Amazon Business Case

Team Members:
Kevin, Alessandro, Zhaochen Ye, Ruoxin Shi,Crystal Xu, Ali

1. Company Introduction

Amazon's Mission, Vision and Value

Mission:

"**We strive to offer our customers the lowest possible prices, the best available selection, and the utmost convenience**."

- affordable prices
- varied selections
- top convenience

Vision:

"Our vision is to be earth's most **customer-centric company**; to build a place where people can come to find and discover anything they might want to buy online."

- customer-centric
- global presence

Value:

- ➔ **Customer Obsession**
- ➔ Ownership
- ➔ Invent and Simplify
- ➔ Learn and Be Curious
- ➔ Hire the Best
- ➔ The Highest Standards
- ➔ Think Big
- ➔ Bias for Action
- ➔ Earn Trust
- ➔ Deliver Results

Core Value

Customer Obsession

“Leaders start with the customer and work backwards. They work vigorously to earn and keep customer trust. Although leaders pay attention to competitors, they obsess over customers.”

- more than customer focus
- customers always dissatisfied, want something better unintentionally
- think ahead, improve services before they have to
- long term needs

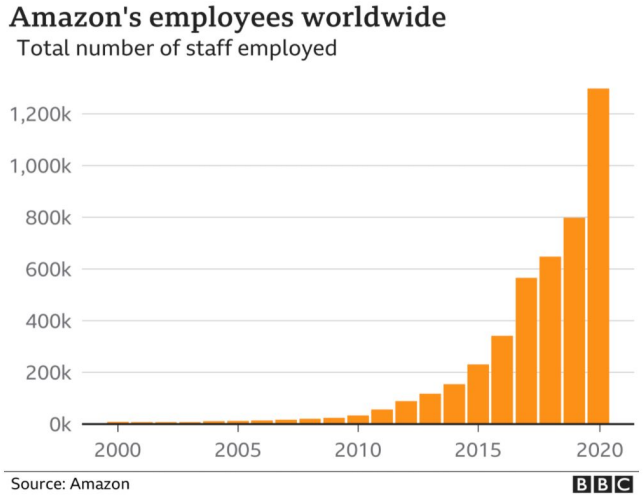
Company Size & Performance

As of Year-End 2020...

Size

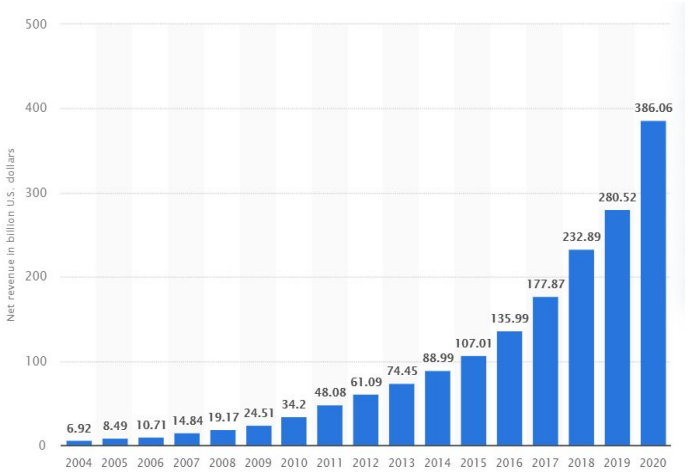
1,298,000 Employees Growing **62.66%** from last year

