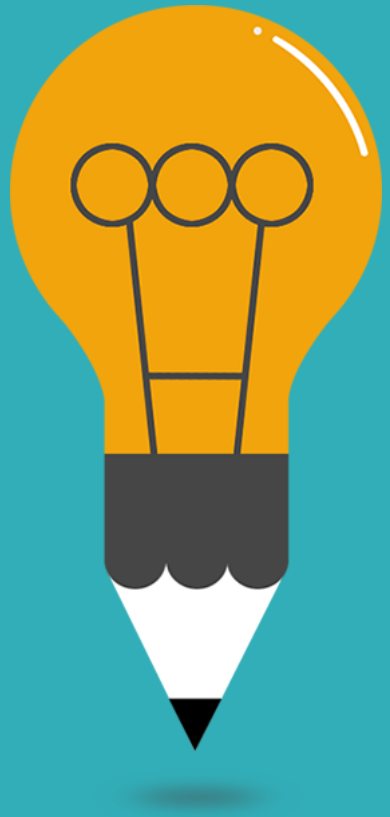
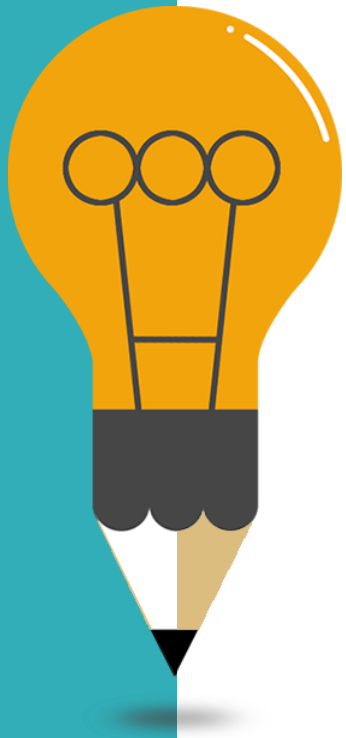
Ji Hei Kim
s3726377



Assessment 3:
Why Not Watch? Business case

MATH2406 Applied Analytics

Content



INTRODUCTION

GOAL

“To find if the WNW’s new algorithm is worth rolling out to all their subscribers by analyzing the results from a recent change they made in their recommendation engine. To find any bias in the data collected and present how this can be correct and to provide recommendations for future A/B Tests”

To perform fundamental statistical analyses (descriptive analysis, hypothesis testing, ANOVA, correlation, and linear regression) in order to identify key findings

Sample Data

01

Date (Interval)

- Period of observation
- 1/7 - 31/7

02

Age (Ordinal)

- Age of the customer
- Min age 18 , Max age 55

03

Gender (Ordinal)

- Gender of the customer
- F for Female, M for Male

04

Social_metric (Ordinal)

- Combined metric based on previous viewing habits
- 0 ~ 10

05

Time_since_signup (Interval)

- No. of months since the customer signed up
- 0 months - 24 months

06

Demographic (Nominal)

- Demographic number
- 1-4

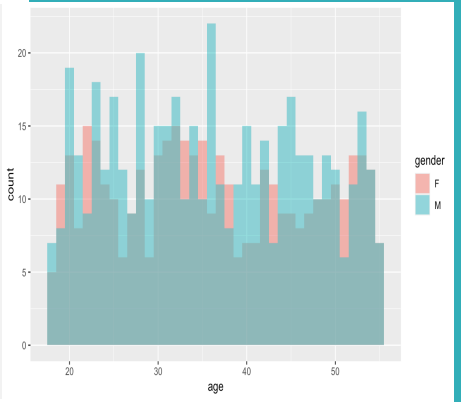
07

Hours_watched (Ratio)

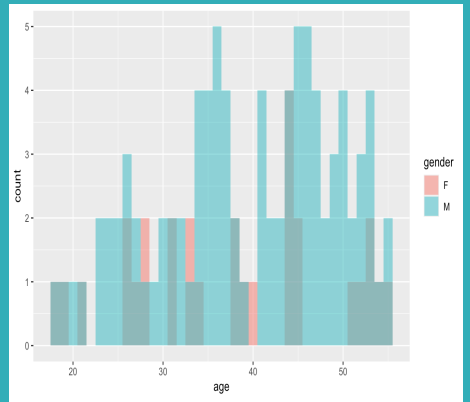
- Number of hours watched in that day
- 0.5 hours - 8.3 hours / per day

Is there any bias in the data?

