BA820 – Alone TV Show Project M2

Team B02

Mika Ismayilli

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

I am investigating whether the survival item selection among the contestants reflects survival strategies and whether or not these strategies are associated with longer survival outcomes. For this analysis, I will use the contestant-item selection level as a primary unit where each contestant's loadout represents a strategic choice.

The success metrics will be reflected in variables like days_lasted, medically_evacuated and tap_out_reason. These metrics capture both endurance and failure modes of the contestants.

This question matters to multiple stakeholders. For example, contestants and their trainers can use insights to refine the preparation and prioritization of items that supports longer survival rather than comfort. Other stakeholders include the producers and rule designers of the TV Show. They can benefit from this analysis to figure out whether or not some items promote dominant strategies and might reduce the suspense and variability. With this, they can adjust the rules and constraints to keep the competitive balance as well as ensure safety. Lastly, this analysis also provides insight into how people allocate the scarce resources under uncertainty, which makes it a realistic case of decision-making with pressure rather than being driven by entertainment.

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

The preliminary exploratory data analysis provided a few important patterns that triggered further item analysis. The most striking finding was the loadout information which indicated a definite survival item preference with some items such as pots, axes, and ferro rods being used much more often than others. The data of the survivalists indicated that the results were highly dispersed, varying in the time of survival; the survival time was skewed with many people dropping out of the game earlier and some surviving continuously longer. This difference in item selection as well as in survival indicated that a particular combination of items could be associated with success and association rule mining was a natural next step. The M1 analysis also determined that the datasets made were relatively clean, but they were required to be standardized in terms of column names and inter-table joining to connect loadouts and survival outcomes. One step I decided to skip from M1 was using a heatmap as a visualization because when I made one it didn't seem to make sense.

M2 added one main data change beyond M1. It made a binary basket format from the loadout data. Each row is one contestant. Each column is one item type. The cells show 1 if that person picked the item. They show 0 if not. This changed the data from long format to wide format. The wide format is needed for the Apriori algorithm. The basket data was joined with outcome data. This added info like days_lasted, medical evacuation status, tap-out reason, gender, and age. Adding these lets us see which item patterns relate to survival success.

3.Analysis & Experiments (~1.5 page)

Method 1: Finding Item Patterns with Apriori

What this addresses: I wanted to find which combinations appear together most often. Not just "50% bring axes," but "40% bring axes AND ferro rods AND pots together."

Why this works: The Apriori algorithm fits perfectly. Each contestant picks exactly 10 items from about 27 options, creating a natural "shopping basket" scenario. The algorithm finds sets appearing together more than random chance predicts. Since contestants make strategic choices, I expected meaningful patterns like bow-and-arrow pairs or fishing-gear clusters.

Parameter experiments:

Low support (10%): Flooded me with meaningless coincidences. Finding two people with "gill net AND slingshot AND rations" doesn't reveal strategy, just random overlaps.

High support (30%): Too restrictive. I only found universal items everyone picks regardless of strategy.

Medium support (20%): The sweet spot. It filtered noise while preserving variety. With 10-item kits from 40+ options, 20% picking the same combo means something.

What happened: Medium support revealed interesting patterns. Basic tools (axes, ferro rods) appeared in tons of combinations. But even "obvious" pairs surprised me like fishing line and hooks had only 15% support. That is because not everyone fishes the same way. Some prefer gill nets and others substitute them for hunting Another surprise was that people rarely combined multiple cutting tools. High support for axes OR saws, but not both. This revealed substitution, contestants pick one cutting tool given only 10 item slots.

The itemset size problem: I expected lots of 3-item and 4-item patterns but mostly got pairs. As you add items, support drops fast. Even if axes appear 60% and pots 50%, the combination might only hit 35%. This taught me survival loadouts aren't rigid formulas and people customize based on skills.

Method 2: Discovering Dependencies with Association Rules

What this adds: Rules show "if A, then probably B" meaning directional information. If you bring a bow, you need arrows.

Key metrics:

Confidence (70%): Measures how often B appears when A is present. I set 70% minimum to avoid weak patterns.

Lift (>1.0): Compares the rule to chance. Lift = 1.0 means random co-occurrence. Lift > 1.0 means real relationship. I focused on lift above 1.2 for strongest dependencies.

Threshold testing: At 60% confidence, too many weak rules. At 80%, only ultra-obvious ones. I believe 70% shows clear patterns without the noise.

