Executive Summary

Throughout this project, I have attempted to minimize the browsing time of Netflix users through finding the optimal configuration of factors that significantly influence the browsing time. During each additional phase of this project, I utilized information from prior phases and experiments.

In Phase 1 : Factor Screening, I first attempted to use a 2^{3-1} fractional factorial design to determine which factors significantly influence the response. However, due to confounding caused by aliasing, I decided to sacrifice efficiency for accuracy and use the 2^3 factorial design. Through my experiments, I found that the factors that significantly influence the browsing time are preview length and match score.

In Phase 2 : Method of Steepest Descent, I wanted to move from the initial region of experimentation towards the vicinity of the optimum. Through my experiments, I found that the optimal browsing time was somewhere in the vicinity of 90 seconds for preview length, and 70 for match score. I also sacrificed some efficiency for accuracy in this phase, by taking some extra steps in Steepest Descent, in order to be more certain of my results.

In Phase 3: Response Optimization, I decided to use a central composite design to fit a full second order response surface model in order to identify the optimum. I chose my high and low factor levels based on a spherical design, since the estimate of the response surface at each condition would be equally precise. After fitting the full second order response surface and plotting the contour plots, I found that the optimum browsing time was when preview length was 70 seconds, and match score was 77%. In addition, the estimated browsing time at the optimum was 11.53 seconds.

Thus, I came to the conclusion that Netflix should utilize preview lengths of 70 seconds and match scores of 77% in order to minimize the browsing time by users.

The Problem

The problem that I will be trying to solve is how to optimize the Netflix homepage by minimizing the browsing time needed for users before they find a show to watch.

Although there are numerous possible factors that influence browsing time, I will be considering 3 specific factors, tile size, match score and preview length. The tile size is the ratio of a show's tile height to the overall screen height. The match score is a prediction of how much a user will enjoy that particular show or movie. The preview length is the duration of a show or movie's preview when a user hovers over its tile.

In order to minimize the browsing time, I will need to find an optimal configuration of these 3 factors that minimize the expected browsing time.

The metric of interest for this problem is *average browsing time*, and the response variable is a continuous measurement for browsing time of a user.

The data for my experiments are simulated from uploading a design matrix to an online response surface simulator that replicates the random assignment of 100 users to each condition and the observation of their response variable.

The region of operability for tile size is (0.1, 0.5). The region of operability for match score is (0, 100). The region of operability for preview length is (30, 120).

I will begin with factor screening in order to determine which factors significantly influence the response variable. I will then use the method of steepest descent as well as conduct a central composite design and use a second order response surface model to find the optimal configuration of the factor levels. I will go over these steps in detail in the following sections. I will be using a 0.01% significance threshold throughout this project since a p-value lower than 0.01 would present significant evidence against the null.

Goals of Response Surface Methodology

Response Surface Methodology aims to conduct a sequence of experiments to obtain an optimal response variable, utilizing information from prior experiments to aid in future ones.

This goal can be achieved by characterizing the relationship between the expected response and a subset of the design factors. The set of possible values that this subset of the factors can take is called the region of operability. This region is where we explore and run our experiments to find the optimal condition.

The function characterizing the relationship between the expected response and a subset of the design factors is unknown, so we fit models to approximate it. We typically use Taylor's Theorem and low order polynomials, especially in small localized regions of experiments, where low order polynomials should well approximate the function. However, second order models are used for response surface optimization since they are capable of modeling concavity/convexity.

An example of response surface methodology is using factor screening to identify factors that significantly influence the response variable, and then following up by optimizing that response with the method of steepest ascent/descent and response surface designs. Each experiment conducted uses information from previous ones and the goal is to find the optimal operating condition.

I began by trying a 2^{3-1} fractional factorial design with design generator Tile.Size = Prev.Length:Match.Score, resulting in the principle fraction design. Being a fractional factorial design, it only has half of the 8 conditions in a full 2^3 design. However, each main effect is aliased with a two factor interaction effect.

The high levels for tile size, match score, and preview length are 0.3, 100, 120 and the low levels are 0.1, 80, 100, respectively. After simulating the data and fitting the model, I found the output provided p-values associated with t-tests of the hypothesis

$$H_0 : \beta = 0 \text{ vs } H_A : \beta \neq 0$$

for each regression coefficient in the model. The p-values for each main effect was significant. However, due to confounding from the aliasing, there is no way to conclude if each main effect was really significant, or if it was due to a two factor interaction effect.

