

Amazon Product Bundling & Customer Recommendation

ISYE 7406 Project Group 115



Ashish Puri

apuri61@gatech.edu



Jagannath Banerjee

jbanerjee7@gatech.edu



Piyush Shivrain

pshivrain3@gatech.edu

Amazon Product Bundling and Recommendation

Objective

Conduct comprehensive analysis of Amazon sales trends and customer purchase behavior and build an application to determine association between the popular products, identify product bundles to improve cross-sell and recommend products to targeted customers.

Research Questions



Which products can be bundled together for up-sell and cross-sell ? (Slide – 11)



Which customer can be recommended for bundled products ? (Slide – 16)



Can we create customer segments for bulk recommendations ? (Slide – 13)



How much revenue increase is anticipated ? (Slide – 17)

Data

The Amazon sales dataset from Kaggle utilized for this effort. Data contains 3204 rows and 9 columns.

Attributes:

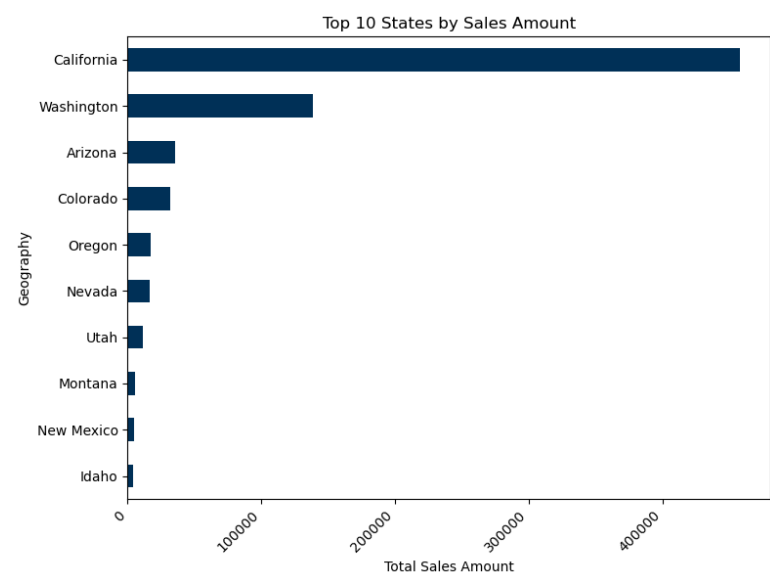
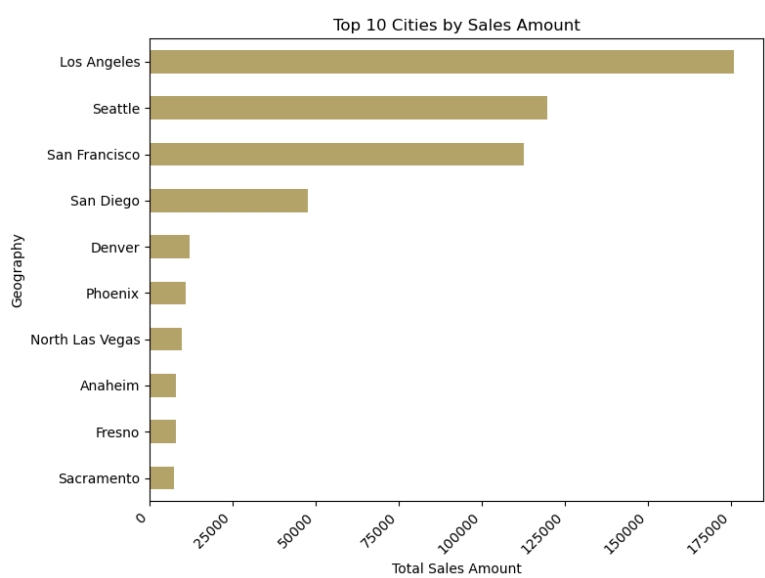
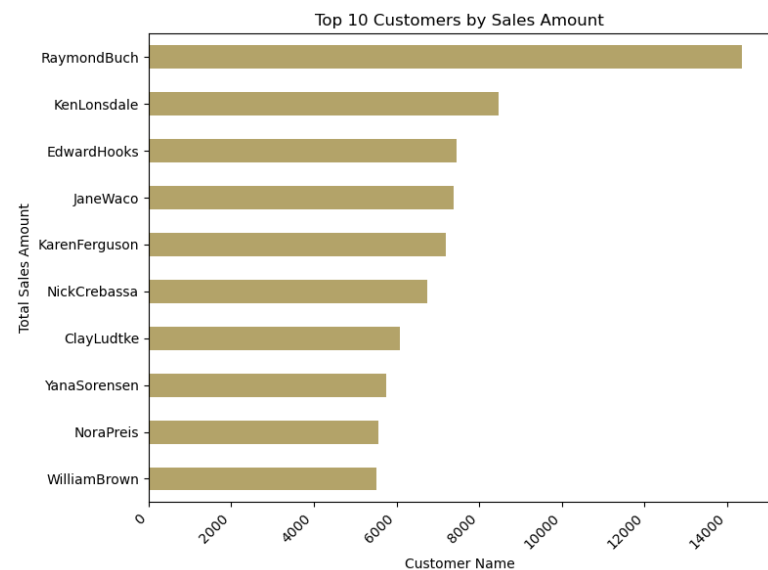
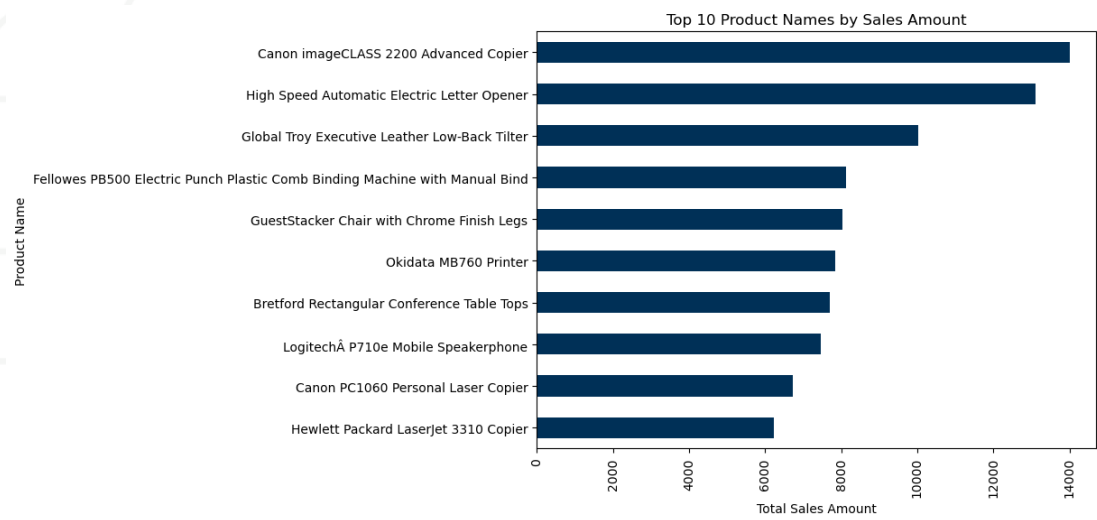
Column Name	Column Description
Order Date	Order Request Date
Ship Date	Shipping Date
Email ID	Email ID of Users
Geography	Location of Orders by Users
Category	Product Category
Product Name	Product Name of Amazon
Sales	Amazon Product Sales
Quantity	How many units of a particular product are available
Profit	Amazon Sales Profit

[Credit : Amazon_Sales_Dataset \(kaggle.com\)](https://www.kaggle.com/datasets/amazon-represents/amazon-sales-dataset)

