



Team 4: Joana, Kevin, Gala

Final Presentation

SENTIMENT ANALYSIS

Bayer Leverkusen Players'
sentiment trends extracted
from online News



Motivation



SENTIMENT


- Fan attraction
- Sponsor attraction
- Motivation for player & team
- Inform talent scouting process



- **Reproduce good sentiment**
- **Counteract bad sentiment**
- **Support players** with high media attention



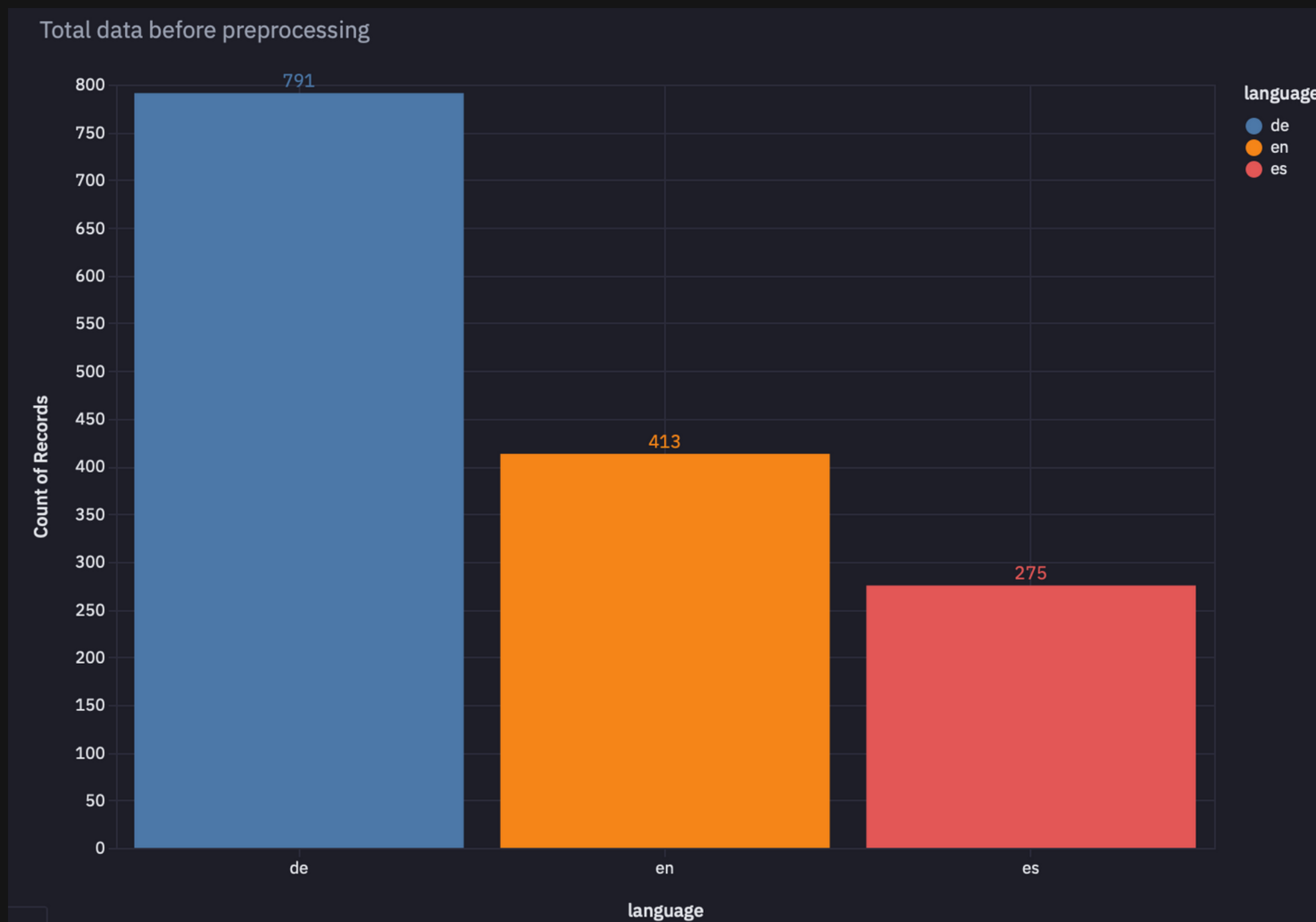
TOPICS

- Fan attraction
 - Sponsor attraction
 - Inform talent scouting process
 - Inform line-up decisions
- 
- Counteract **blind-spots** in news coverage
 - Build a **positioning strategy**





