COME WINE WITH ME

A WINE RECOMMENDATION CAPSTONE PROJECT

by Orla Cronin

Digital Futures

TABLE OF CONTENTS

THE PROBLEM

Wanting to try a new wine but having no idea where to start because the world is wine is vast

THE TASK

I want to be able to give an algorithm the characteristics of a wine I know I like and get back a recommendation that I know I'll enjoy

THE ACTION

Creating a unsupervised learning model to create predictions of wine recommendations based on user inputs

THE RESULT

Giving a demonstration of my function

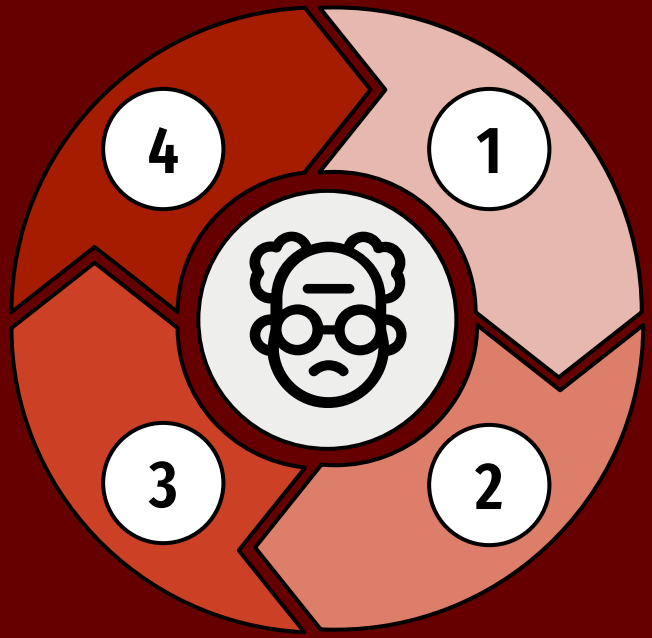
THE PROBLEM

Digital Futures


Mineral
Dairy
Nutty
Floral
Herbal
Savoury
Spice
Wood
Sweet
Fruity
Flint
Stones
Earth
Petrol
Butter
Cream
Yeast
Bread
Toast
Grilled Nuts
Biscuits
Almonds
Violets
Red Roses
Elderflowers
Blossom
Hay
Cut Grass
Asparagus
Green Pepper
Mint
Eucalyptus
Truffles
Game
Bacon
Leather
Ground Coffee
Cinnamon
Pepper
Liquorice
Smoke
Tobacco
Cedar
Oak
Vanilla
Honey
Butterscotch
Toffee
Chocolate
Jam
Fig
Prune
Raisin
Blackcurrant
Blackberry
Plum
Raspberry
Strawberry
Cherry
Lychee
Mango
Apricot
Peach
Banana
Melon
Pineapple
Grape
Gooseberry
Lime
Orange
Grapefruit
Lemon
Pear
Apple

