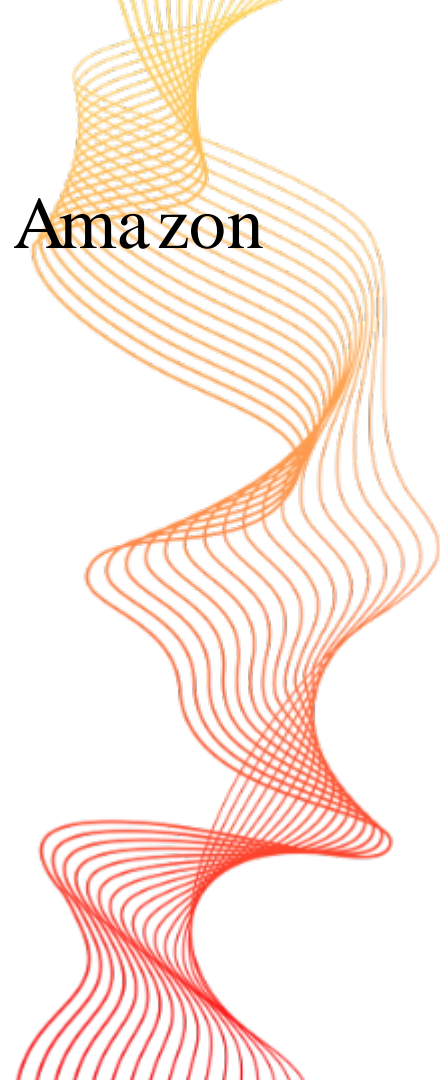




# Creating a Recommendation System for Amazon Marketing Team

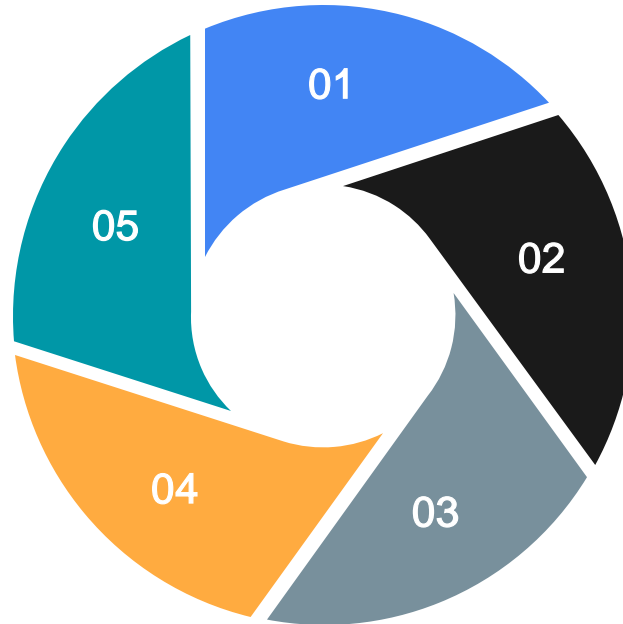
This presentation provides an overview of a project to create a recommendation system for the Amazon marketing team to improve post-purchase email engagement.



# Overview

A collaborative approach was taken, using existing ratings to make recommendations.

The project focuses on post-purchase email engagement for beauty products.



We are a team of internal data scientists at Amazon.

The project aims to create a recommendation system for the Amazon marketing team to send targeted recommendation e-mails.

Personalized, relevant, and timely recommendations are important for customer re-engagement.



