

Tourism Pattern Discovery using Association Rule Mining

Case Study: Nepal Tourism Destination Co-Visitation

'AI and Tourism' – MIT-AI @ Gandaki University

Bidur Devkota, PhD
GCES Pokhara

December 24, 2025

© 2025 Bidur Devkota, PhD. CC BY-NC.

Lab Objectives

After completing this lab, you will be able to:

- Explain **association rule mining** in a tourism context.
- Compute and interpret **support**, **confidence**, and **lift**.
- Apply **Apriori** to mine frequent itemsets and generate rules.
- Filter meaningful rules for decision-making.
- Translate mined rules into **tourism marketing & package** insights.

© 2025 Bidur Devkota, PhD. CC BY-NC.

Problem Context: Tourism Co-Visitation Patterns

- Tourists often visit **multiple destinations** in one trip.
- We want patterns like:

Pokhara → Mustang
- Such patterns help with:
 - Multi-destination packages
 - Cross-selling recommendations
 - Transport & infrastructure planning

© 2025 Bidur Devkota, PhD. CC BY-NC.

Association Rule Mining (ARM)

Association Rule Mining discovers relationships among items in transaction data.

Rule form:

$$A \rightarrow B$$

- A = antecedent (if part), B = consequent (then part)
- In tourism: A and B are **destinations**
- Goal: find rules that are **frequent** and **useful**

© 2025 Bidur Devkota, PhD. CC BY-NC.

Key Metric 1: Support

Support measures how frequently an itemset occurs.

$$\text{support}(X) = \frac{\#\{\text{transactions containing } X\}}{\#\{\text{total transactions}\}}$$

Example: If 150 out of 1000 travelers visited both Pokhara and Mustang:

$$\text{support}(\{\text{Pokhara, Mustang}\}) = \frac{150}{1000} = 0.15$$

Interpretation

Support is the **popularity** of a destination combination.

Key Metric 2: Confidence

Confidence measures how reliable the inference is.

$$\text{confidence}(A \rightarrow B) = \frac{\text{support}(A \cup B)}{\text{support}(A)}$$

Example: If $\text{support}(\text{Pokhara}) = 0.40$ and $\text{support}(\text{Pokhara, Mustang}) = 0.15$:

$$\text{confidence}(\text{Pokhara} \rightarrow \text{Mustang}) = \frac{0.15}{0.40} = 0.375$$

Interpretation

37.5% of Pokhara visitors also visit Mustang.

Key Metric 3: Lift

Lift compares observed co-occurrence against chance.

$$\text{lift}(A \rightarrow B) = \frac{\text{confidence}(A \rightarrow B)}{\text{support}(B)} = \frac{\text{support}(A \cup B)}{\text{support}(A) \cdot \text{support}(B)}$$

- **Lift** > 1: positive association (together more than expected)
- **Lift** = 1: independent
- **Lift** < 1: negative association

Rule Strength

Higher lift \Rightarrow stronger destination linkage.

Apriori Algorithm (Idea)

- Apriori finds frequent itemsets using the **downward closure property**:
If an itemset is infrequent, all of its supersets are also infrequent.
- This reduces the number of candidates dramatically.

Why Apriori for this lab?

Easy to implement, interpretable, and suitable for basket-style tourism data.

Dataset (Transaction View)

Raw table contains:

- **TravelerID**: unique traveler
- **Destination**: destination visited

We convert it into transactions:

$$T_i = \{\text{all destinations visited by traveler } i\}$$

Example

Traveler 101: {Kathmandu, Pokhara, Mustang}

Traveler 102: {Kathmandu, Lumbini}

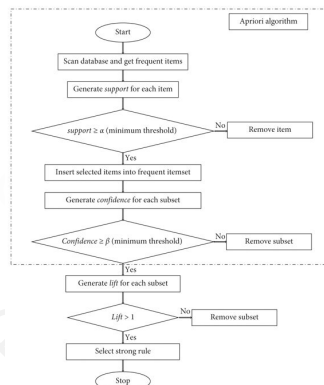
Data Preparation Steps

- 1 Remove rows with missing **TravelerID** or **Destination**
- 2 Standardize destination names (trim, remove trailing commas)
- 3 Group by **TravelerID** to create baskets
- 4 One-hot encode baskets (TransactionEncoder)

Note

Preprocessing quality strongly affects mined rules.

