Course Project Sample

An Analysis of Disney's Performance and Long-Term Strategy

Problem Description

Though many of us recall Disney as the 'Magic Kingdom,' it is still a profit-generating business. The two largest drivers of revenue for Disney are its stellar record of movie releases and subscription revenue from its streaming service "Disney+." In this project, I gathered a total of three datasets. Dataset 1 contains the primary variables related to a movie's box office success or gross revenue generated from Kaggle, with its primary source of data being the box office data of Walt Disney Studios from 1937 to 2016. Datasets 2 and 3 were given to me by Professor Orkun Baycik from Boston University's Questrom School of Business and details attributes of "Disney+' subscribers and Hero/Villian preferences. Recent news sprang to light when Disney reported fourth-quarter earnings that missed expectations, driving its stock price down. With rising competition from Netflix and Hulu, it is crucial to examine if any specific variables affect the media giant's top-line success. Our main point in this project is to answer one essential question: what factors contribute to the success of Disney's two highest revenue streams: movies and streaming?

Data Description

Dataset "Disney Movies Total Gross" comprises a total of 513 responses (observations) and 6 different attributes (variables) for movies. Our key performance metric for the success of said movies is the total gross revenue from box office hits adjusted for inflation [inflation_adjusted_gross]. Within the dataset, there are movies ranging from multiple genres including, but not limited to: 'musical, ' 'adventure,' 'and drama.'

Dataset "Disney+" comprises 933 responses with 6 attributes. This data was collected from an experiment where the attribute Disney+ recommendation system was assigned to different regions and the corresponding number of Disney+ subscribers in the region was reported as data. Attributes of this dataset include [Date], [Region], [PopulationAverageAge], [Average Annual Income (in thousands)], [DisneyPlus Recommendation System], and [Disney+ Subscribers (in thousands)]. The variables we plan to use in this report are [DisneyPlus, Recommendation System], [Disney+ Subscribers (in thousands)], and [Average Annual Income (in thousands)]. [Disney+ Subscribers (in thousands)] has an average of 49,000 subscribers with a range of 18,000 - 80,000 subscribers. The average [Average Annual Income (in thousands)] is $49,000 with a range of $18,000 - $80,000.

Dataset "SurveyData" consists of 459 responses with 5 attributes. This data consists of viewers aged 18-70 and how they feel about the existence of heroes and villains in Disney movies. Attributes include [Responder], [Age], [Retired], [Have Kids Younger than 7], and [Preferences]. The variables we will be using are [Age] and [Preferences]. [Preferences] has two unique response being Villians and Heroes. Villains occur in 239 observations while Heroes occur in 220 observations. [Age] has a mean of 44 and a range from 40 years olds to 70 years olds.

Section 2. Research Questions

1. **Does the average Disney movie from 1937 to 2016 make more than $302,872,154?**

Taking a look at the "Disney Movies Total Gross" dataset, we are mainly interested in the [inflation-adjusted gross] column. One interesting thing to note is that these gross revenues are all from 1937 to 2016. Disney has made a lot more successful movies from 2017 onwards. One of the goals we want to address is, how well these movies hold up in the twenty-first century, specifically from 2017 onward. Taking the top 300 domestic gross Disney movies, from "All-Time Worldwide Box Office for Walt Disney Movies," we filtered by release year, keeping those from 2017 to 2022, leaving us with 38 samples. Then, we took the average of those samples and got $302,872,154. [1]Now we wish to test if the average Disney movie from 1937 to 2016 will make more than $302,872,154. This is important as one benefit from this experiment can be that the previous movies don't hold up well in comparison to today. This would give Disney directors a better insight into what common traits to focus on when directing new movies and what to disregard.

2. **Does genre impact Disney's Movies Gross Income?**

Using the same dataset as the previous experiment, the " Disney Movies Total Gross" dataset, and the same column, the [inflation-adjusted gross] column, we want to identify which factors contribute to Disney's movie revenue stream. We consider the [genre] column as the treatment. Does a change in the genre impact the inflation-adjusted gross of a Disney movie? In other words, does a change in the genre also lead to a change in the mean inflation_adjusted gross? If there exists a mean inflation-adjusted gross that differs from the others, due to genre, we will be able to narrow down the genre/s. If there is data to support the fact that one genre consistently performs better on inflation-adjusted gross compared to another genre, it is only natural for Disney to take advantage of that fact to maximize profit. This experiment is even applicable to the opposite scenarios. If there isn't a difference in the average inflation-adjusted gross when changing the genre, Disney directors don't have to worry about what type of movies they should film. Knowing that a change in one genre to another will not lead to such a

[1] (*All Time Worldwide Box Office for Walt Disney Movies*, n.d.)

significant difference in the mean inflation-adjusted gross, they have free range in all movie genres, with little financial pushback.

3. Do viewers younger than 44 years old viewers 44 years old and older feel differently about of heroes and villains in Disney movies?

In the "SurveyData" dataset, the [age] column and the [preference] column stood out. We want to see how people from different age groups respond to Disney movies with the existence of heroes in comparison to Disney movies with the existence of villains. We segment the age column into two groups, Age group 1 consists of viewers younger than 44 years old, and Age group 2 consists of viewers that are 44 years old and older. The reason why we split the group by 44 is because 44 is the middle age among the samples aged 18-70. It is a splitting point. This question should be answered because Disney can identify whether or not it's worth it to promote movies with heroes and movies with villains differently to Age group 1 vs. Age group 2. It will help Disney promote the right content to the right target audience.

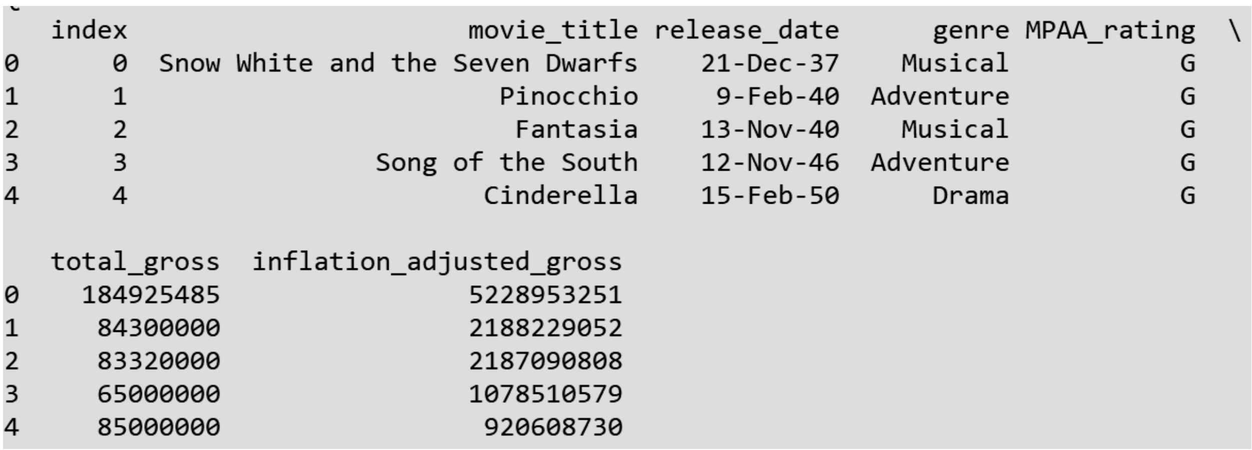
4. Is there a relationship between implementing a generic recommendation system and the number of Disney+ subscribers?

With digital streaming on the rise, Disney+ faces intense scrutiny from investors to surpass its competitors Hulu and Netflix to lead the market. Both competitors have already released an extensive recommendation algorithm on their platforms and Disney created a simple generic recommendation system in response to the pressure. The company must utilize experimental methods to determine whether this simple feature contributes to the subscriber counts. The goal of this experiment is to measure the impact of a new generic recommendation system on the number of disney+ subscribers. We utilized the "Disney+" dataset for this analysis. We specifically looked at column [DisneyPlus Recommendation System] as our independent variable and column [Disney+ Subscribers (measured in thousands)] as our dependent variable. The column [DisneyPlus Recommendation System] held values 'Yes' or 'No' which signifies whether each observation/region was introduced to this new recommendation system. Column [Disney+ Subscribers (in thousands)] measures the number of subscribers every month in thousands. Through this experiment, Disney can quantify the effectiveness of this simple recommendation system and account for factors that affect the number of subscribers which subsequently affect their top-line success. To examine a relationship we used linear regression and to identify a causal relationship we utilized DoWhy.

Section 3. Experiment and Analyses

A.

We will be using the “Disney Movies Total Gross” dataset, here is what our dataset looks like in Figure 1.