THE PROBLEM

THE PROBLEM



Find a Good Wine

You found a wine you really like

Want a Similar Find

You want to try something new but don't know where to start

Confusing Purchase

Wine descriptions are confusing and you end up buying an overpriced bottle

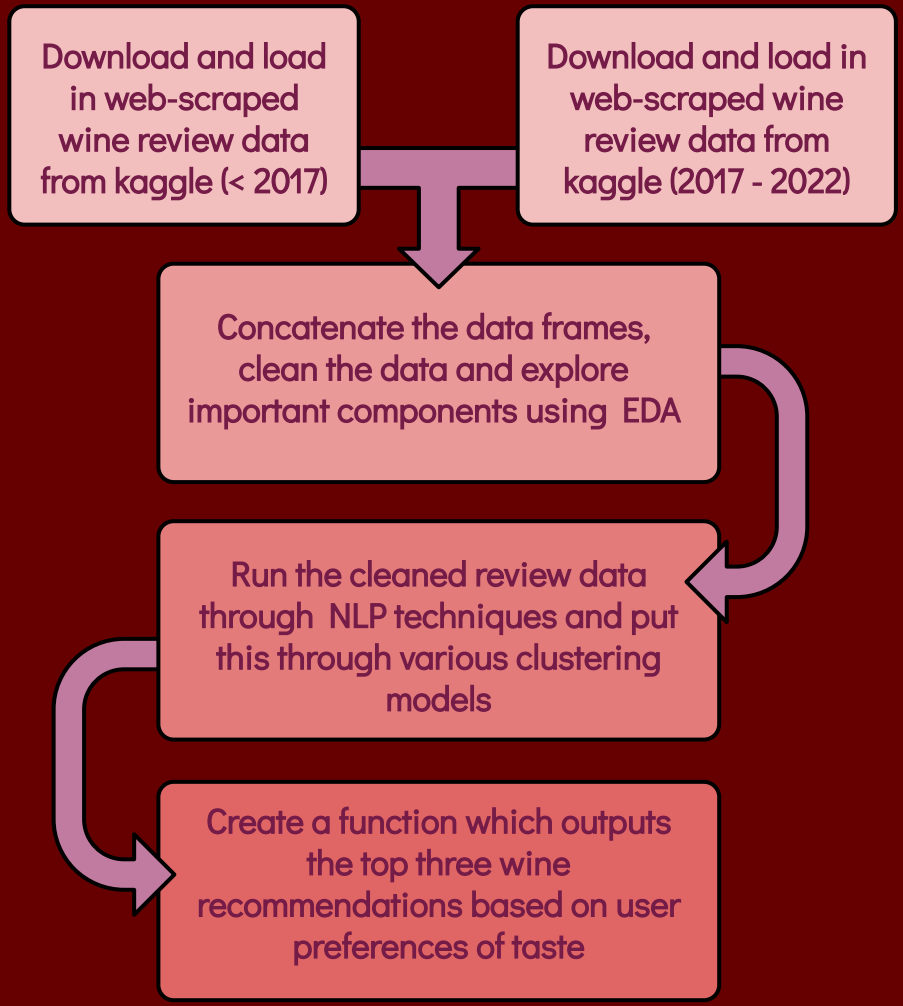
Dissatisfaction

The customer ends up dissatisfied with their purchase, a negative business outcome

THE TASK

Digital Futures

THE TASK


Download and load in web-scraped wine review data from kaggle (< 2017)
Download and load in web-scraped wine review data from kaggle (2017 - 2022)
Concatenate the data frames, clean the data and explore important components using EDA
Run the cleaned review data through NLP techniques and put this through various clustering models
Create a function which outputs the top three wine recommendations based on user preferences of taste