Data - before preprocessing



Total news 1479

German 791

English 413

Spanish 275



Minimum length: 15

Maximum length: 1 578 671

Average length: 10 325.85

Row 403 from the english dataframe



remove similar rows per
player
lower case
delete content patterns
match emojis to
description
remove further noise
(numeric, links, ...)
DATA CLEANED

(remove punctuation)
remove short words
define words to keep &
delete stopwords
extract paragraph
DATA CONDENSED

hero with his stunning winner against Bournemouth. That sound you can hear is Everton fans
 their hopes of extending their incredible top-flight active streak to 70-straight seasons were in
 Goodison Park wild as Bournemouth pushed hard for an equalizer late on but Everton held on
 his 17 games in charge to keep their pride top-flight history intact despite growing turmoil
 \n\n\n\n10 things we learned in the Premier League – Week 38 \n\n\n\n\n\n\n\n\n\n\n\n
 th five at the back from the start and it was too negative. Dyche got it right though. All they
 wasn't a brilliant game. But Everton and Dyche don't care. With loads of injuries they found a
 break as they will be playing at England's top table next season, their 70th consecutive at the
 him the tools to rip things up and rebuild the squad. Just staying up like this can't happen
 size should never be in this situation.\n\n From here at Goodison Park, this is the moment
 ay 28, 2023Sean Dyche: Everton proved they aren't a broken clubJordan Pickford: Toffees
 scored a beauty to be the hero and was brilliant in midfield.Mark Travers: Made some really
 ing Bournemouth chucked at Everton late on.Graphic via FotMob.comHow to watch Everton
 n live analysis! – By Joe Prince-Wright at Goodison Park\|xa0THEY HAVE DONE IT! EVERTON
 t the win they needed against Bournemouth to save themselves from a first-ever Premier
 PW_NBCSports) May 28, 2023SAVE! Jordan Pickford makes a fine save to deny a volley from
 ping deep. Bournemouth having a go. After Pickford was down for a wild getting treatment,
 offees ahead.\n\nABDOULAYE DOUCOURE BLOWS THE ROOF OFF AT GOODISON PARK! \n :
 icked on and it falls to Demarai Gray 6 yards out. He heads it straight at Travers. Then
 0-0 Bournemouth – Half time here at their are loud boos from the home fans.SAVE! Mark
 ff the scoresheet just before the half!\n\n : @USANetwork & @peacock | #EVEBOU
 some poor Everton defending. Excellent from David Brooks there.You get the sense things
 n the box. The atmosphere is very, very quite here.First, as it stands of today – Leicester City
 going down!" from the Bournemouth fans in the away end. Everton's fans are stunned.BIG
 ed into the box but there isn't an Everton player anywhere near it. Huge roars of discontent
 e positions.Demarai Gray with a half chance as he wriggles free but his shot/cross is cleared
 me good chances to whip it in from the right but the quality hasn't been there on the final
 ZQaeTyAs \nMy thoughts on the scene here for Everton v Bournemouth ↪ #EVEBOU
 n fans just want to get this started. This has to be excruciating for them. Meanwhile in the
 hear a wall of noise approaching the stadium from every angle. Great atmosphere building
 njured Dominic Calvert-Lewin. Looks like Coady is at right back and McNeil is at left back.
 ng, the sun is shining and the flares are plentiful! The equation is simple for Everton: win and
 angling, the sun is shining and the flares are plentiful. \n\nAn intense, and pivotal, day awaits
 Wright (@JPW_NBCSports) May 28, 2023Key storylines & in-form players to watchEverton
 er on the final day. Defensively they have issues at full back but they have hung in there and
 c Solanke and Dango Ouattara at the other end of the pitch as there has been a really nice
 3-0 to spark ugly scenes in the away end as the Everton fans turned on their players and
 hilles), Andros Townsend (knee), Dominic Calvert-Lewin (thigh), Nathan Patterson (thigh),
 urnameouth team news, injuries, lineupOUT: Antoine Semenyo (leg), Ryan Fredericks (calf),
 rsonal reasons Billing fit and back in to start Ouattara also in the XI \n\nOur line up for
 ming schedule – Nations League, friendlies, Gold Cup \n\n\n\n\n\n\n\n\n\n\n\nFolarin
 nThat sound you can hear is Everton fans exhaling.\n\nFULL MATCH REPLAY\n\nAt half time
 ight active streak to 70-straight seasons were in real danger. Goodison was full of boos and
 hed hard for an equalizer late on but Everton held on and got the win they needed to stay
 noil behind-the-scenes at the club.\n\nPremier League news\n\n\n\n\n\n\n\n\n\n\n\nPremier
 n Premier League 2022-23 season recap: Video highlights, results, analysis\n\nThey were sent
 fielder and it did in the second half as Doucoure hammered home. It was a brilliant finish in
 I never been relegated from the Premier League and only Arsenal have a longer active top-