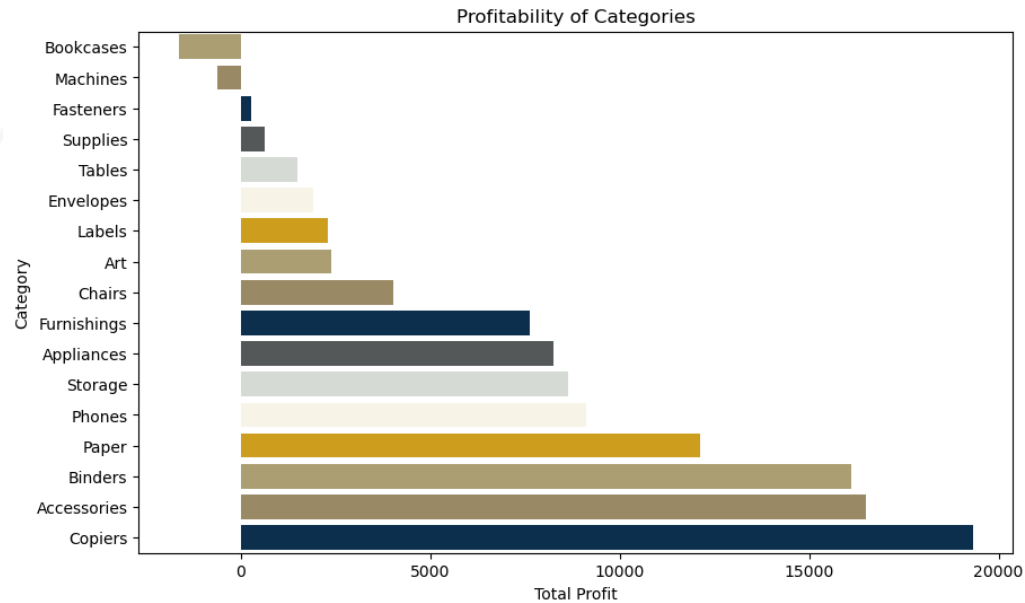
The background of the slide is a faded, sepia-toned photograph of a large, multi-story brick building with many windows, likely a university hall. In the foreground, a group of four people is walking away from the camera on a paved path. The overall aesthetic is academic and professional.

Exploratory Data Analysis

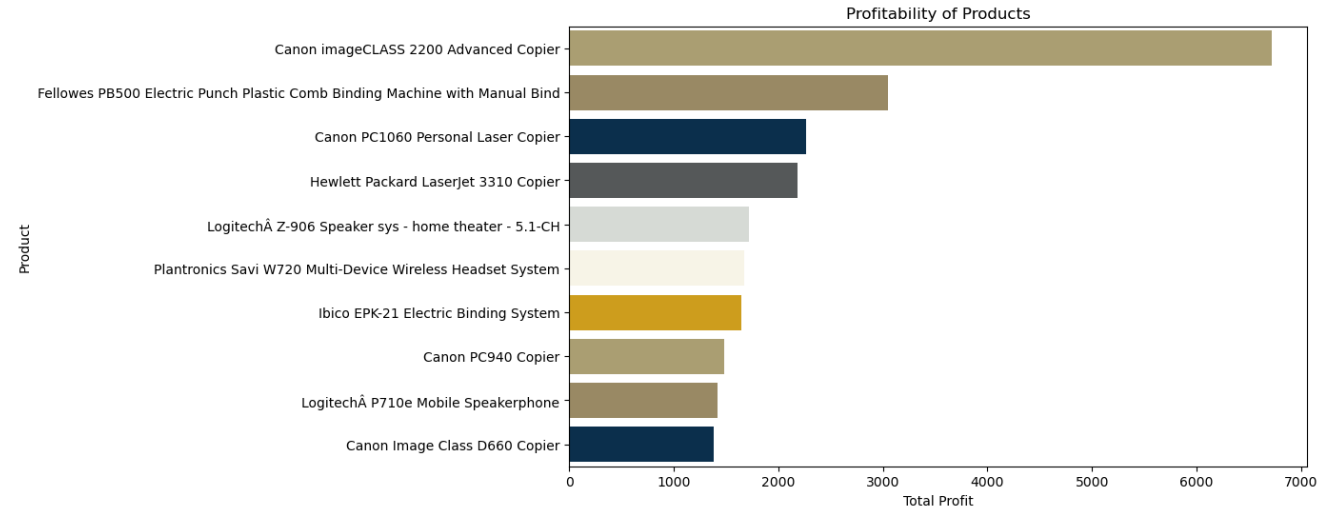
Top Products, Customer & Geography by Sales



