

Wine Recommender -

Utilising Machine Learning &
NLP for Personalised Wine
Recommendations

Soh Sze Ron



Agenda

01

Background &
Problem Statement

02

Data Cleaning &
Exploratory Data Analysis

03

Amazon
Webscraping

04

Modeling

05

Conclusion &
Recommendations

06

Streamlit
Deployment

Background

The Emergence of Recommender Systems

- E-commerce platforms were one of the early adopters
- Music and video streaming services
- Social media platforms
- Rise of Big Data and Machine Learning / AI

Background

- Consumer challenges in finding suitable wines due to complexity and diversity
- 2 datasets from Kaggle (WineEnthusiast) in 2017 & 2020



Problem Statement

As a wine lover, my objective is to develop a tailored recommendation system that addresses individual preferences and streamlines the shopping experience. Leveraging machine learning and NLP methods, I strive to provide readily available personalized wine suggestions, elevating the process of discovering new wines.



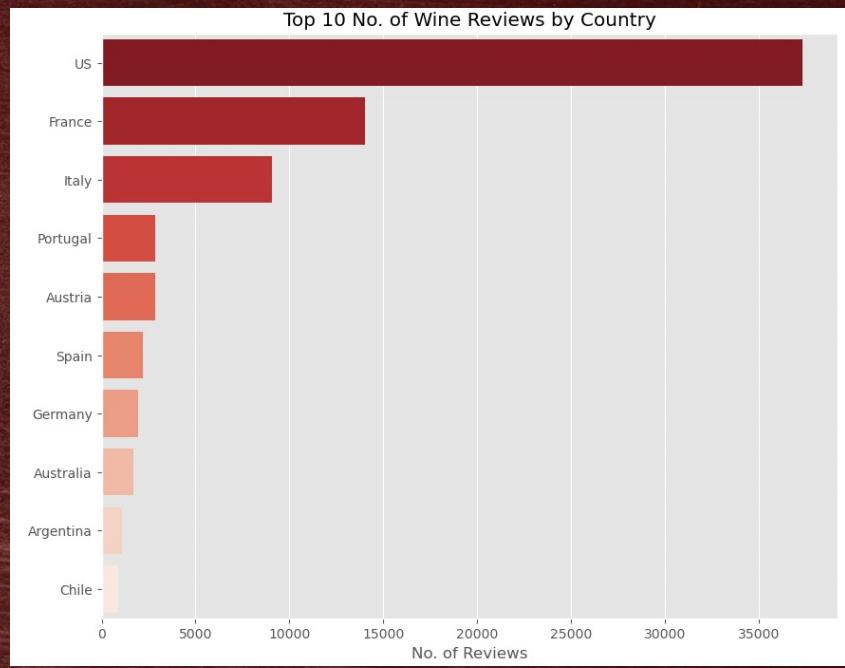
Data Cleaning

- Remove duplicates based on **description**
- Extract **region_1** missing values from **title**
- Drop null values in **country**, **province**, **region_1**, **variety** & **price**
- Impute null values in **taster_name** based on **winery** & **region_1**
- Remove observations with **variety** count of less than 100
- Remove observations below 90 **points**

country	description	designation	points	price	province	region_1	region_2	taster_name	taster_twitter_handle	title
Italy	Aromas include tropical fruit, broom,	Vulkà Bianco	87	NaN	Sicily & Sardinia	Etna		Kerin O'Keefe	@kerinokeefe	Nicosia 2013 Vulka Bianco [Etna]

