

Gift Bundle Discovery

Association Rule Mining with FP-Growth Algorithm

Kristjan Caius Kasuk • Introduction to Data Science • Project ID: C12

Motivation & Methodology

Goal

Identify which gift products are frequently purchased together to inform bundle recommendations.

Dataset

- Online Retail II dataset (UCI Repository)¹
- 1M+ line items → 36,422 transactions (UK only)
- Analyzed top 200 gift products, 26,360 multi-item baskets

FP-Growth Algorithm

Why FP-Growth? Apriori failed due to memory constraints with large itemsets. FP-Growth uses a compact tree structure to find frequent patterns without candidate generation, allowing efficient discovery.

- Min support: 0.5%
- Min confidence: 30%
- Min lift: >1.5 (50% stronger than random)

Threshold Tuning

- Initial thresholds (lift>3, conf>0.5) revealed only color-matching sets.
- Tuned both to lift>1.5 and conf>0.3 to discover cross-category associations.

Diversity-Focused Curation

- Filtered non-gift utility items
- Removed symmetrical pairs ($A \rightarrow B = B \rightarrow A$)
- Selected top 10: 7 top + 3 top cross-category bundles
- Span 6 categories: Food, Kids, Home, Clothing, Storage, Games

Results & Conclusions

Two Distinct Customer Purchase Behaviors:

1. Matching Sets (7 of 10 bundles)

Coordinated color collections with lift 10-18. Customers buying matching items together (e.g., Pink + Green Teacups, Owl + Rabbit Cushions).

2. Cross-Category Bundles [U] (3 of 10)

Themed combinations with lift 8-12. Customers creating gift themes across product types (e.g., Christmas Napkins + Cake Cases for holiday baking).

Key Insights:

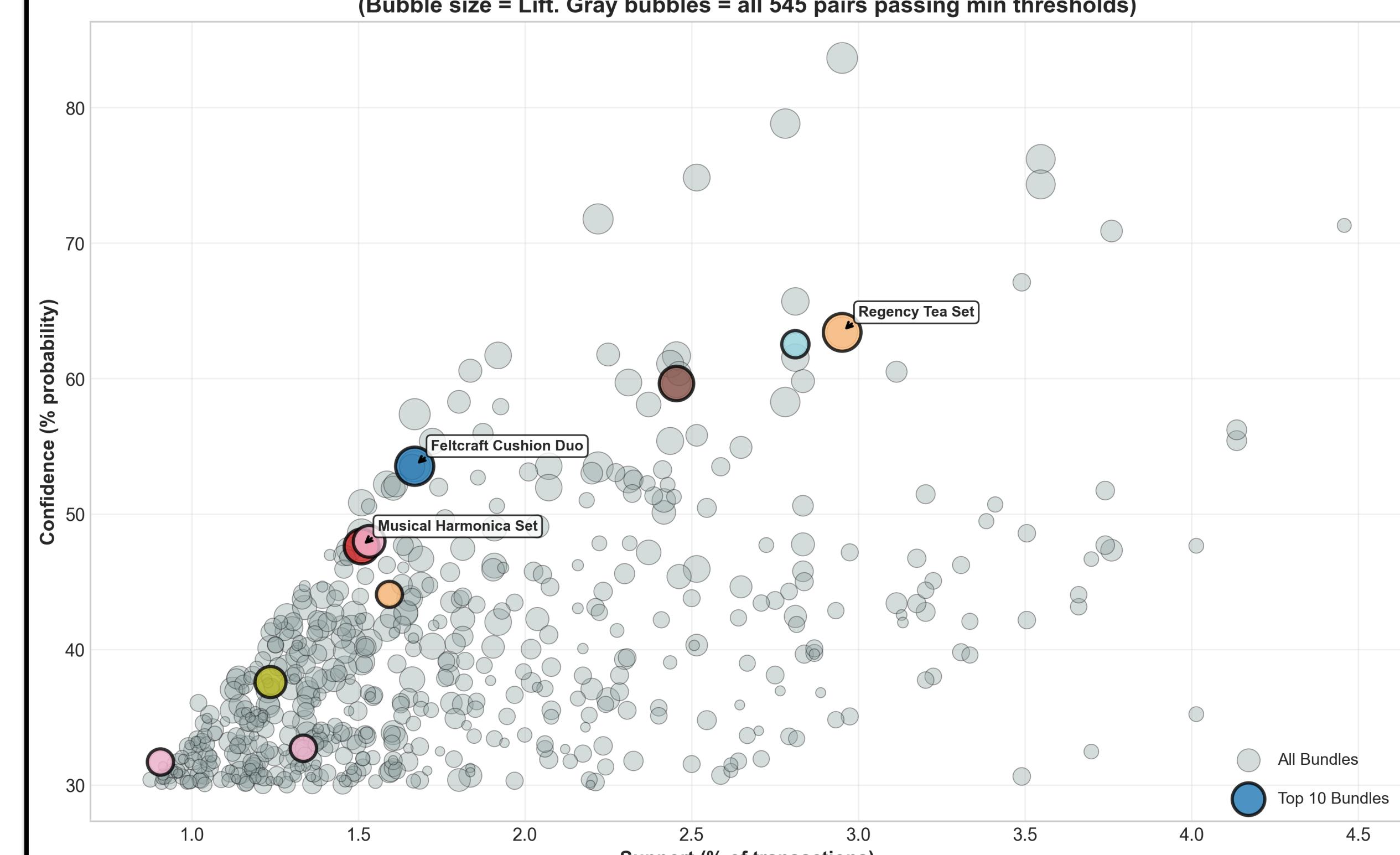
- Threshold tuning critical: Lowering lift 3.0→1.5 revealed many cross-category patterns that higher thresholds completely missed
- A total of 545 strong potential bundles found.
- Larger bundles (3+ products) were rare in the data and had lower confidence. Two item pairs provide clearer, more actionable insights.
- Business value: Matching sets for predictable upsells (high lift), cross-category for creative displays



