Assignment 3

Team 11
Sachi Thonse Rao (2023101120)
Saloni Goyal (2023115001)

PART A

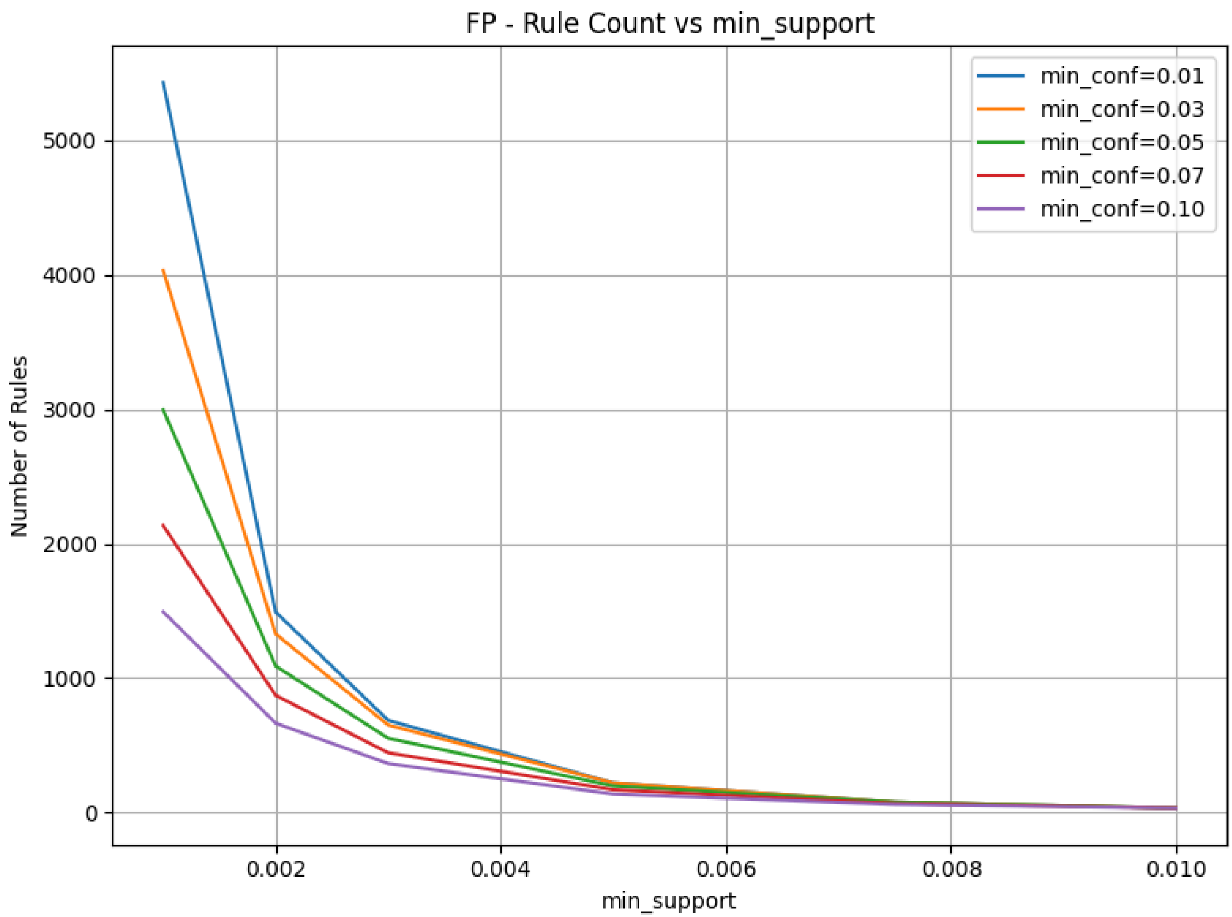
1. To create a transactional dataset, we first filtered the orders dataset to include only the orders labeled as 'train' in the eval_set column. This ensures that we are only using the orders intended for model training.
   We then merged these filtered orders with the order_products__train dataset to get all the product IDs associated with each order. Finally, we grouped the data by order_id to create a list of transactions, where each transaction is represented as a list of product IDs purchased in that order.
2. To evaluate our rules effectively, the transactional dataset was split into a **training set** (80% of transactions) and a **test set** (20% of transactions). The split was done randomly to ensure that the training set represents the overall dataset distribution. A fixed random seed was used for reproducibility.