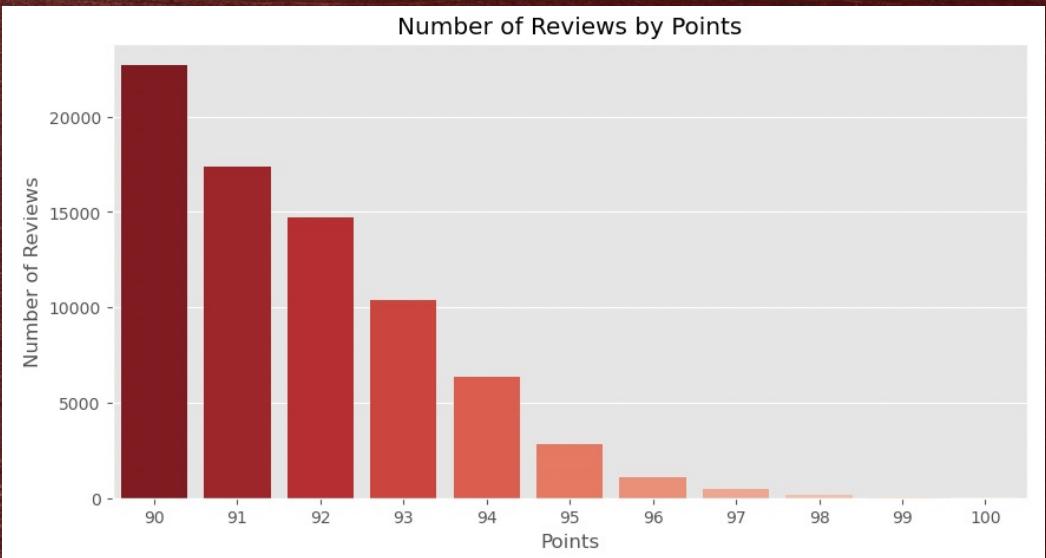
Exploratory Data Analysis

Wines from US, France & Italy make up 79% of our data



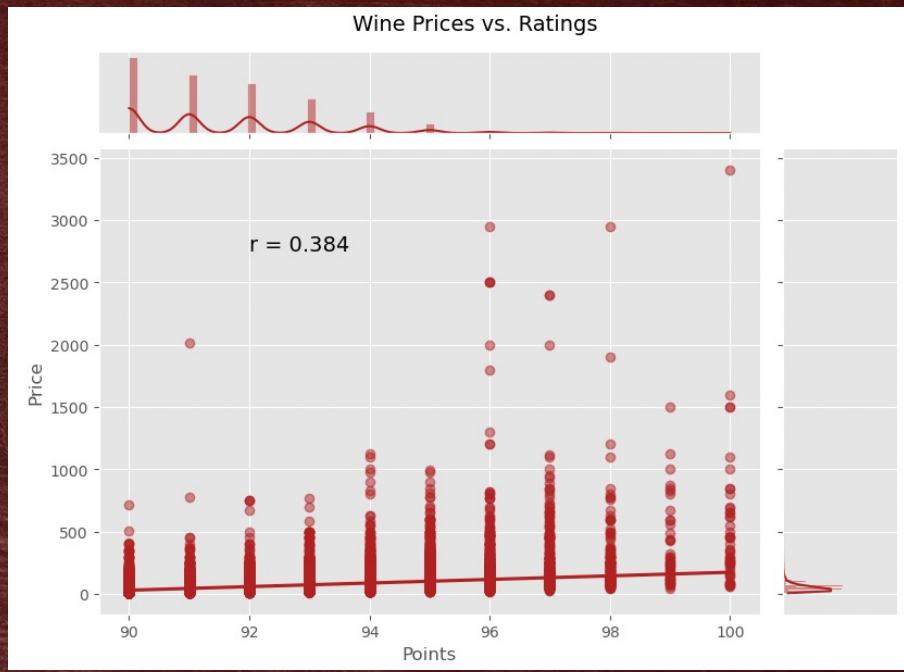
Exploratory Data Analysis

Heavily right-skewed with a median of 91 points



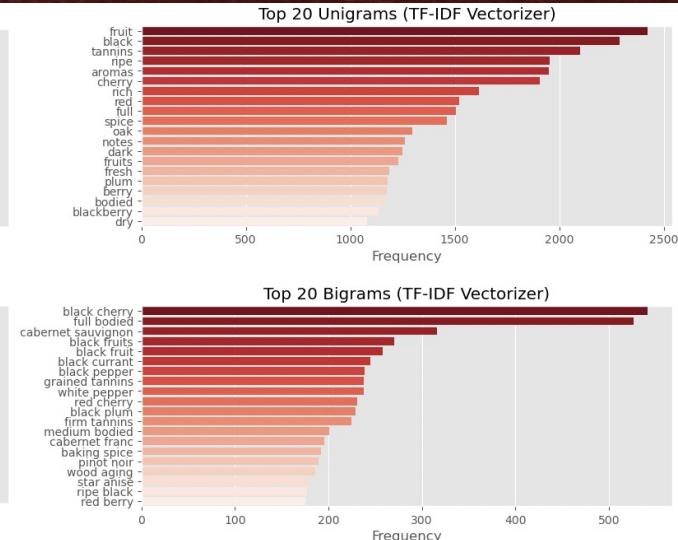
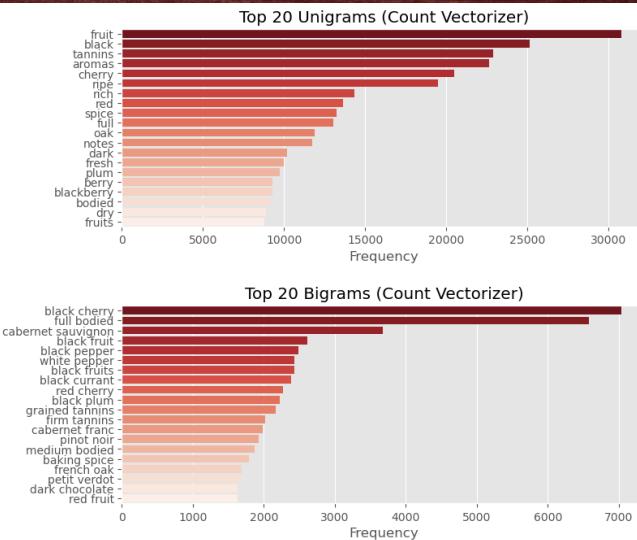
Exploratory Data Analysis

- Moderately positive correlation of 0.384 between wine prices and ratings
- Consumers perceive higher-priced wines as being of higher quality, leading to higher ratings
- Wine prices influence consumer perception and enjoyment of wine



Exploratory Data Analysis

Similar results in both vectorizers; many wine attributes may be observed



Amazon Web Scraping



Roll over Image to zoom in

cloudy bay Sauvignon Blanc 750ML-2021

Cloudy Bay

4.5★ 188 ratings

Price: \$52.00

\$69.33 / litre



Secure
transaction



Returns Policy



Amazon
Delivered

In Stock

Ships from and sold by AmazonFresh.

About this item

- Sauvignon Blanc is the flagship varietal of New Zealand, and is typically a very crisp aromatic wine with notes of grapefruit, tropical fruit and cut grass. The much lauded white wines of New Zealand tend to be slightly sweeter than those from Bordeaux and the Loire Valley, and are hugely influenced by the growing conditions of the country
- The grape originates from Bordeaux, where it is also known as Pouilly Fumé, Sancerre, and Graves, and is one of the world's most widely planted grapes
- Marlborough is the most revered wine producing region in the small country, accounting for 62% of all wine produced in New Zealand
