INFORMATION TECHNOLOGY

Recommendation System with Wines

Prepared by Northwestern University MLDS
On Behalf of Breakthru Beverage Group

Meet Northwestern MLDS Team

Northwestern | McCORMICK SCHOOL OF ENGINEERING

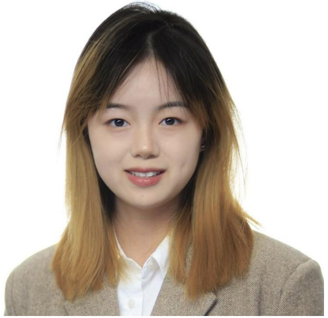
Master of Science in Machine Learning and Data Science

Jackie Kim

Stella Wang

Cathy Wang

Grace Xie



Yumin Zhang

Agenda

01 - Introduction

Objectives & Business Opportunities

02 - Data Exploration

Data Findings & Visualization

03 - Recommender Systems

Content-Based, User-Based, Item-Based, Matrix Factorization, Hybrid

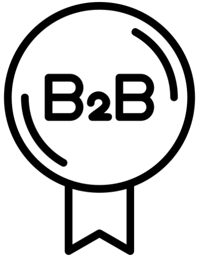
04 - Examples & Considerations

Outputs from Models & Considerations

INFORMATION TECHNOLOGY

Introduction

Developing Wine Recommender System (in USA)



Majority of clients are from actual businesses

19 active Sales Divisions in the USA

On-Premise vs Off-Premise

Business Opportunities

Increased Sales

Optimize Inventory Management

Personalized Client Experience

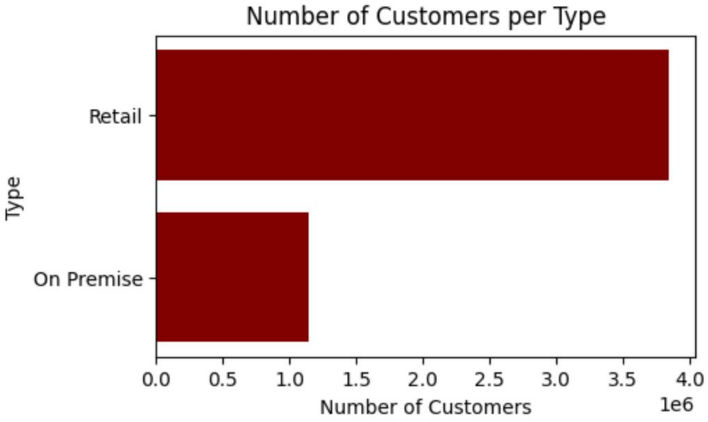
Q: Can an algorithm understand your taste of wine?

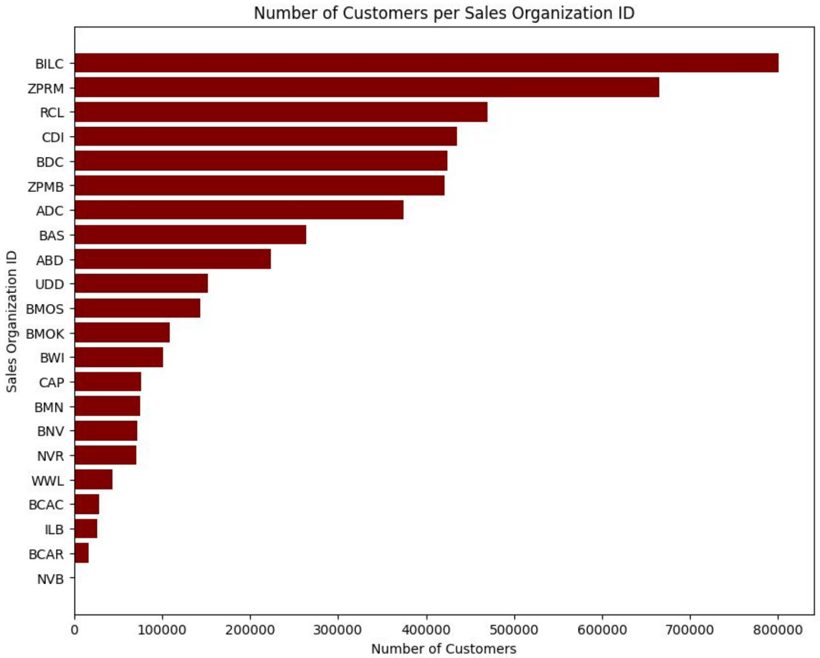
INFORMATION TECHNOLOGY

Data Exploration

Scope of the Data

- **Year:** 2019 ~ 2024 (5 years back from today)
- **# Unique Clients**: 128K
- **# Unique Products**: 67K
- Divide Data into Premise Type and Sales Organizati

