Virtual Work Experience For Viewer Data Analysis And User Experience Enhancement For Stc Tv

The goal

The goal of this project is to analyze viewing behavior among stc tv users, predict future trends, and develop an intelligent recommendation system that enhances the user experience and supports strategic decision-making.

The data source

The analysis is based on detailed viewing logs collected from the stc tv platform.
The dataset includes comprehensive information such as:
  • Type of program (Movie or Series)
  • Program name and classification
  • Viewing duration
  • Streaming quality (SD or HD)
  • User interaction details
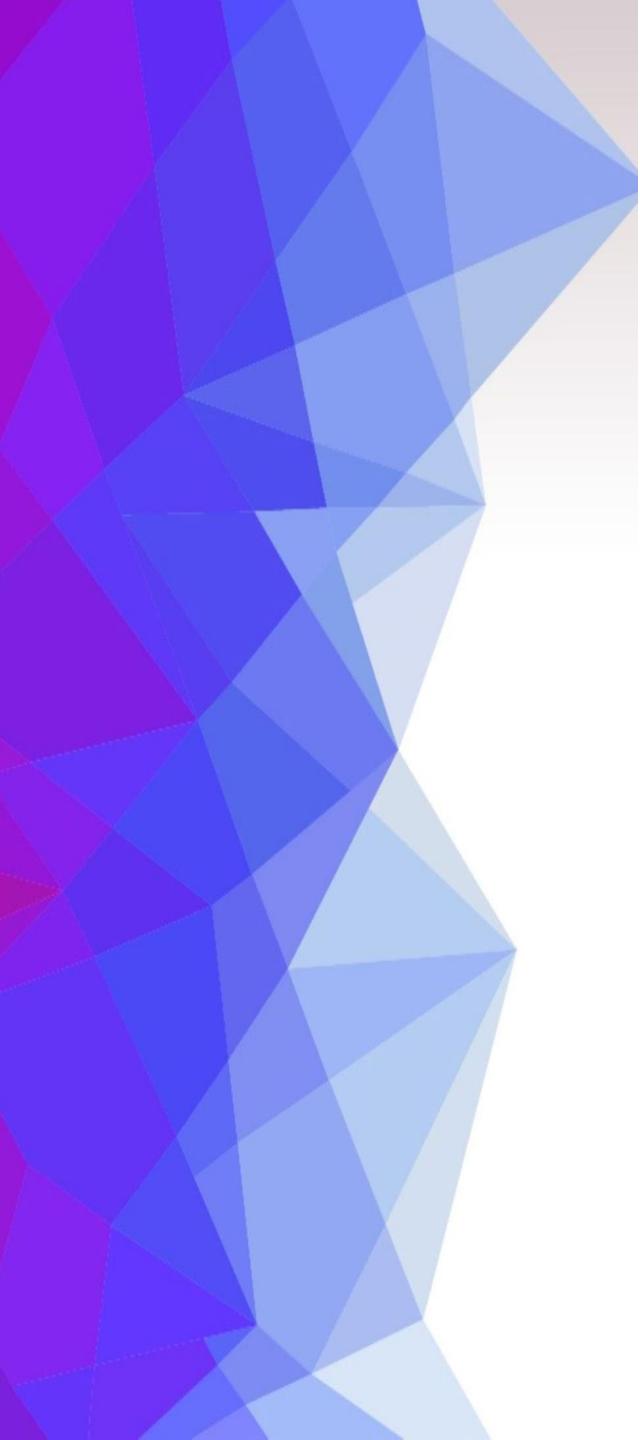
These data points provide a complete picture of user behavior and form the foundation for deeper analysis, forecasting, and recommendation modeling.

Task One

User behavior analysis and study

Introduction

The analysis began by importing and cleaning user viewing data from an Excel .xlsb file.
The primary objective was to understand user viewing behavior and identify interaction patterns with different types of content, in order to support decision-making regarding content selection and quality.
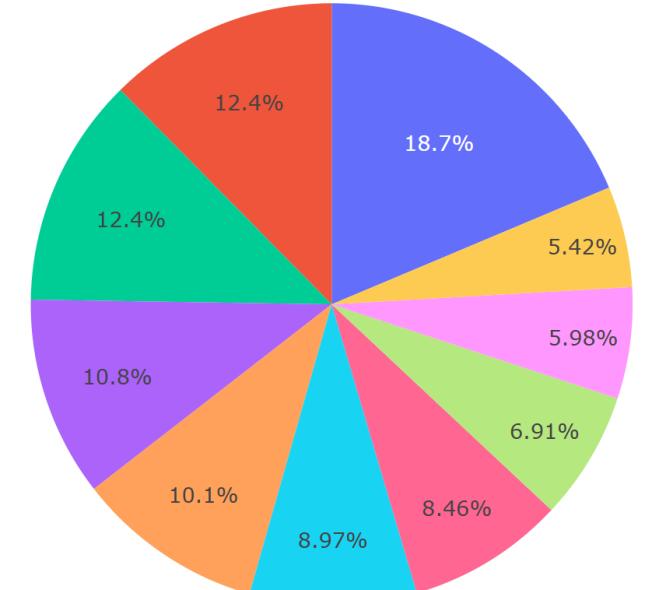
Data Preparation & Cleaning

• The dataset was loaded and columns were converted to their appropriate data types (dates, numbers, text).
• Unnecessary columns (such as indexing fields) were removed.
• Program names were standardized (e.g., removing extra spaces) to avoid duplication caused by minor differences in spelling.

Top-Performing Programs Analysis

• Season and episode details were merged into the program name (especially for series) to treat each episode as an independent entity.

• Calculated number of unique viewers

• Calculated number of views

• Calculated total watch time

10 programs based on total watch time.

A pie chart was used to visualize each program's share of the total watch time.

- The Boss Baby
- The Amazing pider-Man
- The Expendables
- Moana
- Trolls
- Bean
- The murfs
- Hotel Transylvania
- Cloudy With a Chance of Meatballs
- The Man With The Iron Fists

18.7%
12.4%
12.4%
10.8%
10.1%
8.97%
8.46%
6.91%
5.98%
5.42%
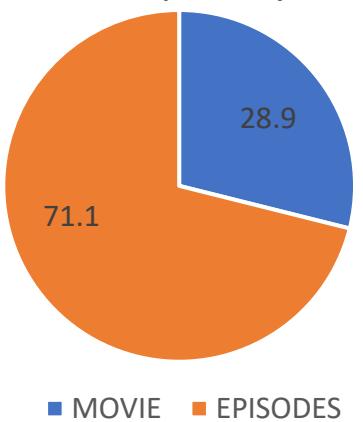
Content Type Analysis (Program Class)

- Data was grouped by content type movies and series
- For each category, we calculated:
- Number of unique viewers
- Number of views
- Total watch time
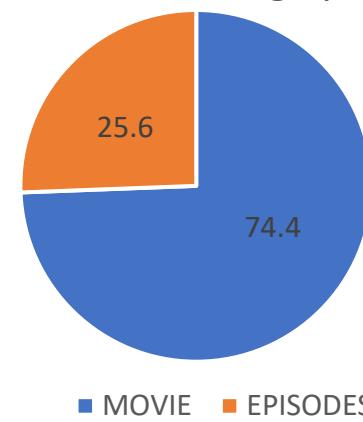
Two pie charts were created:

- One showing the distribution by total watch time
- Another showing the distribution by number of users

duration spent by class

28.9

71.1

■ MOVIE ■ EPISODES
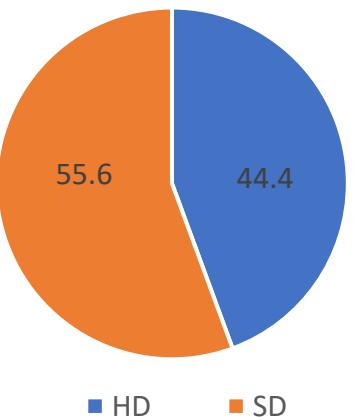
Users watching by class

25.6

74.4

■ MOVIE ■ EPISODES
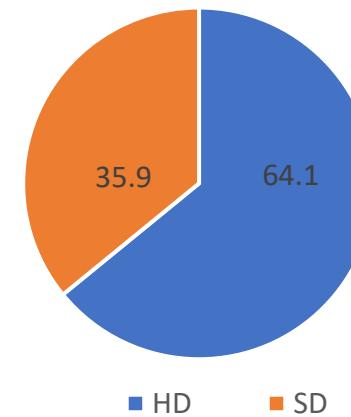
Content Quality Analysis (HD vs SD)

