Winery Recommendation Application
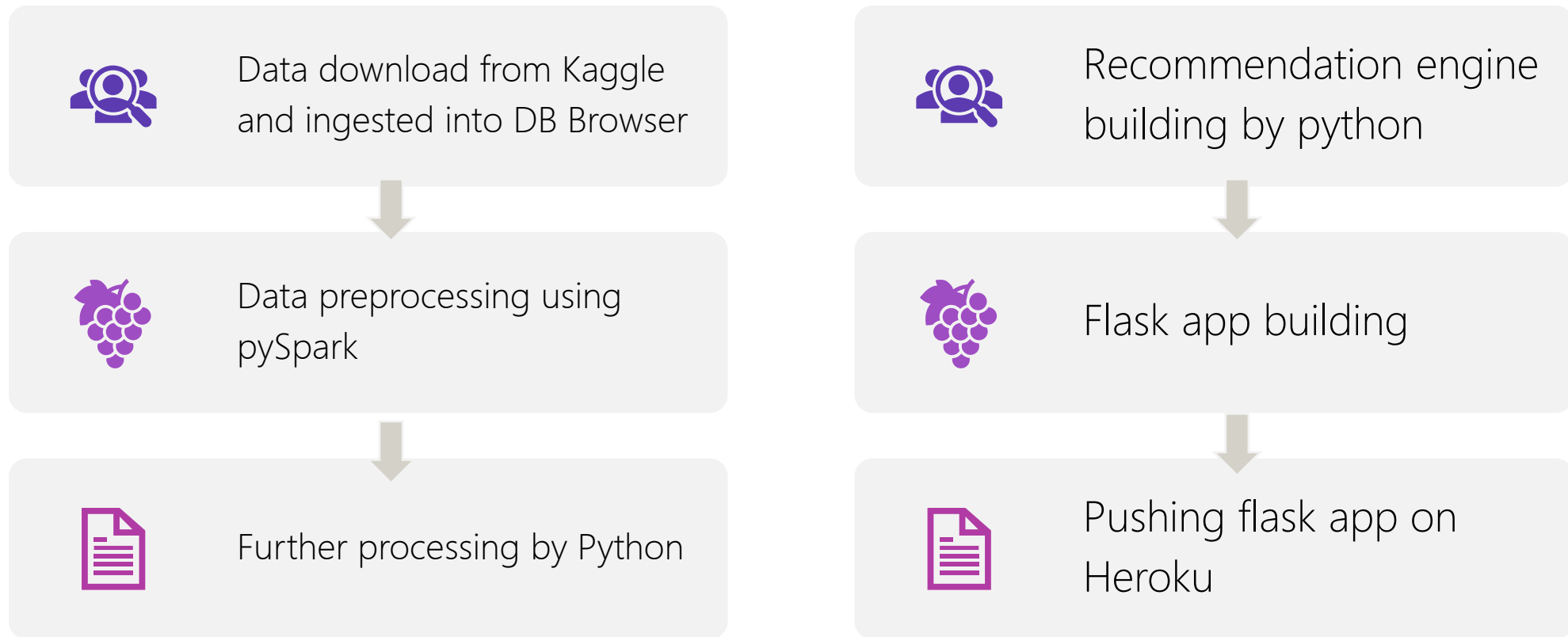
Xinyuan Shu

Sept.6.2022

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 app for them

Design-pipeline

Data download from Kaggle and ingested into DB Browser

Data preprocessing using pySpark

Further processing by Python

Recommendation engine building by python

Flask app building

Pushing flask app on Heroku

Data

Data comes from Kaggle program-wine reviews

Rows:130k wine reviews

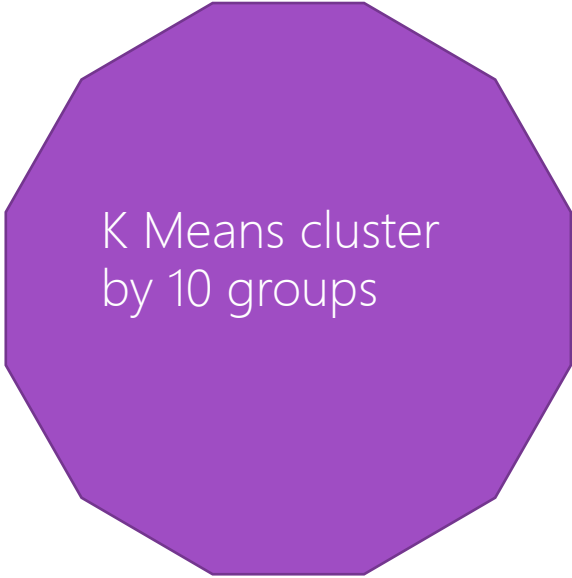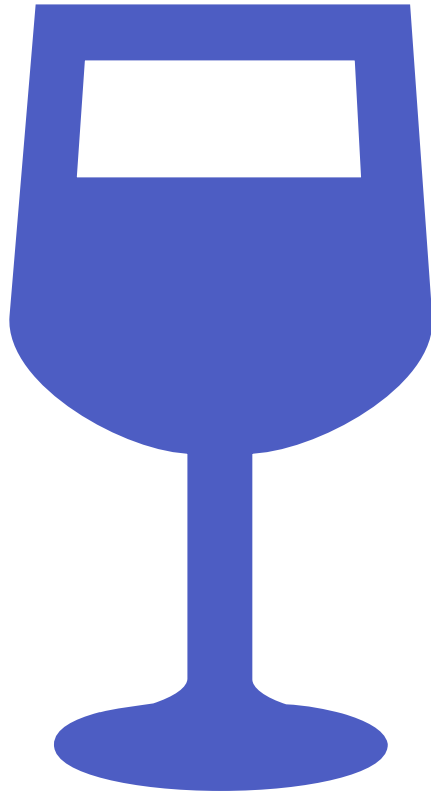Columns(used): description/variety/winery/province/region

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','cluster','description)

Tool Summary

+ Pandas / NumPy / seaborn

+ Flask

+ Kmeans /cosine similarity

+ Matplotlib/WordCloud

+ TfidfVectorizer / RegexpTokenizer /
  Snowball Stemmer / Count Vectorizer

Results

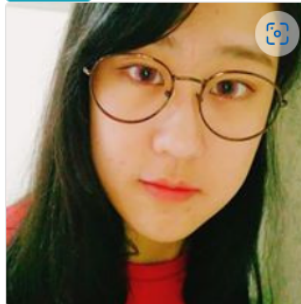# Results

Results

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!

Linkedin

Results

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!

Linkedin

Results

Further Steps

+ Move the whole app to the AWS cloud

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

Thank you