

Wine Recommendation

Bienvenu Thibault
Jun Heesu
Kim Yoonhee
Mahdi Tarraf-Kojok
Thomas
Yang Suyeong



INDEX

Topic Description

Recommendation Process

Dataset & Data Processing

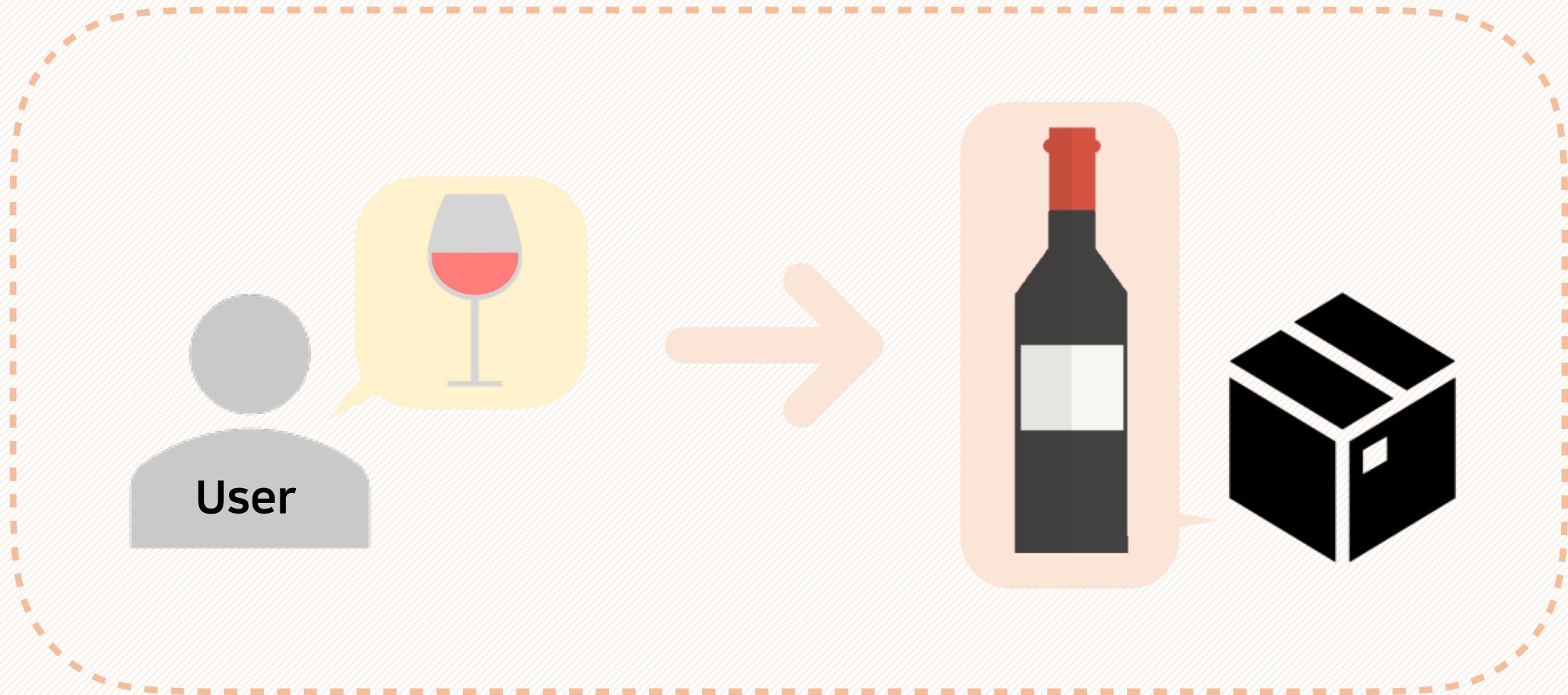
Recommendation Algorithm

Evaluation



Topic Description

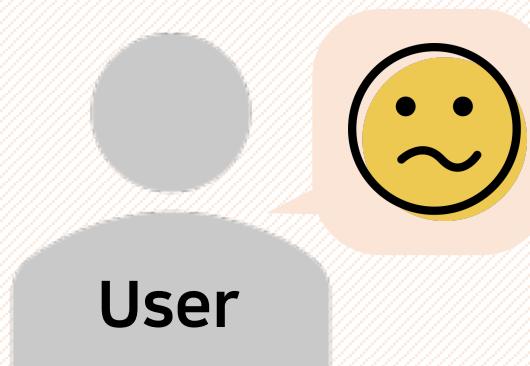
Recommend Wine





Topic Description

Recommend Wine





Topic Description

Recommend Wine

Brut

Demi-Sec

Chablis

Riserva

Viognier



User



Topic Description

Recommend Wine

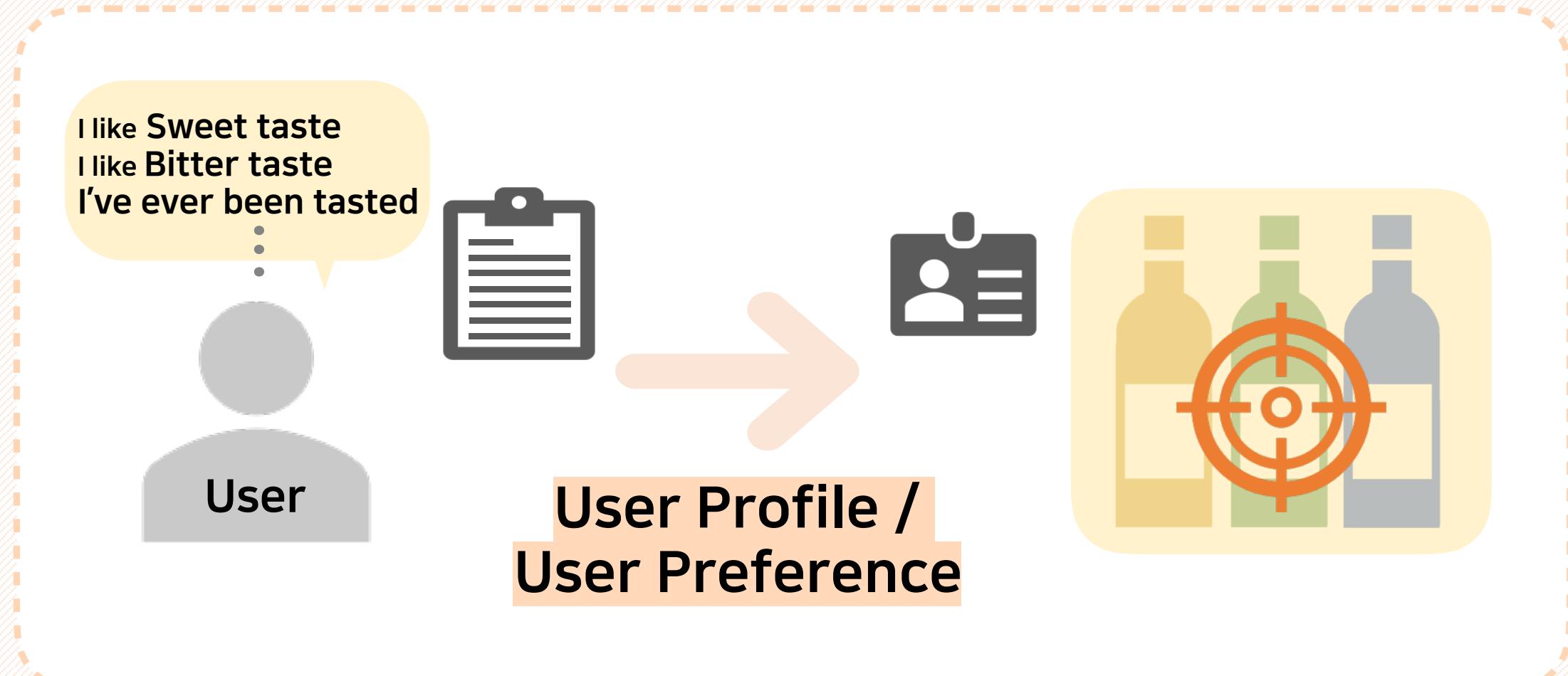
Variety, Vintage, Country, Title
Price, Province, Description