• Focused on the relationship between content quality (HD vs. SD) and the number of viewers.
 • This relationship was analyzed separately for movies and series.
 • Two pie charts were used side by side to compare:
 • Viewers who watched content in HD
 • Viewers who watched content in SD

SERIES/EPISODES

55.6   44.4

■ HD   ■ SD

MOVIE

35.9   64.1

■ HD   ■ SD
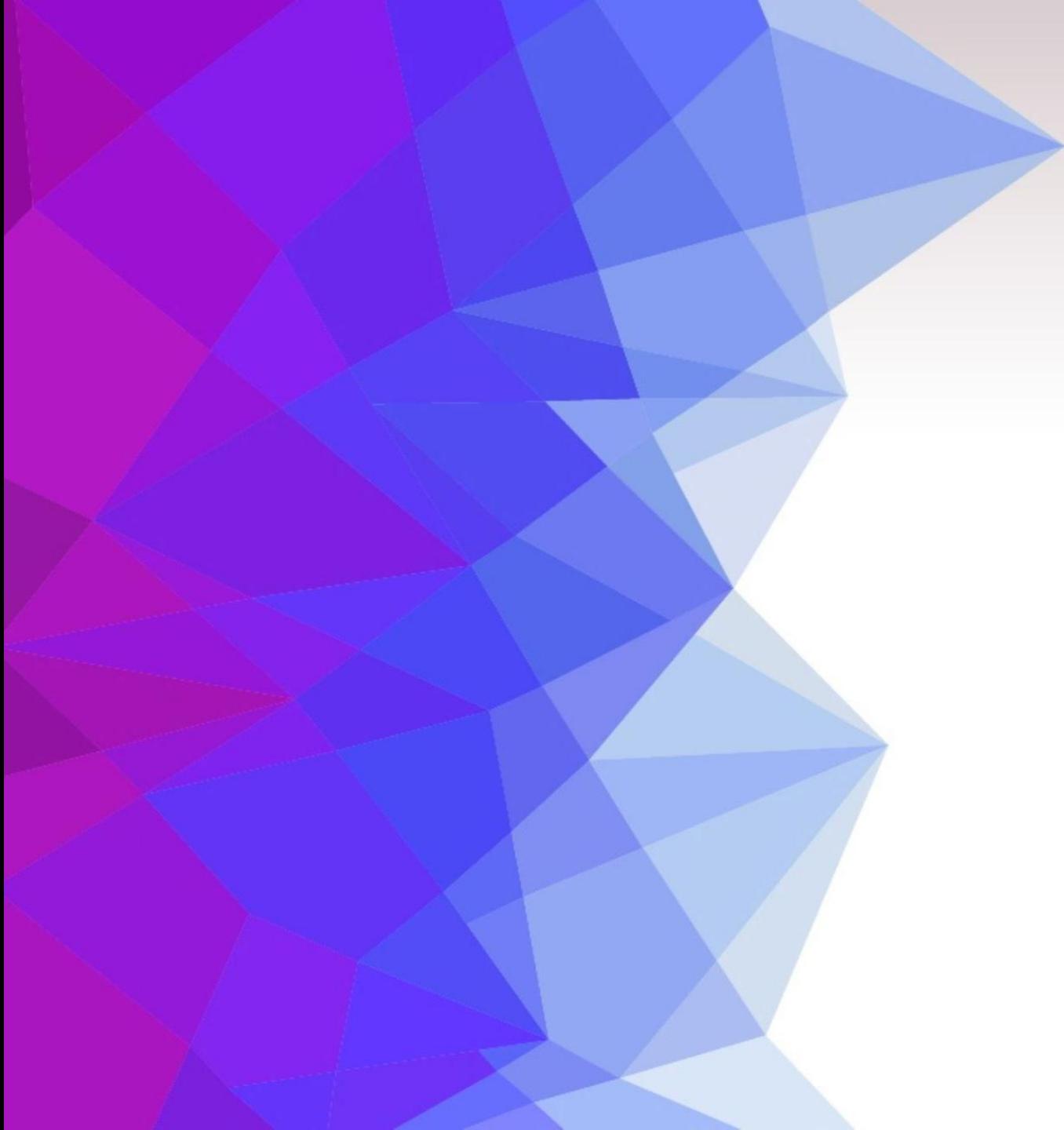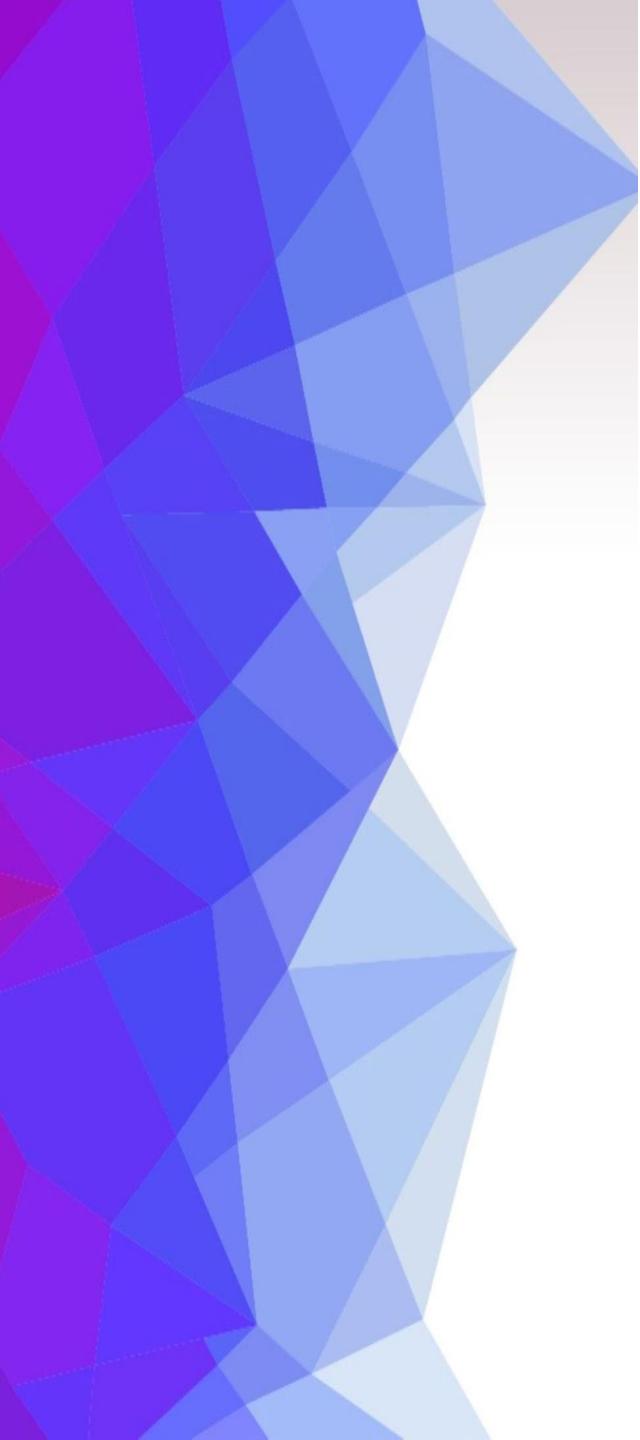
Task Two

