

# Final Project

Ahnaf Ryan

15/08/2021

## Executive Summary

Due to the large size library of movies and TV shows of Netflix, users often experience choice-overload and end up losing interest and not watch anything at all. This problem potentially may increase churn rate hence the objective is to minimize average browsing time by exploring 3 factors- Tile Size, Match Score, Preview Length which are chosen based on the “Top Picks For...” row of the Netflix homepage. Our experimental journey is based on the philosophy of Response Surface Methodology where information gained in one phase is used to inform decisions in next phases. In our first phase, we performed a  $2^{3-1}$  fractional factorial experiment to save cost in our data collection process considering there are many phases where we need to collect additional data. The main effect plots of each of the factor provided evidence that they influence the response. Moreover, we performed formal hypothesis tests and concluded based on the Principle of effect Sparsity that each of our three factors significantly influence the response. With this valuable information, we moved to phase II of our experimental journey where the goal is to locate the vicinity of the optimum. In order to do this, we perform the method of steepest descent which involves two components- 1) locate a small localized region which we believe contains the true optimum by the method of steepest descent algorithm 2) check for curvature which signifies that we are in the vicinity of the optimum. We required 5 steps in our steepest descent algorithm, and we followed it with a formal test of curvature where the p-value ( $2.78e-12$ ) was practically zero. Hence, we confirmed that we are in fact in the vicinity of the optimum. With this valuable information, we set out to phase III where we began by estimating this small localized region with a full second order model. The design chosen was Central Composite Design (CCD) as it made our experimental journey more efficient by using data from phase II. Additionally, the axial conditions were chosen by considering a spherical structure as we wanted our conditions to be equidistant from our center point condition and it ensures precise estimate of each condition. With the design defined, we fitted a full second order model and calculated the stationary point for our estimated model. After converting to natural units, the optimum condition was (0.332,66,90) for tile size, match score and preview length respectively. The estimated browsing time was 12.3 seconds, almost a 20 percent improvement from the default values of factors , which had average browsing time of 15 seconds.

## Introduction

Netflix, being a multibillionaire dollar company, can generate millions of dollars upon improving user experience of their product. In this particular instance, Netflix identifies a problem with their service where wide variety of choices and options for their TV shows and movies create a psychological phenomenon known as decision paralysis for their users. Due to decision paralysis faced by users, they may become overwhelmed with this large number of choices and may use a lot of time browsing for things to watch. Hence, this may lead up to users unsubscribing their accounts and thus churn rate of Netflix may increase. Therefore, if we can improve the average browsing time of users by exploring certain factors such as Tile Size, Match Score and Preview Length of TV shows/movies, we can make the decision of what to watch for the users under the hood. Majority of TV shows and movies in Netflix are popular or highly rated therefore when a user engages in a particular TV show, for instance, they may end up watching this TV show for several seasons which results in subscribing to Netflix for several months. Therefore, minimizing browsing time, will generate revenue for Netflix and improve user activity. Now, the way we tackle this problem, is through a series of experiments where each experiment generates new information for us and allows us to use this information in consequent experiments. This philosophy constitutes the Response Surface Methodology where we perform experiments in three sequential phases:

- Phase I: Factor Screening
  - In this phase, we determine which factors(Tile Size, Match Score, Preview Length) significantly influence browsing time. This is a very important phase in our experimental journey because it eliminates any unnecessary factors. So, we can save resources and be more efficient by working with factors that we find significant.
- Phase II : Method of Steepest Descent
  - After knowing our active factors from Phase I, we would like to know the region where the optimum condition is located. This is the goal of this phase where we use the method of steepest descent algorithm to locate a small localized region that contains the true optimum and confirm it by performing a test of curvature.
- Phase III: Response Optimization
  - By knowing our small optimum region from Phase II, we would like to find the optimum condition by fitting a full second order model to estimate this region. From our estimation, we can provide estimated optimum average browsing time and its confidence interval.

Response Surface Methodology allows us to be very efficient as we can see in every phase, we have used information from the previous phases. This will save us resources/cost and at the same time, provide us with the optimum condition that minimizes average browsing time.

## Factor Screening

We start our experimental journey through factor screening experiment to determine which factors influence our response. This will eliminate the insignificant factors in the next phases of the experimentation process which will help us to be more efficient and may even save costs in our data collection process. The factor screening experiment is constructed through the lens of QPDAC which is Question, Plan, Data, Analysis and Conclusion. We will explain each of the steps and their roles in the factor screening experiment below.

**Objective:** The objective for the factor screening experiment is coherent and self-explanatory. We want to determine which of Tile size, Match Score and Preview Length are significant in influencing browsing time for Netflix users?