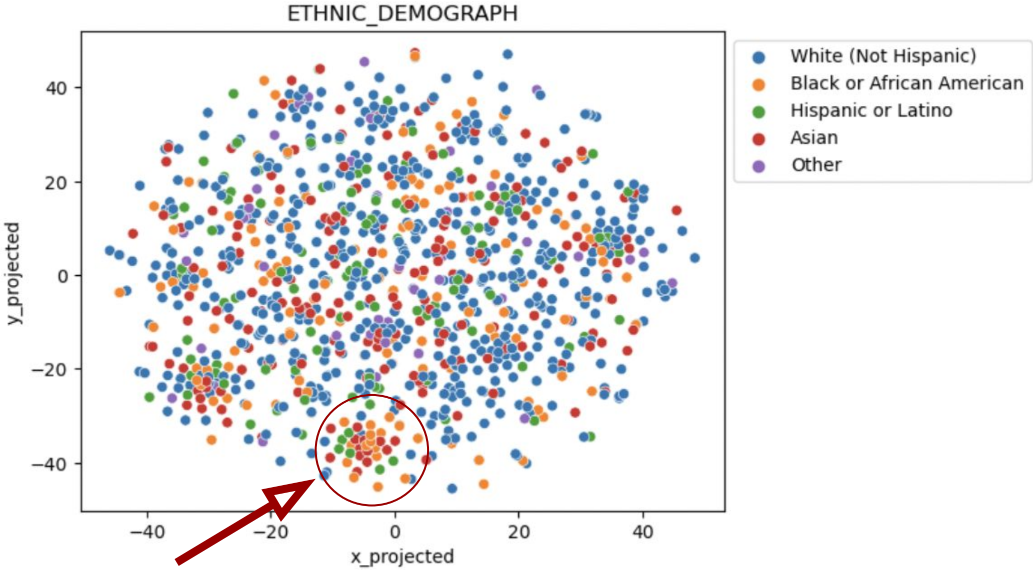


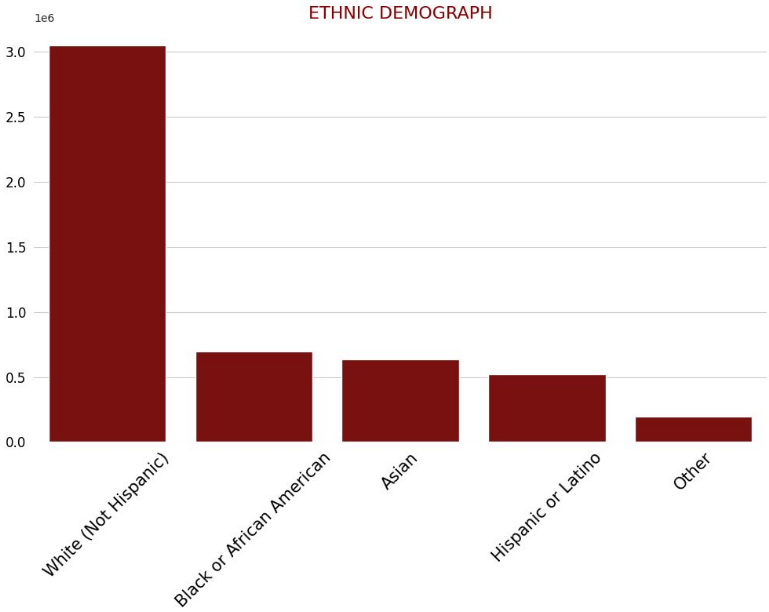


INFORMATION TECHNOLOGY

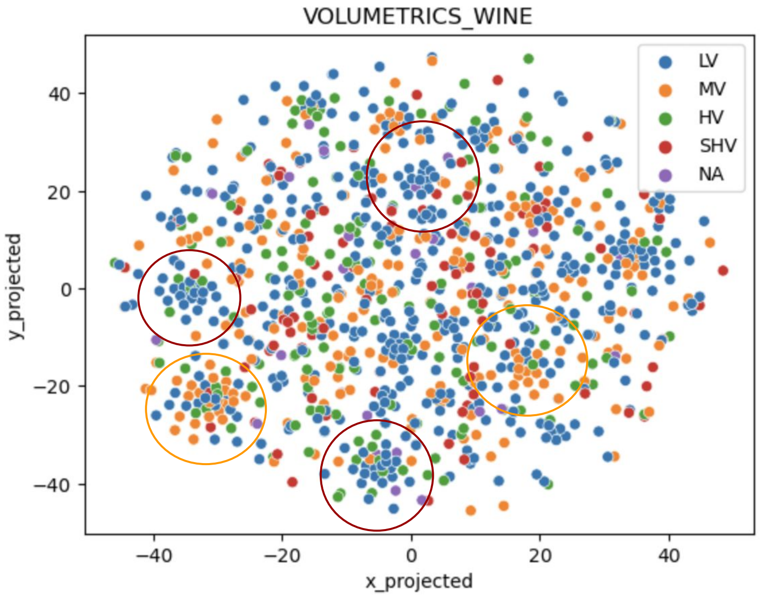
Similarity in Wine Taste Among Ethnic Groups by Clients

T-SNE Clustering

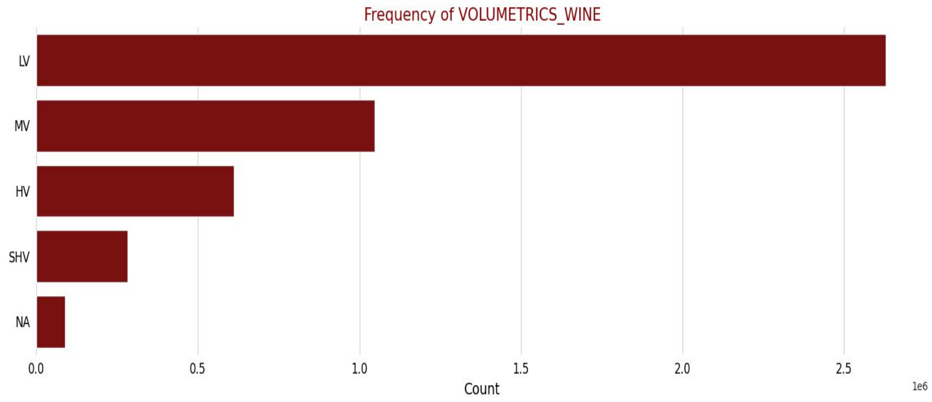




Dominant Quantity and Revenue Proportion from LV Category

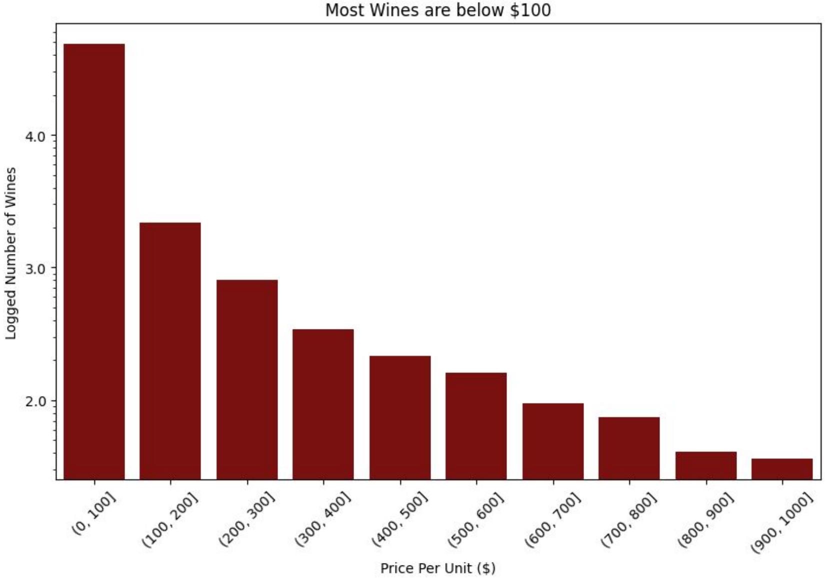


The "LV" category has the highest frequency among wine volume metrics

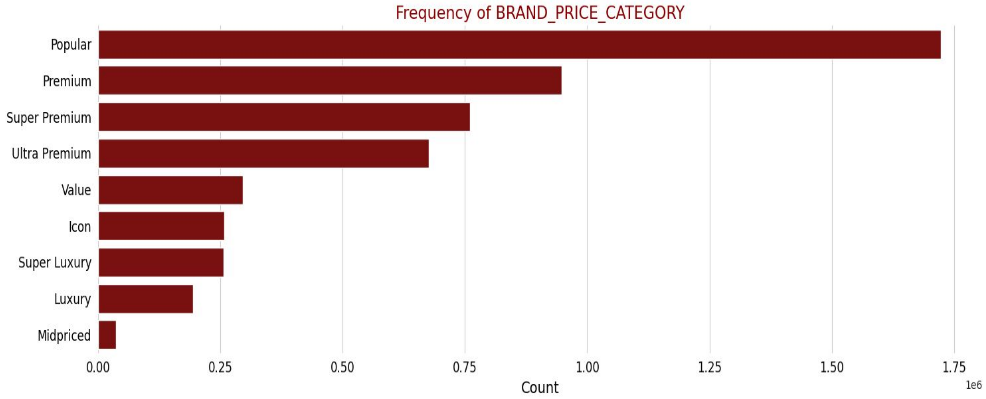


Dominant Trends: Affordable Wines and Popular Categories

Overall, the largest proportion of Wines' "Price Per Unit" demand is at $100 or less.

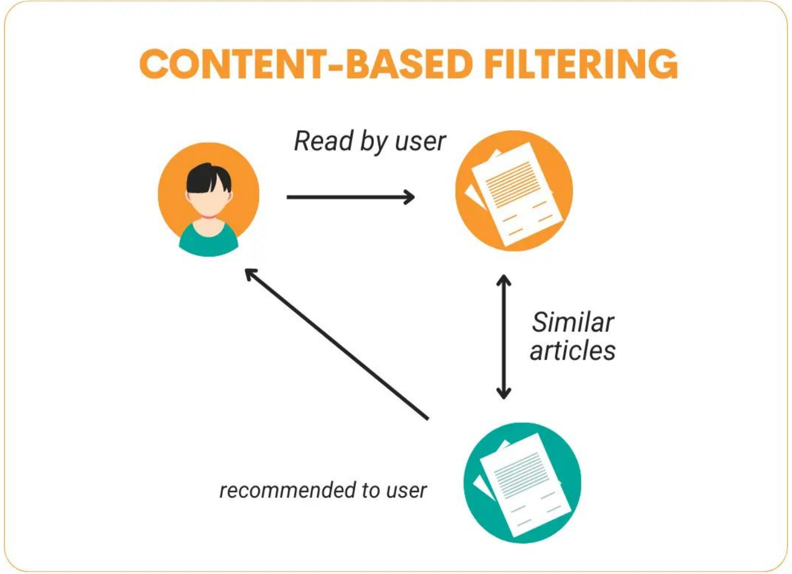


The "Popular" category has the highest frequency among wine brand price categories