Operates in 13 countries with Fulfillment Center



Revenue

$386 Billion, Growing **37.6%** from last year

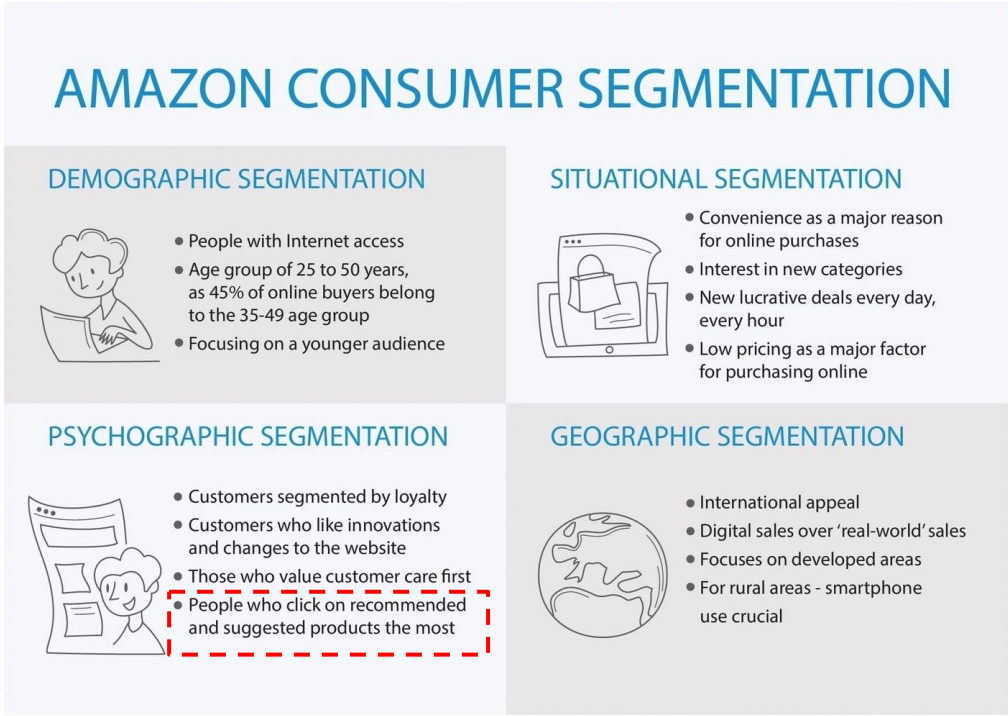


Customer Base (Commerce & Retail)

- Amazon announced that they currently have over **200 million** Amazon Prime subscribers around the world.
- Amazon has **148.6 million** Amazon Prime Members in the US. An increase of **49.05%** from 99.7 million in 2017.

Segmentation:

- Individual / Business
- Demographic: Age, Gender, Occupation
- Geographic: Domestic / International
- Behavioral: Loyalty / Personality
- Situational
- Online / Physical



Major Products and differentiators

Online Stores

amazon.com games music amazonstudios amazonbasics

-Products sold through online stores

-Amazon Brands

-Digital Media Products like ebook, music, video, games and software.

Differentiators:

-Broader Product Assortments

-Guaranteed quality

-Wide user base

-Customer Obsession

AWS

aws

-Sales of compute, storage, database, and other services

- IaaS, PaaS, SaaS

- On-premises and cloud-based

Differentiators:

-Top 1 market share of 33%

-Spans 81 Availability Zones within 25 geographic regions

-Most functionality

-Largest community of customers and partners

-Fastest pace of innovation

Physical Stores

amazon books

amazon pop up

amazon 4-star

amazonfresh

WHOLE FOODS MARKET

amazon go

Differentiators:

-Customer base and market share of the online store

-Branding supermarket chain of WholeFoods

-Inhouse technology support

Subscription Services

amazon music

prime video amazon Prime

-Amazon Prime memberships

-Access to content including digital video, audiobooks, digital music, e-books, and other non-AWS subscription services

Differentiators:

-Bundled Value: Not only offers shopping and delivery benefits, but streaming and digital benefits as well

-Provide exclusive and original digital content

-Access to a great amount of video, songs, books and games

Third-Party Seller Services

-offer programs that enable sellers to sell their products in Amazon stores, and fulfill orders

-Fulfillment by Amazon

-Shipping with Amazon

amazon freight

amazon Prime Air

Differentiators:

-Robust supply chain; Automated, international, highly efficient warehouse and delivery network

-Varias Apps and tools supported by advanced inhouse tech

2. Project Introduction

What are we proposing!

1. How are recommendations made?

- Current algorithm: Geared towards recommending the most popular products on the website based on customer's previous purchases

-Proposed shift: Recommendations should be based on each user's prior purchases and their relationship with the website. Account for your previous purchases and those of similar users.

2. Where does the project fit within Amazon?

-The core values of amazon: Customer-obsession, a bias for action, empower customers through a better online experience!

-The recommendation algorithm aligns with the core values! Essentially shifts the dynamic of product development from being generic to personalized to the customers needs!

- We know that a customer has precious little time! We need to make sure that the likelihood of purchase per minute spent online increases!

But does the online shopping experience require modifications?

The Why!

1. Increase in customer satisfaction
2. Increase in customer loyalty
3. Benefits translate to amazon - upstick in revenues and margins!

35% of all of Amazon's transactions are a consequence of algorithmical product recommendations!

We know that a customer has precious little time and we need to make sure that the customer's time spent online is optimized to deliver preferred outcomes.

How will we accomplish this?

A more personalized approach catered to each customer better captures interest

Increased customer engagement leads to greater time spent online – Copy the blueprint from Instagram and Facebook!

Personalizes the shopping experience as opposed to treating you like a generic buyer

The model will have more room to grow iteratively with each purchase and make more accurate recommendations

With each purchase, the customer is closer to their most desired product

3. Technical Solution

Algorithm

Personalization Recommendation

- Create labels for customers based on their profile and purchasing/searching history; Match customers labels with products labels
- Algorithm can be used:
  - Clustering - K Means, Rule Based

Collaborative Filtering

- Find the closest neighbors from a user of interest and suggest popular items among these neighbors
- Algorithm can be used:
  - KNN - find similar customers
  - Clustering for customer segmentations
  - ANN

Similar Products Recommendation

- Determine the similarity of two products by looking at how often they're present together in the browsing or purchase histories of users; when customers view the products, recommend the associated products or bundles as well
- Algorithm can be used:
  - Association Rule to calculate item-to-item distance matrices
  - Clustering to group products

