SUREC ENGINE

PRIME MOVIE RECOMMENDER SYSTEM

MEET US

DEVELOPMENT TEAM

Mercy Gathoni

Team Lead

Samuel Mijan

Team Member

Linda Kelida

Team Member

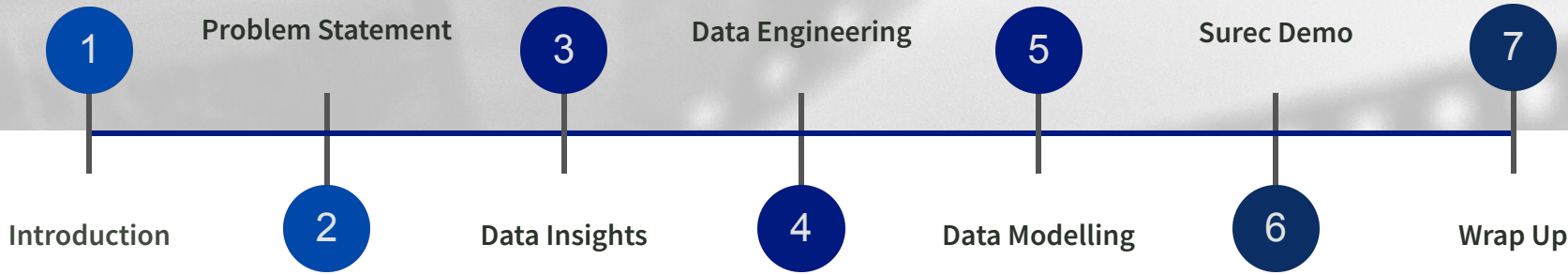
Jessica Njuguna

Team Member

Sipho Lukhele

Team Member

PRESENTATION OUTLINE



INTRODUCTION

DIGITAL

Development of technology has changed the way we live.

ENTERTAINMENT

The movie industry has changed from going to the theatre to home streaming services.

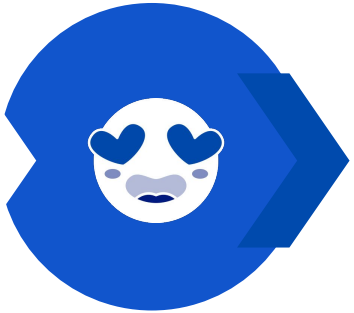
INFORMATION OVERLOAD

The information available is vast and their is need to direct attention to a person’s preference and subject of interest

