Optimizing Browsing Time at NETFLIX

Harshit Singh, Isaac Lo, Tina Lie

Executive Summary

In this report, we conducted an experiment to minimize the Netflix users' browsing time. This browsing time is influenced in part by Tile.Size, Match.Score, Preview.Length and Preview.Type. We are interested in investigating which of these factors significantly influences the average browsing time, and figuring out the optimal value of browsing time. Firstly, we chose some reasonable levels for each factor and constructed a full model including the main effect and all the possible interaction effects. Then, we used a partial F-test to determine if the reduced model we chose is better than the full model, helping us certify the factors that are significant to the browsing time. Our confirmed model is "Browse.Time ~ Prev.Length + Prev.Type + Match.Score + Prev.Length:Match.Score". Using that model, we did some pairwise tests for each condition to find the optimal value of browsing time, and also tried some other values around the optimal location of significant factors to further minimize the browsing time. In conclusion, our optimal location and the optimal value of browsing time is below:

Tile.Size	Match.Score	Preview.Length	Preview.Type	MOI
0.2	82	70	Teaser/Trailer	10.52

Introduction

Problem

Netflix's users have access to a huge library of movies, tv-show, documentaries, etc. When faced with a large number of options to choose from, it becomes harder for users to make a decision, often resulting in longer browsing times.

In this project, we try to minimize the average browsing time by exploring factors that influence the browsing time and finding their optimal values.

Experimentation Methods

We have conducted a series of multivariate experiments to investigate and optimize the relationship between browsing time and Tile Size, Match Score, Preview Length, and Preview Type. We have used a factorial approach to this multi-factor experiment, allowing us to quantify both main effects and interaction effects.

Outline: What to expect in this report?

- A framework for planning and executing an investigation whose results are further analyzed to help in drawing conclusions about the question of interest.
- Defining experimental conditions by selecting design factors and their corresponding levels.
- Screening design factors using statistical techniques like ANOVA to find significant main and interaction terms.
- Visualizing Browse Time with each design factor and significant interaction effects to generate insights on the order of factors to adjust first.
- Iterative pairwise testing to find optimal values of each design factor for the lowest browsing time.
- Based on statistical analysis and business knowledge, we conclude on which design factors and their optimal levels minimize the browsing time, along with possible limitations.

Experiments

Objective

The goal of our experiments is to answer this question: What combination of *Tile.Size*, *Match.Score*, *Preview.Length*, and *Preview.Type* corresponds to the lowest average *Browse.Time*?

Plan

Our series of experiments will work towards lowering the *Browse.Time* for Netflix users. The design factors for this experiment are *Tile.Size, Match.Score, Preview.Length,* and *Preview.Type*. Each condition, composed of levels from the design factors, will consist of a sample of 100 randomly selected Netflix users. We will look at the *Browse.Time* of each Netflix user selected for our experiments. The average *Browse.Time* in each condition will be our metric of interest.

Data

The data collection process will be done in three steps. The first step is the screening experiments. This will involve collecting the most data simultaneously because at least two levels from each factor will need to be tested. The second step is to test different levels of significant factors. These experiments provide information on what direction to adjust the factor levels and a medium amount of data will be collected during the process. The smallest amount of data collected will be during the last step where we fine-tune the factor levels.

Analysis

Screening Experiments

This series of experiments begin with selecting design factor levels for our screening experiments. When selecting these levels we wanted to make sure the levels selected were not too similar and that they made sense. For example, it would not be logical to investigate *Match.Score*=0. With these considerations in mind, our initial conditions consisted of all possible combinations of the following factor levels:

Preview.Length=[30, 60, 90]

Preview.Type=[TT, AC]
Match.Score=[70, 85, 95]
Tile.Size=[0.2, 0.4]

This permutation created 36 unique conditions, each containing 100 observations. A regression model was fitted on these initial conditions to help identify significant design factors. The full model, consisting of main effect terms, 2-factor interaction terms, 3-factor interaction terms, and 4-factor interaction terms indicated that only *Preview.Length, Preview.Type, Preview.Score, Preview.Length:Preview.Type,* and *Preview.Length:Match.Score* were significant predictors of *Browse.Time*, therefore, these five terms will be used for our reduced model 1. The partial F test on the full model vs the reduced model 1 produced a p-value of 0.73, providing evidence that the reduced model 1 is significantly different from the full model.