User Behavior Prediction Model
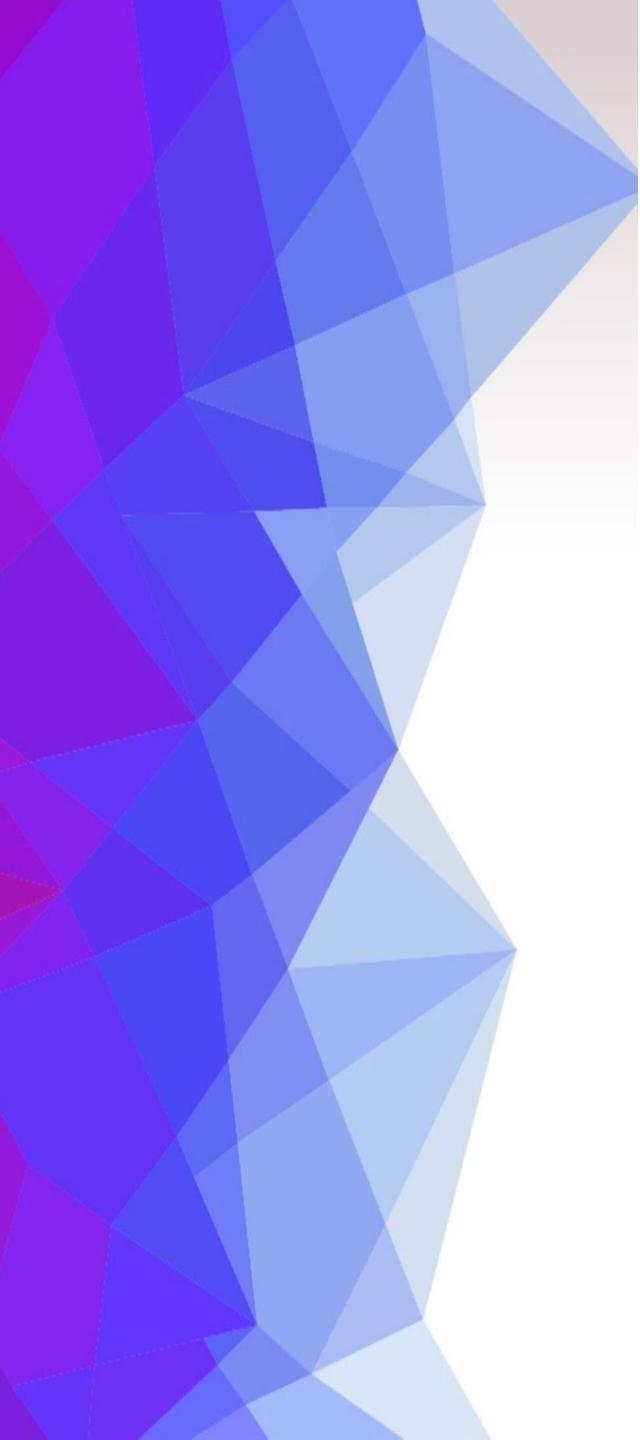
Introduction

In this analysis, we studied the daily viewing time data of STC TV users with the goal of understanding general behavior patterns and forecasting future performance. The process consisted of several stages:

Data Preparation & Cleaning

The data was loaded and examined to ensure it contained no missing or anomalous values.
We used statistical summaries and visualizations to understand the distribution and variability of daily viewing time.

Time Series Decomposition
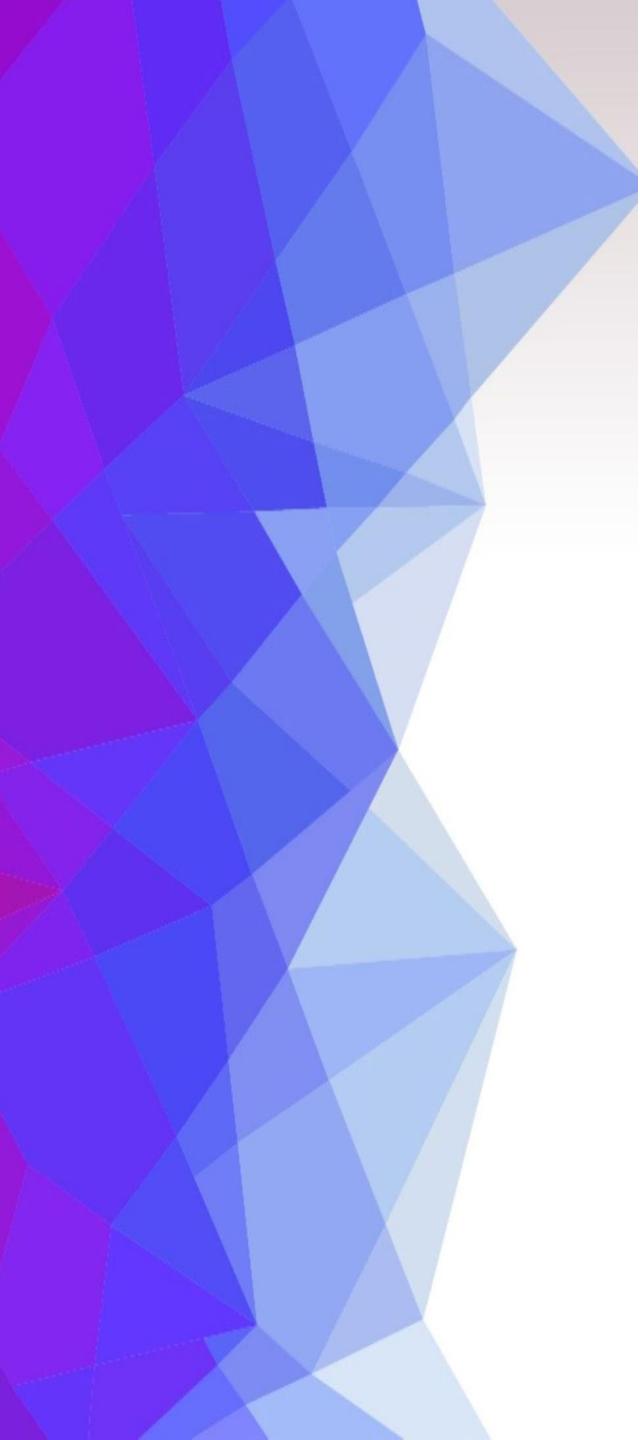
We decomposed the time series into three main components:
 • Trend: Shows whether viewing time is increasing or decreasing over time.
 • Seasonality: A recurring cyclical pattern over a specific period.
 • Noise: Irregular fluctuations not explained by trend or seasonality.