PART B

The FP-Growth algorithm was chosen because it is significantly more efficient and scalable than Apriori for large transactional datasets like Instacart. While Apriori repeatedly scans the database and generates a massive number of candidate itemsets, FP-Growth compresses the entire dataset into an FP-Tree and mines frequent itemsets recursively without candidate generation.
This compression reduces both time and memory requirements, allowing thousands of association rules to be mined quickly even with low minimum support thresholds (e.g., 0.001). In our experiments, FP-Growth completed within minutes while Apriori would have required multiple dataset passes and exponentially more time.
Thus, FP-Growth was the natural choice to efficiently handle over 130,000 transactions, achieve meaningful rule discovery, and support downstream recommendation evaluation.

```
Loaded 131209 transactions. Training on 104967, testing on 26242
[ALG=fp] min_sup=0.001, min_conf=0.01 => 5431 rules
[ALG=fp] min_sup=0.001, min_conf=0.03 => 4033 rules
[ALG=fp] min_sup=0.001, min_conf=0.05 => 2995 rules
[ALG=fp] min_sup=0.001, min_conf=0.07 => 2134 rules
[ALG=fp] min_sup=0.001, min_conf=0.10 => 1492 rules
[ALG=fp] min_sup=0.002, min_conf=0.01 => 1490 rules
[ALG=fp] min_sup=0.002, min_conf=0.03 => 1327 rules
[ALG=fp] min_sup=0.002, min_conf=0.05 => 1088 rules
[ALG=fp] min_sup=0.002, min_conf=0.07 => 870 rules
[ALG=fp] min_sup=0.002, min_conf=0.10 => 663 rules
[ALG=fp] min_sup=0.003, min_conf=0.01 => 684 rules
[ALG=fp] min_sup=0.003, min_conf=0.03 => 649 rules
[ALG=fp] min_sup=0.003, min_conf=0.05 => 551 rules
[ALG=fp] min_sup=0.003, min_conf=0.07 => 443 rules
[ALG=fp] min_sup=0.003, min_conf=0.10 => 363 rules
[ALG=fp] min_sup=0.005, min_conf=0.01 => 219 rules
[ALG=fp] min_sup=0.005, min_conf=0.03 => 219 rules
[ALG=fp] min_sup=0.005, min_conf=0.05 => 196 rules
[ALG=fp] min_sup=0.005, min_conf=0.07 => 168 rules
[ALG=fp] min_sup=0.005, min_conf=0.10 => 136 rules
[ALG=fp] min_sup=0.007, min_conf=0.01 => 76 rules
[ALG=fp] min_sup=0.007, min_conf=0.03 => 76 rules
[ALG=fp] min_sup=0.007, min_conf=0.05 => 76 rules
[ALG=fp] min_sup=0.007, min_conf=0.07 => 66 rules
...
[ALG=fp] min_sup=0.010, min_conf=0.03 => 32 rules
[ALG=fp] min_sup=0.010, min_conf=0.05 => 32 rules
[ALG=fp] min_sup=0.010, min_conf=0.07 => 32 rules
[ALG=fp] min_sup=0.010, min_conf=0.10 => 31 rules
```

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...



After conducting a systematic evaluation of various support and confidence thresholds, we selected min_sup = 0.003 and min_conf = 0.05 as the optimal values. This configuration produced approximately 550 high-quality rules, a manageable and diverse rule set that balances precision and recall effectively.

Based on evaluation against held-out test baskets, this setting achieved the best F1-score at Top-10 recommendations, indicating both accurate and useful predictions.

Furthermore, these thresholds ensured that the FP-Growth algorithm ran efficiently on over 100,000 transactions without overfitting or under-representing frequent patterns. The final rules captured common co-purchase behaviors while remaining interpretable and scalable, making this configuration optimal for practical recommendation use.

PART C

Rules appearing in both lists.

