Winery Recommendation Engine

Xinyuan Shu

July.12.2022

Background

A e-commerce startup company want to design an winery recommendation system to attract more visitors for their websites.

We need to establish a winery recommendation engine for them

Design

By using K-Means, we clustered 130k blind-test reviews into 10 groups

Each group has a special ‘taste’ for wines

Build engine upon clusters by using cosine-similarity

Data

Data comes from Kaggle program-wine reviews

Rows:130k wine reviews

Columns(used): description/variety/winery/province/region

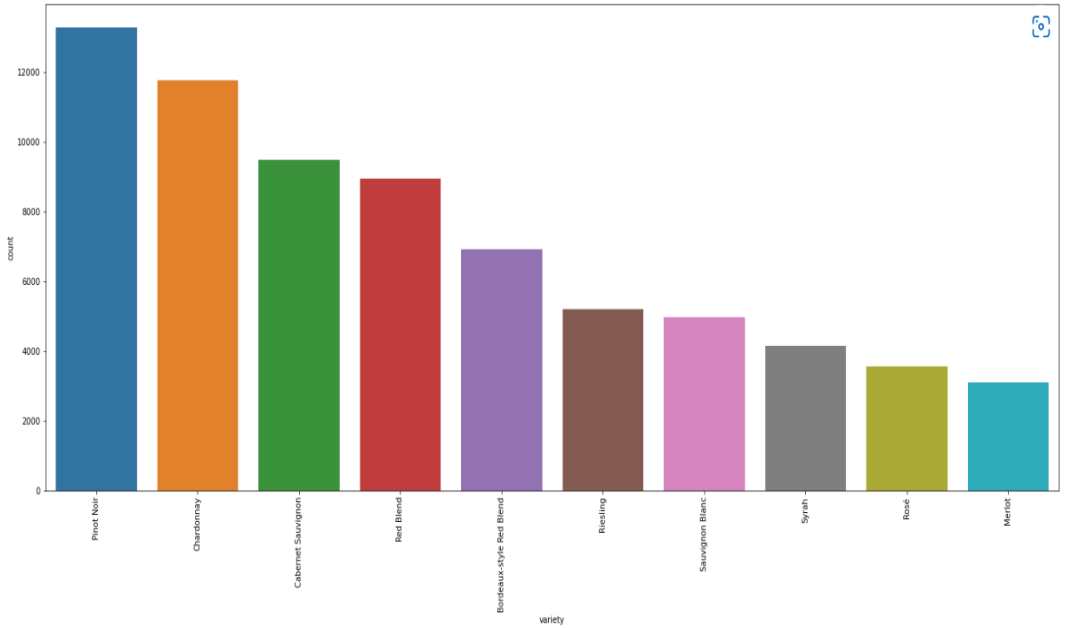