Possible Risk: Cold Start

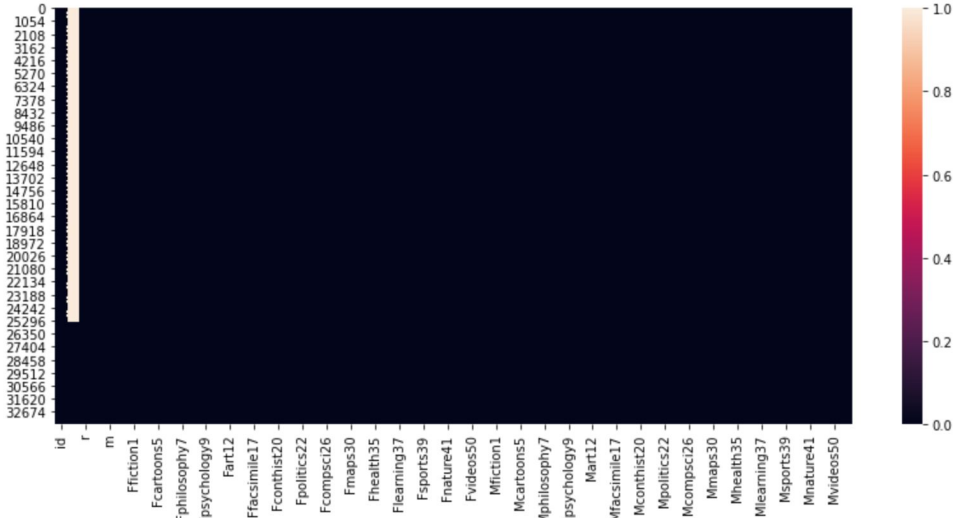
- New Customer
  - Put into proper groups based on initial profile, questionnaires, other social media information.
  - If not enough information, recommend the most popular items.
- New Product
  - Find similar products based on attribute and description and use the recommendation strategies for those products

Data Overview

[29]:

	id	r	f	m	tof	Ffiction1	Mfiction1	Flegends6	Mlegends6	Fphilosophy7	Mphilosophy7	Freligion8	Mreligion8
0	914	194	7	318.894775	1703	1	5.949997	0	0.0	0	0.00000	0	0.000000
1	957	3	14	368.052246	2364	1	12.680038	0	0.0	0	0.00000	0	0.000000
2	1406	1489	15	423.298340	2371	0	0.000000	0	0.0	0	0.00000	0	0.000000
3	1414	155	4	71.217041	1290	0	0.000000	0	0.0	0	0.00000	0	0.000000
4	1546	194	6	442.638184	2188	0	0.000000	0	0.0	3	26.19841	1	19.899994

id: UserId
r: Recency
F: Overall Frequency
m: Overall Monetary
tof: Time on File (in days)
F*: variables are count, e.g., Fclassic3 is the number of classics books the customer has bought in the past
M*: variables are monetary, e.g., Mclassic3 is the amount spent on classics in the past



RFM Analysis

Group the **Recency, Frequency, Monetary** into 4 quantiles

↓

Give scores based on quantiles

↓

Sum the numbers to get RFM score for each customer

↓

Segment the customer based on RFM Scores

	r	f	m	
	mean	mean	mean	count
RFM_Level				
Can't Loose Them	217.2	11.6	552.1	12359
Champions	431.0	4.5	164.9	3804
Loyal	577.6	3.5	112.4	3857
Needs Attention	862.4	1.2	30.5	3089
Potential	535.8	2.1	63.8	4483
Promising	694.9	1.6	44.2	4462
Require Activation	1460.3	1.0	19.0	1659

Insights

- Understand what the **incremental value** of each customer is.
- Group customers and apply different strategies on them
- Having these groups help us understand if a recommendation system is keeping the top class happy and maximizing improvements from those requiring activation.
- We can use these metrics for two key analyses:
  - Business Impact: Did we make sure attrition was minimized for the top class and did we increase the scores for the lowest classes?
  - Modeling: Knowing what class of engagement each user is in is important - we would very much want to include these data points in any modeling efforts downstream.

Improvements

- Create classes based more on which scores of RFM they are high on and which ones need improvement.

K-Means Clustering

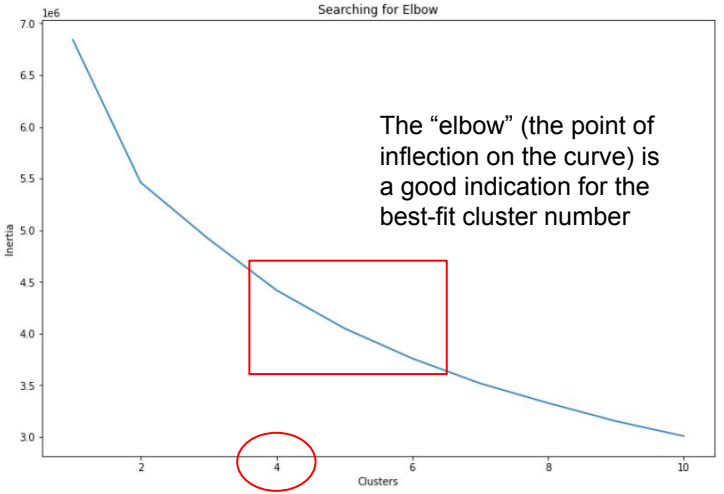
Clustering Frequency of all types of books purchased by every customer into k clusters. each customer belongs to the cluster with the nearest mean.