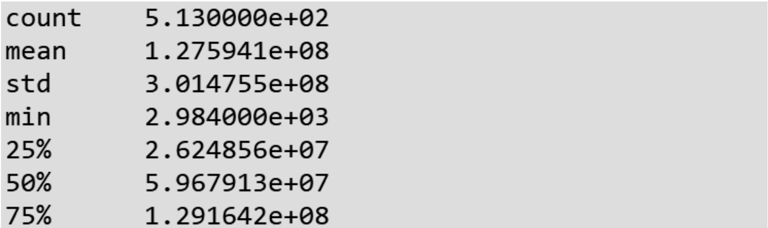


```
   index                     movie_title release_date      genre MPAA_rating  \
0      0  Snow White and the Seven Dwarfs    21-Dec-37    Musical           G
1      1                       Pinocchio     9-Feb-40  Adventure           G
2      2                        Fantasia    13-Nov-40    Musical           G
3      3               Song of the South    12-Nov-46  Adventure           G
4      4                      Cinderella    15-Feb-50      Drama           G

   total_gross  inflation_adjusted_gross
0    184925485                5228953251
1     84300000                2188229052
2     83320000                2187090808
3     65000000                1078510579
4     85000000                 920608730
```

[Figure 1: Dataset for “Disney Movies Total Gross” dataset]

Before starting the Hypothesis Test, we have to check that the data that we are using meets the two assumptions needed for this particular test: independent samples and large enough sample size. For independence, we used the Chi-Squared test on our column of interest, [inflated_adjusted_gross]. With a p-value of 1, we fail to reject the null so we cannot conclude that the samples are dependent, in other words, the samples are independent, meeting the first assumption. For the next one, we took the length of our samples and got 513 samples, which is far greater than the 30 samples required to apply the Central Limit Theorem, so our data is also large enough.

[Table 1: Descriptive Stats of Inflated Adjusted Gross Column]



```
count    5.130000e+02
mean     1.275941e+08
std      3.014755e+08
min      2.984000e+03
25%      2.624856e+07
50%      5.967913e+07
75%      1.291642e+08
```

The descriptive statistics for this column can be seen in Table 1. Table 1 says that the sample size is 513, so in this sample, we have 513 movies with genres. The mean inflated adjusted gross is about $127,594,055.45, with a median inflated adjusted gross of $59,679,131, for movies from 1937 to 2016. The standard deviation of the inflated adjusted gross is about $301,475,535.51. About 75% of the inflated adjusted gross is less than or equal to $129,164,207, and about 90% of them is less than or equal to $267,811,165.80. The minimum inflated adjusted gross is $2,984 and a maximum of $5,228,953,251. The range of the sample adjusted inflated adjusted gross data is $5,228,950,267. The 3rd quartile in the sample data is $129,164,207, and the 1st quartile in the sample data is $26,248,558, making the interquartile range $102,915,649. The variance of the inflated adjusted gross is 9.089e+16. The coefficient of variation is about 236.28 %.

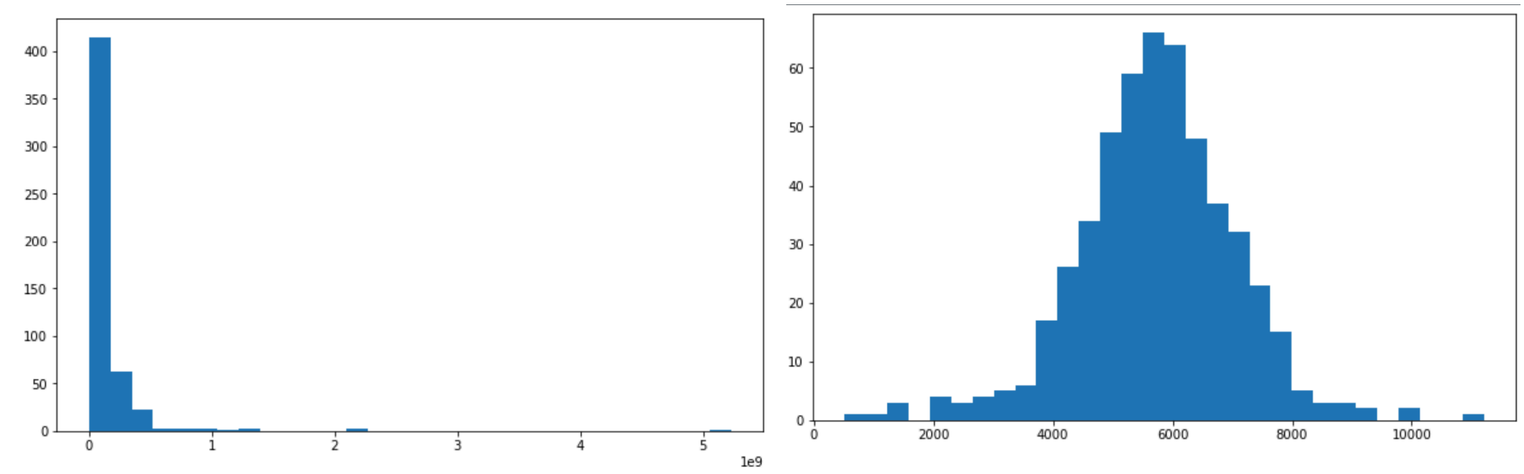
In this experiment, our null hypothesis is, the average Disney movie from 1937 to 2016 grossed less than or equal to $302,872,154. The alternative hypothesis is the average Disney movie from 1937 to 2016 grossed more than $302,872,154. We plug our samples, hypothesis value of $302,872,154, and the alternative=greater,' as we are doing an upper tail test, into the T-statistic function for one sample. We got a test statistic of -13.168437915752799 and a p-value of 1.

With a p-value of 1, which is greater than the significance level of 0.05, we cannot reject the null hypothesis. Therefore, we cannot conclude that the average Disney movie from 1937 to 2016 grossed more than $302,872,154. Unfortunately, this means that we cannot conclude that Disney movies from 1937 to 2016 hold up well in comparison to the average gross of Disney movies from 2017 to 2022. This result might be due to that Disney improved its filming or streaming technology in 2017. Therefore, they might be able to produce higher quality content and attract more subscribers compared to movies released from 1937-2016. To some degree, this is good news. This implies that Disney is moving forward in the right direction in terms of producing the 'right' movies that lead to more successful gross.

To ensure that we have enough samples, we choose a medium effect size and a small effect size to determine the minimum amount of samples we need. For a small effect size of .2, we would need a sample size of 199. For a medium effect size of .5, we would need a sample size of 34. The sample size of our data set is 513 which is greater than the desired sample size for small effect size and medium effect size indicating we have enough sample sizes.

B.

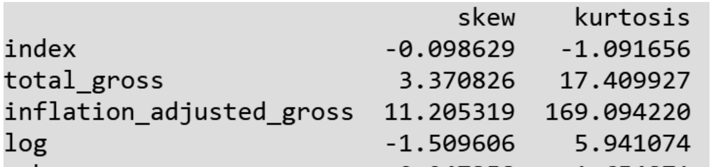
Using the same dataset as before, before beginning the One-Factor ANOVA experiment, we needed to check to see if the dataset meets the three assumptions: normal distribution, independent samples, and equal variances. Starting with normal distribution, we took the data in the [inflation_adjusted_gross] column, as this was the only quantitative column relevant to our experiments. Using the Shapiro-Wilk test, we got a p-value of 6.843213024560958e-40. At a significance level of 0.05, we can safely reject the null hypothesis as there is evidence to support the fact that the [inflation_adjusted_gross] column is not normally distributed. In fact, the distribution of the data was skewed to the right. One way to "fix" the skewed distribution is to normalize the data. We decided to use the log method and then cubed each term, after the log transformation. After applying the normalization to each of the 513 terms, we re-plotted the bar graph and saw that the data was more normalized.



[Figure 2: Distribution of samples before and after normalization]

As shown in Figure 2, on the left side we see that the distribution is skewed to the right with a long right tail. After we normalized the data we see that, visually, the data follow more of a bell curve. To check if it is actually normalized, we aggregate the ['inflation_adjusted_gross'] column checking for the skew and kurtosis.

[Table 2: Normalization Table]



```
                             skew   kurtosis
index                   -0.098629  -1.091656
total_gross              3.370826  17.409927
inflation_adjusted_gross 11.205319 169.094220
log                     -1.509606   5.941074
```

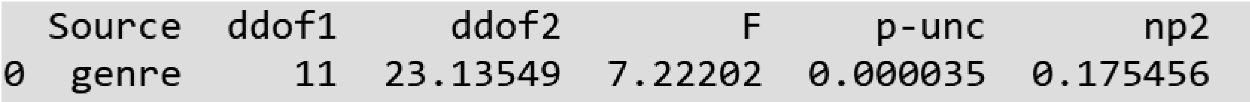
In Table 2, we see that the original skew was 11.205 meaning that the distribution was highly skewed, as it fell into the less than -1 and greater than 1 range. After we normalized that column, calling it [cube], we see that it is -0.247. Since it is between -0.5 and 0.5, we can conclude that the data is relatively symmetric. Therefore, we have passed the first assumption for normality.

For testing for variances, we used the Bartlett test. Since this test, tests for variances across the samples, we broke the [cube] column down by genre. Running a query and filtering by each of the twelve unique genres, we obtained their corresponding [cube] amounts, storing them to a different variable. Once we had the twelve variables, we put them in the test and it produced a p-value of 0.0113. With the same significant level of 0.05, we can safely reject the null hypothesis, which states that each group has the same variances. Therefore, we have evidence to support the fact that there exists at least one variance that is not equal to the rest. In short, the assumption of equal variances is false so we cannot run the typical f_oneway ANOVA test. We have to use Welch's Anova. Welch's Anova will account for the heterogeneity of the variances while the F_oneway assumes that equal variances are true.

