**Presentation Outline**

* Growing up people instill in you that Beauty is only skin deep, beauty is subjective (which is true), or better yet Beauty is in the eye of the beholder. Well today I am here to tell you that Beauty is in the eye of the machine
* With the help of MMLAB, The Chinese University of Hong Kong, we were able apply the Supervised Learning Decision Tree machine learning technique to predict outcomes based on the answers to the decision trees question.
  + For example, this model can provide Directors a high degree of certainty that if someone has specific set of features would that person be considered attractive. All without ever having to meet or see the individual
* So what qualifies as “Attractive” and/or “Unattractive”? According to dataset, these are the factors and how they correlate in regards to attractiveness
  + (Divert Attention to the visualization) So here we are able to see how Young, Arched Eyebrows, Pointy Nose, Oval Face, Rosy Cheeks and High Cheekbones are positively correlated however we can see the negatively correlated factors as well such as Big Nose, Chubby, Double Chin, and others
  + You can also look at the chart to see other correlations as well such as the high correlation between Chubby and Double Chin
* Decision Tree Training & Testing Score
  + With that we decided to utilize two training sets so see which would provide the best outcomes. Using Positively Correlated Features or Negatively Correlated Features
  + We wanted to see what features would allow our machine to predict with the greatest amount of accuracy.
    - By using the positively correlated features out machine acquired a score of 71% where ass Negatively Correlated Features provide a score of 65%
* Gini Impurity & Information Gain
  + (Open Hyper Link for Picture)
  + While building our Decision Tree there are other factors that provide valuable insight, the Gini Impurity score and Information gain
    - Gini Impurity is the best question that reduces uncertainty the most. Gini impurity allows us to quantify how much uncertainty there is at each node.
    - Gini impurity ranges 0 and 1 where lower values indicate less uncertainty at a node and quantifies our chances of being incorrect if we randomly assign a label from a set to an example in that set
      * If Our set and labels are all the same then the gini impurity would be equal to 1
  + Information gain will let us find a question that reduces uncertainty the most. It is a number that describes how much a question helps un-mix the labels at a node. The question that produces the most gain is the question chosen to ask at the node