Profitable Products & Categories

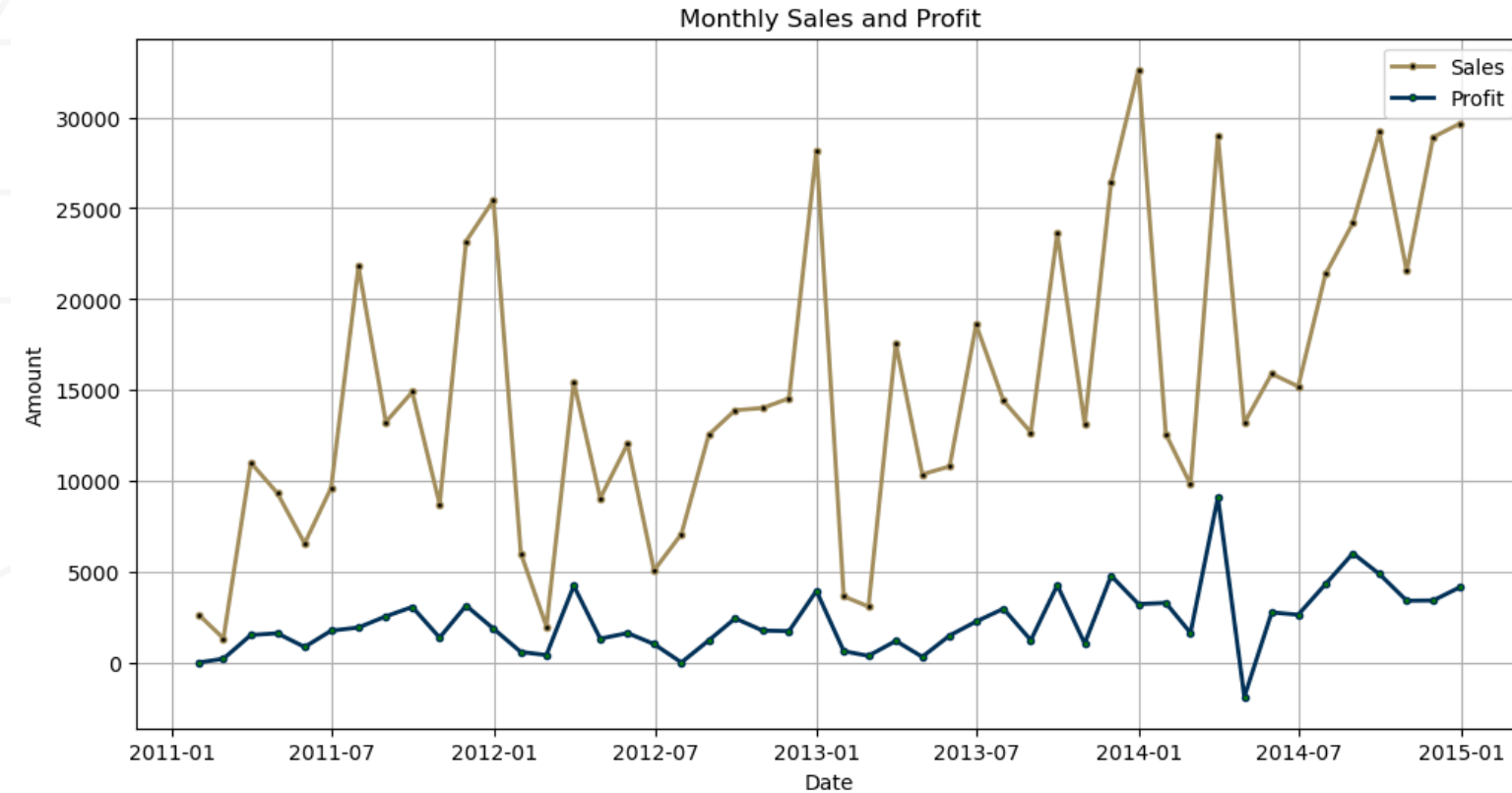


Copiers is most profitable while bookcases seems the least



Cannon Image Class 2200 is most profitable

Profit & Sales Trend



- Both Profit & Sales have increasing trend Year over Year from 2011 to 2014
- Sales & Profit peaked during Dec every year
- Seasonality effects are evident overall

Sales & Profit by Geography

Sales



Profit



- Most Profit & Sales came from California and Washington
- No Sales in Midwest or East coast during 2011-2014



Modelling & Results

Product Association/Grouping

1. We converted the sales data at order and product level to create a list of all receipts over 2011 to 2014.
2. We used association rule mining to determine the association between the products sold. An association rule is an implication expression of the form $X \rightarrow Y$, where X & Y are disjoint item sets.
3. Association Rule Mining has 3 important aspects. Given a rule " $A \rightarrow C$ ", A stands for antecedent, and C stands for consequent.
 - **Support** : Support measures the frequency or the proportion of transactions in the dataset in which a particular combination of items (or itemset) appears together.

$$\text{support}(A \rightarrow C) = \text{support}(A \cup C), \text{range: } [0, 1]$$

- **Confidence** : The confidence of a rule $A \rightarrow C$ is the probability of seeing the consequent in a transaction given that it also contains the antecedent.

$$\text{confidence}(A \rightarrow C) = \text{support}(A \rightarrow C) / \text{support}(A), \text{range: } [0, 1]$$