Factors_fractional	Pvalues_fractional
Preview Length	<2e-16
Match Score	<2e-16
Tile Size	<2e-16

Thus, I decided to sacrifice some efficiency for accuracy and simulated the other 4 conditions of the full 2^3 model, which resulted in the complementary fraction design. Thus, with all 8 conditions, I analyzed the experiment as a full 2^3 design without any confounding. The output provided p-values associated with t-tests of the hypothesis

$$H_0 : \beta = 0 \text{ vs } H_A : \beta \neq 0$$

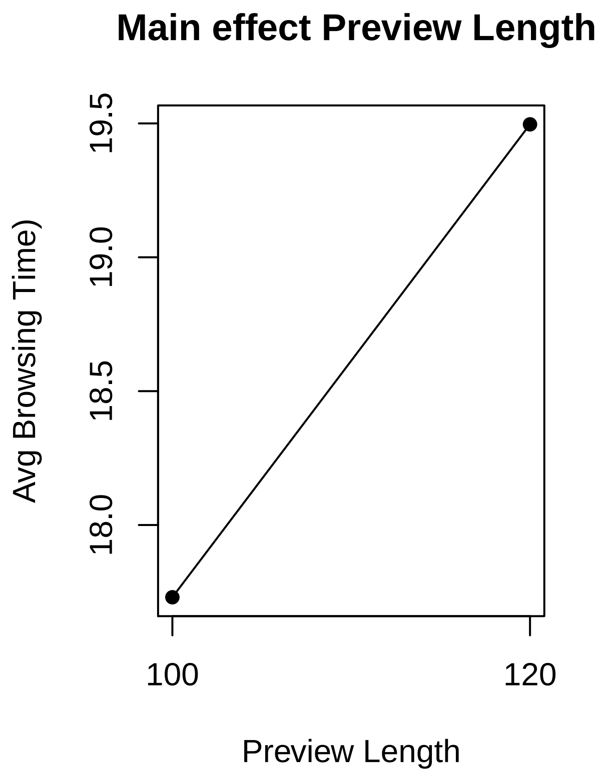
for each regression coefficient in the model. The p-value for tile size was $0.787 > 0.01$ so it is not significant at the 1% level. In addition, all 2 factor interaction effects that included tile size are also not significant at a 1% level since their p-values are greater than 0.01 as well. Thus I concluded that tile size does not significantly influence the response variable and I excluded it in future experiments.

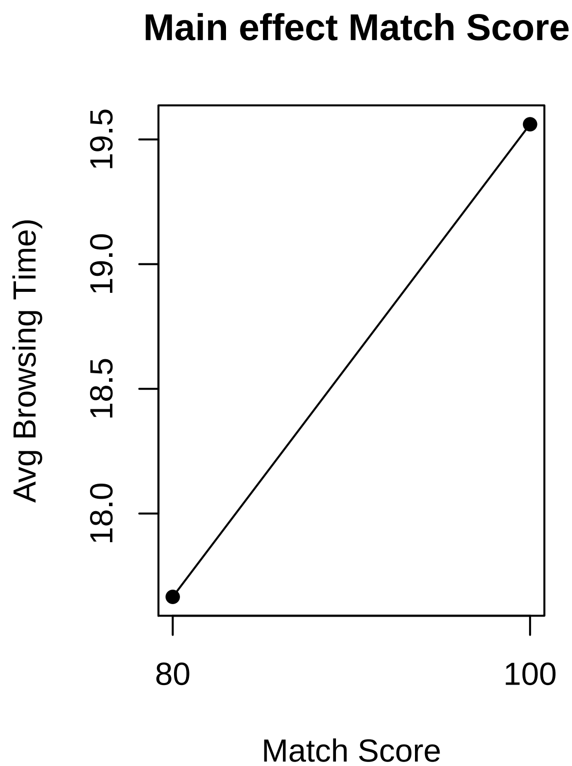
Factors_full	Pvalues_full
Preview Length	<2e-16
Match Score	<2e-16
Tile Size	0.787
Prev.Length:Match.Score	<2e-16
Prev.Length:Tile.Size	0.613
Match.Score:Tile.Size	0.709
Prev.Length:Match.Score:Tile.Size	0.342

We can get the effects for the active factors by multiplying their $\hat{\beta}$ estimates by 2.

