**NLP Project Write-up**

**Winery recommendation system**

**Abstract**

* 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**

*Feature Engineering*

1. Reduce varieties from 1500→10
2. NLP: Tf - idf vectorizer/stem/lemmatize
3. K Means cluster by 10 groups
4. eyeball analysis the 1-1 relationship from clusters to wine varieties
5. Count Vectorizer/Cosine similarity to build engine(Feature: features: region\_1','province','variety','cluster','description)

**Models**

K-means cluster

**Tools**

* Pandas / NumPy / seaborn
* Kmeans /cosine similarity
* Matplotlib/WordCloud
* TfidfVectorizer / RegexpTokenizer /
* Snowball Stemmer / Count Vectorizer
* Matplotlib and Seaborn for plotting

**Communication**

In addition to the slides and visuals presented, I also built a flask app to visualize my engine, LinkedIn link are also pasted.(see next page)

图形用户界面, 文本, 应用程序

描述已自动生成图形用户界面, 文本, 应用程序

描述已自动生成图形用户界面, 应用程序

描述已自动生成