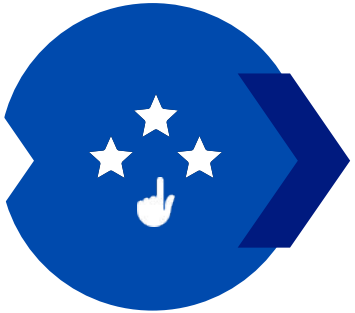
PROBLEM STATEMENT

Streaming Profits

Customer Engagement



Movie Recommendations



Surec App

DATA OVERVIEW

Ratings
Data on ratings of a movie per user

Movies
Titles and Genres of Movies

Movies Metadata
Extra details about movies from IMDB

Genome Tags
Tag names and tag IDs

Links
IMDB ID for each movie

Genome Scores
Relevance of a tag on a movie

DATA OVERVIEW

162K+

Number of Users

48K+

Number of Movies

25

Average Number of ratings per movie

22

Number of Genres

Law of the Lawless

Longest item duration with a 3.6 rating

Woody Allen

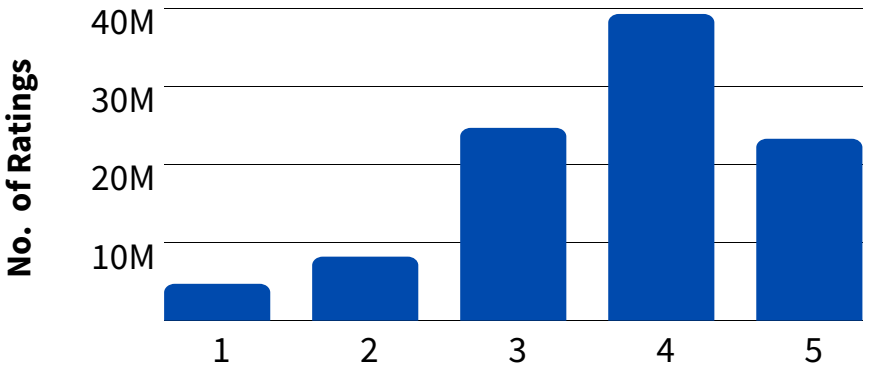
Most popular director

Shawshank Redemption

Highest movie rated

1995 - 2019

Rating year range