THE ACTION

Digital Futures

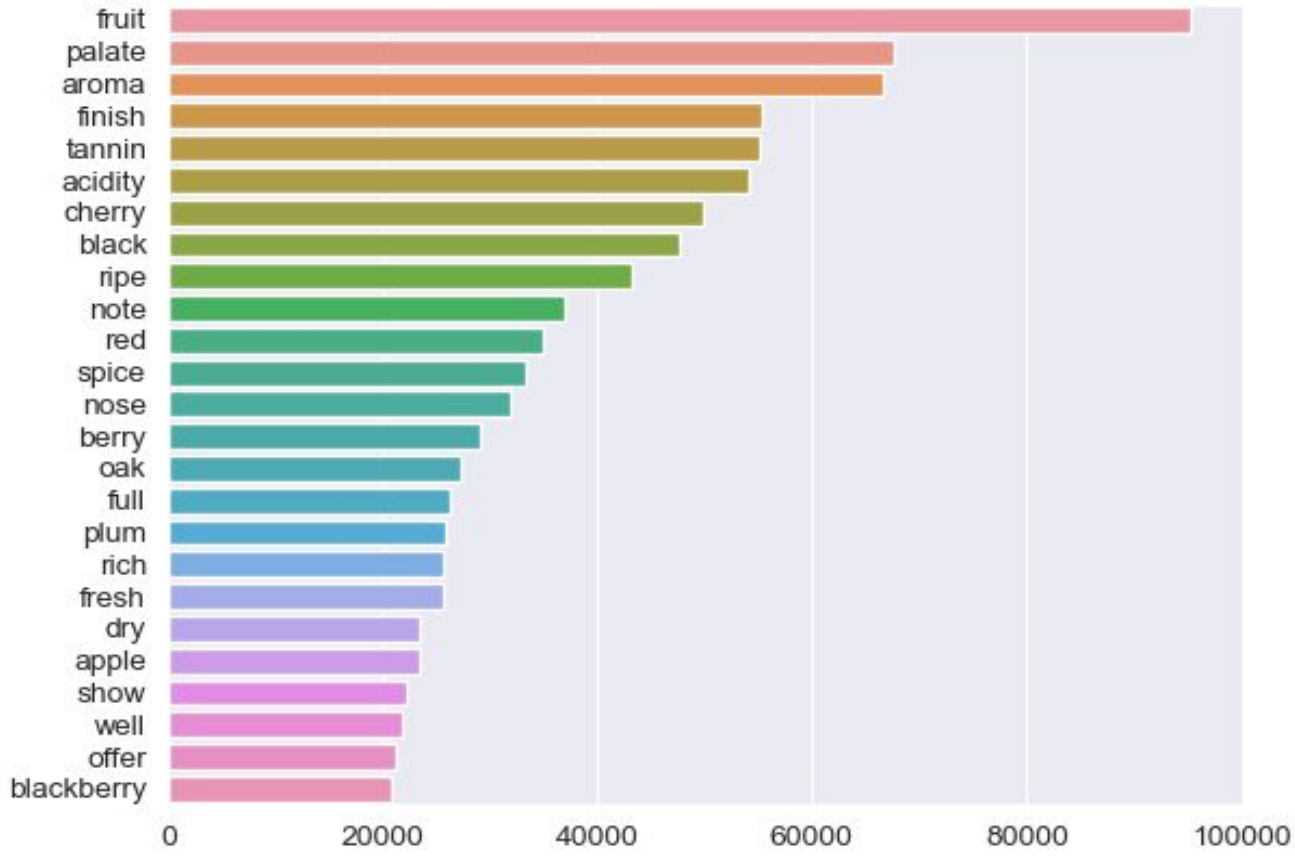
NUMERICAL COLUMN EDA

Attributes Altered	Data Features	
	Prior to Cleaning / EDA	Following Cleaning / EDA
Dropping Columns	14 columns	9 columns
Nulls?	yes	no: imputed or dropped
Dealing with Duplicates	9977 dupes	0 dupes
Removing Outliers	201005 rows	191688 rows