- **Lift** : The lift metric is commonly used to measure how much more often the antecedent and consequent of a rule $A \rightarrow C$ occur together than we would expect if they were statistically independent. If A and C are independent, the Lift score will be exactly 1.

$$\text{lift}(A \rightarrow C) = \text{confidence}(A \rightarrow C) / \text{support}(C), \text{range: } [0, \infty]$$

4. Using the above logic, we extracted the products with strongest associations i.e. Lift > 1.2 and confidence ≥ 0.65

Results - Product Association/Grouping



Which products are associated and can be bundled together for up-sell and cross-sell ?

Focus Product	Associated Product	Lift	Confidence	Support
O'Sullivan 4-Shelf Bookcase in Odessa Pine	GBC Standard Recycled Report Covers, Clear Plastic Sheets	201	50%	0.12%
GBC Standard Recycled Report Covers, Clear Plastic Sheets	O'Sullivan 4-Shelf Bookcase in Odessa Pine	201	50%	0.12%
Tennsco Regal Shelving Units	Staples	14	50%	0.12%
GBC Wire Binding Combs	Carina Double Wide Media Storage Towers in Natural & Black	179	67%	0.12%
Wirebound Four 2-3/4 x 5 Forms per Page, 400 Sets per Book	Staples	14	50%	0.12%
#10 White Business Envelopes, 4 1/8 x 9 1/2	Staples	28	100%	0.12%
Advantus Rolling Storage Box	Great White Multi-Use Recycled Paper (20Lb. and 84 Bright)	215	67%	0.12%
Xerox 1894	Xerox 225	115	50%	0.12%
Boston 16765 Mini Stand Up Battery Pencil Sharpener	Microsoft Natural Ergonomic Keyboard 4000	537	100%	0.12%
Microsoft Natural Ergonomic Keyboard 4000	Boston 16765 Mini Stand Up Battery Pencil Sharpener	537	67%	0.12%

* Our criteria for strong Association is at Least 60% confidence in association and life greater than 1.2

