predictive

YouTube Video Prediction

Will this video be successful?

machine learning

Andre Kaschkow & Raphaela Boll

Agenda

Image sources could be seen in the appendix

1. Intro
YouTube | Rocket Beans | Business Case

2. Methodology & Data
Data Set | Our way

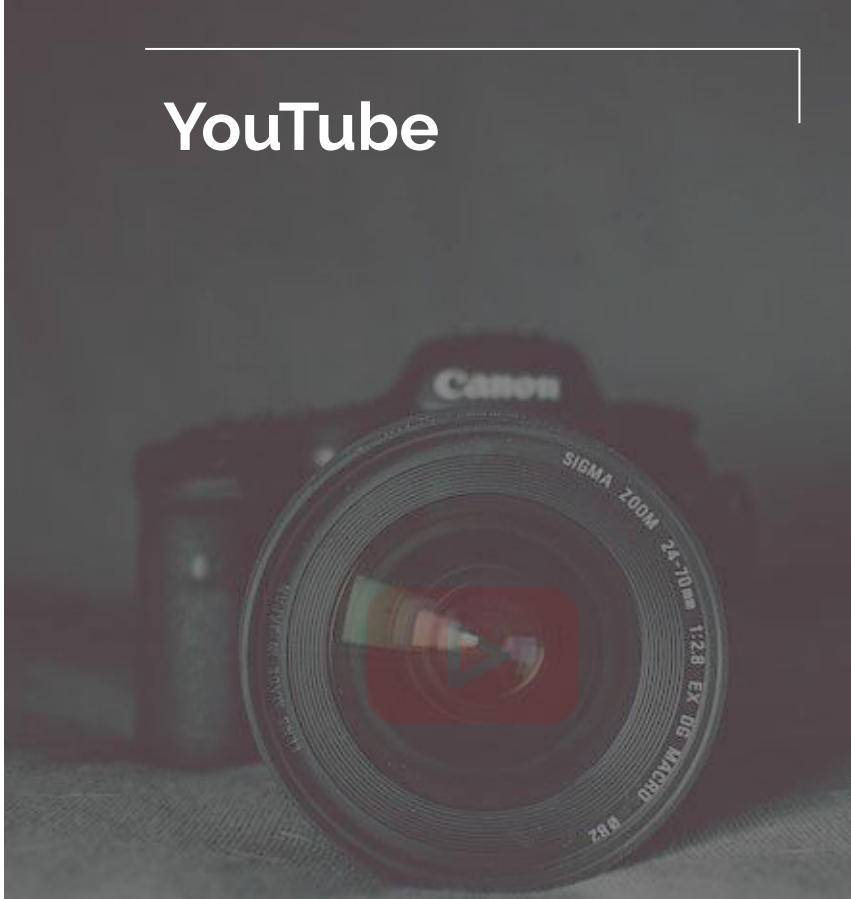
3. Findings & Recommendations
Facts & Figures

4. Achievements
and future work

*“Oh wait, let me check if there is an online tutorial for that on **YouTube**!”*

How long do you usually spend your time watching YouTube and what is your favorite channel?