This method produces exactly k different clusters of greatest possible distinction.

Using “Elbow” Method to find the best-fit k

Visualize Clustering in 2D and remove outliers to find the core of the trend in this section

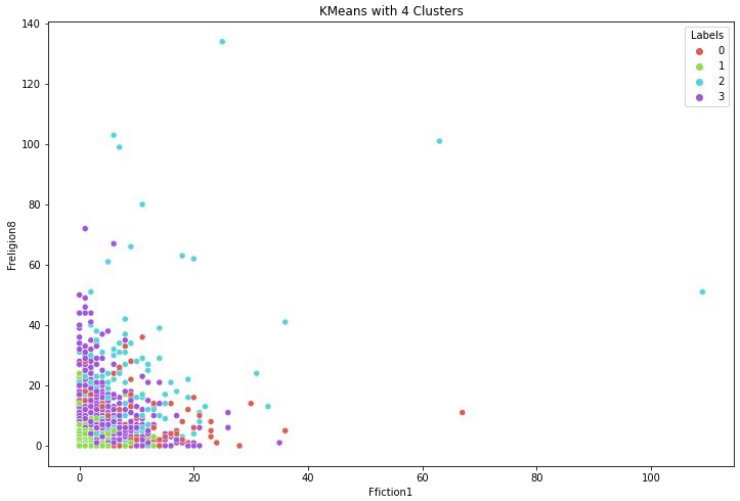
To find the optimal number of clusters k, we will use the “Elbow” method. Inertia measures how well a dataset was clustered by K-Means



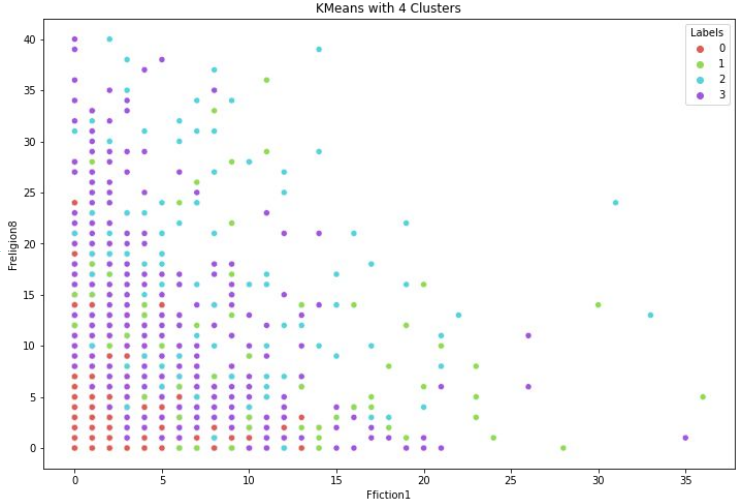
In this case, we choose 4 as our cluster number

K-Means Clustering

Clusters for Fiction1 and Religion8 with outliers

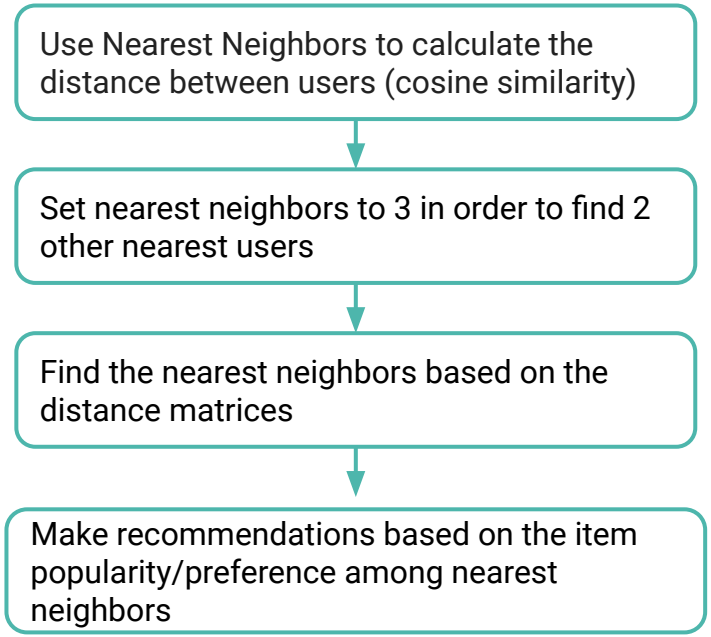


Clusters for Fiction1 and Religion8 without outliers



<u>Purple(cluster 3)</u> seems more into religion than fiction, while **<u>green(cluster 1)</u>** seems to be more on the side of fiction.

