Making the happiest place on earth – Happier!

Disneyland Reviews Sentiment Analysis and Topic Modeling

Metis – NLP Unsupervised
Elisabeth Johnson
08/08/2022

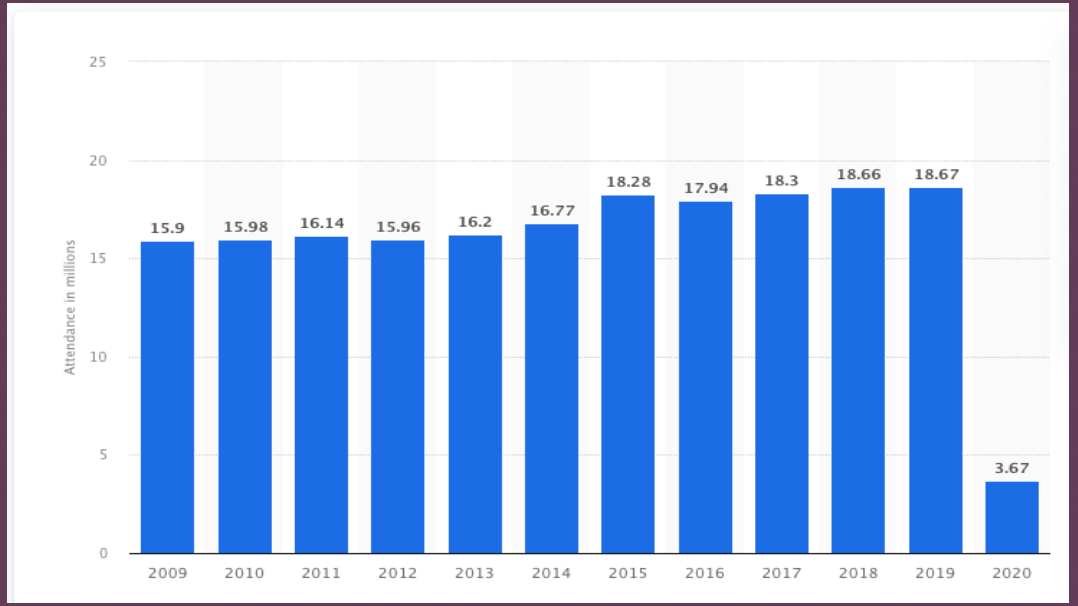
Q: Why Disneyland?

A: Due to COVID and various closures, Disneyland is in the process of recovering historical attendance volumes.

Disneyland Facts

- Established in 1955, Disneyland has the largest cumulative attendance of any theme park in the world.
- As of December 2018, Disneyland has accumulated over 700M visits to the park.
- Disneyland California alone has maintained over 15M visits to the park on an annual basis with the exception of year 2020 due to COVID restrictions.

Disneyland California Attendance



* Image from https://www.statista.com/

Analysis Aim:

Understand how to regain trust and attract more people to Disneyland theme parks!

About the Data

In order to gain a better understanding of how theme park goers feel about Disneyland, I used a dataset containing several client reviews and ratings. This dataset consists of 10,000 documents (randomly chosen from a set of over 40K) that detail the following columns:

1. Review
2. Rating
3. Date
4. Park
5. Review ID

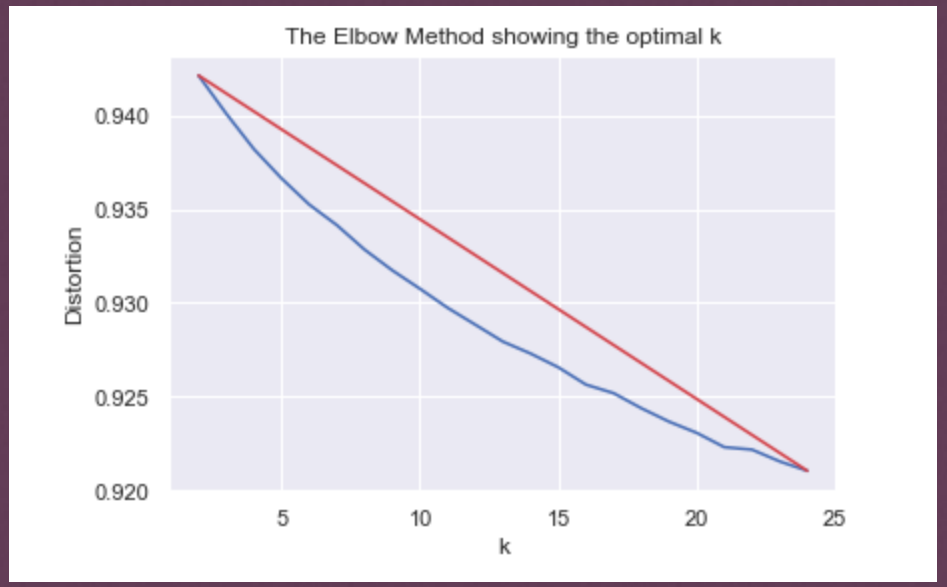
Topic Modeling and Sentiment Analysis Data Preparation