YouTube



about

YouTube about

One billion

hours watched daily

2nd largest

Search Engine after Google

2+ billion

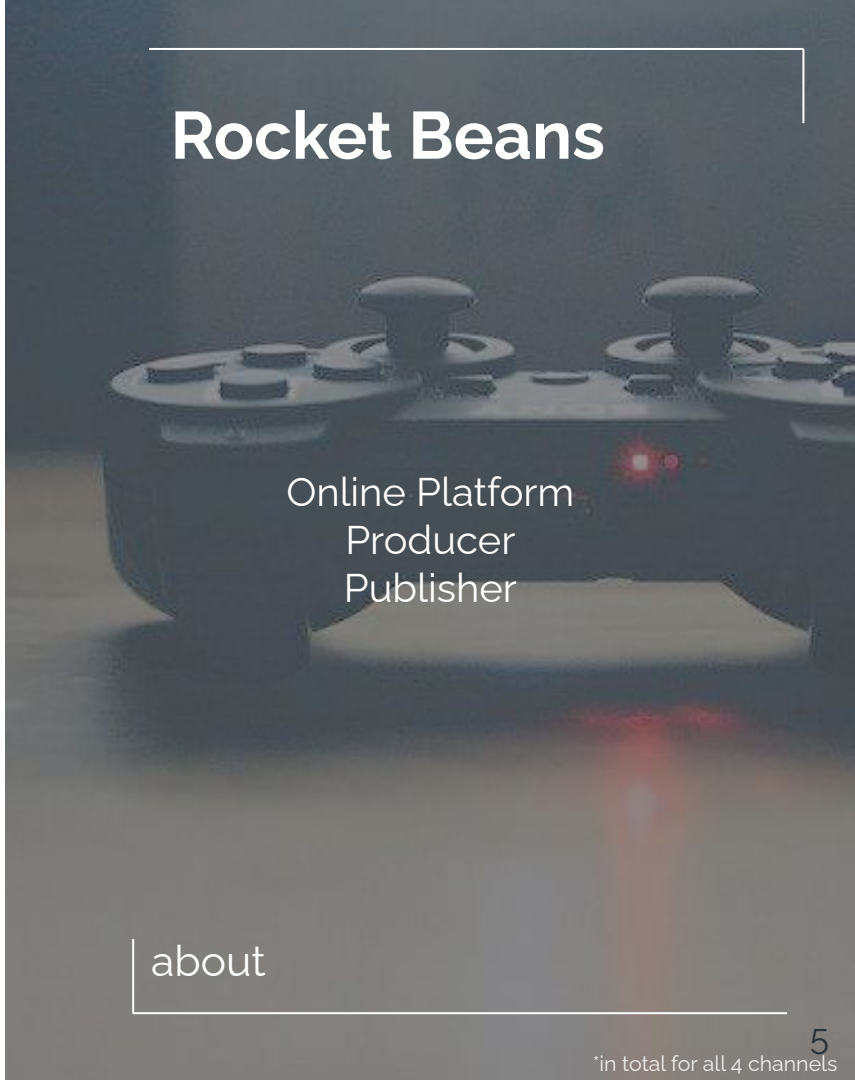
users

4 YouTube **channels**

over **1 million*** **YouTube** subscriber

Daily **content** about **gaming, digital pop culture and entertainment topics**

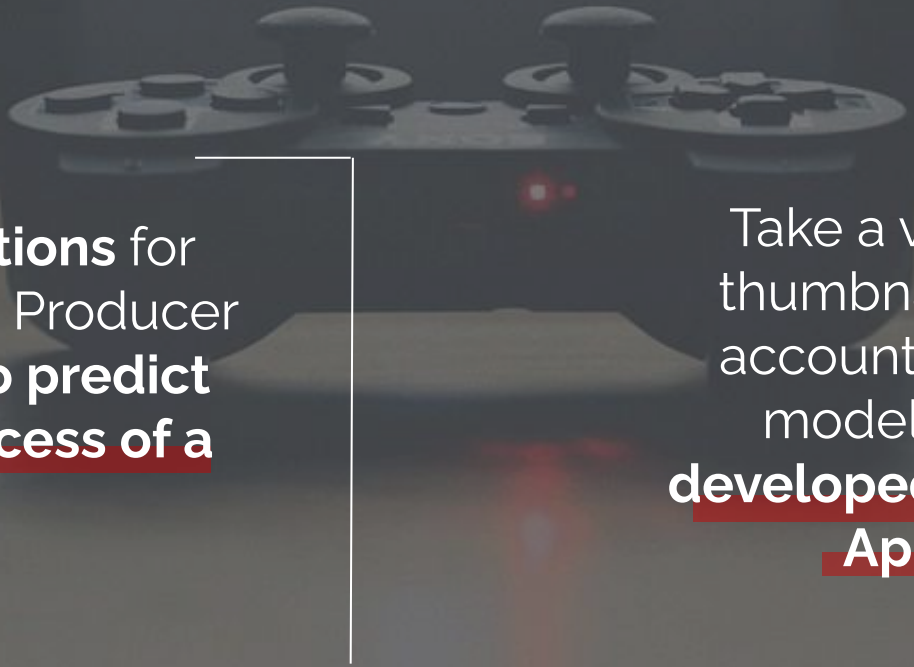
(rocketbeans.tv site | Twitch | YouTube)



*in total for all 4 channels

Business Case

Predict if a Youtube Video is going to be successful



Recommendations for Digital Content Producer and a **model to predict the future success of a video**.