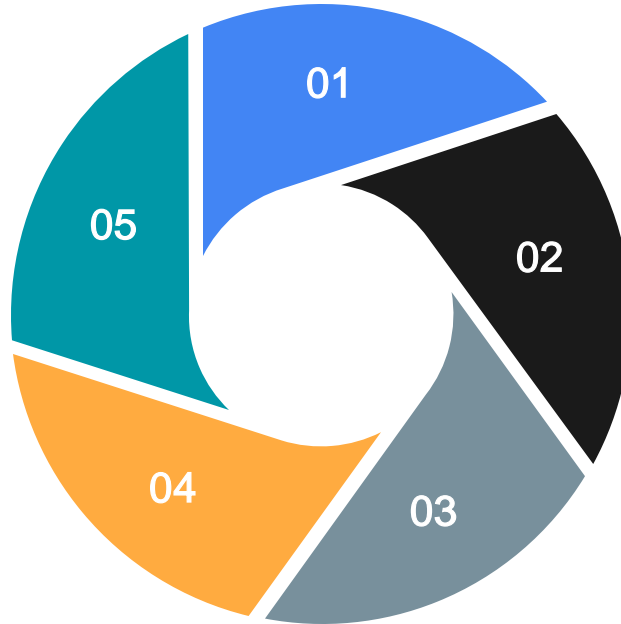
## Business Problem

- Amazon's marketing team wants to improve postpurchase email engagement.
- Customers open postpurchase emails 17% more often than other automated emails.
- Reengaging customers at post-purchase stage can lead to future purchases.
- Content in postpurchase emails must be engaging and relevant.
- A recommendation system can help select personalized product recommendations for customers.

# Data Understanding and Preparation

Reviews were given on a scale of 1-5, with a majority of reviews receiving a rating of 5.

Review data contained 22,363 reviewers and 12,101 unique products.



Data was collected from Amazon Beauty product reviews and meta data.

Two separate JSON files were used: 5-core dataset and dataset with 198,502 reviews.

The smaller dataset was used for this project.



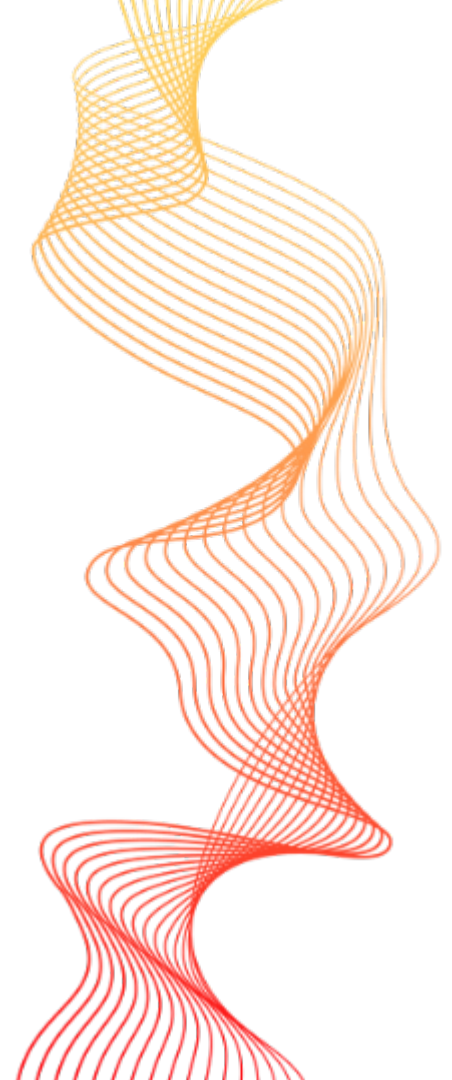
## Review Distributions

- A majority of reviews received under 10 ratings.
- Distribution of product reviews varies.
- The goal is to make accurate recommendations despite the varying review distributions.
- The recommendation system should not recommend products that the customer has already purchased.
- Review distributions impact model performance.



# Presentation: Reviews per Product

This presentation focuses on the distribution of reviews per product and the recommendation system used to return items in the Beauty subcategory.