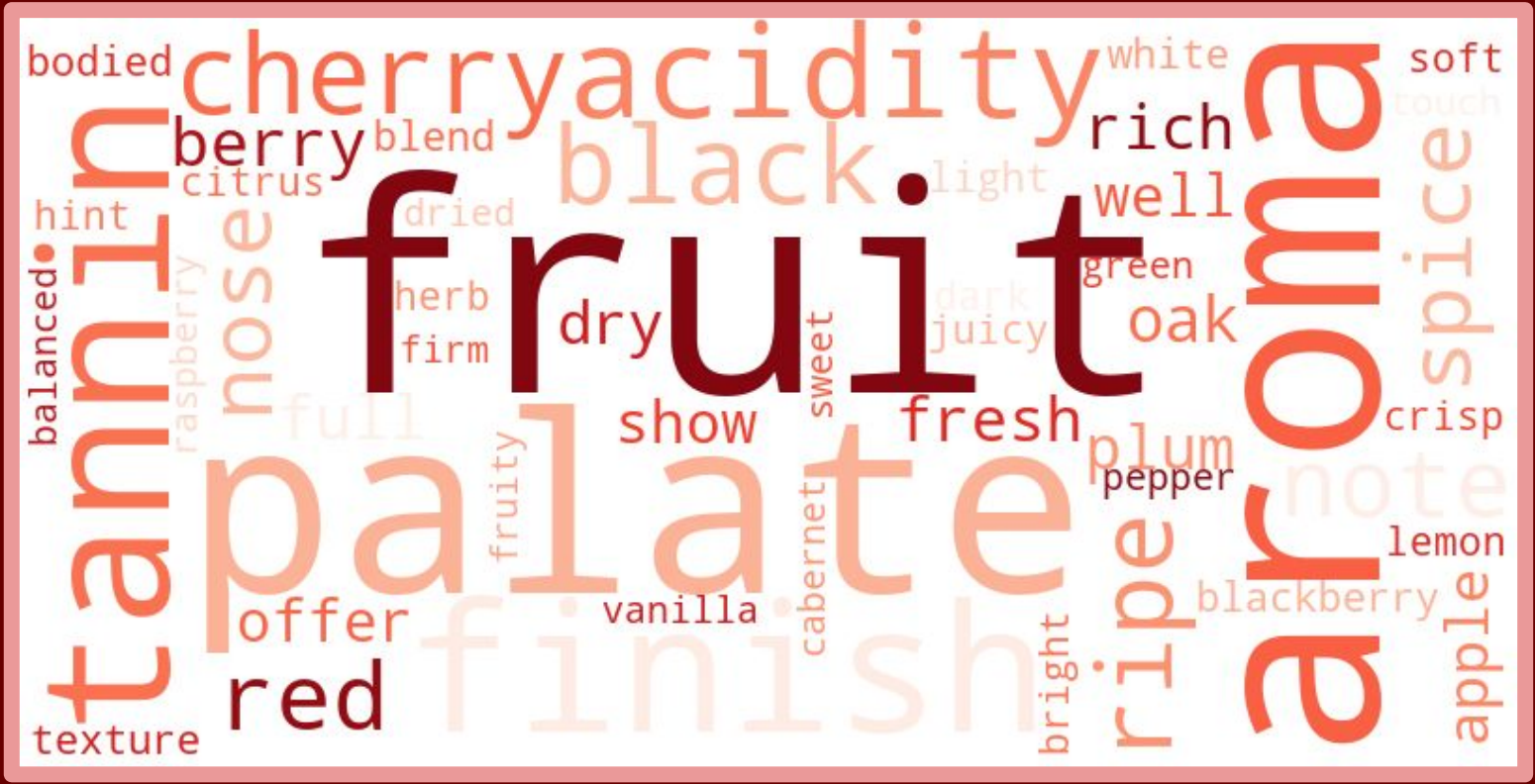
CATEGORICAL COLUMN EDA

★ The next step was to do EDA on the categorical columns, specifically to do some NLP data cleaning.

