



WINE REVIEWS

FINAL PROJECT

Group 3 ម៉ោងការលេខ



WINE REVIEWS

เป็นชุดข้อมูลที่ประกอบด้วยข้อมูลที่
เกี่ยว กับการรีวิวไวน์จากเว็บไซต์
WINEMAG.COM ซึ่งเป็นเว็บไซต์
ข่าวสารและวิจารณ์ไวน์ที่มีชื่อเสียง
ชุดข้อมูลนี้ประกอบด้วย 150,929
ดาวและ 11 คอลัมน์

IMPORT DATA

ຕົວອຍ່າງຂ້ອມູນ

import data

```
wine_data = pd.read_csv('/content/drive/MyDrive/Data Mining/Project Final/winemag-data_first150k.csv')
wine_data = wine_data.iloc[:, 1:]
wine_data.head()
```

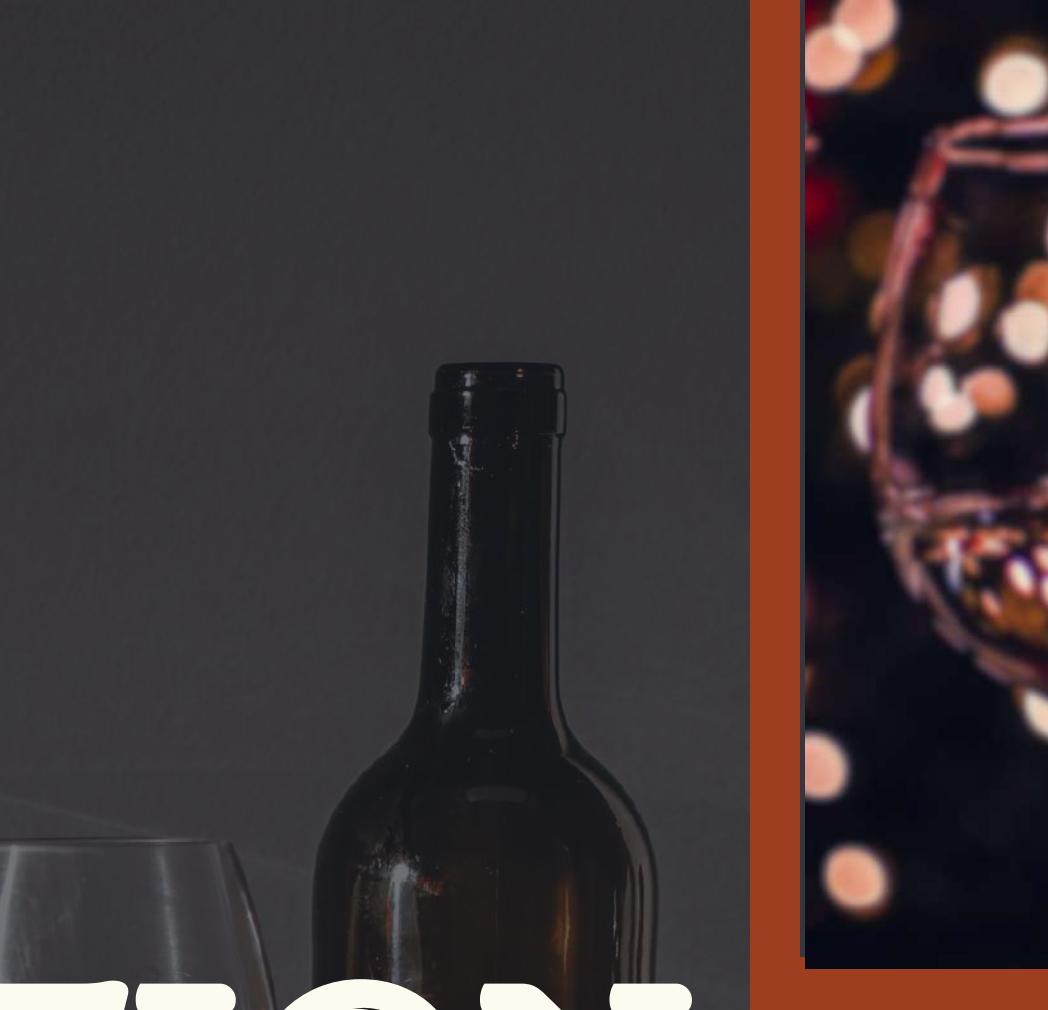
	country	description	designation	points	price	province	region_1	region_2	variety	winery
0	US	This tremendous 100% varietal wine hails from ...	Martha's Vineyard	96	235.0	California	Napa Valley	Napa	Cabernet Sauvignon	Heitz
1	Spain	Ripe aromas of fig, blackberry and cassis are ...	Carodorum Selección Especial Reserva	96	110.0	Northern Spain	Toro	NaN	Tinta de Toro	Bodega Carmen Rodríguez
2	US	Mac Watson honors the memory of a wine once ma...	Special Selected Late Harvest	96	90.0	California	Knights Valley	Sonoma	Sauvignon Blanc	Macauley
3	US	This spent 20 months in 30% new French oak, an...	Reserve	96	65.0	Oregon	Willamette Valley	Willamette Valley	Pinot Noir	Ponzi
4	France	This is the top wine from La Bégude, named aft...	La Brûlade	95	66.0	Provence	Bandol	NaN	Provence red blend	Domaine de la Bégude