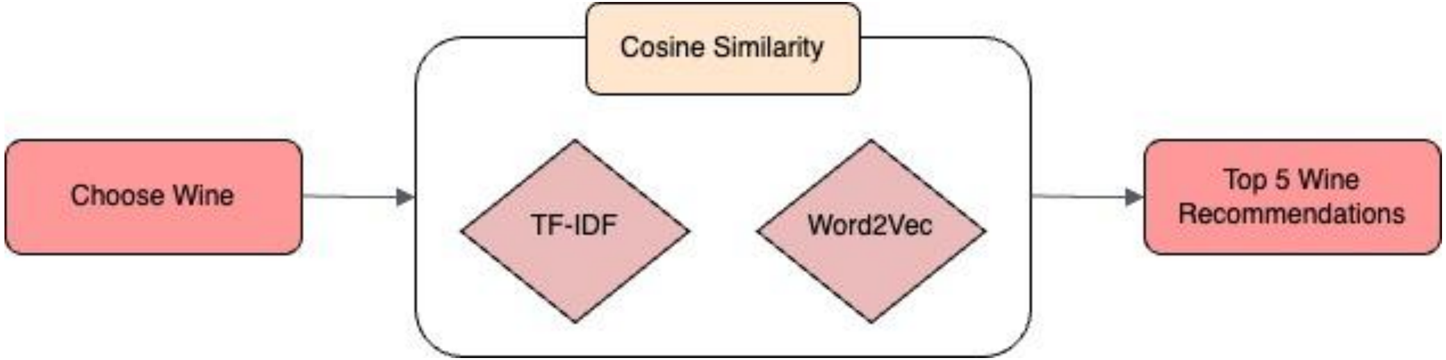
Recommender Systems

Content-Based Filtering (CBF)



- Suggests items to users based on the features and attributes of items they have previously liked or interacted with.
- Matches based on similarity of items (such as item characteristics or descriptions).

CBF Flowchart



- **TF-IDF**: Relevancy/importance of each word to a document in a collection. Unique words have higher score.
- **Word2Vec**: Captures the semantic meaning of each word based on contexts.
  - Ex. Different Appellations will be grouped as similar categories.
- **Cosine Similarity**: Scoring mechanism to measure Top 5 Wine Recommendations

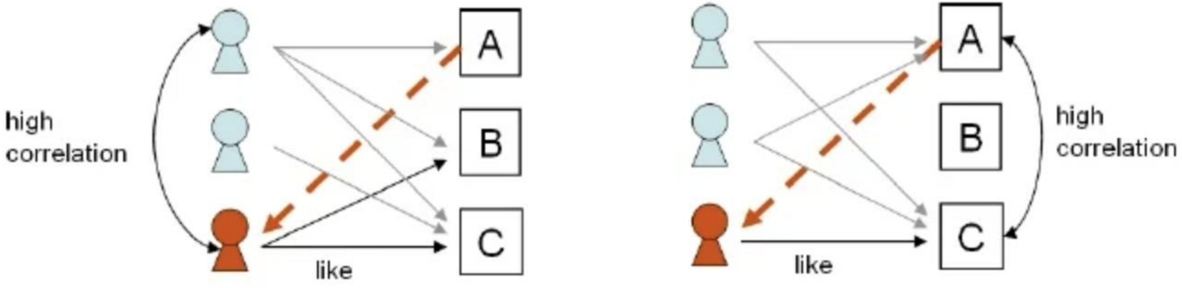
fruity with hints of oak and vanilla

Collaborative Filtering (CF)

The memory-based approach uses user rating data to compute the similarity between users or items.

Typical examples of this approach are item-based and user-based top-N recommendations.

$$r_{u,i} = \operatorname{aggr}_{u' \in U} r_{u',i}$$



User-based filtering (Grouplens, 1994)

Item-based filtering (Amazon, 2001)

Tackling Rating Score

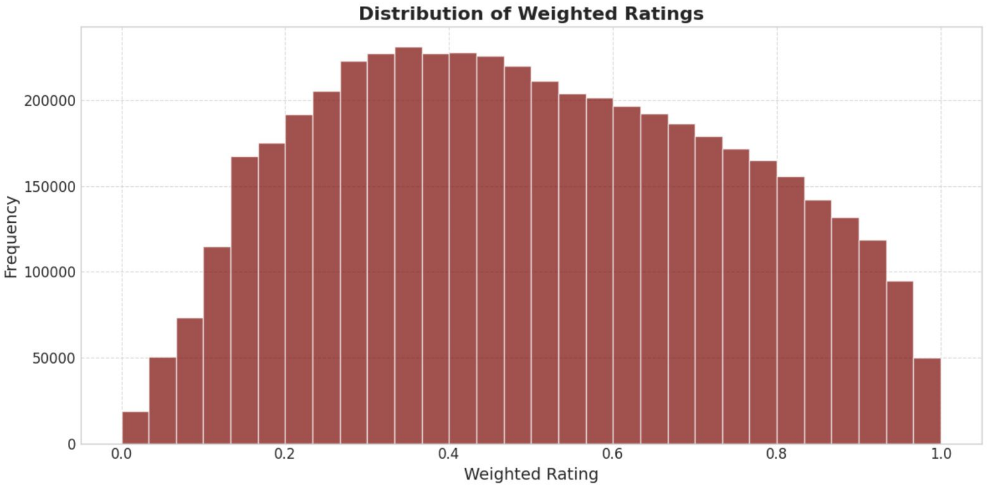
Implicit Rating Score - Quantile Transformer:

rating = mean(sum([Quantity Purchased, Net Revenue, Last Purchased Date]))

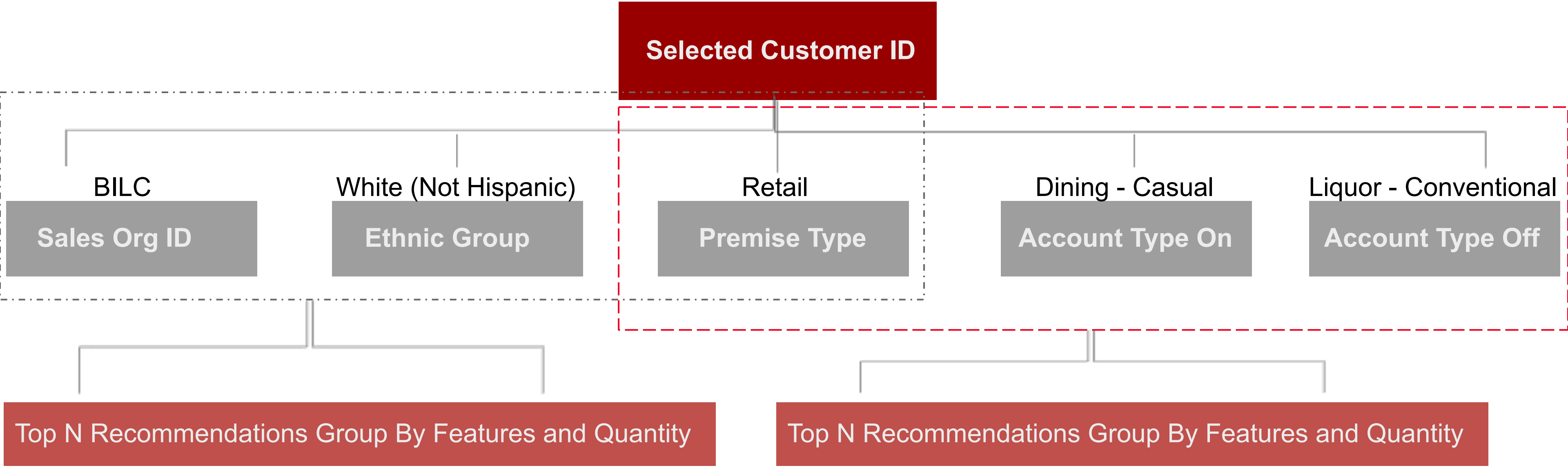
$$\text{Lower Quartile (Q1)} = (N+1) \times \frac{1}{4}$$

