# 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)
* **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)
* **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.

* **Stacked Bar Chart**
* **Heatmap** (advanced): Frequency matrix of all type combinations of dual types

1. **Comparing Pokémon statuses (normal, sub-legendary, legendary, mythical)** to understand how they differ in stats and other attributes.
2. **Uncovering stat distribution trends across generations and types** to identify shifts in design emphasis (e.g., more speed-focused Pokémon).
3. **Investigating the relationship between Pokémon strength and catch rate** — are stronger Pokémon harder to catch?
4. **Generating a type-based performance** overview by calculating average stats grouped by primary and secondary type.
5. **Analyzing Pokémon** **growth rates per type** to observe any trends in development pace.
6. **Top 10 strongest pokemon based on stats**
7. **Top pokemon based on each specific stat**

## 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