CHECK
MISSING

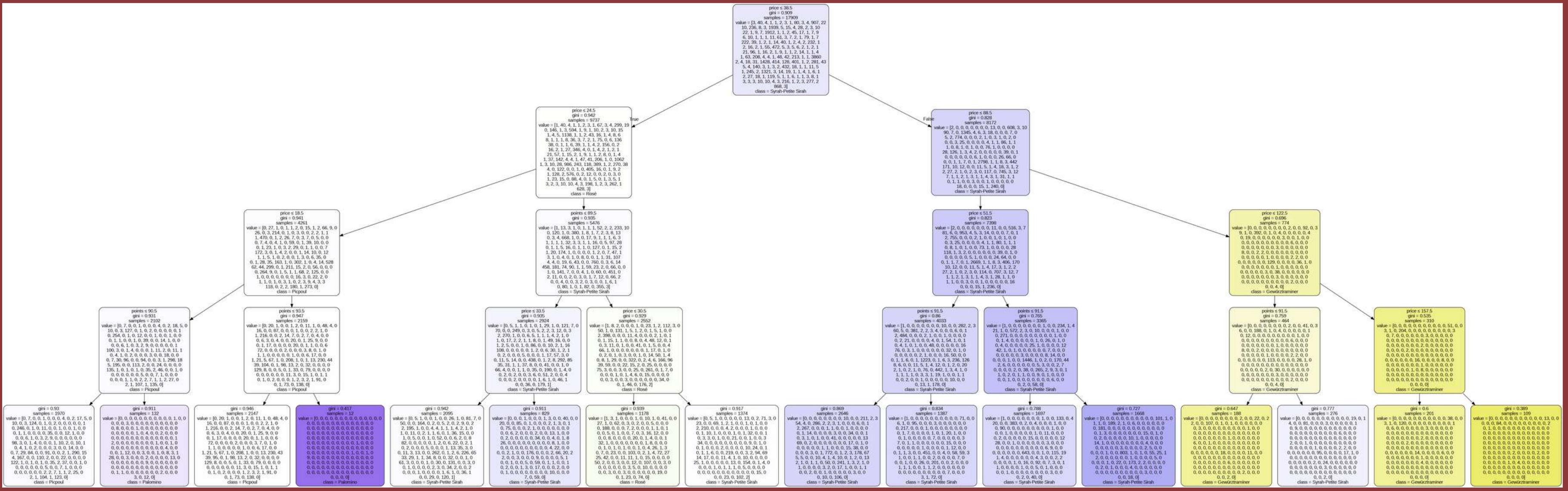
```
Unnamed: 0          False  
country            True  
description        False  
designation        True  
points             False  
price              True  
province           True  
region_1           True  
region_2           True  
taster_name        True  
taster_twitter_handle True  
title              False  
variety            True  
winery             False  
dtype: bool
```