To check the independence between each sample, we refer to the Chi-squared test. Using the data in the [cube] column, we ran the test and got a p-value of 1.0. Since 1.0 is greater than our significance level of 0.05, we fail to reject the null hypothesis that the samples are independent. Therefore, we can conclude that the samples are independent as we cannot conclude that the samples are dependent.

Now that all the assumptions have been checked, we can move on to the next step of the experiment, the Anova test. As previously mentioned, since the variances are not equal, we cannot use the typical ANOVA test, we need to use the Welch ANOVA test. This test accounts for instances when the data variances and the sample sizes are unequal between groups. It also has benefits such that even if the variances and sample sizes are equal, it will produce the same results. The Welch function takes 3 inputs: the dependent variable, the cube values, the groups by which we sort, the genre, and finally, the data frame, which we named Disney.

[Table 3: Welch Anova Results]



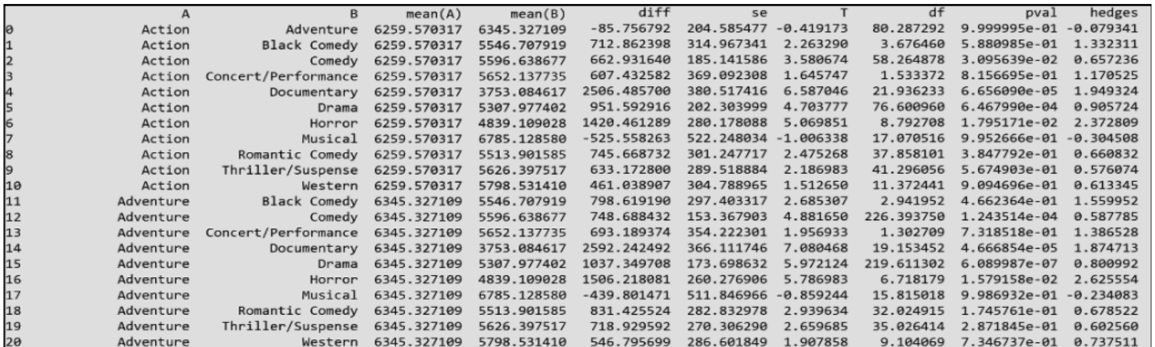
```
  Source  ddof1     ddof2        F     p-unc       np2
0  genre     11  23.13549  7.22202  0.000035  0.175456
```

As seen in Table 3, with a p-value of 0.000035, at a significance level of 0.05, we can reject the null hypothesis, meaning that there exists a difference in the average inflation_adjusted_gross as the genre changes.

Now that we know that there exists an average gross that differs from the other average gross values, as the genre changes, we can move on to identify which genre/s is causing this. The shortfalls of the Welch ANOVA and ANOVA tests in general is that it only tells us that there exists a difference but doesn't specify where that difference is. To figure out which genre is leading to a different average gross, we use the Tukey test. However, the original Tukey test relies on the fact that the sample sizes in each treatment group must be the same; since we have various sample sizes that range from 2(Concert/Performance) to 162(Comedy) per genre. The Games Howell Post- Hoc Tukey test, is a test that accounts for the varying variances.

‘

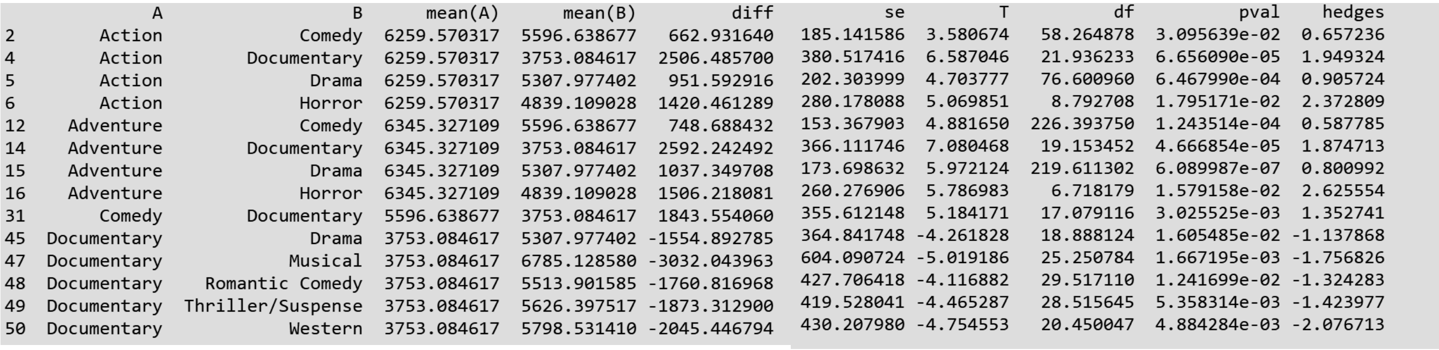
[Table 4: Games Howell Post- Hoc Tukey test Results]



```
                A                    B     mean(A)     mean(B)         diff          se         T          df          pval    hedges
0          Action            Adventure 6259.570317 6345.327109   -85.756792  204.585477 -0.419173   80.287292  9.999995e-01 -0.079341
1          Action         Black Comedy 6259.570317 5546.707919   712.862398  314.967341  2.263290    3.676460  5.880985e-01  1.332311
2          Action               Comedy 6259.570317 5596.638677   662.931640  185.141586  3.580674   58.264878  3.095639e-02  0.657236
3          Action  Concert/Performance 6259.570317 5652.137735   607.432582  369.092308  1.645747    1.533372  8.156695e-01  1.170525
4          Action          Documentary 6259.570317 3753.084617  2506.485700  380.517416  6.587046   21.936233  6.656090e-05  1.949324
5          Action                Drama 6259.570317 5307.977402   951.592916  202.303999  4.703777   76.600960  6.467990e-04  0.905724
6          Action               Horror 6259.570317 4839.109028  1420.461289  280.178088  5.069851    8.792708  1.795171e-02  2.372809
7          Action              Musical 6259.570317 6785.128580  -525.558263  522.248034 -1.006338   17.070516  9.952666e-01 -0.304508
8          Action       Romantic Comedy 6259.570317 5513.901585   745.668732  301.247717  2.475268   37.858101  3.847792e-01  0.660832
9          Action    Thriller/Suspense 6259.570317 5626.397517   633.172800  289.518884  2.186983   41.296056  5.674903e-01  0.576074
10         Action              Western 6259.570317 5798.531410   461.038907  304.788965  1.512650   11.372441  9.094696e-01  0.613345
11      Adventure         Black Comedy 6345.327109 5546.707919   798.619190  297.403317  2.685307    2.941952  4.662364e-01  1.559952
12      Adventure               Comedy 6345.327109 5596.638677   748.688432  153.367903  4.881650  226.393750  1.243514e-04  0.587785
13      Adventure  Concert/Performance 6345.327109 5652.137735   693.189374  354.222301  1.956933    1.302709  7.318518e-01  1.386528
14      Adventure          Documentary 6345.327109 3753.084617  2592.242492  366.111746  7.080468   19.153452  4.666854e-05  1.874713
15      Adventure                Drama 6345.327109 5307.977402  1037.349708  173.698632  5.972124  219.611302  6.089987e-07  0.800992
16      Adventure               Horror 6345.327109 4839.109028  1506.218081  260.276906  5.786983    6.718179  1.579158e-02  2.625554
17      Adventure              Musical 6345.327109 6785.128580  -439.801471  511.846966 -0.859244   15.815018  9.986932e-01 -0.234083
18      Adventure      Romantic Comedy 6345.327109 5513.901585   831.425524  282.832978  2.939634   32.024915  1.745761e-01  0.678522
19      Adventure    Thriller/Suspense 6345.327109 5626.397517   718.929592  270.306290  2.659685   35.026414  2.871845e-01  0.602560
20      Adventure              Western 6345.327109 5798.531410   546.795699  286.601849  1.907858    9.104069  7.346737e-01  0.737511
```

As you can see from Table 4, we need to take a look at the p-value column to determine which pair or pairs of genres produce a significant difference in mean adjusted inflation gross. At a significance level of 0.05, we can conclude that there are only 14 significant rows.