Building the Forecasting Model Using SARIMA

We tested multiple combinations of parameters for the SARIMA model (a statistical model used for forecasting seasonal time series).
The best model was selected based on the lowest AIC value (a metric for model quality).
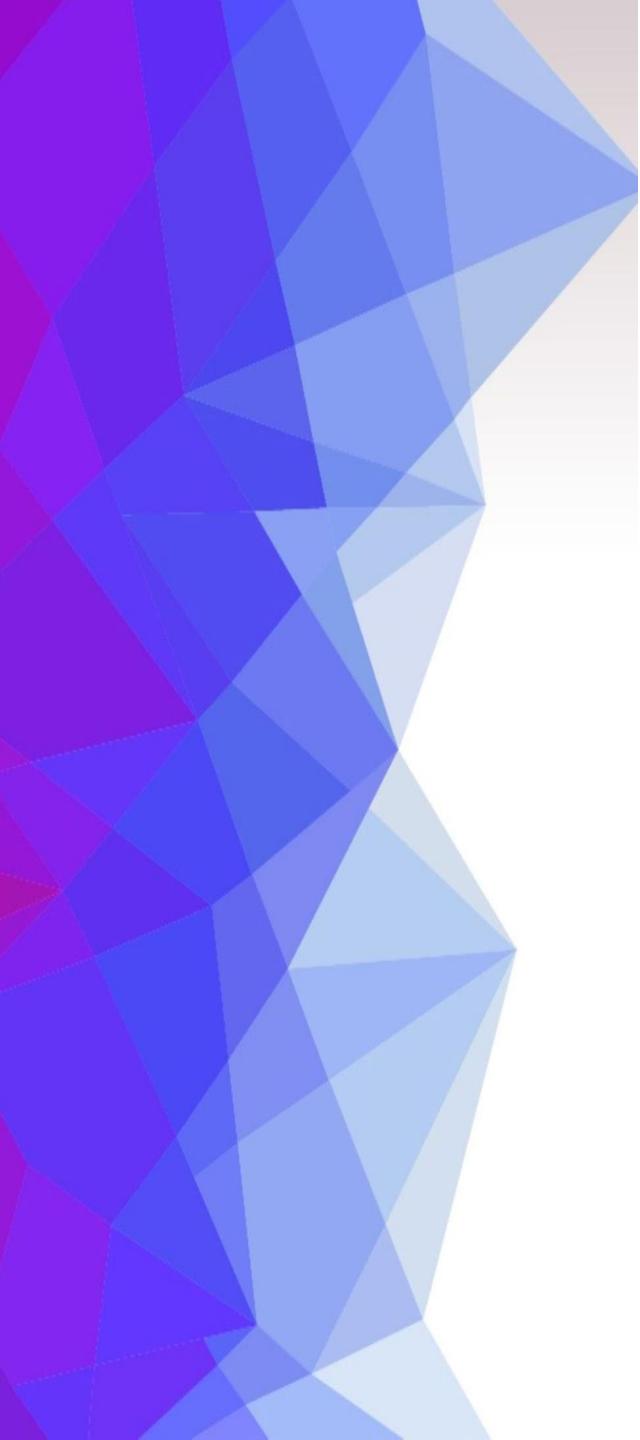
Model Diagnostics
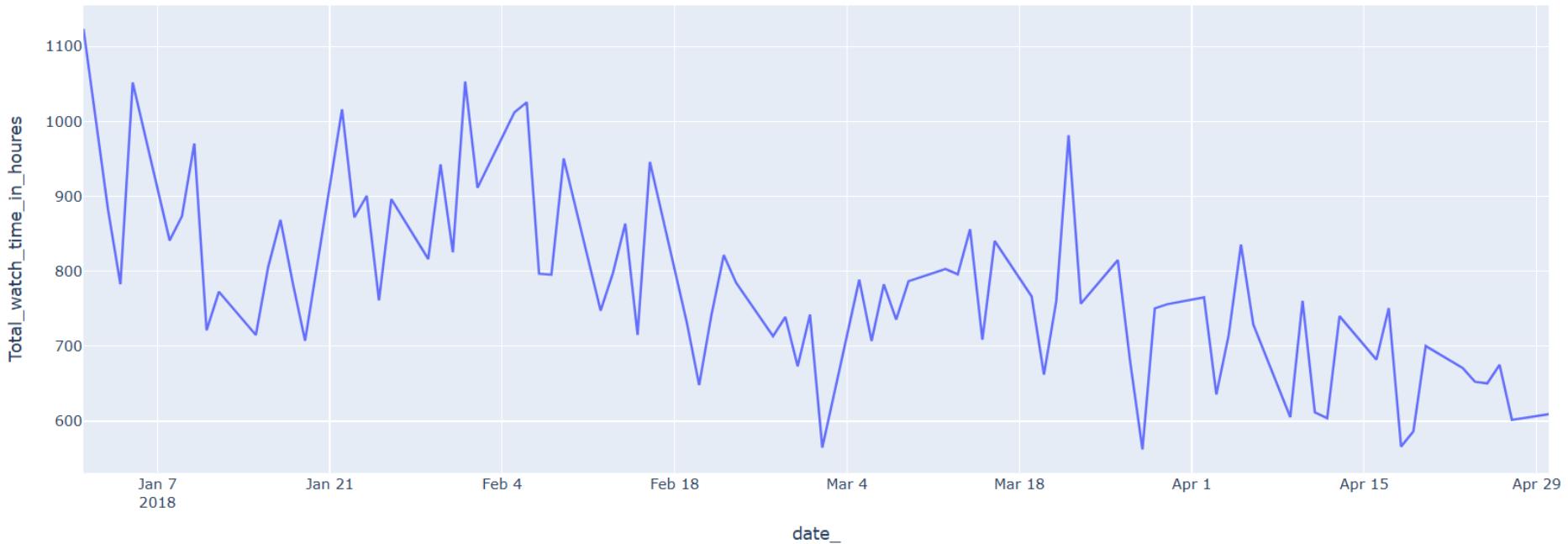
We evaluated the model's quality using diagnostic plots:
The residuals showed a normal distribution, indicating a good model fit.
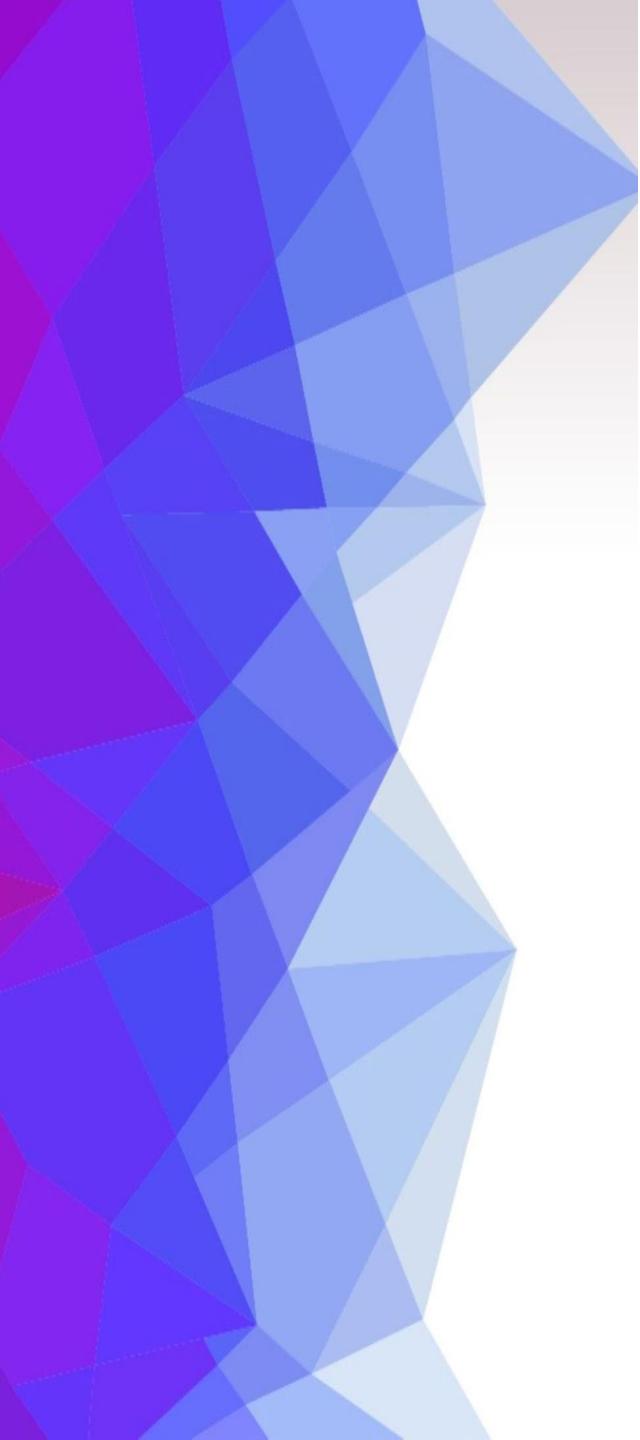
Future Forecasting

The model was then used to forecast viewing time for the next 60 days (two months).