**Plan and Data:** In the planning stage, we want to state our response variable, state the different factors, experiment units and how our experimental conditions are constructed. The response variable is Browsing time and the design factors are Tile size, Match Score and Preview Length.

Now, let us explain the design process and how our experimental conditions are constructed. We have decided to perform a  $2^{3-1}$  fractional factorial experiment to be more efficient and to save cost in our data collection process. Moreover, it is an economical way to investigate 3 factors in  $2^{3-1}$  conditions and since our experimental journey is a sequential process, we need to consider resource allocation for the next processes as well. The low and high levels of these factors for this experiment are shown below:

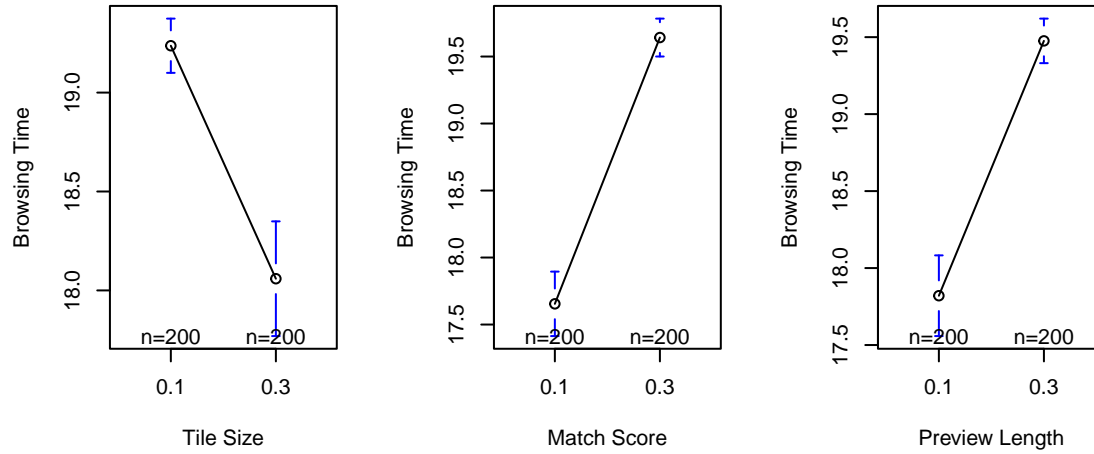
Factor	Low	High
Tile.Size	0.1	0.3
Match.Score	80.0	100.0
Prev.Length	100.0	120.0

For statistical conveniences, let us code our factors with low level as -1 and high level as +1. Also, let us consider the coded factor Tile Size to be  $x_1$ , and Match Score, Preview Length to be  $x_2$ ,  $x_3$  respectively. Since, we are using a fraction of the conditions, we used a design generator  $x_3 = x_1x_2$  to construct the high and low levels of factor Preview Length from Tile Size and Match Score. Hence, the defining relation is  $I = x_1x_2x_3$ . This is used to uncover all aliases which are  $x_1 = x_2x_3$ ,  $x_2 = x_1x_3$  and  $x_3 = x_1x_2$ . Here, shown below, are all the conditions, their coded levels and their average browsing time for visualization purposes:

Condition	Tile Size	$x_1$	Match Score	$x_2$	Preview Length	$x_3$	Average Browsing Time
1	0.3	1	80	-1	100	-1	16.23792
2	0.1	-1	100	1	100	-1	19.40275
3	0.1	-1	80	-1	120	1	19.07108
4	0.3	1	100	1	120	1	19.88041

For each condition as outlined above, we have decided to consider  $n=100$  netflix users who are the experimental units in the experiment. Moreover,  $n=100$  for each condition is sufficient for enough power in our experiment and the users are also selected at random so the proportions are same.

**Analysis:** We will begin with exploratory data analysis where we will observe the main effects plot of our factors.



From the main effect plots above, it seems that browsing time changes significantly when we change the factor level from “low” to “high” for each of the factors. Hence, all of our factors seems to influence Browsing time significantly. Let us perform formal hypothesis tests to confirm this statement.

We are going to fit a linear regression model with linear predictor  $\beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3$ . To test the significance of a factor, we are going to test three hypothesis which are  $H_0 : \beta_j = 0$  for  $j=1,2,3$ . These hypothesis determine whether the main effects are significantly different from zero hence conclude whether the factors influence the response. Since we tested multiple hypothesis, in particular 3, we need to address the multiple testing problem. However, even with the strictest correction which is Bonferroni correction, our adjusted observed p-values( $6.722183e-28$ ,  $2.394960e-62$ ,  $8.359811e-48$ ) for our hypothesis, are practically zero.

