

My Journey into Power BI: Analyzing PUBG Solo Matches to Discover the Best Playstyle Strategy

By: Rey Anthony Rodriguez | 000931509

DATA039- 008

Table of Contents:

- I. Introduction**
- II. Choosing the Dataset: Why PUBG?**
- III. Defining the Study Scope and Limitations**
- IV. Key Questions to Answer**
- V. Extracting and Preparing the Data**
- VI. Data Analysis: Breaking Down the Game**
 - a. Key Metrics and Statistics**
 - b. Understanding Winning Correlations**
 - c. Identifying Player Strategies**
 - d. Key Findings and Visualizations**
- VII. Conclusion**
- VIII. Final Thoughts**

I. Introduction

As a **Data Science student**, I've always been fascinated by the ability of data to **uncover hidden patterns** and **tell compelling stories**. But raw data alone isn't enough—you need the right tools to **transform it into meaningful insights**. That's where **Power BI** comes in.

To apply my learning, I was given a task: **find a public dataset, clean and transform the data, analyze key insights, and present findings using Power BI dashboards**. The goal was to simulate how a data analyst would approach a **real-world business case**.

I wanted a dataset that was **fun, competitive, and strategic**. Something that wasn't just numbers but had real-world implications. That's when I stumbled upon **PUBG match data** on Kaggle, and an idea clicked:

What is the best strategy to win in PUBG solo mode?

II. Choosing the Dataset: Why PUBG?

PUBG (**PlayerUnknown's Battlegrounds**) is a **popular battle royale game** where **up to 100 players** are dropped onto an island, and the last one standing wins. Players adopt different strategies—some play aggressively, hunting down opponents, while others take a more passive approach, avoiding fights and focusing on survival.

I found a **public dataset on Kaggle** that contained detailed match statistics for PUBG players, including:

- **Kills, damage dealt, assists**
- **Distance traveled, time survived**
- **Healing and boosting items used**
- **Final placement in the match**

With this data, I set out to analyze **what really contributes to winning in solo mode**.

III. Defining the Study Scope and Limitations

The **original dataset** contained **over 1 million rows**, covering different game modes (solo, duo, and squad). However, for this analysis, I wanted to focus **only on solo matches**, so I applied the following filters:

- **Only Solo and Solo-FPP matches** (no teams)
- **Reduced dataset from 1,048,575 rows → 179,533 rows**
- **Luckily after filtering the data, the remaining were only distinct players who played one match**. This is a great coincidence to avoid repeating biases.

Another limitation: **Not all matches had 100 players**. Some had as few as 20 players, which meant that match sizes varied. I kept this in mind while analyzing results.

IV. Key Questions to Answer

To structure my analysis, I focused on three main questions:

- 1. How do player actions impact winning in Solo Mode?**
Which factors—kills, movement, healing, etc.—correlate with higher placement?
- 2. What is the best strategy to win?**
Is it better to play aggressively, balanced, or passively?
- 3. Who are the top 5 solo players in the dataset?**
Which players performed the best across key metrics?

V. Extracting and Preparing the Data

Step 1: Connecting the Data to Power BI

I loaded the CSV file into Power BI using the "Get Data" feature. Since my dataset was already in a structured table format, creating relationships between multiple tables was unnecessary.

Step 2: Data Transformation

Using Power Query, I cleaned the dataset by:

- Removing non-solo matches.
- Keeping only relevant columns (kills, damage dealt, travel distance, healing items, etc.).
- Ensuring data types were correct.

This left me with a flat table, which was perfect for my analysis.

