Netflix Final Project
Name: Rayan Roy
Course: Stat 430

Executive Summary:

Netflix is one of the world leaders in the entertainment industry. It provides streaming services to over 209 million paid customers worldwide. The purpose of this project is to optimize Netflix's homepage and minimize a customer's browsing time on its page. Three factors are being explored namely the tile size, match score and preview length. A series of experiments are conducted to determine which factors significantly influence the browsing time and an attempt is made to find an optimal configuration of them, that can minimize the expected browsing time.

In phase I, preview length and match score appear to be the most significant factors influencing the browsing time. In phase II, an initial region of operability is chosen which did not turn out to be in the vicinity of the optimum. As a result, a gradient descent algorithm is run to find the location of the lowest browsing time. It is found out that 75 seconds in preview length and 75% match score leads to a vicinity of the optimum which is confirmed by the curvature test.

In phase III, a central composite design is conducted to find the second order model that estimates the location of the optimum. We find that 71.52 seconds of preview length and 76.21% of match score is the location for lowest optimum with the expected browsing time being 10.144 minutes. A more practical time of 70 seconds of preview length and 76% of match score is recommended to Netflix that achieves an expected browsing time of 10.16 minutes. This value is quite close to the most optimal browsing time.

Introduction:

Netflix is a streaming service that offers a wide variety of TV shows, movies, documentaries. It provides streaming services on phone, tablet, laptop, TV without any ads and are run on a monthly subscription plan. Due to its amazing content, it has been able to retain a large number of customers year over year.

Netflix's homepage is laid out in a grid-like system where TV shows/movies appear as tiles with rows differing based on some categorization. As users, oftentimes it becomes overwhelming to choose which show to watch due to the availabilities of many options. As a result, many lose interest and don't watch anything. To tackle this challenge, Netflix comes up with various techniques to recommend users a variety of shows quickly so that they can choose one easily.

Our goal in this project is to conduct various experiments and determine factors that significantly influence the browsing time. In addition, an optimal configuration of factors will be determined that minimizes the expected browsing time. This will be done by using a web-based simulator in which we will run the experimental designs and obtain the response observations. The browsing time is the length of time that a user spends browsing on Netflix. The three factors of interest are tile size, match score and preview length. Tile size is the ratio of the tile's height to the overall height (smaller ratio corresponds to larger number of tiles shown on the screen) while the match score is the prediction of how likely one will enjoy watching a movie/show based on their viewing history, which is recorded as a percentage (higher percentage indicate higher enjoyment). Lastly, the preview length corresponds to the duration of show/movie's preview in seconds.

The table below summarizes the region of operability for each of these factors and the default values they take on when not being experimented with.

Factor	Code Name	Region of Operability	Default Value
Tile Size	Tile.Size	[0.1,0.5]	0.2
Match Score	Match.Score	[0,100]	95
Preview Length	Prev.Length	[30,120]	75

Table 1: Region of operability and default value for each factor. For the purposes of experimentation, Match.Score must be an integer and the Prev.Length can only be changed in increments of 5 seconds

Response surface methodology (RSM) explores the relationship between several factors and the response variable. Its main goal is to use a sequence of experimental design to obtain an optimal response. In the first phase of RSM (also called as factor screening phase), 2-level factorial experiment are used to determine which among the three factors significantly influence the response. Once the significant factors are found, further experiments are conducted to find the optimized response. Phase two involves using the method of steepest ascent/descent to move from the initial experimentation region to the vicinity of the optimum using the significant factors from phase I. The third phase is the response surface design where the location of the optimum, that minimizes the expected browsing time, is found.

Phase I-Factor Screening:

The factor Screening phase is the first phase of the experiment. The purpose of this phase is to find the most significant factors among a large number of factors that influence the response variable. Two-level (2^k factorial) design are passed into the simulator. This full factorial design, each with two levels, consists of all unique possible combinations of levels for all the factors and allows us to determine the effects of each of the factors and its interactions on the response variable.