Results
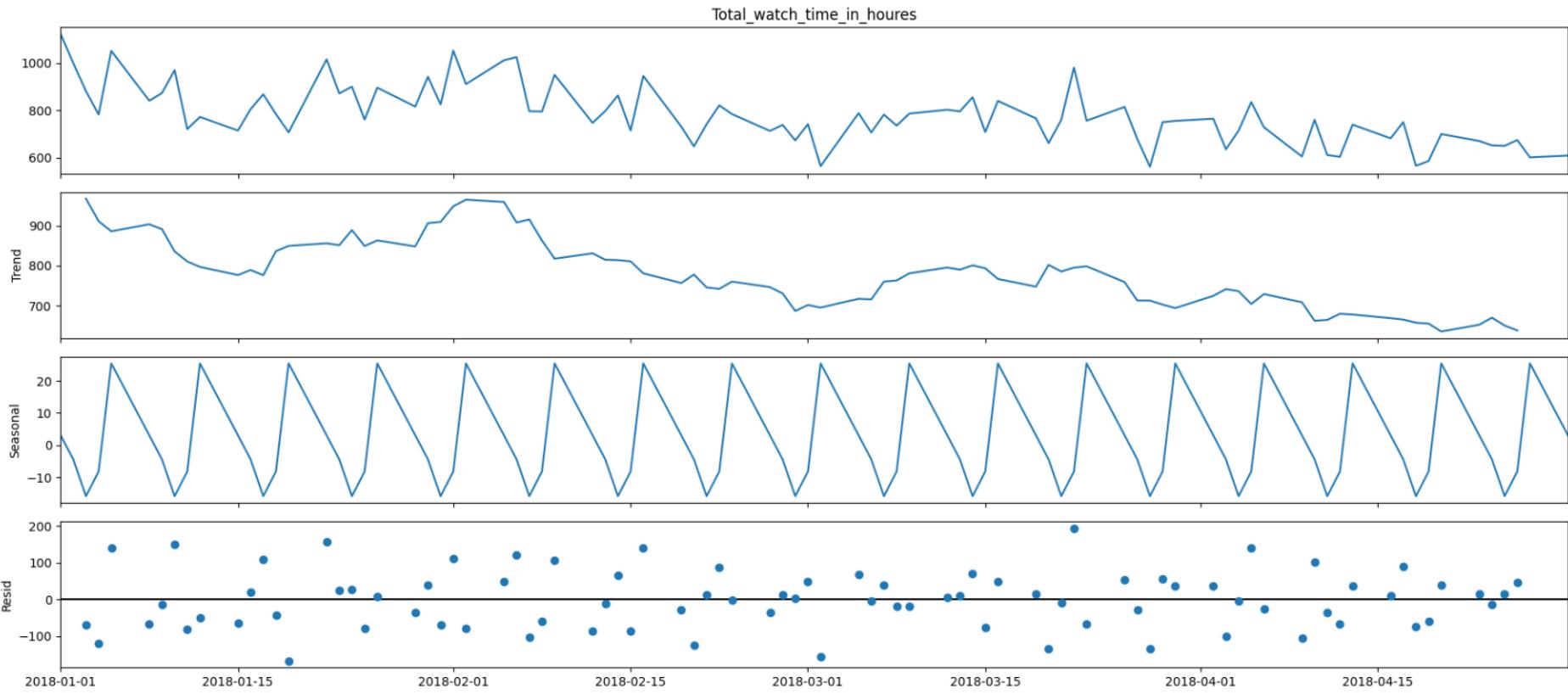
Results
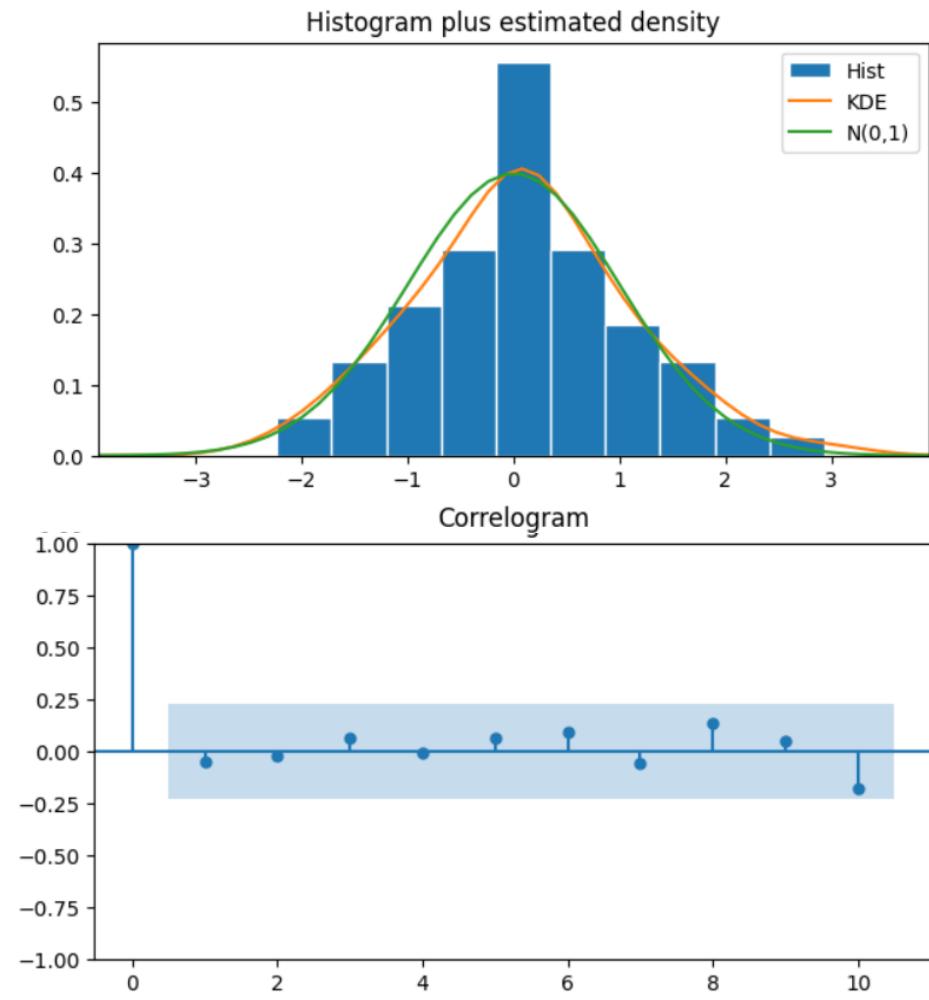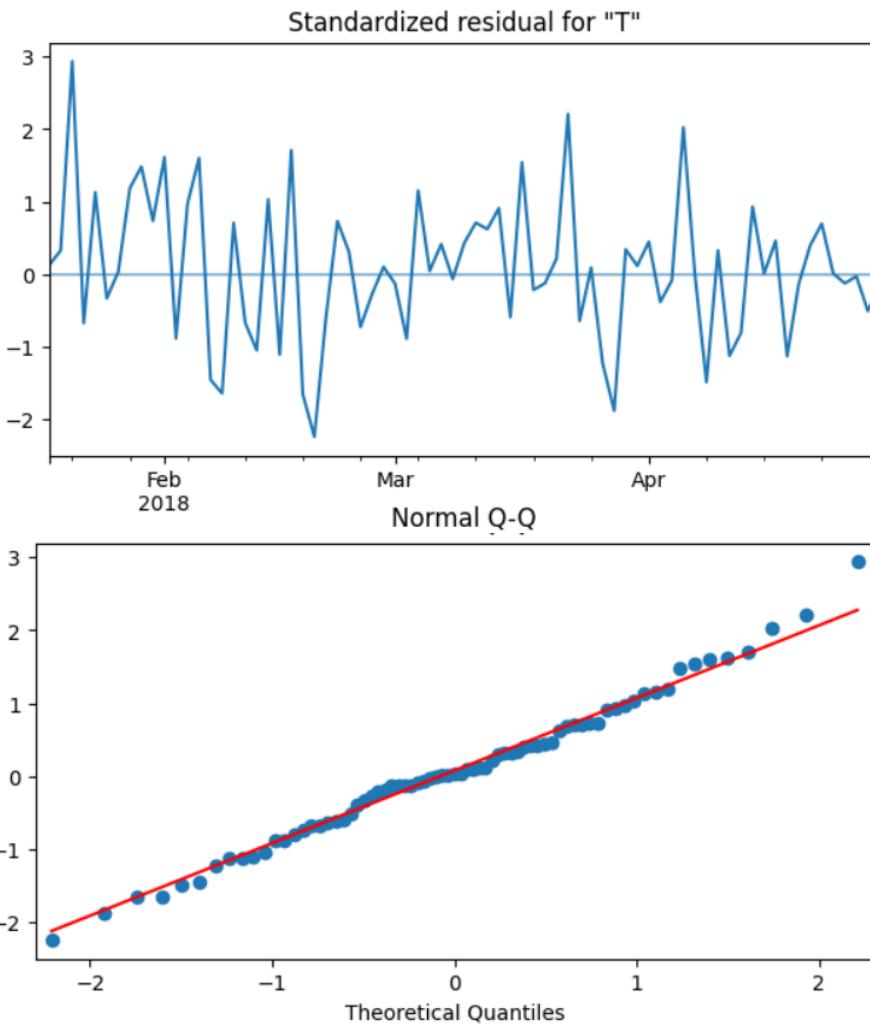
Total_watch_time_in_houres
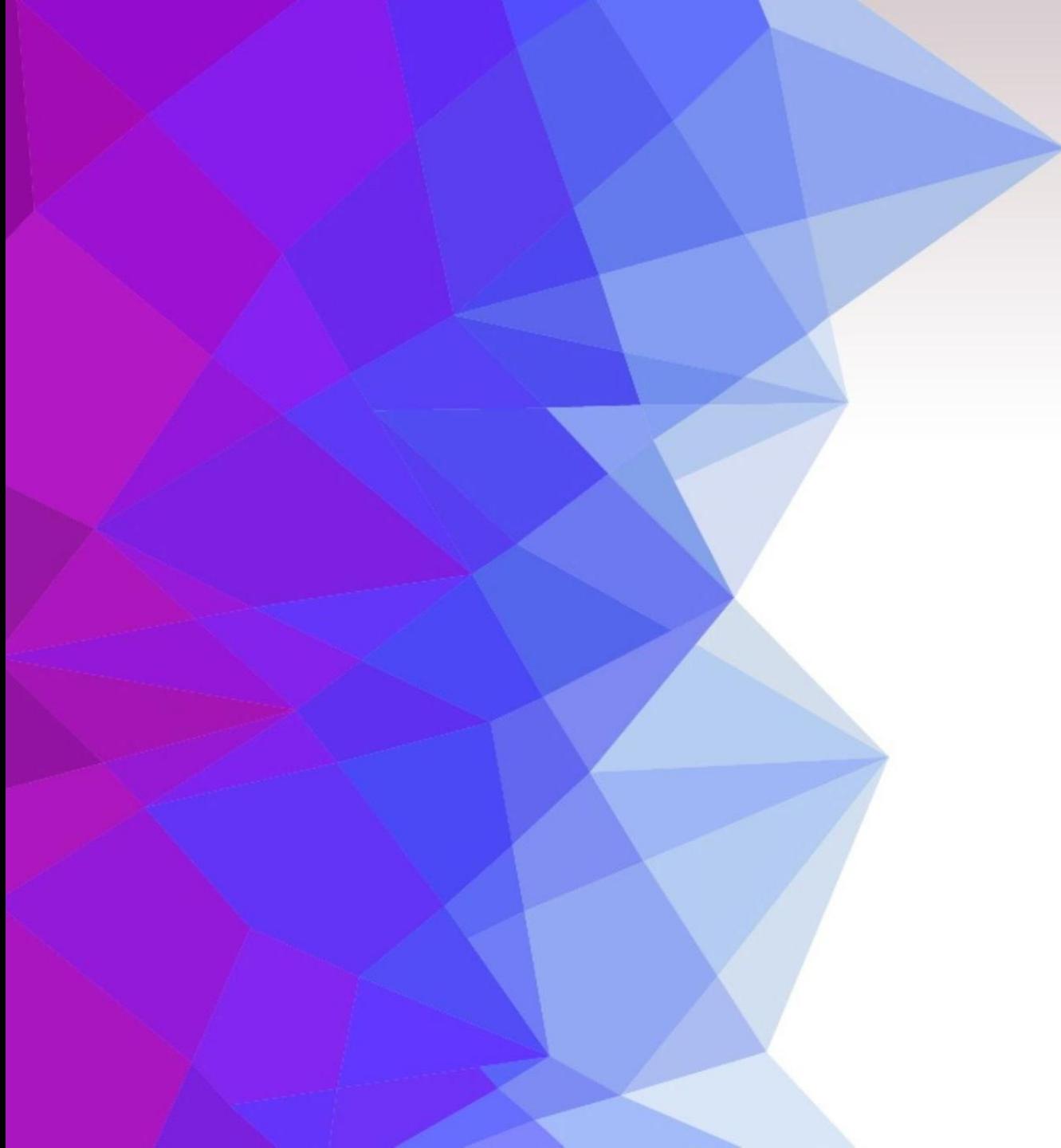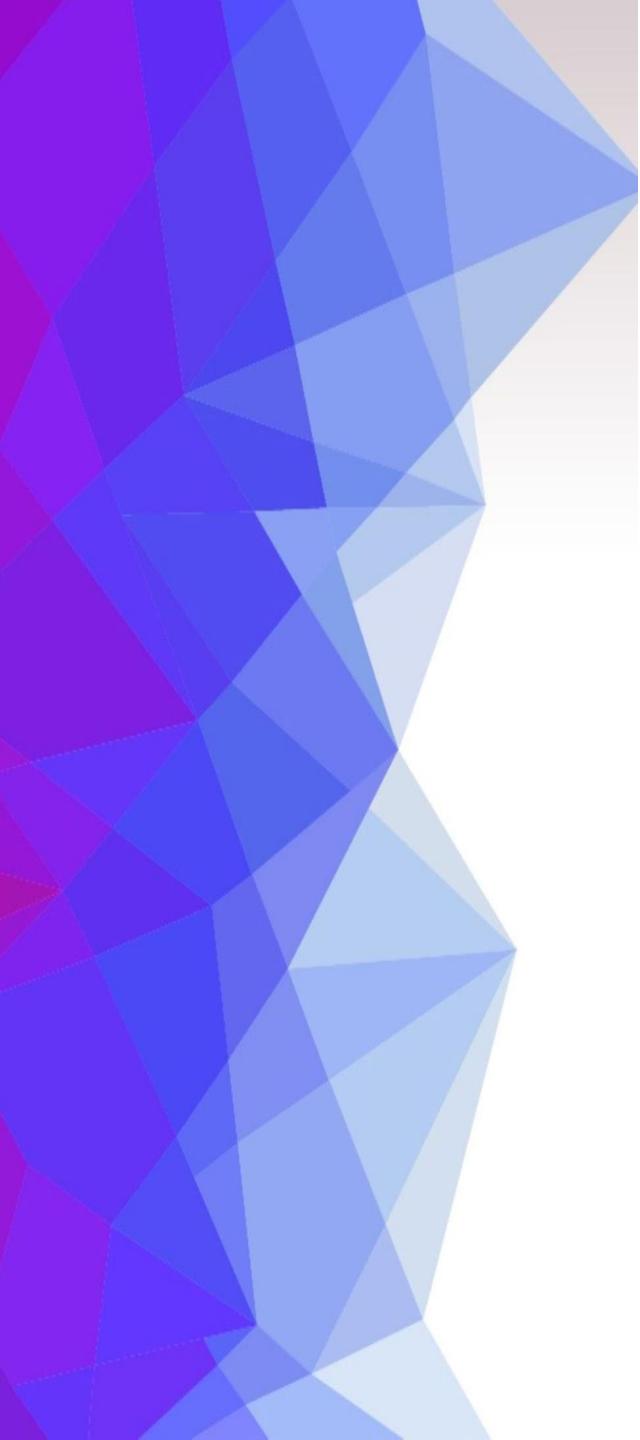
Results