$$\text{Middle Quartile (Q2)} = (N+1) \times \frac{2}{4}$$

$$\text{Upper Quartile (Q3)} = (N+1) \times \frac{3}{4}$$

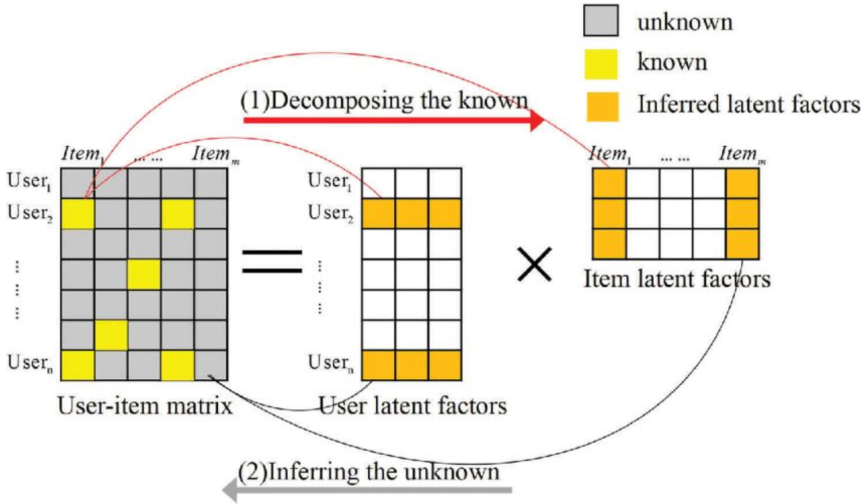


Tackling Cold Start - Popularity Based Model



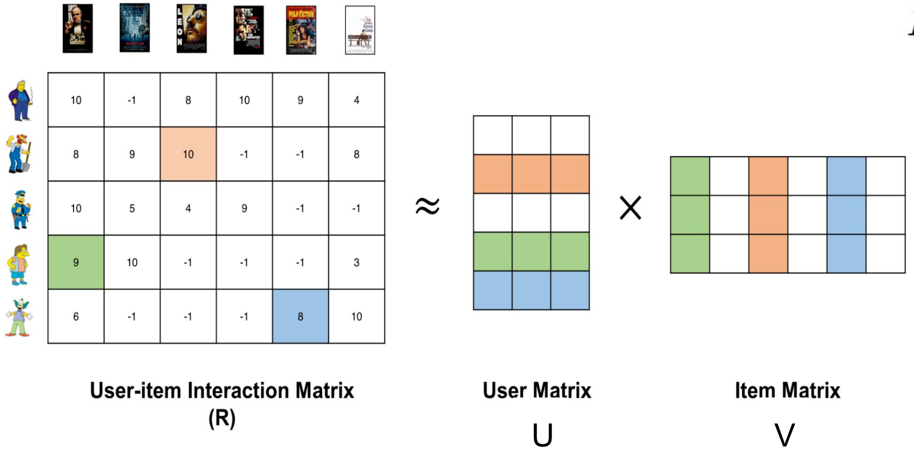
Matrix Factorization

- What is Matrix Factorization?
  - ➢ Linear algebraic model to predict user preferences
  - ➢ Aim to fill in the missing entries in a user-item matrix
- What techniques?
  - ➢ Non-negative Matrix Factorization (NMF)
  - ➢ Combine NMF with Singular Value Decomposition (SVD)
- Why Matrix Factorization?
  - ➢ Interpretability
  - ➢ Sparsity handling
- Challenges
  - ➢ Cold start problem
  - ➢ Computation extensive



Collaborative Filtering - User Based

Singular Value Decomposition



10	-1	8	10	9	4
8	9	10	-1	-1	8
10	5	4	9	-1	-1
9	10	-1	-1	-1	3
6	-1	-1	-1	8	10

SVD factorizes the user-item interaction matrix R into three matrices:

$$R \approx U \Sigma V^T$$

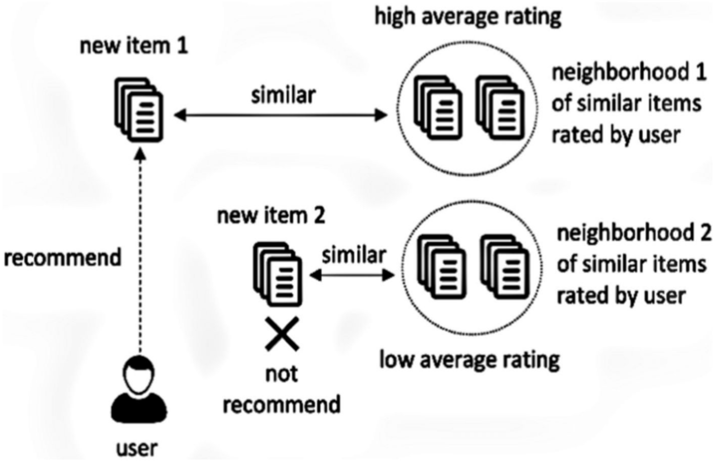
- U is the user-feature matrix.
- Σ is a diagonal matrix of singular values.
- V^T is the transpose of the item-feature matrix.

Collaborative Filtering - Item Based

KNNwithZscore (K-Nearest Neighbors)

- K - maximum number of neighbors to take
- Sim options: Similarity measure configuration

$$\hat{r}_{ui} = \frac{\sum\limits_{j \in N_u^k(i)} \text{sim}(i,j) \cdot r_{uj}}{\sum\limits_{j \in N_u^k(i)} \text{sim}(i,j)}$$

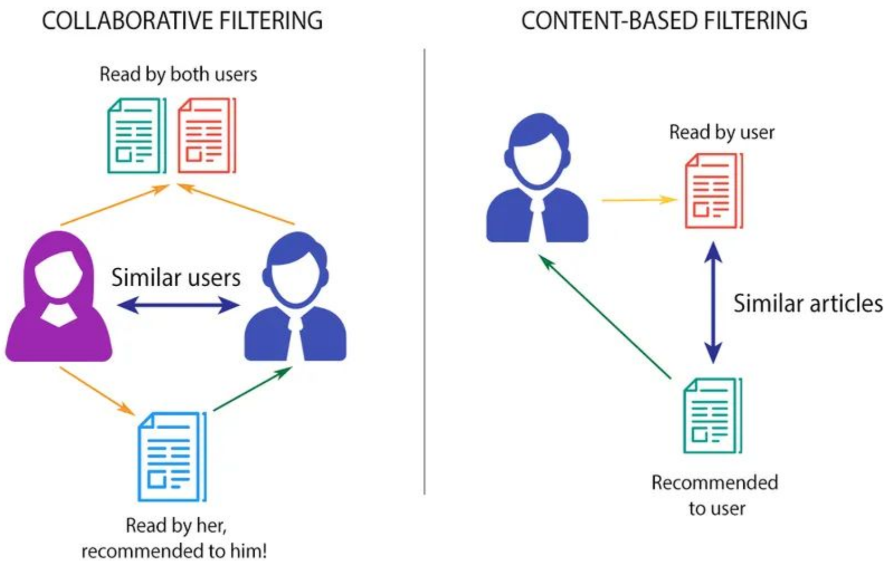


Disadvantages:
- ➔ Scalability
- ➔ Sparsity
- ➔ Computation Complexity

Advantages:
- ➔ Simplicity
- ➔ Interpretability
- ➔ Flexibility

Hybrid Model (LightFM)

Why Hybrid?