Algorithms-1

Reduce varieties
from 1500→10

NLP:
Tf - idf vectorizer
stem/lemmatize

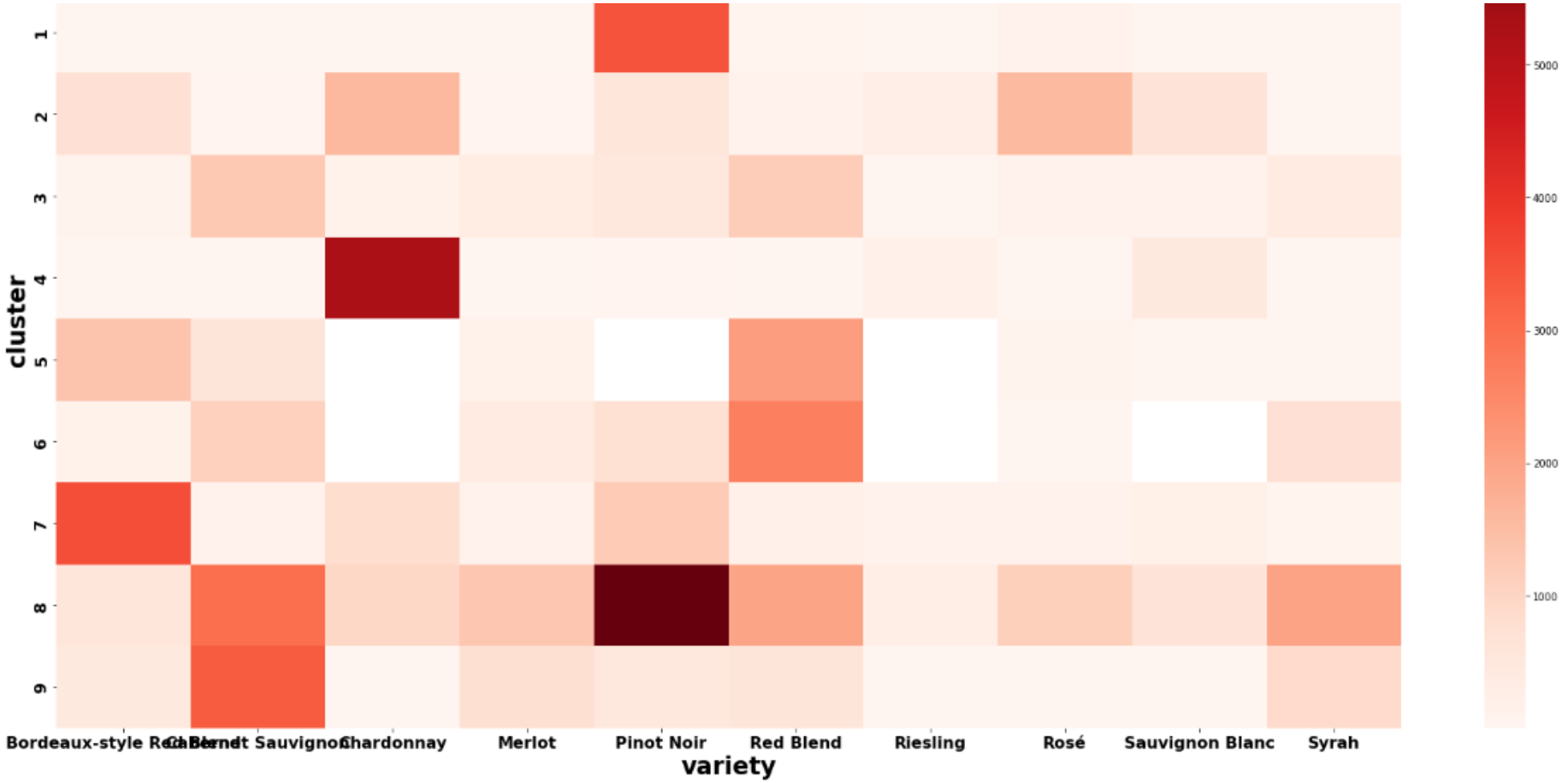
K Means cluster
by 10 groups

Algorithms-2

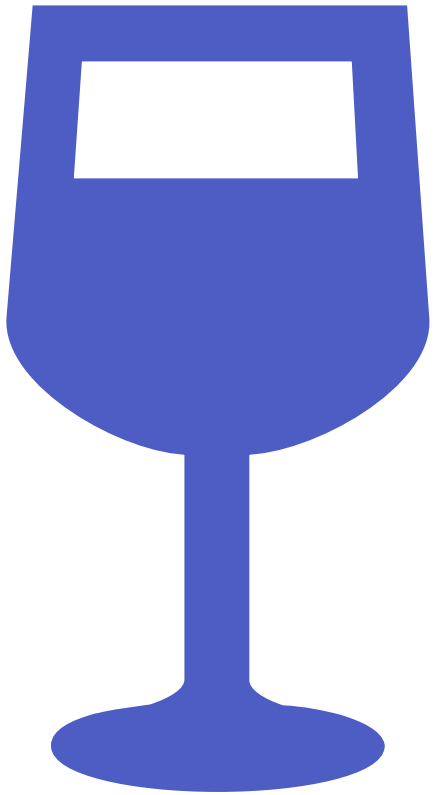


Pinot Noir/ Chardonnay/ Cabernet Sauvignon/ Red Blend/ Bordeaux-style Red Blend/ Riesling/ Sauvignon Blanc/ Syrah/ Rosé/ Merlot

indices of the top ten frequented words in each cluster



Almost 1-1 relationship by eyeball analysis



Algorithms-3

- Engine for the winery
  - Count Vectorizer
  - Cosine similarity

  (features: region_1','province','variety','cluster','description)
- Customer concern