[**Table 5:** Games Howell Post- Hoc Tukey Significant Rows]



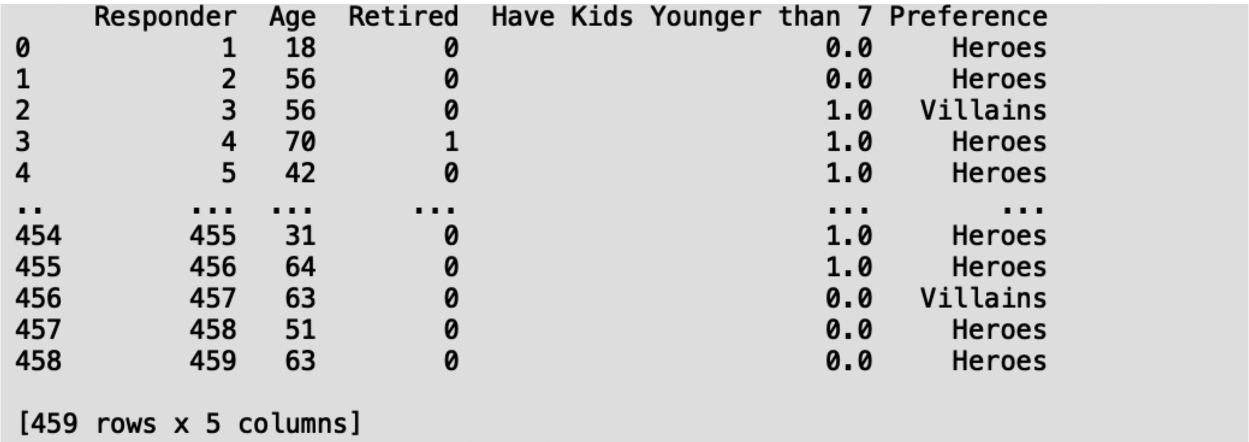
```
              A                 B     mean(A)     mean(B)         diff          se         T          df          pval    hedges
2        Action            Comedy 6259.570317 5596.638677   662.931640  185.141586  3.580674   58.264878  3.095639e-02  0.657236
4        Action       Documentary 6259.570317 3753.084617  2506.485700  380.517416  6.587046   21.936233  6.656090e-05  1.949324
5        Action             Drama 6259.570317 5307.977402   951.592916  202.303999  4.703777   76.600960  6.467990e-04  0.905724
6        Action            Horror 6259.570317 4839.109028  1420.461289  280.178088  5.069851    8.792708  1.795171e-02  2.372809
12    Adventure            Comedy 6345.327109 5596.638677   748.688432  153.367903  4.881650  226.393750  1.243514e-04  0.587785
14    Adventure       Documentary 6345.327109 3753.084617  2592.242492  366.111746  7.080468   19.153452  4.666854e-05  1.874713
15    Adventure             Drama 6345.327109 5307.977402  1037.349708  173.698632  5.972124  219.611302  6.089987e-07  0.800992
16    Adventure            Horror 6345.327109 4839.109028  1506.218081  260.276906  5.786983    6.718179  1.579158e-02  2.625554
31       Comedy       Documentary 5596.638677 3753.084617  1843.554060  355.612148  5.184171   17.079116  3.025525e-03  1.352741
45  Documentary             Drama 3753.084617 5307.977402 -1554.892785  364.841748 -4.261828   18.888124  1.605485e-02 -1.137868
47  Documentary           Musical 3753.084617 6785.128580 -3032.043963  604.090724 -5.019186   25.250784  1.667195e-03 -1.756826
48  Documentary   Romantic Comedy 3753.084617 5513.901585 -1760.816968  427.706418 -4.116882   29.517110  1.241699e-02 -1.324283
49  Documentary Thriller/Suspense 3753.084617 5626.397517 -1873.312900  419.528041 -4.465287   28.515645  5.358314e-03 -1.423977
50  Documentary           Western 3753.084617 5798.531410 -2045.446794  430.207980 -4.754553   20.450047  4.884284e-03 -2.076713
```

As shown in Table 5, we extracted the 14 significant rows to compare the mean differences of each genre pairing, to figure out which genre pairs perform the best for gross. We were able to see that Action performs better than Comedy, Documentary, Drama, and Horror. Adventure outperforms Comedy, Documentary, Drama, and Horror. Comedy and Drama outperform Documentaries. Musical, Romantic Comedy, Thriller/Suspense, and Western outperform Documentaries. When comparing the "winners" of these pairings all of them had insignificant results as they were all larger than the significance level of 0.05. Therefore the best genre pairing would be: (Action and Adventure), (Action, Musical), (Action, Romantic Comedy), (Action, Thriller/Suspense), (Action, Western), (Adventure, Musical), (Adventure, Romantic Comedy), (Adventure, Thriller/Suspense), and (Adventure, Western). This result might be due to Disney putting heavy investment in these six successful genres, making more of these movies instead of the others. It could also be because the characters or actors in these films were more popular from 1937 to 2016. There could be a multitude of other reasons why Disney achieved the most successful gross from these six genres. Another pairing of these genres will not make much of a significant difference in gross. Disney's film directors can now dig deeper into the success factors behind these six genres and replicate them in future productions.

C. Age, Heroes, and Villians (Heterogeneity Test):

We want to see if viewers who are younger than 44 years old and people who are 44 years old and older feel differently about the existence of heroes and villains in Disney movies. In order to find out, we use the heterogeneity test where we made four groups: AG1andHero, AG1andVillain, AG2andHero, and AG2andVillain. The dataset we use is SurveyData as presented in Figure 3.

The null hypothesis here is that viewers who are younger than 44 years old and viewers who are 44 years old and older feel the same about the existence of heroes and villains in Disney movies. The alternative hypothesis is that these two groups of viewers feel differently about the existence of heroes and villains in Disney movies.



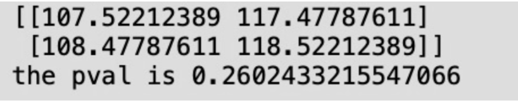
```
     Responder  Age  Retired  Have Kids Younger than 7 Preference
0            1   18        0                       0.0     Heroes
1            2   56        0                       0.0     Heroes
2            3   56        0                       1.0   Villains
3            4   70        1                       1.0     Heroes
4            5   42        0                       1.0     Heroes
..         ...  ...      ...                       ...        ...
454        455   31        0                       1.0     Heroes
455        456   64        0                       1.0     Heroes
456        457   63        0                       0.0   Villains
457        458   51        0                       0.0     Heroes
458        459   63        0                       0.0     Heroes

[459 rows x 5 columns]
```

[Figure 3: SurveyData]

There are a total of 459 responders and 4 variables. “Age” represents the age of the viewer/responder. “retired (0)” represents the people who have not retired, and “retired (1)” represents people who retired. “Having Kids Younger than 7 (0)” means the viewer/responder doesn't have kids that are younger than 7 years old, whereas “having Kids Younger than 7 (1)” means the viewer/responder who has kids that are younger than 7 years old. “Preference (heroes)” represents the viewer who prefers the existence of heroes in the Disney movie. “Preference (Villains)” represents the viewer who prefers the existence of villains in the Disney movie. For this analysis, we are only using the “age” and “preference” columns.

Analysis:

After running the heterogeneity test, we found that the expected cell count for each of the four groups is as shown in Figure 4. Since all the cell count is above 5, the result for this test is reliable.



```
[[107.52212389 117.47787611]
 [108.47787611 118.52212389]]
the pval is 0.2602433215547066
```

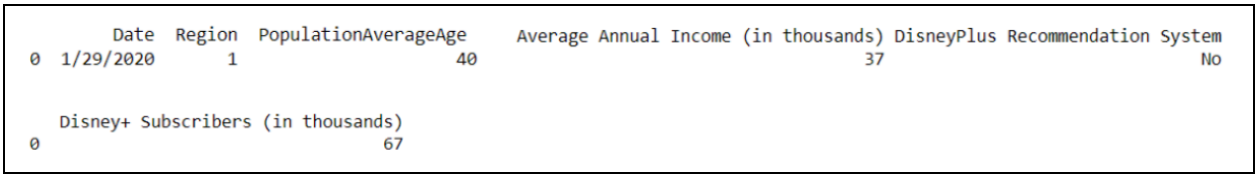
[Figure 4: Expected Cell Count Heterogeneity Test]

The p-value of this heterogeneity test is 0.260, as shown in Figure 4. It is greater than the significance level of 0.1. We cannot reject the null hypothesis. We cannot conclude that viewers younger than 44 years old and viewers who are 44 years old and older feel differently about the existence of heroes and villains in Disney movies. Knowing this, Disney does not have to promote heroes and villains differently to audiences that are younger than 44 and audiences that are 44 years old or older than 44 years old. It can save some extra advertising costs when Disney is trying to target different segments of the audience.

This result from this test might be because the audience in this dataset can be indifferent to whether the movies are about heroes or villains. This conclusion may be a result of the variable "Age" in our dataset. Our dataset only contains adults with the youngest individual being 18 years old and the oldest being 70. This may have affected our results because children/teens (<18 years old) were not included in this variable. If they were included, we might have seen a difference as many children may have a preference for heroes due to their idealistic naivety. However, since this data is not available to us, we can not conclude that individuals in each segment feel differently about the existence of heroes and villains in Disney movies.

