# Data Analysis Project

## Data Analyst: Mauricio S. Tan Jr.

## Client/Sponsor: Pokemon Generation 6 Player

## Purpose:

This project will utilize an exploratory data analysis to uncover the follow:

1. **Identifying top-performing Pokémon types** by analyzing total base stats and individual stat categories (e.g., Attack, Speed, HP).

Insights

* **Primary Dragon-types dominate in base stats:** Pokémon with Dragon as their primary type have the highest average total base stats — exceeding second place by approximately 50 points. This suggests Dragon-types are significantly stronger overall compared to other primary types. (Bar Chart)
* **Dragon-types** have the highest average **attack**, **Psychic-types** have the highest average **sp. Attack**, and **Steel**-**types** have the highest average **Defense.**
* **Top type combinations feature Dragon, Psychic, or Steel:** Type pairings that include Dragon, Psychic, or Steel tend to rank consistently at the top in terms of average total stats, showing these types are commonly associated with high-performing Pokémon. (Radar/Spider Chart)

1. **Visualizing Pokémon population growth over time** by showing how many new Pokémon were introduced in each generation.

Insights

* **Generation 1** introduced the most Pokémon, totaling **192** species. (Bar Chart)
* **Generations 2 to 5** maintained relatively high addition counts, each adding over 100 new Pokémon. (Line Chart)
* Starting from **Generation 6 onwards**, there’s a noticeable decline in the number of new Pokémon introduced — each generation after adds **fewer than 100**. (Line Chart)
  + Gen 6: 85
  + Gen 7: 99
  + Gen 8: 87

1. **Analyzing the frequency and distribution of dual-type Pokémon**, including trends across generations and common type pairings.

Insights

* **Distribution of single and dual-type across generations:** Generations 6 and 7 have higher dual-type Pokémon, while the rest have higher single type Pokémon’s.
* **Frequence of type combos:**
  + Flying/Normal-type Pokémon are the most frequent, followed by Grass/Poison with 14, and in third place is Bug/Flying.
  + Pokémon with Bug and Dark typings have the highest number of unique combinations, followed by Dragon types.

1. **Comparing Pokémon statuses (normal, sub-legendary, legendary, mythical)** to understand how they differ in stats and other attributes.

Insights

* **Legendary Pokémon** are the strongest in terms of classification, leading for all 6 of the stats. **Mythical Pokémon** have better offensive stats than **sub-legendaries**, while **sub-legendary Pokémon** have better defensive stats than the former. **Normal Pokémon** are the weakest.

1. **Uncovering stat distribution trends across generations** to identify shifts in design emphasis (e.g., more speed-focused Pokémon).

Insights

* The **attack stat** is on average higher generation 2, showing that there was a design focus on having higher damage for Pokémon. On the other hand, **speed stat** is the lowest across all generations.

1. **Investigating the relationship between Pokémon strength and catch rate** — are stronger Pokémon harder to catch?

* There is a **moderate to strong negative correlation (-0.62)** between **total base stats** and **catch rate.** This means that stronger Pokémon are generally harder to catch. This aligns with game mechanics, though other factors like rarity, generation, and type of Poké Ball may also affect catch rate.
* There is one notable outlier in terms of strength and catch rate: **Eternamax Eternatus**. Despite being the strongest Pokémon in terms of base stats, it has the **highest catch rate** because it is required to be caught to progress in the storyline of *Pokémon Sword and Shield*. This makes it an exception to the general trend where stronger Pokémon are typically harder to catch.

1. **Generating a type-based performance** overview by calculating average stats grouped by primary and secondary type.

* **Dual-types are generally stronger:** Across all generations, dual-type Pokémon have higher average total stats compared to single-type Pokémon. This indicates that type synergy may be tied to stronger stat distributions. (Grouped Bar Chart)

1. **Analyzing Pokémon** **growth rates per type** to observe any trends in development pace.

* **Medium Fast is the most common growth rate overall**, especially among:
  + **Normal-type Pokémon** (55)
  + **Water-types** (45)
  + **Bug-types** (32)
  + **Electric-types** (34)
* **Medium Slow is the second most common growth rate**, with:
  + **Grass-types** (51)
  + **Fire-types** (34)
  + **Water-types** (37)
* **Normal-type Pokémon are heavily concentrated in Medium Fast**, with very few in other categories, showing the least distribution diversity.
* **Grass-type Pokémon are more evenly spread**, but still highly concentrated in Medium Slow and Medium Fast, suggesting a slightly slower growth pattern.
* **Water-type Pokémon appear across all growth rates**, making them the most **evenly distributed type**, though still favoring Medium Fast and Medium Slow.
* **Dragon-types are mostly in the Slow category** (37), which is consistent with their reputation for being powerful but requiring more experience to level up.
* **Fairy- and Ghost-types mostly follow the Medium Fast growth pattern**, with limited distribution across other rates.
* **Fighting- and Fire-types tend toward Medium Slow**, aligning with their mid-to-late game strength in Pokémon games.
* **Rock- and Steel-types** are among the least distributed, with lower overall counts across all growth rates.
* **Electric-types favor Medium Fast**, with a smaller presence in Medium Slow and Slow, suggesting quicker leveling curves.