- The New Zealand Sauvignon Blanc scene started popping in 1970s, with the high quality Sauvignon Blancs coming out of Marlborough illustrating the great potential for grape growing in New Zealand
- Sauvignon Blanc tends to be fermented in stainless steel vats, which results in a clean, crisp flavor, and will not show any of that butteriness that is often prevalent in well oaked Chardonnay for instance.; When it comes to pairing, think white meat, shellfish, and try to accentuate the herbaceous qualities of the wine with ingredients like parsley and cilantro

Show Less

Name of Wine

Price

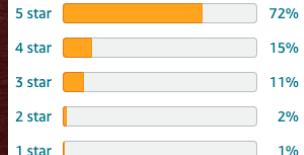
Description

Rating

Customer reviews

4.5 out of 5

188 global ratings



Types of Recommendations

- Popularity-based (Non-personalised)
- Collaborative Filtering using similarity models
- Content-based recommendations

Recommendations - Popularity-based

- Wines are recommended based on their overall popularity among all users
- Recommendations are the same for all users and do not take individual preferences into account
- E.g. Top 10 highest-rated wines or most purchased wines
- Easy to implement and understand, but limited in personalisation

Recommendations - Collaborative Filtering

- Based on the idea that users who have similar preferences in the past will have similar preferences in the future
- 2 types
 - User-based: Recommends wines liked by users who have similar tastes to the target user
 - Item-based: Recommends wines that are similar to those the target user has liked in the past
- Both use similarity metrics (e.g. Cosine Similarity, Pearson correlation)

Recommendations - Collaborative Filtering Only

- New user has no historical rating data
- Hence all recommendations have the same estimated rating