Inequality in Age & Demographic ratio

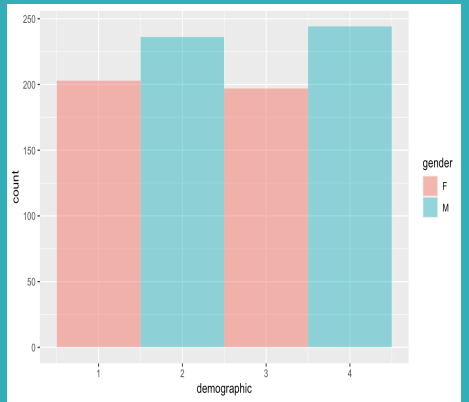


Group A

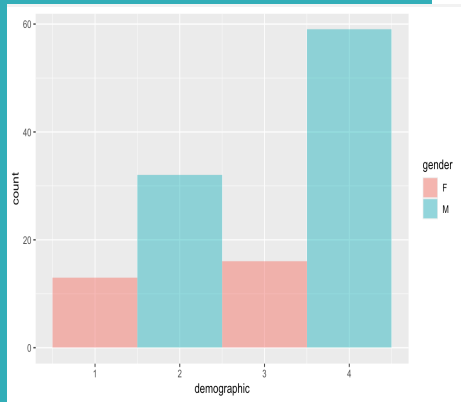


Group B

- There is inequality in age ratio.
- Absence of certain age range in Group B especially for Female



Group A

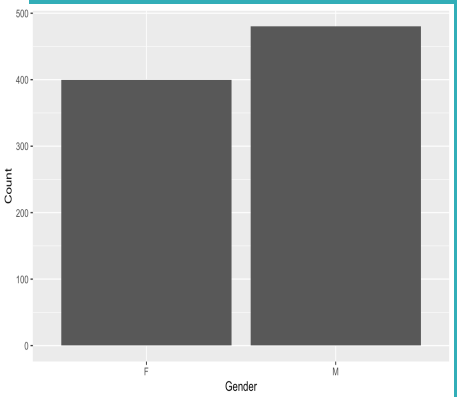


Group B

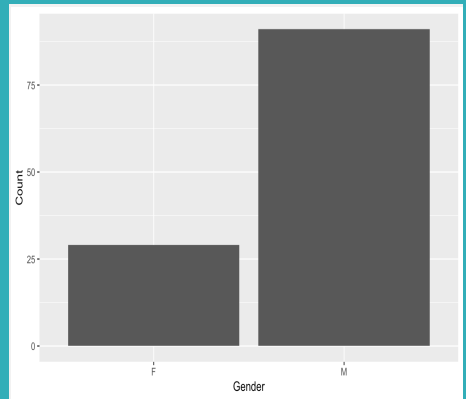
- There is inequality in demographic ratio.
- Group B subscribers from Demographic 4 has the highest number

Is there any bias in the data?

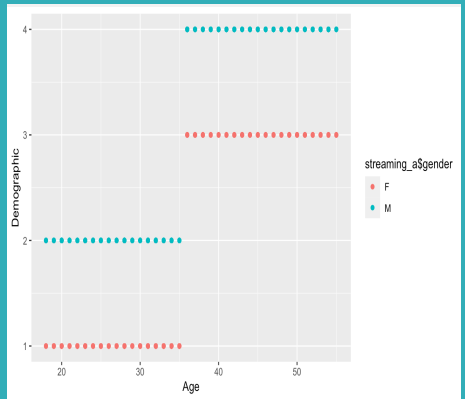
Inequality in Gender & Demographic ratio