1. **Top 10 strongest pokemon based on total stat points**

* The Pokémon with the highest base stat total is **Eternamax Eternatus**, with an unmatched 1125. Following that are **Mega Mewtwo Y**, **Mega Mewtwo X**, and **Mega Rayquaza**, each with a base stat total of 780. **Primal Kyogre** and **Primal Groudon** come next at 770, while **Ultra Necrozma** stands at 754. Rounding out the group are **Zamazenta (Crowned Shield)**, **Zacian (Crowned Sword)**, and **Arceus**, each with a total of 720. These Pokémon represent some of the most powerful in the franchise, often playing key roles in their respective games’ storylines and legendary rosters.

1. **Top pokemon based on each specific stat**

* The highest **HP** stat belongs to both **Blissey** and **Eternamax Eternatus**, each with a value of **255**. The highest **Attack** stat is held by **Mega Mewtwo X** at **190**, while **Eternamax Eternatus** also leads in **Defense** and **Special Defense**, both at **250**. For **Special Attack**, **Mega Mewtwo Y** holds the top spot with **194**. Lastly, the fastest Pokémon is **Deoxys (Speed Forme)**, with a **Speed** stat of **180**.

## Scope / Major Project Activities:

What are the major parts of this project? List out the high-level steps, activities, or stages of the project, and give a brief description for each.

|  |  |
| --- | --- |
| Activity | Description |
| Data Collection | Download and review the Pokemon dataset from Kaggle |
| Data Cleaning (Excel) | Clean dirty data and format data for easier processing in order to maintain data integrity |
| SQL Practice & Exploration | Import dataset into SQL to run queries and uncover basic stats |
| Python EDA & Visualization | Use Python (Pandas, Matplotlib/Seaborn) to generate insights and graphs. |
| Dashboard Design (Tableau) | Create an interactive dashboard showcasing key stats |
| Documentation & Reporting | Summarize methods, visuals, and findings through a github repository and kaggle notebook |
|  |  |

## This project does not include:

Specify the things that this project isn’t responsible for doing (out of scope). For instance, “this project does not involve a summation of 2019 data analysis”

* Machine learning or predictive modeling
* Pokémon battle mechanics or game-based simulations
* Pokemon from Genrations 7 and above
* Text-based data (e.g., Pokémon descriptions or movesets)

## Deliverables:

A specific list of things that your project will deliver.

|  |  |
| --- | --- |
| Deliverable | Description/ Details |
| Cleaned Dataset | CSV or Excel file with cleaned, standardized Pokémon data for use in EDA and Tableau. |
| SQL Queries File | .sql file containing queries used for filtering, grouping, and summary insights. |
| Python EDA Notebook | Jupyter Notebook (.ipynb) showcasing visualizations, summary statistics, and findings. |
| Tableau Dashboard File | Interactive dashboard with key metrics and visuals |
| README/Project Report | Written summary of the project’s purpose, process, and findings, suitable for portfolio. |
| GitHub Repository | A public GitHub repo containing all project files, code, and documentation. |
| Kaggle Notebook | A published version of the Jupyter Notebook on Kaggle for visibility and community review. |

## Schedule Overview / Major Milestones:

The expected schedule for the project. This can be defined by milestones (e.g. “all data is cleaned and processed”), periods of time (“Week 1 / Week 2”), or other ways based on the needs of the project.

|  |  |  |
| --- | --- | --- |
| Milestone | Expected Completion Date | Description/Details |
| Dataset Downloaded & Reviewed | June 18, 2025 | Review column meanings, structure, and types from Kaggle. |
| Data Cleaning in Excel | June 18, 2025 | Handle missing data, rename columns if needed, and prepare for import into Python/SQL. |
| SQL Querying | June 20, 2025 | Practice exploratory queries to understand type counts, averages, and stat distributions. |
| Python EDA and Visualizations | June 22, 2025 | Create visuals (histograms, box plots, correlations, time series, etc.). |
| Tableau Dashboard Finalized | June 24, 2025 | Build interactive dashboard with filters, stat summaries, and key visuals. |
| Capstone Documentation | June 25, 2025 | |  | | --- | |  |   Final write-up |

## \*Estimated date for completion:

This is my “if all goes well and I have everything I need, this is when I’ll be done” date.

June 27, 2025