Take a videos thumbnail into account in our model and **developed a WEB App**

METHODOLOGY & DATA SET

Data Set

Quantity of Data

5047 rows and 31 columns

Further features

engineered from the attributes present

Thumbnails

5047 images

YouTube API Web Scraping

for further Data

Data Set

Label

4 %

likes to views ratio

YouTube Features

Engagement Metrics

(Likes | Dislikes | Comments | Shares | Subscriber | Views)

Video Related Metrics

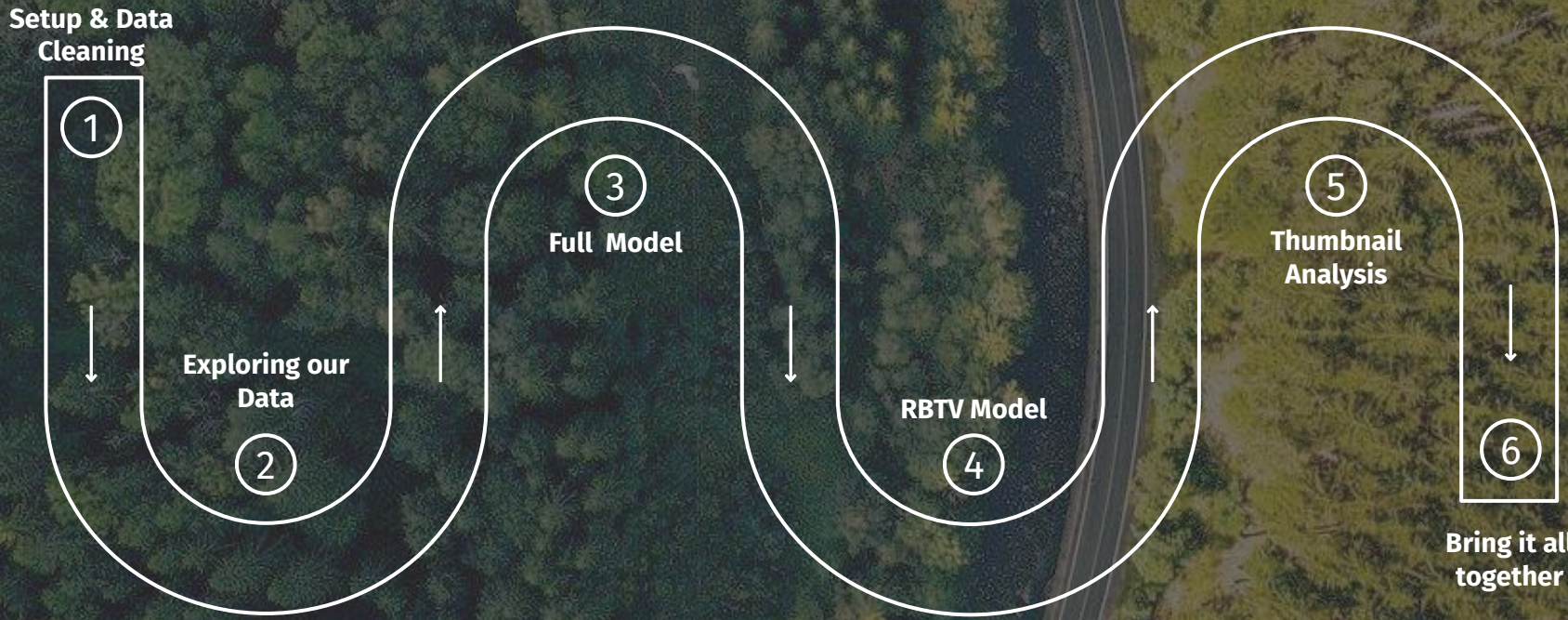
(Channel | Title | Publish Date | Average View Duration)

Rocket Beans Features

Internal Data

(Media length | Media game | Host)

Our Way


Setup & Data Cleaning
1
Exploring our Data
2
3
Full Model
RBTV Model
4
5
Thumbnail Analysis
6
Bring it all together

