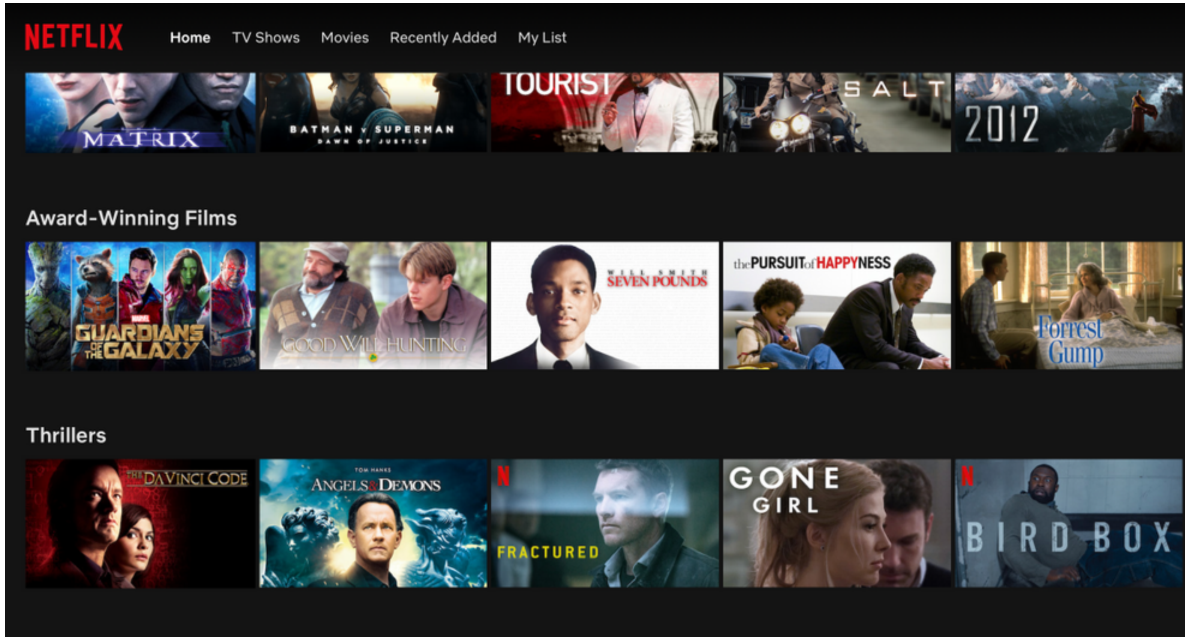
# Netflix Recommender System


Rafael Atanazio
Vincent Giraud
Sandhya Radha Krishnakumar
London Gustofson

# Everything is a recommendation!

**Over 80%** of what people watch comes from recommendations.

Recommendations are driven by **algorithms** and **machine learning**.



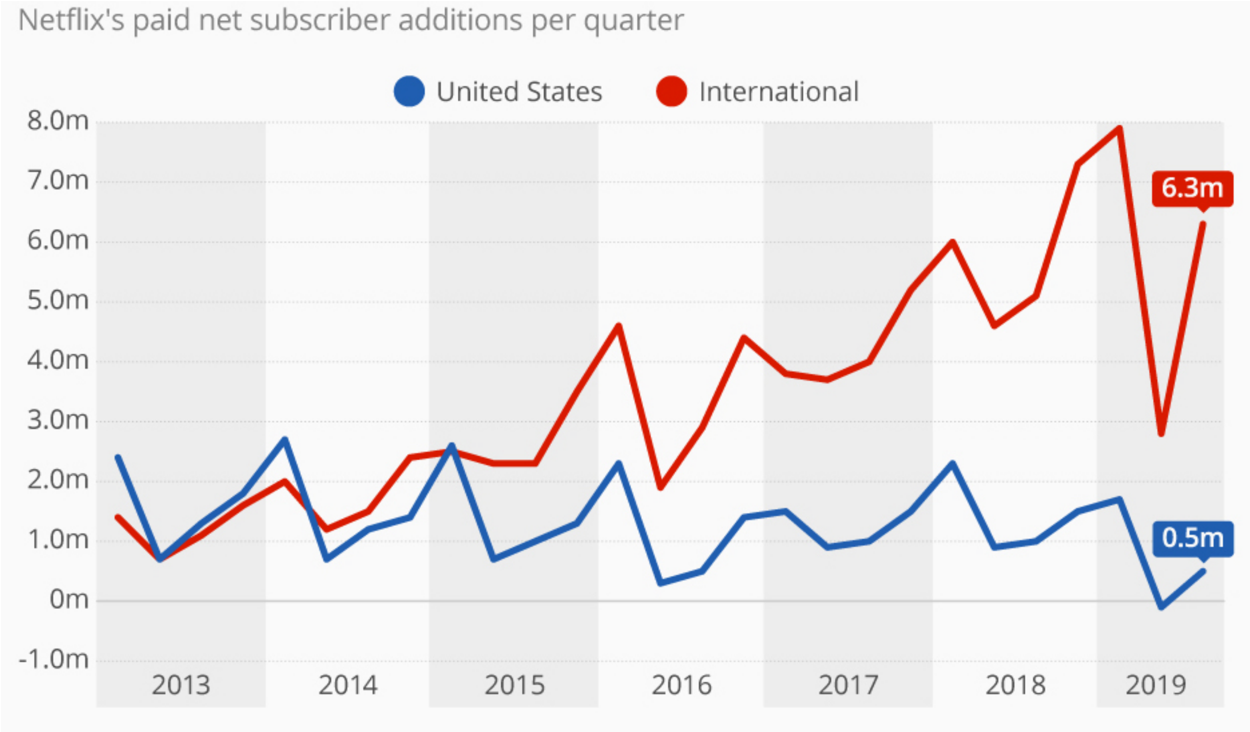
# What is a Recommender System?