Preview Length : $2\hat{\beta} = 1.76624$. Thus, as compared to when preview length is 100, when preview length is 120, we expect the average browsing time to increase by 1.76624 minutes

Match Score : $2\hat{\beta} = 1.85906$. Thus, as compared to when match score is 80, when preview length is 100, we expect the average browsing time to increase by 1.85906. minutes





The main effects plots agree with the results from earlier. Browsing time seems to increase as match score increases from 80 to 100 and preview length increases from 100 to 120.

In conclusion, I have found that tile size does not significantly influence the average browsing time, and I will exclude it in further phases. The factors that I have found that significantly influence browsing time are preview length and match score.

After factor screening, I ended up with 2 factors, preview length and match score, that I had found to significantly influence browsing time.

I then wanted to roughly determine the optimal levels for these 2 factors.

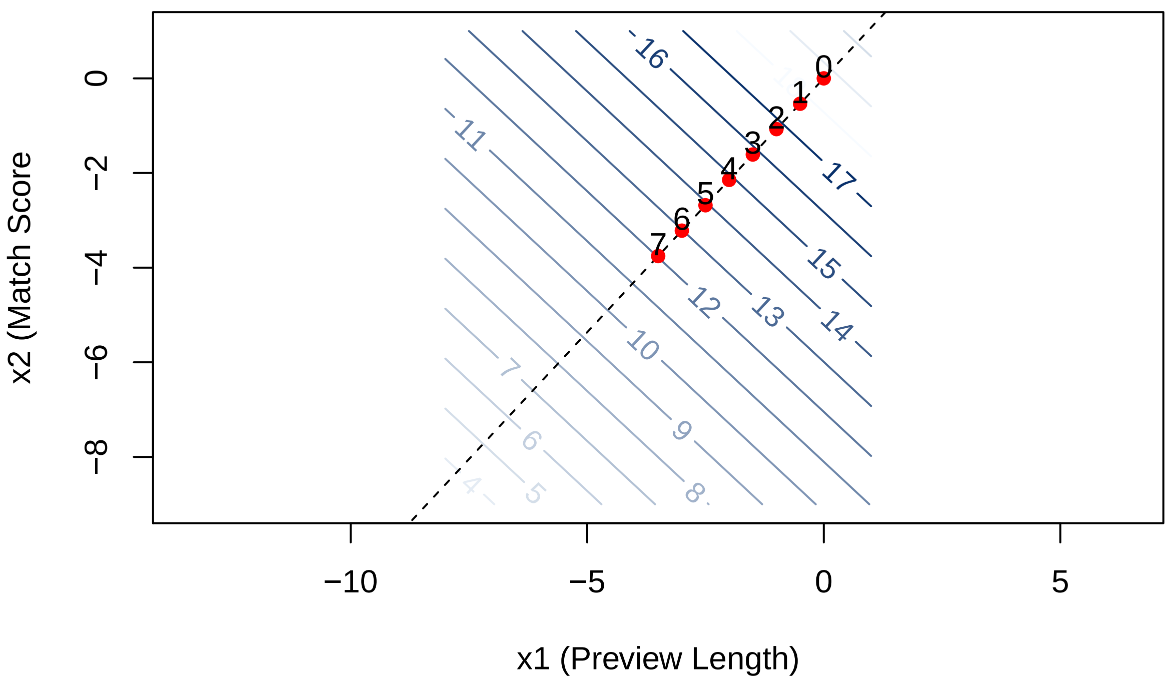
I began with a 2^2 factorial experiment with a center point condition. The initial region of experimentation was the area in which my factor screening took place. Thus, the high level for preview length was 120 and low level for it was 100. The high level for match score was 100 and low level was 80.

I first simulated the data for the center point and combined it with previously simulated data and then performed a curvature test in order to determine whether the initial region was already in the vicinity of the optimum.

I found that although β_{PQ} was significantly different from 0, since its p-value was much lower than 0.01, its size was much larger relative to the other p-values. Thus, there was likely not a pure quadratic curvature in this area and so the initial region was not likely to be in the vicinity of the optimum.