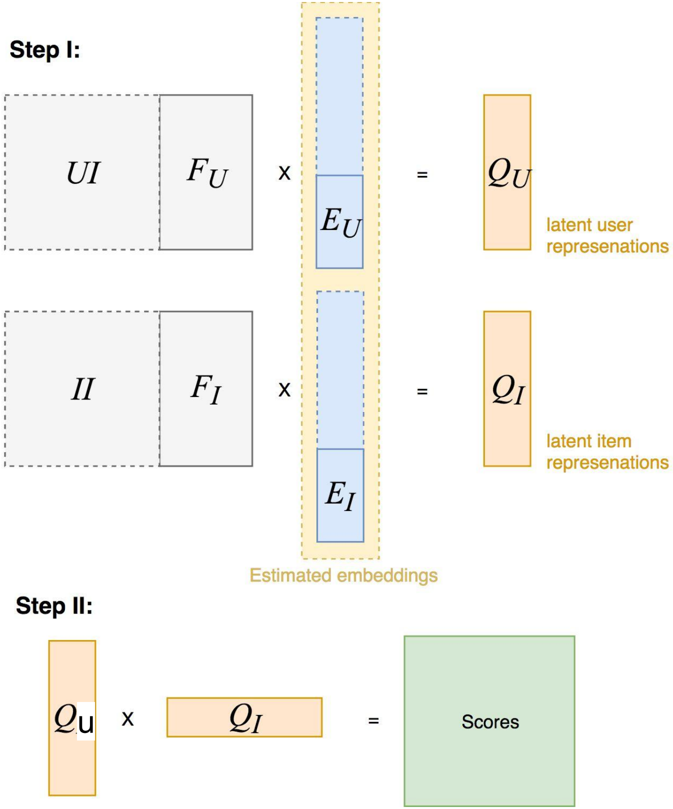


Why LightFM?

- The cold start problem
- WARP (Weighted Approximate-Rank Pairwise) loss

Hybrid Model (LightFM)

UI: user identifiers
II: item identifiers
FU: user features
FI: item features



Enhancing Feature Weight in LightFM for Better Model Performance

Specific Case: Prioritizing Wine Availability (Implemented)

- Goal: Ensure recommendations consider wine availability within the same sales organization and premise type.
- Approach: Repeat the 'availability' feature multiple times in the feature string.

Benefits

- Improved Recommendations: By highlighting key features like availability, the model can better prioritize relevant wines.

Model Outputs & Considerations

Example for Off-Premise Florida Wine Category (CBF)

Chateau La Garde Pessac Leognan 2014 750ml	Mid-End ($37)
Stags Leap Cabernet Sauvignon 2017 750ml	**Mid-End ($50)**
Ken Wright Cellars Canary Hill Pinot Noir 2020 12B 375ml	Mid-End ($24)
Tablas Creek Vineyard Patelin de Tablas Rosé 2022 750ml	Mid-End ($26)
Ken Wright Cellars Freedom Hill Pinot Noir 2020 6B 750ml	Mid-End ($48)
Stags Leap Merlot Napa Valley 2016 750ml	**Mid-End ($27)**

Example from PALACE ITALIAN RESTAURANT

Rank	Item-based CF	Customer-based CF	Hybrid (LightFM)	Matrix Factorization
1	CARPINETO DOGAJOLO RED 750ML	LINDEMANS CAB SAUV BIN 45 750ML	**BAREFOOT ROSE 1.5L**	RWC GIORDANOS P GRIGIO 750ML
2	EDNA VALLEY CHARD PARAGON 750ML	LIBERTY CREEK CHARD TETRA PK 12B 500ML	YELLOW TAIL SANGRIA RED 750ML	TURNING LEAF MERLOT 750ML
3	CARLO ROSSI CHARD 4L	**BAREFOOT CAB SAUV 1.5L**	CYT CASILLERO DIABLO CARMENERE 750ML	RWC GIORDANOS CAP SAUV 750ML
4	COASTAL EST P NOIR 750ML	APOTHIC CA ROSE 750ML	COPPOLA DIAMOND CAB SAUV 20 750ML	ARMAND DE BRIGNAC BRUT ROSE GB 6B 750ML
5	FLEUR DE MER ROSE ST TROPEZ 16 750ML	BREWER CLIFTON CHARD STA RITA 6B 750ML	**BAREFOOT FRUITSCATO MANGO 1.5L**	ARMAND DE BRIGNAC ROSE GFT BAG 6B 750ML

Example from WOLFES WINE SHOPPE LLC

Rank	Item-based CF	Customer-based CF	Hybrid (LightFM)	Matrix Factorization
1	LIBERTY CREEK CHARD 1.5L	SUTTER HOME P GRIGIO 24B 187ML	APOTHIC CA WT 750ML	QUARTER CUT CAB SAUV BRBN BBL 750ML
2	B SIDE CAB SAUV 750ML	**BELLETTI PROSECCO 750ML**	19 CRIMES SAUV BLOCK 750ML	**BELLETTI PROSECCO 750ML**
3	YELLOW TAIL P NOIR 1.5L	**PETIT COCO 750ML**	BURLWOOD PACIFIC FRUIT SWT MANGO 750ML	WINKING OWL P GRIGIO BIB 6B 3L
4	BERINGER MAIN & VINE WT ZIN 6/4PK 187ML	QUARTER CUT CAB SAUV BRBN BBL 750ML	BARTENURA MOSCATO ASTI CAN 12/4PK 250ML	**PETIT COCO 750ML**
5	MERIDIAN CHARD S BARBARA 750ML	SUTTER HOME WT ZIN 24B 187ML	KORBEL CHAMP BRUT 750ML	BURLWOOD PACIFIC FRUIT SWT STRAWB 750ML

Summary Table of Each Model

	Content-Based Filtering	Item-based CF	Customer-based CF	Hybrid (LightFM)	Matrix Factorization
Pros	• Minimal Cold-Start Issue by limiting to Products. • Rating Score not required	• Straightforward integration process • Direct performance insights • Adaptable to changes	• Reduce dimensionality • Handling missing data • Find latent representations of users and items	• Combines strengths of multiple models • Handles sparse data and cold start	• Ensures interpretability (values are all non-negative) • Suitable for handling sparse data
Cons	• Limited wine variety • Potential user preference mismatch • Non-clear evaluation metrics	• Cold start problem • Scalability • Sparsity of memory based	• Cold start problem • Interpretability	• Complexity in implementation • Requires tuning	• Cold start problem • Computational extensive for large datasets

Important Considerations

Opportunity Cost in Computation & Time

- **Hybrid Model** - the best performance & prediction but will take longest to run.
- **CBF Model** - Fastest but wine variety will not be as diverse.
- **User CF, Item CF, MF CF Model** - In-between above two models.

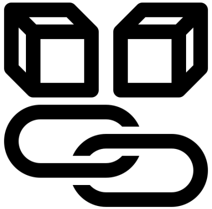
Cautions

- **Data/Model Drift** - Consistent training, once per day or week, before predicting is imperative.
- **No Explicit Score** - Generally makes model performs better. Minimized the impact using Implicit Synthetic Score.

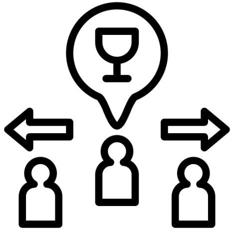
Next Step

- **A/B/n Test** - Test multiple models to provide recommended wines and collect ratings and feedback.
- **Validation** - Requires corroboration from subject matter experts to confirm whether clients converted with the recommendations.

Next Practicum Project Considerations



Price Elasticity in Supply Chain



Demand Forecast for Products

Beer Recommender System

Thank you!