Collaborative Filtering - KNN



Optimization Methods Instead of 3 neighbors, find the optimal n

- We can experiment with **different distances** (i.e Euclidean vs Manhattan).
- The metric we will use to identify the optimal number of neighbors is comparing the recommendation performance in terms of revenue.

Next Steps Instead of k=3 KNN, we can experiment with more sophisticated proven methods:

- **Market-Basket (Apriori)** This would be more relevant to our use-case compared to KNN because it fits the context of our analysis. It creates a network of likes and dislikes between users and items, whereas KNN just compares the distances between users.
- **Reinforcement Learning** This would be good to implement because it creates a sense of context for the algorithm to operate. Context can significantly help improve the precision score and thus drive business value

What additional data do we need?

- Searching/clicking history of users
- Detailed purchasing history of Amazon customer
- Nice to have if possible:
  - External searching data (Google)
  - 3rd party customer profile data
  - Alexa (If we ignore the time frame)

Efficiency Testing

- Controlled Group
  - Seperate the users into 2 groups, one with new AI based recommendation system, one with old recommendation system.
  - Compare the success metrics of two groups
- A/B Testing

Success Metrics

- ❑ Increase CTR (Click through rate) = Click / View
- ❑ Increase CVR (Conversion rate) = Order / Click
- ❑ Increase Merchandise Volume / 1000 recommendations
- ❑ **% of revenue through recommendations**

4. Build the team

Recommendation Engine Improvement Team

In order to improve Amazon's recommendation engine, we will need to hire for a couple of different positions. However, given Amazon's size and mastery of digital transformation, we can assume some of the necessary skillsets have already been centralized.

What Amazon likely already has:

Data Engineers and Stewards: There are likely teams of people who work on maintaining Amazon customer data, so we will not need to hire additional engineers to add on to this work

What this team needs:

Project Manager: Someone to lead projects across a lot of different stakeholders to bring our research into production.

Data Scientist: Someone with a strong business and statistical background to develop / improve the models we plan to use.

Machine Learning Engineer: Someone with strong engineering background who can take the DS's work and know how to productionalize it in Amazon's tech stack.

Recommendation Engine Improvement Budget

In total, our development costs will likely be around $900k / year. We are saving a lot of money by leveraging shared services and by developing strong partnerships within the organization.

Role	How many	When	How Long?	Cost	Hiring strategy
Data Scientist	2	Discovery	Production	$150k	Internal
Data Steward	0	Discovery	Development	$0	Internal/partner
Data Engineer	0	Development	Production	$0	Internal/partner
Domain Expert	1	Discovery	Production	$150k	Partner
Software Developers	2	Development	Prod/post prod	$150k	Partner
Project Manager	1	Discovery	Production/post	$150k	Interal/Partner
Total				**$900k**	

5. Project Execution

Planning

A Data Science team's planning schedule might look something like this (Gantt chart):

Week 1: Brainstorm and Planning for prioritizing aspects of the recommendation system to research over the next couple of quarters.

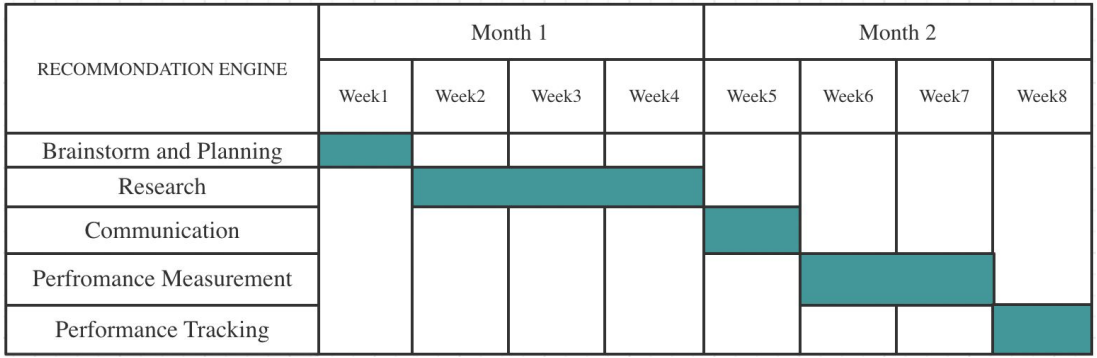
Week 2-4: Individuals will conduct their research in order to provide requirements for engineering team to implement changes.

Week 5: Socialize requirements cross functionally in a core meeting (involving engineering, business stakeholders, market team etc.)

Week 6-7: Provide requirements, set up automated tracking to measure performance after launch

Week 8: After implementation, validate and track performance.

Repeat!



Tools and Platforms

An Amazon Data Scientist's work might look something like this.