User

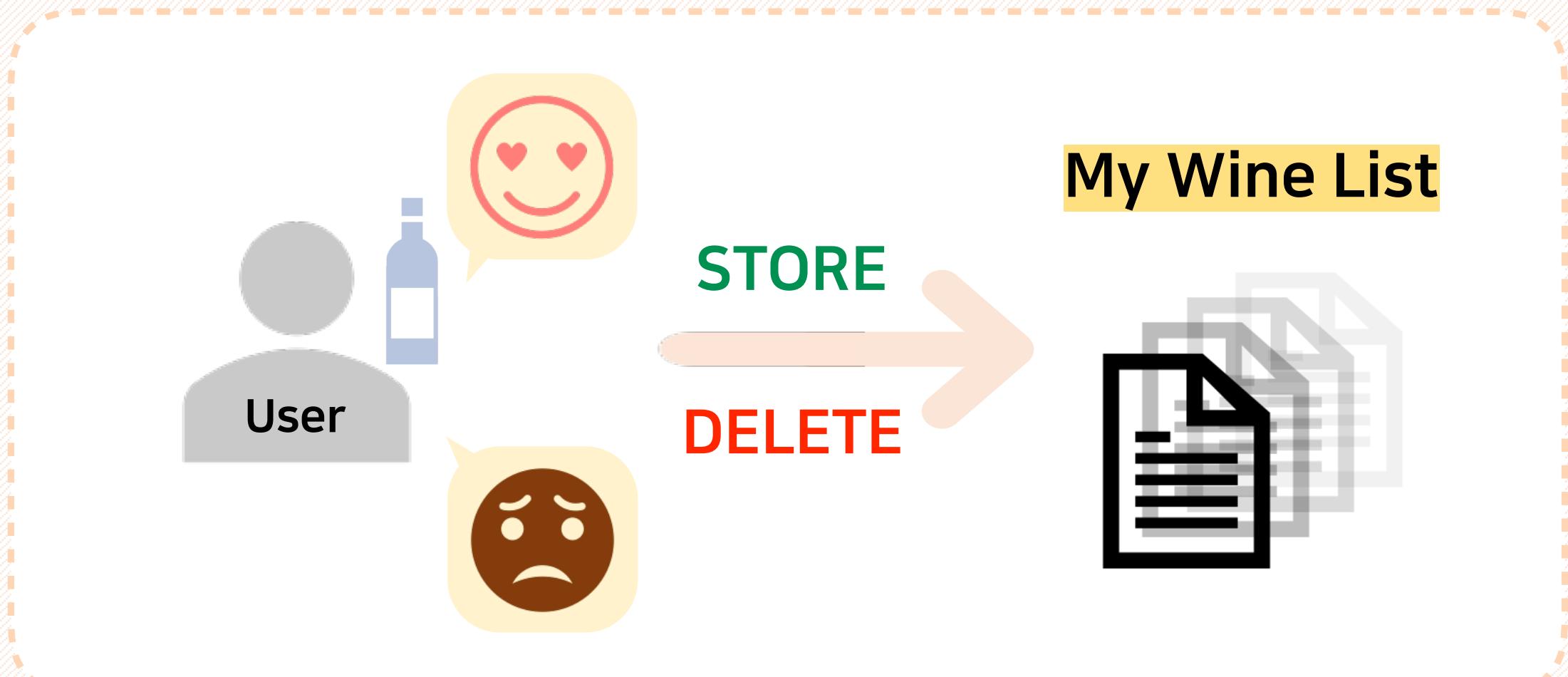


Recommendation Process



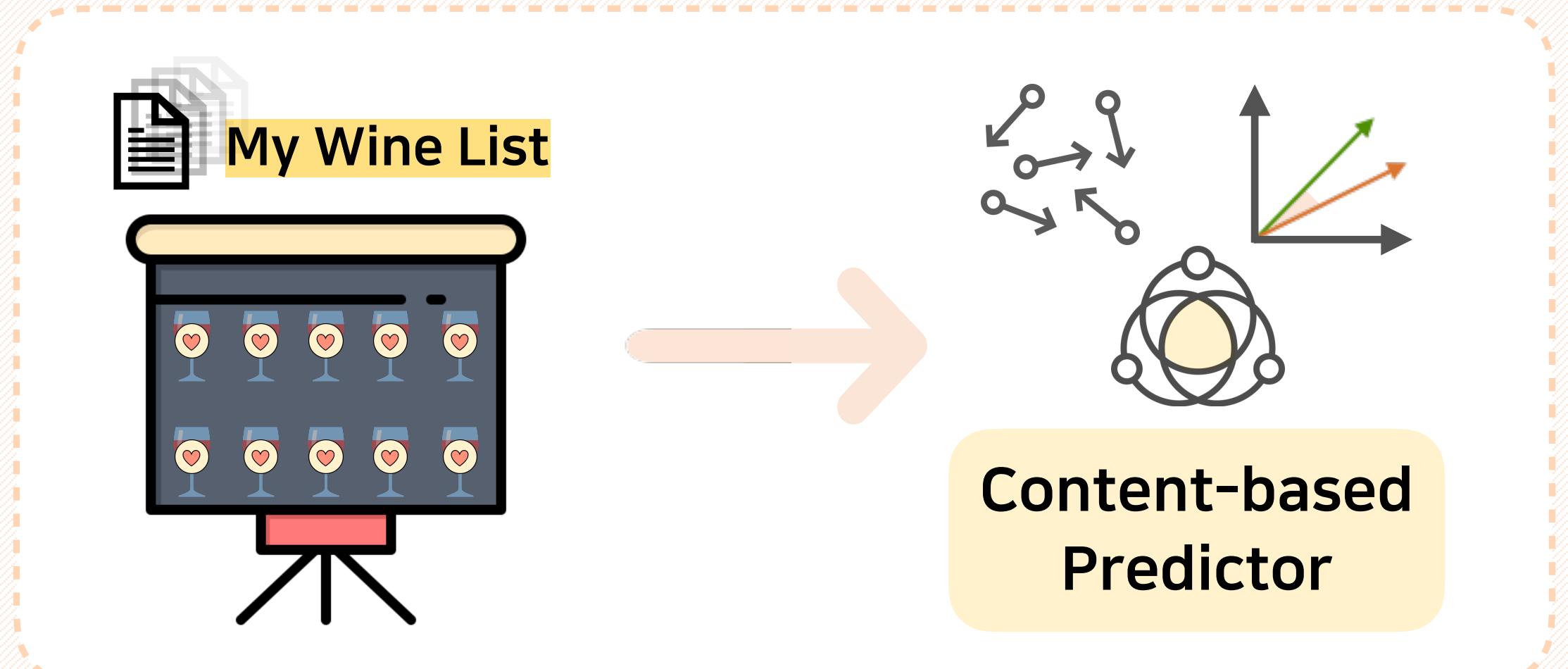


Recommendation Process





Recommendation Process





Dataset & Data - Processing

24998 kinds of wine

	vintage	country	country_lookup	continent	county	
1	1919-01-01	Spain	Kingdom of Spain	Europe	Cava	
2	1929-01-01	Italy	Italy	Europe	Vernaccia di San Gimignano	
3	1929-01-01	Italy	Italy	Europe	Sangiovese di Romagna Superiore	
4	1934-01-01	Portugal	Portugal	Europe	No data.	
5	1945-01-01	France	France	Europe	Rivesaltes	
6	1952-01-01	Portugal	Portugal	Europe	No data.	
7	1963-01-01	Portugal	Portugal	Europe	No data.	
8	1974-01-01	Portugal	Portugal	Europe	No data.	
9	1978-01-01	Portugal	Portugal	Europe	No data.	
10	1980-01-01	Portugal	Portugal	Europe	No data.	
11	1984-01-01	Portugal	Portugal	Europe	No data.	
12	1987-01-01	US	United States of America	North America	Sonoma Valley	
13	1989-01-01	France	France	Europe	Champagne	
14	1989-01-01	Portugal	Portugal	Europe	No data.	
15	1990-01-01	France	France	Europe	Champagne	
16	1991-01-01	US	United States of America	North America	Sonoma County	
17	1992-01-01	US	United States of America	North America	Carneros	
18	1992-01-01	US	United States of America	North America	Anderson Valley	
19	1992-01-01	Austria	Austria	Europe	No data.	

About this file

LAST UPDATED 11 months ago

OWNER @markpowell

CREATED 11 months ago

SIZE 1.89 MB

Displaying 12 columns, 24,997 rows in table **wines**

vintage ⓘ
Year that the grapes were harvested

country ⓘ

country_lookup ⓘ

continent ⓘ

county ⓘ

Growing region

designation ⓘ

points ⓘ
95-100 Classic: a great wine 90-94 C ?

Vintage Country Province Price Title Variety