Customer Grouping

1. This is a customer level information. We grouped customer to calculate the Recency, Frequency and Monetary Value over 2011-2014
 - Recency (R) refers to how recently a customer has made a purchase or engaged with the business.
 $R_i = (\text{Max_Date} - \text{Purchase_Date})$
 - Frequency (F) measures how often a customer makes purchases or interacts with the business over a specific period.
 $F_i = \text{count of all transaction for the customer}$
 - Monetary (M) value represents the total amount of money spent by a customer over a certain period.
 $M_i = \text{sum of Sales \$ for the customer.}$
2. We calculated relative recency, frequency and monetary values by create a quartile of 1 to 5 for each of them.

Recency	Frequency	Monetary
R-Tier-1 (most recent)	F-Tier-1 (most frequent)	M-Tier-1 (highest spend)
R-Tier-2	F-Tier-2	M-Tier-2
R-Tier-3	F-Tier-3	M-Tier-3
R-Tier-4	F-Tier-4	M-Tier-4
R-Tier-5 (least recent)	F-Tier-5 (only one transaction)	M-Tier-5 (lowest spend)

3. We summed relative recency, frequency and monetary values quartile scores to have a customer score and ranked the customer in ascending order to get the most valued customers for targeted marketing.

Results - Customer Grouping



Can we create customer segments for bulk recommendations ?

Customer Name	Recency	Frequency	TotalSpend	Rel_Frequency	Rel_TotalSpend	Rel_Recency	Rqm_Score	Rank	Segment
LaurelBeltran	250	14	2836.04	4	5	5	14	1	2
SandraGlassco	46	10	2567.151	3	5	5	13	2	0
AlanDominguez	98	5	4353.34	2	5	5	12	3	0
AlanSchoenberger	371	9	2017.184	3	5	4	12	4	0
ArthurGainer	523	9	2382.179	3	5	4	12	5	0
...
MauriceSatty	965	4	83.102	1	1	2	4	682	1
BenjaminPatterson	1312	3	145.66	1	1	1	3	683	1
ChristineKargatis	1188	3	140.708	1	1	1	3	684	1
PatrickJones	1274	4	137.504	1	1	1	3	685	1
PaulLucas	1141	3	88.57	1	1	1	3	686	1