Task Three

Building a Personalized Recommendation Model

Introduction

The data from the Jawwy TV platform was analyzed to understand user viewing behavior and to recommend similar content that enhances the viewing experience and increases engagement.
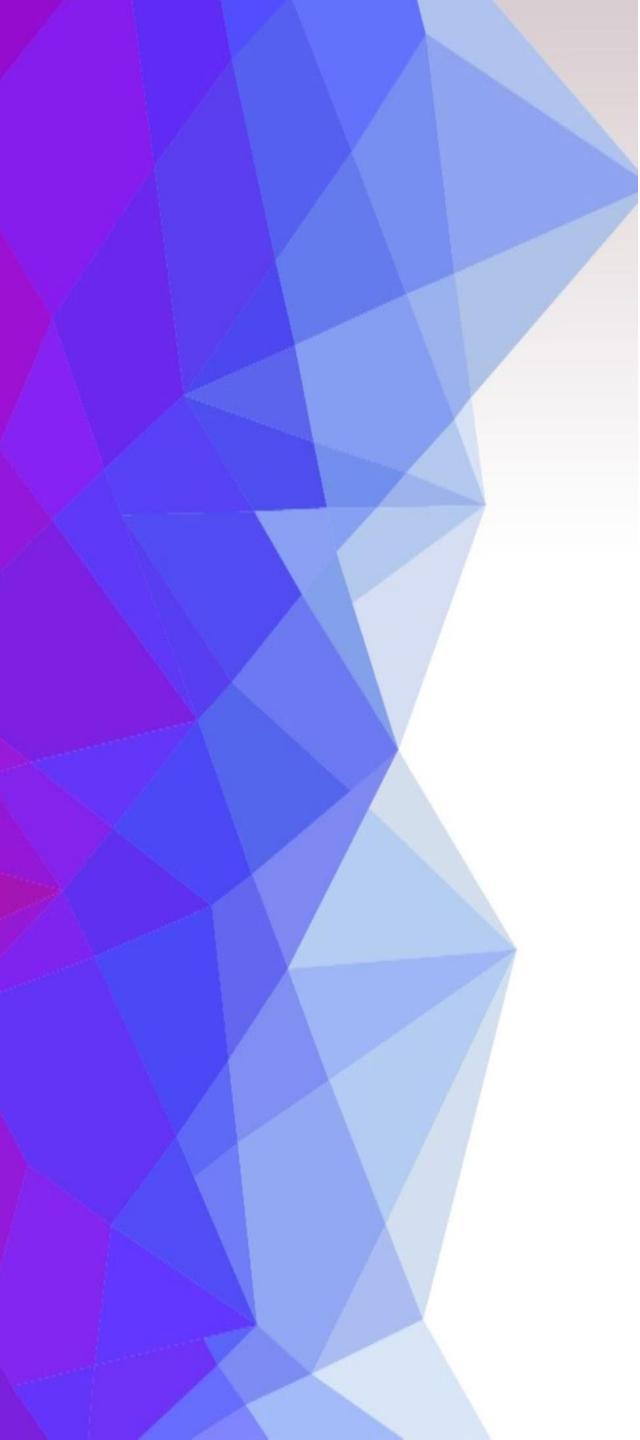
Data Preparation & Cleaning