<https://data.world/markpowell/global-wine-points/workspace/file?filename=Wines.xlsx>



Dataset & Data - Processing

Province?



Variety?

Vintage?

Newbie of wine **cannot be familiar** of terms of wine



Dataset & Data - Processing

120,000 kinds of wine-descriptions

The screenshot shows the Kaggle interface for the "Wine Reviews" dataset. At the top, there's a navigation bar with "kaggle" and a search bar. Below the search bar are links for Competitions, Datasets, Notebooks, Discussion, Courses, and more. On the right side of the header are a bell icon and a user profile icon.

The main content area features a large image of purple grapes. Overlaid on the image are the dataset title "Wine Reviews", a subtitle "130k wine reviews with variety, location, winery, price, and description", and the author information "zackthoutt • updated 2 years ago (Version 4)".

Below the image, there are tabs for "Data" (which is selected), "Kernels (2,281)", "Discussion (24)", "Activity", and "Metadata". There are also buttons for "Download (174 MB)" and "New Notebook".

A message box indicates "Your Dataset download has started. Show your appreciation with an upvote" and shows a count of 1870 upvotes, accompanied by a series of small profile icons.

At the bottom of the main section, there are metrics: "Usability 7.9", "License CC BY-NC-SA 4.0", and "Tags food and drink, alcohol, critical theory".

The page is divided into two main sections: "Description" and "Context". The "Description" section is currently empty. The "Context" section contains a detailed narrative about the creator's motivation for creating the dataset, mentioning the documentary "Somm" and their plan to use deep learning to predict wine variety based on descriptions.

