Pokémon Data Analysis Report

# Introduction

This project focuses on analysing a dataset of Pokémon and their alternate forms using a structured approach to data cleaning, modelling, and visualisation. Pokémon are fictional creatures from a popular media franchise, each with unique combinations of attributes, including elemental types, statistical values (such as Hit Points and Attack), and special “forms” or categories like Mega Evolutions and Legendary Pokémon. The goal of this assignment is to demonstrate professional skills in data analysis using Excel and Tableau, uncover patterns in Pokémon distribution and power, and communicate findings in clear, accessible language suitable for a non-technical audience (Bulbapedia, 2024; Serebii.net, 2024).

# Dataset Description and Data Preparation

The dataset was compiled from reputable public sources including Serebii.net, Bulbapedia, and Kaggle, ensuring a comprehensive list of Pokémon, their forms, types, and stats (Kaggle, 2024; Serebii.net, 2024).   
A key challenge in this analysis is that many Pokémon have multiple “forms”—variations such as Mega Evolutions, Alolan and Galarian forms, or regional variants. Each form may have distinct stats or even types, which can make analysis complex if not handled carefully. For this reason, each unique form was treated as a separate entry in the data, allowing for more granular comparisons.  
  
Data cleaning and preparation steps performed in Excel included:  
- Removal of duplicates: Ensured each Pokémon form (not just species) appeared only once.  
- Standardisation of text fields: The TRIM and PROPER functions were used to make names and types consistent.  
- Handling of missing data: All missing stats were either filled in from trusted sources (Bulbapedia, 2024; Serebii.net, 2024) or marked as “Unknown” if not available.  
- Creation of helper columns: These included “Form” (Regular, Mega, Alolan, etc.), “Category” (Legendary, Mythical, Mega, Regular), and normalised “Type 1” and “Type 2” columns for easier aggregation.  
- Validation of data types: Ensured numeric columns (HP, Attack, Defence, etc.) were correctly formatted for analysis.  
  
This cleaning process was essential to ensure that comparisons were meaningful and all relevant Pokémon forms were included in the analysis.

# Data Modelling and Analytical Approach

The cleaned Excel dataset was imported into Tableau for visual analysis and modelling. In Tableau, calculated fields were created to group, filter, and compare Pokémon based on key features:  
- “Is Mega Form”: Used the formula IF CONTAINS([Name], "Mega") THEN "Mega" ELSE "Regular" END to easily split the dataset into Mega Evolutions and Regular forms.  
- “High Total”: Marked Pokémon with total base stats ≥ 600, using IF [Total] >= 600 THEN "High Total" ELSE "Normal Total" END, to highlight the most powerful Pokémon.  
- Type breakdowns: Visuals aggregated by “Type 1” (primary type) and “Category” for flexible insights into the distribution of Pokémon.  
  
Two main dashboards were created to facilitate exploratory analysis:  
- Dashboard 1 (“Mega vs Regular Overview”) compares the frequency of Mega and Regular Pokémon, both overall and split by primary type.  
- Dashboard 2 (“Type & Category Overview”) shows the count of Pokémon by elemental type and by category (Regular, Mega, Legendary, Mythical).  
  
Interactive filters (e.g., by Type 1) were added to allow users to explore specific subsets of the data. Each dashboard includes concise, plain-English text explanations for accessibility to anyone, regardless of prior Pokémon knowledge.  
  
Stories were built in Tableau to present the findings as a narrative, guiding the reader through key visuals, findings, and summary conclusions.

# Key Findings and Insights

The analysis produced several clear patterns and insights:  
  
- Distribution of Types: Water and Normal are the most common primary types among Pokémon, while types such as Fairy, Dragon, and Flying are less frequent (Kaggle, 2024).  
- Mega and Special Forms: Mega Evolutions are relatively rare compared to Regular Pokémon, but are distributed across a variety of types. Their inclusion in the analysis reveals that they often represent some of the most powerful forms in the Pokémon universe (Serebii.net, 2024).  
- Powerful Pokémon (“High Total”): Pokémon with total base stats of 600 or higher—often Mega, Legendary, or Mythical forms—tend to cluster in certain types, notably Dragon, Psychic, and Fairy. These high-stat Pokémon are both rare and strategically important in competitive play (Bulbapedia, 2024).  
- Category Breakdown: The majority of Pokémon are in the Regular category, with a smaller proportion classed as Mega, Legendary, or Mythical. However, analysis of calculated fields showed that Mega and Legendary Pokémon have significantly higher average stats.  
- Data Cleaning and Modelling: The creation of helper columns and calculated fields was critical to the accuracy of this analysis. By ensuring each Pokémon form and category was clearly defined and separated, the resulting insights are more reliable and meaningful.  
  
This combination of careful cleaning, thoughtful modelling, and visualisation made it possible to uncover trends and relationships that would otherwise be hidden in a raw dataset.

# Communication and Storytelling

A core aim of this project was to make the analysis accessible to those unfamiliar with Pokémon. Each step was supported by:  
- Plain-English explanations in dashboard text boxes, captions, and story slides.  
- Visual clarity: Charts are clearly titled, axes are labelled, and legends are provided for all color-coded fields.  
- Summary findings: Each story concludes with an “Insights & Summary” slide, listing key takeaways in a clear, readable format.  
  
These approaches demonstrate both data analysis skill and effective communication—an essential combination in professional analytics roles (Kaggle, 2024).

# Conclusion

Through robust data cleaning in Excel, precise modelling and calculated fields in Tableau, and a focus on clear communication, this project delivered a comprehensive analysis of Pokémon and their forms. The approach ensured that every form was fairly represented, trends were revealed, and complex information was presented in an understandable way. The techniques used here—especially handling of categorical data and calculated fields—are directly applicable to real-world business intelligence scenarios, making the project both an academic and a practical success.

# Appendix: References (APA Style)

Bulbapedia. (2024). Pokémon Database. https://bulbapedia.bulbagarden.net/  
Serebii.net. (2024). Pokémon Stats and Forms. https://www.serebii.net/  
Kaggle. (2024). Pokémon with stats. https://www.kaggle.com/datasets/abcsds/pokemon-with-stats