SHEPHERDS AND SEARCHING FOR DIVERSITY

Examining Content Diversity in Recommendation Systems

Leo Schell Villanueva

@seanthesheepman

282	3.5M	61.2M
Following	Followers	Likes

Farming and working dogs
Scotland 🏴
Merch in link below

https://linktr.ee/Seanthesheepman

AUDIENCES WANT MORE DIVERSITY

LEO SCHELL VILLANUEVA

Data Scientist

Texas A&M University
B.A .Communication
Concentration Media Audiences

Chapman University
M.F.A. Screenwriting

Former Writer/Producer

AGENDA

1. **Business Understanding:**
   Why content diversity?
2. **Data Overview:**
   Netflix Competition Data
3. **Modeling:**
   PVR, V2V, U2U, Trending
4. **Diversity Analysis:**
   Recommender System Results
5. **Looking Ahead:**
   Modeling Timeline

BUSINESS PROBLEM

AUDIENCES DON'T CARE ABOUT ACCURATE PREDICTIONS

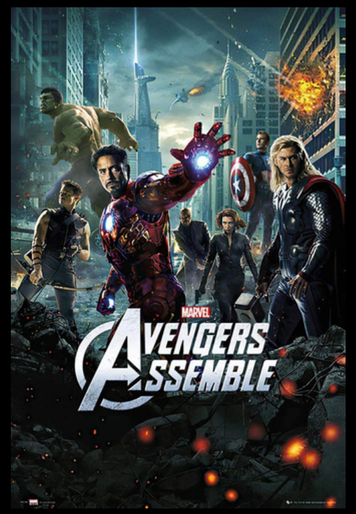

LARSON JACKSON MENDELSOHN MOUNTAIN PAUL LYNCH CRAIN BENING GREGG LAW
MARVEL STUDIOS
CAPTAIN MARVEL
IN REAL 3D MARCH 8 AND IMAX


IRON MAN
STARK INDUSTRIES


DOWNEY JR. PALTROW CHEADLE JOHANSSON ROURKE
IRON MAN 2
COMING SOON
IN THEATRES AND IMAX


EDWARD NORTON
LIV TYLER
TIM ROTH
WILLIAM HURT
THE INCREDIBLE HULK
JUNE 13, 2008


COURAGE IS IMMORTAL
THOR
FROM THE STUDIO THAT BROUGHT YOU "IRON MAN"
COMING SOON
IN DIGITAL 3D AND IMAX 3D ALSO IN 2D IN SELECTED THEATRES