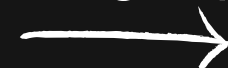


Data - for topic clustering

DATA RAW

{'content': 'Bayer Leverkusen muss im Kampf um den Einzug ins Halbfinale der Europa League auf den argentinischen Weltmeister Exequiel Palacios verzichten. Der Mittelfeldspieler, der beim 1:1 im Viertelfinal-Hinspiel gegen Union Saint-Gilloise mit einer Oberschenkelverletzung zur Pause ausgewechselt worden war, wurde nicht rechtzeitig für das Rückspiel in Brüssel fit. Er steht nicht im Kader. Am Mittwoch war Palacios ins Training zurückgekehrt.\nAmine Adli, der das Training am Mittwoch vorzeitig beenden musste, sitzt zumindest auf der Bank. ...

paragraph



kampf einzug halbfinal europaleagu
argentinischen weltmeist verzichten
mittelfeldspiel viertelfin hinspiel
union saint gillois
oberschenkelverletzung paus
ausgewechselt worden war wurd
rechtzeitig ruckspiel brussel fit steht
kader mittwoch train zuruckgekehrt



Data - for sentiment

DATA RAW

{\content\': \Football\\nFlorian Wirtz\\'s goal for Bayer Leverkusen in the Europa League represents the next step on the road to recovery for the German teenager; Adam Bate was in the BayArena to see it and discuss the player\\'s vast potential with sporting director Simon Rolfes\\nComment and Analysis\\n @ghostgoal\\nFriday 17 February 2023 00:01, UK\\nIt was the moment that Florian Wirtz had been waiting for, the moment that the BayArena had been waiting for, and when it came it was special. The first goal for Bayer Leverkusen's wunderkind since his long-term knee injury lit up their game against Monaco.\\nWirtz touched the ball nine times in scoring it. Ten if you included the toe-poked volley to team-mate Exequiel Palacios in the build-up. But those nine touches - the ball never moving more than a metre from his boot until the finish - showed why he is so special.\\nThat quick shift of the ball from one foot to the other had been seen minutes earlier when his previous dribble had been the catalyst for Moussa Diaby\\'s equaliser. There is a calmness to his game, an ability to glide without breaking stride, that only the best possess.\\nIt felt cruel on Wirtz that it was not the winne...