Customers are ranked and placed in appropriate segment for recommendation

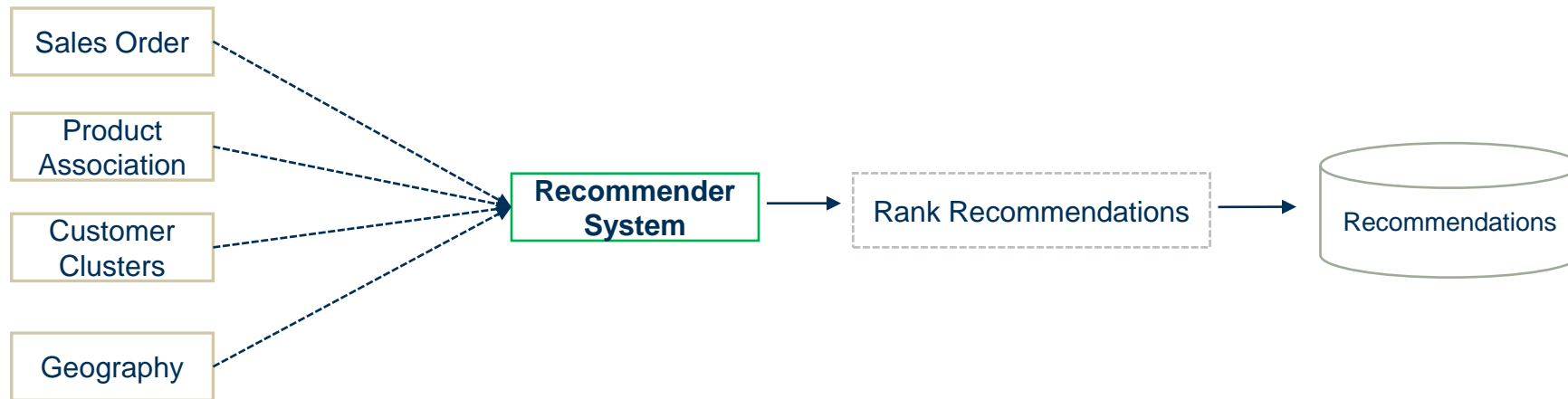
Results - Customer Grouping

Similar Customers were clustered using K-Mean clustering in 5 groups using Recency, Frequency and Monetary value. Below are the cluster summary:

Segment	Recency (days)	Frequency	Total Spend (\$)	Insights	Marketing Strategy
0	370.69	6.27	2039.29	Recent purchasers with high total spend. Likely regular customers contributing significantly to revenue and profit.	Maintain loyalty with personalized recommendations, special discounts, and targeted marketing campaigns.
1	853.85	4.5	682.61	Moderate engagement and expenditure but haven't made recent purchases.	Re-engage with personalized reminders, limited-time discounts, and showcasing new products.
2	963.9	10.32	1682.67	High frequency of transactions but haven't made recent purchases. Moderate expenditure.	Reactivate with targeted campaigns, personalized messages, and referral incentives.
3	63.35	2.08	365.14	Recent shoppers with moderate engagement and expenditure. Likely new or occasional shoppers.	Implement first-time buyer programs, welcome discounts, and recommend popular products to increase purchase frequency and amount.
4	874.46	9.23	5620.29	Loyal customers with recent purchases, high engagement, and expenditure.	Strengthen loyalty with exclusive rewards, early access, and feedback programs.

Customer Recommendation

- For Customer Recommendation, we have used Content-based filtering method as our recommender engine. Content-based filtering uses item features to recommend other items similar to what the user likes, based on their previous actions or explicit feedback.
- We built user profile using Clustering of customers and their buying pattern.
- We used association rule to understand the product association
- Based on the association and customer buying pattern, we recommended product to customer.
- We rank the customers with highest probability using R, F & M Quantile score.



Results - Customer Recommendations



Which customer can be recommended for bundled products ?

Customer	Product Recommendation	Recommendation Comments
JustinRitter	GBC Standard Recycled Report Covers, Clear Plastic Sheets	Customer bought product O'Sullivan 4-Shelf Bookcase in Odessa Pine and not GBC Standard Recycled Report Covers, Clear Plastic Sheets.
KeithHerrera	GBC Standard Recycled Report Covers, Clear Plastic Sheets	Customer bought product O'Sullivan 4-Shelf Bookcase in Odessa Pine and not GBC Standard Recycled Report Covers, Clear Plastic Sheets.
MarkHaberlin	O'Sullivan 4-Shelf Bookcase in Odessa Pine	Customer bought product GBC Standard Recycled Report Covers, Clear Plastic Sheets and not O'Sullivan 4-Shelf Bookcase in Odessa Pine.
PatrickGardner	O'Sullivan 4-Shelf Bookcase in Odessa Pine	Customer bought product GBC Standard Recycled Report Covers, Clear Plastic Sheets and not O'Sullivan 4-Shelf Bookcase in Odessa Pine.
FredMcMath	Staples	Customer bought product Tensco Regal Shelving Units and not Staples.
ScottWilliamson	Staples	Customer bought product Tensco Regal Shelving Units and not Staples.
AlanBarnes	Carina Double Wide Media Storage Towers in Natural & Black	Customer bought product GBC Wire Binding Combs and not Carina Double Wide Media Storage Towers in Natural & Black.
ChadSievert	Staples	Customer bought product Wirebound Four 2-3/4 x 5 Forms per Page, 400 Sets per Book and not Staples.
PenelopeSewall	Staples	Customer bought product Wirebound Four 2-3/4 x 5 Forms per Page, 400 Sets per Book and not Staples.
CynthiaDelaney	Great White Multi-Use Recycled Paper (20Lb. and 84 Bright)	Customer bought product Advantus Rolling Storage Box and not Great White Multi-Use Recycled Paper (20Lb. and 84 Bright).
DennyJoy	Xerox 225	Customer bought product Xerox 1894 and not Xerox 225.
JohnLucas	Xerox 225	Customer bought product Xerox 1894 and not Xerox 225.
KalycaMeade	Boston 16765 Mini Stand Up Battery Pencil Sharpener	Customer bought product Microsoft Natural Ergonomic Keyboard 4000 and not Boston 16765 Mini Stand Up Battery Pencil Sharpener.

