

DisneyLand Reviews Analysis

Natural Language Processing and Unsupervised Learning

By: Mai Aljuaid & Basma Alduaiji



Table Of Contents



1

Introduction



2

Workflow



3

Data & Design



4

Tools



5

Data Visualization



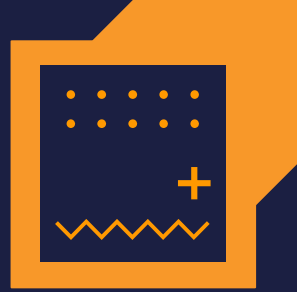
6

Models



7

Conclusion



Introduction



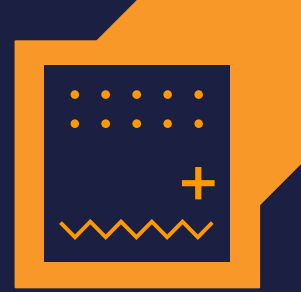
Introduction

- Walt Disney is the pioneer of animated cartoon films.
- Disneyland Resort in Anaheim, California.
- Opened on 1955.
- Disneyland branches have More than 52 million visitors yearly.



DisneyLand Branches

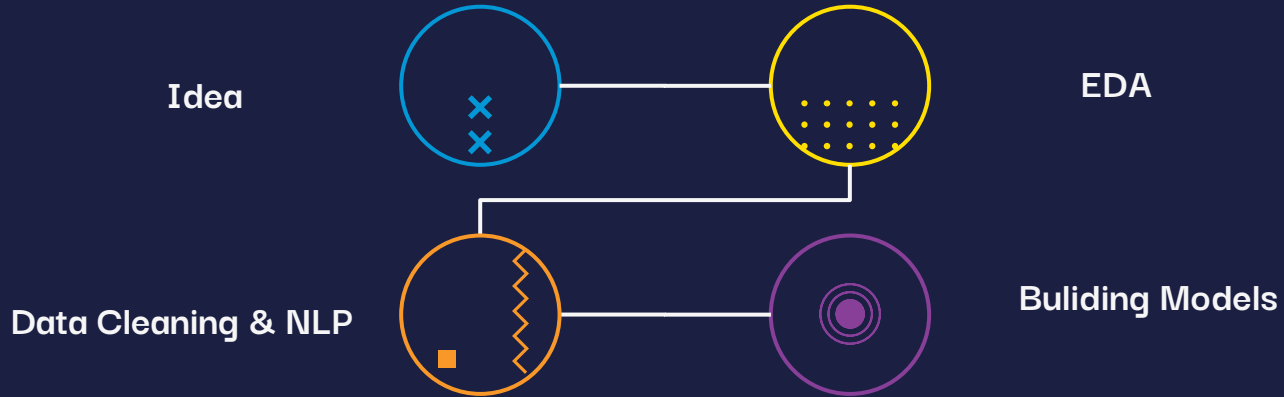


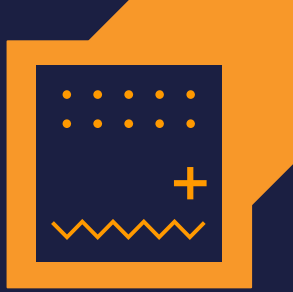


Workflow



Workflow





Data & Design





Data & Design

20,000

Text Reviews

5

Columns

Data Cleaning

Checking for
Null values

Duplicate
Rows

Renaming
Columns

Drop **unnecessary**
columns







Natural Language Preprocessing



**Remove
Frequent Words**



**Remove Rare
Words**



Stemming



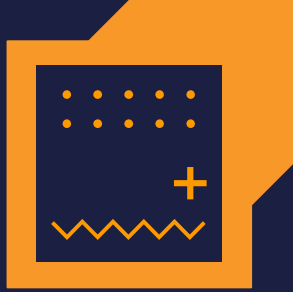
Lemmatization



**Spell
Correction**



**Remove
Emoji**



Tools





Tools

Numpy

Pandas

Matplotlib,
Seaborn

Sklearn

NLTK

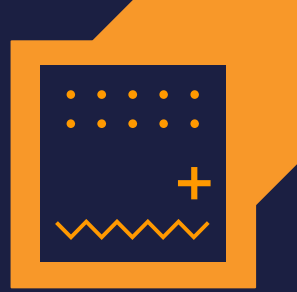
Plotly

Emoji