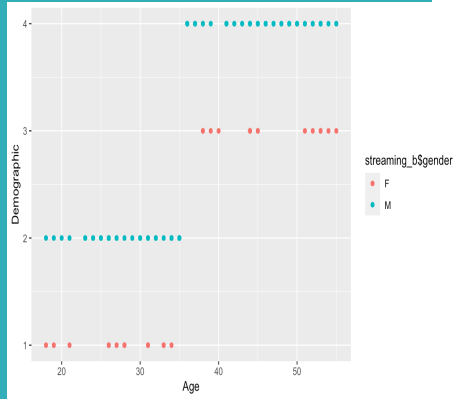
Group A



Group B



Group A



Group B

- There is inequality in gender ratio.
- For group A, 480 males and 400 females
- For group B, 91 males and 29 females

- For Group A & B, certain demographic has only has certain age groups.
- Demographic 1 & 2 only has age group between 18-35 and Demographic 3 & 4 has age group between 35 over.

How could any bias be corrected?

Remove the extreme values from the data

- E.g., the very small data close to 0 or very maximum values

Sufficient sample size

- Calculate the minimal sample size before launching the test
- Assess the test results only after test reaches the minimal sample size

Sampling is completely randomised

- Includes all demographics (Gender, age, etc.)
- Random sampling means that any customers of WNW has the same probability to be chosen to see a variation of A/B test

Sufficient period of time

- Observation period should be long enough to ensure the sample represents true population

DATA ANALYSIS:

A/B Testing / Regression

Two-sample hypothesis test

01

Null hypothesis $H_0 : \mu_1 = \mu_2$

Two groups A and B have the same efficacy, i.e. that they produce an equivalent number of hours watched in that day.

02

Alternative hypothesis $H_A : \mu_1 \neq \mu_2$

There is a difference in number of hours watched between two groups, i.e. that A and B have different efficacy.

03

Statistical significance

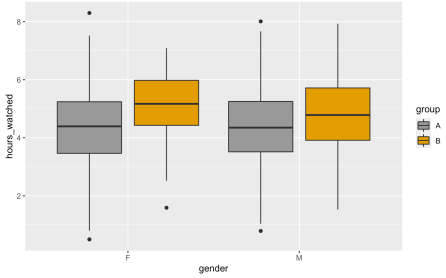
The statistical significance is measured by the p-value, i.e. the probability of observing a discrepancy between our samples at least as strong as the one that we actually observed.