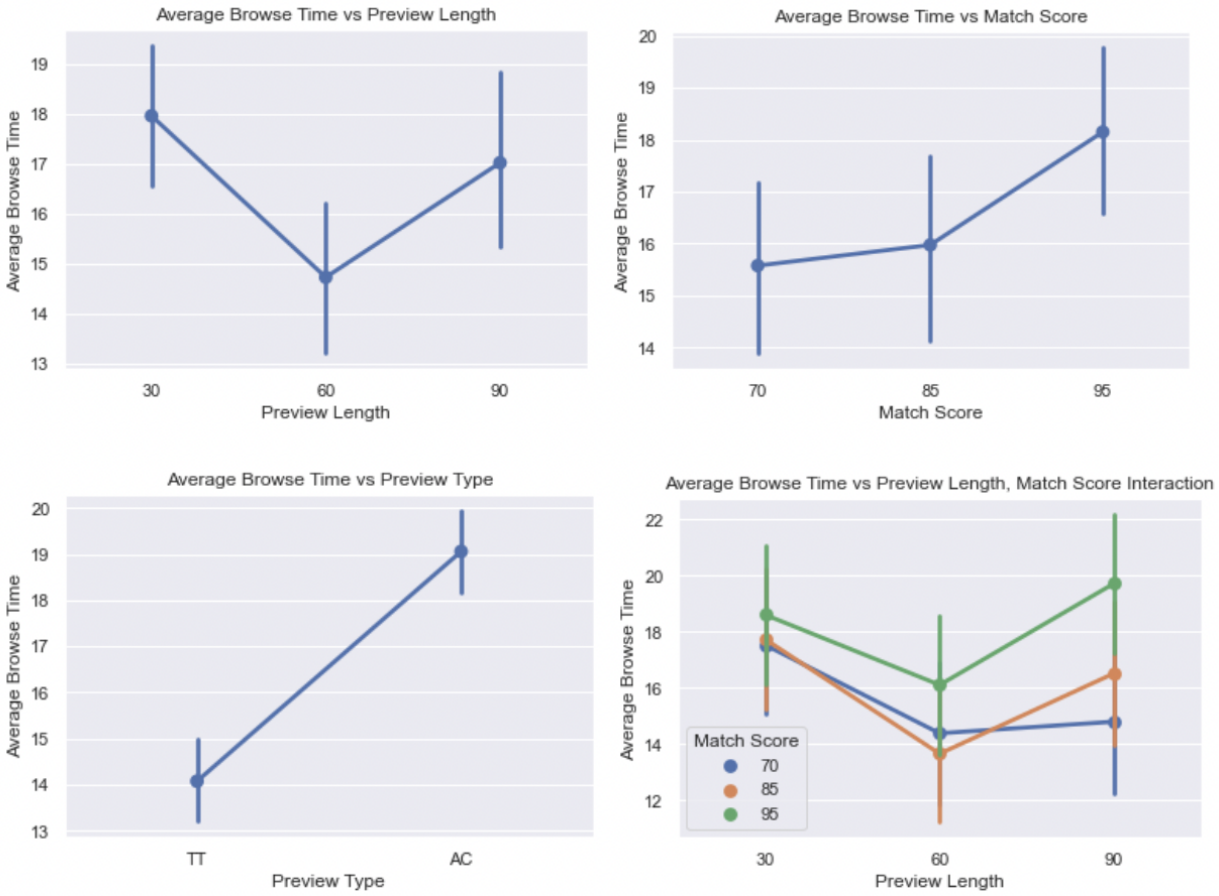
The reduced model 1 showed that the *Preview.Length:Preview.Type* term was not significant to the response variable. Another partial F test was done on the reduced model 1 vs the reduced model 2, which exempted *Preview.Length:Preview.Type.* The p-value of this test was 0.27, therefore, reduced model 2 was the better model.
All terms in reduced model 2 were significant to *Browse.Time* giving us our final model: "*Browse.Time ~ Preview.Length + Preview.Type + Preview.Score + Preview.Length:Preview.Type".*

It's important to note here that *Tile.Size* is not relevant to *Browse.Time*. *Tile.Size* will be held fixed and not be optimized for.