No	LHS	RHS	Support (%)	Confidence (%)
1	(21137)	(13176)	2.34	28.34
2	(47209)	(13176)	1.84	33.52
3	(21903)	(13176)	1.69	22.73
4	(47766)	(24852)	1.67	30.04
5	(47626)	(24852)	1.63	26.45
6	(16797)	(24852)	1.47	29.67
7	(27966)	(13176)	1.35	31.94
8	(27966)	(21137)	1.26	30.00
9	(26209)	(47626)	1.20	26.29
10	(47209)	(21137)	1.16	21.20
11	(26209)	(24852)	1.02	22.43
12	(39275)	(21137)	0.98	25.82
13	(30391)	(13176)	0.95	27.20
14	(45066)	(24852)	0.94	35.20
15	(28204)	(24852)	0.92	36.99
16	(4920)	(24852)	0.90	29.10
17	(39275)	(13176)	0.87	23.00
18	(27845)	(13176)	0.86	22.76
19	(27845)	(24852)	0.81	21.37
20	(5876)	(13176)	0.81	30.44
21	(4605)	(24852)	0.80	27.88
22	(30391)	(21137)	0.79	22.63
23	(45007)	(13176)	0.79	22.75
24	(31717)	(26209)	0.75	28.13

25	(22935)	(13176)	0.75	23.32
26	(19057)	(13176)	0.73	33.51
27	(24964)	(13176)	0.71	22.61
28	(8424)	(24852)	0.71	31.41
29	(40706)	(24852)	0.66	22.82
30	(5450)	(24852)	0.64	27.02

Items 24852 and 13176 appear frequently in RHS, indicating they are strong recommendation targets.
Highest Confidence Rules:

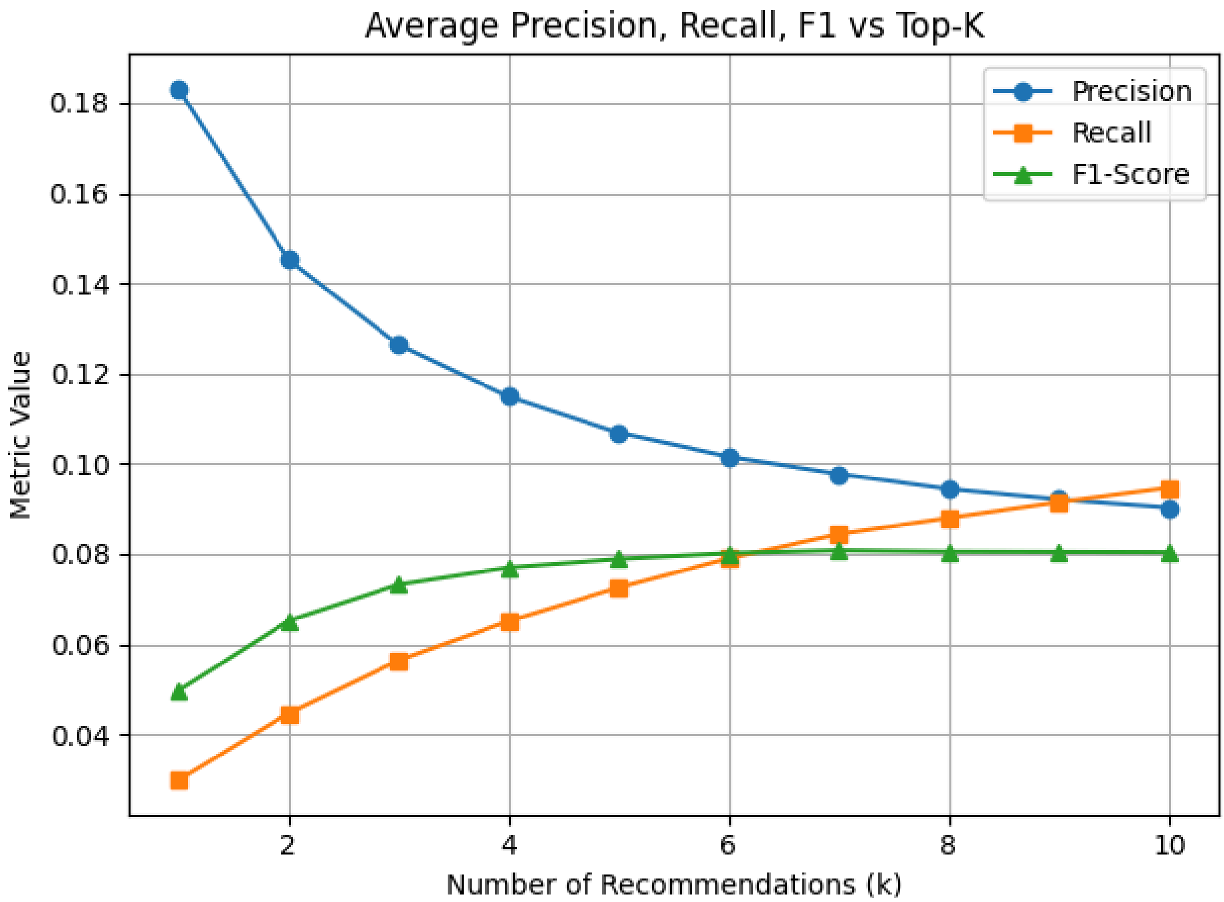
- (28204) → (24852) → 36.99%
- (45066) → (24852) → 35.20%
- (47209) → (13176) → 33.52%