<https://www.kaggle.com/zynicide/wine-reviews/data>



Dataset & Data - Processing

Screenshot of a Kaggle dataset page for "Wine Reviews".

The page features a header with the Kaggle logo, a search bar, and navigation links for Competitions, Datasets, Notebooks, Discussion, Courses, and more. A notification bell icon and user profile picture are also present.

The main content area displays the dataset title "Wine Reviews" and a description: "130k wine reviews with variety, location, winery, price, and description". It shows a photo of dark grapes and a post by user "zackthoutt" updated 2 years ago (Version 4).

Below the title, there are tabs for Data (selected), Kernels (2,281), Discussion (24), Activity, and Metadata. A "Download (174 MB)" button and a "New Notebook" button are also visible.

A message indicates "Your Dataset download has started. Show your appreciation with an upvote" and shows a count of 1870 upvotes. Below this, there are icons for users to upvote.

At the bottom, there are metrics: Usability 7.9, License CC BY-NC-SA 4.0, and Tags food and drink, alcohol, critical theory.

The page includes sections for "Description" and "Context". The "Context" section contains a detailed narrative about the user's motivation to create a predictive model for wine identification based on blind tasting, drawing inspiration from the documentary "Somm".

Vintage Country Province Price Title Variety





Dataset & Data - Processing

Description of taste

description

Aromas include tropical fruit, broom, brimstone and dried herb. The palate isn't overly expressive, offering unripened apple, citrus and dried sage alongside brisk acidity.

This is ripe and fruity, a wine that is smooth while still structured. Firm tannins are filled out with juicy red berry fruits and freshened with acidity. It's already drinkable, although it will certainly be better from 2016.

Tart and snappy, the flavors of lime flesh, and rind dominate. Some green pineapple pokes through, with crisp acidity underscoring the flavors. The wine was all stainless-steel fermented.

Pineapple rind, lemon pith and orange blossom start off the aromas. The palate is a bit more opulent, with notes of honey-drizzled guava and mango giving way to a slightly astringent, semidry finish.

Fresh and tannic, with rustic, earthy, herbal characteristics. Nonetheless, if you think of it as a pleasantly unfussy country wine, it's a good companion to a meal.

Herbs and, in this case, horseradish. In the mouth, this is fairly full bodied, with tomatoey acidity. Spicy, herbal flavors complement dark plum fruit, while

Here's a bright, informal red that opens with aromas of candied berry, white pepper and savory herb that carry over to the palate. It's balanced with fresh acidity and soft tannins.

This dry and restrained wine offers spice in profusion. Balanced with acidity and a firm texture, it's very much for food.

Savory dried thyme notes accent sunnier flavors of preserved peach in this brisk, off-dry wine. It's fruity and fresh, with an elegant, sprightly footprint.

This has great depth of flavor with its fresh apple and pear fruits and touch of spice. It's off dry while balanced with acidity and a crisp texture. Drink now.

Soft, supple plum envelopes an oaky structure in this Cabernet, supported by 15% Merlot. Coffee and chocolate complete the picture, finishing strong at the end, resulting in a value-priced wine of attractive flavor and immediate appeal.

This is a dry wine, very spicy, with a tight, taut texture and strongly mineral character layered with citrus as well as pepper. It's a food wine with its almost crisp aftertaste.

Slightly reduced, this wine offers a chalky, tannic backbone to an otherwise juicy explosion of rich black cherry, the whole accented throughout by firm oak and cigar box.

This is dominated by oak and oak-driven aromas that include roasted coffee bean, espresso, coconut and vanilla that carry over to the palate, together with plum and chocolate. Astringent, drying tannins give it a rather abrupt finish.

Building on 150 years and six generations of winemaking tradition, the winery trends toward a leaner style, with the classic California buttercream aroma cut by tart green apple. In this good everyday sipping wine, flavors that are both approachable and complex.

Zesty orange peels and apple notes abound in this sprightly, mineral-toned Riesling. Off dry on the palate, yet racy and lean, it's a refreshing, easy quaffer with wide appeal.

Baked plum, molasses, balsamic vinegar and cheesy oak aromas feed into a palate that's braced by a bolt of acidity. A compact set of saucy red-berry and plum flavors features tobacco and peppery accents, while the finish is long and satisfying.

Raw black-cherry aromas are direct and simple but good. This has a juicy feel that thickens over time, with oak character and extract becoming more apparent. A flavor profile driven by dark-berry fruits and smoldering oak finishes with a touch of smoke.

Desiccated blackberry, leather, charred wood and mint aromas carry the nose on this full-bodied, tannic, heavily oaked Tinto Fino. Flavors of clove and woodspice sit on top of blackberry fruit, then hickory and other forceful notes come through in the finish.

Red fruit aromas pervade on the nose, with cigar box and menthol notes riding in the back. The palate is slightly restrained on entry, but opens up to riper notes of cherry and plum specked with crushed pepper. This blend of varietals is a solid wine.

taster_name

Kerin O'Keefe

Roger Voss

Paul Gregutt

Alexander Peartree

Paul Gregutt

Michael Schachner

Kerin O'Keefe

Roger Voss

Anna Lee C. Iijima

Roger Voss

Virginie Boone

Roger Voss

Virginie Boone

Kerin O'Keefe

Matt Kettmann

Anna Lee C. Iijima

Description of Wine

associate

Flavor of Variety



Dataset & Data - Processing

[Description]

“A full, ... with deep aromas of licorice, earth, tar, prunes and a lovely creamy note. Plum and blackberry flavors with soy and smoke accents fill the palate, and as the wine closes it displays good length with fruit and spice flavors.”

Extracting words about for intuitive understanding of wines for newbies



Dataset & Data - Processing

[Description]

“A full, ... with **deep aromas of licorice, earth, tar, prunes** and a lovely **creamy** note. **Plum and blackberry** flavors with **soy and smoke** accents fill the palate, and as the wine closes it displays good length with **fruit** and **spice** flavors.”