**A recommender system (RS) helps users that have no time to evaluate the overwhelming, number of alternatives offered by a web site.**

**In their simplest form, RS recommend to their users personalized and ranked lists of items**

# Business Problem

# Why are the Subscription rates going down?




- **The content behind in what is currently premiering**

  Hulu, or Amazon have quicker rates of premiering the recent content

# Content Providers moving to alternative platforms

- **More competitors**

  recommendation engine needs to be updated to be more personalized

# Business Solution

***The Prototype***

# Preparing the data

- **Tableau prep**

  To deal with *inconsistent* data and other *data issues*

# Building the Algorithm

Build a data frame with ratings and movies data

Build a Rating Matrix with Users as rows and Movies as columns

Build Similarity Matrix for the users

Find the closest Users

Compute Prediction

Find the highest prediction

# Build a data frame with ratings and movies data

Filter

| | movieId | userId | rating | title |
|---|---|---|---|---|
| 1 | 1 | 73302 | 5.0 | Toy Story (1995) |
| 2 | 1 | 113839 | 2.0 | Toy Story (1995) |
| 3 | 1 | 65457 | 2.0 | Toy Story (1995) |
| 4 | 1 | 83356 | 3.5 | Toy Story (1995) |
| 5 | 1 | 39527 | 5.0 | Toy Story (1995) |
| 6 | 1 | 68116 | 4.5 | Toy Story (1995) |
| 7 | 1 | 91115 | 5.0 | Toy Story (1995) |
| 8 | 1 | 46945 | 3.5 | Toy Story (1995) |
| 9 | 1 | 130425 | 2.5 | Toy Story (1995) |
| 10 | 1 | 25436 | 3.0 | Toy Story (1995) |

- **Create a data frame**

  Merge the *ratings* and the *movies* data

# Build a rating matrix with users and movies

| | 'burbs, The (1989) | 10,000 BC (2008) | 101 Dalmatians (1996) | 102 Dalmatians (2000) | 12 Angry Men (1957) | 13 Tzameti (2005) | 17 Again (2009) | 2 Days in the Valley (1996) | 2 Fast 2 Furious (Fast and the Furious 2, The) (2003) |
|---|---|---|---|---|---|---|---|---|---|
| 8405 | 0.0000000 | 0.375 | 0.0 | 0.0000000 | 0 | 0.0 | 0.25 | 0.0000000 | 0.0 |
| 34576 | 0.0000000 | 0.000 | 0.0 | 0.0000000 | 0 | 0.0 | 0.00 | 0.0000000 | 0.0 |
| 74142 | 0.0000000 | 0.000 | 0.0 | 0.0000000 | 0 | 0.0 | 0.00 | 0.0000000 | 0.0 |
| 82418 | 0.0000000 | 0.000 | 0.0 | 0.0000000 | 0 | 0.0 | 0.00 | 0.0000000 | 0.0 |
| 83090 | 0.0000000 | 0.000 | 0.0 | 0.0000000 | 0 | 0.0 | 0.00 | 0.0000000 | 0.0 |
| 118205 | 0.3333333 | 0.000 | 0.0 | 0.0000000 | 0 | 0.0 | 0.00 | 0.0000000 | 0.0 |
| 121535 | 0.0000000 | 0.000 | 0.0 | 0.3333333 | 0 | 0.0 | 0.00 | 0.5555556 | 0.0 |
| 125794 | 0.0000000 | 0.000 | 0.5 | 0.0000000 | 0 | 0.5 | 0.00 | 0.3333333 | 0.5 |
| 131904 | 0.0000000 | 0.000 | 0.0 | 0.0000000 | 1 | 0.0 | 0.00 | 0.0000000 | 0.0 |