FINDINGS IN DATAFRAME

Facts & Figures

5047
Videos

109
Shows

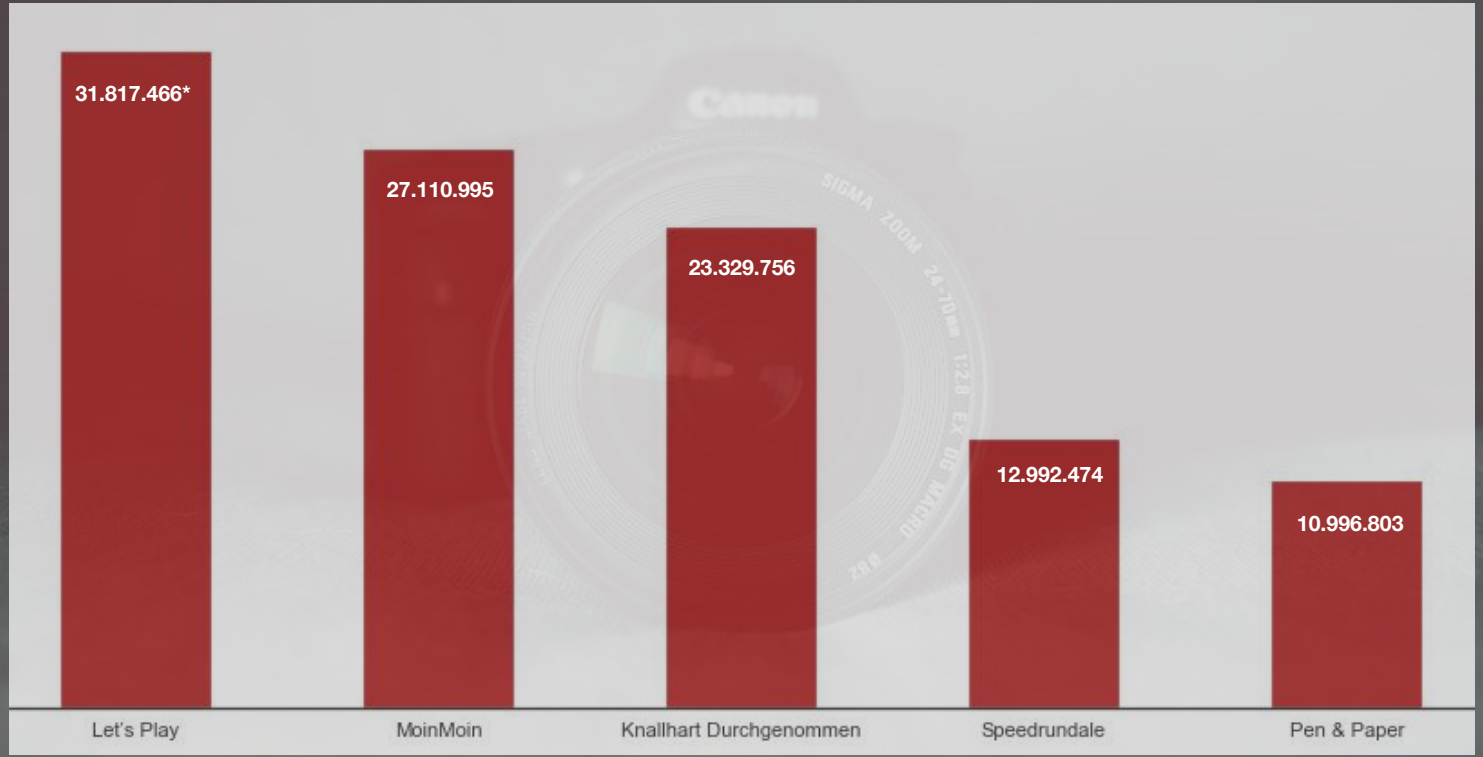
76 min.
average video length

35.810
average views

1.154
average likes