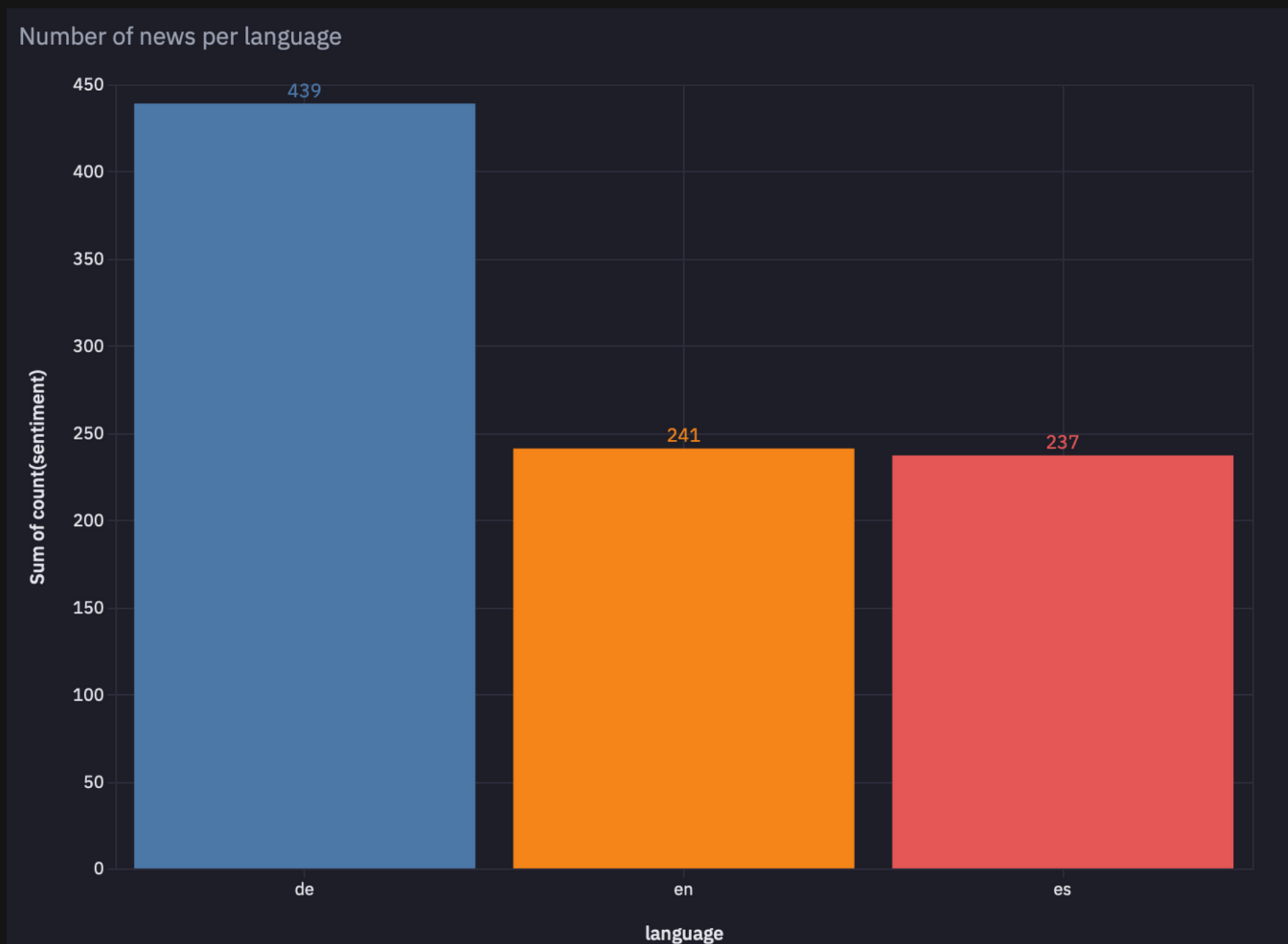
sentence



ten if you included the toe
poked volley to team mate
palacios in the build up



Data - after preprocessing



*Data used for sentiment

Total news 1479

German 791

English 413

Spanish 275



Total news: 917

German: 439 (47%)

English: 241 (42%)

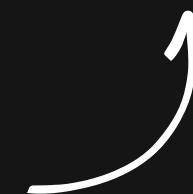
Spanish: 237 (13%)