Extracting words about for intuitive understanding of wines for newbies



Dataset & Data - Processing

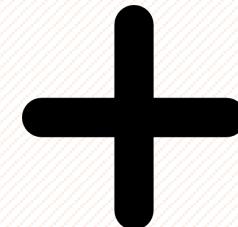
5 categories of FLAVOR

Acidity	Tannin	Body	Sweet	Fruit
Acid	Tannin	Full-Bodied	Sweetness	Grapefruit
Acidity	Auster	Big Body	Sweetish	Apple
	Astring		Sweetest Supersweet	Lemon



Dataset & Data - Processing

Acidity	Tannin	Body	Sweet	Fruit
Acid	Tannin	Full-Bodied	Sweetness	Grapefruit
Acidity	Auster	Big	Sweetish	Apple
	Astring	Body	Sweetest Supersweet	Lemon



Spicy, Earth, Smoky, Flowery, Herb

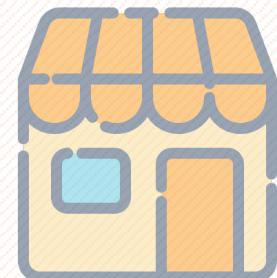


Dataset & Data - Processing

Result can ignore popularity of wine



**That is so rare !
I cannot find anywhere !**





Dataset & Data - Processing

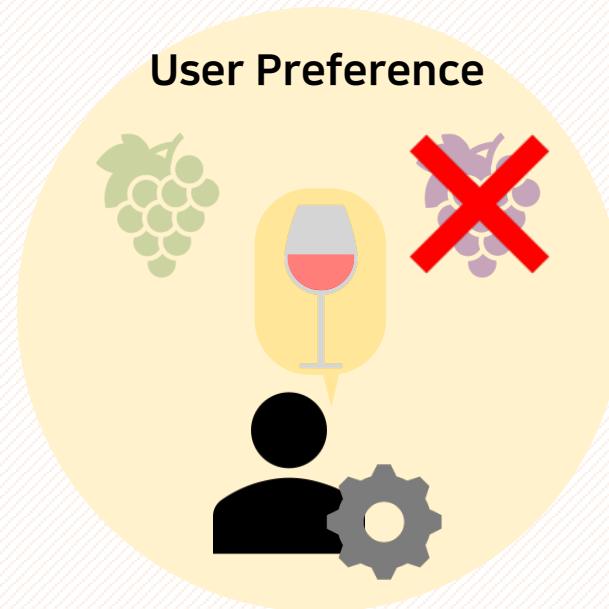
Create new dataset about **100 famous top wines**



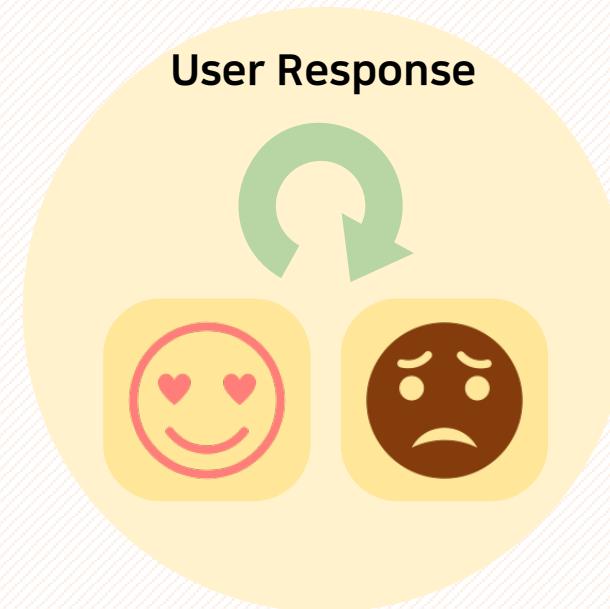


Algorithm Organization

Main Recommendation



**Knowledge-Based
Recommendation**

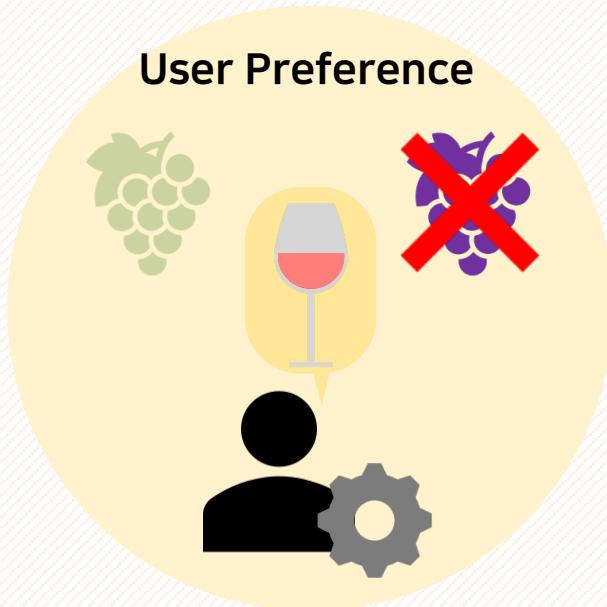


**Content-Based
Recommendation**



Algorithm Organization

Knowledge - Based



Question.

“Have you ever tasted wine before?”

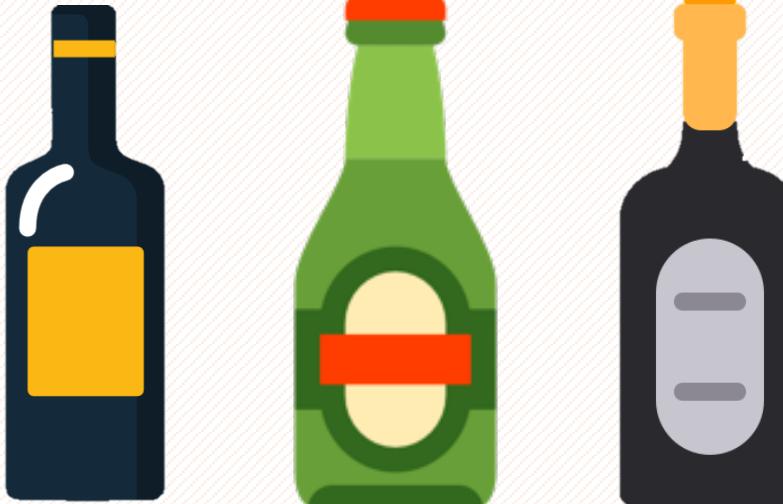


Algorithm Organization

Question.

