DON'T CHOOSE, CHILL

MOVIE RECOMMENDATION SYSTEM

Are you still searching?

TABLE OF CONTENTS

01

INTRO

Quick Introduction

INTRO 01

HIBAH AHMED

LILY HU

CHRISTINA MACKIEWICZ

JENNIFER PERES

ANZHELIKA SUCHKOVA

01 INTRO

Are you still searching?

Recommend

Continue Searching

INTRO 01

WHY DID WE CHOOSE THIS TOPIC?

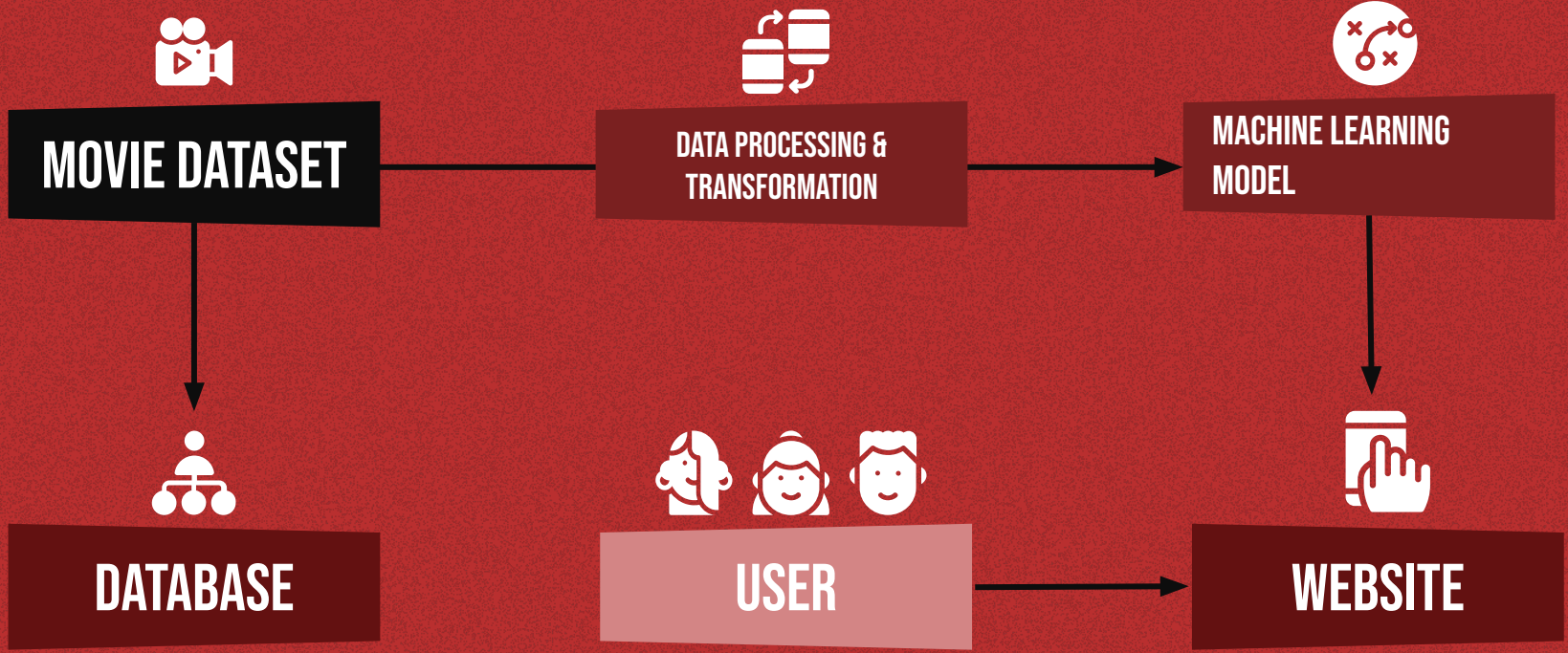
Sometimes it’s hard to pick a good movie…

