

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 app for them

Design-pipeline



Data download from Kaggle and ingested into DB Browser



Recommendation engine building by python



Data preprocessing using pySpark



Flask app building



Further processing by Python



Pushing flask app on Heroku

Data



Data comes from Kaggle program-wine reviews



Rows:130k wine reviews



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

Algorithms

Reduce varieties from 1500→10

NLP: Tf - idf vectorizer stem/lemmatize

K Means cluster by 10 groups



Algorithms

+ Engine for the winery

Count Vectorizer

Cosine similarity

(features:

region_1','province','variety','clust er','description)

Tool Summary

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

Home

Welcome to the winery recommender

Enter the winery you like to get recommendations.

Ponzi

Enter



Xinyuan Shu

I'm a Data Science student who loves to play with machine learning.

Click the link below to know about me!

Home

"Ponzi" is a great choice.

Here are some more like this

Three sticks

Etude

Pierre andré

Antiquum farm

Betz family



Xinyuan Shu

I'm a Data Science student who loves to play with machine learning.

Click the link below to know about me!



Home

Welcome to the winery recommender

Enter the winery you like to get recommendations.

Test

Enter



Xinyuan Shu

I'm a Data Science student who loves to play with machine learning.

Click the link below to know about me!

Home

Sorry! This winery is not in our database. Please check if you spelled it correct. Or try with another winery.



Xinyuan Shu

I'm a Data Science student who loves to play with machine learning.

Click the link below to know about me!







Shu

Xinyuan

Q

马上加入





Xinyuan Shu

Data Science& Analytics|UW-Madison Econ Student 美国 威斯康辛州 麦迪逊 7 位好友

加入领英,建立联系

Harvard College Association for U.S.-China Relations...

🗑 美国威斯康星大学麦迪逊分校

ℰ 个人网站 ☑

看过本页的会员还看了



Xingyuan Su

TEST ENGINEER at Bookham Inc

深圳



Cheng Li

US Corporate Credit

波士顿地区



Skylar Meng

Emory University, QSS-Informatics; Baylor University, BS in Statistics

美国



Hannah Chu

Biocl ☑ ⊕ Ø Ø y Stud

Further Steps

+ Move the whole app to the AWS cloud

+ Try to write an API to gain Data from the source page