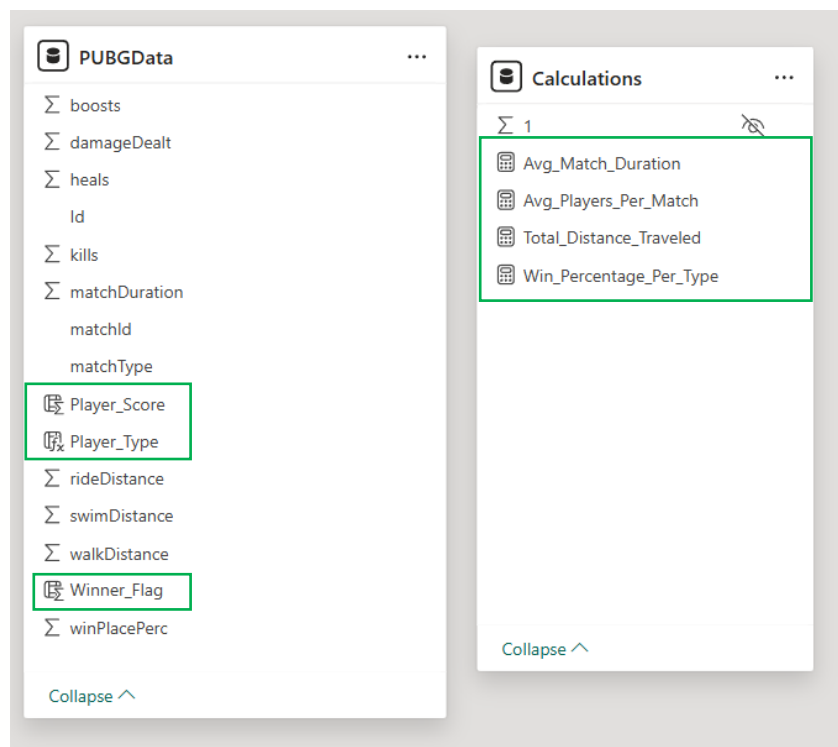


Figure 1 Table View

VI. Data Analysis: Breaking Down the Game

a. Key Metrics and Statistics

To understand the dataset better, I calculated some key measures:

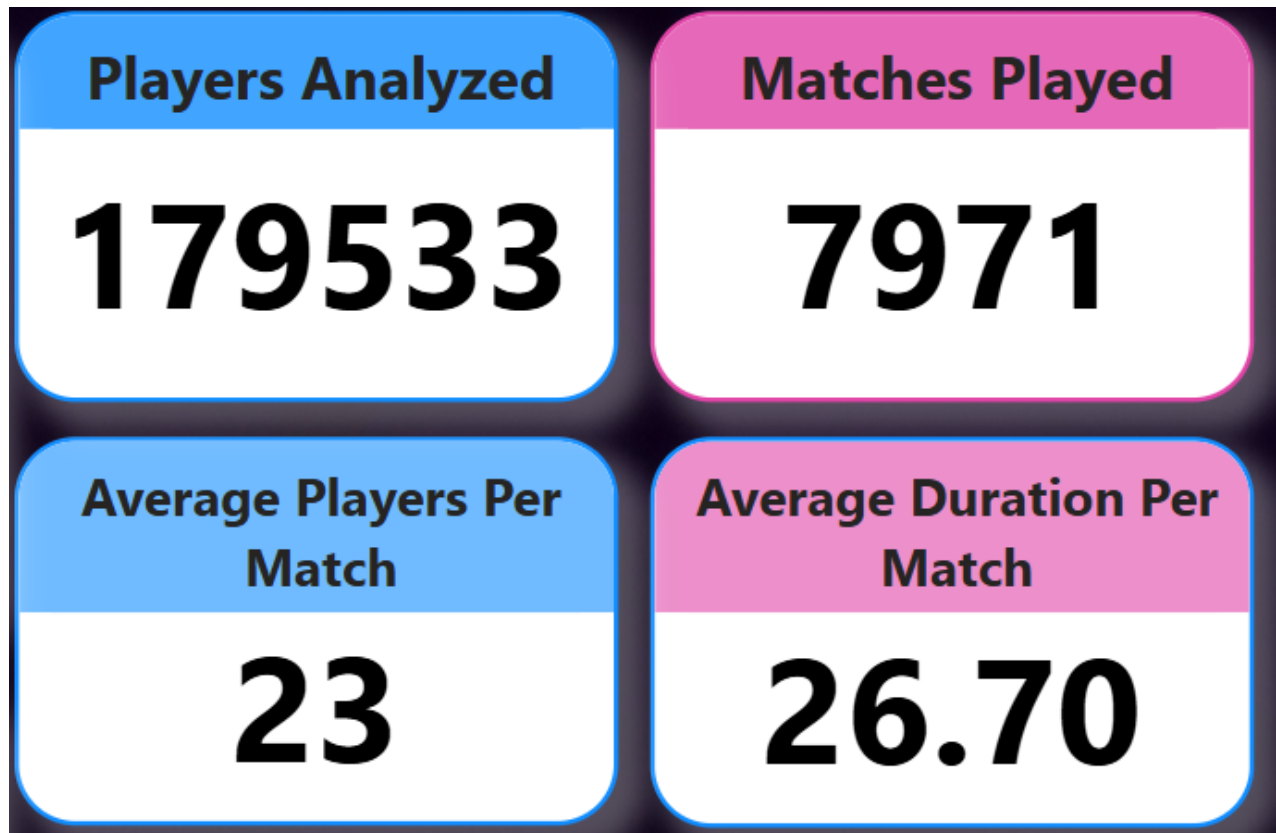


Figure 2 Key Metrics and Statistics

- **Total Players Analyzed:** 179,533
- **Total Matches:** 7,971
- **Average Players Per Match:** 23
- **Average Match Duration:** 1602.19 seconds (~27 minutes)

b. Understanding Winning Correlations

To determine **which factors impact winning**, I created scatter plots to visualize the relationship between different gameplay actions and a player's final ranking (**WinPlacePerc**).

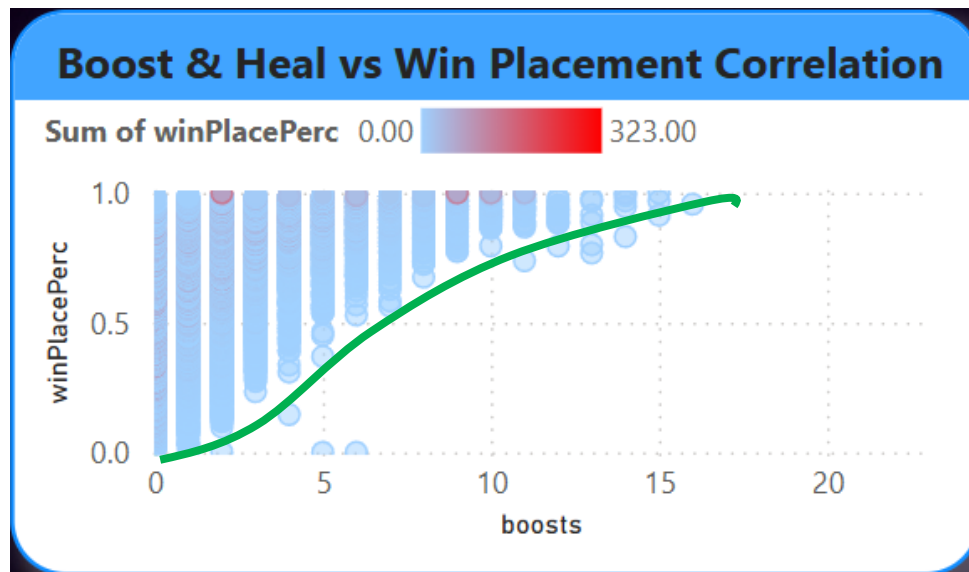


Figure 3 Correlation of Boosts and Heal to Win Placement

- **Boosts and Heals vs. Placement:** Players who used more healing and boost items tended to survive longer and place higher.