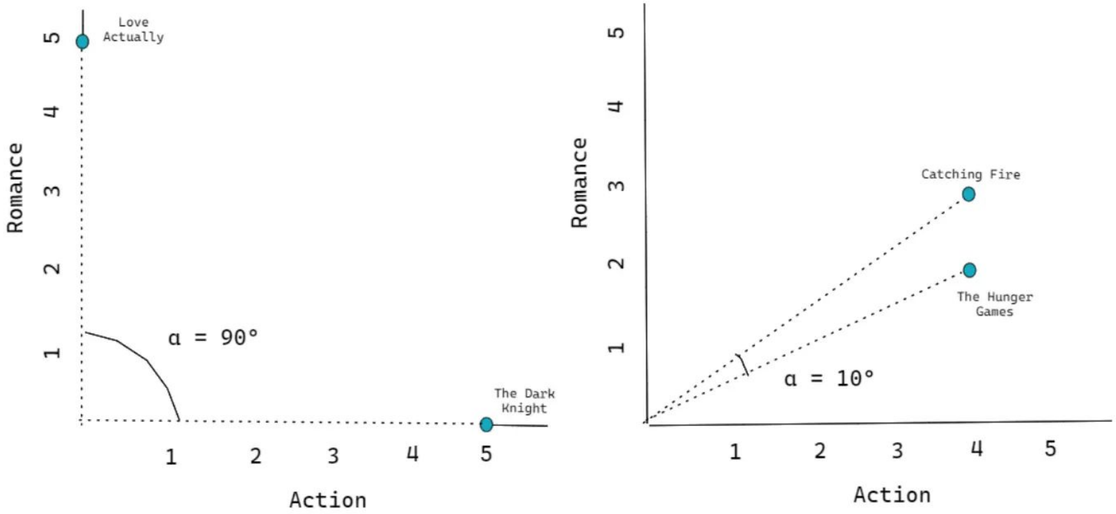
Appendix

Details in Formula

Cosine Similarity

$$cos(x, y) = \frac{x.y}{||x|| * ||y||}$$

- closeness between two group of words by the angle.



TF-IDF (Term Frequency-Inverse Document Frequency)

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

- ***tf_x,y*- -** the frequency of term **x** within document ***y***
- ***N* -** total number of documents
- ***df_x* -** total number of documents which contain term ***x***

Benefit: Measures the relevancy/importance of each word to a document in a collection or corpus.

Word2Vec

$$\max_\theta \sum_{(w,c)\in D} \log P(c|w;\theta)$$

- θ - parameters of the model, specifically the word embeddings.
- D - is the training corpus consisting of pairs (w,c), where w is the target word and c is the context word.
- $P(c|w;\theta)$ **-** probability of the context word c given the target word w, parameterized by θ

Benefit: Captures the semantic meaning of each word based on contexts.

Augmentation (For Content-Based Filtering)

$$[\text{Similarity}(d_1, d_2) = \alpha \cdot \frac{\text{TFIDF}(d_1) \cdot \text{TFIDF}(d_2)}{\|\text{TFIDF}(d_1)\| \|\text{TFIDF}(d_2)\|} + (1 - \alpha) \cdot \frac{\text{Word2Vec}(d_1) \cdot \text{Word2Vec}(d_2)}{\|\text{Word2Vec}(d_1)\| \|\text{Word2Vec}(d_2)\|}]$$

Factors taken to account for the data

- Allowed an option to filter by Sales Organization and Premise Type.
    - **Reason:** If duplicate wines are included for different sales org, it creates a noise in the recommender system.
    - Above filter significantly reduces the random noise.

High-End for Off-Premise Florida Wine Category (CBF)

Domain Ponsot Close de Roche Grande Cru 2020 750 ML	High-End ($750)
Groth Cabernet Sauvignon 2017 1.5L	High-End ($106.50)
Groth Cabernet Sauvignon 2016 1.5L	High-End ($106.50)
Quinta Noval Port Nacional 2021 6B 750 ML	High-End ($1089)
Dows Port Vintage 2016 6B 750 ML	High-End ($106.50)
Chateau Montelena Cabernet Sauvignon 2018 6B 750 ML	High-End ($150)

Matrix Factorization (NMF)

$$\min_{S_o, S_r \geq 0} \left\| S_o - WH \right\|$$

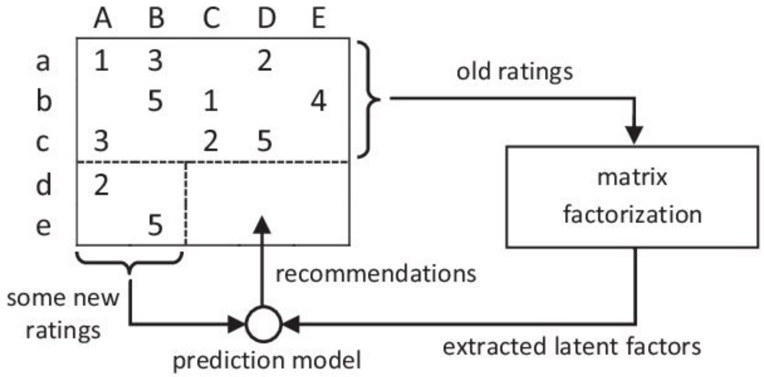
- where *So* is the original spectrum, and *W* and *H* are 2D non-negative matrixes.
- NMF the following iterative learning rules are used to find the linear decomposition:

$$H_{a\mu} \leftarrow H_{a\mu} \sum_i H_{a\mu} \frac{V_{i\mu}}{(WH)_{i\mu}}$$

$$W_{ia} \leftarrow W_{ia} \sum_\mu \frac{V_{i\mu}}{(WH)_{i\mu}} H_{a\mu} \qquad W_{ia} \leftarrow \frac{W_{ia}}{\sum_j W_{ja}}$$

Collaborative Filtering - User Based

SVD (Singular Value Decomposition)



SVD factorizes the user-item interaction matrix R into three matrices:

$R \approx U\Sigma V^T$

- U is the user-feature matrix.
- Σ is a diagonal matrix of singular values.
- V^T is the transpose of the item-feature matrix.

$$\hat{r}_{ui} = \mathbf{u}_u^T \mathbf{v}_i$$

$$\min_{U,V} \sum_{(u,i)\in\mathcal{K}} (r_{ui} - \hat{r}_{ui})^2 + \lambda(\|U\|^2 + \|V\|^2)$$