# Reviews per Product

Majority of users rated under  
10 products

Distribution of the number of  
reviews completed by each  
user

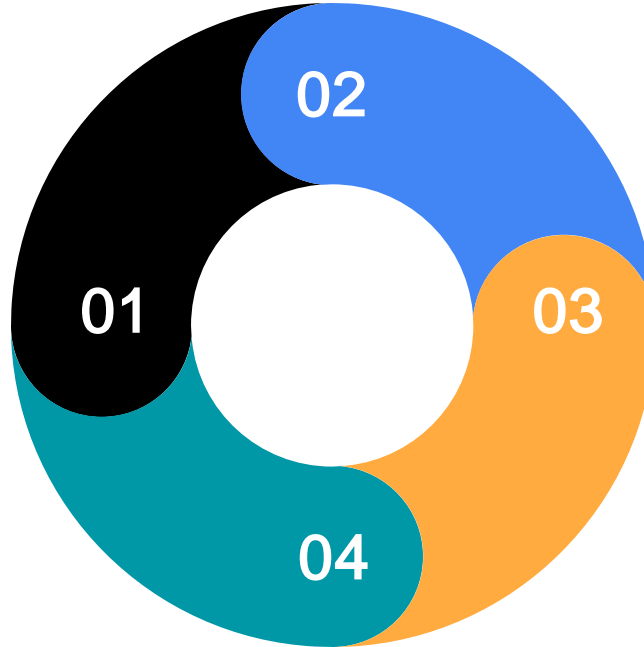
Insights about user  
engagement and activity



# Ratings per User

259,204 unique products  
in Meta Data

Focused on Beauty  
subcategories for  
recommendation system



Identified six  
subcategories: Skin Care,  
Tools & Accessories,  
Makeup, Hair Care, Bath &  
Body, and Fragrance