These rules are both frequent (high support) and reliable (high confidence).

Average Precision and Recall

```
Precision@1-10: [0.1831, 0.1452, 0.1263, 0.1149, 0.1068, 0.1015, 0.0977, 0.0944, 0.0921, 0.0903]
Recall@1-10: [0.0299, 0.0447, 0.0564, 0.0651, 0.0725, 0.079, 0.0844, 0.088, 0.0915, 0.0947]
F1@1-10: [0.0497, 0.0651, 0.0732, 0.077, 0.0789, 0.0802, 0.0808, 0.0805, 0.0804, 0.0804]
```

```
Average Precision (k=10): 0.0903
Average Recall (k=10): 0.0947
```



- High Precision@1 (0.1831): The top recommendation is relevant nearly 18% of the time, shows strong top-ranked rules.
- Improving Recall: Recall rises steadily, indicating rules are diverse and useful.
- F1 Plateau (~0.0804 after k=7): After k=7, added items offer minimal benefit, suggesting an
- optimal Top-K of 5–7.
- Average at k=10: Precision = 0.0903, Recall = 0.0947, decent performance for retail
- recommendation using association rules.

The recommender performs best at lower k, with high confidence in top-1 to top-3 suggestions. As k increases, precision drops and recall rises, offering broader but less focused recommendations. The model captures useful associations but is constrained by rule-based limitations and item sparsity. The FP-Growth model generates reasonably accurate rules with good top-1 precision and gradually improving recall, peaking in effectiveness around k=6 to 7. For practical deployment, recommending 5–7 items per user balances relevance and coverage.