Implementation Pipeline (Lab Workflow)



Destination Popularity Segmentation

Destinations can be categorized as:

- **High-growth markets**: Top destinations by traveler count
- **Stable markets**: Remaining destinations

Why this matters

Rules can be interpreted differently for:

- Growth → Stable (gateway upsell)
- Stable → Growth (market expansion)

Apriori Settings Used (Example)

Typical parameters in this lab:

- Minimum support: **0.03** (fallback **0.02** if no itemsets)
- Max itemset length: **3**
- Minimum lift for rule generation: **0.8**

Tuning Tip

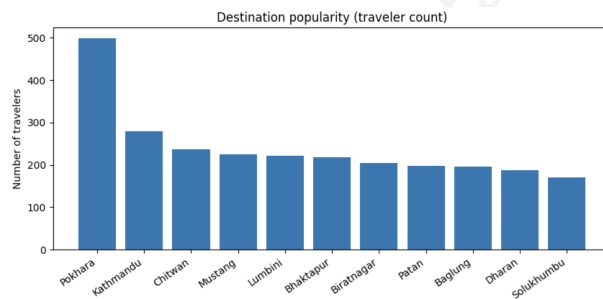
Lower support finds more rules, but can produce noisy/rare patterns.

Traveller Data

TravelerID Destination

TR_0001	Lumbini
TR_0001	Pokhara
TR_0001	Chitwan
TR_0002	Baglung
TR_0002	Dharan

Destination Popularity



Regional Pair Frequency

Destination A Destination B Both count Support

Pokhara	Mustang	113	0.113
Pokhara	Chitwan	111	0.111
Kathmandu	Patan	66	0.066
Kathmandu	Bhaktapur	52	0.052
Lumbini	Chitwan	46	0.046

Example Rule & Interpretation

Rule:

Pokhara → Mustang

- Support: 0.150
- Confidence: 0.375
- Lift: 2.50

Interpretation

Pokhara and Mustang are co-visited **2.5× more than expected by chance**, suggesting a strong travel corridor.

© 2025 Bidur Devkota, PhD. CC BY-NC.

Meaningful Rules

About one-third of travelers who visit Patan also visit Kathmandu, and this happens about 19 percent more often than expected by chance, making it a meaningful but moderate tourism association.

- 6.6 percent of travelers visited both Patan and Kathmandu
- 33 percent of Patan visitors also visited Kathmandu
- Lift > 1: positive association (19% above chance)
- Representativity = 1.0, indicating the rule is fully representative of the dataset

antecedents	consequents	antecedent support	consequent support	support	confidence	lift	representativity
(Patan)	(Kathmandu)	0.198	0.279	0.066	0.333333	1.194743	1.0
(Mustang)	(Pokhara)	0.224	0.499	0.113	0.504464	1.010950	1.0
(Chitwan)	(Pokhara)	0.236	0.499	0.111	0.470339	0.942563	1.0
(Lumbini)	(Pokhara)	0.221	0.499	0.094	0.425339	0.852384	1.0
(Kathmandu)	(Pokhara)	0.279	0.499	0.117	0.419355	0.840390	1.0

© 2025 Bidur Devkota, PhD. CC BY-NC.

Meaningful Rules

Business Insights:

- **Pokhara is a major tourism hub:** Appears as the *consequent* in **4 out of 5** filtered association rules.
- **Patan → Kathmandu shows the strongest association:** This rule has the **highest lift value**, indicating the strongest relationship beyond chance.
- **All filtered rules involve major destinations:** No weak or non-significant destinations appear in the final rule set.
- **Confidence values are moderate:** Ranging from approximately **33% to 50%**, indicating reasonable conversion likelihood.

antecedents	consequents	antecedent support	consequent support	support	confidence	lift	representativity
(Patan)	(Kathmandu)	0.198	0.279	0.066	0.333333	1.194743	1.0
(Mustang)	(Pokhara)	0.224	0.499	0.113	0.504464	1.010950	1.0
(Chitwan)	(Pokhara)	0.236	0.499	0.111	0.470339	0.942563	1.0
(Lumbini)	(Pokhara)	0.221	0.499	0.094	0.425339	0.852384	1.0

© 2025 Bidur Devkota, PhD. CC BY-NC.