ກໍາการຕຽບສອບ Missing
ແລະ
au Missing ໂດຍການ Dropna

CLASSIFICATION



DecisionTree



- ใช้คอลัมน์ POINTS และ PRICE เป็น FEATURES และคอลัมน์ VARIETY เป็น TARGET แบ่งข้อมูลออกเป็น TRAINING SET และ TEST SET โดยแบ่ง 80:20

ได้ค่าเท่ากับ **ACCURACY : 0.2541725060517263**

จึงสรุปได้ว่า โมเดล DECISION TREE นี้มีความแม่นยำในการคำนายน้อยมาก
อาจจะมีปัญหารือว่า OVERFITTING หรือว่าโมเดลไม่สามารถแยกแยะข้อมูลได้ดีพอสมควร



- ใช้คอลัมน์ POINTS และ PRICE เป็น FEATURES และคอลัมน์ VARIETY เป็น TARGET แบ่งข้อมูล ออกเป็น
- TRAINING SET และ TEST SET โดยแบ่ง 80:20

จาก NAIVE BAYES ได้ค่าความแม่นยำ (ACCURACY) เท่ากับ 0.19747738565422346 ซึ่งถือว่าต่ำมากๆ โดยจะต้องพิจารณาปรับปรุงโมเดล หรือเปลี่ยนโมเดลอื่นๆ เพื่อสามารถทำนายชนิดไวน์ได้อย่างแม่นยำขึ้น

NAIVE BAYS



k-Nearest Neighbor (k-NN)



- เลือกเฉพาะคอลัมน์ 'price', 'points', และ 'variety' เพื่อนำมาใช้ในการ Train Model
- เลือกคอลัมน์ 'price', 'points' เป็น features และให้ 'variety' เป็น target

โมเดล k-Nearest Neighbor (k-NN) โมเดลนี้มีความแม่นยำประมาณ 99% ในการคำนายนิดของใบบิลด้วยข้อมูล points, price และ variety ที่ใช้ในการ Train Model ซึ่งมีความแม่นยามาก

ได้ค่า ACCURACY เก้าก้าบ 0.9993867546495006

ACCURACY

DecisionTree

0.2541725060517263

k-Nearest
Neighbor (k-NN)