“What is your three favorite wines?”

Answer.



Get user contents

Switched to **Content Based Recommendation**





Algorithm Organization

Question.

Which one do you like ? Sweet? Tannin?

Question.

Do you like acidity?

Question.

Do you like fruity or not?

⋮



Algorithm Organization

Question.

Which one do you like ? Sweet?

Question.

Do you like acidity?

Question.

Do you like fruity or not?

⋮



Extract user preference



Algorithm Organization

User Preference



25000 data



100 top famous wine

3 wines

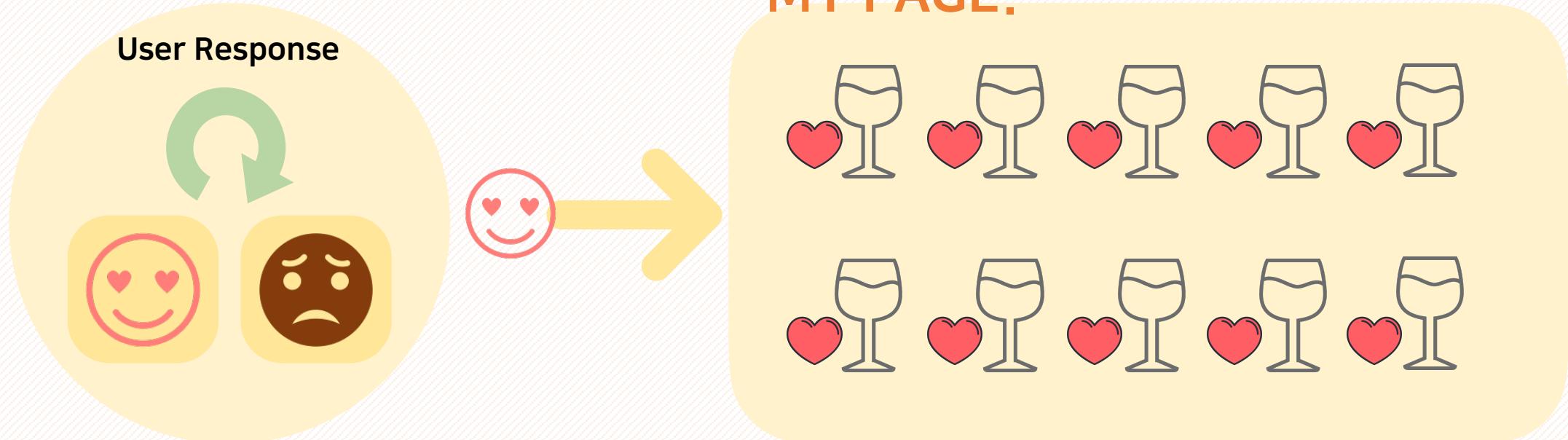
2 wines

List of 5 wines to recommend



Algorithm Organization

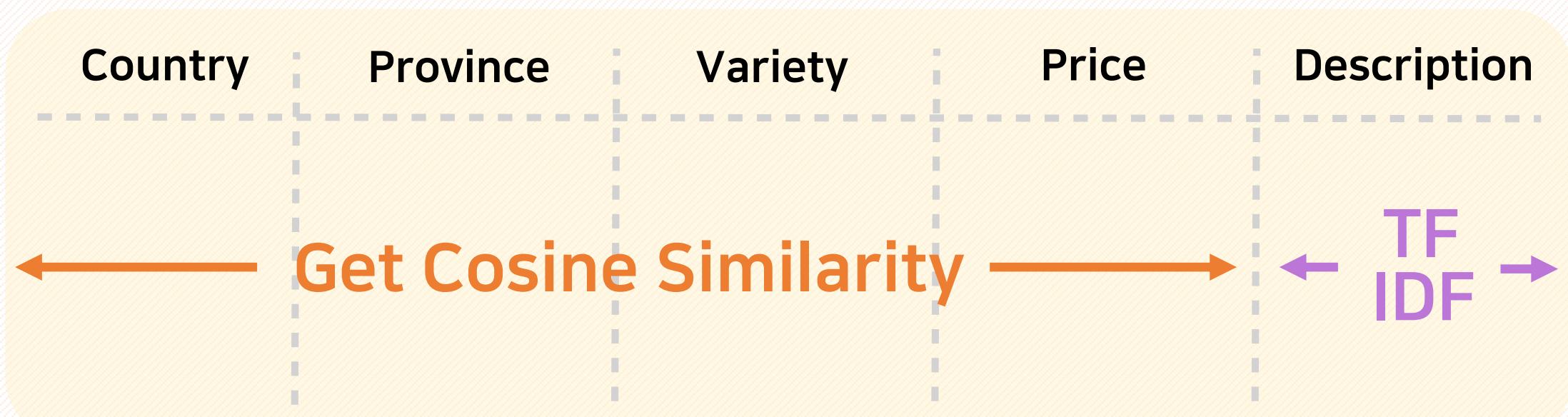
Content - Based





Algorithm Organization

Content - Based





Algorithm Organization

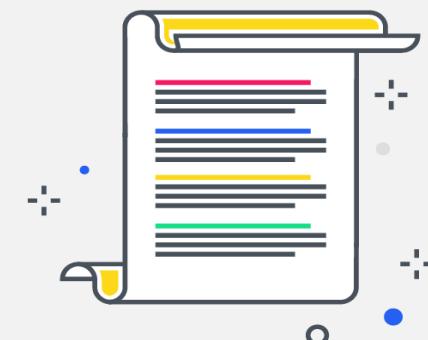
Content - Based

TF-IDF

Frequency of a term across multiple documents



Frequency of a term across a single document