Factor	Low	High
Tile.Size	0.1	0.3
Match.Score	80	100
Prev.Length	100	120

Table 2: Different factors and their low and high level values

In our experiment, a full 2^3 factorial design is used since there are 3 design factors of interest for us namely preview length, match score and tile size. The table above shows the factors that were used for the factor screening process with the high and low levels of each factor. We chose the full factorial experiment instead of a fractional factorial design because we think that $2^3 = 8$ is a reasonable number of conditions. We also get a lot of information about each of the factors.

The design consists of 8 rows, one for each condition, where each row consists of different combinations of the factors with their levels. The data for the design was collected via a simulator based on our student ID. It generated $n = 100$ browsing times (recorded in minutes) for each condition in the design. Since the design consisted of 8 different conditions, a total of 800 rows was generated. Each row represents a single user of the Netflix with the browse time and condition they were in. The browsing time only records the time spent scrolling and searching.

After fitting in the linear regression model with all the 3 factors, the p-values for both the preview length and match score are close to 0 while the tile size has a p-value of 0.262 which is above the 5% significance level. Furthermore, a reduced model with tile size and match score was fit and an Anova test is conducted which shows that preview length and match score appear to be significant. Thus, the factors that most significantly influence the browsing time are match score and preview length.

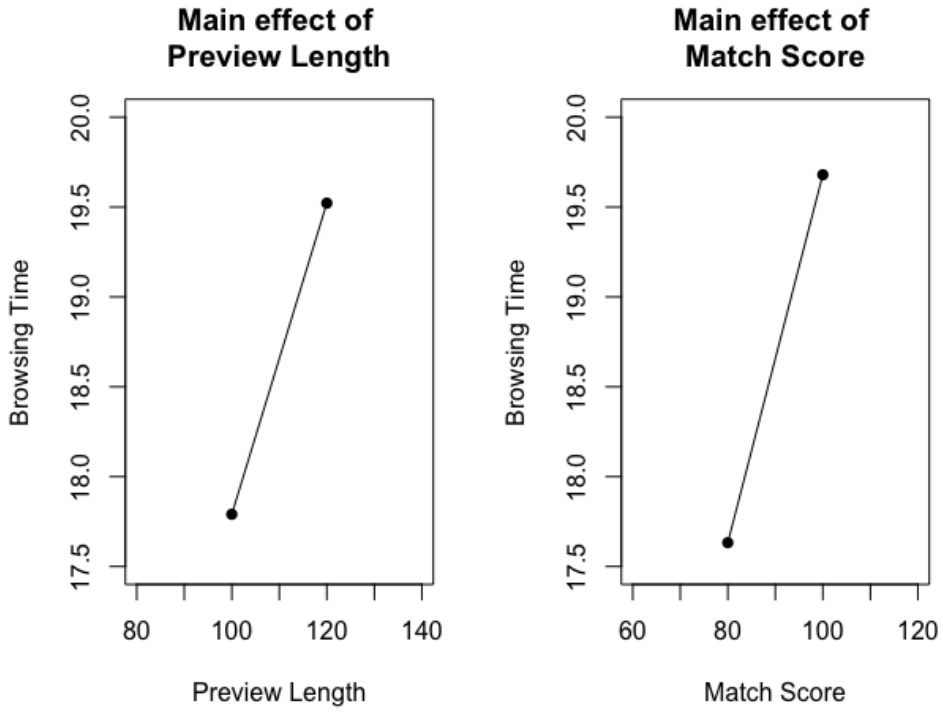

Fig. 1: Plot Main effects of preview length (left) and plot of main effects of match score (right)

The above two plots depict the main effect plots of the preview length and match score. From the main effect plot of preview length, it can be concluded that the browsing time increases as we move from a preview length of 100 seconds to preview length of 120 seconds. Similarly, from the main effect of match score, we can conclude that the browsing time increases as we move from a match score of 80% to a match score of 100%.