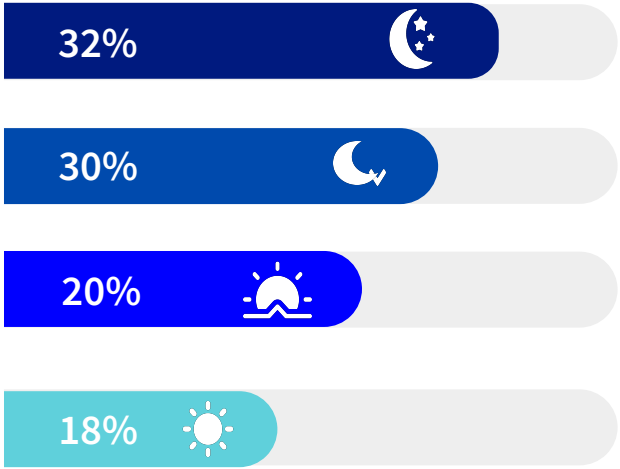


DATA INSIGHTS

TIME AND RATINGS

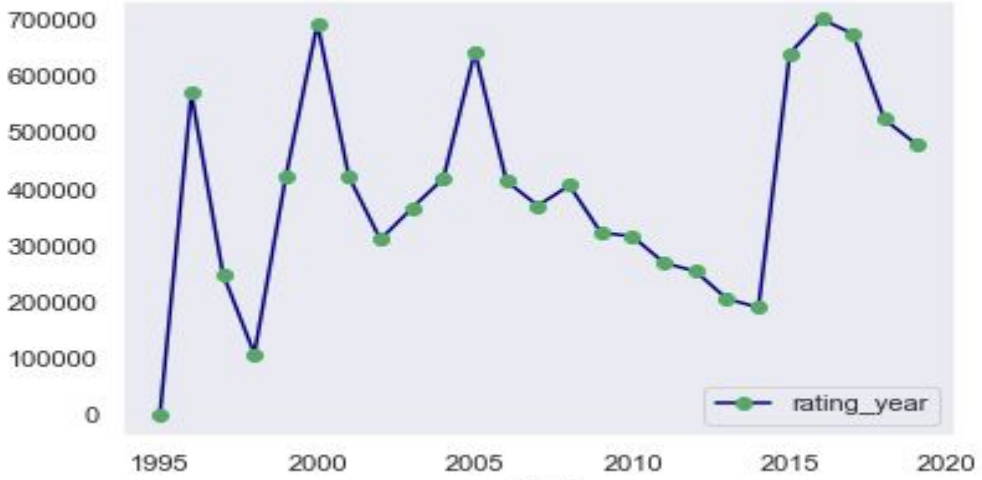
Ratings per time of day

Most movies were rated at night, with few ratings during the daylight hours

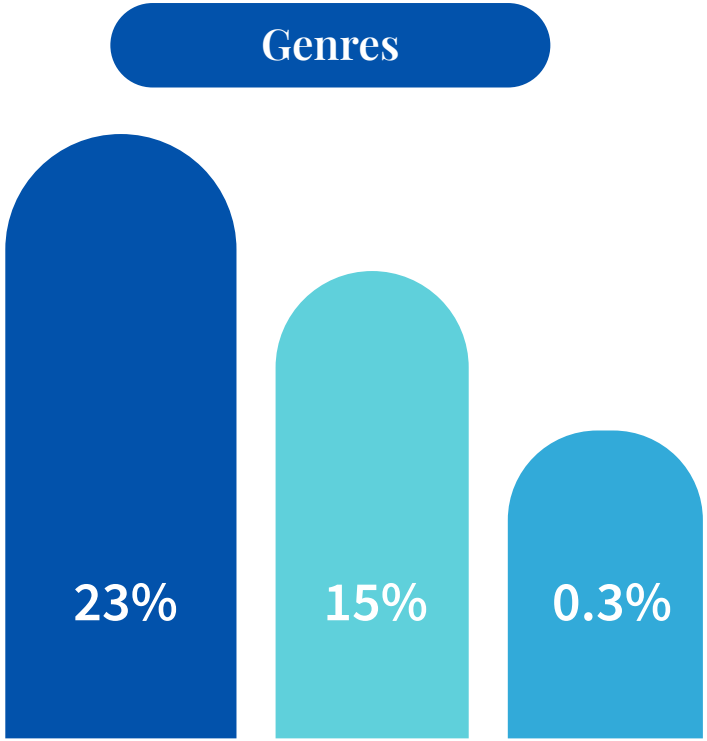


Ratings over Year

No of ratings peaked in 2000 and 2016. Currently on a dip.



DATA INSIGHTS



Drama

Most Frequently Rated Genre

Comedy

Second most frequently rated genre

Musicals

Among the least rated genre

DATA INSIGHTS

Missing Data

71%

Budget

44%

Keywords

40%

Runtime

DATA PREPROCESSING

Merge Datasets

Combined Ratings data with Movies Metadata

Null Values

Dropped the columns with the two highest null values. Imputed the rest with the model