MARVEL
AVENGERS ASSEMBLE


CHRIS HEMSWORTH
NATALIE PORTMAN
TOM HIDDLESTON
ANTHONY HOPKINS AS ODIN
MARVEL
THOR
THE DARK WORLD
OCTOBER 30


FROM THE STUDIO THAT BROUGHT YOU AVENGERS ASSEMBLE
MARVEL
IRON MAN 3
APRIL 26


THE
DATA
NETFLIX

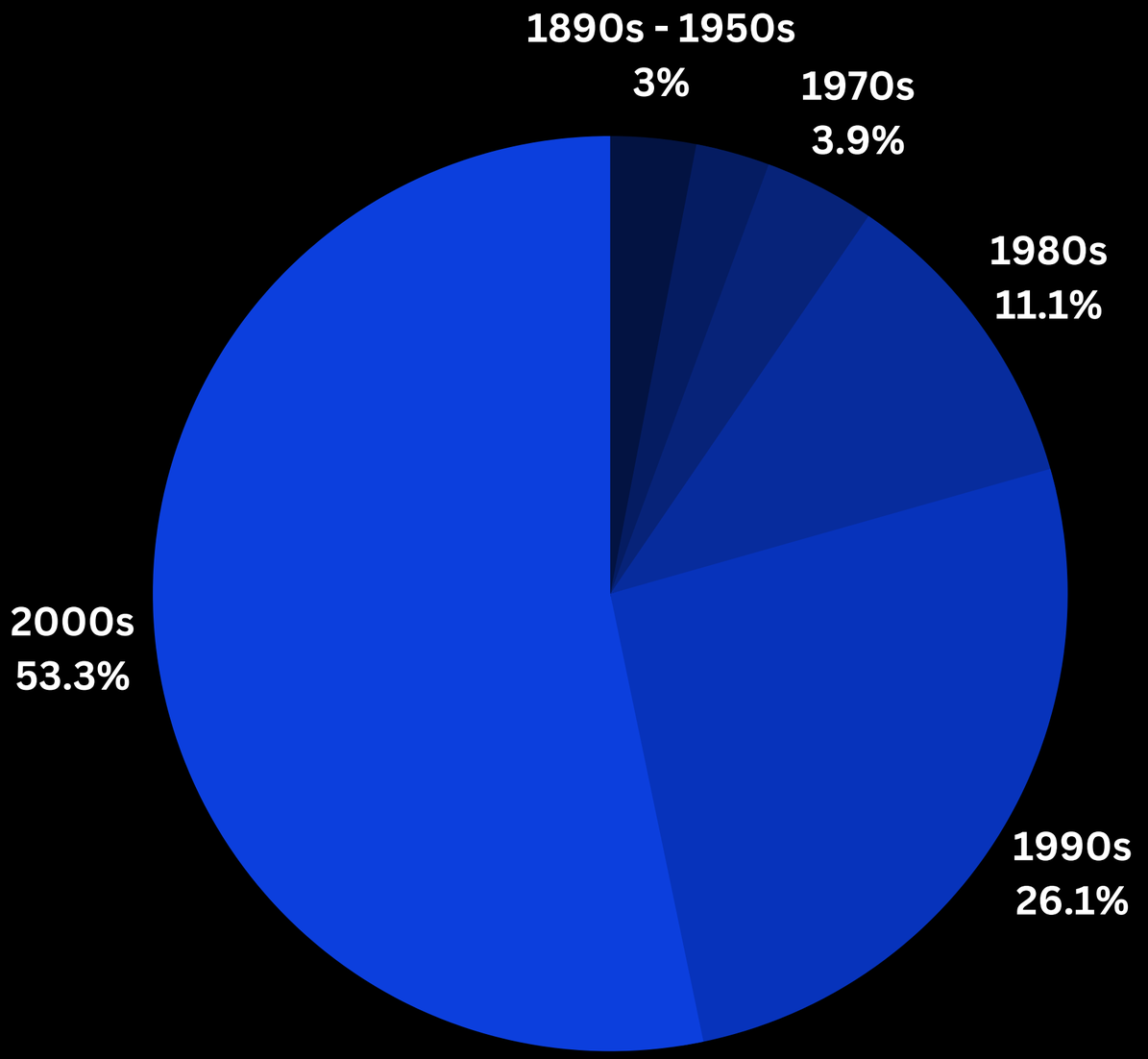
1M RATINGS FROM 290K USERS

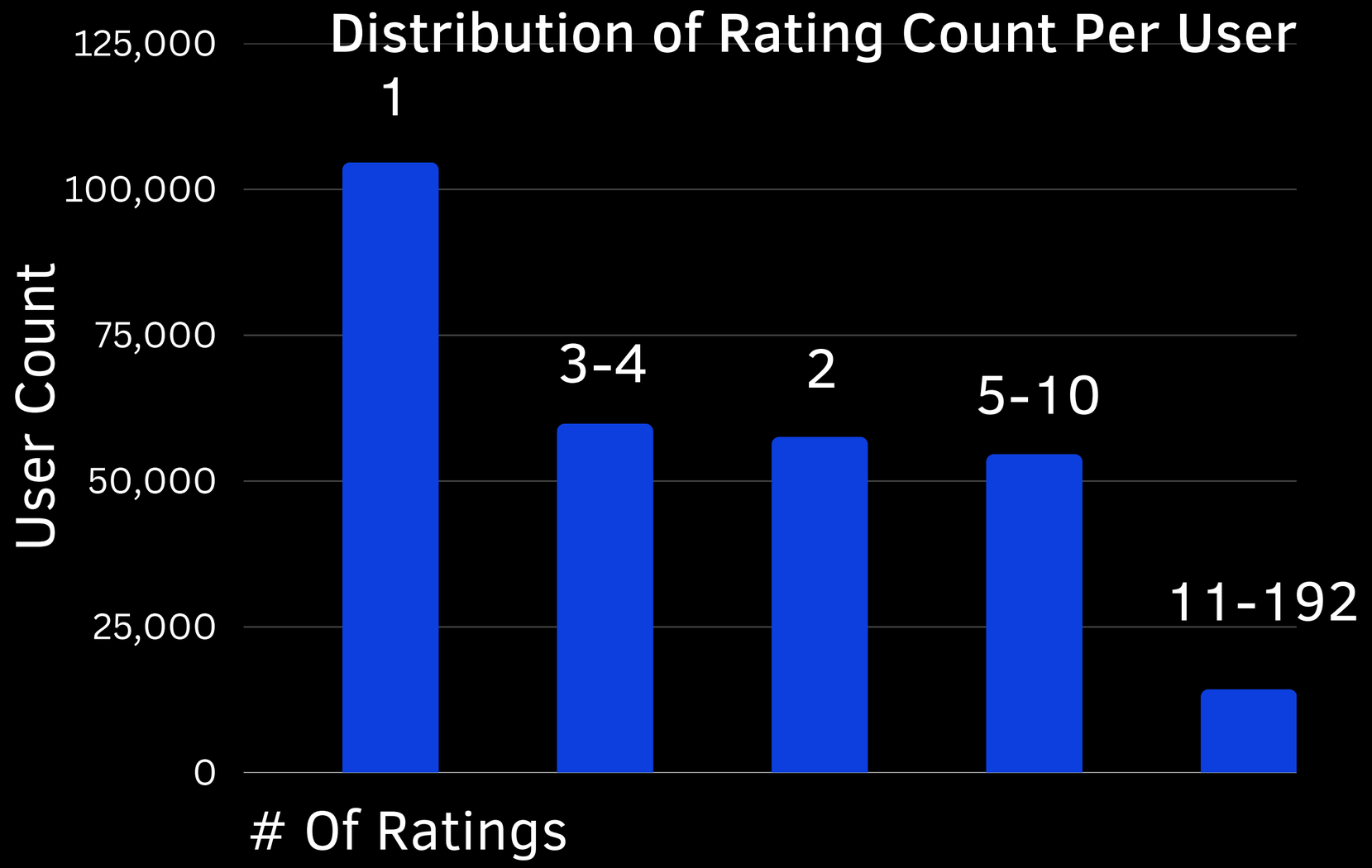
- 2006 Netflix Competition Data
- Sampled from 100M Observations
- Oct 1999 - Dec 2005
- 17,770 Movies, Shorts, and Shows
  - Year 1898 to 2004

