**FINAL PROJECT #2:**

**Project Problem and Hypothesis**

* ***What's the project about? What problem are you solving?***  
  TV viewership has been seeing a shift from live viewing for quite some time now. The narrative has always been that people are watching live television on a traditional screen less frequently, and that mobile and tablet were the new screen of choice.  
  We are now seeing a pretty big adoption of OTT or over-the-top devices. It offers the freedom and agency of a streaming mobile device as well as the high-quality viewing experience of a television screen. While it is still making up a small percentage of viewership, it is rapidly growing and has become the fastest growing means through which people are accessing TV content.Viacom currently has various brands’ TVE apps on AppleTV, Chromecast, and Roku. We also currently announced an exclusive partnership with Roku that gives us the rights to sell our own Roku ad inventory across our own channels (MTV, VH1, Nick, NickJr, Comedy Central, Nogging) as well as additional channels on the platform.   
  Selling and pricing this ad inventory proves to be a challenge. Inventory is small because raw volume of viewership remains small. However, because the viewing experience is more immersive and engaging, it begs a higher CPM. Couple that with the increasing desire of advertisers to target their ads to relevant audience targets makes Roku a rather difficult ad environment from a publisher perspective.  
  The ability to predict where certain audience segments will show up will help Viacom as a publishers to price more effectively as well as craft better media plans to provide advertisers with the most bang for their buck.
* ***Where does this seem to reside as a machine learning problem? Are you predicting some continuous number, or predicting a binary value?***This seems to reside as a machine learning problem because we are hoping to be able to predict where (which programs) certain audience segments are most likely to watch again, so that we can sere them with a relevant ad.We are looking to predict how many people of a certain audience segment are likely to view a particular program.
* ***What kind of impact do you think it could have?***  
  Currently, our Delivery team does very elementary modeling in excel to project ad inventory and pricing. In a world where advertisers are increasingly interested in reaching specific audience targets, this is not sufficient. If we were able to predict where certain segments of users are most likely to watch, we can build out better pricing models.
* ***What do you think will have the most impact in predicting the value you are interested in solving for?***Past viewership of a certain show.

**Datasets**

* In the initial phase of our relationship with Roku, we have received a sample of viewership data at the individual level. Our data strategy team has already paired that viewership data with Acxiom (a third-party data provider) to segment those individuals out to particular core demographics.

**Domain knowledge**

* ***What experience do you already have around this area?***  
  I currently work on the Digital Ad Sales Research team, so I am very used to having to prove that a certain network or show is right place for their messaging.  
  However, I have never done anything on the individual level and am not necessarily familiar with Roku data, as I typically work with cookie data.
* ***Does it relate or help inform the project in any way?***  
  Yes. I am very accustomed to creating audience segments using cookie data, and have tried to use excel modeling to project out video views/general traffic. However, it tends to be incredibly manual and intuitively driven, rather than data-driven.
* ***What other research efforts exist?***  
  Not much, really. Viacom has an exclusive deal with Roku. However, many media companies are coming out with their own data products to differentiate themselves in the marketplace. Viacom, even, has a product called Vantage that allows advertisers to buy a heavily targeted media plan on television (something that has traditionally only been available via digital advertising).  
  The Vantage team uses data science and machine learning to inform an algorithm that can predict what shows will have the highest concentration of a certain target. We are looking to extend this capability to our digital and OTT offerings. However, Nielsen TV data only provides data at the household level, whereas digital and OTT typically offer information at the individual level – that, I believe, will be the difference/distinction.

***Project Concerns***

* ***What questions do you have about your project? What are you not sure you quite yet understand? (The more honest you are about this, the easier your instructors can help).***  
  I am still in the process of getting data and some additional direction from my team, so have yet to see actual data. I also wonder if the scope of my project is too wide and I should reverse the model (which audience segments show up with the highest concentration to a particular show), but not sure if that would necessarily simplify it.  
  I also worry that my project topic is too complicated and that I may not be advanced enough to complete the project.  
    
  If by chance, I am not able to get the Roku data already paired with Acxiom data, does having to do that match myself complicate things?Would it be smarter to look at just one particular audience segment?
* ***What are the assumptions and caveats to the problem?***  
  There is an assumption about what constitutes viewership. Currently, we are tracking 1 second of a content view as a stream/video view, therefore making that person a viewer of a certain network/program. We are assuming that prior viewership can predict future behavior, but not looking at whether someone watched one full episode, more than one full episode, etc.   
  I also believe that this is not full Roku data, but only our network data as well as a sample of the kind of inventory we’d have the right to sell.
* ***What are the risks to the project?***
  + ***What's the cost of your model being wrong? (What's the benefit of your model being right?)***The benefit of my model being right would be to offer better targeting capabilities for advertisers looking to tap into our OTT ad inventory. From a revenue perspective, it could help make our pricing models more efficient (being able to project that a certain program may not work for an advertiser will allow us to price it more appropriately).
  + ***Is any of the data incorrect? Could it be incorrect?***To be honest, not entirely sure yet as I have yet to get access to the data that I need.

***Outcomes***

* ***What do you expect the output to look like?***  
  For one audience segment, I expect to get a return of a list of shows and their respective composition percentages for the audience segment.
* ***What gain do you expect from your most important feature on its own?***
* ***How complicated does your model have to be?***  
  I’m hoping not that complicated… but, I do think that if I have to data match Roku data with Acxiom on my own, the model may have be a bit more complicated.
* ***How successful does your project have to be in order to be considered a "success"?***Not quite sure what you are asking here… Purely mathematically speaking, I am hoping for a high accuracy %.
* ***What will you do if the project is a bust (this happens! but it shouldn't here)?***  
  Hope that this doesn’t happen, but we may have to revisit a definition of a view and try to bring in an additional factor around viewership behavior.