Business Benefit



How much revenue opportunity is anticipated ?

Customer	Product Recommendation	Unit Price (\$)	Average Purchase/Year	Opportunity (\$)
JustinRitter	GBC Standard Recycled Report Covers, Clear Plastic Sheets	\$ 3.24	10 \$	32.40
KeithHerrera	GBC Standard Recycled Report Covers, Clear Plastic Sheets	\$ 3.24	10 \$	32.40
MarkHaberlin	O'Sullivan 4-Shelf Bookcase in Odessa Pine	\$ 102.00	5 \$	510.00
PatrickGardner	O'Sullivan 4-Shelf Bookcase in Odessa Pine	\$ 102.00	5 \$	510.00
FredMcMath	Staples	\$ 28.40	5 \$	142.00
ScottWilliamson	Staples	\$ 28.40	5 \$	142.00
AlanBarnes	Carina Double Wide Media Storage Towers in Natural & Black	\$ 80.98	6 \$	485.88
ChadSievert	Staples	\$ 28.40	10 \$	284.00
PenelopeSewall	Staples	\$ 28.40	10 \$	284.00
CynthiaDelaney	Great White Multi-Use Recycled Paper (20Lb. and 84 Bright)	\$ 5.98	100 \$	598.00
DennyJoy	Xerox 225	\$ 20.74	6 \$	124.44
JohnLucas	Xerox 225	\$ 20.74	6 \$	124.44
KalycaMeade	Boston 16765 Mini Stand Up Battery Pencil Sharpener	\$ 11.66	20 \$	233.20
Total Opportunity				\$ 3,502.76

Considerations:

- Top10 customer product combination is considered
- Average Purchase/Year is calculated by taking average qty sold over 2011-2014
- Unit Price is calculated dividing Sales Price/Qty
- Opportunity (\$) = Unit Price * Average Purchase/Year

Conclusion

- We considered the Amazon sales dataset from Kaggle contains sales data from 2011-2014.
- With the Sales data we found there are several product being sold together and have strong associations like Wirebound Four 2-3/4 x 5 Forms per Page, 400 Sets per Book & Staples, Xerox 1894 & Xerox 225,m #10 White Business Envelopes, 4 1/8 x 9 1/2 & Staples. (Slide-11)
- We segmented customers based on Recency, Frequency and Monetary value and observed a possibility of 5 grouping – Recent purchase & High spend, Moderate engagement & spend, High frequency & moderate expenditure, Recent with moderate expenditure and Recent Purchase with high frequency customer. We created targeted selling strategy for these customers .(Slide 14)
- Using the product and customer combination, we created targeted recommendations for customers.
- Finally, we calculated the up-sell and cross-sell revenue opportunity with our marketing effort and realize a potential gain of **\$3,500 from top 10 customer** with only few combinations.

Future Work

- We can consider can larger data set to identify more profitable combinations.
- We can build an application for recommendation and receive customer feedback to improve the recommendation.
- Try out collaborative filtering methodology for improving recommendations.

References

<https://www.upwork.com/resources/what-is-content-based-filtering>

<https://developers.google.com/machine-learning/recommendation/content-based/basics>

https://en.wikipedia.org/wiki/Recommender_system

<https://rasbt.github.io/mlxtend/>

<https://www.turing.com/kb/content-based-filtering-in-recommender-systems>