**Content Consumption Choices Impacted by the Pandemic   
Research Proposal**

April 11th, 2021

Brian Moon, Kelly Wen, Ryan Goding

**Overview**

**Context:** Large streaming content providers such as Netflix have vast amounts of data on viewing habits that help them plan future streaming content. All providers execute on these plans, costing them millions of dollars based on analyzing trends to maintain & produce shows, series, or movies. However, due to the pandemic, there is a potential change in behavior in content consumption. As more people are spending more time at home due to mandated lockdown policies in their country, they have more time for indoor activities such as watching more content. This drastic event can invoke changes in behaviors and habits due to emotional stress and therefore this may affect the type of content their viewers decide to consume. This change may or may not have a long lasting effect on consumption that will need to be analyzed for future planning.

**Problem:** Strategies that are forecasted and executed prior to the pandemic may not be sufficient to determine future content planning due to the potential changes in behavior or demand caused by the pandemic.

**Impact:** There is a potential risk of wasting time and money producing and maintaining content based on prior research and analysis that is no longer relevant. As consumers re-evaluate their choices, companies need to be ready to support this change in their services. If the strategy isn’t executed based on research on this event, companies will be at risk of losing consumers and market share.

**Objective & Design:** The purpose of this study is to identify trends that can show us changes in behavior caused by the pandemic. At a high level, we plan to look at what genre and content were successful at the start of the pandemic due situational stresses of that year, comparing this to prior pandemic data. The content that was popular in the pandemic may or may not be popular in the post-pandemic era, and vice versa. More details on how this data will be used can be seen in the body of this research design report. What we plan to achieve at the end of the study to provide recommendations to help plan future content based on the analysis.

**Benefit:** This research would benefit large streaming companies as well as other media providers. And because the pandemic is still underway with some regions still implementing social distancing orders, no current literature exists on media consumption habits during the pandemic. This research will help fill these current literature deficiencies.

**Audience:** Our intended audience is large streaming companies executives who define the strategy on content. Our research would help their strategy to retain and grow their customer base by providing content that customers prefer. The research outcomes will better inform streaming companies on the type of content to produce, keep, and market in the future.

**Ethics:** In order to protect consumers, personal information of individuals will not be shown. The users will not be subjected to a treatment or control study, as this is purely gathering data of their own behavior choices on content consumption. The benefit for the viewers is that large consumption companies can provide content that viewers want to consume. We will work with content creators/streaming companies in order to get the proper data. However, the research design will be solely executed and planned by the research team.

**Research Question**

Did the pandemic cause a significant change in streaming consumption behavior compared to previous years?

Sub-questions:

* What genres are new subscribers of the pandemic vs. regular subscribers watching?
* How does content consumption in genres and popular shows change across different countries?

Definitions:

* Lockdown: stay at home orders mandated by each country/region
* Pandemic: global crisis due to the spread of coronavirus (covid 19)
* Content: television series, movies, documentaries for all genres
* New subscribers: subscribers who signed on during the pandemic
* Pre-pandemic subscribers: subscribers who had a subscription prior to the pandemic

**Data**

To perform this analysis, we will need two types of data. First, streaming company's internal data that captures content demand trends. Second, data on lockdown periods for each country/region, which would allow us to map the changes in content consumption to lockdowns.

1. The Large Streaming Company’s Content Data
   1. Data that shows content consumption of subscribers over time.
   2. The data should be on the date, subscriber, and content level and should contain data fields that convey information on the content watched by the subscriber, such as:
      1. Date
      2. Subscriber ID
      3. Content ID
      4. Genre, or content tags
      5. Watch time
      6. Country content produced in
      7. Country of subscriber and other demographic information
   3. Data should be collected since whenever the data is available to one year after stay-at-home orders are lifted in regions we are trying to conduct analysis
   4. This is company’s internal data, so there would be no additional costs to obtain
2. Lockdown Data
   1. Data that shows for each country, on which dates lockdowns were enforced. The data will be on date and country level and have a flag variable showing whether stay-at-home order was enforced on that date. Thus, the data fields to be included are:
      1. Date
      2. Country
      3. Flag variable showing whether stay-at-home was enforced
   2. Data should be collected from the first date stay-at-home orders were implemented for the countries we are including in this study.
   3. This data is expected to be public, and will thus incur no additional costs to obtain.