Number of wines with desired attribute(s): 17

Top 10 Recommendations for new user with Desired Attributes:

wine_id	estimated_rating	title	points
0 87272	91.70127	Rosenhof 2015 Orion Eiswein Grüner Veltliner (...)	95
1 115007	91.70127	Dr. Loosen 2009 Ürziger Würzgarten Auslese Rie...	94
2 122068	91.70127	Tapteil Vineyard 2011 Argus Bone Dry Riesling ...	92
3 185700	91.70127	Emile Beyer 2016 Tradition Gewurztraminer (Als...	92
4 6595	91.70127	Fiddlehead 2014 Gooseberry Sauvignon Blanc (San...	92
5 95923	91.70127	Qupé 2015 Sawyer-Lindquist Vineyard Marsanne (...)	91
6 192142	91.70127	Dr. Heidemanns-Bergweiler 2016 Bernkasteler al...	91
7 150610	91.70127	Lincourt 2014 Courtney's Chardonnay (Sta. Rita...)	91
8 164023	91.70127	Weingut Binz 2015 Nackenheimer Engelsberg Beer...	91
9 206665	91.70127	Gérard Bertrand 2019 Côte des Roses Rosé (Lang...	91

Recommendations – Content-based

- Use the attributes of wines to find similar items
- E.g. User likes a particular wine with specific characteristics (e.g., fruity, dry, full-bodied), system will recommend other wines with similar attributes
- This approach considers the user's tastes and preferences, offering a more personalised experience

Recommendations – Hybrid

Combining **both** approaches:

- Calculate cosine-similarity between wine's tokens & user attributes (content-based)
- Predicted ratings for each wine (collaborative filtering)

wine_id	combined_rating	title	points	variety
0	8775	92.849384 Lincourt 2014 Courtney's Chardonnay (Sta. Rita...)	91	Chardonnay
2	95923	92.588226 Qupé 2015 Sawyer-Lindquist Vineyard Marsanne (...)	91	Marsanne
4	115007	92.565362 Dr. Loosen 2009 Ürziger Würzgarten Auslese Rie...	94	Riesling
5	164023	92.402026 Weingut Binz 2015 Nackenheimer Engelsberg Beer...	91	Riesling
6	185700	92.385756 Emile Beyer 2016 Tradition Gewürztraminer (Als...	92	Gewürztraminer
7	6595	92.374515 Fiddlehead 2014 Goosebury Sauvignon Blanc (San...	92	Sauvignon Blanc
8	34864	92.354258 Beresan 2007 Sémillon (Walla Walla Valley (WA))	90	Sémillon
9	206665	92.273454 Gérard Bertrand 2019 Côte des Roses Rosé (Lang...	91	Rosé

Modeling - Surprise Library

Surprise Library

- Used for building and evaluating collaborative filtering-based recommendation systems
- Algorithms used:
 - Normal Predictor
 - Baseline Predictor
 - k-NN Basic
 - k-NN with Means
 - k-NN with Z-score
 - k-NN Baseline
 - SVD
 - Non-negative Matrix Factorisation (NMF)
 - Co-Clustering

Modeling – Metrics used



RMSE (Root Mean Square Error)

Difference between predicted ratings and actual ratings



Precision@k

Proportion of relevant recommendations among the top-k recommendations



Recall@k

Proportion of relevant recommendations that are included in the top-k recommendations