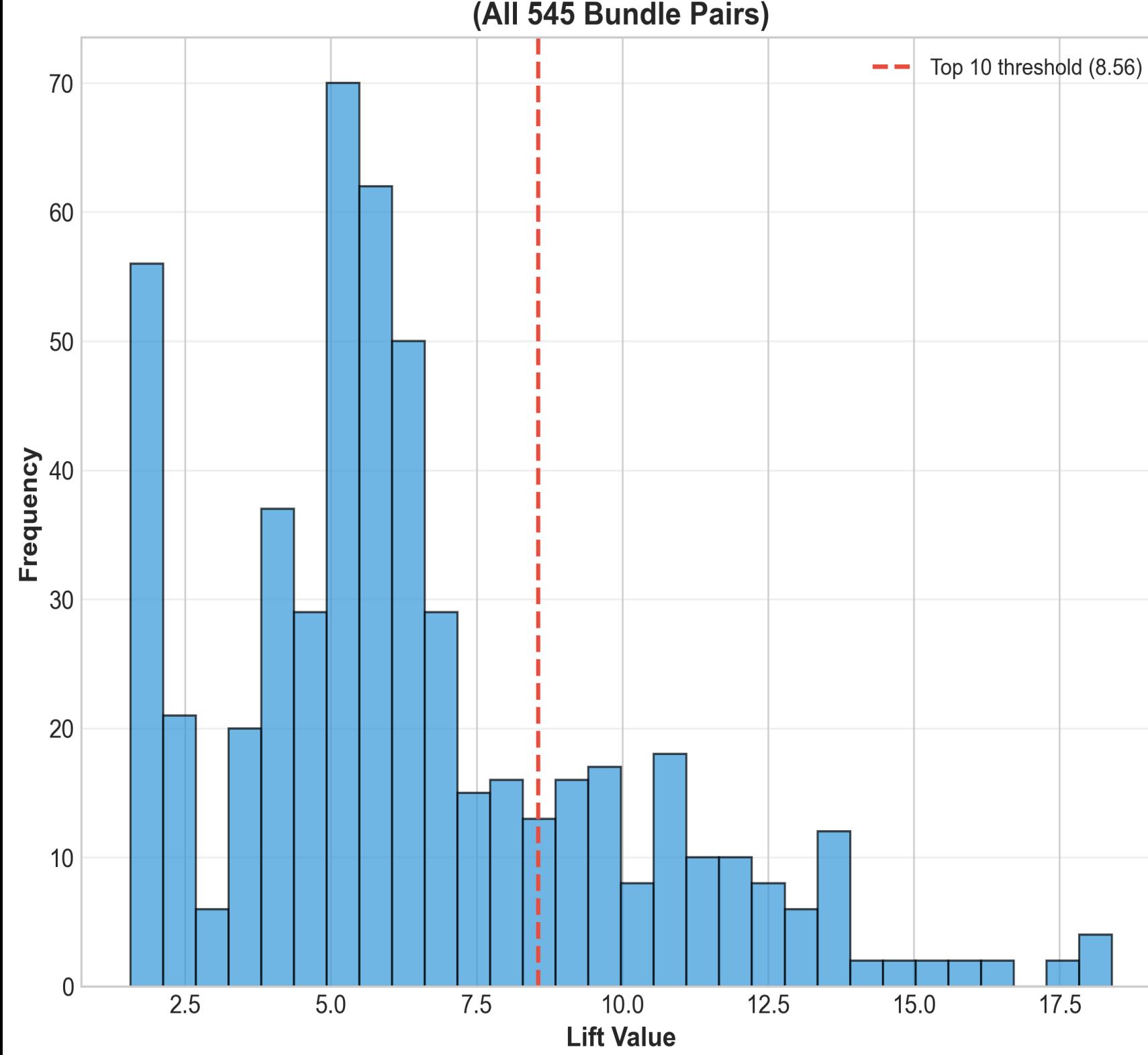
Top 10 Gift Bundle Details

#	Bundle Name	Product 1		Product 2	Lift	Conf.
1	Feltcraft Cushion Duo	FELTCRAFT CUSHION OWL	+	FELTCRAFT CUSHION RABBIT	18.40	54%
2	Regency Tea Set	GREEN REGENCY TEACUP AND SAUCER	+	PINK REGENCY TEACUP AND SAUCER	17.97	63%
3	Musical Harmonica Set	RED HARMONICA IN BOX	+	BLUE HARMONICA IN BOX	15.32	48%
4	Character Lunch Box Pair	DOLLY GIRL LUNCH BOX	+	SPACEBOY LUNCH BOX	14.98	60%
5	Novelty Plasters Duo	PLASTERS IN TIN CIRCUS PARADE	+	PLASTERS IN TIN SPACEBOY	12.70	48%
6	Snack Box Collection	ROUND SNACK BOXES SET OF 4 WOODLAND	+	ROUND SNACK BOXES SET OF 4 FRUITS	12.19	38%
7	Vintage Card Game Bundle	VINTAGE HEADS AND TAILS CARD GAME	+	VINTAGE SNAP CARDS	9.44	63%
8	Cozy Hot Water Bottle Set [U]	RED WOOLLY HOTTE WHITE HEART	+	KNITTED UNION FLAG HOT WATER BOTTLE	8.78	33%
9	London-Themed Gift Duo [U]	PACK OF 12 LONDON TISSUES	+	LUNCH BOX I LOVE LONDON	8.75	32%
10	Festive Baking Bundle [U]	60 CAKE CASES VINTAGE CHRISTMAS	+	PAPER CHAIN KIT VINTAGE CHRISTMAS	8.56	44%

