

WINE & SPIRIT RECOMMENDATIONS: A SHOPPING HISTORY ANALYSIS

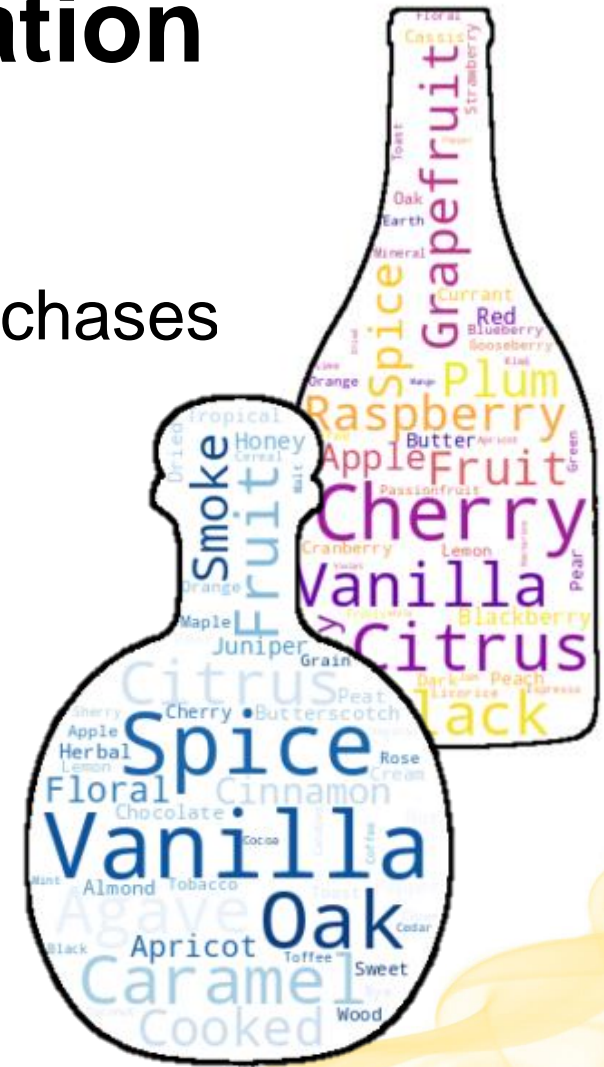
MSML 651: Big Data Analytics
Thomas Marmo

Evolving Trends in Wine & Spirits Retail

- Online alcohol retail boomed in 2020
 - Curbside pick-up
 - Delivery
- Last mile delivery popularized
 - Drizly
 - DoorDash
 - GrubHub
- The New Normal

Leveraging Data for Personalization

- Time to reassess how a customer shops
- Recommendation strategy only accounts for purchases
 - Product pages visited are left out of the equation
- No salespeople in a digital storefront
 - Personal recommendations
 - Product tastings
- Results in...
 - Repetitive purchasing behavior
 - High-ticket window shopping



Dataset

```
root
|-- CUSTOMER_ID: long
|-- DATE: string
|-- PRODUCT_ID: integer
|-- PURCHASE: integer
|-- PRODUCT_NAME: string
|-- DIRECT: integer
|-- DEPT_1: string
|-- DEPT_2: string
|-- DEPT_3: string
|-- CHAR_1: string
|-- CHAR_2: string
|-- UNIT_PRICE: double
|-- MARGIN_RATE: double
|-- POPULARITY: double
```

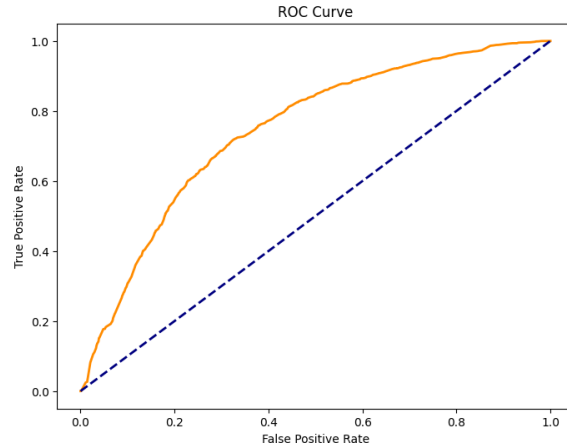
- Joined Three Tables
 - Customer browsing data
 - Customer purchase data
 - Product data
- Supplemental Calculations
 - Gross Margin Rate
 - Low-Level Dept. Popularity
- Limited Selection of Products and Markets
 - Less than 5000 SKUs
 - No scraping data