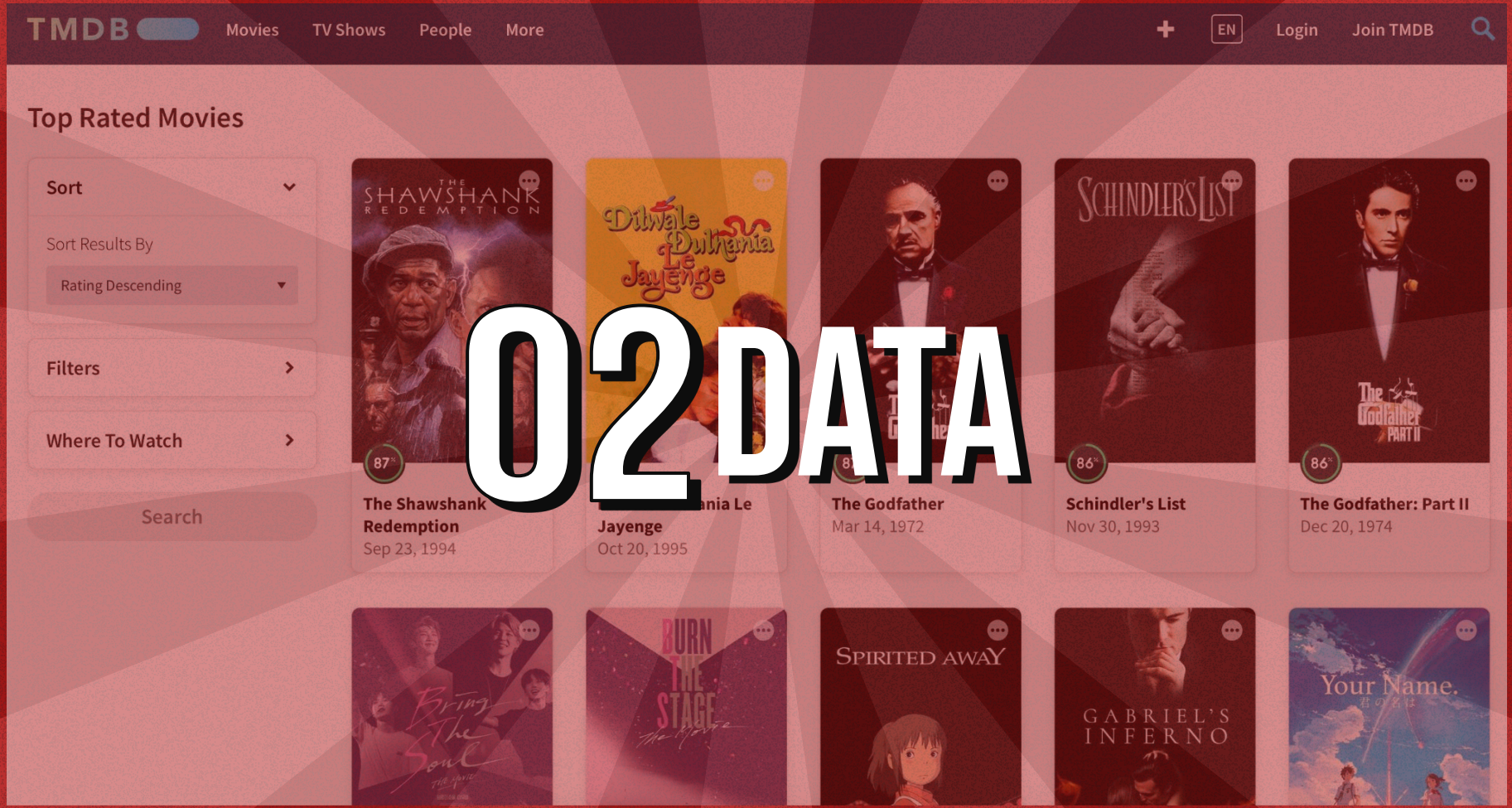
WHAT DOES THE MODEL DO?

Recommends similar movies based on the input

01 INTRO

PROJECT STEPS




TMDB
Movies
TV Shows
People
More
EN
Login
Join TMDB
Top Rated Movies
Sort
Sort Results By
Rating Descending
Filters
Where To Watch
Search
THE SHAWSHANK REDEMPTION
87%
The Shawshank Redemption
Sep 23, 1994
Dilwale Dulhania Le Jayenge
Oct 20, 1995
The Godfather
Mar 14, 1972
SCHINDLER'S LIST
86%
Schindler's List
Nov 30, 1993
The Godfather PART II
86%
The Godfather: Part II
Dec 20, 1974
Bring The Soul
BURN THE STAGE
SPIRITED AWAY
GABRIEL'S INFERNO
Your Name.
02 DATA

02 DATA

DATA SELECTION

TOP 5,000

THE MOVIE DATABASE (TMDB)

★ ★ ★ ★ ☆

2 DATASETS

CREDITS

3 COLUMNS

MOVIES

19 COLUMNS

4800 MOVIES

20,000 DIFFERENT WORDS

used to describe most popular films

02 DATA

QUESTIONS WE HOPE TO ANSWER WITH THE DATA:

- What data would work best for this analysis?
- What type of machine learning model would work best?
- Can we make predictions based on keywords?
- What kind of input do we need from the user?