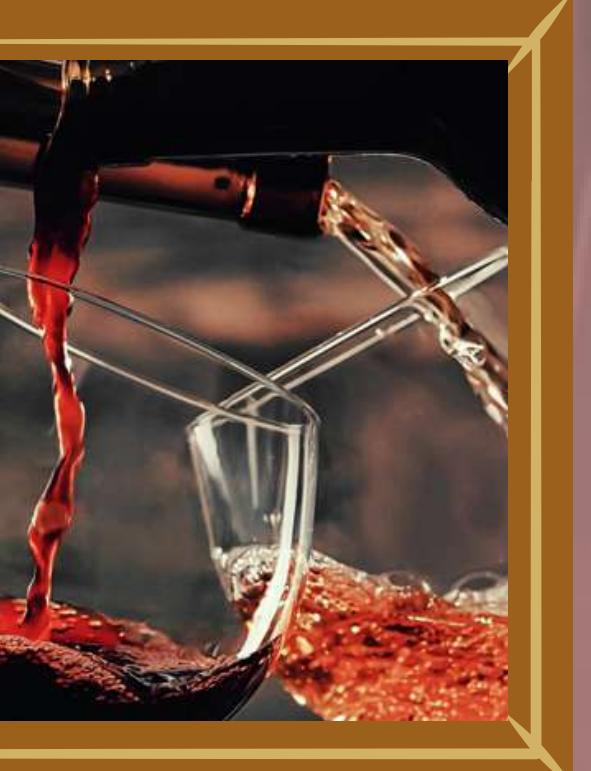
0.9993867546495006

Naive Bays

0.19747738565422346

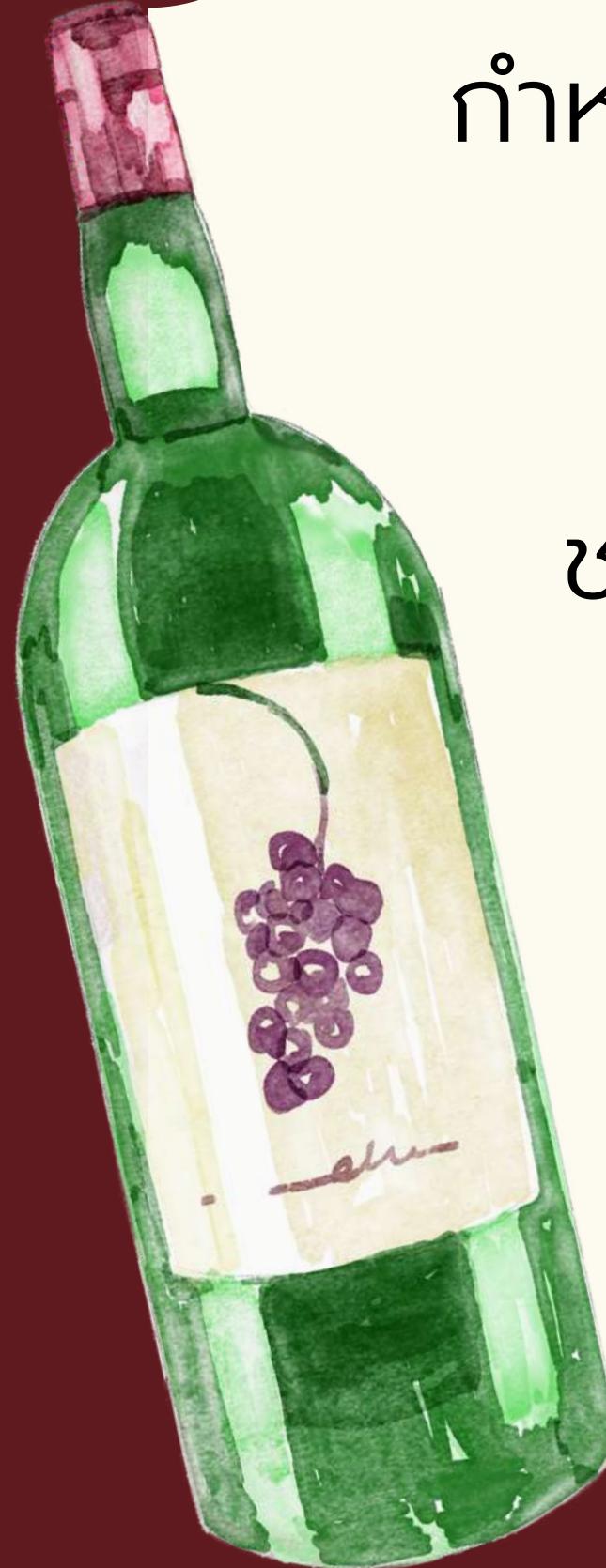


Association Rules



กำหนดเงื่อนไขในการกรองข้อมูลโดยต้องเป็นไวน์ที่มาจาก
ประเทศ US และมีคะแนน 90 ขึ้นไป

บันดขององค์ประกอบที่ใช้ในการผลิตไวน์ (variety)
และชื่อโรงงานผลิตไวน์ (winery)



	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
126	(Fumé Blanc)	(Robert Mondavi)	0.001782	0.003323	0.001011	0.567568	170.804935	0.001005	2.304816
127	(Robert Mondavi)	(Fumé Blanc)	0.003323	0.001782	0.001011	0.304348	170.804935	0.001005	1.434939

พิจารณค่า confidence ของ antecedents (Fumé Blanc) และ consequents (Robert Mondavi) คือ 0.567 หมายความว่า ไวน์ Fumé Blanc จะถูกซื้อพร้อมกับ Robert Mondavi ในอัตราส่วน 0.567 ของรายการที่ระบุไว้ในข้อมูล



	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
126	(Fumé Blanc)	(Robert Mondavi)	0.001782	0.003323	0.001011	0.567568	170.804935	0.001005	2.304816
127	(Robert Mondavi)	(Fumé Blanc)	0.003323	0.001782	0.001011	0.304348	170.804935	0.001005	1.434939