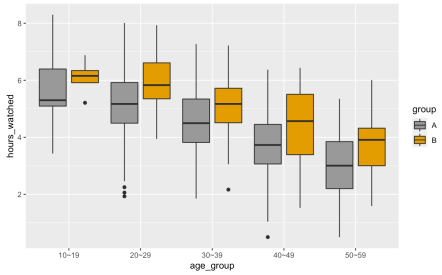
04

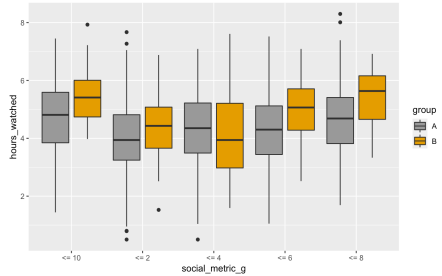
Two-tailed test

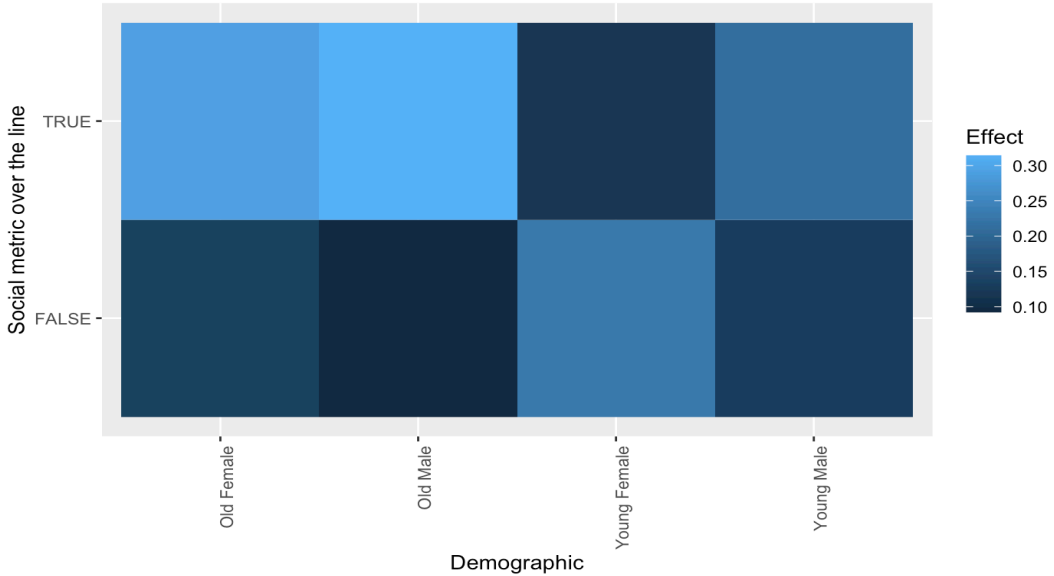
A two-tailed test has been chosen since no reason to know a priori whether the discrepancy between the results of A and B will be in favor of A or B.

A/B Test Result



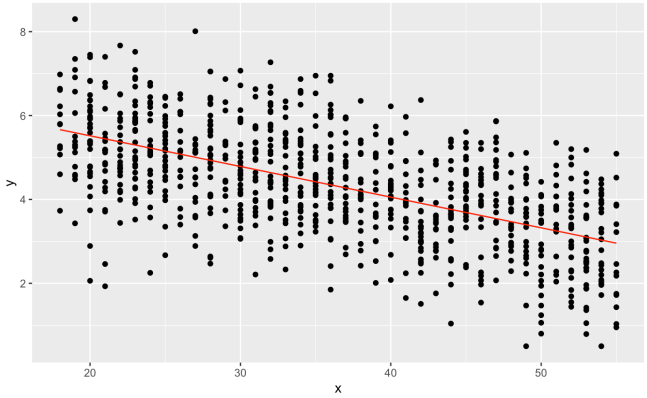


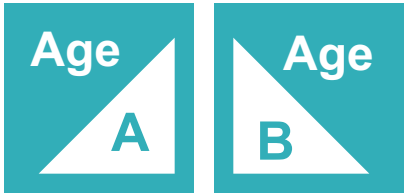


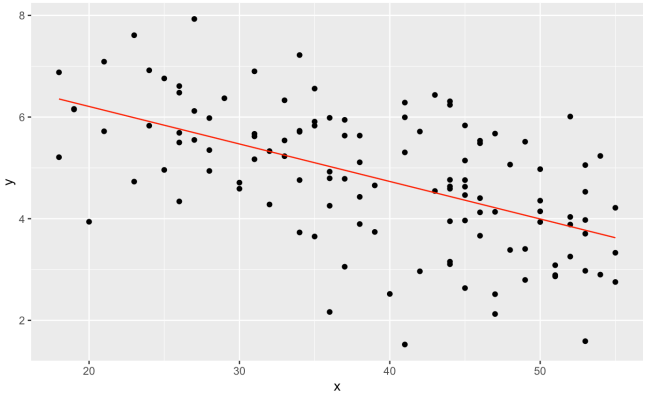


- Demographic defined by the categories: **gender**, **above**(social metrics over 5), and **young**(age below 29), **old**(age above 30)
- Increased hours watched in Group B
- Significant effect shows especially for those whose *social metric belongs 6-10* with an *older demographic*.