**Conclusion:** Since we performed a fractional factorial experiment, we must be aware of the confounding effects in our experiment, in particular, all of our main effects are confounded with interaction effects as detailed in the design process. However, based on the Principle of effect sparsity, we know that these interaction terms are less likely to be significant than our main effect terms. Our adjusted p-values, found in analysis portion of the experiment above, are extremely close to zero hence they are all significant. So, based on the Principle of effect sparsity, we conclude that all main effects significantly influences our response. Hence, Tile Size, Match Score and Preview Length are all significant factors.

## Phase II: Method of Steepest Descent

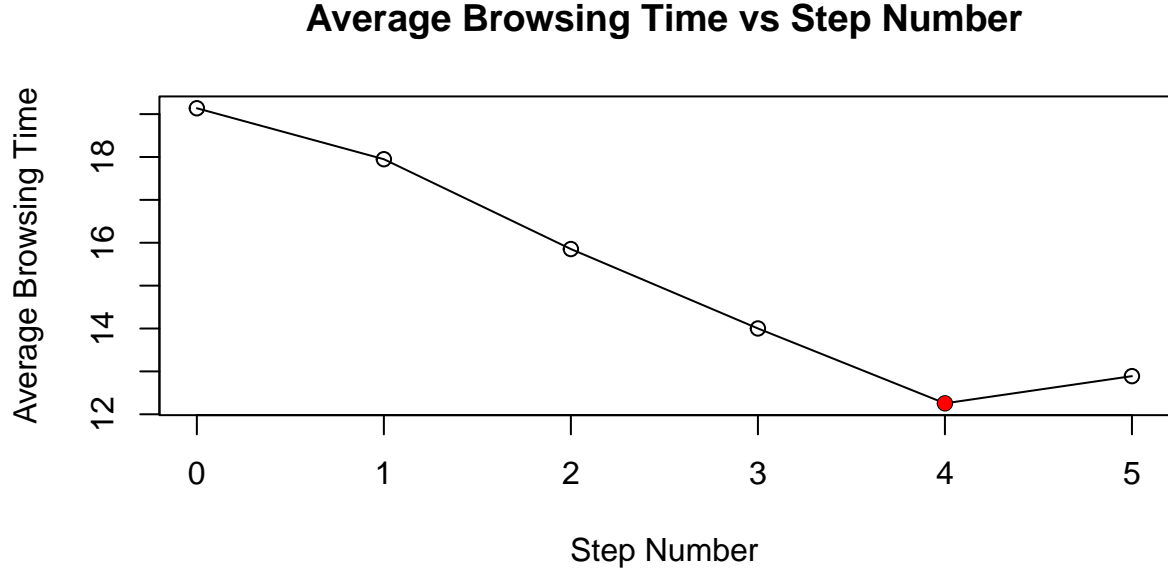
The true response surface can be estimated quite well in small localized region with a full second order polynomial model. If we consider a small localized region that contains the true optimum, we can estimate our response surface very well with a second order model. However, we need to move into this region of optimality before we fit the second order model. In our factor screening experiment, we concluded that all our factors significantly influence our response. With this information, we perform a Method of Steepest Descent experiment with these active factors to move quickly into the small localized region of optimality. Similar to Phase I, we are going to explain the experiments through the lens of QPDAC.

**Objective:** The objective is to locate the vicinity of the optimum and then confirm it with a test of curvature.

**Planning, Data and Analysis:** As mentioned in our introduction, one of the goals of response surface methodology is to pass on information from one experiment to another. Hence, from the factor conditions defined in our factor screening experiment, we can construct our center point condition and just collect data for this center point condition only as we already collected data for the other conditions from Phase I. This is a benefit of the response surface methodology. Since our goal is to minimize browsing time, we are performing method of steepest descent. Now, with this data, we have estimated a first order model to use in our method of steepest descent algorithm as the gradient of this estimated surface gives us the direction of the optimum region and allows us to move quickly into that region. The algorithm is performed with a fixed step size which was chosen from the factor Preview Length to ensure that Preview Length changes by 5 seconds only as it is a requirement from the Netflix Server for purposes of experimentation. In every iteration of our algorithm, we constructed a new condition with the help of our gradient and step size and calculated average browsing time for that condition.

Step	Tile.Size	Match.Score	Prev.Length	Browse.Time
0	0.2000000	90.00000	110	19.13674
1	0.2355729	83.99822	105	17.94830
2	0.2711458	77.99643	100	15.85279
3	0.3067187	71.99465	95	14.00117
4	0.3422916	65.99286	90	12.25457
5	0.3778645	59.99108	85	12.88932

The table, above, reports the browsing time for each of our steps and is visualized in the plot below.