Compare
TF_IDF score

$$w_{x,y} = tf_{x,y} \times \log \left(\frac{N}{df_x} \right)$$



Algorithm Organization

Content - Based

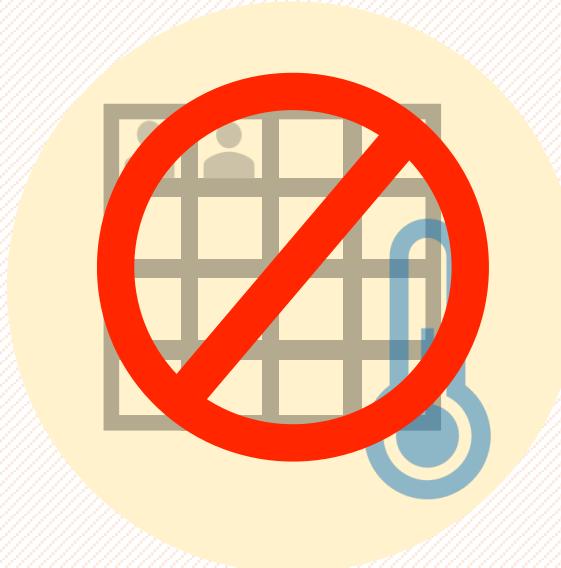
Final Score



$(\text{Cosine-Similarity} * 0.5) + (\text{TFID} * 0.5)$



Algorithm Organization



NO Cold Start



New (NO First Rater)



Unique tastes



Algorithm Organization



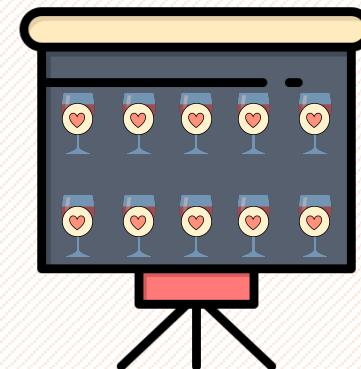
Beginner



Have Experience



Questions



My Wine List

Knowledge based



25000 data

3 wines



100 top famous wine

2 wines

Content based

Similar 5 wines



Evaluation

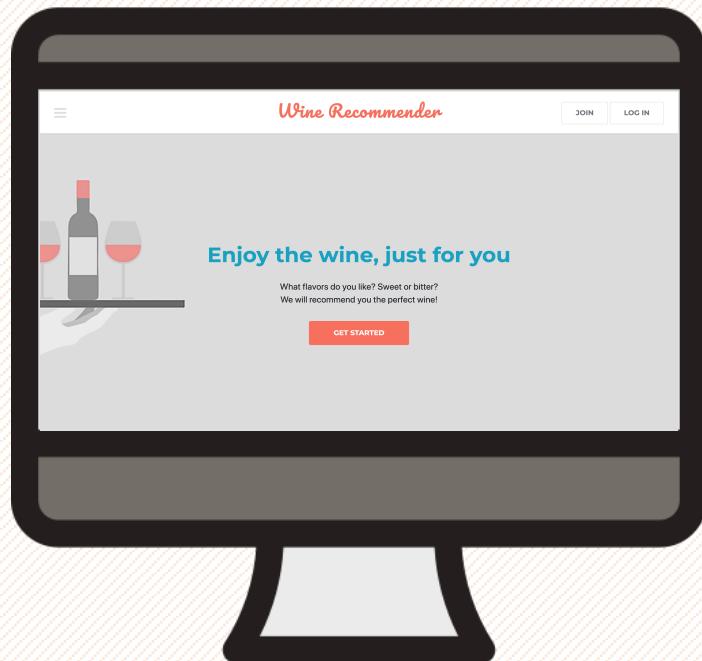


How can we evaluate our recommend system ?



Evaluation

How can we evaluate our recommend system ?



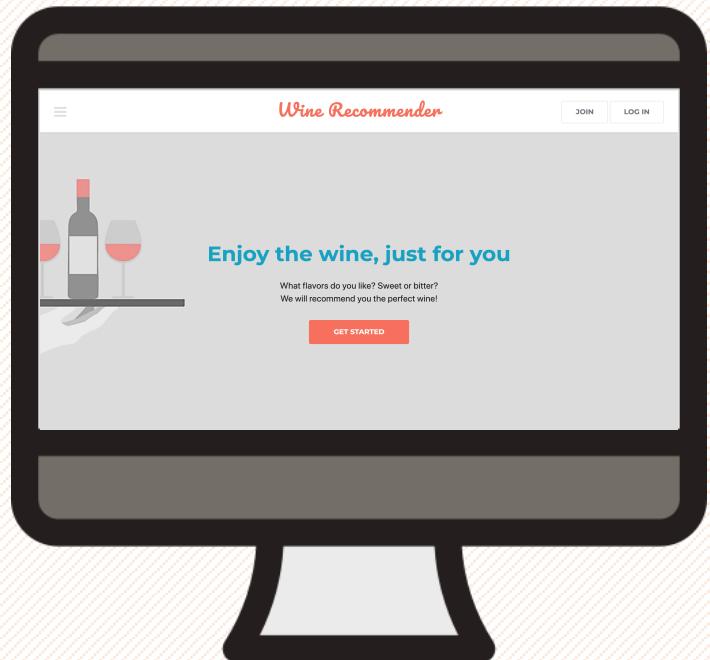
Answer.

**From the users !
Let's get evaluation like RATING**



Evaluation

How can we evaluate our recommend system ?