Linear Regression



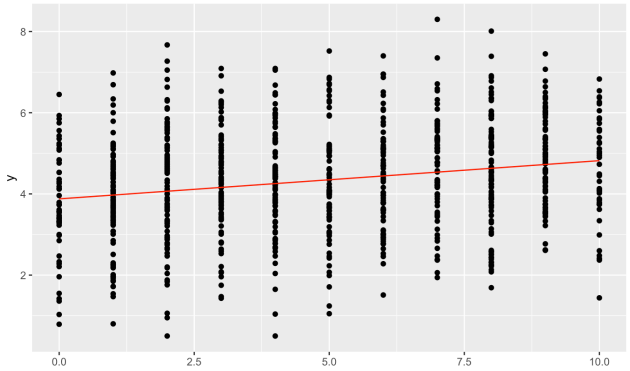


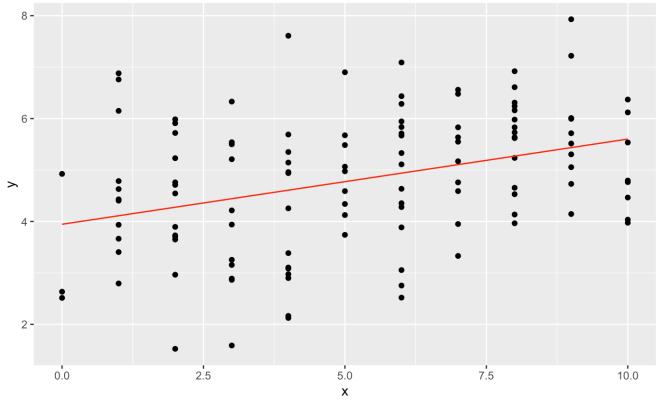



```
Call:
lm(formula = y ~ x, data = data_df)

Residuals:
    Min      1Q  Median      3Q     Max
-3.5142 -0.7242 -0.0030  0.7474  3.0046

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.980111   0.126542   55.16   <2e-16 ***
x           -0.073137   0.003356  -21.79   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.067 on 878 degrees of freedom
Multiple R-squared:  0.3511,    Adjusted R-squared:  0.3503
F-statistic:   475 on 1 and 878 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = y ~ x, data = data_df)

Residuals:
     Min       1Q   Median       3Q      Max
-3.13400 -0.76369  0.02025  0.78792  2.23803

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.68412    0.40368  19.035  < 2e-16 ***
x           -0.07378    0.01004  -7.351 2.81e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.105 on 118 degrees of freedom
Multiple R-squared:  0.3141,    Adjusted R-squared:  0.3083
F-statistic: 54.04 on 1 and 118 DF,  p-value: 2.806e-11
```