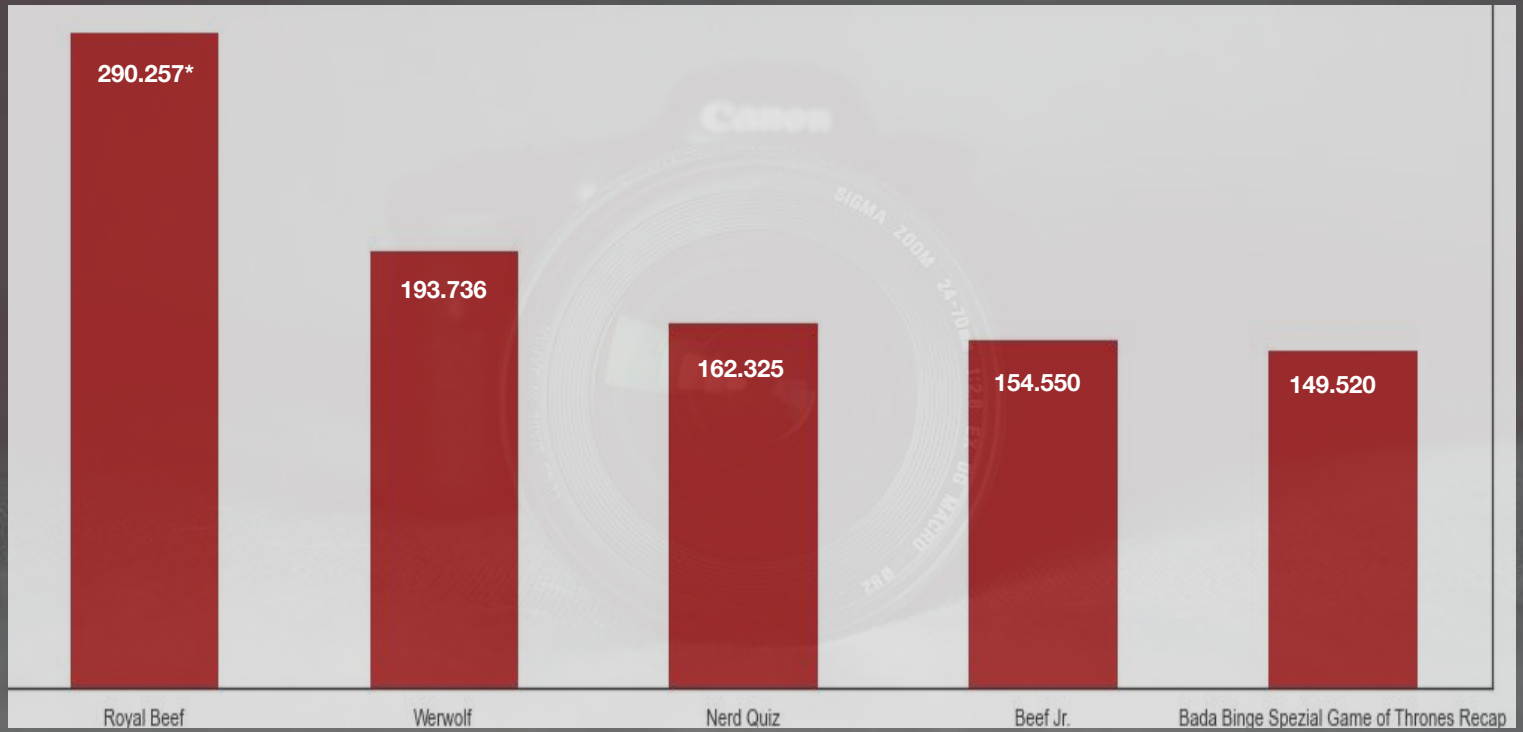
23 min.
average watch time

Highlights of all Rocket Beans Shows



*Total views

Highlights of all Rocket Beans Shows (average views)