1. Exploring the Data Structure
 • User and program data were loaded from a binary Excel file (XLSB).
 • The dataset included program names, user ratings, and user IDs.
 • Missing data was checked and handled to ensure analysis quality.

Transforming Data into a Rating Matrix

- A User-Item Matrix was created to represent interactions between users and programs.
- Each row represents a program, each column represents a user, and each cell contains the user's rating for that program.
- Missing values were filled with 0 to represent no interaction.

user_id_maped	1	5	9	11	15	17	20	26	28	30	...	34259	34261	34263	34265	34267	34269	34271	34273	34277	34280
program_name																					
#FollowFriday	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10 Days in a Madhouse	0.0	0.0	0.0	0.0	1.5	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
100 treets	0.0	0.0	0.0	1.0	2.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
101 Dalmatians	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
102 Dalmatians	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

5 rows × 11578 columns

Analyzing Similarity Between Programs

• The K-Nearest Neighbors (KNN) algorithm with Cosine Similarity was applied.

• The goal was to identify programs that are similar based on user ratings.
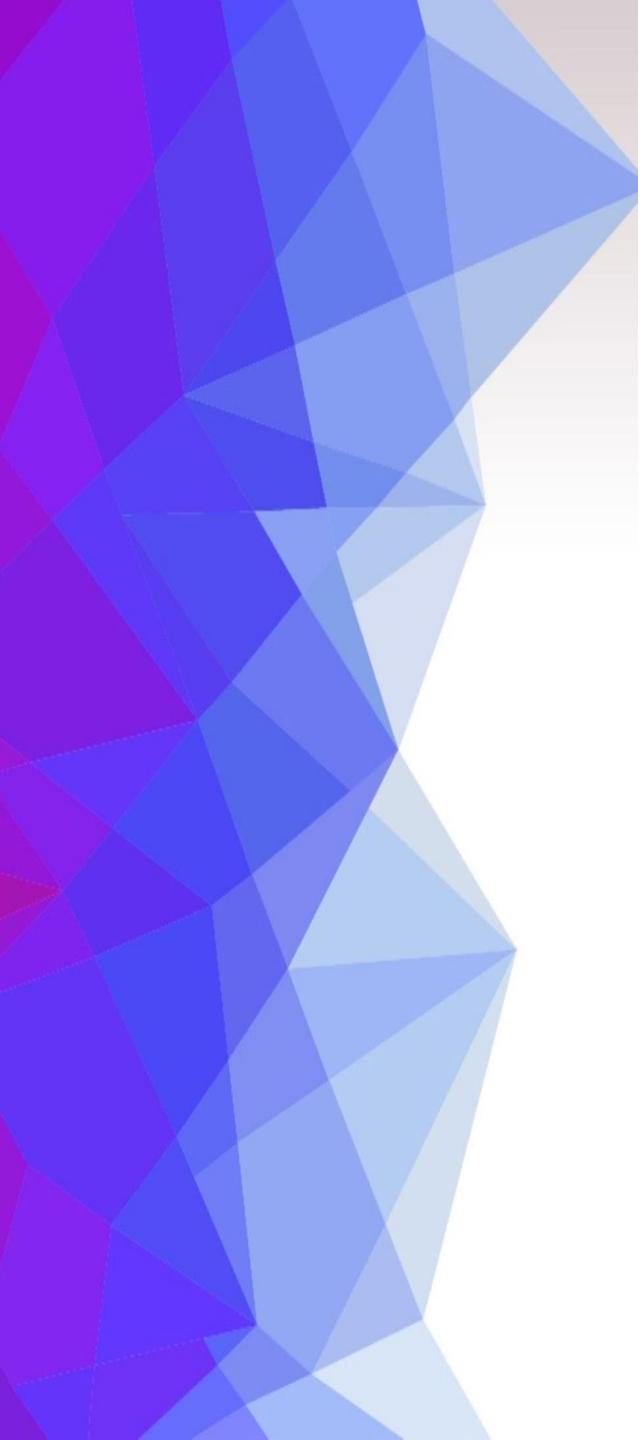
Generating Personalized Recommendations

For example, when selecting the program "Moana", the system identified the 5 most similar programs based on user interaction:

• Trolls
• Surf's Up
• The Mermaid
• The Bods Baby
• The Jetsons & WWE

```
Recommendations for Moana:

1: Trolls, with distance of 0.42764217010640215:
2: Surf's Up : WaveMania, with distance of 0.4705763355181768:
3: The Mermaid Princess, with distance of 0.5066377099343184:
4: The Boss Baby, with distance of 0.551442834662541:
5: The Jetsons & WWE: Robo-WrestleMania!, with distance of 0.5610577907608365:
```

Recommendations Based on Analysis Results
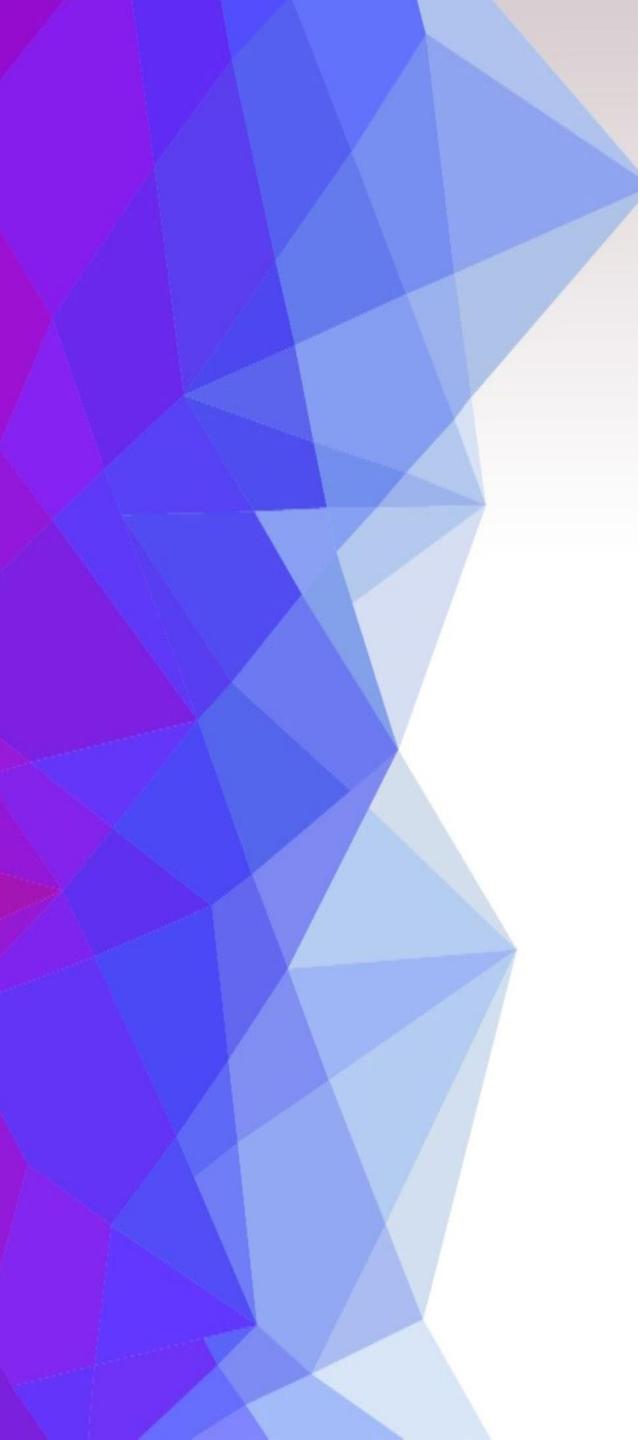
Movies vs. Series

Observation:
• Movies attract a larger number of viewers.
• Series have higher total watch time due to multiple episodes and longer duration.