Phase II-Method of Steepest Descent:

The method of steepest descent is the second phase of the experiment. In this phase, we move from the initial region of experimentation to the vicinity of the optimum. We saw in phase I, that preview length and match score were the most significant factors.

The initial region of experiment is chosen to be [75, 105] for preview length where 75 seconds is the low level and 105 is the high level and [69,99] for match score where 69% is the low level and 99% is the high level. It was known that the region of operability in phase I (*Table 2)* was not in the vicinity of the optimum. A little shift in the region of operability is made so that we can reach the vicinity of the optimum in the least number of steps.

A 2^2 factorial experiment with initial region of experiment and center point (90 seconds, 84%) condition is conducted. Different conditions as shown in the table below are submitted in the simulator and browsing time is collected. The summary of the data is as follows:

Condition	Preview Length (seconds)	Match Score (%)	Average Browsing Time (min)
1	75	69	15.83861
2	105	69	18.48309
3	75	99	16.38265
4	105	99	18.98763
5	90	84	14.82871

Table 3: Average browsing time for different conditions in the initial region of experiment

A curvature test is performed to determine if the experimental region was already in the vicinity of the optimum. The p-value we find is 0.187 which is greater than 0.05 level of significance. Hence, we know that we haven't reach the vicinity of the optimum.

We embark upon the path of steepest descent. The first step is chosen as 90 seconds preview length and 84% match score because that is the center point condition we get from the curvature test. The gradient defines the direction of steepest descent and is determined to be [0.9913, 0.666] with the step size λ as 0.3362648. The step size controls whether the gradient descent converges to a minimum slowly or quickly. In our case, it is based on preview length and determines the decrement of the match score. The step size and gradient were kept constant throughout the descent. Data is collected based on each step and the browsing time is determined. Note that the match score was rounded to the nearest integer after each step size. The following table shows the average browsing for each step.

Step	Preview Length (seconds)	Match Score (%)	Average Browsing Time (min)
0	90	84	14.82871
1	85	81	12.81002
2	80	78	11.06648
3	75	75	10.11692
4	70	72	10.50166

Table 4: Average browsing in each step size for phase II

The plot below shows the visualization of the step number and the average browsing time.

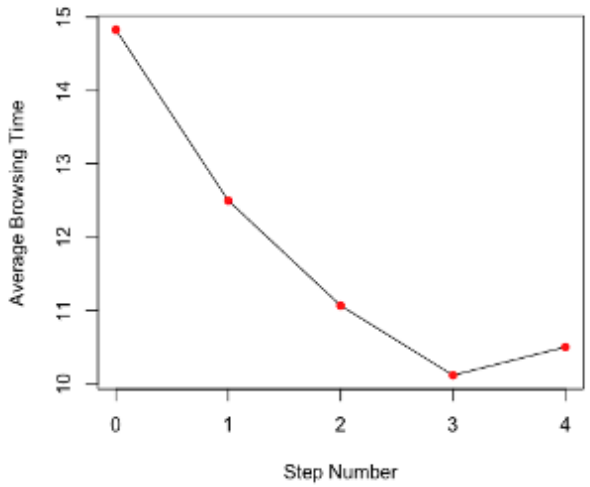

Fig. 2: Plot of Step number and average browsing time

From the graph above, we can see that *Step Size 3* corresponds to the lowest observed average browsing time. We stop at step size 4 since we see the average browsing time increases as we go from step size 3 to step size 4. Hence a test of curvature is performed in this region to see if we reached the vicinity of optimum. We do a 2^2 factorial experiment with a centre point (75 seconds, 75%). Data with the below conditions was submitted in the simulator and the average browsing time in each condition is reported in the table below.

Condition	Preview Length (seconds)	Match Score (%)	Average Browsing Time (min)
1	60	60	15.06717
2	90	60	13.20048
3	60	90	11.89288
4	90	90	15.35557
5	75	75	10.11692

Table 5: Average browsing with different conditions for test of curvature

A linear regression model was fitted and the p-value of pure quadratic effect, x_{PQ}, is 1.65e-131 which is less than significance level of 0.05. This indicates the presence of a quadratic curvature.