As you can see from the plot, we observe the lowest average browsing time in step 4 (marked in red). Hence, we stopped our experiment in step 5 as the condition in step 5 did not improve our average browsing time. We are now confident that we are in the vicinity of the optimum and ready to move to the next process in our phase which is the test of Curvature. We perform the test of Curvature by associating a  $2^2$  factorial experiment with a center point condition. We have used the condition in step number 4 in method of steepest descent algorithm as the center point condition as it gave us the minimum average browsing time in our algorithm. Moreover, for the low and high levels of the other factors, we made sure in consideration from step 3 and step 5 that the values are wide enough to include the quadratic curvature. Below, given are all the conditions defined for the test of Curvature.

Condition	Tile Size	x1	Match Score	x2	Preview Length	x3	Average Browsing Time
1	0.38	1	62	-1	80	-1	12.23632
2	0.30	-1	70	1	80	-1	10.43110
3	0.34	0	66	0	90	0	12.31301
4	0.30	-1	62	-1	100	1	14.62285
5	0.38	1	70	1	100	1	15.16943

Note that similar to Phase I,  $x_3$  is aliased with  $x_1$  and  $x_2$  with  $x_3 = x_1 x_2$  as it will help us to save resources. Hence, we just collected data ( $n=100$ ) for 4 conditions around this center point. Now, with these 5 conditions we cannot separately estimate all of the coefficients in a second order model hence instead we fitted a modified model with linear predictor  $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{PQ} x_{PQ}$  where  $\beta_{PQ}$  simultaneously estimates the coefficients of the second order terms and  $x_{PQ}$  is 0 when  $(x_1, x_2)$  is (0,0) or 1 otherwise. The reason for using a modified linear model is to save resources. However, we must be aware that our hypothesis conclusion assumes that second order terms if considered in a full second order model, have the same sign and we are not in a saddle point. This is a very practical route to consider as fitting a full second order model requires collecting additional data when we have to allocate resources to other phases as well.

**Conclusion:** After performing the test of overall curvature, we find that the p-value is  $2.78e-12$ , which is practically zero. Therefore, we do not need to worry about the assumption error. Since our p-value is practically zero, we conclude that there is a quadratic curvature and we are in fact in the vicinity of the optimum.

### Phase III: Response Optimization

In Phase II, we identified a small, localized region which we believe contains the true optimum. As we mentioned before, our true response surface is estimated well in small localized region by a full second order model. Hence, let us fit a full second order model to estimate this region we identified and find the stationary point of our estimated response surface. As a result, we can infer about the expected minimum browsing time from our stationary point. As before, we are going explain our experiment through the lens of QPDAC.

**Objective:** The objective is to locate the optimum operating condition in the small localized region we identified in Phase II.

**Plan, Data and Analysis:** It is a wise and practical decision to use central composite design (CCD) to estimate our full second order response surface model because we already have the two-level factorial conditions, and the center point conditions from our phase II experiment. Hence, we just need to gather data for our axial, or star, conditions. This is a great way to save resources and be more efficient in our data collection process. In particular, the central condition was chosen from step 4 in our steepest ascent algorithm as it was the optimal condition in phase II and the low and high levels were chosen considering step 3,4,5 such that it is wide enough to include the quadratic curvature captured in step 4 and 5. Additionally, the axial conditions are captured by considering a spherical design with  $a = \sqrt{2}$  because all of our conditions are equidistant from the center point conditions. Moreover, the estimate of our response will be equally precise at each condition. Based on the conditions defined, we collected data and below given are the average booking rate for each of the conditions we defined.

Condition	Tile Size	x1	Match Score	x2	Preview Length	x3	Average Browsing Time
1	0.380	1.0	62	-1.0	80	-1	12.23632
2	0.300	-1.0	70	1.0	80	-1	10.43110
3	0.340	0.0	60	-1.5	90	0	13.27241
4	0.284	-1.5	66	0.0	90	0	12.16762
5	0.340	0.0	66	0.0	90	0	12.31301
6	0.396	1.5	66	0.0	90	0	12.10909
7	0.340	0.0	72	1.5	90	0	12.48825
8	0.300	-1.0	62	-1.0	100	1	14.62285
9	0.380	1.0	70	1.0	100	1	15.16943

Note, since Match Score is an integer, the value of a is rounded up to be 1.5. Based on the conditions defined and data collected, we fitted a full second order model. We, then, proceeded to calculate the stationary point of our estimated full order model and found out that the natural optimum conditions are (0.332,66, 90) for tile size, match score and preview length respectively.

**Conclusion:** The estimated average browsing time from our model is 12.3 seconds with confidence interval [12.09459 12.50825]. This is a massive improvement of approximately 20 percent from our default set of values of the factors which gave an average browsing time of 15 seconds. One limitations is that we assumed that this is the global optimum which is a practical assumption as otherwise we need to collect additional data and perform Phase II again to see if we can find another valley that may contain another optimum small localized region. However, a 20 percent improvement in average browsing time will generate Netflix more revenue than it was generating before which is a positive step for the company.