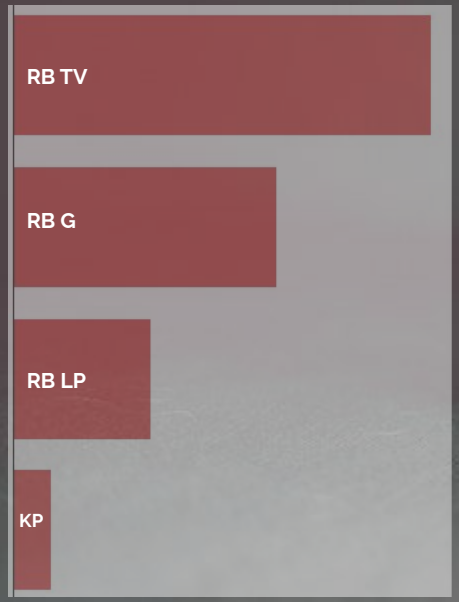
*avg views

Top 5 YouTube Videos with most views of all Rocket Beans Shows

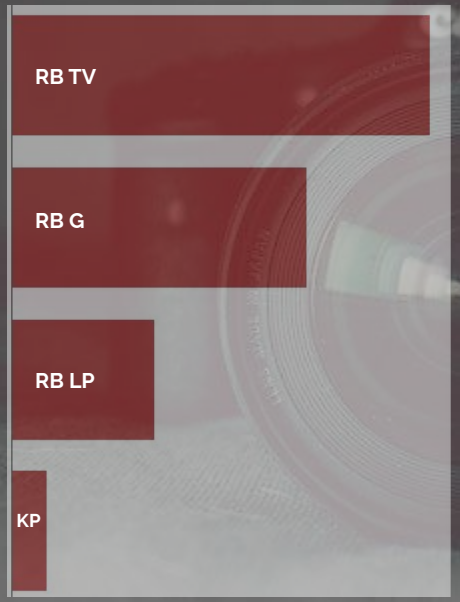
*Total views

Engagement per Channel

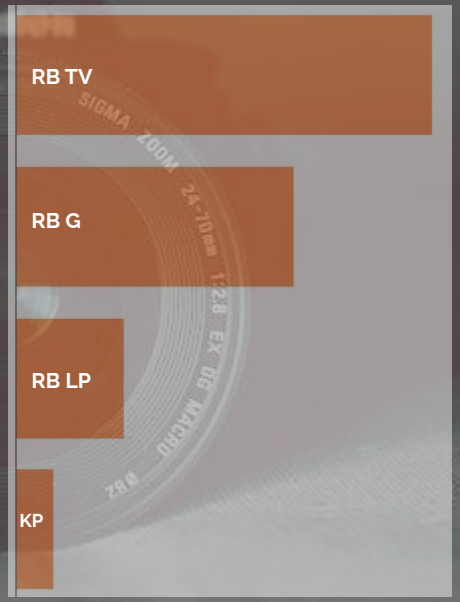
Likes



Dislikes



Shares



Comments

RECOMMENDATIONS

Recommendations

Take care of
words and punctuations
in title & description

Combination of
Hosts
matter

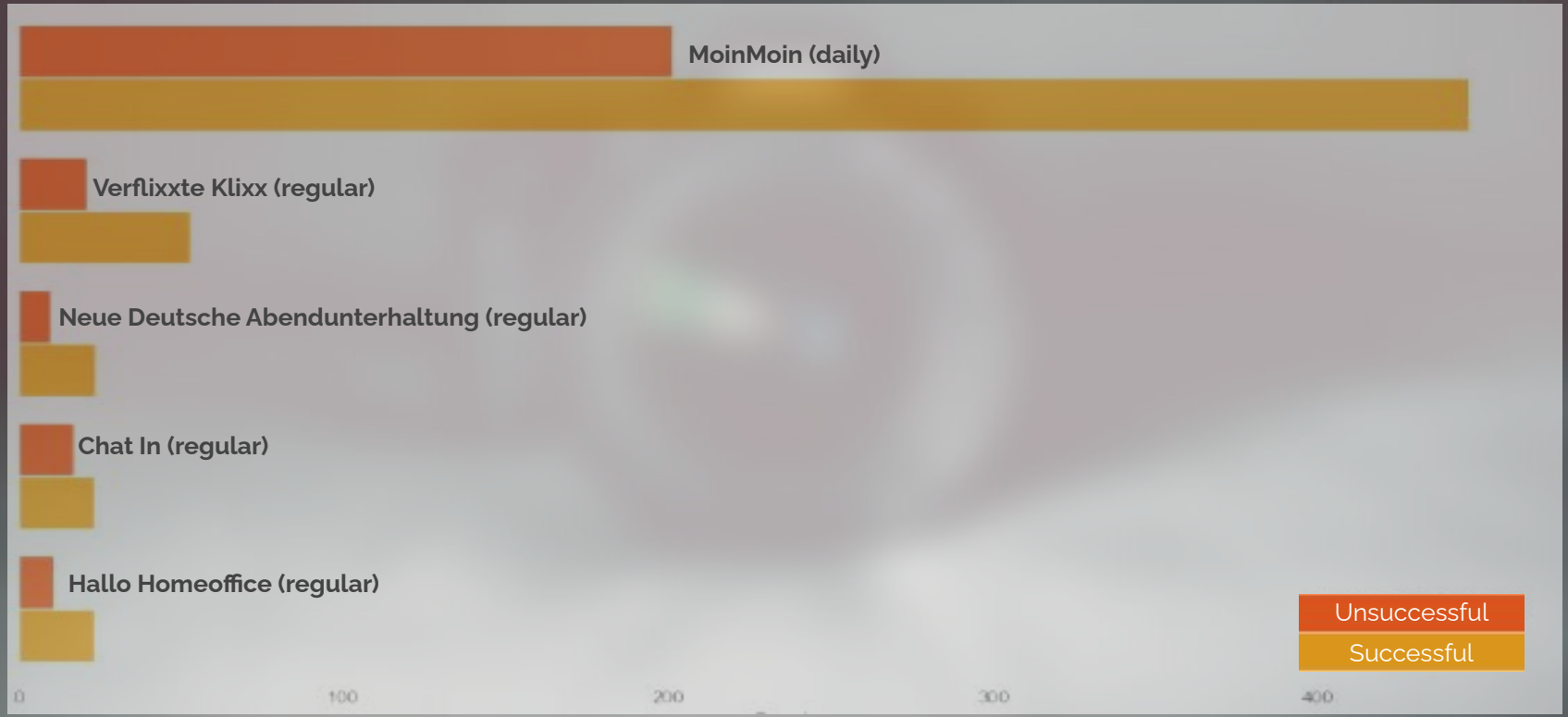
Keep
regular formats

Decide on a good
time to publish
your content

Video content
is important

RBTV Keep regular formats

Selected shows with good like/view ratio



The Combination of **Hosts** matters

 # 11,14,31,33,16,42

best combination of hosts concerning their views (936.455)

Most frequent host

in videos

Host # 33

(316 times alone / 994 in total)

Got

27.686 avg views

Mostly we do have videos with only

one host (1816 times)

Compared to other videos their are not so successful (24.596 avg views)