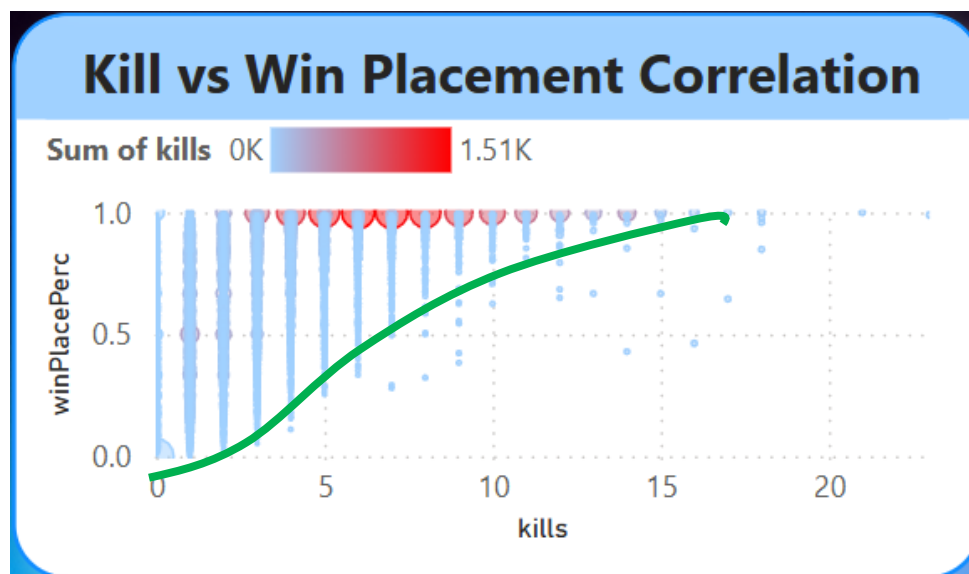


Figure 4 Correlation of Kills to Win Placement

- **Kills and Damage Dealt vs. Placement:** A clear trend showed that **the more kills and damage a player inflicted, the better their final ranking.**

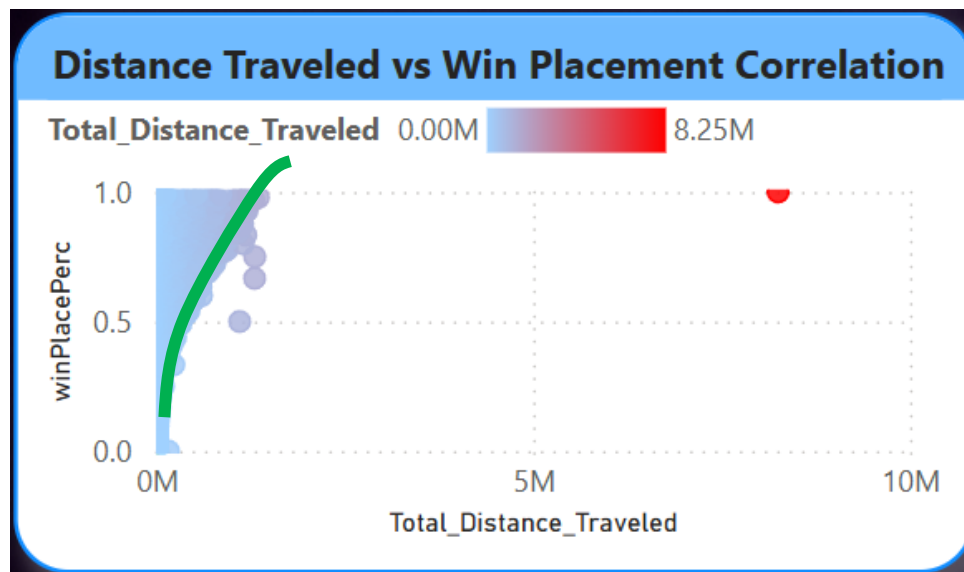


Figure 5 Correlation of Distance Traveled to Win Placement

- **Travel Distance vs. Placement:** Interestingly, players who moved more (walking, riding, or swimming) had better placements, suggesting that active movement plays a role in survival.

c. Identifying Player Strategies

Based on these correlations, I developed a **Player Strategy Model** by categorizing players into three types:

- **Aggressive Players:** High kills, high damage dealt, high movement.
- **Balanced Players:** A mix of offensive and defensive playstyles.
- **Passive Players:** Low kills, minimal movement, focusing purely on survival.

To assign a score to each player, I created a **calculated column in DAX**:

```

1 Player_Score =
2 VAR Kill_Score = PUBGData[kills] * 3
3 VAR Damage_Score = DIVIDE(PUBGData[damageDealt], 100, 0)
4 VAR Move_Score = DIVIDE(PUBGData[walkDistance] + PUBGData[rideDistance] + PUBGData[swimDistance], 100, 0)
5 VAR Heal_Score = PUBGData[heals] + PUBGData[boosts]
6
7 VAR Total_Score = Kill_Score + Damage_Score + Move_Score + Heal_Score
8
9 RETURN
10 Total_Score

```

Figure 6 Player Score Calculation

This formula ensured that players were grouped based on their overall engagement in the match.

