BA820 – Project M2

Cover Page

- **Project Title: Survival Strategies, Attrition Patterns, and Audience Engagement in the Alone TV Series**
- **Section and Team Number: A1 – Team 08**
- **Student Name: Marcus Shi**

1. Refined Problem Statement & Focus (~0.5 page)

This milestone focuses on understanding whether participant exits in the television series *Alone* follow identifiable and recurring pathways rather than occurring as isolated or random events. Specifically, I investigate the following question:

Are there identifiable exit pathways—such as early voluntary withdrawal versus later medical evacuation—that differ meaningfully in timing and survival duration?

This question builds directly on the framing established in M1, which emphasized heterogeneity in survival outcomes and exit reasons. The core problem definition remains unchanged; however, the focus has been refined from descriptive exploration to structural pattern discovery. Rather than asking what exits occur, this milestone asks how exit characteristics co-occur and whether these co-occurrences form consistent pathways.

Several assumptions from M1 were tested and revised in this process. An initial assumption that early exits would be dominated by voluntary or psychological withdrawal was partially invalidated. Early exploratory experiments in M2 revealed that medical evacuation frequently co-occurs with very short survival durations, challenging the notion that medical exits are primarily late-stage events. Conversely, another assumption—that prolonged survival would be associated with higher medical risk—was reinforced by the strong association between medical evacuation and longer survival duration bins.

Overall, this milestone shifts the analytical lens from outcome prediction to pathway discovery, treating exit timing, exit cause, and medical involvement as interacting components of a participant's survival trajectory rather than independent attributes.

2. EDA & Preprocessing: Updates (~0.75 page)

The exploratory analysis conducted in M1 established several key patterns that directly motivated the analytical approach used in this milestone. Survival duration was shown to be strongly right-skewed, with most participants exiting early and a small subset enduring substantially longer. Exit reasons also displayed structure, with psychological and hunger-related exits occurring earlier and medical evacuations tending to appear later in the survival timeline. These findings suggested that exit outcomes are not independent events, but instead reflect interacting dimensions of timing, cause, and endurance.

To support association rule mining in M2, additional preprocessing steps were required. Because association rules operate on discrete item co-occurrence rather than continuous variables, survival duration was discretized into coarse bins (0–7, 8–21, 22–45, 46–70, and 71+ days). These bins were chosen to reflect meaningful survival phases observed in M1 while maintaining sufficient support given the small sample size.

Exit outcomes were also decomposed into multiple attributes rather than treated as a single categorical label. Specifically, exit timing (early vs. not early), medical evacuation status, and exit category were encoded as separate tokens. Early experiments that treated exit outcome as a single label produced sparse or trivial rules, motivating this more granular transaction design.

Finally, item frequency inspection was conducted to ensure that key attributes appeared frequently enough to justify a minimum support threshold of 5%. No additional data cleaning beyond what was performed in M1 was required. These preprocessing choices were essential to balancing interpretability with statistical stability in the subsequent analysis.

3. Analysis & Experiments (~1.5 page)

Method Selection and Analytical Rationale

To investigate whether participant exits in *Alone* follow structured and recurring pathways rather than occurring randomly, I applied association rule mining using the Apriori algorithm. This unsupervised approach is well suited for identifying interpretable co-occurrence patterns among categorical attributes without requiring predefined outcome labels.

The objective of this milestone is pathway discovery rather than prediction. Association rules allow exit characteristics—such as timing, medical involvement, and survival duration—to overlap and interact, reflecting the reality that participants may share partial similarities rather than belonging to mutually exclusive groups. Compared to clustering methods, which require distance metrics and a fixed number of clusters, association rules provide transparent and easily interpretable results that are appropriate given the small sample size.

The analysis focuses on rules of the form:

{exit timing, medical status, exit category} → {survival duration bin}

These rules directly support the goal of characterizing exit pathways rather than forecasting outcomes.

Transaction Design and Feature Construction

Each participant was treated as a single transaction composed of discrete exit-related tokens. Transactions included season identifier, survival duration bin, exit timing (early vs. not early), medical evacuation status, and exit category derived from the recorded reason for tapping out or winning.

Survival duration was discretized into coarse bins (0–7, 8–21, 22–45, 46–70, and 71+ days) to reflect the strong right-skew observed in M1 while maintaining sufficient item support. Early exits were defined as occurring within the first seven days, consistent with observed attrition patterns.

Initial experiments treated exit outcome as a single categorical label. This design produced either trivial identity rules or extremely sparse itemsets. Decomposing exit outcomes into multiple attributes proved necessary to expose meaningful structure and avoid masking interactions among timing, medical status, and exit cause.

Frequent Itemset Mining and Rule Generation

Transactions were one-hot encoded and frequent itemsets were mined using the Apriori algorithm with a minimum support threshold of 5%, selected after inspecting individual item frequencies. This threshold balanced interpretability with statistical stability given the limited sample size.