พิจารณาค่า lift ของ antecedents และ consequents เท่ากับ 170.804935 หมายความว่า การซื้อ Fumé Blanc และ Robert Mondavi พร้อมกันจะมีความน่าจะเป็นที่จะเกิดขึ้นเป็นจำนวน 170.804935 เท่ากับความน่าจะเป็นที่จะซื้อ Robert Mondavi และ Fumé Blanc พร้อมกัน โดยไม่คำนึงถึงกันเลย



K-MEANS CLUSTERING





เลือกคอลัมน์ description เป็นการแยกกลุ่มของรีวิวไวน์ตามลักษณะเด่นของรสชาติและกลิ่นของไวน์ เพื่อให้ผู้บริโภคสามารถเลือกซื้อไวน์ที่ตรงตามความต้องการและรสชาติของตนได้

โดยทำการแบ่งกลุ่มรีวิวไวน์ที่มีลักษณะคล้ายกันอยู่ในกลุ่มเดียวกัน แล้วเลือกเฉพาะคำที่เป็นคำเด่น (top terms) ที่อยู่ในกลุ่มนั้นๆ มาแสดงในกราฟ



ຕັວອຢ່າງຄຳ TOP TERMS ກ່ອຍຸໃນແຕ່ລະ CLUSTER

Cluster 0:

fruity, ripe, cherry, blackberry, plum, berry

Cluster 1:

citrus, lemon, apple, crisp, acidity, tart

Cluster 2:

oak, vanilla, spice, chocolate, black, tannin

Cluster 3:

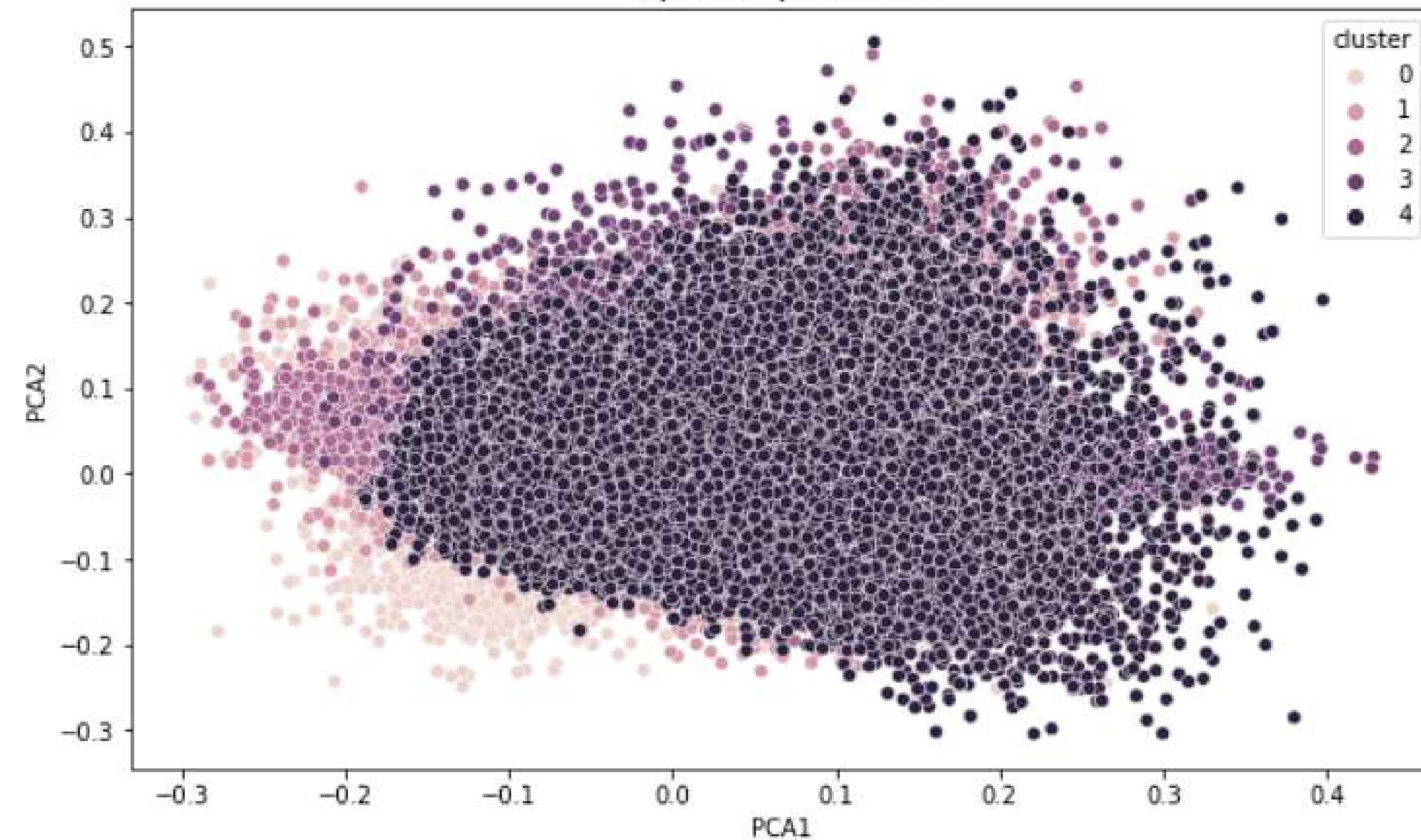
peach, apricot, floral, honey, pineapple, tropical