02 DATA

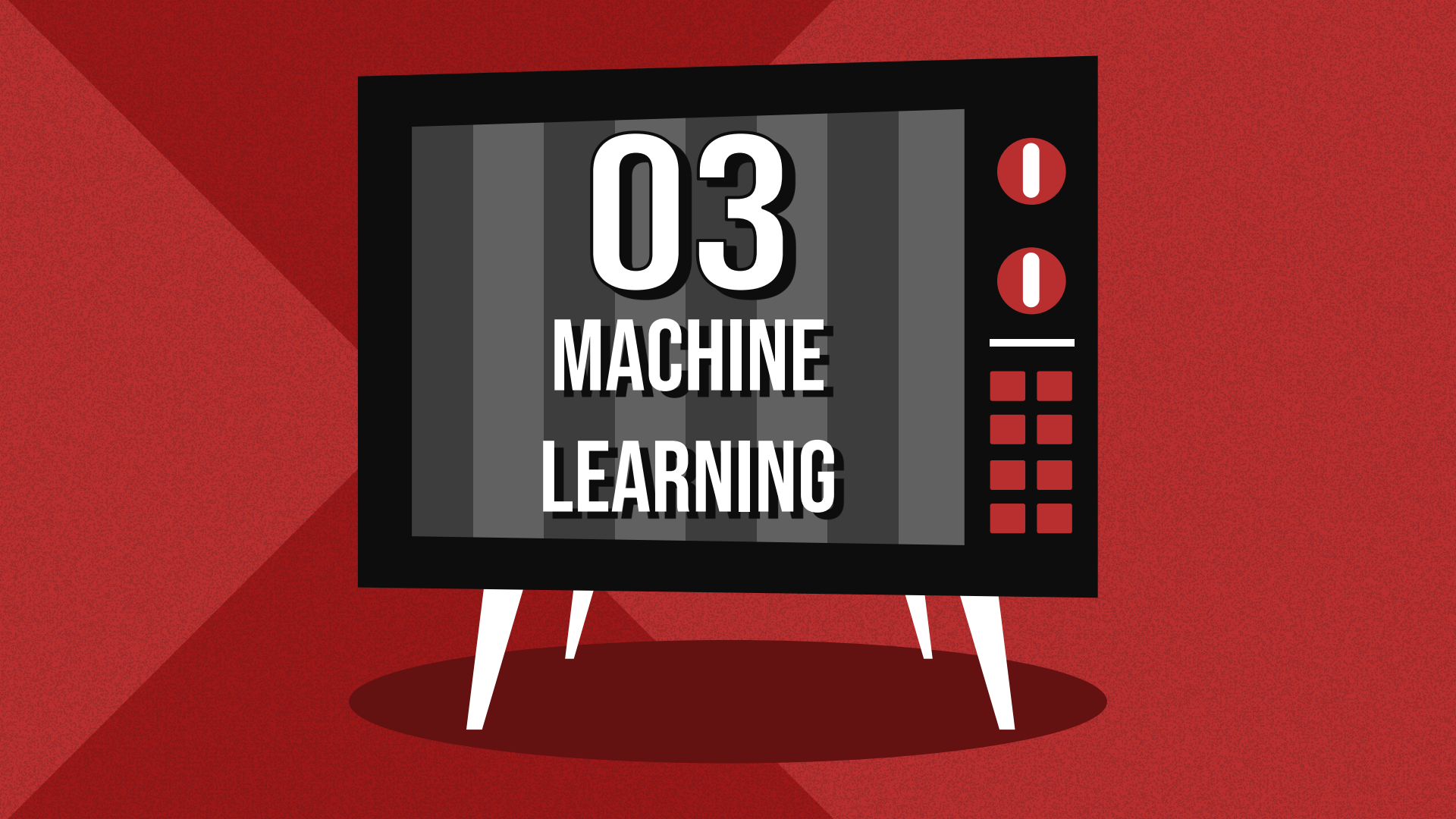
DATA PROCESSING STEPS

I. The CSV was imported into a pandas dataframe using Jupyter Notebook
II. Combined 2 datasets into 1
III. Removed unnecessary data
IV. Kept:
`'movie_id','title','overview','genres','keywords','cast','crew'`
V. Dropped null rows from dataframe
VI. Converted several columns to strings
    A. Genre, keywords, production company, and cast, crew, overview
    B. Added to a string/column name tag and created a new dataframe:
        1. Movie id, title, and tags