Importance of  
subcategories for  
personalized  
recommendations

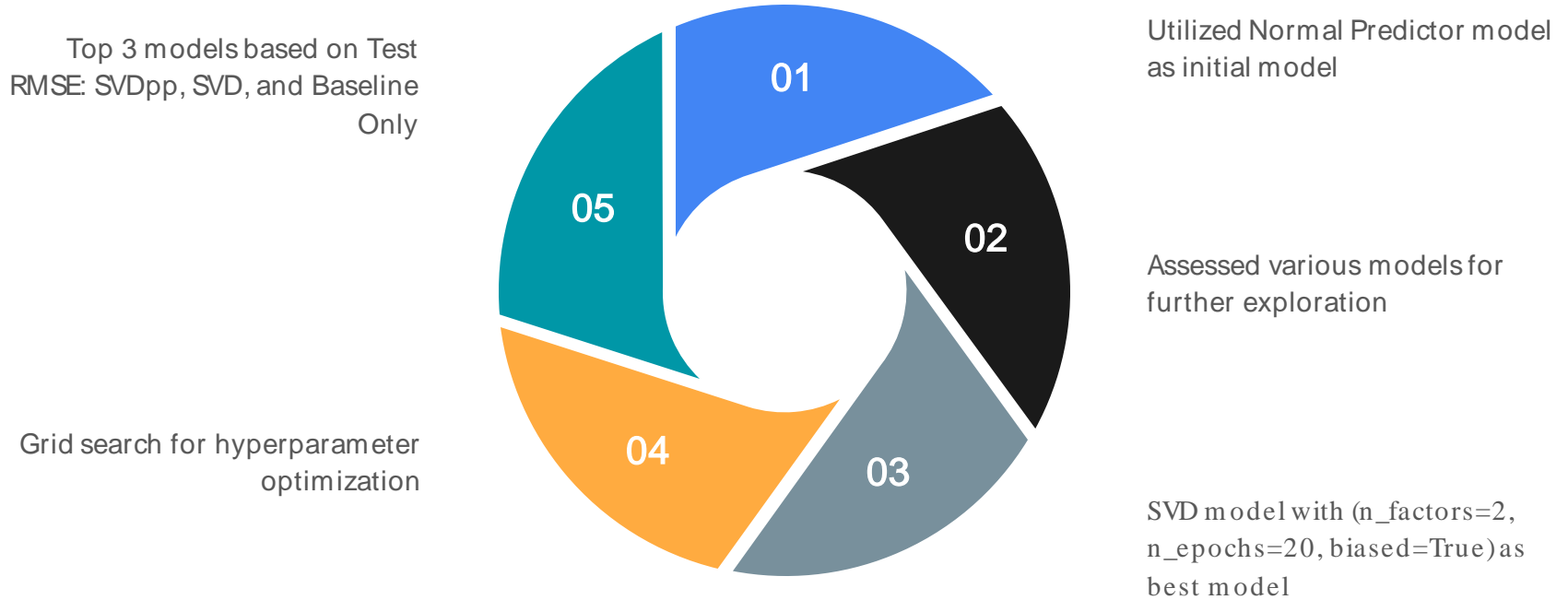




## Data Cleaning and Selection

- Selected columns for surprise model: reviewerID, asin, and overall
- Unique reviewer ID, product ID, and rating
- Scale of 15 for overall rating
- Importance of relevant data selection for accurate predictions