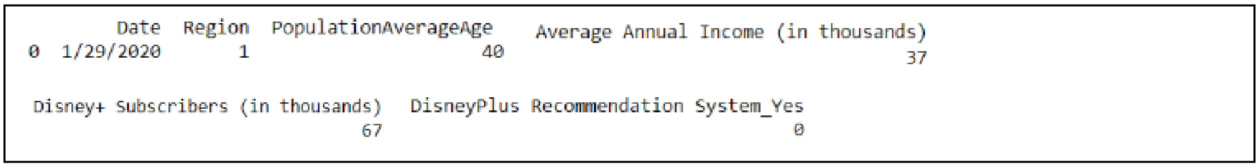
D. Income, Disney+ Recommendation System, and Subscriber Count (Regression and ATE)

We ran a linear regression to test the relationship between the existence of the recommendation system and the Disney+ subscriber count. For the experiment, we created a control group and a treatment group to measure the impact of this new recommendation system. We converted our treatment variable [DisneyPlus Recommendation System] into a dummy which enables us to run a single regression that compares the two groups rather than running two separate regressions. Dummy variables act as on-and-off switches for our treatment variables. This resulted in the creation of a new column: [DisneyPlus Recommendation System_Yes]. In this new column, a "1" indicates regions that were exposed to the new recommendation system, and a "0" are regions that were not given this new feature as shown in Figure 6 compared to the previous dataset in Figure 5.



```
        Date  Region  PopulationAverageAge    Average Annual Income (in thousands) DisneyPlus Recommendation System
0  1/29/2020       1                    40                                      37                               No

   Disney+ Subscribers (in thousands)
0                                  67
```

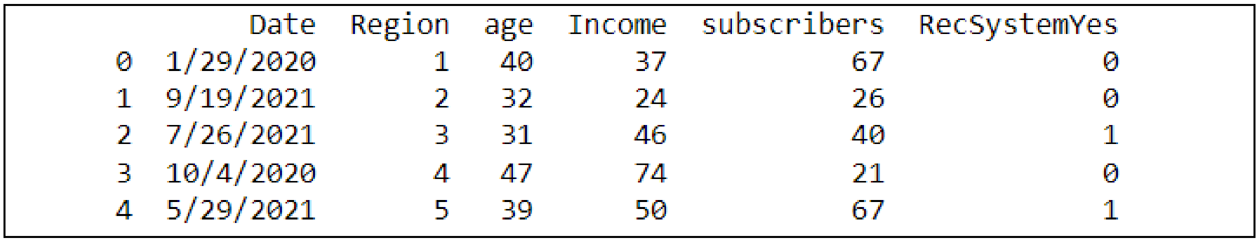
[Figure 5: Dataset "Disney+" before creating a dummy variable]



```
        Date  Region  PopulationAverageAge    Average Annual Income (in thousands)
0  1/29/2020       1                    40                                      37

 Disney+ Subscribers (in thousands)  DisneyPlus Recommendation System_Yes
                                 67                                     0
```

[Figure 6: Dataset "Disney+" after creating the dummy variable]

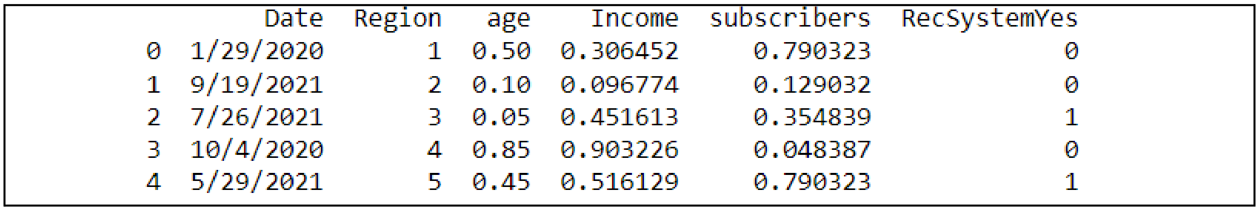
After creating the dummy, we renamed our data columns for convenience. Some of the data column names had spaces such as [PopulationAverageAge] and were inconsistent with spelling [Disney+ Subscribers] and [DisneyPlus Recommendation System_Yes]. Our renamed dataset appears below in Figure 7:



```
          Date  Region  age  Income  subscribers  RecSystemYes
0    1/29/2020       1   40      37           67             0
1    9/19/2021       2   32      24           26             0
2    7/26/2021       3   31      46           40             1
3    10/4/2020       4   47      74           21             0
4    5/29/2021       5   39      50           67             1
```

[Figure 7: Dataset "Disney+" after renaming the columns]

Prior to running the regression, we normalized our dataset because we had multiple numerical variables measured in different units. In this dataset, [PopulationAverageAge] is measured on a scale from 1-100,   [Average Annual Income] is measured in thousands, with a higher number meaning a greater annual income, and [Disney+ Subscribers] is measured in thousands with a higher number suggesting a greater number of people being subscribed to Disney's streaming services. We must normalize our data to even out the discrepancy in the units measured and generate reliable confidence intervals and prediction intervals. Our new dataset is highlighted below in Figure 8:

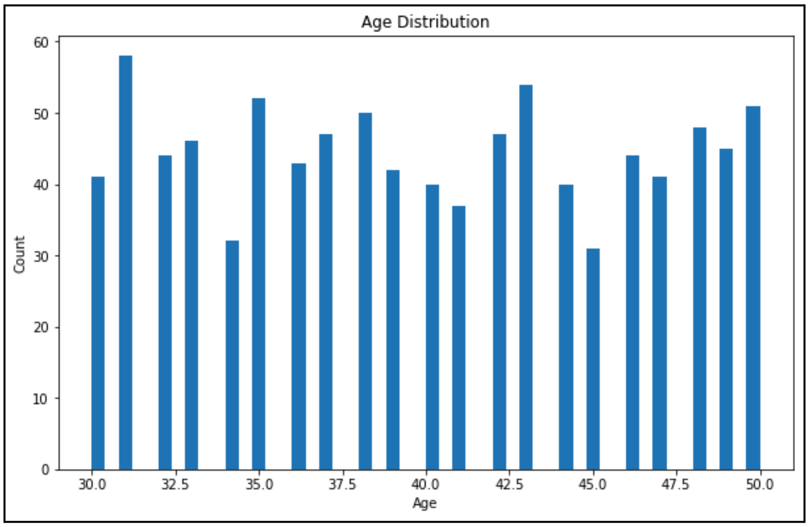


```
          Date  Region   age    Income  subscribers  RecSystemYes
0  1/29/2020       1  0.50  0.306452     0.790323             0
1  9/19/2021       2  0.10  0.096774     0.129032             0
2  7/26/2021       3  0.05  0.451613     0.354839             1
3  10/4/2020       4  0.85  0.903226     0.048387             0
4  5/29/2021       5  0.45  0.516129     0.790323             1
```

[Figure 8: Dataset “Disney+” after normalization of the data]

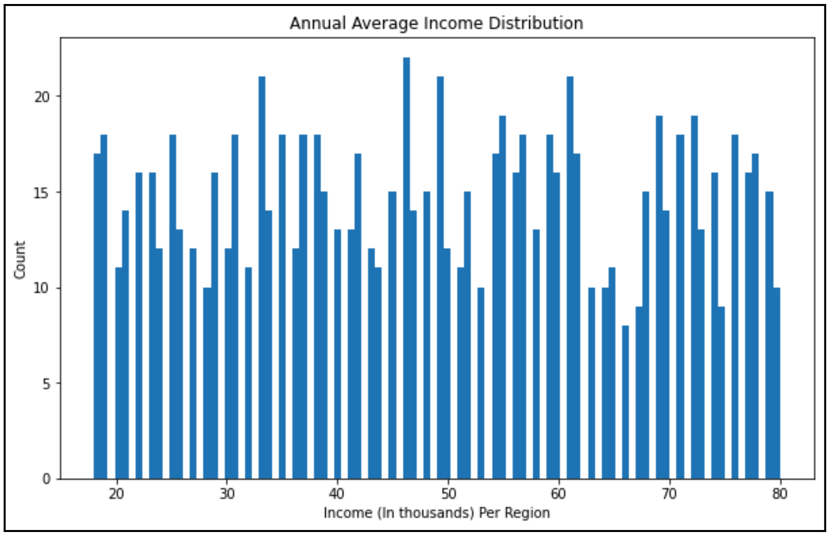
Data plot distribution after normalization:

Our numerical variable [Age] shows a normal distribution after the conversion in Figure 9.



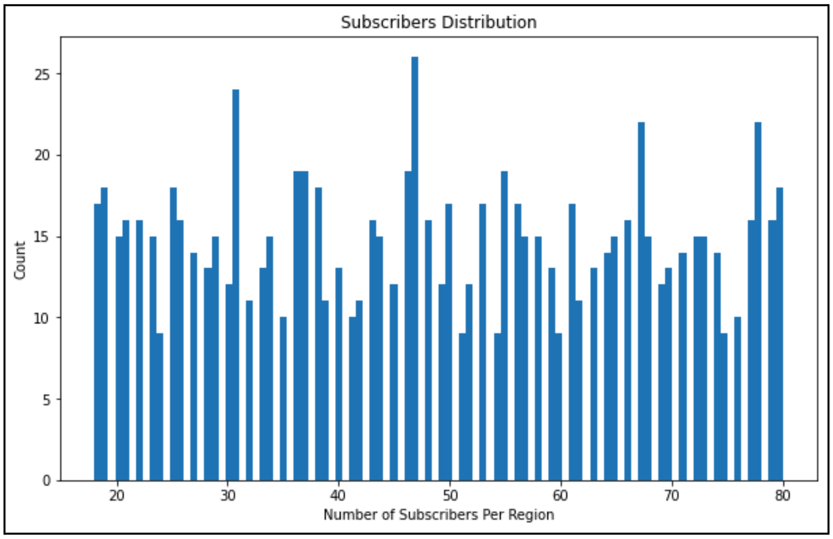
[Figure 9: Normal distribution of age]

Our numerical variable [Income] shows a normal distribution after the conversion in Figure 10.