- **Dataframe -> Matrix**

  With *Users* as rows and *Movies* as columns

# Build Similarity Matrix for the users

| | 8405 | 34576 | 74142 | 82418 | 83090 | 118205 | 121535 | 125794 | 131904 |
|---|---|---|---|---|---|---|---|---|---|
| 8405 | 1.00000000 | 0.02565105 | 0.022947181 | 0.017240876 | 0.02111051 | 0.04351798 | 0.05207691 | 0.01719920 | 0.03757748 |
| 34576 | 0.02565105 | 1.00000000 | 0.047299854 | 0.025584419 | 0.02321188 | 0.01881706 | 0.02589055 | 0.02274442 | 0.02901758 |
| 74142 | 0.02294718 | 0.04729985 | 1.000000000 | 0.003504749 | 0.03969629 | 0.02264896 | 0.00360150 | 0.03238302 | 0.02706820 |
| 82418 | 0.01724088 | 0.02558442 | 0.003504749 | 1.000000000 | 0.02063901 | 0.02548678 | 0.03421702 | 0.02779233 | 0.01275883 |
| 83090 | 0.02111051 | 0.02321188 | 0.039696294 | 0.020639007 | 1.00000000 | 0.01815405 | 0.02862490 | 0.03835906 | 0.02316084 |
| 118205 | 0.04351798 | 0.01881706 | 0.022648956 | 0.025486775 | 0.01815405 | 1.00000000 | 0.05404819 | 0.02708391 | 0.03607282 |
| 121535 | 0.05207691 | 0.02589055 | 0.003601500 | 0.034217017 | 0.02862490 | 0.05404819 | 1.00000000 | 0.02439537 | 0.02319427 |
| 125794 | 0.01719920 | 0.02274442 | 0.032383022 | 0.027792332 | 0.03835906 | 0.02708391 | 0.02439537 | 1.00000000 | 0.02186883 |
| 131904 | 0.03757748 | 0.02901758 | 0.027068201 | 0.012758828 | 0.02316084 | 0.03607282 | 0.02319427 | 0.02186883 | 1.00000000 |

- **User based collaborative method - Cosine similarity**

# Find the closest Users and build the ratings matrix

| | userId |
|---|---|
| 1 | 121535 |
| 2 | 118205 |
| 3 | 131904 |
| 4 | 34576 |
| 5 | 74142 |

| | 'burbs, The (1989) | 102 Dalmatians (2000) | 12 Angry Men (1957) | 2 Days in the Valley (1996) | 2012 (2009) | 25th Hour (2002) | 28 Days Later (2002) |
|---|---|---|---|---|---|---|---|
| 34576 | NA | NA | NA | NA | NA | NA | 0.08977868 |
| 74142 | NA | NA | NA | NA | NA | NA | NA |
| 118205 | 0.1305539 | NA | NA | NA | NA | 0.1740719 | NA |
| 121535 | NA | 0.1041538 | NA | 0.1562307 | 0.1562307 | NA | NA |
| 131904 | NA | NA | 0.1690986 | NA | NA | NA | NA |

- **Ratings matrix of the closest users**

# Compute the highest prediction

```
                                                     titles
1                                  West Side Story (1961)
2                                      Set-Up, The (1949)
3                         Shake Hands with the Devil (1959)
4                          Shawshank Redemption, The (1994)
5                                State of the Union (1948)
6                                    Stunt Man, The (1980)
7  Tae Guk Gi: The Brotherhood of War (Taegukgi hwinalrimyeo) (2004)
8                                         Town, The (2010)
9                Unknown, The (a.k.a. Alonzo the Armless) (1927)
10                        Silence of the Lambs, The (1991)
```

- **Using weighted value**

  Similarity values * ratings


You give us 3 seconds,
we’ll give you the world.


NETFLIX
You give us three seconds...we'll give you the world
Choose a Movie
Toy Story (1995)
Rating 1-5
1.0
Choose a Movie
Toy Story (1995)
Rating 1-5
1.0
Choose a Movie
Toy Story (1995)
Rating 1-5
1.0
Choose a Movie
Toy Story (1995)
Rating 1-5
1.0
Choose a Movie
Toy Story (1995)
Rating 1-5
1.0
Submit My Ratings

# Business Gain

Users: Better Personalization

Becoming a bigger and more powerful platform could help in winning new movies rights, subscribers

Netflix: Increase in loyalty from Users and Content providers, leading to higher data and revenue generation

Content Providers: Better visualization of real size of audience and the target audience for their content

# The next steps..

| | |
|---|---|
| **Processing time** | Improve the *processing time* and include more *ratings* |
| **Analysis Methods** | Include *regression, clustering* and other computation methods |
| **Compute RMSEs** | Finding the *least* RMSE value methods |
| **Machine learning** | Deploy methods like *gradient descent* |

NETFLIX

Questions?