Association rules were generated using confidence as the primary metric, with a minimum threshold of 0.6, and ranked by lift to identify relationships that occurred more frequently than expected under independence. The analysis prioritized rules whose consequents involved survival duration bins, medical evacuation, or exit categories, as these directly address the research question.

Several strong and interpretable rules emerged, particularly those linking early exit timing with short survival durations and medical evacuation with longer survival bins.

Iterations, Surprises, and Limitations

The most effective aspect of this analysis was the transaction design, which allowed exit pathways to emerge organically rather than being imposed through predefined labels. Using coarse survival duration bins was especially effective, as it balanced clarity with sufficient item support.

One unexpected finding was the strength of the association between early exit timing and medical evacuation. I initially expected early exits to be dominated by voluntary or psychological withdrawal, but the rules revealed that medical evacuation frequently co-occurred with very short survival durations. Another surprising result was the consistent association between medical evacuation and longer survival duration bins, suggesting that medical exits often represent late-stage failure after prolonged endurance rather than early weakness.

This analysis is constrained by the small sample size and the sensitivity of association rules to discretization choices. Results should be interpreted as descriptive co-occurrence patterns rather than causal relationships, and rare exit categories remain difficult to analyze due to limited support.

4. Findings & Interpretations (~0.75 page)

The association rule analysis reveals that participant exits in Alone follow structured and recurring pathways, rather than occurring randomly. Exit timing, medical involvement, and

survival duration consistently co-occur, indicating that exits reflect distinct stages of the survival process rather than independent outcomes.

A dominant pattern is the strong association between early exit timing and very short survival durations. Participants exiting within the first seven days overwhelmingly fall into the shortest survival bins, confirming that early attrition represents a distinct and unstable phase of the competition. Contrary to initial expectations, early exits are not driven solely by voluntary or psychological withdrawal. Medical evacuation frequently appears alongside early timing and short survival durations, suggesting that some early exits reflect acute physical failure or rapid environmental stress rather than discretionary decisions.

A contrasting pathway emerges among participants who survive longer. Medical evacuation is strongly associated with extended survival durations, indicating that medical exits often occur after prolonged endurance. This suggests that as participants persist, cumulative physical strain increases the likelihood of medically enforced removal. Endurance therefore functions as both a marker of successful adaptation and a source of escalating risk.

Finally, certain exit categories—such as family or personal withdrawals—rarely appear in early survival bins. These exits tend to occur after participants have already invested substantial time in the competition, reinforcing the idea that exit causes align with specific survival stages rather than occurring interchangeably.

Overall, these findings support a reframing of exit behavior in Alone as a progressive process involving early instability, sustained adaptation, and late-stage physical risk. Even with a small sample size, association rule mining reveals meaningful structure in how exit timing, cause, and survival duration jointly shape participant outcomes.

5. Next Steps (~0.25 page)

This milestone identifies structured exit pathways based on timing, medical involvement, and survival duration, but it does not explain how participant decisions contribute to these outcomes. In particular, survival strategy choices—such as loadout composition—have not yet been incorporated into the exit pathway analysis.

The next stage of this project will integrate loadout data to examine whether specific item combinations are systematically associated with different exit pathways. Unsupervised methods such as clustering or similarity analysis can be used to identify implicit strategy types, which can then be linked to early exit, prolonged endurance, or medical evacuation patterns identified in this milestone.

Future work may also explore season-level context, including environmental conditions and early-drop rates, to assess whether exit pathways vary across seasons. These extensions will move the analysis from pathway identification toward a more complete understanding of how strategic decisions and endurance jointly shape survival outcomes.

Appendix

Shared GitHub Repository (Required)

GitHub Repository: https://github.com/Marcusshi/BA820-A1-08

My individual work for Project M2 is contained in the following files:

- M2/Marcus/BA820_A1_08_M2_Marcus.ipynb
- M2/Marcus/BA820_A1_08_M2_Marcus_Report.pdf

Supplemental Material (Highly Recommended)

Additional supporting material is included within the M2 notebook, including:

- Item frequency tables used to select minimum support thresholds for Apriori.
- Examples of high-lift association rules linking exit timing, medical evacuation status, and survival duration bins.
- Sensitivity checks demonstrating the impact of alternative survival duration binning schemes.

Process Overview

At a high level, the analytical pipeline for Project M2 followed these steps:

1. Load and minimally clean survivalist-level data.
2. Discretize survival duration and decompose exit outcomes into timing, medical status, and exit category.
3. Represent each participant as a transaction of exit-related attributes.
4. Apply association rule mining (Apriori) to identify frequent co-occurrence patterns.
5. Interpret high-confidence and high-lift rules as distinct exit pathways.

Use of Generative AI Tools

Generative AI tools were used in a limited and transparent manner to assist with code structuring, markdown organization, and phrasing clarity. All analytical decisions, feature design, interpretation of results, and conclusions were independently developed and validated by me. No generative AI tools were used to generate or modify empirical results.

https://chatgpt.com/share/69895f85-d3a8-800e-8600-86f7098e8332