Using the above data, we will be able to observe the impact of the pandemic on streaming consumption behavior by comparing consumption pattern changes as lockdowns continue. In addition to the time series analysis, we will be able to analyze the consumption behavior changes by region, by seeing how content consumption in regions/countries that were more heavily affected by the pandemic compared to the regions/countries that were less affected.

**Study Design**

In order to understand the impact of the pandemic, we need to be able to isolate the pandemic and lockdown variables.The study design will look at time series and panel on content consumption. We will continue to monitor and collect the data for 5 years after the pandemic to identify whether there are any lasting behaviors.There are 4 scenarios where we will try to compare pre and post treatment, the treatment as the pandemic and/or lockdown. Within each study, we will look for trends and patterns in genre and content. In the first two studies, we are comparing neighboring countries/states in order to understand the effects of the lock down on content. The last two are identifying changes in behaviors and how long it lasts, one identifying changes globally, another looking at changes in consumption with multiple lockdown periods.

* Observe content consumption trends for neighboring countries/states, one with a lock down, one without.
  + Different states/countries had different responses for the pandemic. However, the most common response was the lockdown. One country that did not implement a lockdown will be the control group and one country who did implement a lock down will be the treatment. Ideally, we can find neighboring countries to further control variation in our analysis because neighboring countries tend to have similar demographics and socio-economic norms. One example of this is that Sweden never implemented a lockdown while Norway has.
* Observe content consumption for neighboring countries/states that have different lock down periods.
  + Similar to the first study, we want to look at neighboring countries/states with different timing of lockdowns. By doing so, we can continue to observe the lockdown’s effects and the impact of it’s timing. As an example, California had an earlier lockdown period than most of the other states in the US.
* Observe change in genre and popular content over time for globally and individual countries.
  + By observing trends over time leading up to the pandemic, we can get an idea of the pandemic effects on the entire world’s consumption habits. During lockdown orders, do they watch more comedic shows to distract themselves? Do they consume more science fiction shows to educate themselves on the world crisis? This will be observed in different countries as well.
* Observe trends for one country/region with multiple lockdown periods over time.
  + The lockdown is implemented to reduce the spread if conditions are not improving. When a lockdown is over, we assume that most of the activities such as eating restaurants and going to work are going to resume. However, if conditions get worse again, another lockdown might be implemented. Studying content consumption with a country that went through multiple lockdowns, can tell us the lasting impact of the pandemic/lockdown on consuming content. We can estimate how quickly people return to their pre-pandemic/lockdown behaviors based on the content they watch. Then with another lockdown, we can observe yet another behavior change. Will they watch the same type of shows they watched during the first lockdown period?

**Sample**

Our sample will include all subscribers in each country or state we plan to conduct the study on. Because we want to be able to analyze the global population, our sample is determined by the availability of data based on criteria for each study. Therefore, our sample size can be as large as the country or states available for each scenario design. We should also be mindful if availability of data is dependent on any systematic causes, since that would determine to what extent we can generalize our study results.

We may want to segment new subscribers and pre-pandemic subscribers since people who started service post lockdown may be significantly different from the pre-existing users. For example, people who joined the streaming service post lockdown may be people who are less likely to watch the company’s content, but only decided to do so with the extra time with lockdown. The content they choose to consume may be different from those consumed by pre-existing users. How we should treat these new subscribers would depend on data investigation results, which would try to determine if the content consumption of this group is noticeably different. Depending on the investigation, we may want to simply flag these new subscribers in the data, or exclude them for our main studies yet use them to answer sub-research questions introduced in the *Research Question* section.

