**Detailed Documentation Report for Netflix Watching Behavior Project**

**Github account:**

<https://github.com/Matrasulov/BDA_Final>

**Team members:** Akbarkon Matrasulov 12204550

Landraud Robin 12235243

Lavigne Nicolas 12235239

Muminjon  Mukhammadov 12214678

Datasets used for project:

<https://www.kaggle.com/datasets/arnavsmayan/netflix-userbase-dataset>

<https://www.kaggle.com/datasets/shivamb/netflix-shows>

**Introduction**

This project aims to analyze Netflix watching behavior using two main datasets: one obtained from Kaggle and another from a survey conducted among 20 participants. The objective is to understand the viewing habits, preferences, and patterns of Netflix users, using big data analysis techniques.

**Project Details**

1. **Presentation Slides:** Visual representations of the analysis and findings
2. **Code implementation:** Jupyter notebook (BDA\_SurveyAnalysis.ipynb) containing the code for data analysis and visualization.
3. **Detailed Report:** Documentation of the Project

**Dataset**

Primary Dataset(Kaggle)

**Origin**: Kaggle

This dataset includes comprehensive data on Netflix viewing behavior, covering various aspects such as movie genres, watching frequency, preferred times, and device usage.

**Link:** Kaggle Dataset

**Secondary Dataset(Survey)**

**Origin:** Conducted survey among 20 people

This dataset consists of responses from a survey designed to capture detailed insights into individuals' Netflix watching habits. The data includes attributes such as gender, frequency of watching, preferred time of day, and more.

**Methodology**

Data collection

* Kaggle Dataset: Downloaded directly from Kaggle
* Survey Dataset: Collected responses using a structured survey form. Initial data was partially generated using OpenAI’s ChatGPT, whereas remaining data(code) manually entered for the last 10 participants.

**Data structure:**

The survey data is structured as a list of dictionaries , where each dictionary represents a single participant’s response.

Exmaple code:

data = [

{'Name': 'Remy', 'Gender': 'Male', 'Frequency': 'Every day or two days', 'Time\_of\_Day': 'Evening',

'Movies\_per\_Week': 4, 'Preference': 'With wife', 'Genres': 'Action, Comedy, Drama, Horror, Science Fiction, Romance, Documentary, Other',

'Favorite\_Movies': 'The Shawshank Redemption', 'Selection\_Method': 'Watchlist',

'Favorite\_Actors': 'Robin Williams, Morgan Freeman', 'Importance\_of\_New\_Releases': 'Very important',

'Rewatch\_Frequency': 'Often', 'Participation\_in\_Discussions': 'Yes', 'Device': 'TV app',

'Satisfaction': 'Somewhat dissatisfied', 'Recommendation\_Likelihood': 'Likely'},

...

]

**Attributes/Insights**

* **Name:** Respondent’s name
* **Gender:** We can analyze the differences in movie-watching frequency, preferred genres, favorite actors, selection methods, satisfaction levels, and recommendation likelihood between males and females. This analysis could provide insights into gender-specific preferences in movie consumption.
* **Frequency:** Determine the frequency distribution of movie watching among respondents. This could help understand how often people engage with movies and whether there are any patterns based on gender or other factors.
* **Time\_of-Day:** Preferred time for watching. movies
* **Movies\_per\_Week:** Number of moview watched every week
* **Preference:** Watching preference
* **Genres:** Identify the most popular movie genres among respondents. This information can be useful for content creators, streaming platforms, and movie studios for decision-making related to content production and distribution
* **Favorite\_Movie:** Favorite movies
* **Selection\_Method:** Explore how people choose the movies they watch, whether it's through recommendations, specific actors or directors, random selection, or other methods. Understanding the selection process can help improve recommendation algorithms and marketing strategies.
* **Importance\_of\_New\_Releases:** how newly released movies are important
* **Rewatch\_Frequency:** Analyze how often people rewatch movies. Understanding rewatch behavior can provide insights into the enduring appeal of certain films and inform decisions about content acquisition and licensing.
* **Participation\_in\_Discussion:** Determine the level of engagement in movie-related discussions among respondents. This insight can be useful for building community features on streaming platforms and social media engagement strategies.
* **Device:** Preferred device for watching
* **Satisfaction:** Satisfaction with Netflix
* **Recommendation\_Likelihood:** Likelihood of recommending Netflix

Code implementation

Libraries Used

* **Pandas:** For data manipulation and analysis
* **Matplotlib:** For Creating static, intereactive and animated visualization
* **Seaborn:** For statical data visualization

**Dataset**

From this dataset, we can derive various insights and conduct different analyses related to movie-watching habits and preferences:

**Analysis and Findings**

**Viewing habits**

* Frequency of watching movies varies widely among respondents with a significantnumber watching every day or every two days
* Eveing are most preferred time for watching movies

**Preferecnes**

* Action, Comedy and Drama are the most popular genres
* A majority prefer watching movies with family or partnet

**Device Usage**

* TV apps are most commonly used devices for watching Netflix

**Satisfaction and Recommendations**

* While satisfaction lelvels vary, a majority are likely to recommend Netflix to others

**Conclusion**

This project provides valuable insights into Netflix watching bahavior, highlighting key trends and preferences among users. The combinations of big data analysis from Kaggle and detailed survey responses offers clear understandings of viewing habits, which can be useful for content recommendation and improving user experience.

**Appendicas**

* **Presentation slides**
* **Jupyter Notebook**

**References**

* [1st Kaggle dataset](https://www.kaggle.com/datasets/shivamb/netflix-shows)
* [2nd Kaggle dataset](https://www.kaggle.com/datasets/arnavsmayan/netflix-userbase-dataset)

**Future Work**

* Expand the survey to larger sample size for more robutus analysis
* Incorporate machine learning techniques for predective analysis of viewing habits

**Contribution**

Robin Landraud:

* Dataset search and analysis
* Présentation

Nicolas Lavigne:

* Survey creation
* Project proposal creation
* Survey chart creation
* Presentation

Akbaron Matrasulov:

* Survey analysis
* Documentation report
* Presentation

Muminjon  Mukhammadov:

* Presentation