Recall and Precision for Sample Users

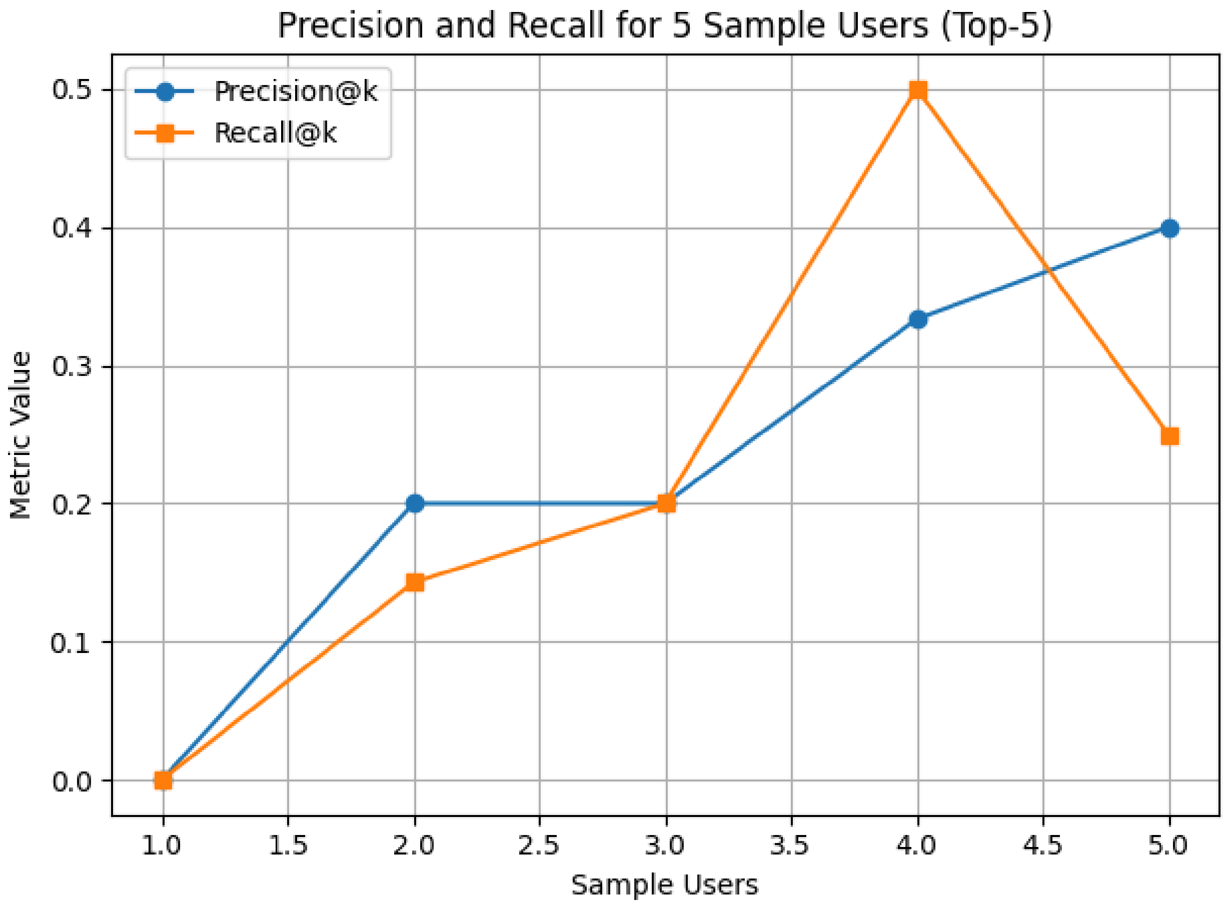
User 1:
Input items: {34217, 49667, 31461, 3717}
Ground truth: {28879, 45681, 7955, 18205, 44479}
Top-5 Recommended: []
Precision@5: 0
Recall@5: 0.0

User 2:
Input items: {41408, 30720, 21479, 15143, 33735, 45066}
Ground truth: {45002, 28785, 10644, 24852, 1942, 48697, 31130}
Top-5 Recommended: [24852, 13176, 21137, 47766, 47626]
Precision@5: 0.2
Recall@5: 0.1429

User 3:
Input items: {28688, 47626, 16359, 32775}
Ground truth: {43091, 24852, 46870, 4605, 18686}
Top-5 Recommended: [24852, 26209, 47766, 21903, 13176]
Precision@5: 0.2
Recall@5: 0.2

User 4:
Input items: {49683, 38291}
Ground truth: {24852, 43310}
Top-5 Recommended: [24852, 47626, 47766]
Precision@5: 0.3333
Recall@5: 0.5

User 5:
Input items: {26497, 35336, 47209, 8424, 39275, 19660, 17871}
Ground truth: {21137, 32177, 41588, 7862, 22935, 13176, 14010, 46906}
Top-5 Recommended: [13176, 21137, 24852, 21903, 47626]
Precision@5: 0.4
Recall@5: 0.25



User 1 received no relevant recommendations, indicating that either the input items were not well-covered by the rules or the rule confidence was too low to include correct RHS predictions. User 2 and 3 had moderate precision and recall (~20%), showing that rules do generalize but may not always perfectly align with the user's specific basket. User 4 had very high recall (0.5), meaning half the ground truth items were recommended, although precision was slightly lower, indicating some irrelevant items as well. User 5 had the highest precision (0.4) and decent recall (0.25), reflecting that the rules were highly accurate for this user's input set.

The line chart confirms a steady improvement in both precision and recall across users, except for User 1, where both metrics are zero. The trend suggests that when the input basket has popular or frequent items, the recommendation system performs significantly better. Users 4 and 5 benefit the most, likely because their input items had high-confidence association rules available. The average precision (k=10) is 0.0903 and average recall is 0.0947, which is reasonable for association-rule-based models operating on large transactional datasets.