**Variables and/or Intervention**

The variables that we will compare over the years will be genre, content, countries/states, time(year or month) and subscriber profile. The variables explored will come from the dataset outlined above. The categorization of genres will give us a sense of emotional needs before and during the pandemic. Comparing across countries/states will give us an idea of how different countries/states consume content and how it might or might not have changed during the pandemic. The subscriber profile can help us track the number of subscribers over time, determining whether the pandemic has caused a change in number of subscriptions.

| **Variable Name** | **Research Question** |
| --- | --- |
| Independent Variable(s): Lockdown, Pandemic | Did the pandemic cause a significant change in streaming consumption behavior from previous years? |
| Dependent Variable: Streaming genre consumption habits |
| Independent Variable(s): Lockdown,Pandemic | What genres are new subscribers of the pandemic watching vs. regular consumers watching? |
| Dependent Variable: Streaming genre consumption habits of new subscribers and regular consumers |
| Independent Variable(s): Lockdown, Pandemic | How does content consumption change across different countries? |
| Dependent Variable: Streaming genre consumption habits |

**Statistical Methods**

While conducting our analysis, there are a few things we need to keep in mind.

1.There can be incremental changes in content consumption over time, but we theorize that the pandemic and lockdown has caused a significant change in content consumption. The significant change in consumption will indicate that there is a change in behavior in overall consumers. Therefore, our null hypothesis is:

“There is no significant change in content consumption patterns due to the lockdown orders during the pandemic.”

The alternative hypothesis is:

“There is a significant change in content consumption patterns due to the lockdown orders during the pandemic”.

This hypothesis will be used for each observational study listed above.

2. Because there is a multi-collinearity in our variables: pandemic and lockdown, we will use factor analysis as a way to investigate the two variable relationships. The analysis will help us determine how much each variable explains the other variable, the pandemic to lockdown and vice versa. Overall, we can determine the magnitude of each variable’s impact to the change in content consumption/behavior.

3. In each design, we will perform our analysis by two groups: new subscribers and pre-pandemic subscribers. This allows us to analyze behaviors between the two groups and compare impacts.

**Potential Risks**

Two potential risks that we may encounter during this project are:

1. The proposed research may not be aligned with the streaming company’s content vision. The content planning departments may have other planning strategies that we are unaware of and may not involve using historical viewing data or patterns. While this may occur, if a clear insight with a high level of confidence can be provided, they could use our research as a data point for their decisions or even change their future content planning strategy.
2. The degree of lockdown orders could vary among countries and even states. What the lockdown actually entails and how strictly the orders are enforced (i.e. is it recommended or forced, under what conditions can you leave home, and etc.) could substantially affect people’s sentiments and the content they choose to consume, but may not be easily observable in the data and factor into the model. This is a risk since it means that the treatment groups may not actually be receiving the same treatment.

**Deliverables**

Our anticipated timeline and deliverables for executing this research design is as follows:

Study deliverables timeline:

1. Collect and explore data outlined above.
2. Conduct study design and use stated statistical methods on the data to help answer our research questions and provide insight on consumption trends during the pandemic.
3. Determine level of confidence in insight. If the confidence is lower than required then we need to determine if more time or data is required to provide that level of confidence.
4. Once research questions and confidence is at a good point, then develop presentation deliverables.

Presentation deliverables:

* Presentation to senior executives of large streaming companies such as Netflix on findings and recommendations
  + A slide deck to help communicate our findings to any non-technical audience or decision makers interested in the outcome.
* Research paper on study will be published to the public.

**Statements of Contribution**

* Kelly Wen: In the research proposal, I will contribute to the study design, statistical methods, and overview section. In addition, I will also review other sections and submit the assignments on or before their proper due date.
* Brian: I will contribute to writing the data, sample, and potential risks sections, as well as reviewing other sections. I will also correspond and schedule any help needed with the Professor.
* Ryan : I will write the overview, research question, and deliverable sections, as well as reviewing and providing feedback on all other sections. I also contributed to the risk and variable sections. I will also serve as a project manager role for coordinating necessary meetings, and setting milestones as well as creating and presenting slides for my respective sections.