What worked: This revealed complementary tools like paracord with snare wire, bows with arrows, pots with ferro rods. I saw strategic specialization: some went all-in on fishing, others on hunting. Very few combined both since there are not enough item slots. Lift around 1.0-1.1 meant weak coincidences. Lift above 1.3 showed genuine dependencies.

The big failure: I tried outcome-based rules like "if axe AND pot, then days_lasted > 50." This failed because association rules handle categorical data, not continuous outcomes. With about 150 contestants, I lacked statistical power because survival depends on skill, location, and weather, not just gear. Massive confounding variables exist. I realized these methods answer "what goes with what," not "what causes success," and focused on pattern discovery instead.

Technical challenges: The algorithm generated redundant reciprocal rules, requiring manual deduplication. Interpreting lift was tricky. 1.5 sounds small but means items appear together 50% more than chance. Multi-item rules were rare because 3+ item sets had low support, suggesting custom kits rather than rigid formulas.

Surprises: I expected lift values of 2.0+ for obvious pairs like pot and ferro rod. Actual lifts were mostly 1.1-1.4. Even "logical" pairings aren't universal. Popular items (axes at 60%+) generated weak rules because they appear everywhere. The most interesting rules came from moderately popular items (20-40% support).

Key Takeaways

These methods excel at discovering "what goes with what" in gear patterns, revealing diverse, flexible strategies rather than rigid formulas. The sweet spot: medium support (20%) and moderate confidence (70%), balancing detail against noise. However, these methods can't predict success; they're descriptive, showing what people do, not necessarily what works best.

3. Findings & Interpretations (~0.75 page)

Three Key Discoveries

1. Universal Necessities vs. Strategic Choices

The data reveals a clear split. Nearly all contestants (~95%) bring fishing gear and pots which are survival fundamentals, not strategic choices. Sleeping bags (~85%), axes (~85%), and saws (~85%) complete the universal category. Below 80%, items reflect personal strategy: bow and arrows (~65%), paracord (~60%), rations (~60%). (Figure 1) The real strategic decisions happen in that final 30-40% of loadouts where contestants can differentiate based on strengths.

2. Predictable Clusters with Built-in Flexibility

Association rules revealed strong pairing patterns. The core survival kit—axe, pot, sleeping bag, trapping wire—appears together with 76% reliability and lift of 2.16 (more than twice random chance). Ferro rods pair with pots at 73% reliability. Hunting gear follows the same pattern at 71% reliability.

But even strong patterns max out around 75-80%. That remaining 20-25% represents genuine diversity—people adapt based on skills rather than following rigid formulas. For producers, this balance makes for interesting watching experiences: enough patterns to see strategies, enough variation to keep outcomes unpredictable.

3. Rare Items Signal Specialized Expertise

Items chosen by under 5% of contestants—sharpening stones, slingshots, scotch-eyed augers—represent specialized strategies. These aren't bad choices; they let people leverage unique skills. Success comes from matching gear to abilities, not copying templates.

Broader Implications

These patterns mirror resource-constrained decision-making everywhere. There are table-stakes requirements (pots, fishing gear) and differentiating choices (bows vs. gill nets). The diversity of successful approaches challenges "one-size-fits-all" thinking, suggesting multiple valid strategies exist when you match approach to strengths.

Limitation: These methods show what people choose and which items pair together—not what actually works. Association rules reveal correlation, not causation.

4. Next Steps (~0.25 page)

What Hasn't Been Done

I’ve identified which items contestants choose and how they pair together, but I haven't connected these patterns to survival outcomes. The analysis is descriptive—showing what people do, not whether it works. I also haven't examined how strategies vary by location or season.

Planned Analyses

Link items to outcomes: Use survival analysis or regression to determine which items or combinations actually extend survival time, controlling for age, experience, and location.

Segment by environment: Re-run association rules separately for arctic, temperate, and tropical locations to find environment-specific patterns.

Analyze failure modes: Cross-reference item combinations with tap-out reasons to identify which items prevent specific failures.

Unresolved Questions

Do rare items give advantages or represent mistakes? Is there an optimal loadout, or does success depend on matching gear to skills? Which items are overpacked based on fear rather than utility?

Justification

My findings show 20-25% variation even in strong patterns, suggesting room for optimization. Association rules revealed what people use. Survival analysis will show what people should use, enabling actionable recommendations beyond descriptive insights.