Tool Summary

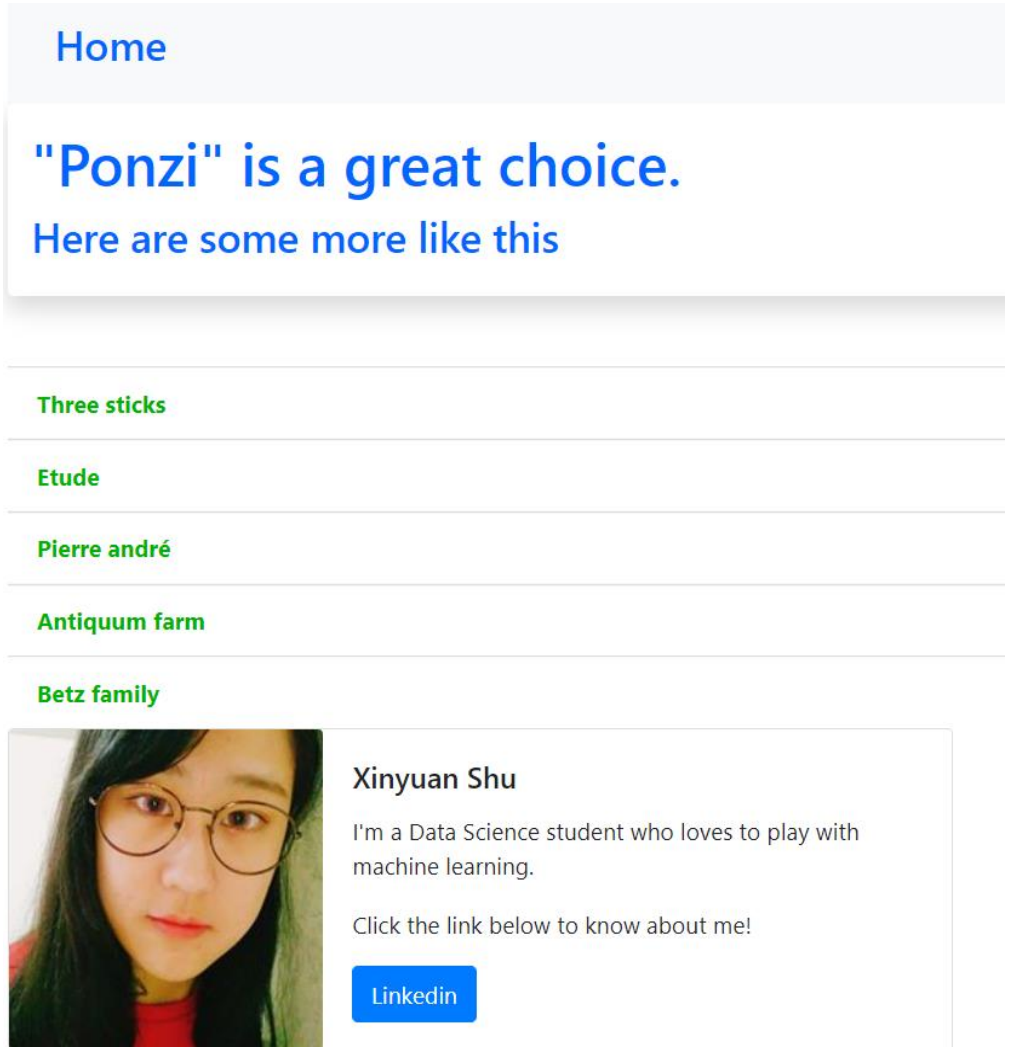
+ Pandas / NumPy / seaborn
+ Kmeans /cosine similarity
+ Matplotlib/WordCloud
+ TfidfVectorizer / RegexpTokenizer / Snowball Stemmer / Count Vectorizer

Results

+ Flask app
+ Three-sticks/Ponzi comparison:

  Main wines:

  pinot noir/ chardonnay




Thank you