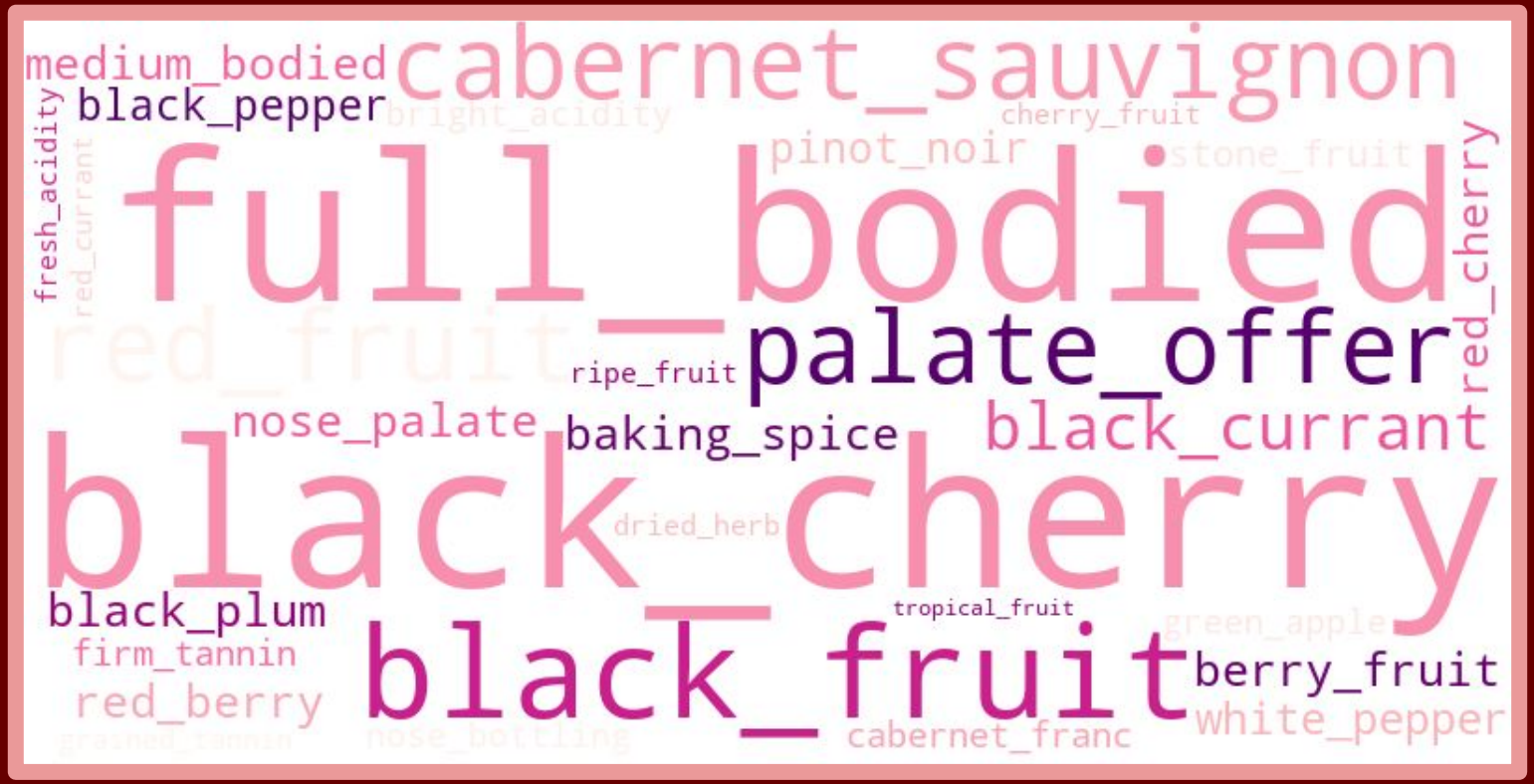
WORD COUNT


fruit
palate
aroma
finish
tannin
acidity
cherry
black
ripe
note
red
spice
nose
berry
oak
full
plum
rich
fresh
dry
apple
show
well
offer
blackberry
0
20000
40000
60000
80000
100000