Meaningful Rules

Strategic Recommendations Based on Filtered Rules:

- **Priority 1:** Develop **Patan–Kathmandu combined tour packages** (highest lift and strongest association).
- **Priority 2:** Create **Mustang–Pokhara adventure packages** (positive lift with the highest confidence).
- **Priority 3:** Design **cross-promotional campaigns** for other routes leading to Pokhara.
- **Strategic consideration:** For rules with **lift < 1**, focus on improving connections through better transport integration or redesigned travel packages.

antecedents	consequents	antecedent support	consequent support	support	confidence	lift	representativity
(Patan)	(Kathmandu)	0.198	0.279	0.066	0.333333	1.194743	1.0
(Mustang)	(Pokhara)	0.224	0.499	0.113	0.504464	1.010950	1.0
(Chitwan)	(Pokhara)	0.236	0.499	0.111	0.470339	0.942563	1.0
(Lumbini)	(Pokhara)	0.221	0.499	0.094	0.425339	0.852384	1.0
(Kathmandu)	(Pokhara)	0.279	0.499	0.117	0.419355	0.840390	1.0

© 2025 Bidur Devkota, PhD. CC BY-NC.

Filtering Meaningful Rules (Business Thresholds)

Example criteria to keep useful rules:

- Support ≥ 0.025
- Confidence ≥ 0.30
- Lift ≥ 1.0

Why filter?

To remove weak rules and focus on actionable tourism patterns.

antecedents	consequents	antecedent support	consequent support	support	confidence	lift	representativity
(Patan)	(Kathmandu)	0.198	0.279	0.066	0.333333	1.194743	1.0
(Mustang)	(Pokhara)	0.224	0.499	0.113	0.504464	1.010950	1.0

© 2025 Bidur Devkota, PhD. CC BY-NC.

Insight Categories (Tourism Interpretation)

- 1 **Emerging Market Corridors**: growth \leftrightarrow growth
- 2 **Gateway Patterns**: growth \rightarrow stable
- 3 **Market Expansion**: stable \rightarrow growth
- 4 **Regional Focus Pairs**: strong local circuit pairs

© 2025 Bidur Devkota, PhD. CC BY-NC.

Tourism Insight

The following insights are derived from filtered, high-confidence association rules.

Emerging Market Corridors:

- Travelers who visit ['Mustang'] are 50% likely to also visit ['Pokhara'] (Lift=1.01)

Market Expansion (Stable \rightarrow Growth):

- ['Patan'] \rightarrow ['Kathmandu'] (Confidence=33%, Lift=1.19)

Regional Focus Pairs:

- Patan \rightarrow Kathmandu: Confidence=33%, Lift=1.19
- Mustang \rightarrow Pokhara: Confidence=50%, Lift=1.01
- Chitwan \rightarrow Pokhara: Confidence=47%, Lift=0.94
- Lumbini \rightarrow Pokhara: Confidence=43%, Lift=0.85
- Kathmandu \rightarrow Pokhara: Confidence=42%, Lift=0.84

Figure: 2025 Tourism Insight

© 2025 Bidur Devkota, PhD. CC BY-NC.

Strategic Recommendations (From Rules)

- Create **multi-destination packages** from high-confidence pairs.
- Run **cross-selling campaigns**: recommend *B* when traveler searches *A* (high lift).
- Improve **transport links** between frequent co-visited destinations.
- Segment marketing: promote **stable markets** to visitors of **high-growth markets**.
- Design **regional circuits** using high-support pairs.

© 2025 Bidur Devkota, PhD. CC BY-NC.

Lab Tasks

- 1 Run Apriori with multiple **min_support** values.
- 2 Generate rules and compute **support, confidence, lift**.
- 3 Filter rules using thresholds and interpret top 5 rules.
- 4 Propose one tourism package based on mined rules.

© 2025 Bidur Devkota, PhD. CC BY-NC.

Discussion Questions

- Why can a rule have high confidence but low lift?
- What happens when **support** is set too low?
- Do association rules imply **causation**? Why or why not?
- How can ARM support Nepal tourism policy and marketing decisions?

© 2025 Bidur Devkota, PhD. CC BY-NC.

Conclusion

- Association rule mining reveals **co-visitation patterns** in tourism data.
- Apriori efficiently finds frequent destination combinations.
- Support, confidence, and lift help identify **actionable rules**.
- Insights guide packages, marketing, and infrastructure planning for Nepal tourism.

© 2025 Bidur Devkota, PhD. CC BY-NC.