Elizabeth Earl

DSC680-T301

Milestone 1 – Proposal and Data Selection

**Topic:**

The project I wanted to work on for this first project is an expansion on a project I started on in another course. The project is focused on Disney Reviews and how reviews impact their company in a variety of branches throughout.

**Business Problem:**

Reviews are a beneficial way for a company like Disney to try and predict how future guests will feel in similar situations. Analyzing review data will help Disney grow different branches in their business such as theme parks, standalone, online sales, etc. In recent years Disney began shifting to an online only sale as well as made changes in ticketing schemes throughout parks.

The aim of my project is to discover which branches’ changes have made the most positive impact to Disney’s company. Some potential questions I will consider:

* Ticketing reservations have lessened the number of guests at the park but is this causing an increase in annual passes/expensive tickets?
* How has the closure of physical stores impacted online sales?
* Which Park brings the most positive experience according to guests?
* What is different about the reigning park to cause such positive reactions?

**Datasets:**

The dataset I’m using can be found here: [Disneyland Reviews | Kaggle](https://www.kaggle.com/arushchillar/disneyland-reviews?msclkid=1c5b9354a7c711ec968f3c333c2e4ac5) The dataset includes reviews of three Disney parks, California, Paris, and Hong Kong. These three parks will help determine which park is more successful (in terms of guest experience). Insight on other potential datasets will be explained under Challenges/Issues.

\*\*I may potentially use film review dataset as well. This dataset can be found: [Disney+ Shows and Movies - Exploratory Analysis | Kaggle](https://www.kaggle.com/code/shivamb/disney-shows-and-movies-exploratory-analysis)\*\*

The reason for this dataset is I will use the reviews to determine guest reaction to Disney and reflect it by the year of the film.

**Methods**: - What analysis methods will you use to complete this project? Note: this is just a proposal, your project can adapt as you work on it

I want to run some predictive models on my data. Some models I am considering are:

* Random tree classifier
  + I’m predicting this classifier might end up becoming a decision tree classifier, both are being considered
* Logistic Regression

I will rely heavily on sentiment graphs and matrixes to have a better (visual) understanding on how views are split. These visual graphs will help guide my models and help determine which will be most beneficial based on data trends (how negative and positive reviews are spread out).

**Ethical Considerations:**

Working with a company like Disney it becomes a touchy topic getting information. I must make sure the privacy of Disney and the consumers alike are protected. This becomes a thin line to work with as I will have to ensure I do not manipulate data in a way that will benefit one party over the other.

**Challenges/Issues**:

As already stated prior, the datasets might be a potential problem as Disney is very private about their data. I am considering simply using park reviews and potentially film reviews to see how people react to films by the years.

**References:**

A lot of the datasets are from Kaggle which will be used heavily to discover more datasets. As mentioned, Disney is very private with their information and although I will use as many recourses provided by Disney’s site to analyze (any available) data.