

# Disneyland NLP Write-Up

Elisabeth Johnson

## Abstract

The purpose of this analysis is to build a model that most accurately helps stakeholders understand the areas in which theme park goers may be enjoying or not enjoying Disneyland for the purpose of attracting more people to Disneyland Parks. In this analysis I performed Sentiment analysis, Topic Modeling, K-Means clustering and Logistic regression.

The dataset used has over 43K rows.

## Design

For this project I used several python libraries (listed below) along with PowerPoint in order to visualize my data and construct various models.

In order to account for runtimes with a large amount of data, I sampled 10K documents from the original 43K documents.

### Libraries Used

- Pandas
- Numpy
- SkLearn
- Seaborn
- Matplotlib
- ImbLearn
- NLTK
- Plotly
- SK-Learn

## Data

The data I used for this project was obtained from Kaggle and can be found [here](#).

## Algorithms

For this project I used Exploratory Data analysis in addition to several vectoring techniques in order to better understand the sentiments in my corpus.

## Communication

For this project I chose to communicate my findings via python visualizations, Streamlit and PowerPoint.