Linear Regression

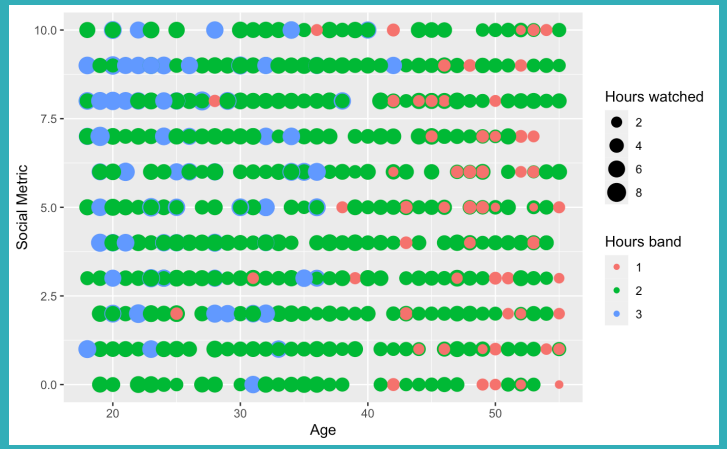


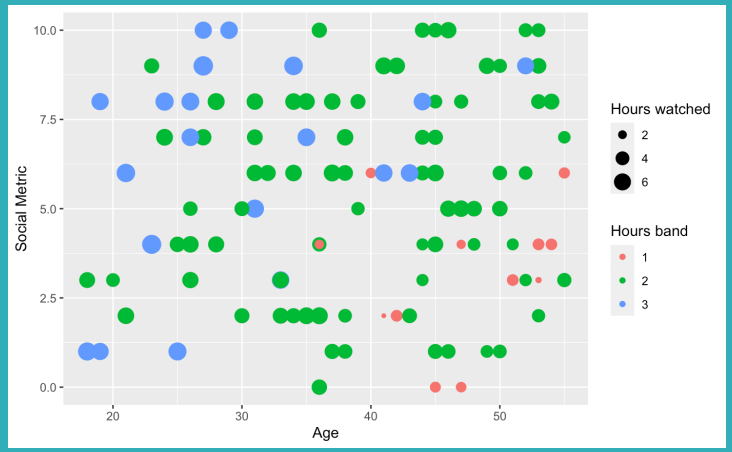



```
Call:
lm(formula = y ~ x, data = data_df)

Residuals:
    Min      1Q  Median      3Q     Max
-3.7544 -0.8255  0.0273  0.8349  3.7632

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.87785    0.08295  46.747  < 2e-16 ***
x            0.09414    0.01449   6.495 1.39e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.294 on 878 degrees of freedom
Multiple R-squared:  0.04585,   Adjusted R-squared:  0.04476
F-statistic: 42.19 on 1 and 878 DF,  p-value: 1.386e-10
```

```
Call:
lm(formula = y ~ x, data = data_df)

Residuals:
    Min      1Q  Median      3Q     Max
-2.8525 -0.8858  0.1861  0.7691  3.0019

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.94574    0.23817  16.567  < 2e-16 ***
x            0.16558    0.04003   4.136 6.65e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.247 on 118 degrees of freedom
Multiple R-squared:  0.1266,    Adjusted R-squared:  0.1192
F-statistic: 17.11 on 1 and 118 DF,  p-value: 6.654e-05
```

Multiple Regression





A B

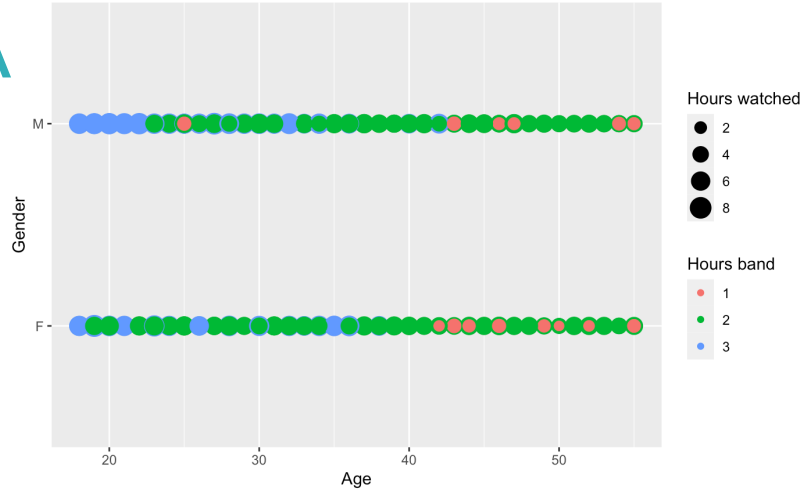
Age + Social metrics
(+ Gender??)