[Figure 10: Normal distribution of Income]

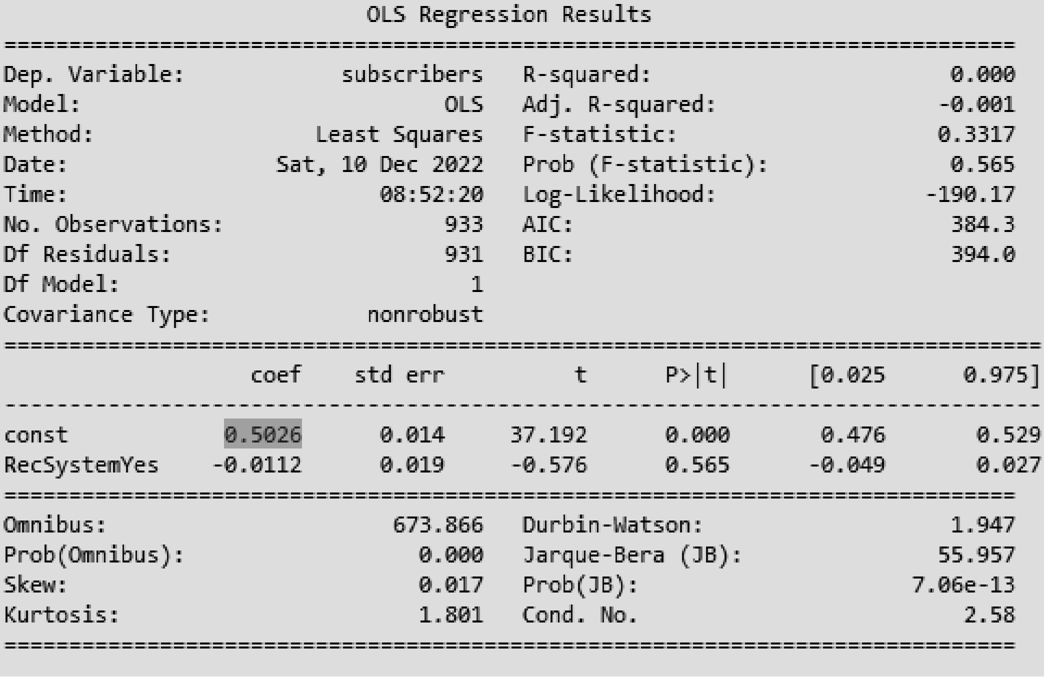
Our numerical variable [Subscribers] shows a normal distribution after the conversion in Figure 11.



[Figure 11: Normal distribution of Subscribers]

The control group consisted of regions that had responded '0' under the [DisneyPlus Recommendation System_Yes] column while the treatment group had regions that responded '1.' Our results can be found below:

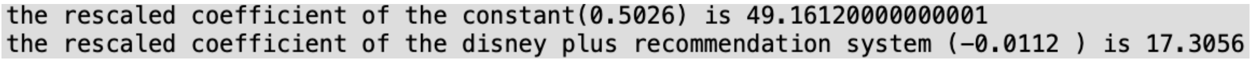
[Table 6: Recommendation System on the Number of Subscribers Regression]



```
                            OLS Regression Results
==============================================================================
Dep. Variable:            subscribers   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                 -0.001
Method:                 Least Squares   F-statistic:                    0.3317
Date:                Sat, 10 Dec 2022   Prob (F-statistic):              0.565
Time:                        08:52:20   Log-Likelihood:                -190.17
No. Observations:                 933   AIC:                             384.3
Df Residuals:                     931   BIC:                             394.0
Df Model:                           1
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const            0.5026      0.014     37.192      0.000       0.476       0.529
RecSystemYes    -0.0112      0.019     -0.576      0.565      -0.049       0.027
==============================================================================
Omnibus:                      673.866   Durbin-Watson:                   1.947
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               55.957
Skew:                           0.017   Prob(JB):                     7.06e-13
Kurtosis:                       1.801   Cond. No.                         2.58
==============================================================================
```

Table 6 can be formatted into a regression equation: **Disney+ Subscribers (in thousands) = 0.5026 - 0.0112 *DisneyPlus Recommendation System_Yes**

The mean number of Disney+ subscribers in our control group is 0.5026. Due to the normalization of our data, the coefficients are within the range (0,1) thus we rescaled each coefficient back to its true range to better interpret the regression model.



```
the rescaled coefficient of the constant(0.5026) is 49.16120000000001
the rescaled coefficient of the disney plus recommendation system (-0.0112 ) is 17.3056
```

[Figure 12: Rescaled coefficients]

This means that on average, the number of subscribers in regions without the recommendation system is 49.1612 or approximately 49,160 subscribers. The difference between the treatment and control groups is the average treatment effect (ATE). Our ATE in this regression is -0.0112. This means that introducing the new recommendation system decreased the number of subscribers by - 0.0112 or -17,305 (scaled within the true range) and the number of subscribers in regions with this new feature is roughly 48.46669 as part of our experimental group. With Walt Disney Studios charging $7.99/month for Disney+

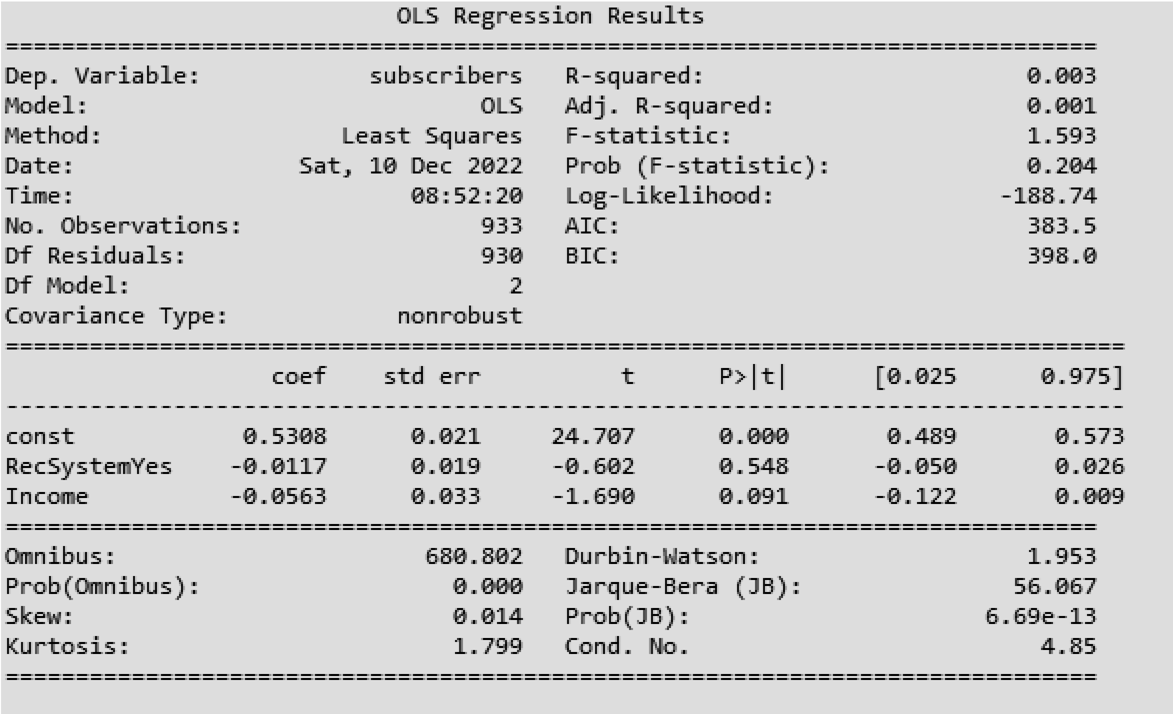
subscriptions, this could mean that the company is losing on average $96,012,226 per region with this new feature.[2] The coefficient of -0.0112 suggests that there is a negative relationship between an increase in subscriber count and the extra benefit of the recommendation system.

It is important to note that the p-value of this regression is 0.565 which is above the alpha value of 0.05. This indicates that the results are not statistically significant. The standard error for the independent variable is 0.019 which shows how much the mean number of subscribers in regions in our sample deviates from the actual mean number of subscribers in the population. This number is very close to the desired value of 0 (no deviation from the population) which implies if the experiment was replicated, it is very likely we would get similar results. The r-squared value is 0.000, which means that almost 0% of the variance in the Disney+ subscriber count can be explained by our independent variable (the recommendation system). The low r-squared value means a high variance around the regression line and that there may be other factors influencing the region's decision to subscribe to Disney+.

