**Proposal By Group One (2 options)**

Team Members:- Daixuan Chen, Vyas Pranali, Wentao Gao Amy, Nimmikrishna Babu

**Option1:**

**Exploring the Preference of Netflix Movie Viewers through Web Scraping**

**Background and Problem Statement:**

Netflix is a leading global streaming platform that provides a wide range of movies and TV shows on its platform. As a streaming platform, Netflix collects a vast amount of user viewing history data, which contains users' preferences for various movie and TV show genres. By analyzing and mining these data, we can obtain valuable insights into audience preferences.

The research problem in this proposal is: If viewers like a certain type of movie, what other types of movies are they more likely to enjoy?

**Methodology and Procedure:**

To answer this question, we will use web scraping techniques to obtain user viewing history data from the Netflix platform. In this data, we will identify users who have rated a certain type of movie (e.g., horror) and collect which other types of movies these users have also rated.

**The specific steps are as follows:**

Identify the target website for web scraping: We will choose the Netflix platform as our target website and use web scraping techniques to obtain user viewing history data.

Build the dataset: We will write a web scraping program using the Python programming language, which will call the Netflix platform's API to obtain user viewing history data. We will then organize this data into a dataset.

Data preprocessing: We will preprocess the dataset, including deduplication, handling missing values, removing outliers, and so on.

Data analysis: We will use data analysis techniques, including clustering, classification, association analysis, etc., to discover hidden relationships in the user viewing history data.

Results presentation and interpretation: We will use visualization tools to present the results and provide interpretations and analysis. We will identify some interesting conclusions from the results and further explore the reasons and effects of these conclusions.

**Expected Results:**

This study will answer the question of which other types of movies viewers are more likely to enjoy when they like a certain type of movie. We will obtain some interesting conclusions and further explore the reasons and effects of these conclusions. These conclusions will provide important insights for optimizing Netflix's movie recommendation system and personalized recommendation for users.

Database reference:

<https://www.geeksforgeeks.org/scrape-imdb-movie-rating-and-details-using-python/>

<https://developer.fandango.com/rotten_tomatoes>

Option2:

**The most popular hotel style in New York**

——by analyzing data from TripAdvisor.com

With the increasing popularity of online travel platforms, TripAdvisor has become an essential tool for travelers to research and compare hotels, read reviews and ratings, and make informed decisions about their accommodations.

Our plan involves scraping data from TripAdvisor.com to collect information about hotels in New York, including their ratings, reviews, amenities, prices, and other relevant details. **By analyzing this data, we can determine which hotel style is the most popular among travelers and what factors contribute to its popularity.** This information will be useful for tourists, travel agencies, and hotels in New York, who can use it to improve their services and attract more customers.

Our research will involve the following steps:

1. Collecting Data: We will use web scraping tools to collect data from TripAdvisor.com about hotels in New York. We will gather information such as hotel name, rating, review, and other relevant details.
2. Data Cleaning: The data collected will be processed and cleaned to remove any irrelevant or duplicated data, ensuring accuracy and reliability.
3. Data Analysis: The cleaned data will be analyzed to identify trends, patterns, and insights that will help us determine the most popular hotel in New York. We will use statistical methods to compare the ratings and reviews of different hotels, and to identify the most frequently mentioned amenities and services.
4. Report and Recommendations: We will prepare a detailed report that will summarize our findings and provide recommendations for hotels in New York. The report will also include visual representations of the data and insights to make it easier for our clients to understand.

In conclusion, our plan is to analyze data from TripAdvisor.com to identify the most popular hotel in New York. We believe that this business proposal has the potential to provide valuable insights into the travel and hospitality industry in New York and help hotels attract more customers.

Database source:

TripAdvisor based on HW2.