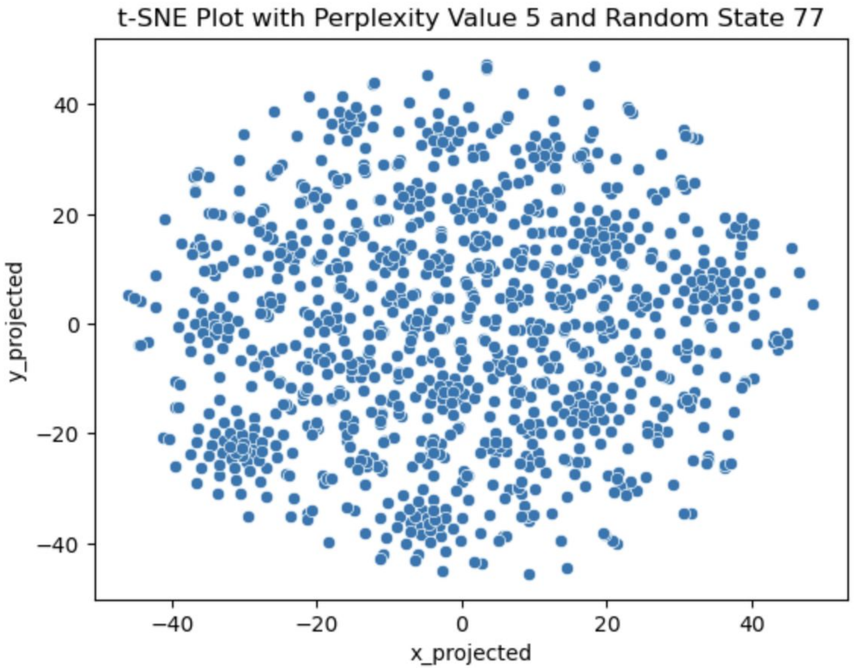
LightFM (hybrid) - WARP Loss Function

$$L = \sum_{u,i,j} \max(0, 1 - (s_i - s_j)) \cdot rank(i)$$

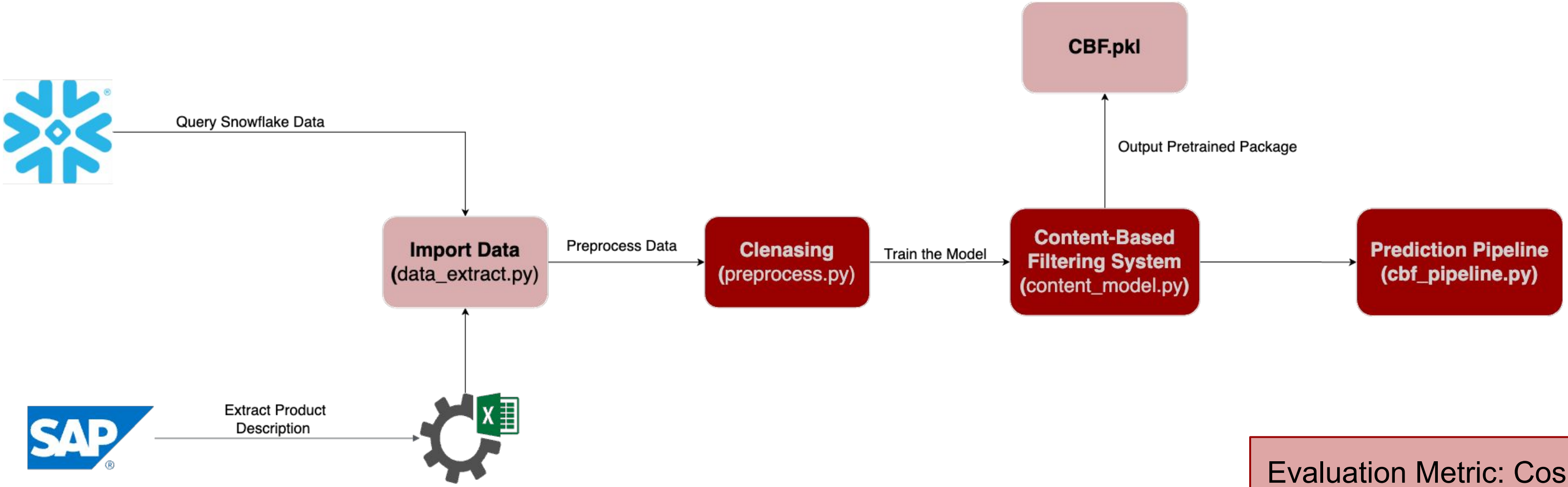
- u is a user.
- i is a positive item that the user has interacted with.
- j is a negative item that the user has not interacted with.
- s_i and s_j are the model's scores for items i and j, respectively.
- $rank(i)$ is a weighting function that increases the penalty based on the rank of the positive item i when compared to negative item j.

Summary Table for Performance Metrics

Metrics	Content-Based Filtering	Item-based CF	Customer-based CF	Hybrid (LightFM)	Matrix Factorization
RMSE (range: 0 to ∞; the lower the better)	N/A - Cosine Similarity (depends on wine)	On premise: 0.223 Off premise: 0.249	On premise: 0.196 Off premise: 0.184	N/A	On premise: 0.221 Off premise: 0.196
AUC (range: 0 to 1; the higher the better)	N/A - Cosine Similarity (depends on wine)	N/A	N/A	On premise: 0.903 Off premise: AUC: 0.938	N/A

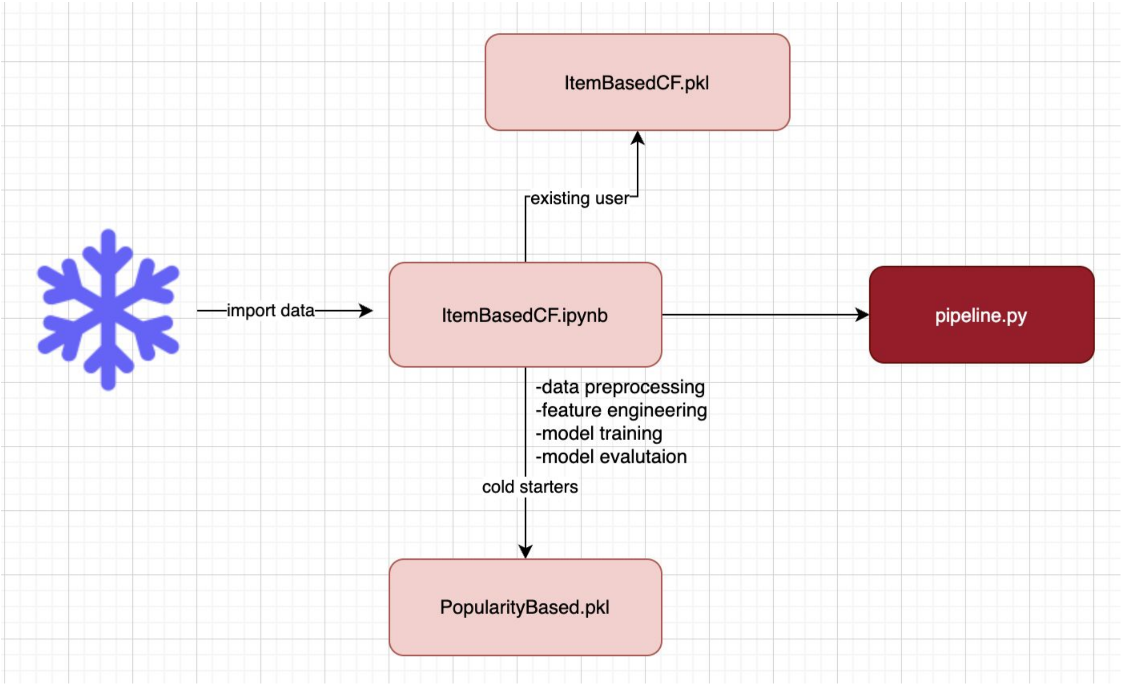

t-SNE Plot with Perplexity Value 5 and Random State 77
y_projected
x_projected
40
20
0
−20
−40
−40
−20
0
20
40