Regular
Expression

Folium &
WordCloud

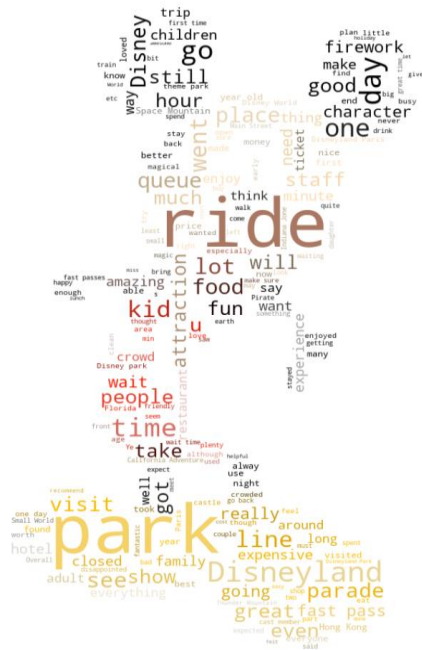




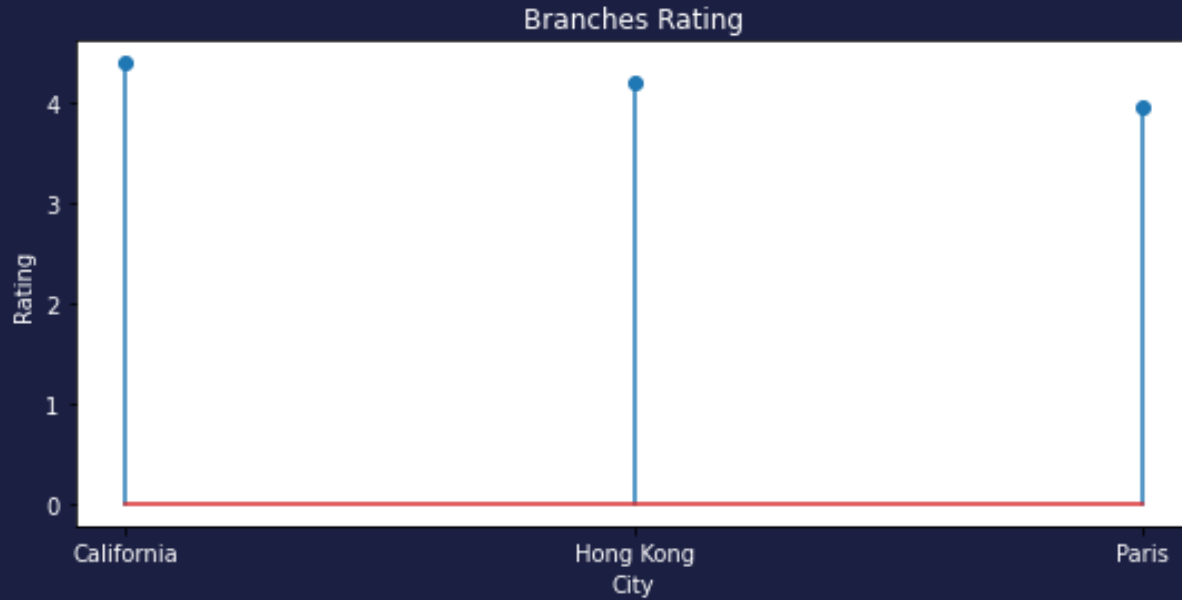
Data Visualization



Word Cloud Graph



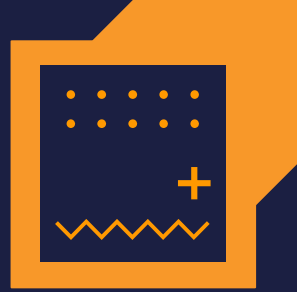
Branches Rating





Top 7 Reviewers Countries





Models



Experiments

K-means

LDA

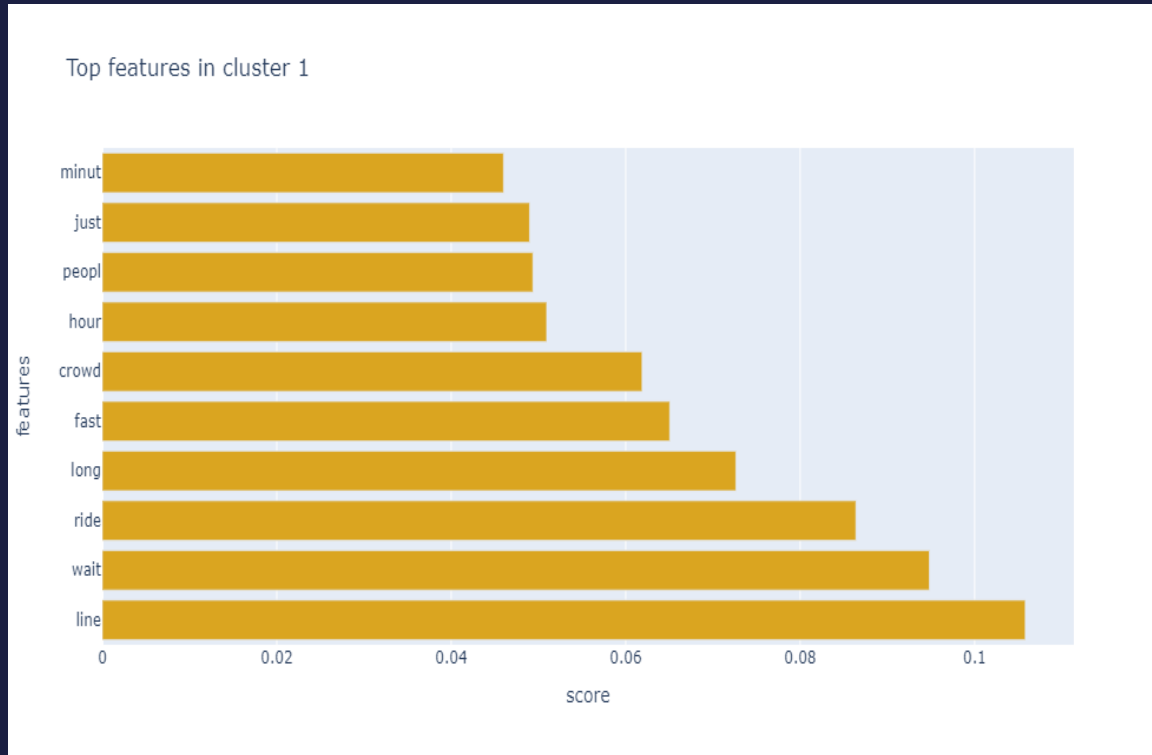
NMF

Supervised
Learning

NMF
Supervised
Learning

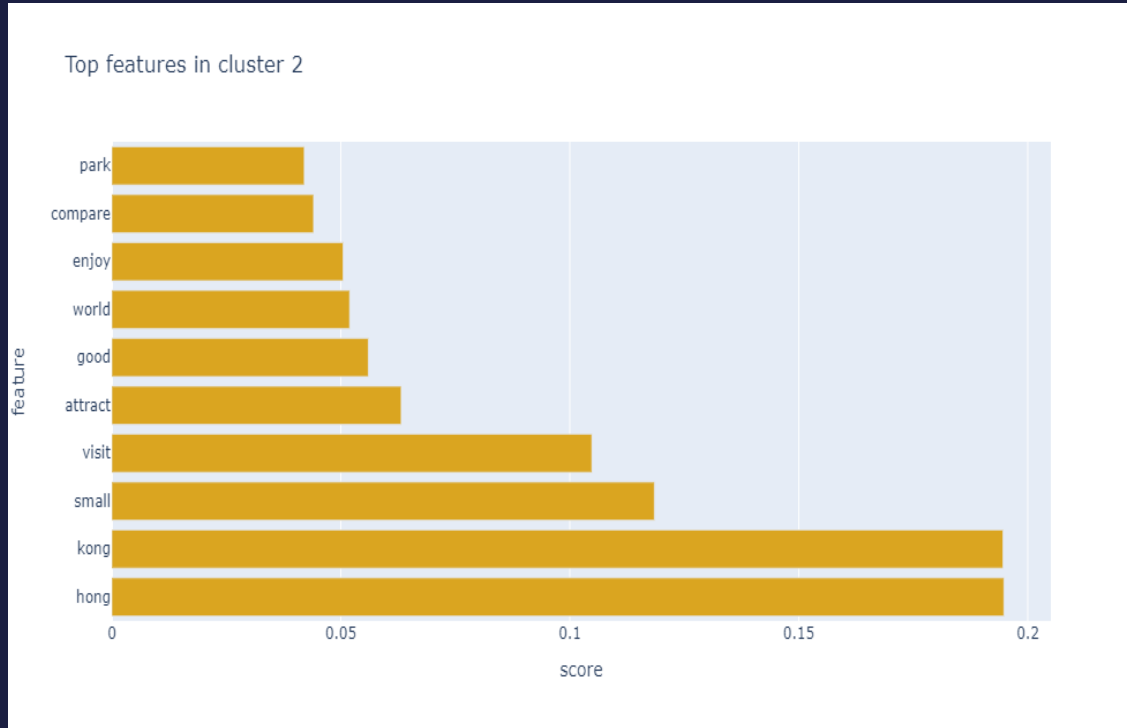


K-Means Clustering Result



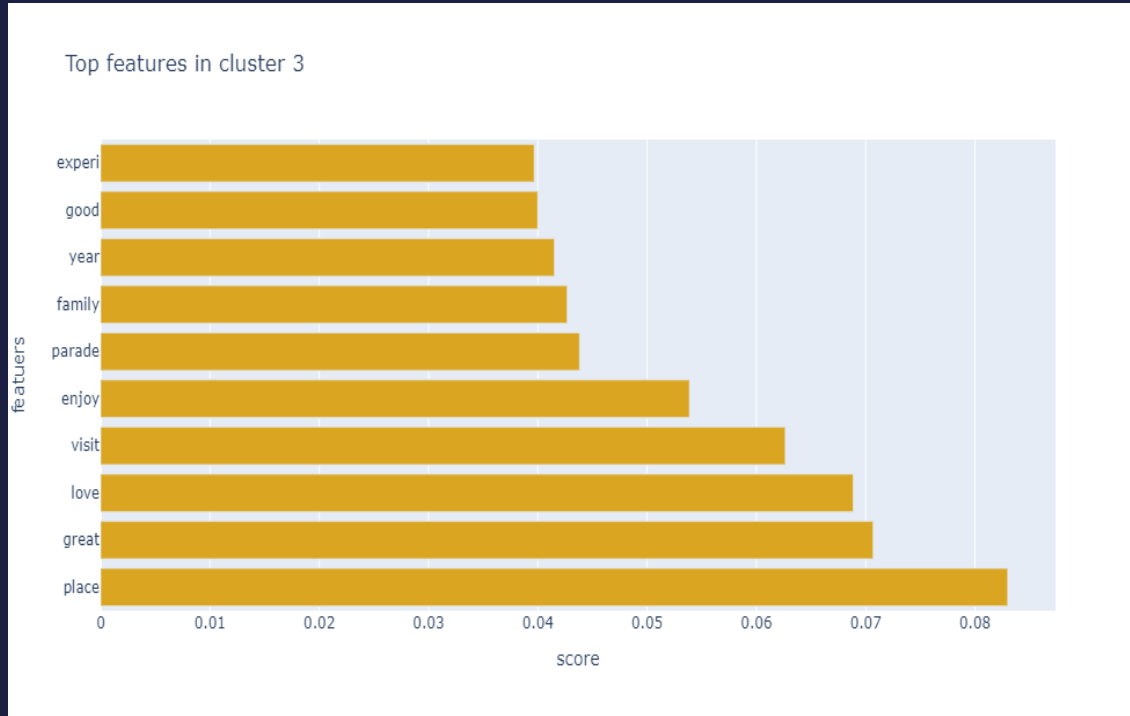
Ride

K-Means Clustering Result



Hong Kong Branch
Comparing

K-Means Clustering Result