Exploring Significant Design Factors

The reason to explore the significant design factors is to see which factor influences *Browse.Time* the greatest. Looking at the range of *Browse.Time* for each factor and significant interaction effects will provide the best insight for what factor to adjust first. The plots below visualize these ranges and the factor effects on

Browse.Time.



Based on the plots, *Preview.Type* has the greatest effect on *Browse.Time* with a range of 5 minutes. A variance F test and student's t-test provides evidence that *Preview.Type* = TT significantly reduces browse time more than AC. *Preview.Type* will be held at TT. *Preview.Length* influences the response variable the second most, followed by *Match.Score*. Therefore, we will optimize for *Preview.Length* then *Match.Score.*

Searching For the Best Condition

From the screening experiments, using pairwise testing, we found that the condition *Tile.Size*=0.2, *Preview.Type*=TT, *Preview.Length*=60, *Match.Score*=85 had the lowest average browsing time at 11.22 minutes. The second-lowest condition had an average browsing time of 11.84 minutes. A variance F test and student's t-test were conducted and concluded that 11.22 is significantly less than 11.84. Therefore we will test new levels around *Preview.Length*=60 and *Match.Score*=85 while holding *Tile.Size*=0.2 and *Preview.Type*=TT.

The objective now is to test moderately above and below the current levels of *Preview.Length* and *Match.Score* to navigate towards the optimal levels. The new levels

tested are *Preview.Length* = [50, 70] and *Match.Score* = [75, 90] which produce 4 new conditions. Pairwise testing the new conditions with the current best condition we discovered that *Tile.Size*=0.2, *Preview.Type*=TT, *Preview.Length*=70, *Match.Score*=85 have the lowest average browsing time at 11.11 minutes, however, not significantly lower. Furthermore, this informs us that increasing *Preview.Length* from 60 results in a better MOI.

Next, *Preview.Length*=75 and *Preview.Length*=70 were tested while holding *Match.Score*=85, and found that *Preview.Length*=75 is 0.23 minutes longer, however, not significantly longer. We believe that we have found the optimal minimum of average browsing time for *Preview.Length* at 70 seconds, as of result, we'll hold *Preview.Length*=70 and now adjust *Match.Score*.

We tested *Match.Score*=85 against a lower level: *Match.Score*=80 and received an average browse time of 10.69 minutes, 0.42 lower than our current best condition. Using a variance F test and student's t-test, there was significant evidence that 10.69 minutes is lower than 11.11 minutes, hence, making *Match.Score*=80 the better level.

Lastly, we tested similar levels of the current best *Match.Score* to fine-tune the score and discovered that *Match.Score*=82 had the lowest average browsing time at 10.52 minutes, however, it is not significantly less than the current 10.69 minutes.

In conclusion, we found *Tile.Size*=0.2, *Preview.Type*=TT, *Preview.Length*=70, *Match.Score*=82 to be the optimal condition with an estimated average *Browse.Time*=10.52 minutes and 95% confidence interval of (10.36 minutes, 10.74 minutes).

Conclusion

According to our experiment, we conclude that the table shown below is the optimal location of the four factors and the optimal value of browsing time is 10.52 minutes.

Tile.Size	Match.Score	Preview.Length	Preview.Type	MOI
0.2	82	70	Teaser/Trailer	10.52

Limitations

We only choose a limited number of levels for the continous factors to conduct our experiment, which may cause the inaccuracy of our result. The region of operability is limited, while there are possibilities that the true optimal MOI lies out of our chosen region, leading to the inaccuracy of our experiment result. Meanwhile, we did not take the sample size into consideration for this experiment and set n = 100 for every condition. The sample size may not be enough for us to generate statistically significant results.