02 DATA

DATA PROCESSING STEPS

V. Using WordNetLemmatizer changes to the words to its root form, applied to our tags column
VI. Using CountVectorizer, utilized stop words to *(will come back to)*
VII. Create a vector tag which transforms to an array
VIII. Calculating cosine similarity from one movie to another
IX. Display the recommendations


03
MACHINE
LEARNING

CONTENT BASED FILTERING

2ND MACHINE LEARNING MODEL

3RD MACHINE LEARNING MODEL

03 MACHINE LEARNING

WHICH MACHINE LEARNING MODEL DID WE USE?

A

CONTENT BASED FILTERING

Machine Learning technique that uses similarities in features to make decisions.

B

2ND MODEL

It's the fourth-brightest object in the night sky. It was named after the Roman god of the skies

C

3RD MODEL

Mercury is the closest planet to the Sun and the smallest one in the Solar System

04
DATABASE

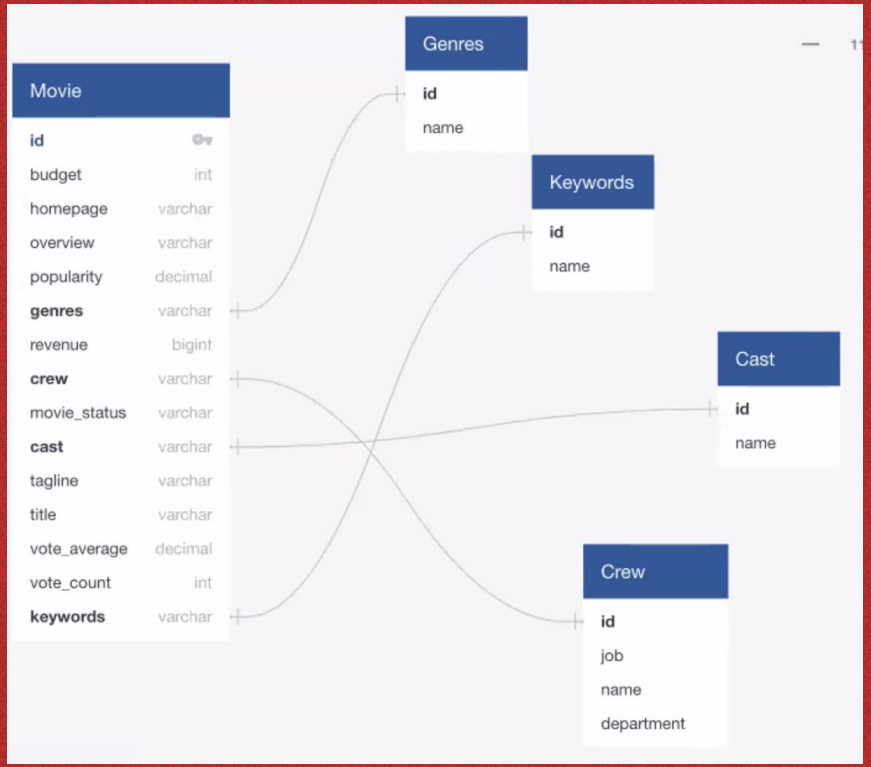
CREATING TABLE

Query Editor | Query History

```
1  CREATE TABLE movie (
2    budget int DEFAULT NULL,
3    homepage varchar(1000) DEFAULT NULL,
4    movie_id int NOT NULL,
5    overview varchar(1000) DEFAULT NULL,
6    popularity decimal(12,6) DEFAULT NULL,
7    revenue bigint DEFAULT NULL,
8    movie_status varchar(50) DEFAULT NULL,
9    tagline varchar(1000) DEFAULT NULL,
10   title varchar(1000) DEFAULT NULL,
11   vote_average decimal(4,2) DEFAULT NULL,
12   vote_count int DEFAULT NULL,
13   PRIMARY KEY (movie_id)
14 );
15
16 select * from movie
```