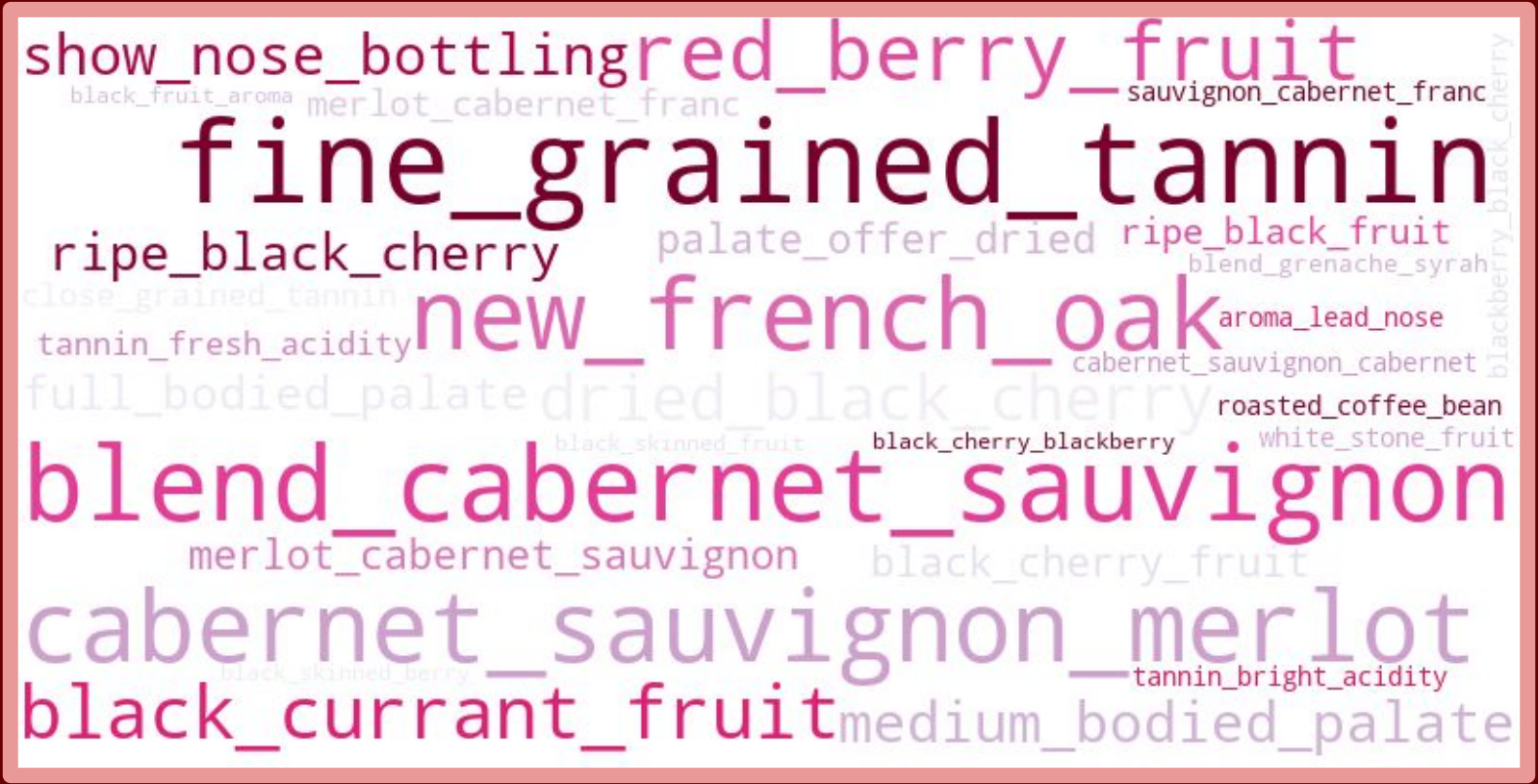
WORD CLOUD


bodied
cherryacidity
white
soft
berry
blend
rich
black
touch
citrus
light
well
hint
dried
fruit
aroma
tannin
spice
green
herb
dry
oak
nose
dark
juicy
balanced
raspberry
firm
sweet
full
show
fresh
crisp
palate
plum
note
fruity
pepper
cabernet
lemon
offer
vanilla
blackberry
finish
ripe
apple
red
bright
texture

