Data-Driven Player Insights: Fortnite

Overview and Objective

This small project analysed a public Kaggle Fortnite performance dataset to practice data cleaning, transformation and analysis skills. Using Python, I engineered custom metrics and functions to categorise players, revealing trends in eliminations and accuracy, alongside the influence of behavioural factors. The output is a cleaned dataset ready for data visualisation.

Summary of Insights

1. Those with more eliminations exhibited better placement.
2. Higher eliminations did not equal higher headshot accuracy.
3. Time of day had no impact on elimination performance.
4. Reported mental state ("high" vs "sober") had no impact on elimination performance.

Analytical Approach

1. Data Preparation

- Identified and handled missing values, removed duplicates, and detected impossible values (e.g., negative placements, extreme eliminations).
- Standardised column names to ensure consistency and avoid errors during analysis.

2. Data Transformation

- Constructed custom metrics (e.g., elimination level, accuracy groupings) for comparative analysis.
- Added engineered variables as new columns to the cleaned dataset while preserving the raw CSV.
- Applied conditional logic and Boolean expressions to segment performance by contextual factors such as time of day and reported mental state.
- Employed aggregation functions to quantify relationships (e.g., eliminations vs. time of day, accuracy vs. eliminations).

3. Data Aggregation & Analysis

- Aggregated and summarised performance data using Python; calculating rounded averages, proportions, value counts, and correlations between key metrics.
- Key analyses included:
  - Calculated minimum, maximum, and average eliminations to establish performance thresholds for categorising players.
  - Categorised players by elimination levels to segment performance using IF/ELIF/ELSE.
    - *Note: Equivalent logic can be implemented using a SQL CASE query.*

   - Determined percentages of players in each elimination category.
   - Calculated headshot accuracy, then compared to elimination categories.
   - Explored average eliminations by time of day and reported mental state (“sober” vs “high”) to identify potential behavioural patterns and their impact on performance.

Key Insights

1. The Trend Between Placement and Eliminations

- Players with more eliminations tended to place better in games, as expected.
- Beyond eliminating opponents, higher-elimination players may gain greater map control and resource advantage, indirectly improving placement.

2. Headshot Accuracy vs. Eliminations

Analysis revealed a counterintuitive trend: “high” elimination category players had the lowest average headshot accuracy, while “low” category players had the highest. Possible explanations:

- **Playstyle:** High-elimination players may use rapid, aggressive firing to secure kills quickly, reducing precision; low-elimination players take fewer, more carefully calculated shots.
- **Risk engagement:** High-elimination players may seek more chaotic or risky combat situations (e.g., multiple opponents, high-pressure scenarios), sacrificing accuracy for elimination volume.
- **Excluded contextual factors:** Weapon choice, strategy, whether they had a teammate, or situational factors outside the dataset could better explain this trend.

3. Performance Variations at Different Times of Day

- Average eliminations varied slightly across different times of day, with no significant trends observed.
- Conceptually, performance at different times of day may be influenced by the following factors: player alertness, availability, or engagement levels, as well as external influences such as fatigue or internet conditions.

4. The Impact of “Mental State” on Performance

- Analysis revealed no meaningful performance differences between players reporting as “sober” versus “high.”
- This binary classification is overly simplistic and relies on self-reported data, introducing ambiguity: individual perceptions of what constitutes “high” versus “sober” may vary considerably.
- Conceptually, players reporting as “high” could experience greater relaxation or reduced performance anxiety, whereas “sober” players might benefit from enhanced focus, accuracy, or situational awareness. But this cannot be reliably distinguished given the dataset’s limitations.

Limitations and Recommendations

Dataset and metric limitations may influence result interpretation, suggesting opportunities for improved collection and analysis:

- **Limited and potentially biased sample (~90 players):** Expand data collection to include a larger, more representative cohort, spanning novices to expert players.
- **Binary mental state measurement lacks nuance, and qualitative factors are absent:** Capture more granular mental state data and integrate qualitative measures (e.g., strategy, enjoyment, solo vs team) to contextualise performance outcomes.
- **Single-day data limits reliability:** Implement longitudinal data collection (≥30 days) to account for day-to-day variation in player performance.
- **Time-of-day trends may be influenced by external factors:** Record contextual factors (e.g., fatigue, internet performance) that could influence observed performance variations over time.
- **Unspecified game mode:** Document gameplay conditions (e.g., building mode, weapon selection), as individual performance may vary across different playstyles.
- **Custom elimination categories may not generalise:** Validate curated metric thresholds against larger or benchmark datasets to ensure broader applicability.
- **Statistical significance was not assessed:** Perform tests (e.g., t-tests, ANOVA) to determine whether observed differences in performance are meaningful.

Conclusion

Through cleaning and feature engineering, this dataset was transformed into a reliable foundation for analysis. The results confirmed expected relationships—such as higher eliminations observed with better placements—while also surfacing counterintuitive findings, such as players performing well in eliminations also showcasing lower headshot accuracy. This underlines the importance of analysing multiple metrics together, rather than a single indicator. Extending data collection across a longer timeframe would enable more robust longitudinal analysis and open the door to predictive modelling of player performance. The dataset is limited in scope and not representative of the broader Fortnite community; however, it serves as a sandbox to demonstrate transferable skills in data cleaning, transformation and feature engineering.