To account for the variance in our first regression, we added a covariate to raise the r-squared. We selected the column [Average Annual Income (in thousands)] to be our covariate because income impacts the ability of households to purchase discretionary goods such as streaming subscriptions. As income increases, households have the cash flow to subscribe to the Disney+ platform. Our new regression measures the efficacy of the recommendation system on the number of Disney+ subscribers per region controlling for the region's average annual income. Our results can be found below:

[2] (*Disney+ | Stream Disney, Marvel, Pixar, Star Wars, National Geographic, and More*, n.d.)

[Table 7: Recommendation System on the Number of Subscribers Regression Controlling for Income]



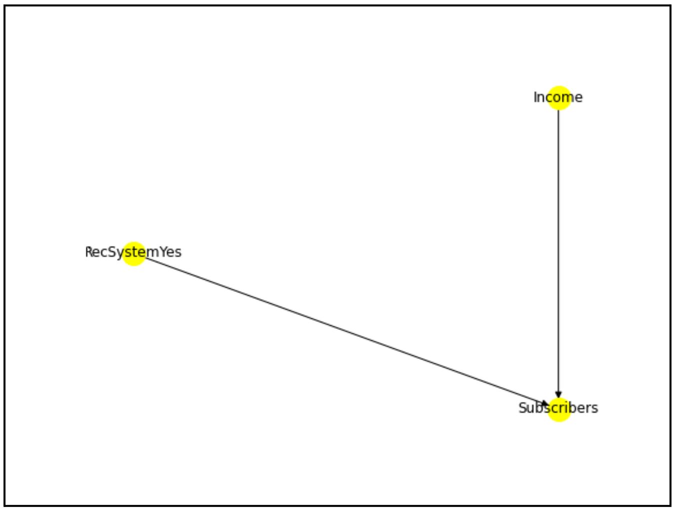
```
                            OLS Regression Results
==============================================================================
Dep. Variable:            subscribers   R-squared:                       0.003
Model:                            OLS   Adj. R-squared:                  0.001
Method:                 Least Squares   F-statistic:                     1.593
Date:                Sat, 10 Dec 2022   Prob (F-statistic):              0.204
Time:                        08:52:20   Log-Likelihood:                -188.74
No. Observations:                 933   AIC:                             383.5
Df Residuals:                     930   BIC:                             398.0
Df Model:                           2
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const            0.5308      0.021     24.707      0.000       0.489       0.573
RecSystemYes    -0.0117      0.019     -0.602      0.548      -0.050       0.026
Income          -0.0563      0.033     -1.690      0.091      -0.122       0.009
================================================================================
Omnibus:                      680.802   Durbin-Watson:                   1.953
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               56.067
Skew:                           0.014   Prob(JB):                     6.69e-13
Kurtosis:                       1.799   Cond. No.                         4.85
==============================================================================
```

Table 7 can be formatted into a regression equation: **Disney+ Subscribers (in thousands) = 0.5308 - 0.0117 *DisneyPlus Recommendation System_Yes - 0.0563*Average Annual Income (in thousands).**

The mean number of Disney+ subscribers in our new control group is 0.5308. This means that on average, the number of subscribers in regions controlled for the income and without the recommendation system is 50.909 or approximately 50909 subscribers when scaled to its true range. Our ATE in this regression is - 0.0117. This means that introducing the new recommendation system decreased the number of subscribers by - 0.0117 or -17274.6 (true range) when controlling for the average annual income and the number of subscribers in regions with this new feature is roughly 50.1842 or 50,184.2. The coefficient - 0.0117 signifies that there is a negative relationship between an increase in subscriber count, annual average income, and the extra benefit of the recommendation system. Our ATE decreased from -0.0112 to - 0.0117 which signifies that when the annual average income across regions is the same, the recommendation system still produces more harm than benefit for the company though less harm than when the average annual income is different. It is important to note that when controlling for the annual income, it decreased the number of subscription cancellations from approximately 17305 to 17274 (a 31 subscriber difference per region). This implies that there may be other factors outside of both the recommendation systems and income levels affecting a person's choice to subscribe. Since adding the covariate income reduced the unsubscription amount, it signals to Disney that it is too early to give up on the recommendation system as there are still other variables to evaluate.

It is important to note that the p-value of this regression is 0.204. The new p-value is lower than the previous p-value of 0.565 yet it is still above the desired alpha value of 0.05. This indicates that the new results are not statistically significant. The standard error for the independent variable is 0.019 compared to the 0.019 from the first regression. Much like the prior regression, an std error lower than 0.05 indicates that our regression is representative of the entire population and that replicating this experiment would yield the same results. The adjusted r-squared value is 0.003, which means that 0.3% of the variance in the Disney+ subscriber count can be explained by our variables [DisneyPlus Recommendation System_Yes] and [Average Annual Income (in thousands)]. The low r-squared value means a high variance around the regression line and we need to explore other factors that can influence a region's choice in subscribing to Disney+.

We plotted a causal diagram to show the relationship between our three variables.

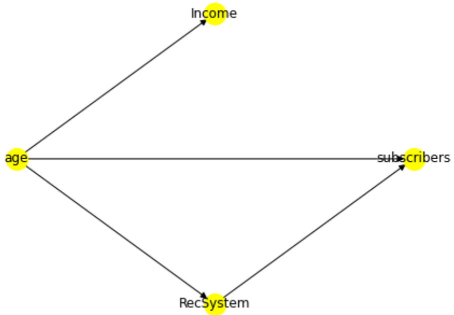


[Figure 13: Casual Diagram with Income, RecSystemYes, and Subscribers]

Figure 13 shows the relationship between the three variables: [DisneyPlus Recommendation System_Yes], [Disney+ Subscribers (in thousands)], and [Average Annual Income (in thousands)]. We have renamed column [Disney+ Subscribers (in thousands)] to [Subscribers], column [DisneyPlus Recommendation System_Yes] to [RecSystemYes], and column [Average Annual Income (in thousands)] to [Income]. This diagram shows that our covariate 'Income' is not impacted by our treatment variable 'RecSystemYes.' Annual average income is not influenced by whether or not the region receives the recommendation feature and therefore making it a good control.

E. Disney recommendation system's causal relationship with subscribers amount (DoWhy):

In this case, we use dataset 3 (Disneyplus) to find out the factors that are impacting the response variable: Disney+ subscribers. The treatment we use is whether or not a Disney+ recommendation system is implemented. The research question we are trying to answer is: does implementing the Disney recommendation system cause an increase in Disney+ subscribers amount? Figure 14 below is the causal diagram that displays the relationship among the variables.



[Figure 14: Causal relationship with Income, Age, RecSystem, and subscribers]

After running backdoor linear regression, we found that the [Figure 14: Causal relationship with Income, Age, RecSystem, and subscribers]

It means that if the Disney recommendation system is implemented, it will cause Disney+ subscribers to drop by 695. After running the propensity score matching method, we found that the causal effect is 2.495. It means that if the Disney recommendation system is implemented, it will cause Disney+ subscribers to increase by 2495. The difference between these two methods is 3190 subscribers.

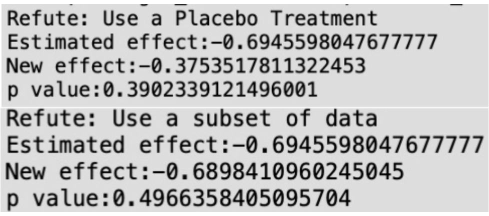
For the backdoor linear regression, as shown in Figure 15, when we use a subset of data to test the validity of our assumption, it shows that the new effect is -0.690. It is close to the original estimate of -0.695. The p-value is 0.500, which is greater than the 0.05 significance level. It means that our initial assumption is valid. When we use a placebo treatment to test the validity, it shows that the new effect is 0.375. This number is close to 0. At the same time, the p-value is 0.390, which is greater than the 0.05

significance level. It means that our initial assumption is valid. Both of the refuting methods show the validity of our assumption and estimate using the backdoor linear regression.

For the propensity score matching, as shown in Figure 16, when we use a subset of data to test the validity of our assumption, it shows that the new effect is 0.736. The p-value is 0.255, which is greater than the 0.05 significance level. It means that our initial assumption is valid. When we use the placebo treatment to test the validity, it shows that the new effect is - 0.274, which is close to 0. The p-value is 0.459. It is greater than a 0.05 significance level. It means that our initial assumption is valid. Both of the refuting methods show the validity of our assumption and estimate using the propensity score matching regression.

