**To submit: Code + Datafolio + Presentation**

**9-29:** Figure out possible questions and topics we can answer

**10-3:** Finalize presentation and datafolio

**Homework:**

* Create a 2 minute pitch for the pipeline for TWO IDEAS

**Project Philosophy:**

* Steps
  + Building familiarity with data & project
    - Research about industry and similar projects
  + Machine Learning Analysis
  + Slides Developer
* Tangible/Actionable insights from data analysis

**Topics (and subquestion):**

* Before/After Time-Series Analysis
  + Oscars
    - Ratings before and after Oscar nomination/winning
    - Oscar nomination prediction factors e.g. movie genre, genre popularity in a specific year, rating before nomination, movie information such as director, racial factors of actors
    - Relationship between budget/revenue and nominations
    - Oscar winners careers before/after
      * Compare with oscar nominees careers
    - Biases in oscar nominations/winners--selected subjectively by a panel
  + “The Netflix Effect”: A) Measuring movie trends and B) Find the movie unicorns (cheap to obtain b/c they’re old/independent, but capable to blow up on the platform)
    - Problem: Netflix is giving out movies like crazy ← research starting point
    - Ratings before and after a movie gets onto Netflix
      * Do old movies that get onto Netflix get a ratings/viewership bump?
      * What are the types of movies that get a ratings/viewership bump after they get onto Netflix?
    - How long after a movie has been released should Netflix buy rights to it to maximize revenue?
    - New movies vs older movies
      * How much higher are newly released movies vs older movies
    - Find the unicorns that NETFLIX can create
      * Price to success ratio
      * Anomaly detection system
        + Finding which movies tend to be unicorns and which ones fail the most, investigate outliers

Separate unicorn movies and predict them

* Rating Variation (based on multiple factors)
  + Movies and their respective rating scores +
  + Between different genres
  + Between different types of reviewers
* Genre of movie/film
  + Shifting trends of genres in movies and popular taste
  + Time series to determine how consumer preferences change over time
  + Correlation between consumer preference and movie budget etc.
* Viewership ratings
  + Determine how certain viewership ratings affect reviews/budget

Variables we could try to predict

* Movie revenue
* IMDB user rating

**Time-series analysis**

**Ratings between genres/ratings between viewers**

**Unicorn detection and analysis**

**Adjusting data--variance between movie genres, avid movie goers, etc.**

**Movie Lense data → Time between user reviews grouped by movie → How often are movies reviewed → Movie popularity → Before and after it gets on Netflix → RDD analysis?**

**Which movies are best to put on Netflix?**

**“Reliability”/Consistent of ratings score based on genres**

**Improve recommender systems --- analyze difference between casual movie goers and avid movie goers ratings across movie genres.**