Planning: MS Planner, Excel, MS Teams - a centralized team project management tool.

Preparation of Data: SQL & AWS to pull data from Amazon's data environment for research.

Research Phase: To best improve the recommendation engine, using a combination of Jupyter Notebooks and H2O.ai service will help us do a lot of modeling research over a couple of weeks.

Socializing Phase: Use GitHub / Notebook PDF to communicate findings and provide requirements to the engineering team. Python sometimes is explicitly necessary for MLE tasks.

Deployment, Validation and Tracking: Setting up an Airflow with AWS / Tableau to automate data refresh on the communicated changes to measure performance.

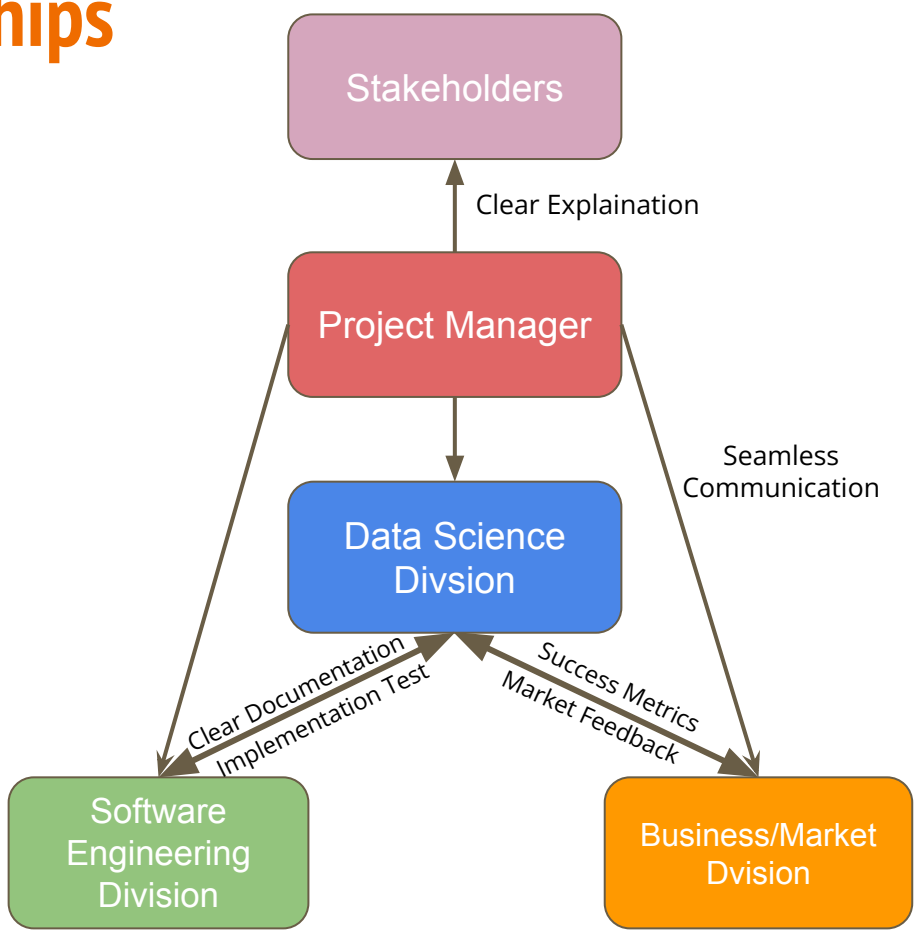
Developing Strong Partnerships

Leadership: Understand what their goals are, and make sure your work aligns with those strategic objectives. Having the confidence to hold your ground to inform or correct them to make sure they are making sound decisions and reasonable expectations.

Business Stakeholders: Explain concepts clearly without arrogance. Don't be pedantic or over technical.

Software Engineers: Provide clear documentation & requirement, understanding technical environment and timelines. Receiving implementation test results from them.

Business/Market Division: Provide clear success metrics. Receiving market feedback from them for a deeper market understanding.



6. Project Value

Driving Business Value

Is there economic opportunity through an improved algorithm?

- **As a refresher, the 'hope' of the recommendations is that users will add items to their shopping cart and increase the size of each purchase.**
- **While there are multiple interrelated goals associated with a better recommendation engine, from an economic perspective driving incremental revenue is the primary goal.**

Goal

- ❑ **Increase revenue**
- ❑ Decrease cost
- ❑ Better user experience
- ❑ More activity on the website
- ❑ Competitive edge over competitors

Success Metrics

- ❑ Increase CTR (Click through rate) = Click / View
- ❑ Increase CVR (Conversion rate) = Order / Click
- ❑ Increase Merchandise Volume / 1000 recommendations
- ❑ **% of revenue through recommendations**

The Results are In!