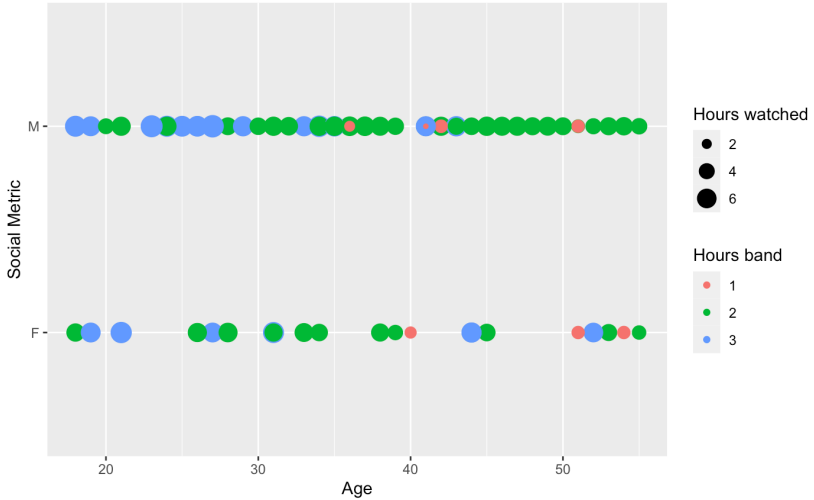
```
Call:
lm(formula = hours_watched ~ age + social_metric, data = streaming_a)

Residuals:
    Min      1Q  Median      3Q     Max
-3.6244 -0.6361 -0.0271  0.6988  2.8773

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    6.535941   0.137147  47.657  < 2e-16 ***
age           -0.072279   0.003262 -22.157  < 2e-16 ***
social_metric  0.084869   0.011619   7.305 6.25e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.037 on 877 degrees of freedom
Multiple R-squared:  0.3883,    Adjusted R-squared:  0.3869
F-statistic: 278.3 on 2 and 877 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = hours_watched ~ age + social_metric, data = streaming_b)

Residuals:
     Min       1Q   Median       3Q      Max
-2.65282 -0.61812  0.06309  0.68267  1.80700

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    6.840745   0.391174  17.488  < 2e-16 ***
age           -0.075783   0.008972  -8.446 9.47e-14 ***
social_metric  0.176314   0.031714   5.560 1.73e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9874 on 117 degrees of freedom
Multiple R-squared:  0.4574,    Adjusted R-squared:  0.4482
F-statistic: 49.32 on 2 and 117 DF,  p-value: 2.92e-16
```

Multiple Regression

Group A & B : Age + Gender … (+ Social metrics??)