Modeling – Results

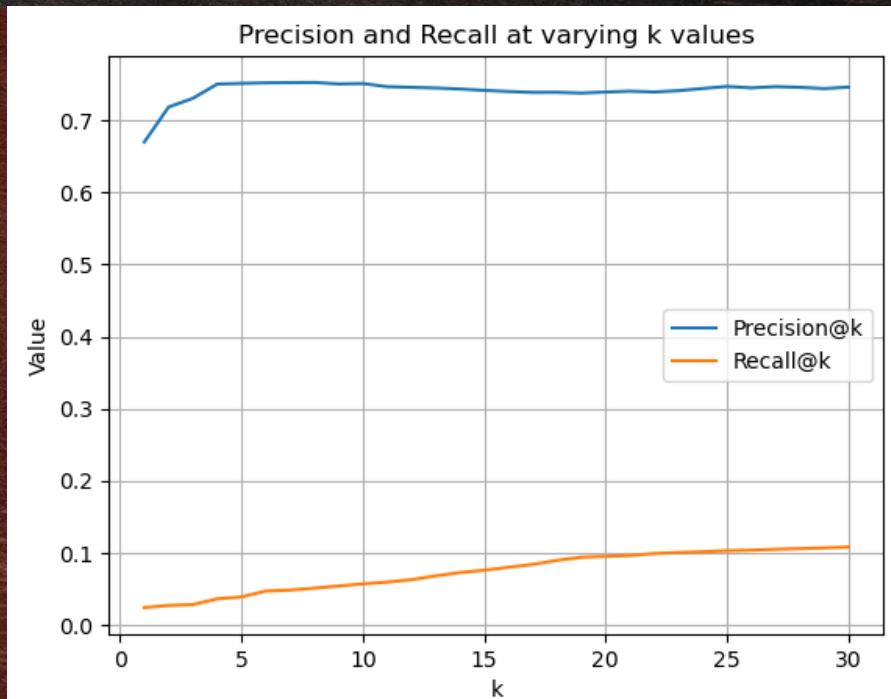
Algorithm	RMSE	Precision@10 (%)	Recall@10 (%)
Normal Predictor	2.176	57.0	16.3
Baseline Predictor	1.614	72.4	9.6
k-NN Basic	1.632	59.9	19.8
k-NN with Means	1.632	57.5	19.2
k-NN with Z-Score	1.632	58.9	19.9
k-NN Baseline	1.614	72.1	9.2
SVD	1.615	73.8	6.6
SVD (tuned)	1.615	72.7	8.8
NMF	1.632	59.7	19.1
Co-Clustering	1.632	57.3	18.5

Modeling -varying k-values

As k increases:

- Precision@10 hovers around 0.7
- Recall@10 increases marginally to 0.1

We will not vary the value of k for our model as no significant improvement in performance is observed.



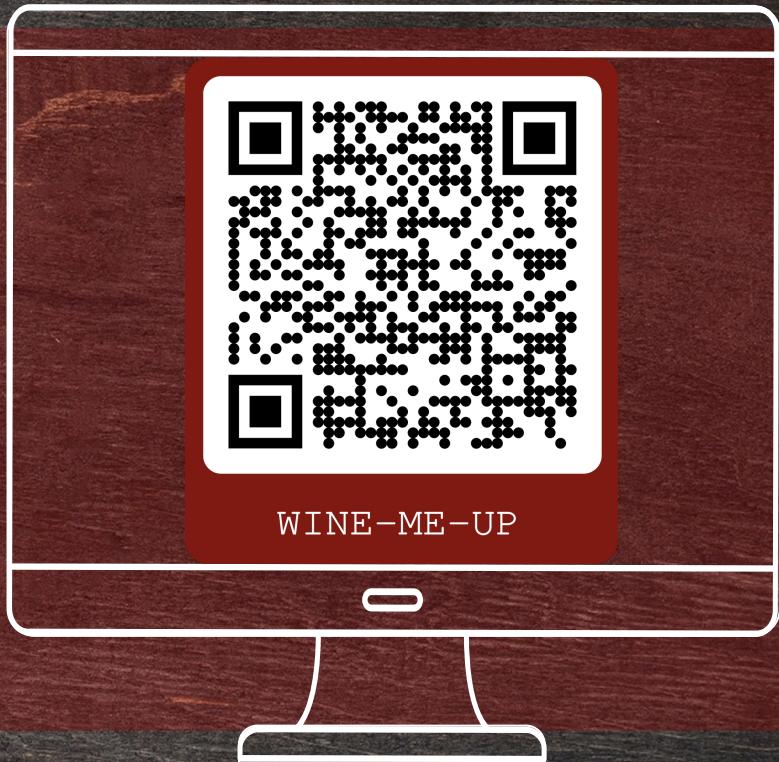
Conclusion

- SVD is the best choice for this project
- Combining both content-based and collaborative filtering approaches results in a more effective hybrid recommendation system

Recommendations

- Continuously updating dataset with new user ratings and wine information to ensure system remains relevant and accurate
- Incorporating user feedback to refine and improve the recommendations over time
- incorporating the Implicit library
- Implement advanced NLP techniques such as attention mechanisms and transformers using PyTorch

Streamlit Demo



<https://wine-me-up.streamlit.app>

Thank you!

CREDITS: This presentation template was created by **Slidesgo**,
including icons by **Flaticon** and infographics & images by **Freepik**