyweight) to analyze how weight influences factors like base stats, abilities, and evolutionary patterns, and identify any correlations between weight class and competitive performance.
Goal 6: Investigating the descendants of Pokemon species by tracing the evolutionary paths of Pokemon mentioned in the evolves from column, identifying

any evolutionary trends or patterns that emerge across generations and types of Pokemon.

Goal 7: Exploring the characteristics and attributes of legendary Pokemon, including their base stats, abilities, and rarity, and analyze how they differ from non-legendary Pokemon in terms of gameplay mechanics and competitive viability.

Group 8: Analyzing Pokemon rankings to understand the factors influencing a Pokemon's rank, such as its base stats, evolutionary stage, or abilities, and identify any patterns or trends in how Pokemon are ranked across different generations or types.