Diversity in the Data

VIDEO BY DECADE

Vast majority released from 2000 - 2005





Diversity in the Data

USER ENGAGEMENT

Diversity in the Data

KEY DEMOGRAPHIC REPRESENTATION

Minority Top Billed Cast, Director, or Writer

12.4% Minority Representation

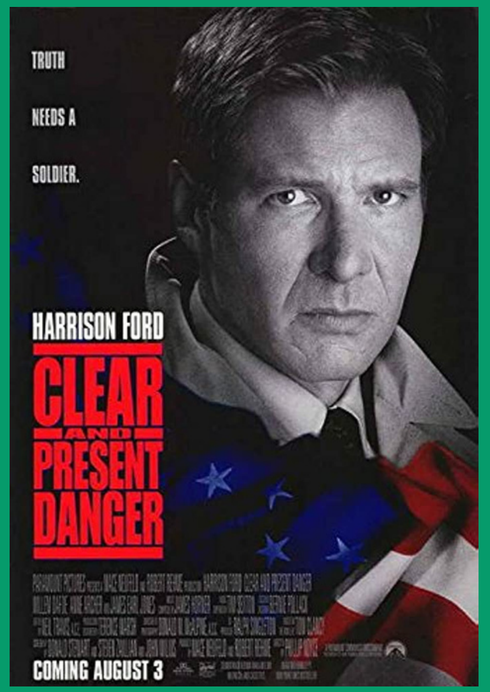

MODELING
PROCESS
THE CROWN
NARCOS
NETFLIX

THE MODELS

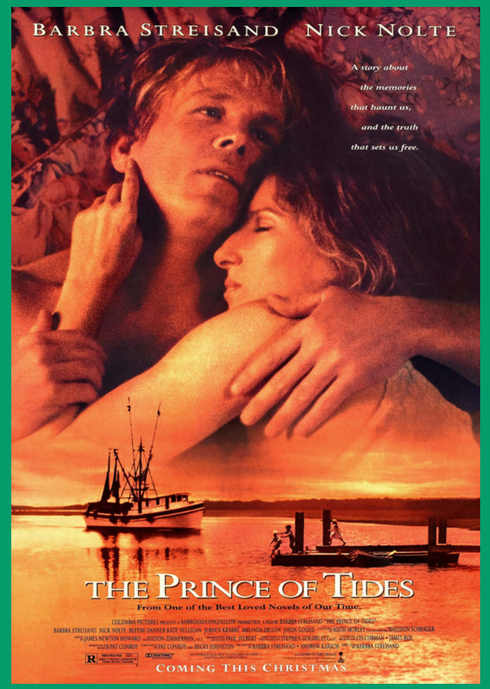
- Personalized Video Ranker
- Top Video Ranker
- Trending Now
- Video-To-Video
- User-To-User


DIVERSITY
ANALYSIS

1994

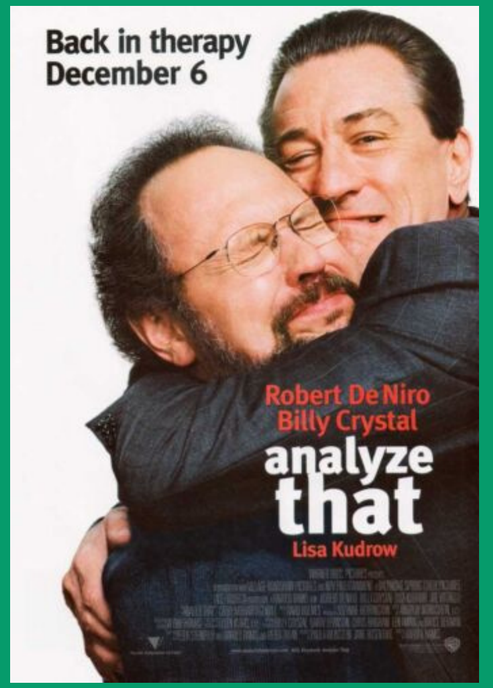


1991



1991

2002



USER PROFILE

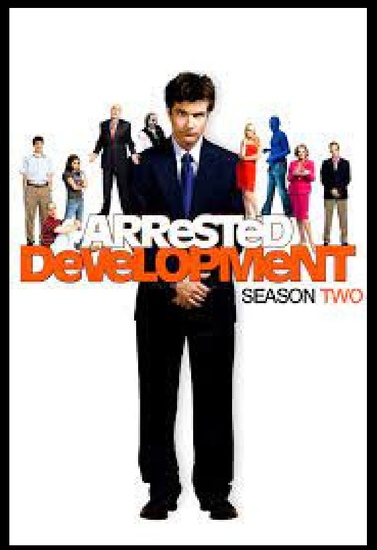
Customer #2407458

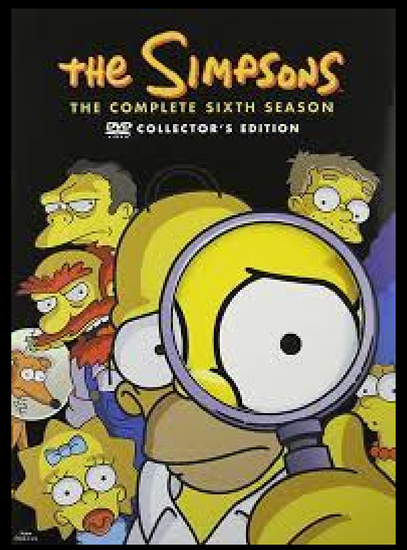
First Review: Nov 5th, 2005

Avg Rating: 3.25

BIAS BASELINE

ERROR = 0.76





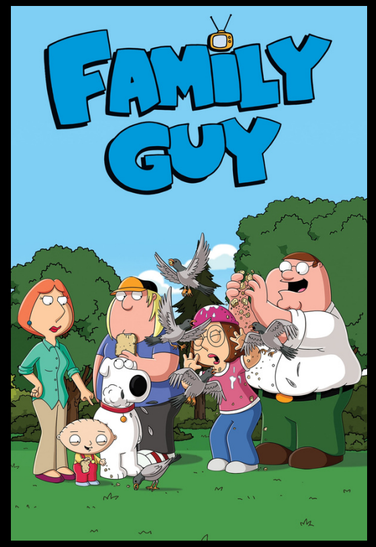
Prediction Similarity:

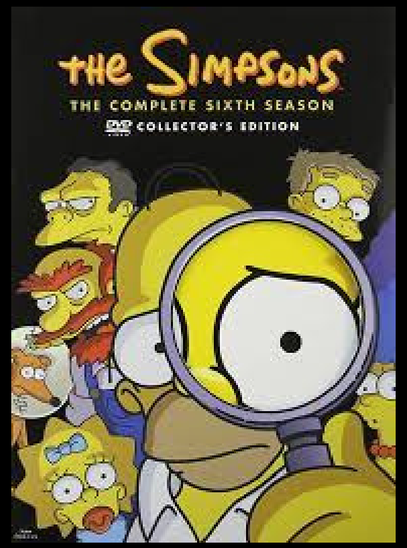
99.5%	99.7%	99.5%	99.9%	99.4%

Customer #2407458

PERSONALIZED

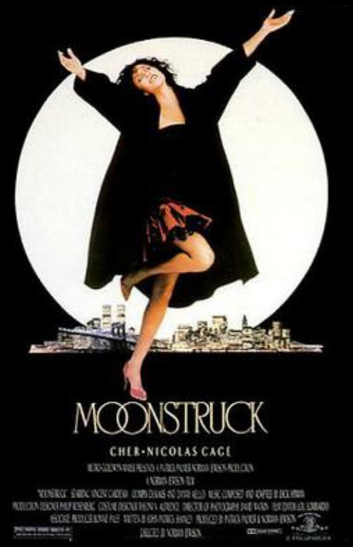
ERROR = 0.70

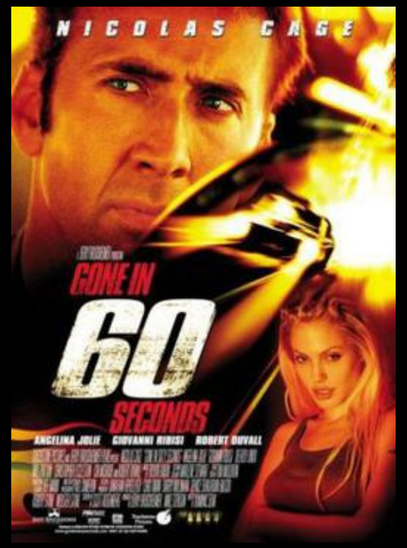




Customer #2407458

VIDEO TO VIDEO

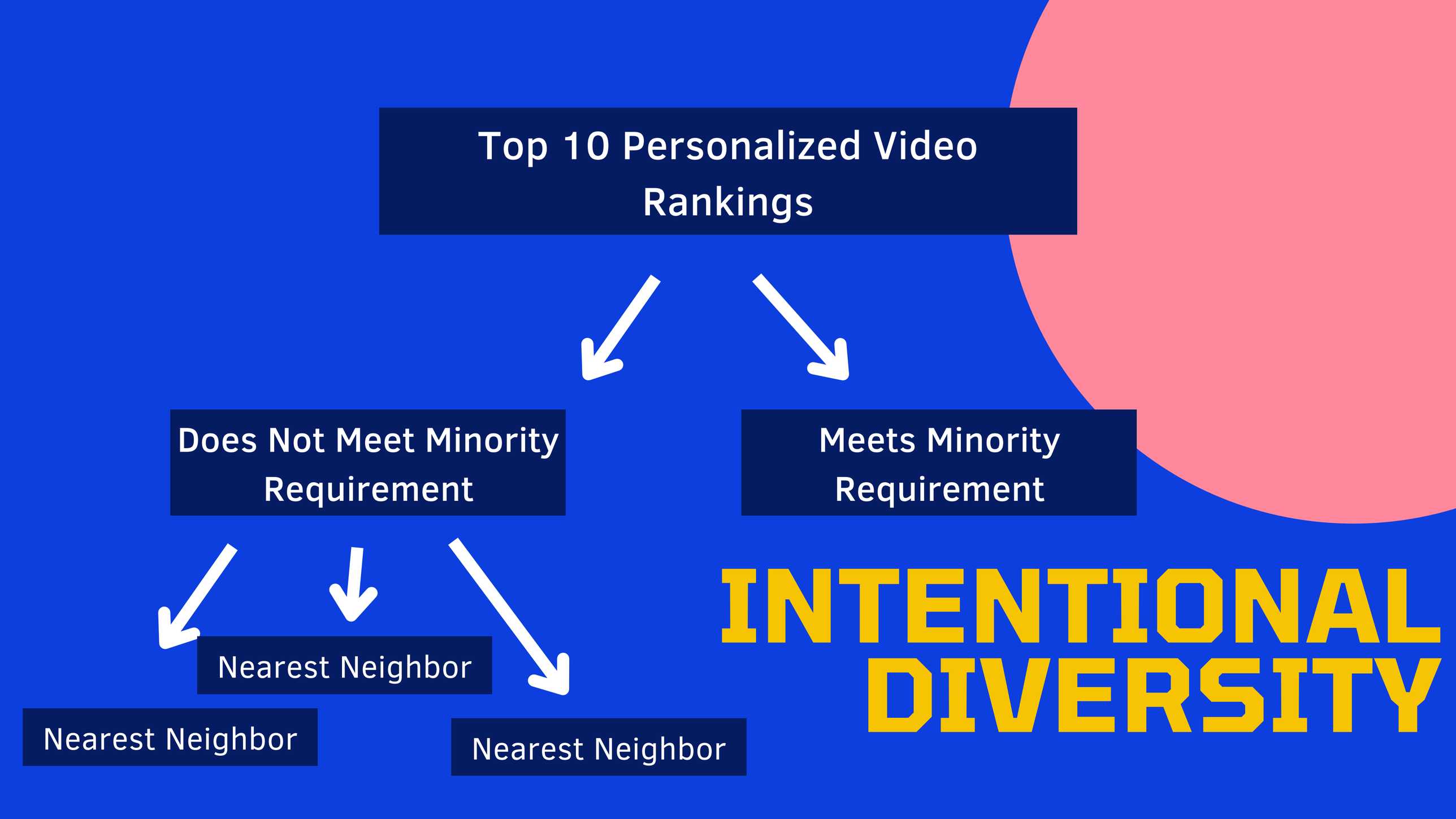




Customer #2407458

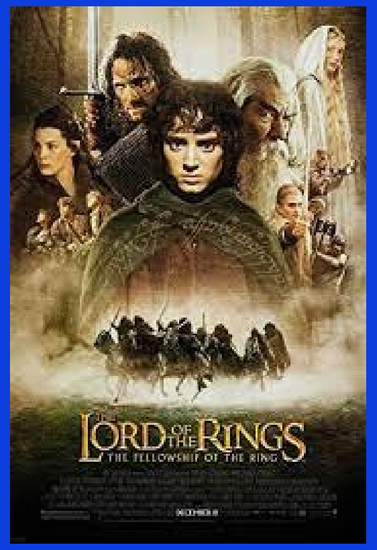
INTENTIONAL DIVERSITY

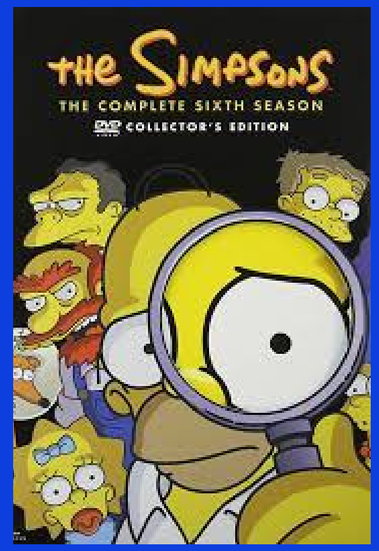
INTENTIONAL DIVERSITY


Top 10 Personalized Video Rankings
Does Not Meet Minority Requirement
Meets Minority Requirement
Nearest Neighbor
Nearest Neighbor
Nearest Neighbor
INTENTIONAL DIVERSITY

INTENTIONAL DIVERSITY

PERSONALIZED





Customer #2407458

NEXT STEPS

LOOKING AHEAD

Deep Learning
Working with Industry Standard Models and Regularization Techniques

Scraping Data:
More contemporary data

Big Data:
Getting used to working with bigger data sets

ANY QUESTIONS?

github.com/leo-schell

schell.v.leo@gmail.com