Take care of

Words and Punctuations in Title

A good video title could capture viewers attention instantly.

Titles with

20 words &

12 punctuations

are **most successful when it comes to average views** (51.023 and 634.545)

Take care of

Your Description Text

A good description text shouldn't be too short.

Descriptions with

168 words

are **most successful when it comes to average views** (727.310)

Publish Time

Decide on a good time to publish your content

Most videos were published on a

Thursday (955)

Sunday

(59.915 avg. views)

Is the best day to publish
followed by Saturday & Friday

Most videos were published at

8 pm but videos which were

published at **8 am** got the

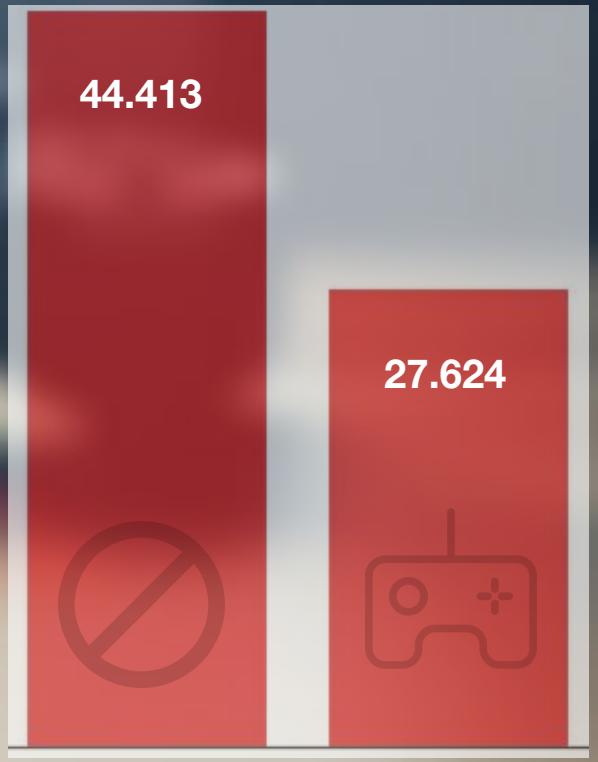
most views (74.084 avg.)

Game related or not

Video content is important

If a video is **not game related**

we do have one and a half times more views in average.