Phase III-Response Optimization:

A following response surface experiment like the central composite design (CCD) is conducted to estimate the full second order model that can be fitted to locate the coordinates of the optimum. In our experiment, CCD consists of the 4 factorial conditions with 2^2 factorial design, the centre point condition (1 condition) and the axial conditions (4 conditions) that lie outside the region of the factorial condition.

The low and high level for preview length is chosen to be 60 and 90 seconds respectively. Similarly, the low and high for match score was 60% and 90% respectively. They were chosen based on the centre point condition of 75 seconds and 75% which we obtained from Phase II to be in the vicinity of the optimum.

The axial condition in this experiment is $\alpha = \sqrt{2} \approx 1.414$ since we want the axial conditions at an equal distance from the centre point like the factorial condition. The value of the axial condition is based on the number of factors being tested (i.e., 2 in our case). This design is chosen so that the estimate of the response surface at each condition is precise. The preview length and match score are 96.21315 seconds and 96.21315% based on the coded units of -1.414 and 1.414. Since preview length can only increase in increment of 5, we will choose 95 seconds for preview length and 95% for match score which corresponds to $\alpha = 1.3$. Data is collected by submitting the conditions in the simulator and browsing rates are obtained. The summary of the data is as follow:

Condition	Preview Length (seconds)	Match Score (%)	Average Browsing Time (min)
1	60	60	15.06717
2	90	60	13.20048
3	60	90	11.89288
4	90	90	15.35557
5	95	75	14.14685
6	55	75	12.03228
7	75	95	14.81442
8	75	55	14.47136
9	75	75	10.11692

Table 6: Average Browsing Time in each condition for CCD

A second order linear regression model was fit using the data from above conditions and various estimates were obtained. A contour plot of fitted response surface is shown below.

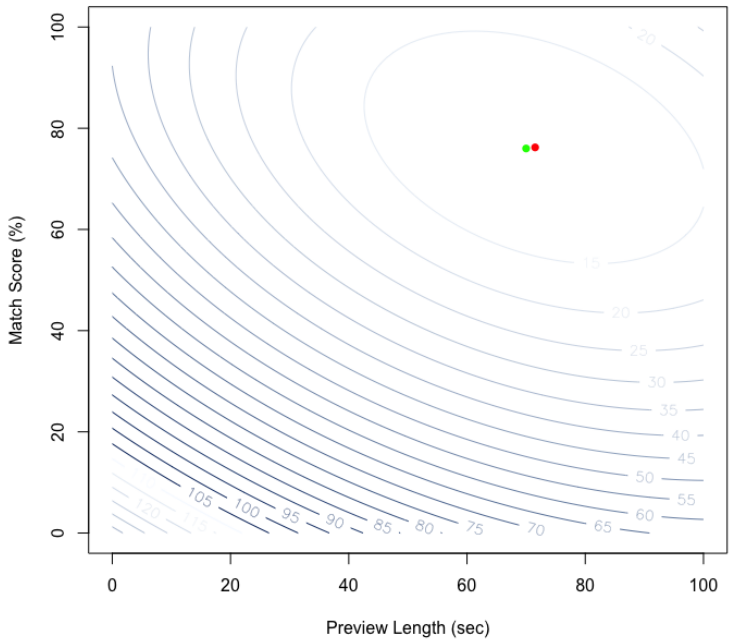

Fig 3. Contour Plot of Preview Length and Match Score

The location of the optimum (i.e., factor levels that minimize expected browsing time) is 71.52282 seconds of preview length and 76.21429% of match score (represented by the red dot above). The estimate of the expected browsing time is 10.144 minutes with the 95% confidence interval being (9.9523,10.3358).

A slightly less optimal but more practical configuration that Netflix could incorporate would be 70 seconds of preview length and 76% of match score (represented by the green dot above). This would help achieve an expected browsing time of 10.16 minutes with a 95% confidence interval of (9.975,10.348). This value is quite close to the most optimal browsing time.