Overall loss: 36.19%



Approach - text clustering

on DE data



APPROACH

Clustering (KMeans) and topic modeling (LDA)
Cluster and topic size 2-15

CHALLENGE

Small Clusters → very general not precise
Many clusters → very specific and topics double

RESULT

Cluster size of 8 and aggregate into 3 central topics

Bundesliga News

Game Situations
and Match
Performance

Europa League





Game Situations and Match Performance

[illegible]



Approach - sentiment

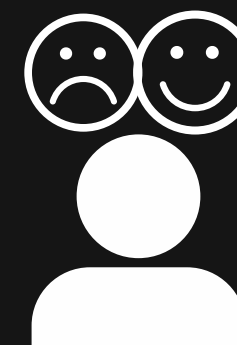
MODEL USED

- Selected model

nlptown/bert-base-multilingual-uncased-sentiment

- multilingual & uncased
- BERT based
- Neural network layers
- The model analyzes the text's structure, meaning, and give a sentiment score: 1 to 5

on DE, EN, ES data





Approach - Wrap up

Preprocessing per language
Take relevant data

WORD FREQUENCY

- Normalization: Stemmatization
- Representation: TfidfVectorizer

SENTIMENT ANALYSIS

- Normalization: Tokenization
- Representation: Transformer-based
- Models: multilingual & BERT per language on different datasets

TEXT CLUSTERING

- Normalization: Stemmatization
- Representation: TfidfVectorizer
- Label data
- Models: KMeans, LDA

Evaluate accuracy
Create Dashboard in HEX, including matchdata

WHY WAS IT CHOSEN?

- requirements from news data
- three languages
- interest in text clustering from Bayer04
- limited time