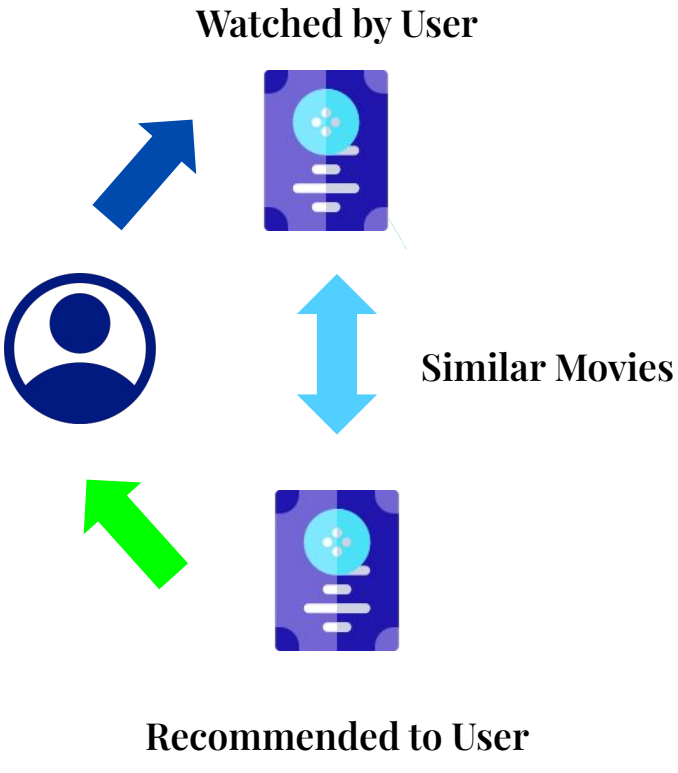
Create Word Bank

Combined the metadata into a soup of words

Vectorization

Converted the word bank into features with relevant scores

CONTENT BASED FILTERING



01

Cosine Similarity

Created a matrix with the similarity movies have with each other

02

Build a Function

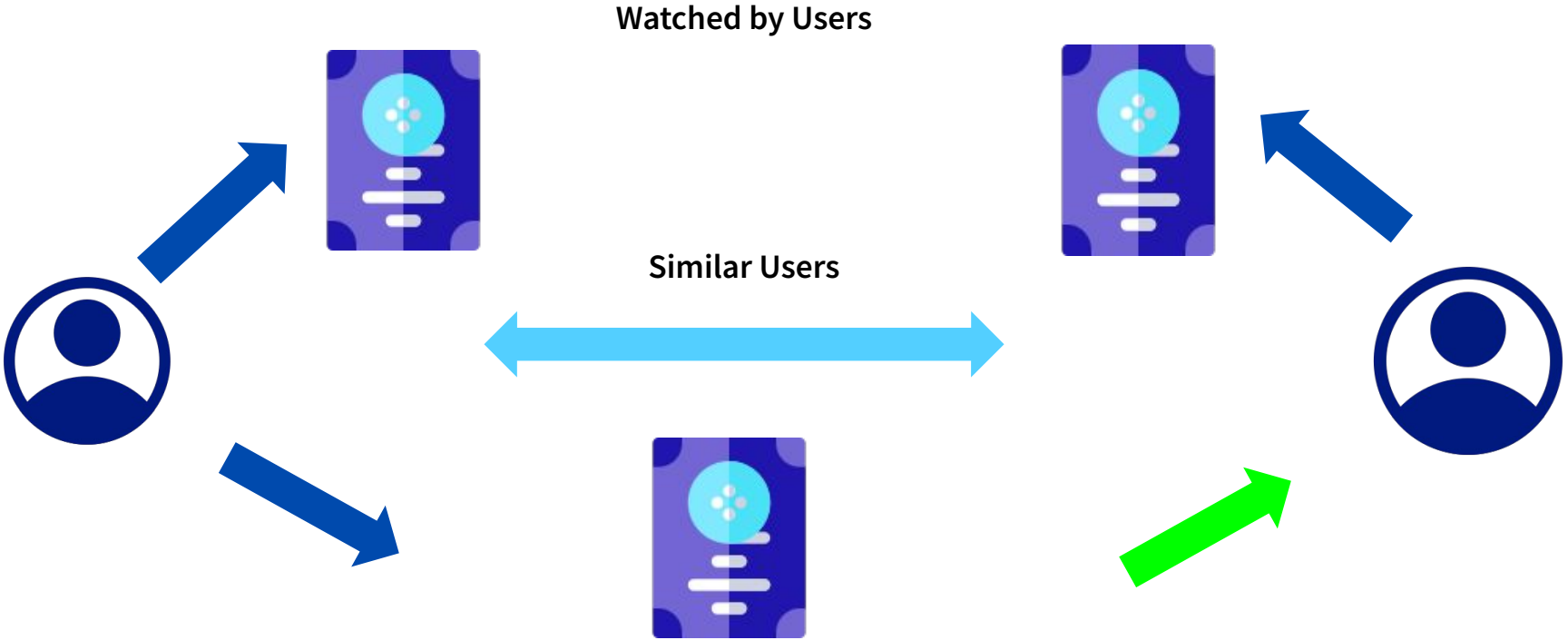
Built a custom function, that takes in a movie and checks the similarity score

03

Recommends Movies

The function recommends the top 10 movies that the user will enjoy

COLLABORATIVE FILTERING



COLLABORATIVE FILTERING MODELS

Singular Value Decomposition(SVD)

Factorizes one matrix into three matrices

0.79

SVDpp

This model is an extension of SVD that takes into account implicit ratings

0.91

Baseline Only

Model that provides predictions without complex computations

0.86

SUREC APP DEMO

WRAP UP

Task

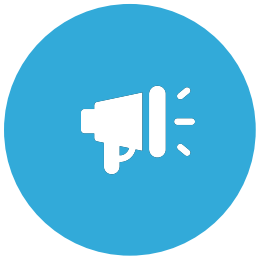
Build a recommender system using collaborative and content filtering.

Result

Surec Engine is user-friendly and has additional features that customers will love.

WRAP UP

Opportunities



Ratings Campaign

Boost rating statistics to improve recommendations

Drama Content

Increase entertainment users' love

Time Factor

Implement recommendations based on time

Thanks

Questions?

Contact:

info@surecteam.com

+254 704 118713

www.surecengine.com