Methodology Using PySpark

- PySpark enabled an end-to-end ML pipeline
 - Categorical features were string indexed
 - Feature columns were identified & assembled into a vector
 - An ML model was defined
- Random Training & Testing Split (80/20)
- PySpark Evaluation Libraries

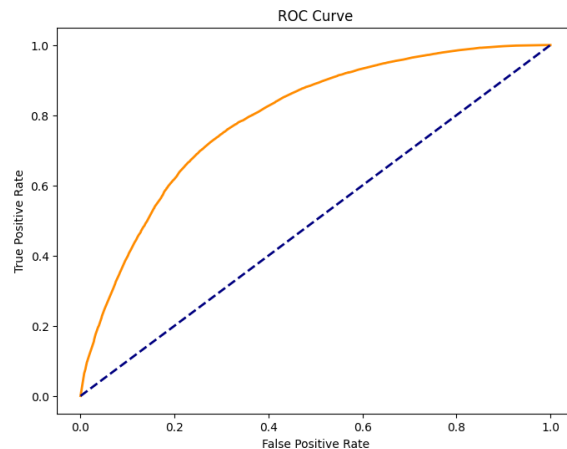
CUSTOMER_ID	DATE PURCHASE	PRODUCT	PRODUCT_NAME	DIRECT	DEPT_1	DEPT_2	DEPT_3	CHAR_1	CHAR_2	UNIT_PRICE	MARGIN_RATE	POPULARITY
8000020000000	6/17/2023	1	94641750 Kim Crawford Sauv...	0	Wine	White Wine	Sauvignon Blanc	Citrus	Grapefruit	20.99	0.04	0.24
8000020000000	6/30/2023	1	239796750 Ch de Laubade 21Y...	1	Spirits	Brandy & Cognac	Armagnac	Dried Fruit	Caramel	120.99	0.42	0.06
8000020000000	12/22/2022	1	127093750 Monkey Shoulder S...	0	Spirits	Scotch	Blended Scotch	Apricot	Apricot	30.99	0.17	0.11
8000030000000	3/3/2023	1	104139750 Ch Megyer Tokaji ...	1	Wine	White Wine	Furmint	Apple	Floral	17.99	0.6	1.0
8000030000000	12/17/2022	1	164292750 Iter Pinot Noir C...	1	Wine	Red Wine	Pinot Noir	Blueberry	Blackberry	18.99	0.53	0.03
8000030000000	12/17/2022	1	224527750 Agitator Bourbon ...	1	Wine	Red Wine	Cabernet Sauvignon	Toast	Spice	28.99	0.59	0.01
8000030000000	11/7/2023	0	224527750 Agitator Bourbon ...	1	Wine	Red Wine	Cabernet Sauvignon	Toast	Spice	28.99	0.59	0.01
8000040000000	4/6/2023	1	636750 Johnnie Walker Black	0	Spirits	Scotch	Blended Scotch	Smoke	Oak	33.99	0.11	0.25
8000040000000	4/6/2023	1	2812750 Jameson Irish Whi...	0	Spirits	Irish Whiskey	Irish Whiskey	Spice	Vanilla	24.99	0.1	0.66
8000040000000	4/6/2023	1	228903750 Gallant Watermelo...	1	Spirits	Vodka	Flavored Vodka	Watermelon	Spice	16.99	0.33	0.02

Results



- Logistic Regression Pipeline

- |-- Area Under ROC: 0.75
- |-- Accuracy: 0.73
- |-- Precision: 0.72
- |-- Recall: 0.73
- |-- F1 Score: 0.71



- Gradient Boosted Tree Pipeline

- |-- Area Under ROC: 0.79
- |-- Accuracy: 0.76
- |-- Precision: 0.75
- |-- Recall: 0.76
- |-- F1 Score: 0.75

Conclusion & Next Steps

- Machine learning models can predict purchasing behavior
 - Proven on customers with both browsing history and transactional history
- Next steps are to test this on non-transactional users
 - Users that should see the most impact from personalized recommendations
- Look into Scattertext, since everyone has a distain for word clouds