BI-GRAM WORD CLOUD



TRI-GRAM WORD CLOUD

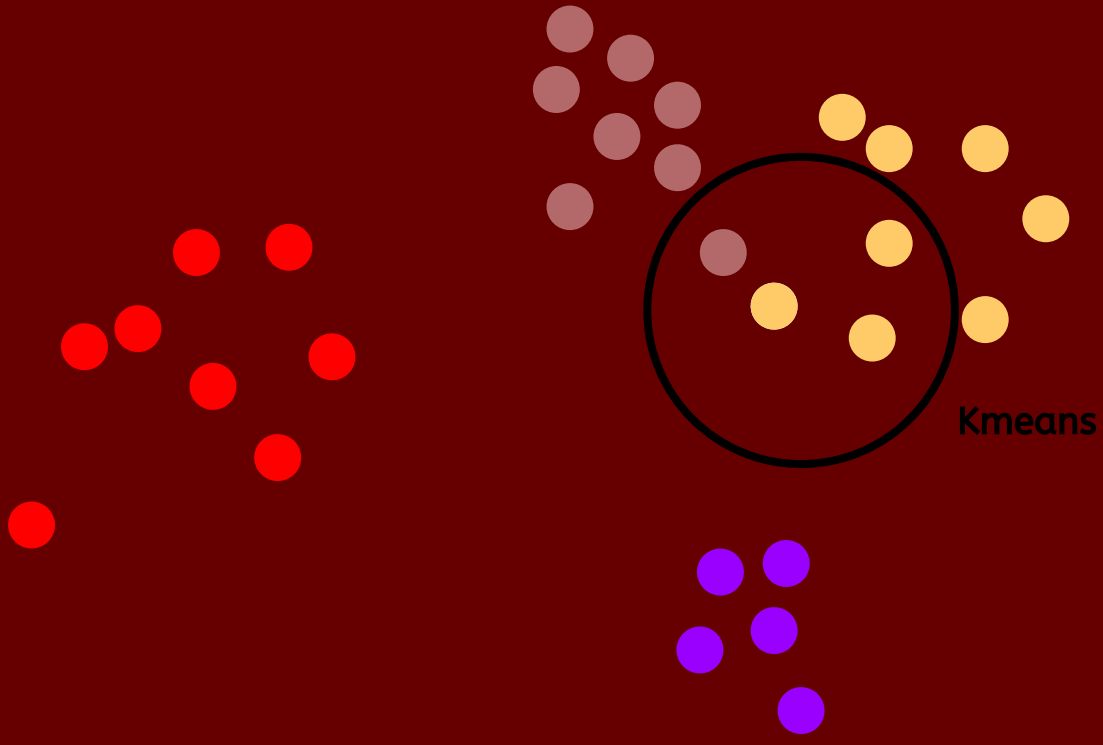


VECTORISATION

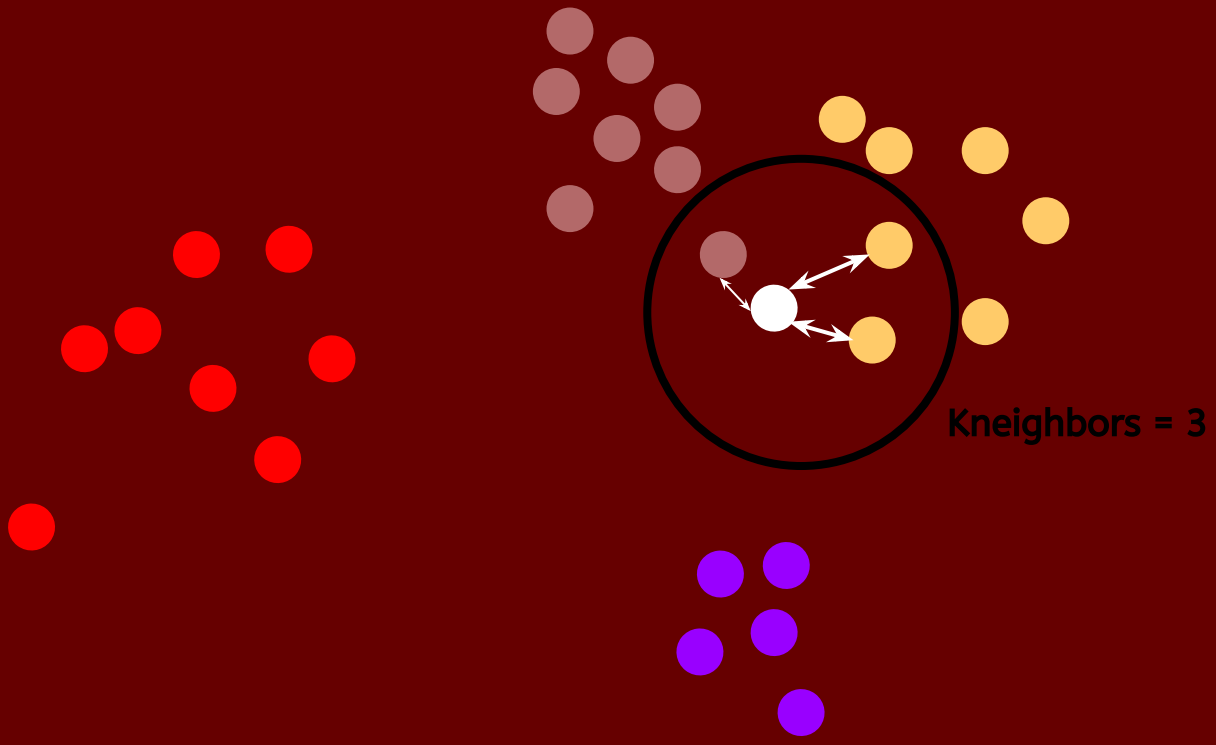
★ TF-IDF vectorisation created a weighted matrix of wine descriptors which could be fed into the clustering models