04 DATABASE

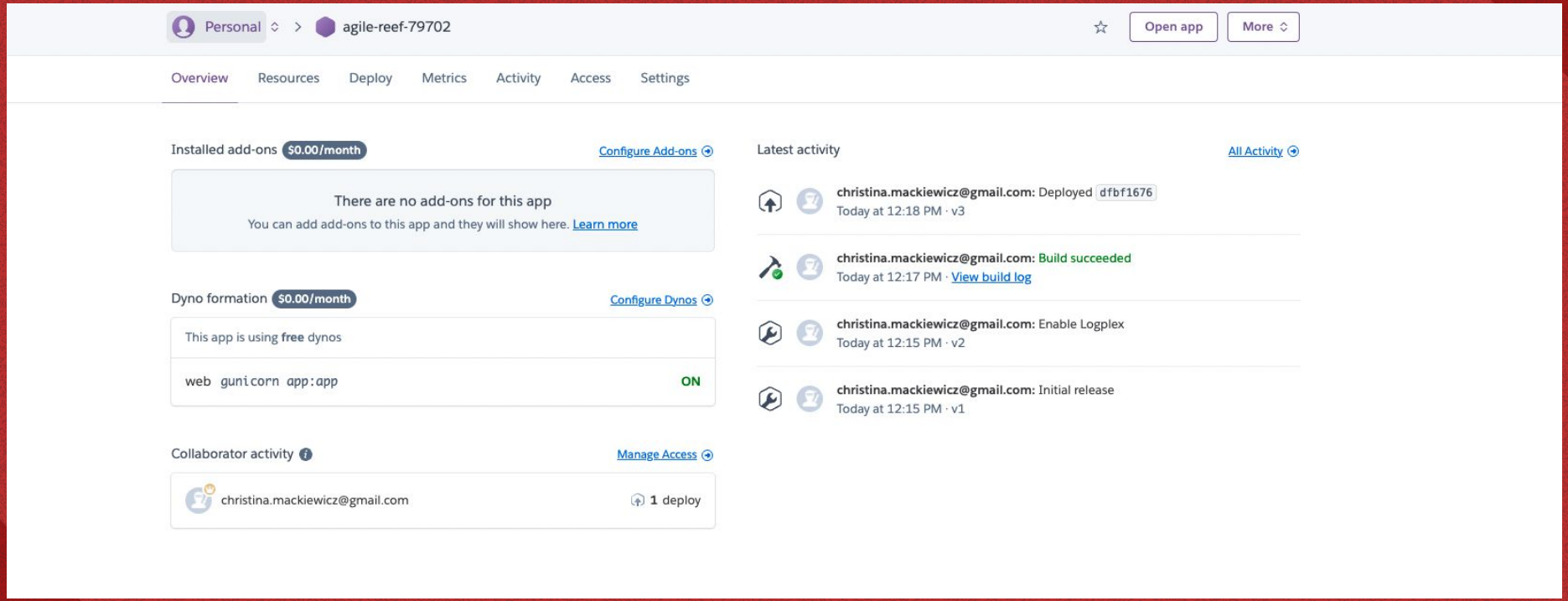
DIAGRAM



05
WEBSITE

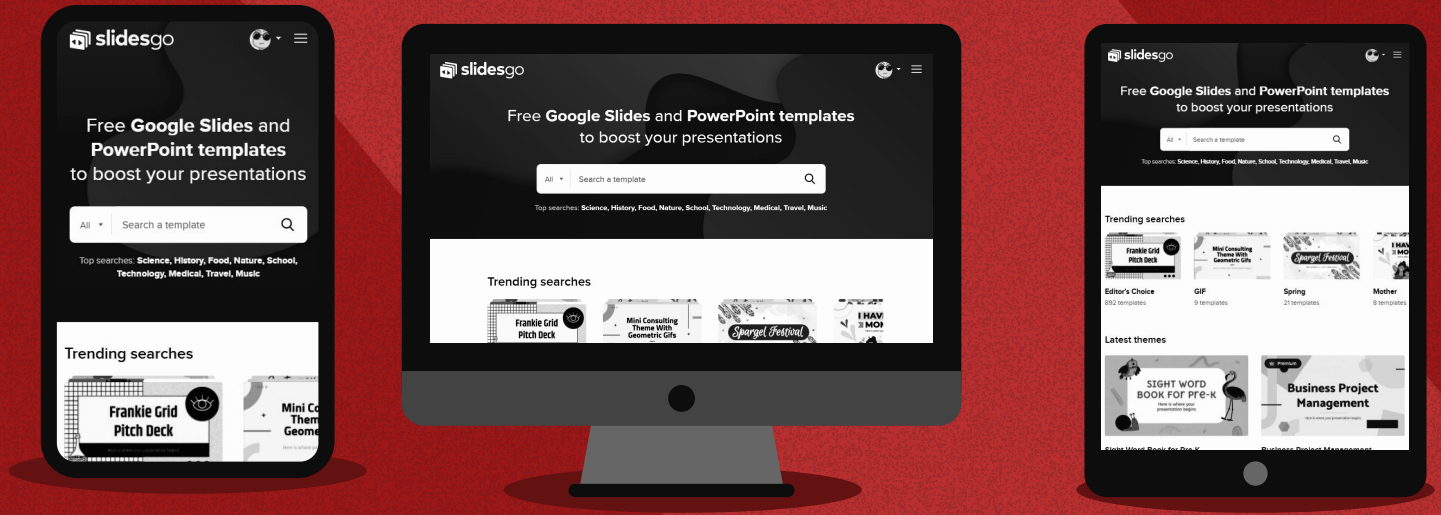
05 WEBSITE

HEROKU



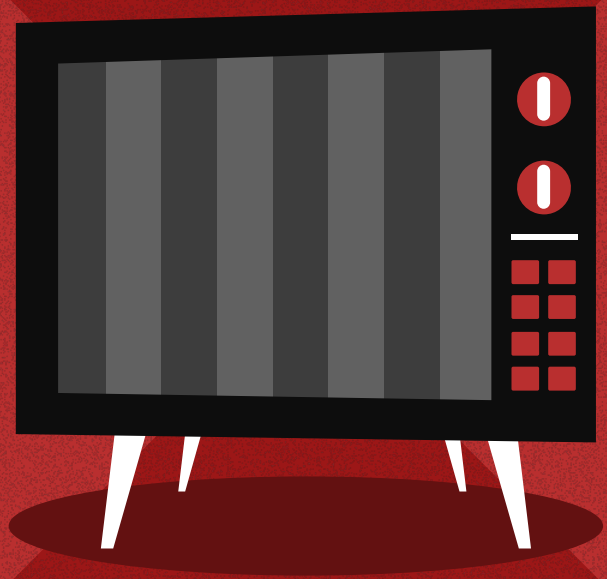
05 WEBSITE

WEBSITE



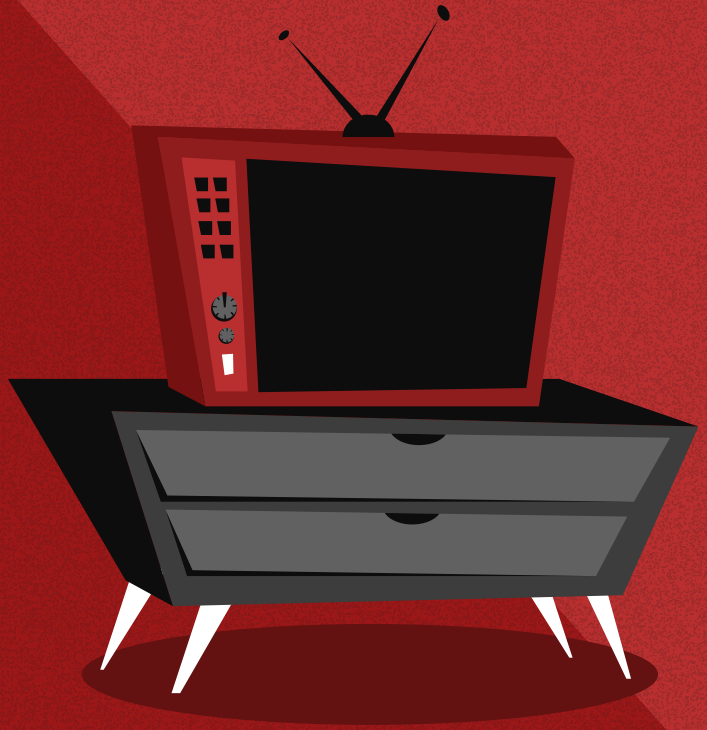
These are the examples of how it will look like on different platforms

06 THE END



RESOURCES

- Python
- Pandas
- SQL
- PGAdmin
- SciKit Learn
- Jupyter Notebook
- Heroku



THANK YOU FOR FOLLOWING ALONG!

Do you have any questions?

Fonts & colors used

This presentation has been made using the following fonts:

Bebas Neue
(https://fonts.google.com/specimen/Bebas+Neue)

Lato
(https://fonts.google.com/specimen/Lato)