THUMBNAIL SUCCESS PREDICTION

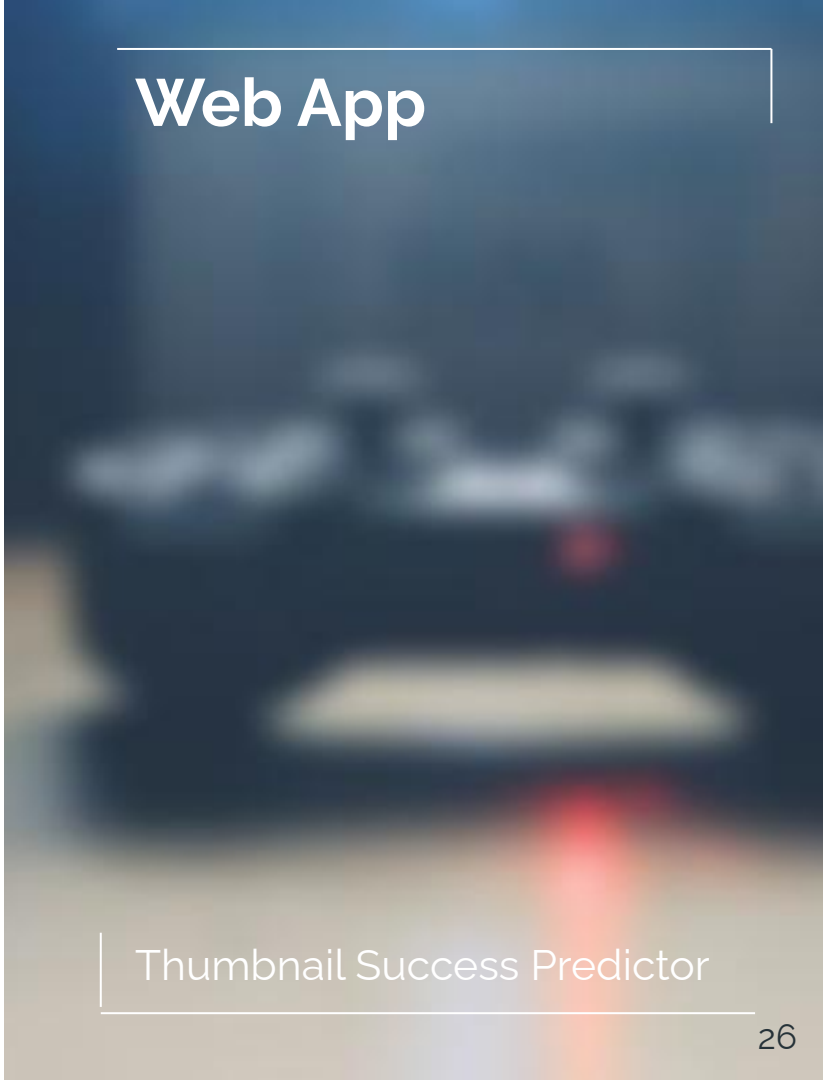
Web App

CNN model
trained on 3000 images
for classification if a

thumbnail

will lead to a **successful**
video **or not.**

Let's take a look at the
Web App.



Thumbnail Success Predictor

ACHIEVEMENTS
&
FUTURE WORK


87% ACCURACY
for our Regression Model
75 % ACCURACY
for our CNN
1 WEB APP
for prediction

Future Work

- Impressions
- Revenue Metrics
- Comment Analyzing
- Further analysis of individual host-importance
- Improve our CNN (more images)

THANK YOU

for your attention

Any questions?

APPENDIX

Tools

Python / Pandas / NumPy / scikit-learn / Matplotlib / Seaborn / Plotly / Logistic Regression / AdaBoost / XGBoost / Random Forest / Support Vector Machines / CNN / TensorFlow / Keras / YouTube API / YouTube Wrapper / Spacy / HTML / Flask / Git

Greatest Challenges

At the beginning:

- just like it is in real world business, you don't always have all the information immediately - we got the dataset little by little
- We solved it just by starting (working with what we've got)

During the project:

- Finding a business case out of the data
- Data Leakage (all YouTube features)
- Curios to start with text analyzing - realization that most NLPs are for englisch text structures
- Unfortunately too few images for our SVM/CNN (need to improve it to better detect successful images)

At the end:

- Taking a step back - bring it all together

What we've learned

- Even if the topic or business case is not what you are normally interested in or spending your free time with - **you can be interested in anything**. Dive deeper into the topic and you'll find so many interesting things and gain your knowledge
- **If you believe you can do anything** - never thought to build a CNN or a short Web App with HTML
- **Think outside the box** - data leakage is annoying but you can solve it
- **Be pragmatic** - just make assumptions concerning your business case
- **More routine** in working with python and pandas and going through an EDA
- **Finding an end** - every idea leads to another so you can't discover everything (80 / 20)

Image Sources

https://pixabay.com/de/photos/video-controller-videospiel-336657/

https://pixabay.com/de/photos/lager-handel-monitor-gesch%C3%A4ft-1863880/

https://pixabay.com/de/photos/kamera-digital-fotografie-isoliert-819359/

https://pixabay.com/de/photos/feuerwerk-menschen-festival-nacht-4768501/