**Not enough
rating data**

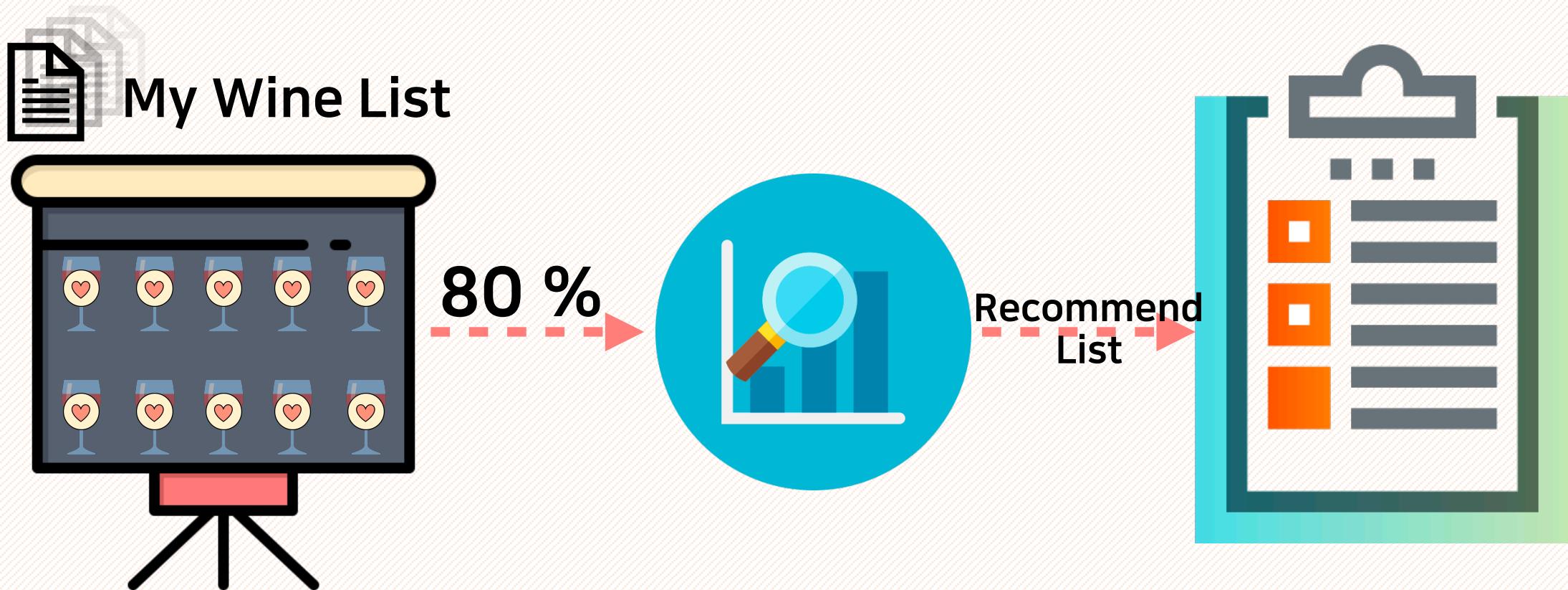


Not many users



Evaluation

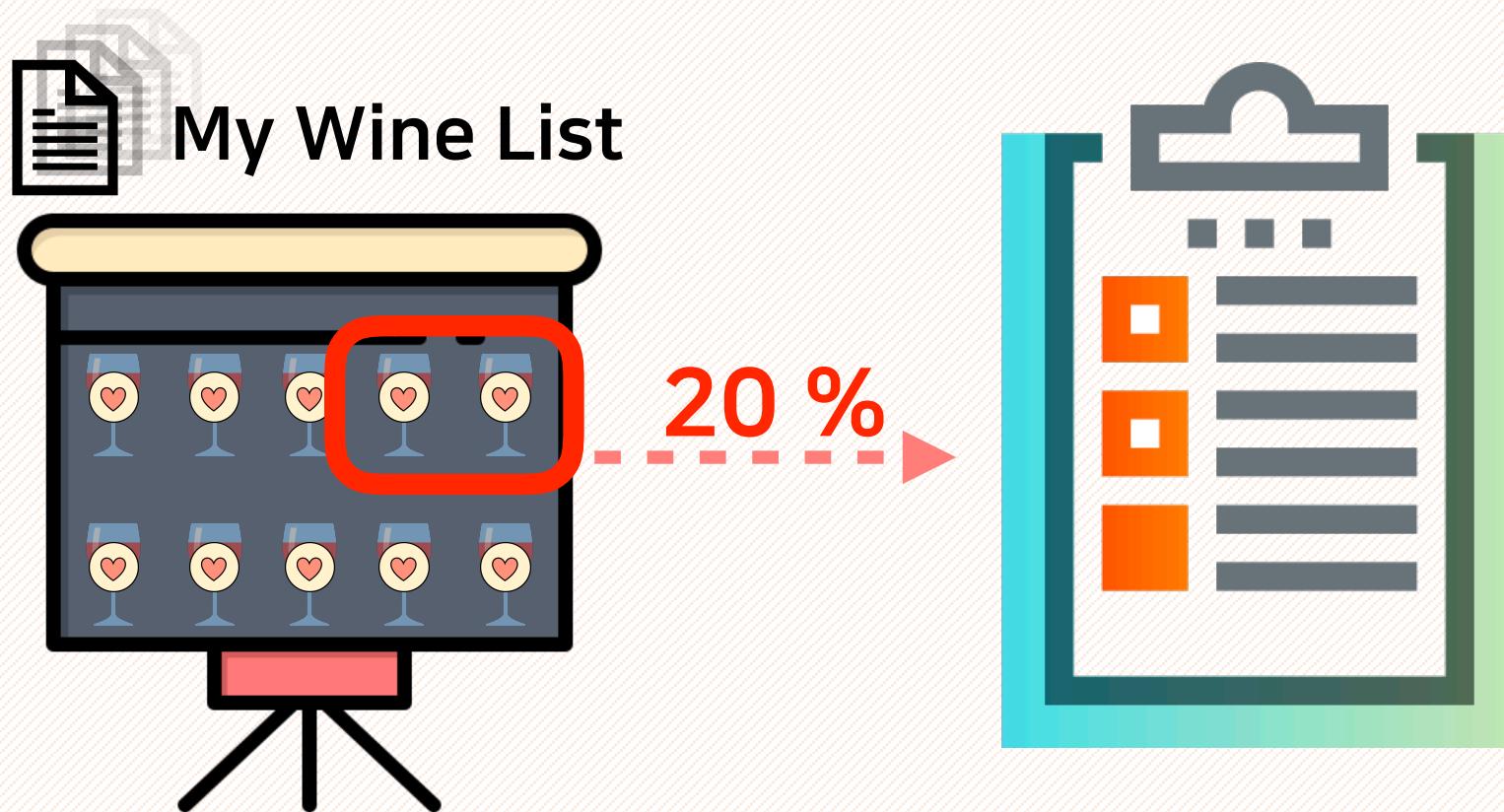
How can we evaluate our recommend system ?





Evaluation

How can we evaluate our recommend system ?





Thank you

<https://bi-wine-recommender.herokuapp.com/index.html>