1. Clean data by removing stop words, numbers, punctuation etc.
2. K-Means modeling using PCA decomposition after vectorization of data.
3. LSA Topic Modeling using Single Value Decomposition and sentiment analysis
4. Streamlit Implementation

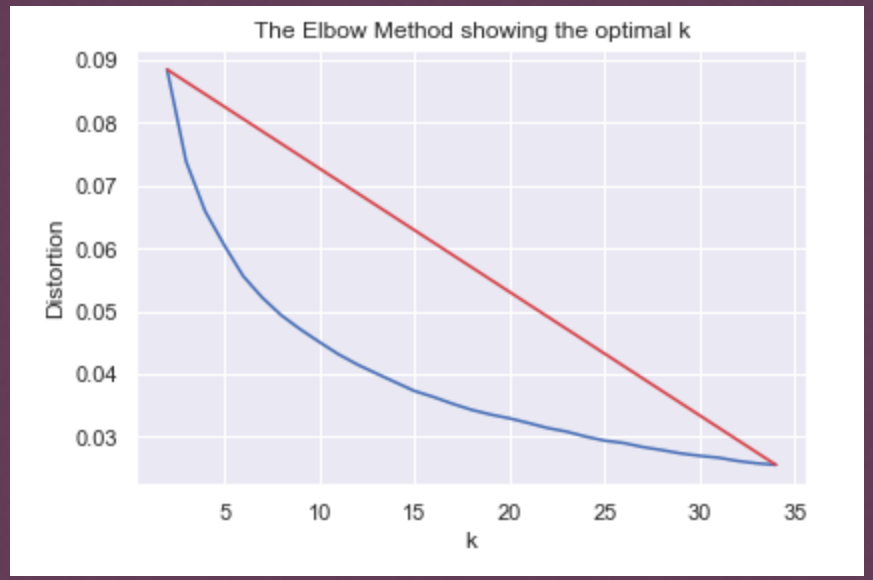
K-Means Modeling

Choose appropriate K Using the "Elbow Method" and PCA Decomposition

Components = .95



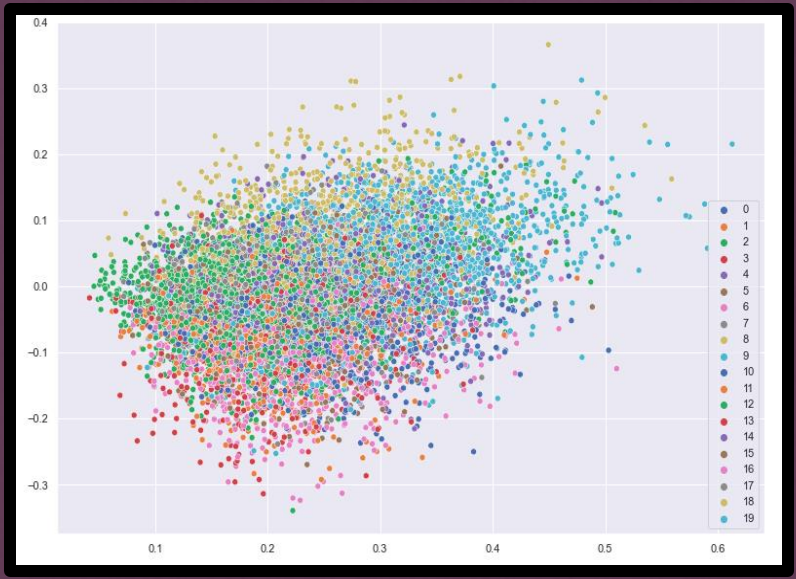
Components =2



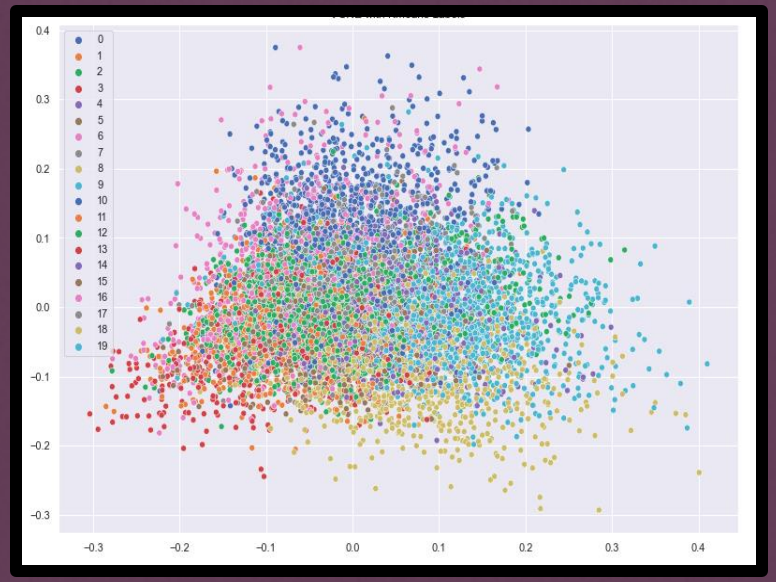
Choose K= 20 and Components = .95.