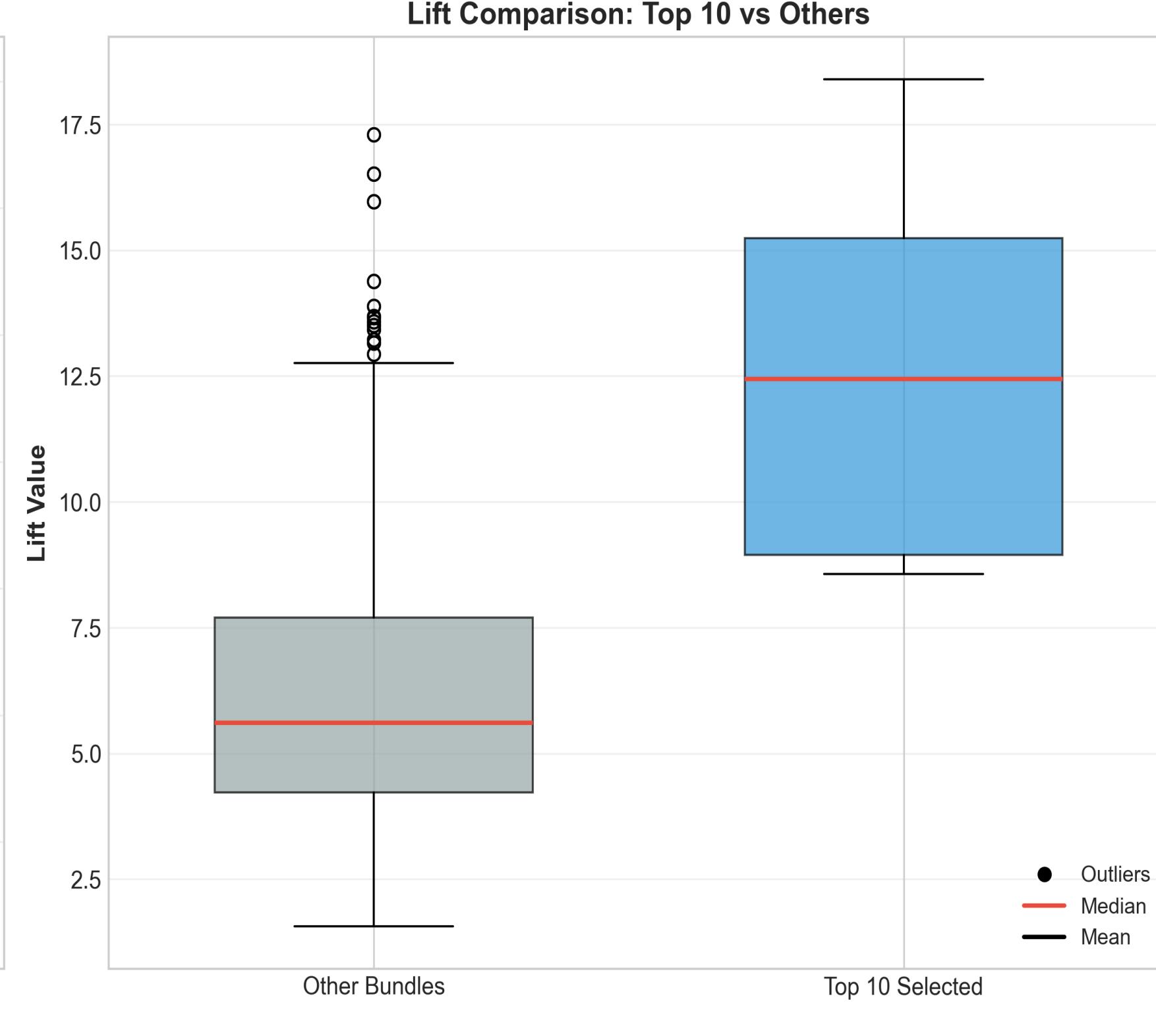
Bundle Quality: Support vs Confidence
(Bubble size = Lift. Gray bubbles = all 545 pairs passing min thresholds)