Recommendations:
• Leverage the high popularity of movies to attract new users through homepage placement and promotional campaigns.
• Invest in producing or licensing more high-quality series, as longer watch time indicates higher user retention and engagement.
• Consider introducing mini-series to cater to users who prefer short watch times but enjoy longer story arcs.
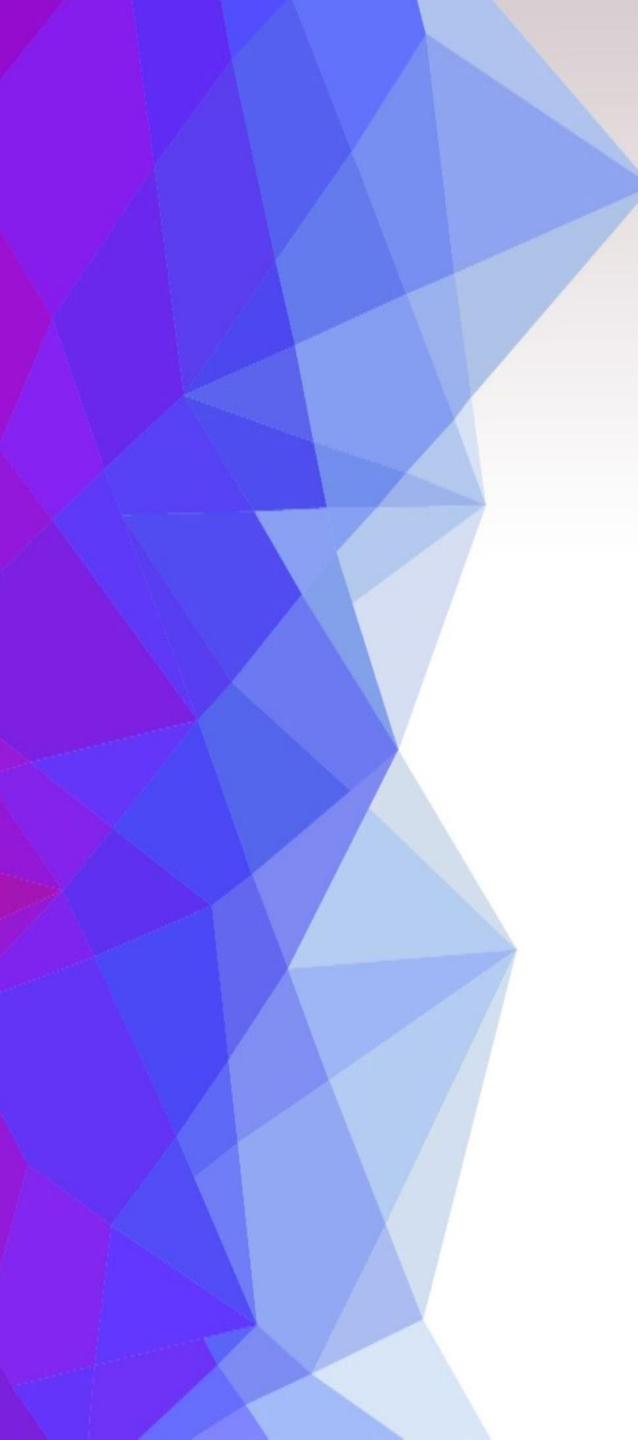
Viewing Quality (HD vs. SD)
Observation:
• HD is more popular for movies.
• SD is more common for series.
Recommendations:
• For movies, continue prioritizing high-quality HD streaming and encourage users to switch to HD when supported by their devices.
• For series, investigate why SD is preferred – whether it's due to internet limitations or device constraints.
• If technical, improve adaptive streaming performance.
• If related to data usage, offer a "medium quality" option as a compromise.
• Consider offering small incentives for users to try HD for series (e.g., first episode available in HD).
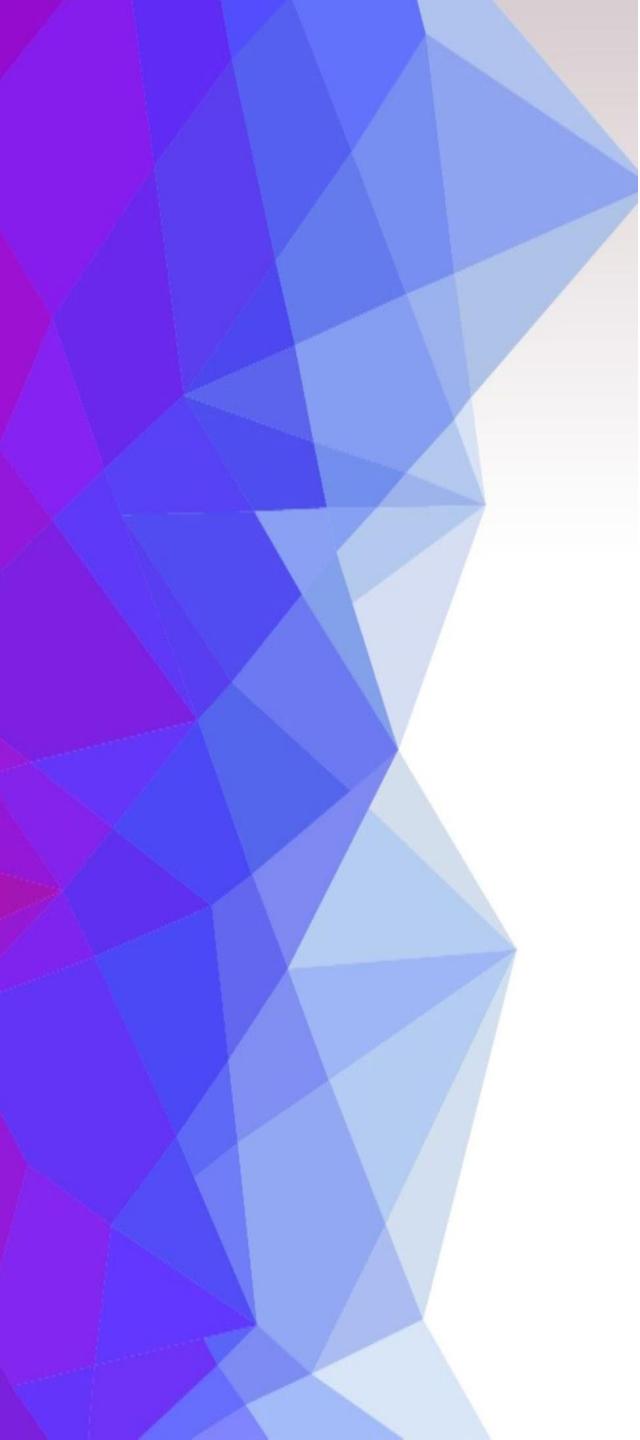
Top 10 Programs by Watch Time
Observation:
Popular programs include The Boss Baby, Moana, Trolls, The Amazing Spider-Man, and others.
Recommendations:
 • Highlight top-performing programs in marketing campaigns and homepage recommendations.
 • Analyze common characteristics of these programs (genre, audience, story style) to guide future content acquisition.
 • Create playlists or "content bundles" of similar shows to increase total watch time through automatic recommendations.

Recommendations Based on Analysis Results

Peak Viewing Period (22 March 2018)
Observation:
• A significant spike in viewing was recorded on 22 March 2018, identified by the forecasting model.
Recommendations:
• Investigate the reason for this spike (e.g., event, holiday, new content release).
• If seasonal patterns exist (e.g., higher viewership in March or during holidays), use them to schedule new content releases and marketing efforts.
• Use forecast results to allocate resources (servers, marketing, content) during expected peak periods.