index	TF-IDF values:	
	cherry	chocolate
1	0.000	0.455
2	0.000	0.000
3	0.138	0.000
4	0.000	0.202868

CLUSTERING



CLUSTERING



★ DEMO ★

(please noone break it)

Digital Futures

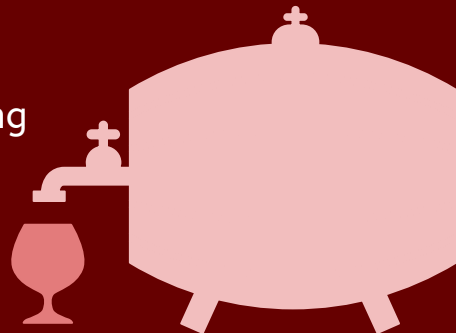
FUTURE PROOFING

IMPROVEMENTS

- Creating separate models for each wine subtype
- Incorporating prices into the user selection customise output further
- As an extension, create a wine-food pairing system

LIMITATIONS

- Inexperienced user inputs of wine characteristics could lead to inaccurate recommendations
- Being unsupervised its difficult to interpret the accuracy of results
- Include error handling

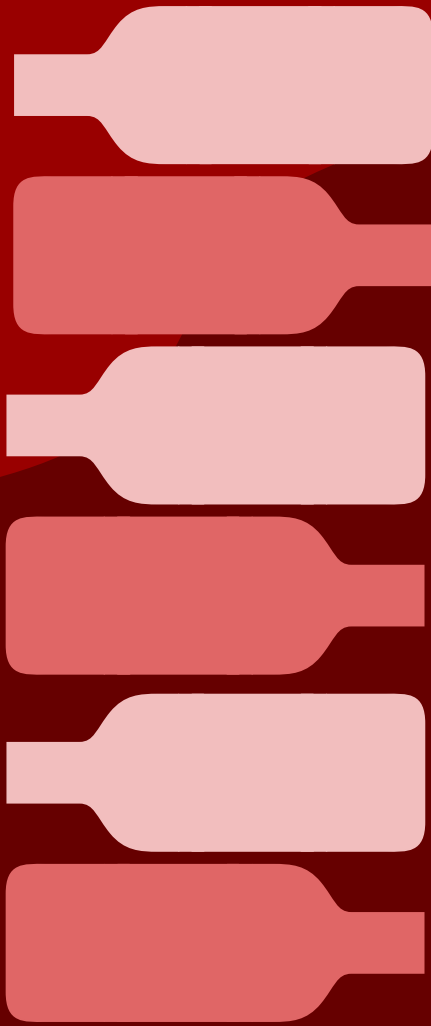


SUMMARY:

Objective: develop an easy to use wine recommendation system to help consumers navigate the over-complicated world of wine flavours

Solution: built an unsupervised clustering algorithm that can recommend best-fitting wines based on customer preferences

Findings: were my aims met? Did i achieve a easy to use wine recommendation system? Did i do this well?



ANY QUESTIONS?

thank you guys :)

Digital Futures