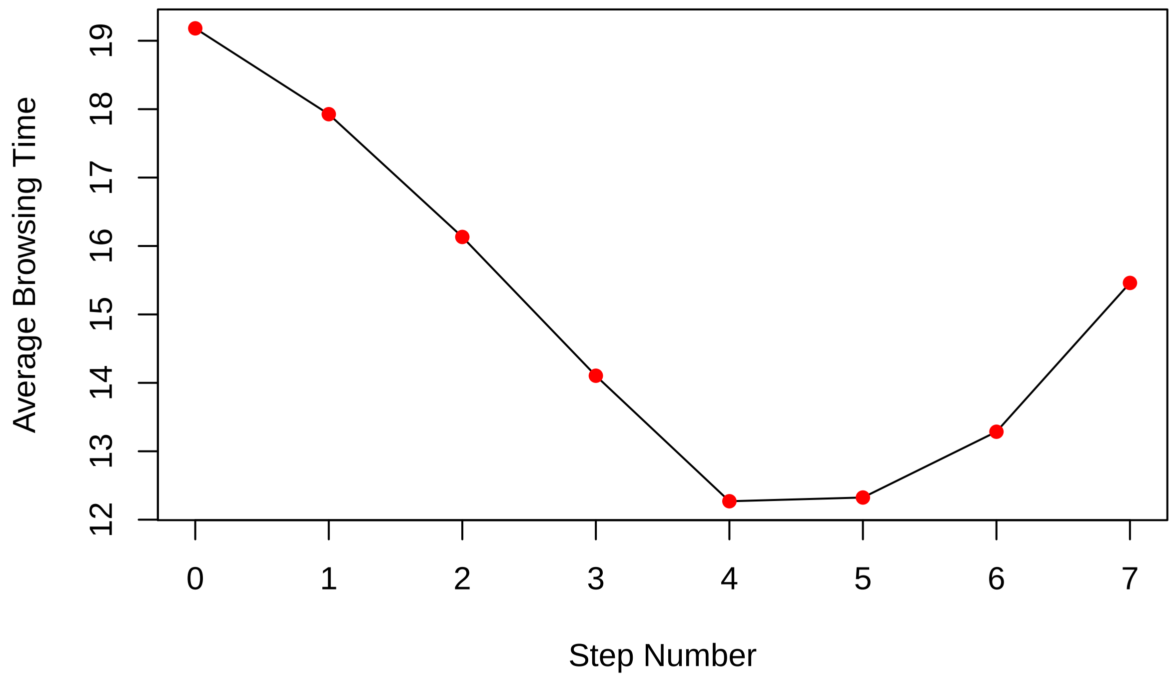
My next step was to use the method of steepest descent in order to determine roughly where in the x-space the optimum was lying.

I first fitted the first order model to determine the direction of the path of the steepest descent. I also found the 2D contour plot, with the gradient on it and we see that the starting point is (0,0).



Since preview length can only be changed in increments of 5, I decided that the steps should be 5 seconds long in preview length. I then converted it from natural units to coded units. I got that $\Delta x_1 = 0.5$ and so $\lambda = \frac{0.5}{\hat{\beta}_1} = 0.5661741$.

I took one step at a time, until I found the lowest observed average browsing time, and then I took a few more to make sure that I had found the lowest, since the following steps ended up with higher average browsing time.



I found that step 4 corresponded to the lowest observed average browsing time. It corresponded to a preview length of 90 seconds and match score of 70%.

In order to determine if I reached the vicinity of the optimum, I decided to perform another test of curvature in this region. I ran another 2^2 factorial experiment with a center point and simulated the data. My new high level for preview length was 100, and low level was 80. My new high level for match score was 80, and low level was 60.

This time, β_{PQ} was significantly different from 0 and the p-value for the pure quadratic was <2e-16 < 0.01, so I rejected the null hypothesis that the pure quadratic was 0. Thus, I concluded that I was in the vicinity of the optimum. The optimal browsing time was somewhere in the vicinity of 90 seconds for preview length, and 70% for match score.