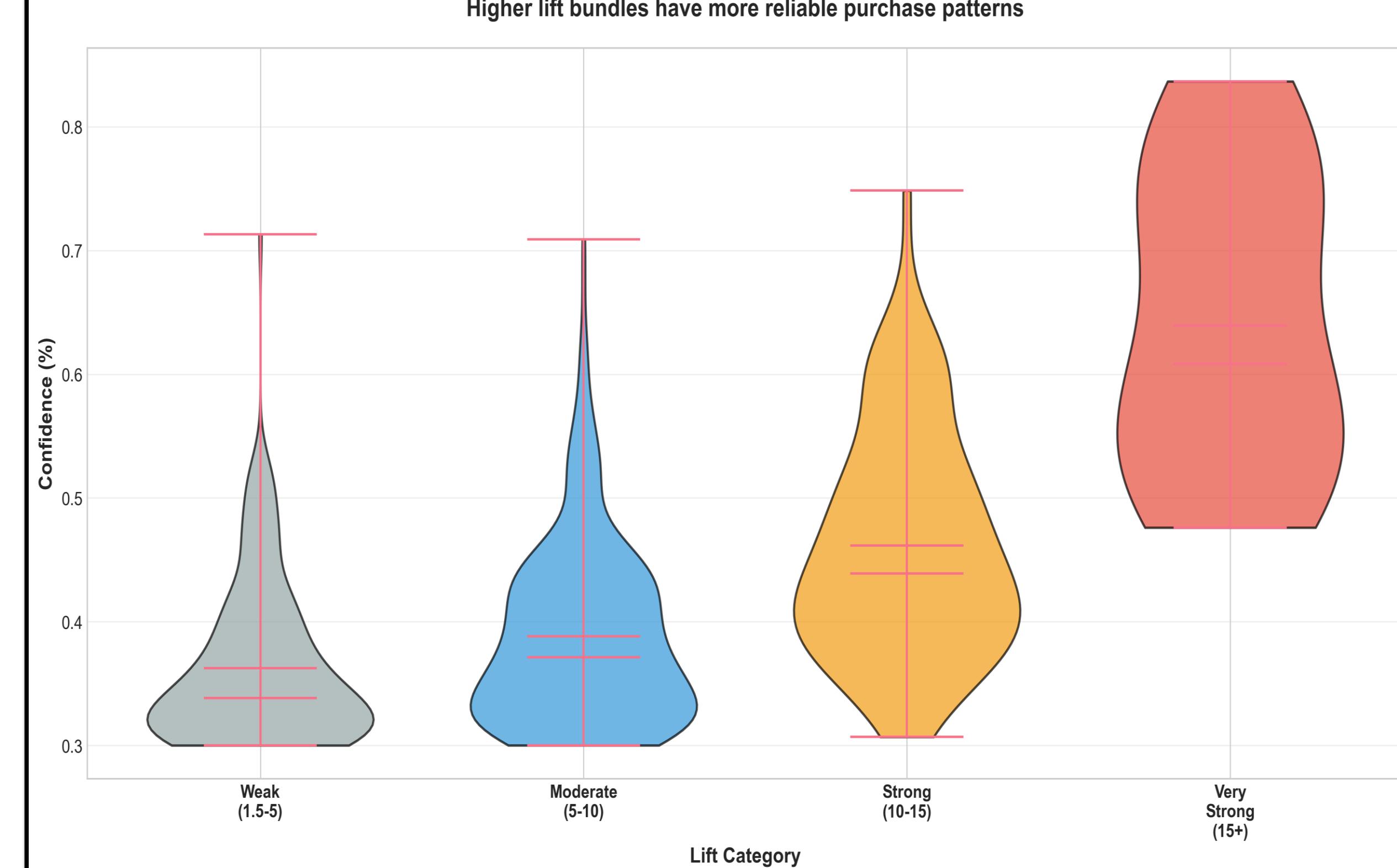
Distribution of Association Strength
(All 545 Bundle Pairs)



Lift Comparison: Top 10 vs Others



Pattern Quality by Strength
Higher lift bundles have more reliable purchase patterns



¹Chen, D. (2019). Online Retail II. UCI Machine Learning Repository. <https://doi.org/10.24432/C5CG6D>