Appendix

Shared GitHub Repository (Required)

https://github.com/MikaIsmayilov/B02_Alone_TVShow_Unsupervised_Project

Supplemental Material (Highly Recommended)

Figure 1 – Top 20 Most Popular Items

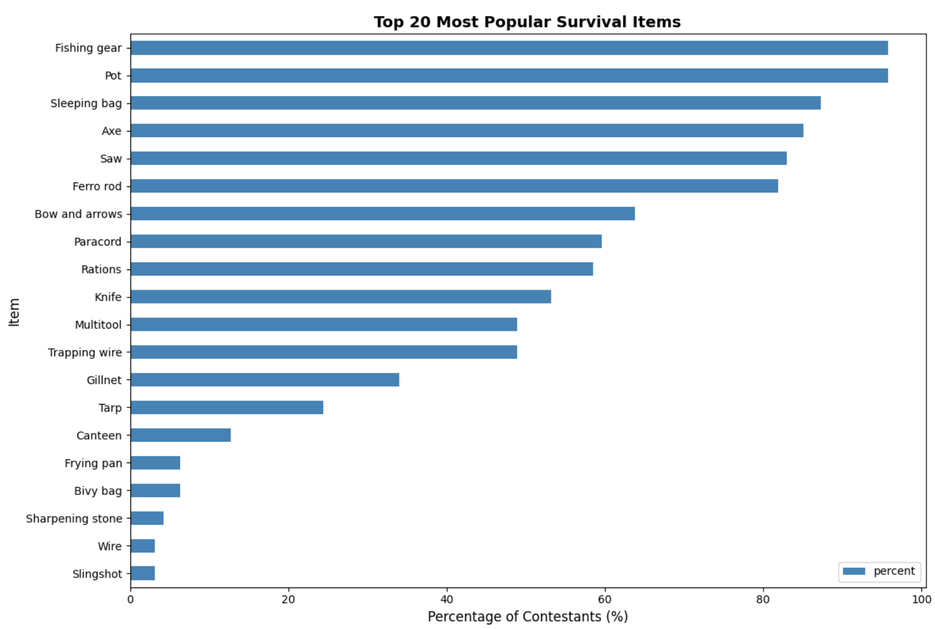


Figure 2 – Pairwise Plot for Association Rule Metrics

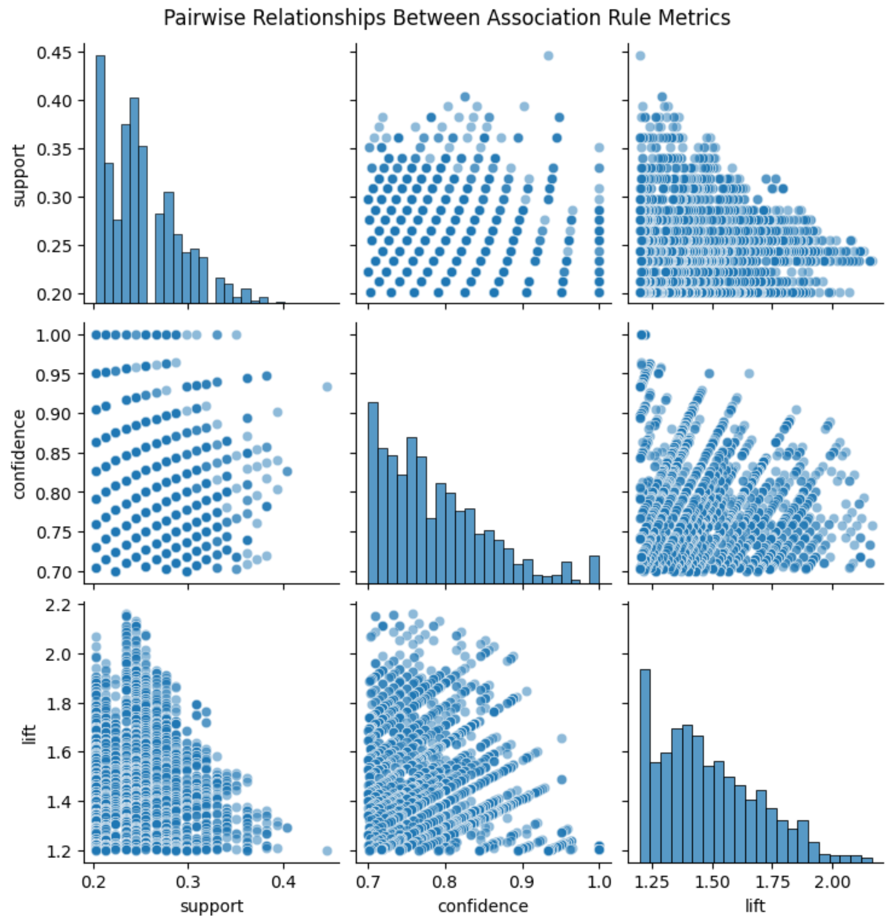
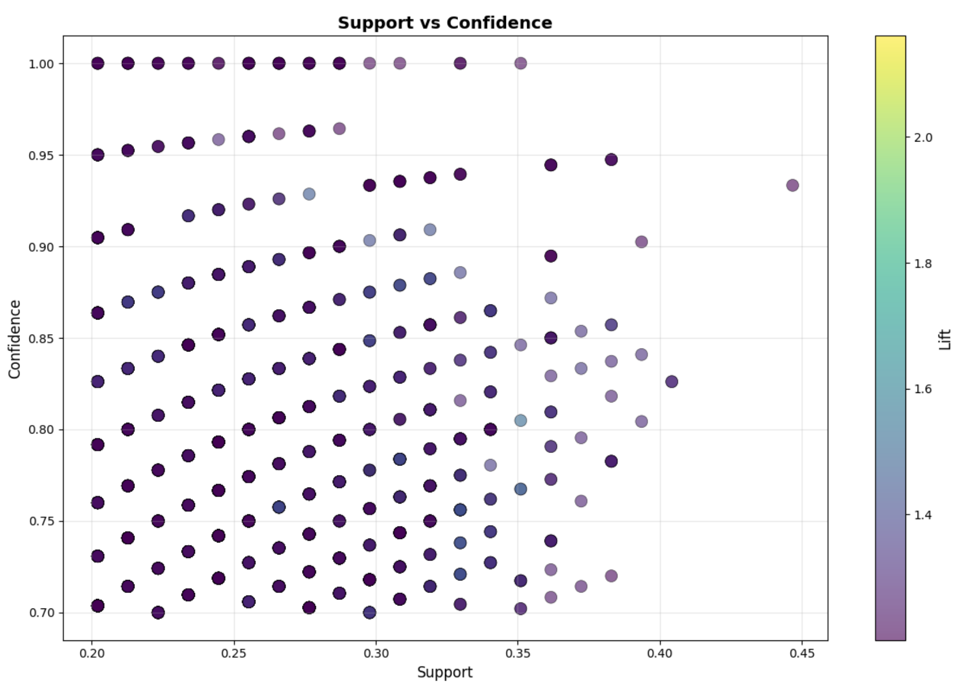


Figure 3 - Support vs Confidence Scatter Plot



Top 10 Most Popular Items:

	item	count	percent
0	Fishing gear	90	95.74
1	Pot	90	95.74
2	Sleeping bag	82	87.23
3	Axe	80	85.11
4	Saw	78	82.98
5	Ferro rod	77	81.91
6	Bow and arrows	60	63.83
7	Paracord	56	59.57
8	Rations	55	58.51
9	Knife	50	53.19

Bottom 10 Least Popular Items:

	item	count	percent
17	Sharpening stone	4	4.26
18	Wire	3	3.19
19	Slingshot	3	3.19
20	Ground sheet	2	2.13
21	Hammock	2	2.13
22	Salt	2	2.13
23	Soap	2	2.13
24	Rope	1	1.06
25	Scotch eyed auger	1	1.06
26	Shovel	1	1.06

Process Overview

Step 1 — Data Review / Refresher

Step 2 — Data Preprocessing

Step 3 — Feature Engineering

Step 4 — Apply Unsupervised ML

Step 5 — Evaluate & Interpret

Use of Generative AI Tools

For this step, I used ChatGPT for brainstorming, teaching me the unsupervised methods, as well as debugging my codes and giving me ideas for changing it. I also used the native Colab Gemini for code generation and additional debugging.

Association Rule and Apriori Chat: https://chatgpt.com/share/6989133e-4f2c-8002-987b-beac1cfb89b5

Strategy Brainstorming: https://chatgpt.com/share/698913bc-d8e8-8002-b3c4-c1a0bb3a397c

Step-by-step brainstorming for M2: https://chatgpt.com/share/6989142f-3268-8002-953a-fe4cd7b12250

Debugging and coding chat: https://chatgpt.com/share/698a6821-6f54-8002-a20e-854849599a25