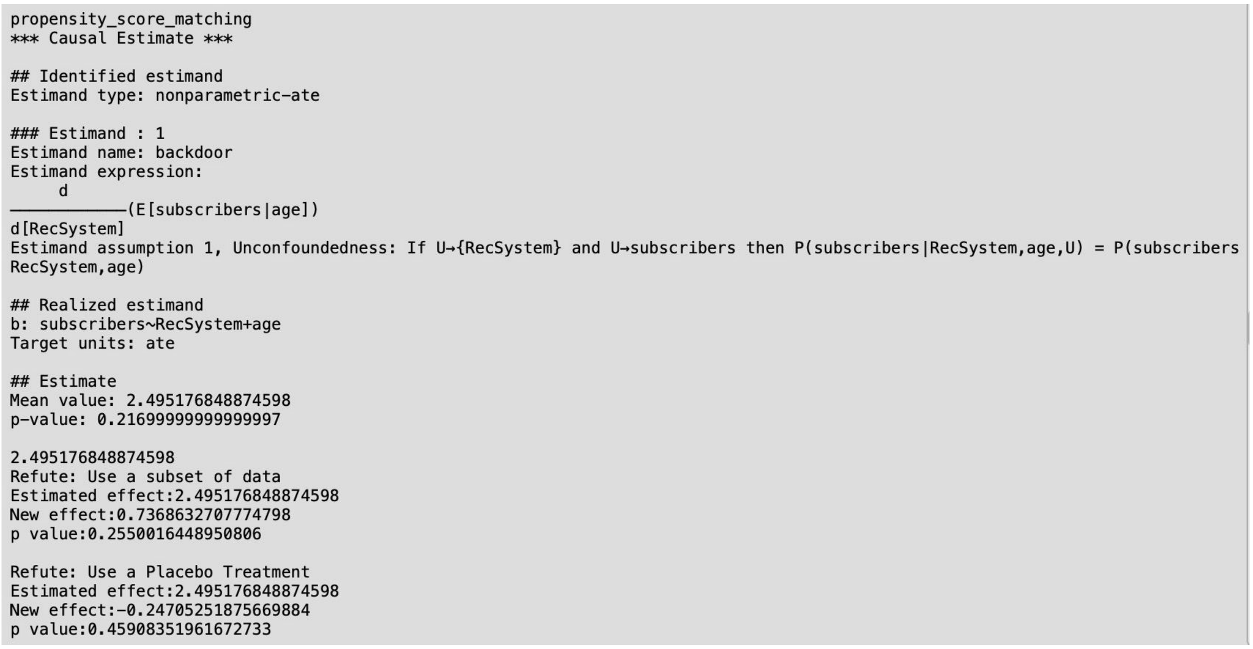
Since both estimates say that our initial assumption is valid, we can conclude that there is a causal relationship between the recommendation system and the number of subscribers of Disney+. In the end, we decided to use the casual estimate from the propensity score matching method. This is because the data is not randomized, we do not have access to the pre-treatment data, and not all the relationships we are analyzing are linear. Therefore, this indicates that having a recommendation system increases the number of Disney+ subscribers by 2,495. This result might be because the recommendation system successfully identifies subscribers' preferences in shows. The subscribers are more likely to stay subscribed to Disney+ as a result of that. Because of this, the existing subscribers also might be more willing to recommend Disney+ to potential subscribers.

This could be important to Disney because now Disney+ can invest more in its recommendation system and perhaps recommendation system in other regions.



```
Refute: Use a Placebo Treatment
Estimated effect:-0.6945598047677777
New effect:-0.3753517811322453
p value:0.3902339121496001
Refute: Use a subset of data
Estimated effect:-0.6945598047677777
New effect:-0.6898410960245045
p value:0.4966358405095704
```

[Figure 15: Backdoor Regression Result]



```
propensity_score_matching
*** Causal Estimate ***

## Identified estimand
Estimand type: nonparametric-ate

### Estimand : 1
Estimand name: backdoor
Estimand expression:
     d
────────────(E[subscribers|age])
d[RecSystem]
Estimand assumption 1, Unconfoundedness: If U→{RecSystem} and U→subscribers then P(subscribers|RecSystem,age,U) = P(subscribers
RecSystem,age)

## Realized estimand
b: subscribers~RecSystem+age
Target units: ate

## Estimate
Mean value: 2.495176848874598
p-value: 0.21699999999999997

2.495176848874598
Refute: Use a subset of data
Estimated effect:2.495176848874598
New effect:0.7368632707774798
p value:0.2550016448950806

Refute: Use a Placebo Treatment
Estimated effect:2.495176848874598
New effect:-0.24705251875669884
p value:0.45908351961672733
```

[Figure 16: Propensity Score Matching Result]

Section 4. Conclusions and Discussion

Since, we cannot reject the null hypothesis, in our hypothesis testing experiment, we can't conclude that the average Disney movie from 1937 to 2016 grossed more than $302,872,154. Implying that their performance doesn't hold up as well as the 2017 to 2022 movies' gross. There are many factors that may come into play when determining the reason for this result. One interesting thing to look at is the economic trends from 1937 to 2016. Even though we account for the inflation over the years, disposable income back then may be different than now. That may explain why some people were unable to watch these movies back then, impacting the gross. One potential limitation of our experience is that we only considered the Domestic gross, not the international gross as well. This may be an issue in the future if Disney, an international corporation, tries to implement this experiment with an international sample.

Though we were able to conclude that Action, Adventure, Musical, Romantic Comedy, Thriller/ Suspense, and Western were the genres that perform better in terms of inflation-adjusted gross. There is one drawback to our experiment, we were unable to narrow it down to fewer genres. So, our conclusion is very broad. One way to address this is to add another treatment. It would also be interesting to see if the ratings of each movie may also play a role in the gross and genre. This would mean we need a two-factor Anova test, as we consider two treatments: genre and MPAA_rating. This would hopefully give Disney directors a more defined set of parameters to consider when filming. In other words, they would theoretically be able to see which genre and rating they should focus on to get the best outcome of gross.

After running the heterogeneity test, we cannot conclude that viewers younger than 44 years old and viewers who are 44 years old and older feel differently about the existence of heroes and villains in Disney movies. The extension of our work would be to segment the age group differently, for example, we might want to see if there is any difference between viewers younger than 30 years old and viewers that are 30 years old and older when responding to movies with heroes vs movies with villains. The limitation of this test is that this sample data only includes people who are 18-70. There are no responders that are younger than 18. Because of this, it might influence the results when we group them by viewers who are younger than 44 years old. Something else that is interesting to analyze is using other variables in the dataset "SurveyData". For example, we can try to see if viewers who have kids younger than 7 vs viewers who do not have kids younger than 7 feel differently about movies with heroes vs. movies with villains. To accomplish this, we can run the heterogeneity test again.

After running a regression to test the relationship between Disney+'s recommendation system and the number of subscribers, the new feature proves to be a bust. Rather than increasing the number of subscribers, it showed a negative relationship between the two variables. In regions where the recommendation system was introduced, the number of subscribers decreased. The p-value is also significantly higher than the desired value of 0.05 meaning that the regressions were not statistically significant. To conduct further research on the effectiveness of a recommendation system, Disney could collect the number of click rates per feature offered on the app. The recommendation feature could prove to be the most visited feature on the app despite the general decrease in subscribers. This would indicate an overall issue with the subscription service. Outside of this recommendation system, Disney can also test the price sensitivity of its subscribers by collecting data on how users feel about a +$1/-$1 change in the monthly subscription price. There is a multitude of factors that influence a user's decision to subscribe to a streaming service and Disney should take advantage of experiments to better improve its system.

The results of the DoWhy test indicate that implementing Disney+'s recommendation system increases Disney+'s subscriber count by 2,495. We recommend that Disney+ look into implementing a recommendation system in more regions and perform ATE or DiD to collect more data on the effect of the recommendation system on subscriber count by region. If the recommendation system is consistently shown to improve the number of Disney+ subscribers, Disney+ should consider implementing the recommendation system in all regions. A potential limitation to running propensity score matching is limiting the remaining confounding variable that may still be present and therefore lead to biased results.

An interesting question to analyze in the future would be to see if there is a causal relationship between Disney+'s recommendation system and the income of Disney+.

Walt Disney Studios' mission "to entertain, inform, and inspire people around the globe through the power of unparalleled storytelling" left an imprint on every person's childhood. Though making movies is a form of art, it is undeniable that every company is driven by bottom-line returns. Within this project, we attempted to find factors that impact Disney's two highest revenue streams in hopes of helping the company dominate the market. In the end, we have reached the conclusion that Disney+ should continue the recommendation system and consider implementing a recommendation system in other regions. Since genres action, adventure, musical, romantic comedy, and western, thriller/suspense performed the best out of 12 genres in the "Disney Movies Total Gross" dataset, Disney should produce the majority of its movies under these 6 genres. With the heterogeneity test proving people under and above the age of 44 are indifferent toward villains/heroes, Disney can focus on promoting its movies to the population as a whole rather than targeting each segment differently. All of these techniques unveiled the result of Disney's current production strategies and imply a potential for improvement using our findings. In the future, we may see Disney pivot to a more complex recommendation algorithm better suited for its global audience.

Reference Page

All time worldwide box office for Walt Disney Movies. The Numbers. (n.d.). Retrieved December 9, 2022, from https://www.the-numbers.com/box-office-records/worldwide/all-movies/theatrical-distributors/walt-disney

Disney+: Stream Disney, Marvel, Pixar, Star Wars, National Geographic, and more. Stream Disney, Pixar, Marvel, Star Wars, Nat Geo. (n.d.). Retrieved December 9, 2022, from https://www.disneyplus.com/home