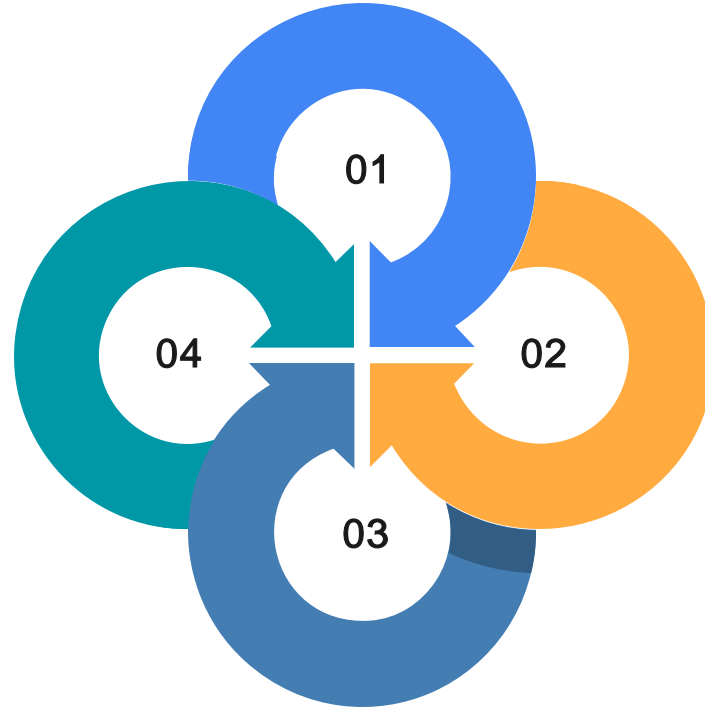
# Model Selection and Evaluation



# Final Collaborative Filtering Models

Order recommendations by  
highest predicted\_rating

Input reviewerID and number  
of recommendations



Returns ASIN, Product Name,  
Description, Image, and  
predicted\_rating

Facilitates personalized  
recommendations for users

# Additional Final Model

01

Useful for promoting specific items at certain times

02

Input reviewerID, number of recommendations, and product category

03

Returns ASIN, Product Name, Description, and Image

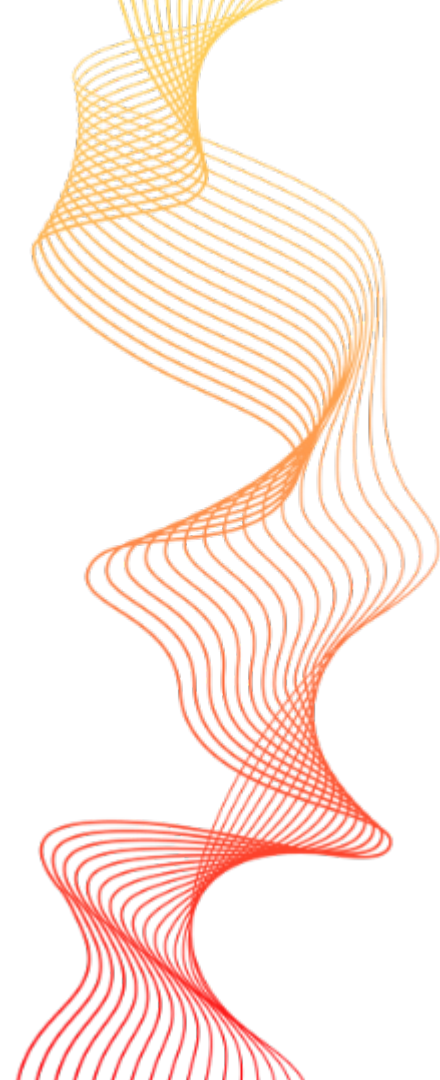
04

Enhances targeted marketing strategies



# Final Model Evaluation

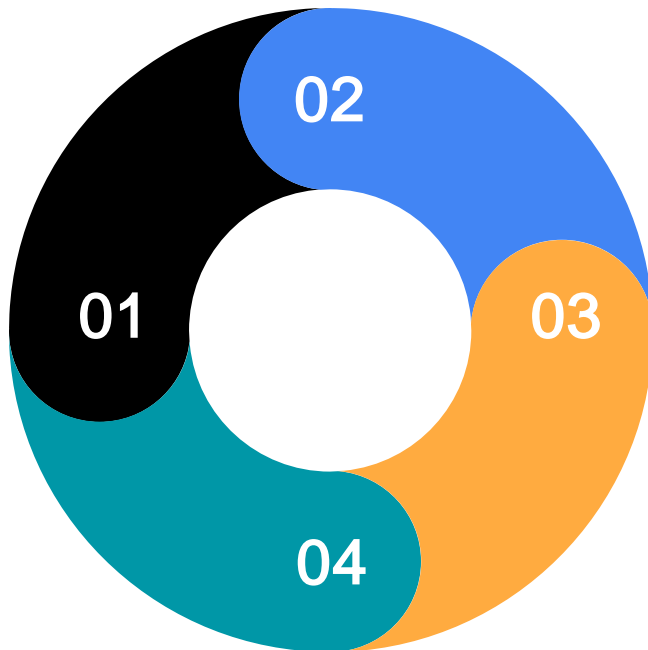
This presentation provides an overview of the final recommendation model and its evaluation.



# Model Evaluation

The final recommendation model using SVD yielded a RMSE of 1.08.

This means that, on average, our predicted review scores for Amazon buyers were 1.08 points off of the true value of review scores.




This score is more than half a point drop from our baseline model.

On a review scale of 1-5, this improvement is considered significant.



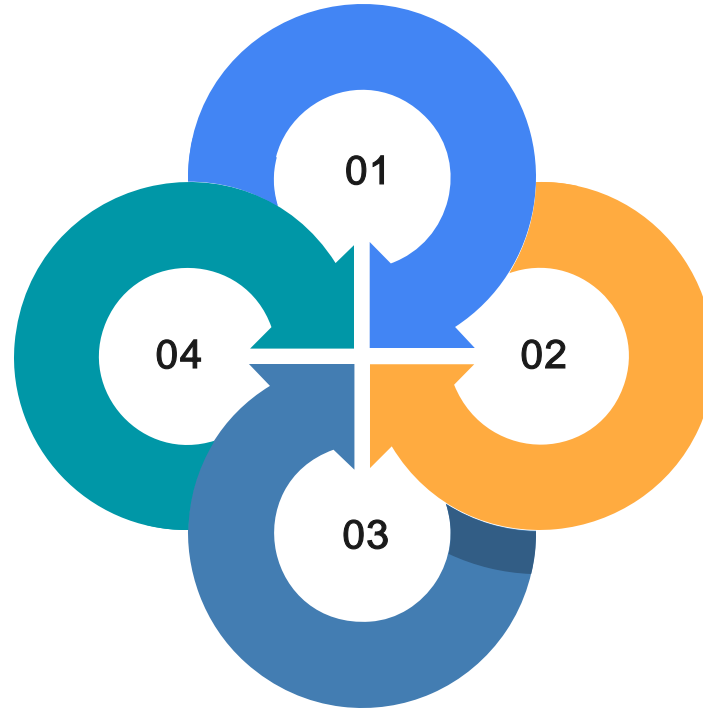
## Personalization Features

- No repeat products: The model will not recommend items that the buyer has already purchased.
  - Prioritizes the best match for the buyer: It will deliver the top predicted reviews for a buyer.
  - Subcategory filtering: Allows filtering beauty products based on six subcategories.
  - Image retrieval: Converts URL to an image and delivers alongside recommended titles and descriptions.
- 

# Limitations

Cold start problem: Prior reviews from users are needed to offer recommendations.