d. Key Findings and Visualizations

1. Distribution of Player Types (Pie Chart)

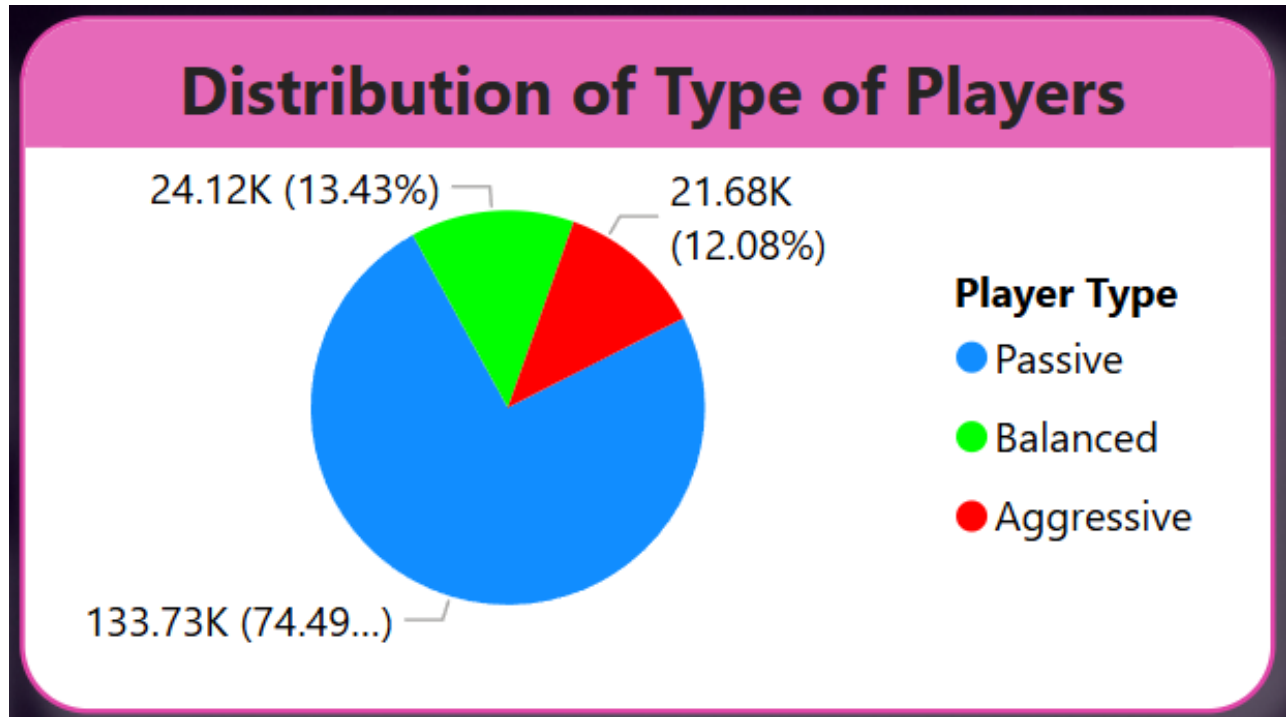


Figure 7 Distribution of Type of Players

- 75% of players were **passive**.
- 13% were **balanced**.
- 12% were **aggressive**.

This showed that **most players preferred a passive survival strategy**.

2. Win Rate by Player Type (Clustered Bar Chart)

To understand which strategy worked best, I calculated the **win percentage for each type**:

```
1 Win_Percentage_Per_Type =  
2 VAR TotalWinners = CALCULATE(COUNT(PUBGData[Id]), PUBGData[Winner_Flag] = 1)  
3 VAR TotalPlayers = CALCULATE(COUNT(PUBGData[Id]))  
4 RETURN  
5 | IF(TotalPlayers = 0, BLANK(), DIVIDE(TotalWinners, TotalPlayers))  
6
```

Figure 8 Win Percentage Per Type Calculation

The results:

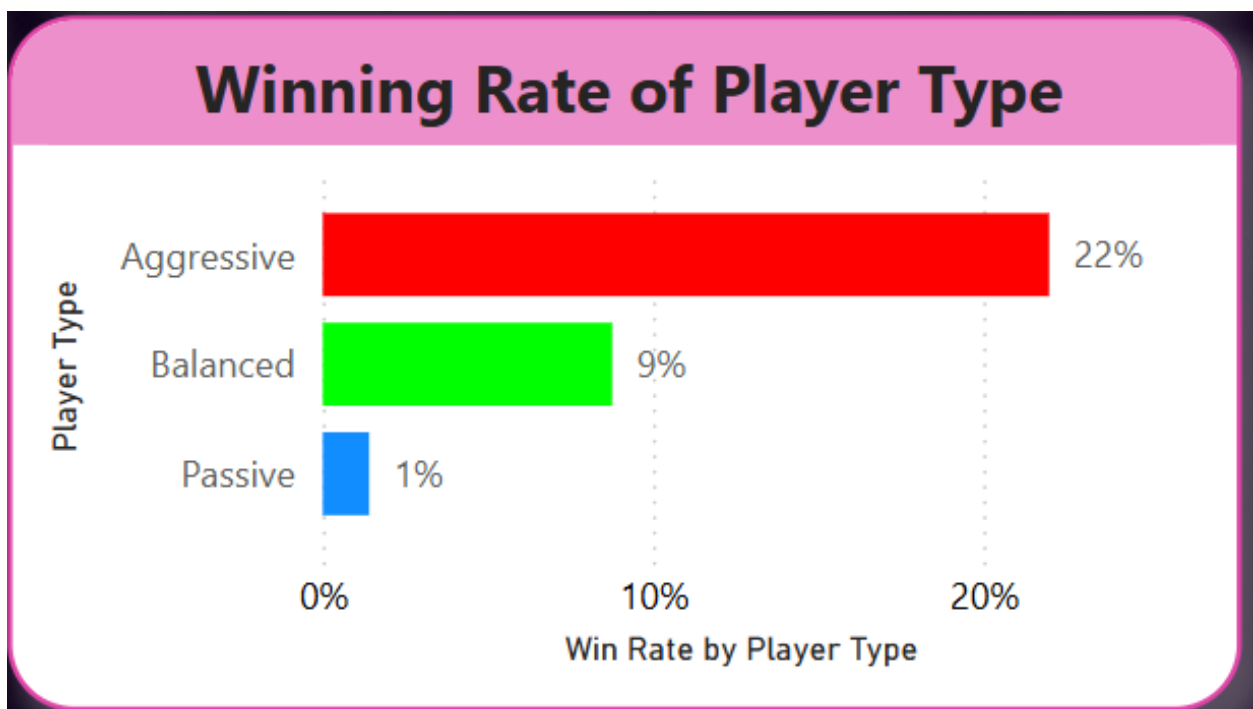


Figure 9 Win Percentage per Player Type

- **Aggressive players: 22% win rate**
- **Balanced players: 8% win rate**
- **Passive players: 1.5% win rate**

This proved that **aggressive playstyles lead to the most wins.**

3. Top 5 Players (Table Visualization)

Top 5 Players based on Score		
Id	Player_Score	Winner_Flag
f09d0c99127850	387.78	1
c3fabfce7589ae	305.11	1
6c74f0dc7ac84f	304.35	0
9d0b5863b94a5f	300.80	1
30b291c24dfdc0	287.49	0
Total	WIN	LOSE

Figure 10 Top 5 Solo Players

To further illustrate this point, I examined the top 5 players based on their Player Score—which factors in kills, damage dealt, travel distance, and item usage. These were then filtered to show only players who won their matches.

As you can see, the players who scored highest across these metrics also ended up winning. This reinforces our finding that Aggressive behavior, characterized by more kills, higher damage, and more active movement, leads to higher chances of winning.

VII. Conclusion: What's the Best Strategy to Win?

From my analysis, the best strategy for winning PUBG solo mode is:

- **Be aggressive** – Engage in fights, take down enemies.
- **Deal high damage** – The more you attack, the better your placement.
- **Move around the map** – Players who traveled more had better placements.
- **Use healing items strategically** – Survivability matters.

While **passive play** is the most common approach, it has the lowest success rate. Players who **actively engage and take risks are more likely to win.**

Future Considerations

One limitation of this study is that it **doesn't account for player skill level.** Future analyses could include metrics like headshot accuracy or survival time to refine the findings further.

VIII. Final Thoughts: My Journey with Power BI

This project was an incredible learning experience. Not only did I improve my **Power BI skills**, but I also gained valuable insights into **data storytelling**.

By exploring a **real-world dataset**, I learned how to:

- Clean and transform data in **Power Query**
- Create **DAX calculations** to categorize players
- Visualize and analyze trends using **Power BI charts**
- Draw conclusions based on **data-driven insights**
-

This was just the beginning. With Power BI, I'm excited to explore **even more datasets** and uncover new stories hidden in the data!

BONUS: This is the completed dashboard!