Implementation Pipeline: Content-Based Filtering



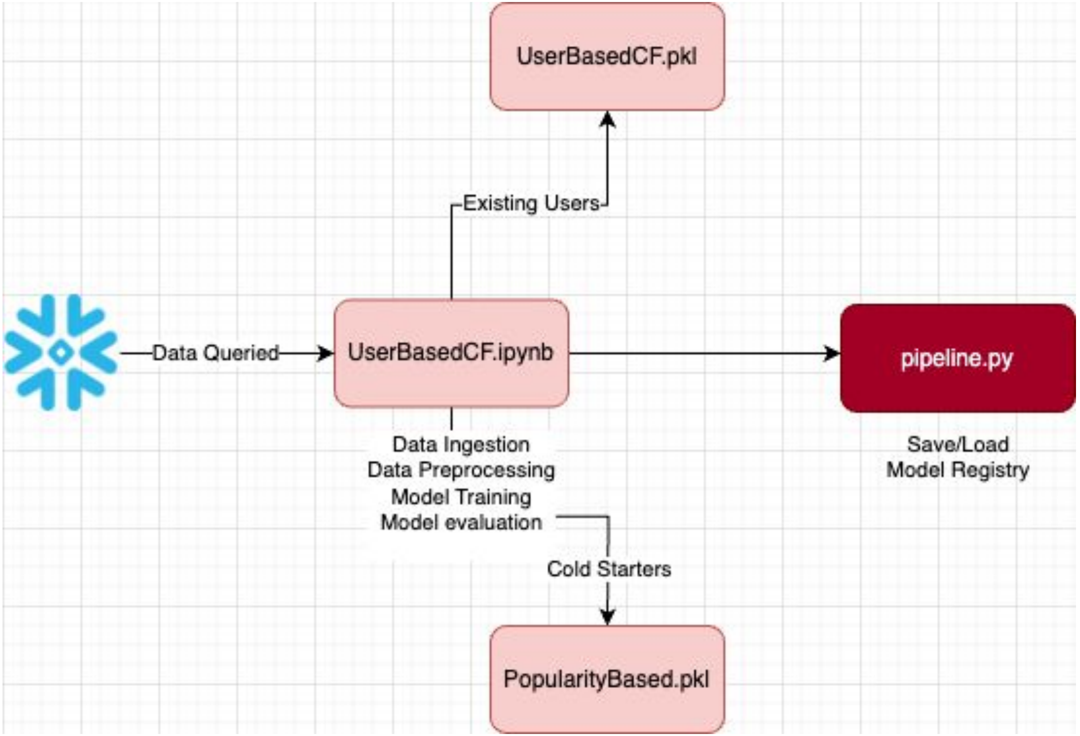
Evaluation Metric: Cosine Similarity. This depends on Chosen Wine

Implementation Pipeline: Item-based CF



Evaluation Metric (RMSE)
Train RMSE:0.1654
Test RMSE: 0.2443

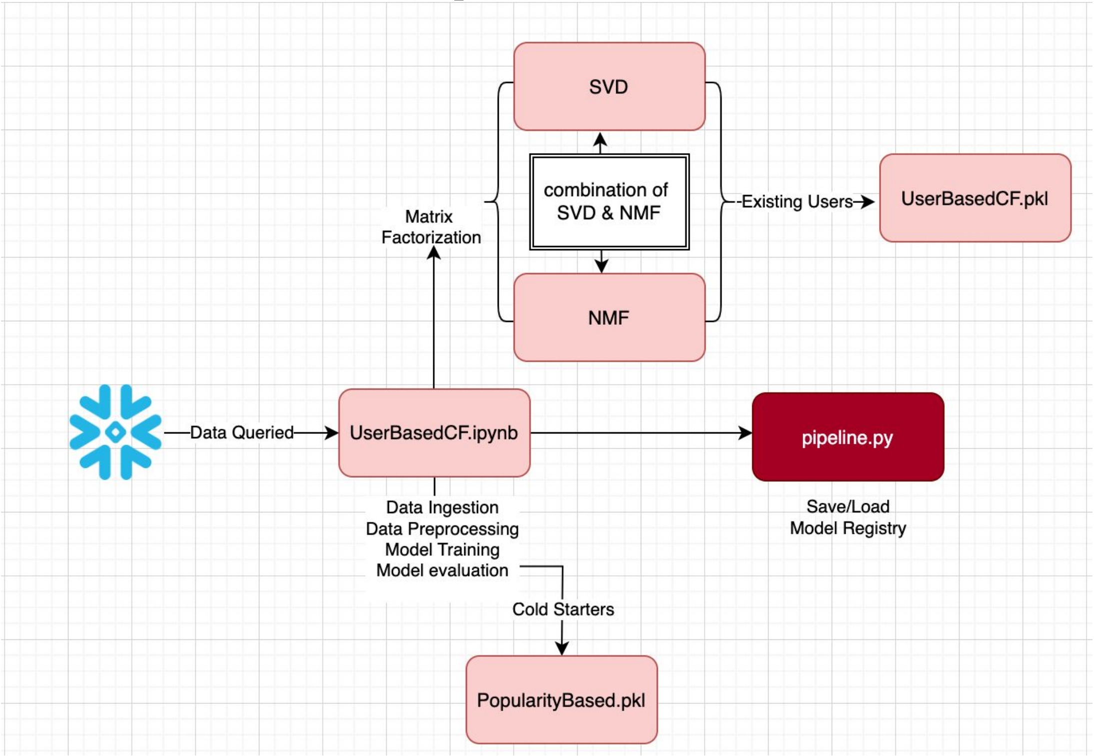
Implementation Pipeline: User-based/MF CF



Evaluation Metric (RMSE)
Train RMSE: 0.1921
Test RMSE: 0.1920

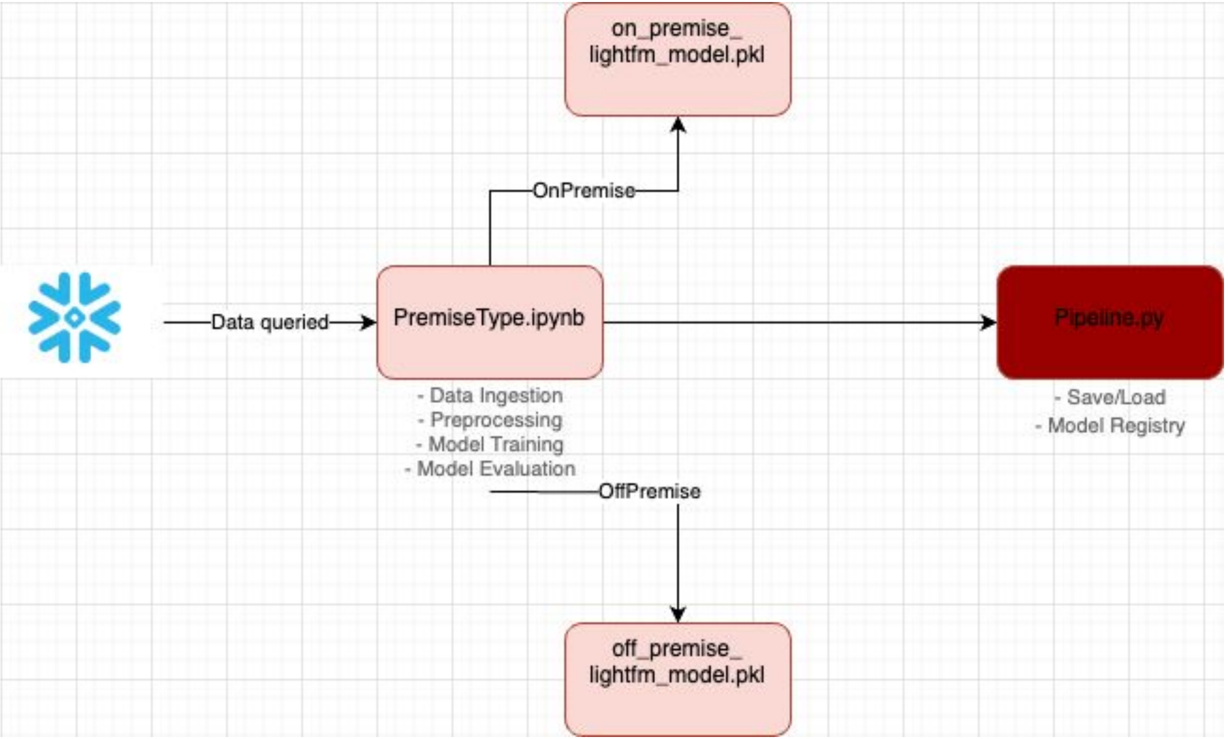
INFORMATION TECHNOLOGY

Implementation Pipeline: NMF+SVD CF



Evaluation Metric (RMSE)
NMF RMSE: 0.234
combined RMSE: 0.20

Implementation Pipeline: LightFM



Evaluation Metric (AUC)
On premise:
Train AUC: 0.942
Test AUC: 0.903
Off Premise:
Train AUC: 0.979
Test AUC:0.938