Insights:

- **Click-through rate remained consistent with baseline.**
- **Conversion increased 2.2%, greater than our hypothesized 1%.**
- **This implies that the algorithm doesn't necessarily increase the chances of customers to explore recommendations; however, the increased conversion rate implies the new algorithm's recommendations are more relevant to the customer.**
- **Our development costs were under initial assumption.**

Conclusion:

- **We outperformed our hypothesis assumption of 97% ROI by 38% pts!**
- **This equates to $2.35M in annual economic value.**
- **This aligns with Amazon's core value of 'Customer Obsession' while also adding economic value for shareholders = SUCCESS!!**

Hypothesis Business Case:

Is the result good enough?

Each month more than 197 million people around the world get on their devices and visit Amazon.com.

That's more than the entire population of Russia.

- **Key Inputs and Assumptions on 197M monthly visitors:**
  - Base click-through rate on a recommendation is 10%
  - Base conversion rate on a recommendation is 10%
  - Avg order value = $25
  - # of non-recommended purchases is the same; therefore all recommendation-driven orders is incremental revenue and value

Assumptions:	
Monthly visitors	197,000,000
Base rec eng CTR	10.0%
% CTR Improvement	1.0%
Base rec eng CVR	10.0%
% CVR Improvement	1.0%
PF CTR	10.1%
PF CVR	10.1%
Contribution Margin	25.0%
Est Development Costs	$1,000,000

Base Scenario:	
# of rec eng clicks	19,700,000
Base rec eng orders	1,970,000
AOV	$25
Base Monthly Revenue	**$49,250,000**
Pro-Forma 'PF' Scenario:	
PF rec clicks	19,897,000
PF rec eng orders	2,009,597
AOV	$25
PF Monthly Revenue	**$50,239,925**
Δ in Monthly Revenue	$989,925
Incremental Contribution	**$247,481**
Annualized Contribution	**$2,969,775**
Net Annual Economic Value*	**$1,969,775**
ROI	***97.0%***

** Assume new algorithm is replaced by new project annually for continues improvement*

Conclusion: Assuming we can improve key metrics by 1% or more there is significant economic value in the project.

Actual Results:

Metrics	Assumptions	Actuals
Monthly visitors	197,000,000	197,000,000
Base rec eng CTR	10.0%	10.0%
% CTR Improvement	1.0%	0.0%
Base rec eng CVR	10.0%	10.0%
% CVR Improvement	1.0%	2.2%
PF CTR	10.1%	10.0%
PF CVR	10.1%	10.2%
Contribution Margin	25.0%	25.0%
Development Costs	$1,000,000	$900,000

Base Scenario:	Hypothesis	Actuals	Difference
# of rec eng clicks	19,700,000	19,700,000	0
Base rec eng orders	1,970,000	1,970,000	0
AOV	$25	$25	$0
Base Monthly Revenue	**$49,250,000**	**$49,250,000**	**$0**
Pro-Forma 'PF' Scenario:			
PF rec clicks	19,897,000	19,700,000	-197,000
PF rec eng orders	2,009,597	2,013,340	3,743
AOV	$25	$25	$0
PF Monthly Revenue	**$50,239,925**	**$50,333,500**	**$93,575**
Δ in Monthly Revenue	$989,925	$1,083,500	$93,575
Incremental Contribution	**$247,481**	**$270,875**	**$23,394**
Annualized Contribution	**$2,969,775**	**$3,250,500**	**$280,725**
Net Annual Economic Value*	**$1,969,775**	**$2,350,500**	**$380,725**
ROI	***97.0%***	***135.1%***	***38.1%***

** Assume new algorithm is replaced by new project annually for continuous improvement*

THANK YOU

Sources

https://www.bbc.com/news/business-57109282

https://www.statista.com/statistics/234488/number-of-amazon-employees/

https://www.digitalcommerce360.com/article/amazon-sales/

https://markets.businessinsider.com/news/stocks/amazon-worlds-most-valuable-retailer-1029364143

https://www.organimi.com/organizational-structures/amazons-corporate-structure/

https://www.nytimes.com/2017/06/16/business/dealbook/amazon-whole-foods.html

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

https://en.wikipedia.org/wiki/Amazon_(company)

https://www.thestreet.com/investing/stocks/here-s-are-15-of-the-most-important-dates-in-amazon-s-history-14135911

https://www.statista.com/statistics/1109330/walmart-ecommerce-sales-by-division-worldwide/

https://www.statista.com/statistics/226793/e-commerce-revenue-of-alibabacom/#:~:text=The%20revenue%20of%20Alibaba%20Group,dollars)%20in%20financial%20year%202020

https://www.statista.com/statistics/1240178/amazon-physical-store-sales-united-states/

https://www.macrotrends.net/stocks/charts/HD/home-depot/revenue

https://www.statista.com/statistics/1109288/global-revenue-of-costco-by-region/

https://statista.com/statistics/233725/development-of-amazon-web-services-revenue/#:~:text=Amazon%20Web%20Services%3A%20Annual%20revenue%202013-2020&text=In%202020%2C%20Amazon%20Web%20Services,dollars%20with%20its%20cloud%20services.