Cluster 4:

earthy, spice, herbal, tannic, blackberry, currant



Top Terms per Cluster



Our Group

633020440-9

633020447-5

633021013-3

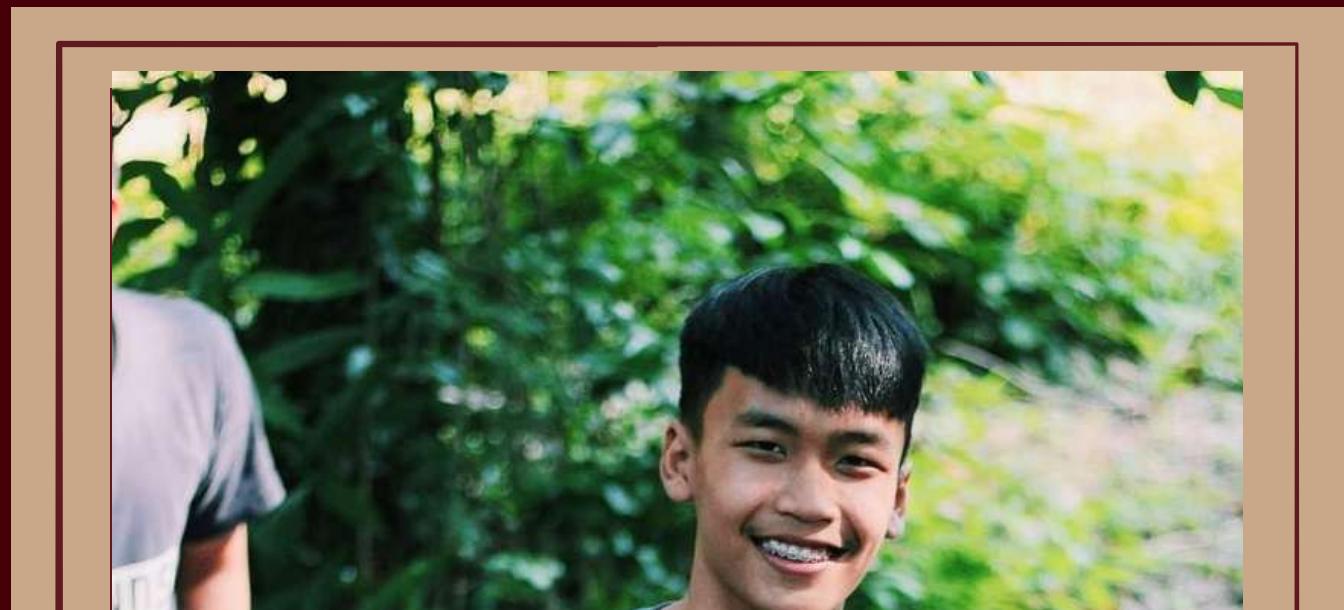
633021020-6

นางสาวณัฐริดา สรีเมงาม

นางสาวศุภัสรา คุริรัง

นายเฉลิมเกียรติ คำชนะ

นายพิทวัส พิมพุสิกธ์



REFERENCE

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

**THANK YOU
FOR
ATTENTION!**