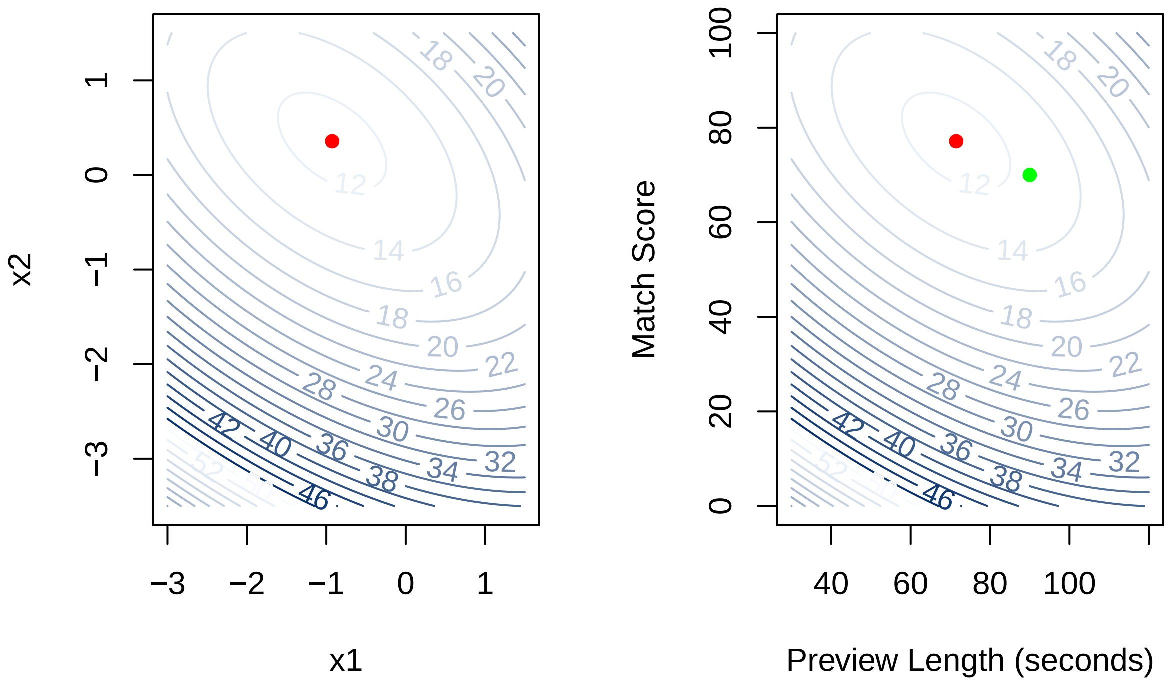
I decided to follow up my results with a response surface experiment so that a full second order model could be fit and the optimum identified. I have already identified the significant factors though factor screening, and found a rough idea of where the optimum lies through the method of steepest descent.

The response surface experiment that I decided to use was the central composite design. A central composite design facilitates estimation of the full second order response surface model, and hence identification of the optimum.

I chose the high and low levels of the factors based on the center point being my estimate of the optimum from steepest descent in phase 2. I also chose the high and low levels based on my selection of a spherical design in order to ensure the estimate of the response surface at each condition is equally precise.

Condition.Num	Prev.Length	Match.Score	Browse.Time
1	70	50	15.72111
2	70	90	15.82121
3	110	50	15.90275
4	110	90	15.93098
5	90	70	15.91739
6	120	70	15.73697
7	60	70	15.95454
8	90	100	15.91128
9	90	40	15.84565

I intended to perform axial conditions with $a = \sqrt{2}$, but the corresponding preview times and match scores were messy. Thus, in the interest of defining experimental conditions with more convenient levels, I let $a = 1.5$, yielding the preview lengths and match scores in the table above. I then generated data simulating 100 users randomized into each of these 9 conditions, and recorded their browsing time.



I then fit the full second order response surface by fitting the second order regression model.

From the coefficients in the output, I plotted the contour plot in coded units and found that the stationary point was located at x_1 = -0.9266421 and x_2 = 0.3571865.

I then converted the contour plot to natural units. The corresponding stationary point is when preview length is 71.47 (70) seconds and match score is 77.14% (77%), represented by the red point. The green point is my rough estimate of the optimal conditions from using the method of steepest descent in phase 2. It suggested a preview length of 90 seconds and a match score of 70%. As can be seen, it is somewhat close to the true optimum.

The estimated browsing time at the optimum is 11.53 seconds and a 95% confidence interval is given by (11.3855,11.675).

Thus, Netflix should utilize preview lengths of 70 seconds and match scores of 77% in order to minimize the browsing time by users.