Results - overall

HEX



Performance evaluation - sentiment

	f1-score	support
DE	0.57	30
EN	0.47	30
ES	0.43	30
Overall	0.49	90

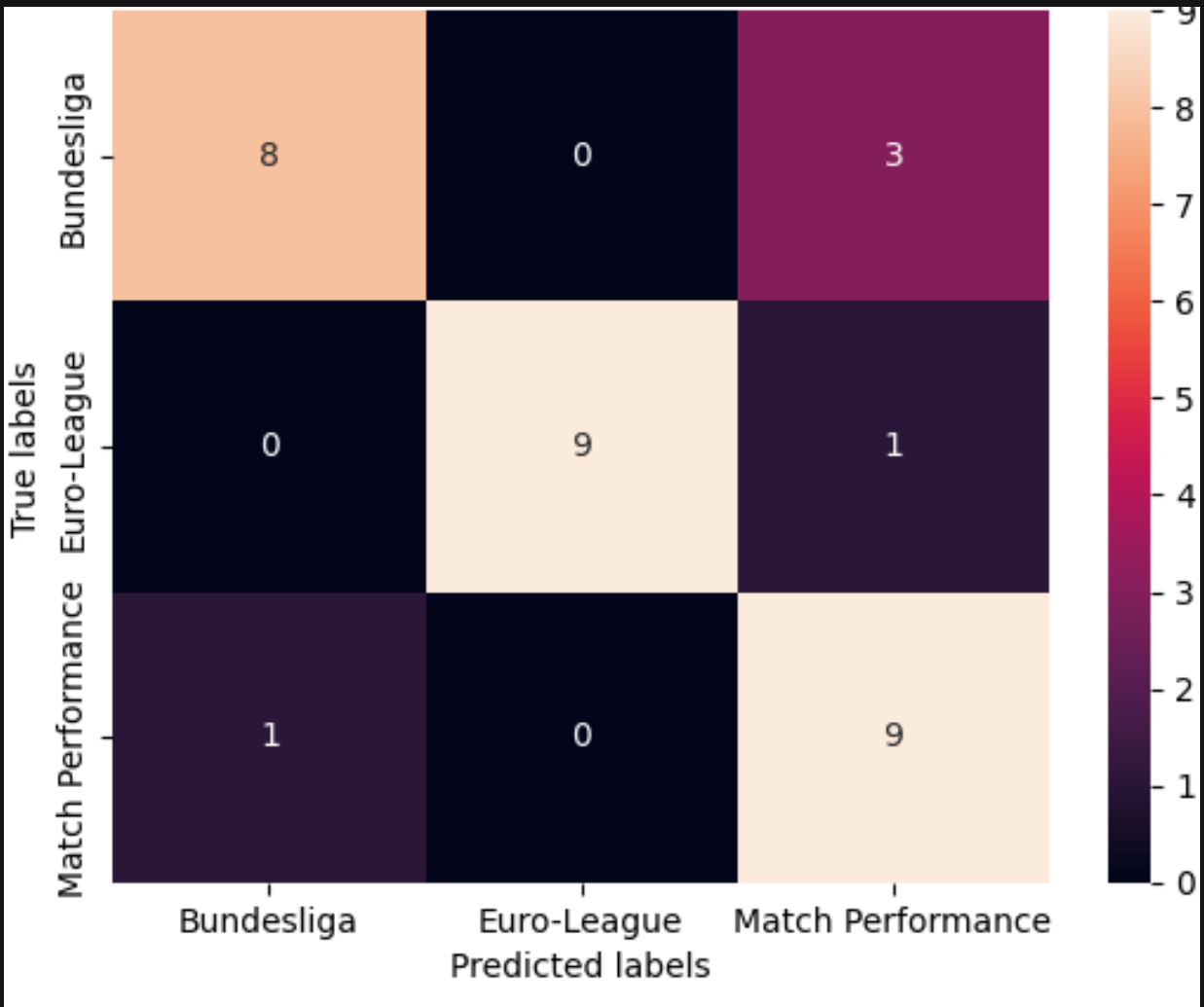
Manual labeling

Mapping scores changed

	precision	recall	f1	support
negativ (1)	0.41	0.58	0.48	24
neutral (2-4)	0.52	0.73	0.61	33
positiv (5)	0.60	0.18	0.28	33
Accuracy and overall			0.49	90



Performance evaluation - text clustering



Manual labeling

	precision	recall	f1	support
BL-News	0.89	0.73	0.80	11
Europa-Leage	1.00	0.90	0.95	10
Situation & Perf.	0.69	0.90	0.78	10
Accuracy and overall			0.84	31



*Results form kMeans



Next Steps for Bayer04 - sentiment

IMPROVE MODEL PERFORMANCE

- Fine tune model
- Try different models e. g. VADER or NewsSentiment
- Combine multiple sentiment models to create an ensemble

IMPROVE EVALUATION RELIABILITY

- Calculate confidence intervals
- see for both





Next Steps for Bayer04 - text clustering

- Perform for all languages
- Try different normalisation e.g. Tokenization
- Try different embeddings
- Try different models e.g. Transformers
- Look at overlapping clusters/topics
- Gather more specific topics

FOR BOTH:

- Get more and better data from the API
- Balance out the classes
- Pre-Label more datapoints in order to archive more accurate accuracy scores





Achievements in respect to motivation



SENTIMENT

- Fan attraction
 - Sponsor attraction
 - Motivation for player & team
 - Inform talent scouting process
- ↓
- **Reproduce good sentiment**
 - **Counteract bad sentiment**
 - **Support players** with high media attention



GENERAL: NUMBERS PER ...

SENTIMENT



Having a good sentiment performance in all three languages



Showing differences between players



Showing changes over time, specifically related to matches



Achievements in respect to motivation



- Fan attraction
 - Sponsor attraction
 - Inform talent scouting process
 - Inform line-up decisions
 - Counteract **blind-spots** in news coverage
 - Build a **positioning strategy**
- ↓

TOPICS

- ☑ Identifying text cluster in German data
- ☑ Showing differences between players
- ☑ Showing changes over time, specifically related to matches
- ☑ Comparing topics and sentiment
- ✂ Get more specific clusters to have more actionable insights



**THANK YOU
FABIAN AND FELIPE
FROM BAYER 04!**



THANK YOU!

Questions and Feedback are welcome

Team 4: Joana, Kevin, Gala