Misclassified products: Dataset often miscategorizes products in subcategories, leading to recommendations of misclassified products.



Skewed dataset: Nearly 60% of all reviews were rated 5 points, which may affect the accuracy of lower rated items.

Indiscriminate reviewers: The model does not handle reviewers who rate all products the same.

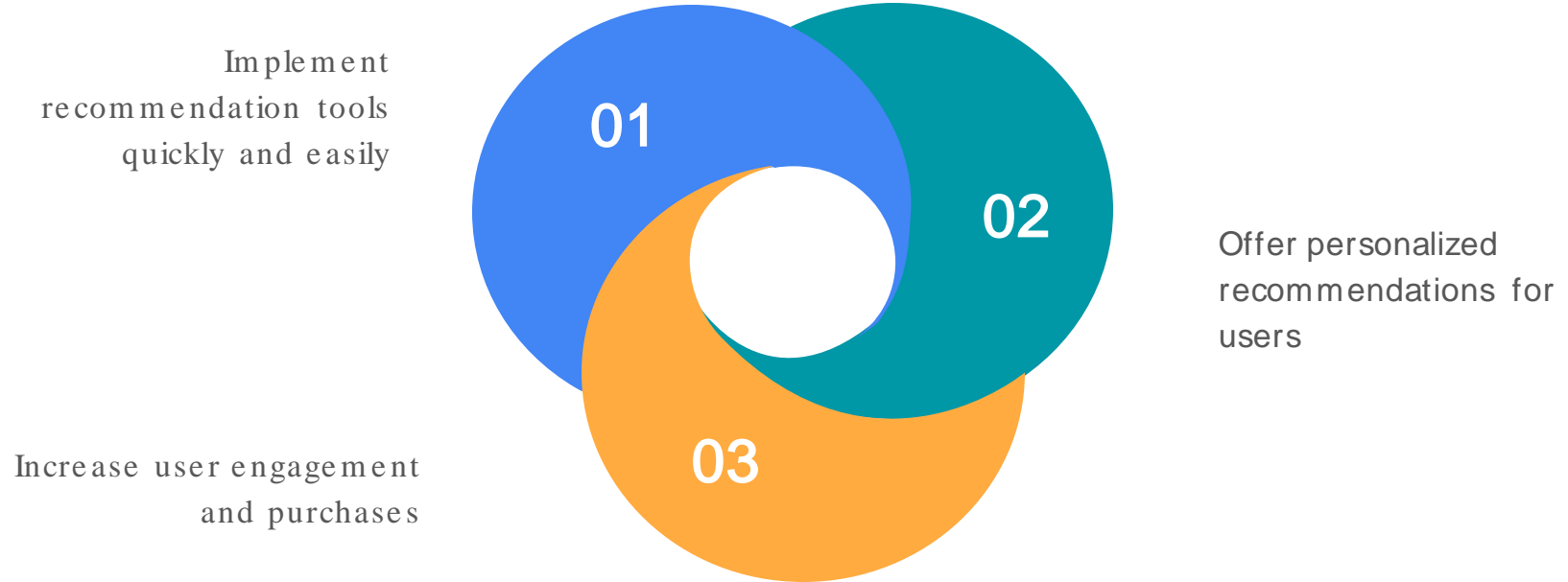




## Next Steps

- Investigate dataset bias: Determine if high ratings are consistent with Amazon buyer behavior.
- Address indiscriminate reviewers: Send a separate survey to these buyers to capture their preferences.
- Solve the cold start problem: Add a contentbased approach to offer recommendations without prior reviews.
- Implement standardized classification: Improve the accuracy of subcategory predictions by implementing a standardized

# Conclusion





Thank you for your time and attention 😊