K-Means Clustering Visualizations using SVD and PCA Decomposition.

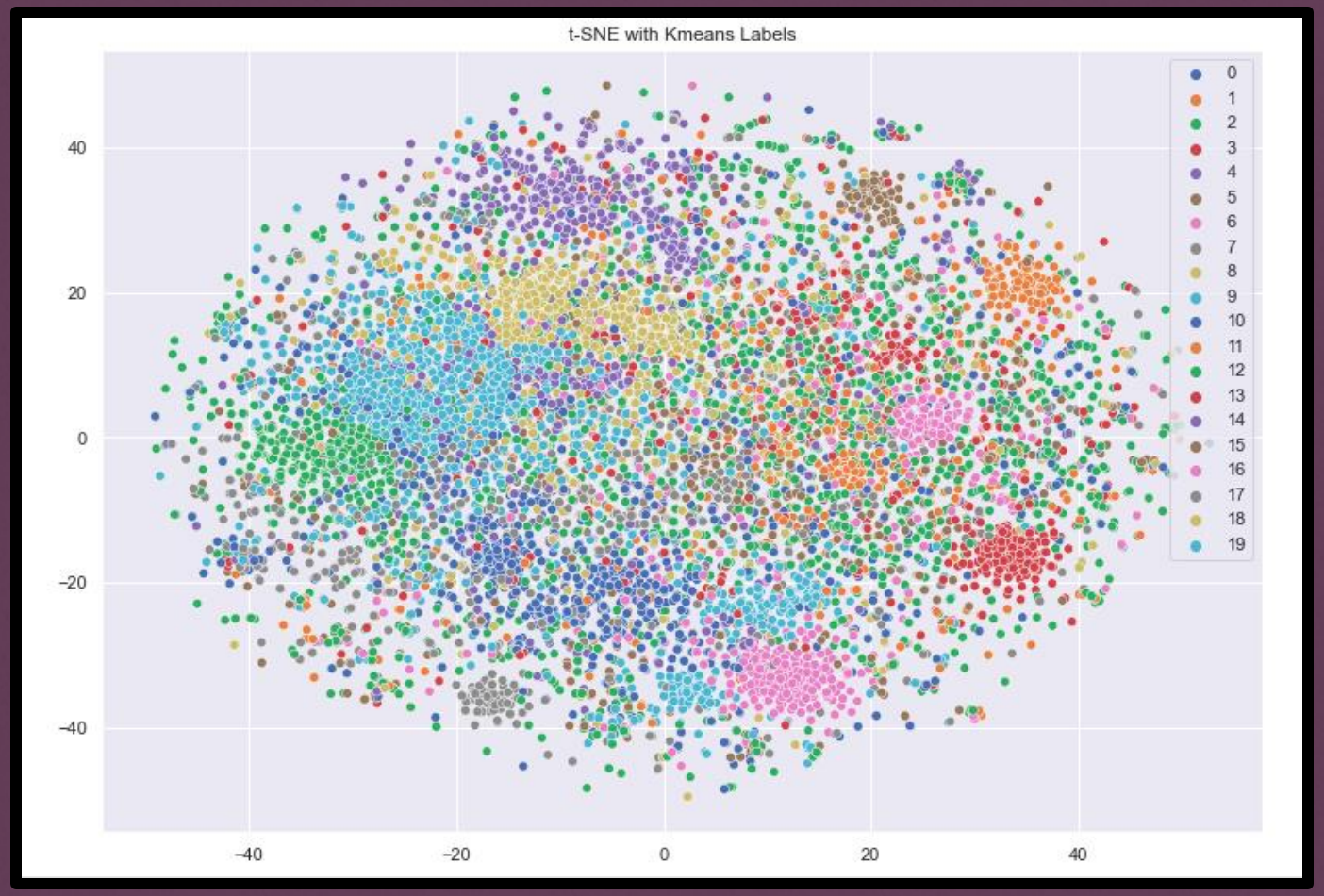
SVD Decomposition Model



PCA Decomposition Model



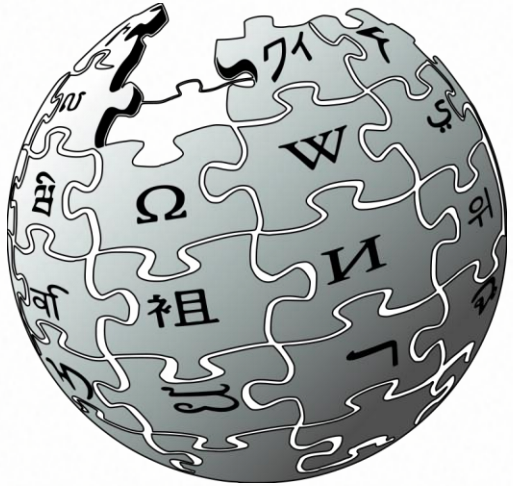
TSNE K-Means Clustering



TSNE may be used to understand the volume of various topics in our corpus.

References

kaggle.com




See Streamlit for
Topic Modeling
Demo and More
Sentiment analysis.