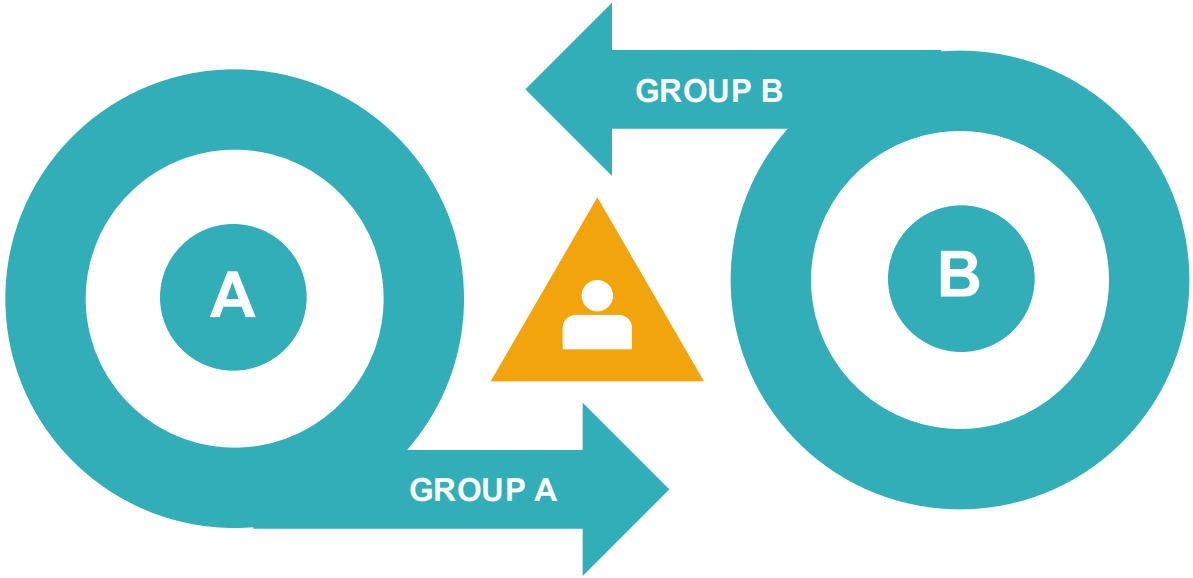



```
Analysis of Variance Table

Model 1: hours_watched ~ age + gender
Model 2: hours_watched ~ gender + age + social_metric
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    877 1000.1
2    876  942.9  1    57.254 53.192 6.769e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Analysis of Variance Table

Model 1: hours_watched ~ age + gender
Model 2: hours_watched ~ age + gender + social_metric
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    117 144.20
2    116 114.05  1    30.153 30.668 1.935e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Conclusion

After implementing new recommendation engine there has been an **increase** in overall number of hours watched.
Positive effects on customers over the age of 30 with high social metrics



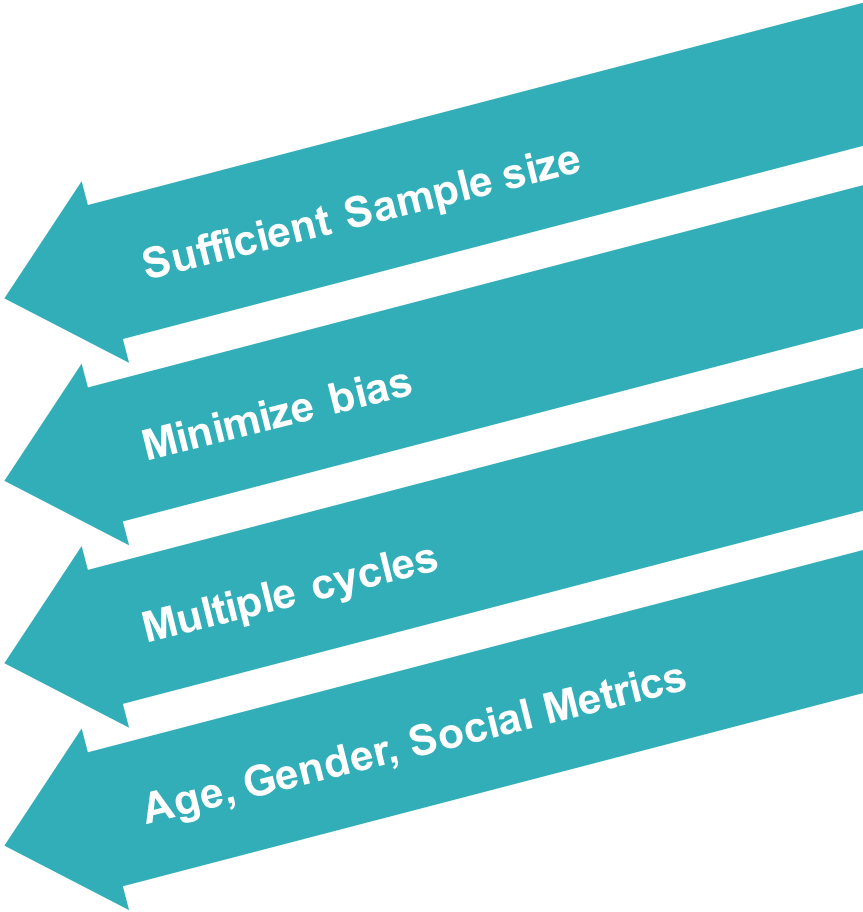
Relationship found between **age + gender + social_metric** and hours watched.
Predict hours watched using a new recommendation engine based on subscribers age, gender and their social metric.

Improvements to be made for future A/B Test:

1. Satisfy minimum sample size to make a confident call
2. Randomise sampling to gain full representative of the population
3. Have sufficient length of time for observation to conclude the effects of the treatment

RECOMMENDATION

Based on the analysis of given sample data, it is worth rolling out new recommendation engine to all subscribers



Thank you