https://www.entrepreneur.com/article/366056

https://www.statista.com/statistics/478176/google-public-cloud-revenue/

https://blog.iotechnologies.com/as-amazon/

https://www.comparably.com/companies/amazon/mission

https://www.thevalueengineers.nl/amazons-empire/

https://www.amazon.jobs/en/principles

https://backlinko.com/amazon-prime-users

https://www.lexisclick.com/blog/customer-obsession-the-secret-to-amazons-success

https://www.contino.io/insights/whos-using-aws

References

https://cxl.com/blog/product-recommendations/

Appendix

Appendix - Technical Solution Summary

Comparison for all 3 analysis:

- **RFM analysis:**
  - more useful as a **data transformation tool** rather than a **decision engine**.
  - Knowing the RFM score is useful but doesn't translate well to a business decision by itself.
- **K-Means clustering:**
  - useful to divide the data into groups. The groups themselves do not provide meaningful decisions
  - data points from the cluster might be helpful if we passed them further to combine with another model.
- **Item-Based Collaborative filtering:**
  - one of the most popular recommendation algorithms in use
  - logic is easy to explain to business stakeholders and has proven results across a variety of industries.
  - does not explain its individual decisions very well.

What will you do when all 3 methods don't produce relevant results?

- All 3 methods are unsupervised that provide insights but don't really help make new decisions in a prescriptive or cognitive manner.
- Include the outputs of these unsupervised algorithms in a supervised algorithm for making product recommendations.
- Initial data works with supervised algorithm but add features like RFM score and cluster should drastically improve model performance.

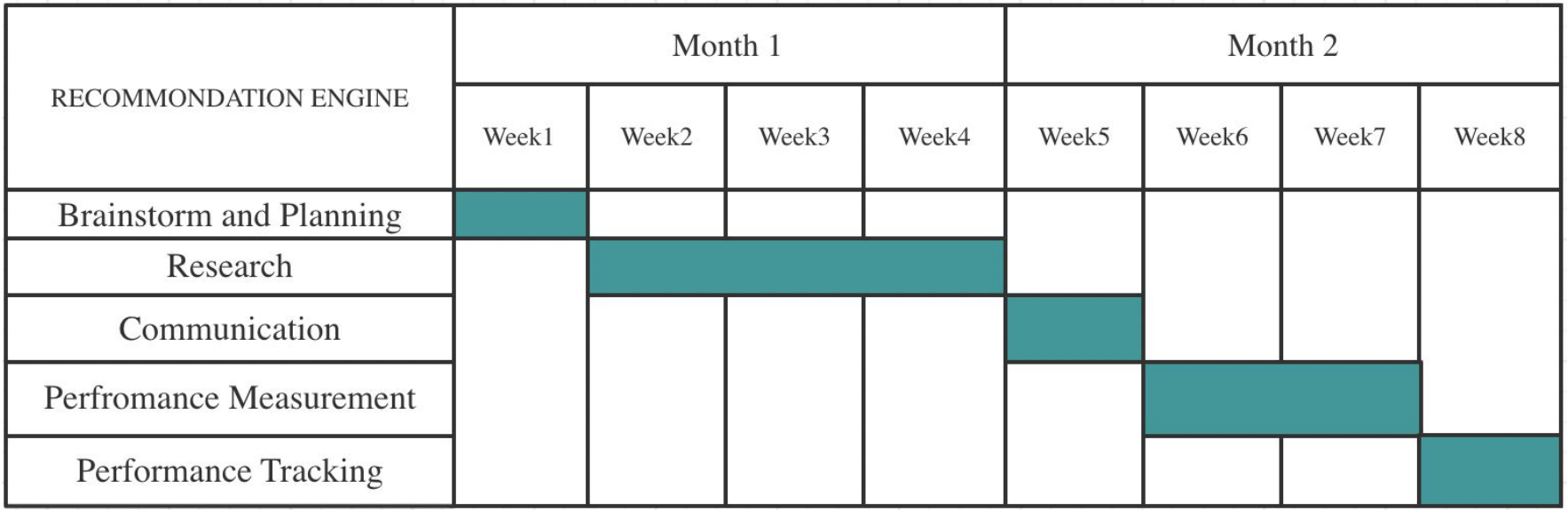
What are some other recommendation techniques?

- **Market-Basket Analysis (Apriori):** One of the more popular recommendation algorithms is market-basket. This would be more relevant to our use-case because it fits the context of our analysis. It creates a network of likes and dislikes.
- **Reinforcement Learning:** it creates a sense of context for the algorithm to operate. Adding that context will help us make a more precise recommendation for the customer.

What other data set can you use for this project?

- Search history from the platform (or externally), help boost the precision scores of our recommendations.
- External Social Media data
- Customer profile: Age, location, education, job…

Gantt Chart


RECOMMONDATION ENGINE
Month 1
Month 2
Week1
Week2
Week3
Week4
Week5
Week6
Week7
Week8
Brainstorm and Planning
Research
Communication
Perfromance Measurement
Performance Tracking