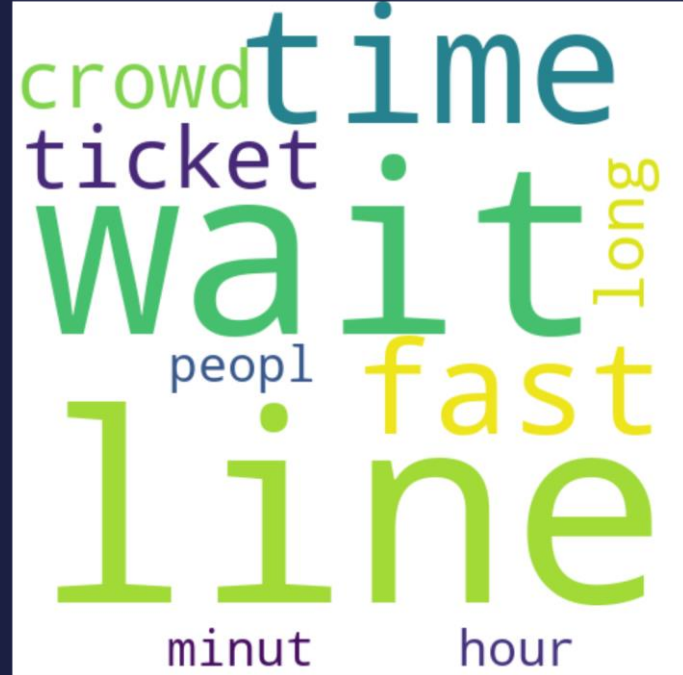
Great Experience

NMF Topics Modeling



A word cloud representing a topic. The words are arranged in a vertical stack. The largest word is 'place' in green. Below it is 'great' in purple. Then 'love' in purple. Then 'good' in green. At the bottom is 'firework' in yellow. Other words include 'make' in green, 'parade' in blue, and 'really' in blue.

place
make great
love
good parade
really
firework



A word cloud representing a topic. The words are arranged in a vertical stack. The largest word is 'wait' in green. Below it is 'line' in green. Then 'fast' in yellow. At the bottom is 'hour' in blue. Other words include 'time' in blue, 'crowd' in green, 'ticket' in purple, 'people' in blue, 'long' in yellow, and 'minute' in blue.

time
crowd ticket
wait long
people fast
line
minute hour

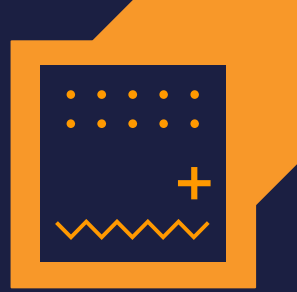
Scores

Classification

	LR	KNN	RF
Val	0.61	0.52	0.55
Test	0.59	0.52	0.54

NMF and Classification

	LR	KNN	RF
Val	0.83	0.80	0.81
Test	0.83	0.81	0.82



Conclusion



Conclusion

.....

- NMF is good in displaying topics.
- The results show that people love Disneyland despite the long waiting hours in rides.



× ×



x
x



+

Thanks