Group 7: Salita Epifania, Shayna Chernak, Hima Gharat

UPENN-PHI-DATA-PT-09-2020-U-C

Project 2 Proposal

Project Title: Netflix Data Analysis

Project Description/Outline:

Netflix is arguably one of the most popular online streaming websites around. Since its creation in 1997, it is safe to say that the platform has changed greatly- what started off as a DVD rental service has transformed into a streaming platform used by millions. This growth led us to wonder about how exactly Netflix has evolved since its creation.

We have focused on two aspects of Netflix’s transformation:

1. Is Netflix increasingly focusing on shows rather than movies in recent years?
2. Which of these do consumers prefer: shows or movies?

We will be extracting this data using Pandas and uploading it to PostGreSQL.

Datasets to be used

<https://www.kaggle.com/shivamb/netflix-shows?select=netflix_titles.csv>

<https://www.kaggle.com/vikassingh1996/netflix-movies-and-shows-plotly-recommender-sys>

<https://www.kaggle.com/vikassingh1996/netflix-movies-and-shows-plotly-recommender-sys/comments>

<https://www.kaggle.com/aalhendi/netflix-movies-and-tv-shows-ratings>

Breakdown of Tasks:

1. Preprocessing data

2. Extract

3. Transform

4. Load