



FAKULTÄT FÜR
INFORMATIK

TRIWIZARD: Trend Recognition in the WissZeitVG

Project Co-Ordinator: Dr. David Broneske, Saijal Sahania

Team Members: Moinam Chatterjee, Krupa Kapadia, Behnam Ensan, Pravin Pandey

What is WissZeitVG & Why #IchBinHanna ?

- ❖ A federal law known as the "Wissenschaftszeitvertragsgesetz" (WissZeitVG) was established in 2007, setting up the basis for fixed-term employment in higher education.
- ❖ This regulation establishes a maximum period of fixed-term employment for academic staff members in German universities to 6 + 6 years.
- ❖ #IchbinHanna is a spontaneously emerging Twitter trend that responds to an explanatory video by the Federal Ministry of Education and Research on the Science Fixed-Term Contract Act (WissZeitVG).



Motivation

- ❖ To us as a team the project comes at the right time as we continue in our journey in the world of Data Science. We feel that the project offers a unique opportunity to work on real-world data.
- ❖ We are motivated to be able to use our theoretical knowledge of Data mining and text analysis here.
- ❖ The topic of the project is very close to our hearts as it affects us directly with the available opportunities in the scientific world and the corporate world.

Research Questions

Q1 What are the prominent hashtags in the tweets extracted?

- ❖ What is the concurrency between the hashtags?

Q2 What latent topics are found in the tweets?

- ❖ How do the topics change over time?

Q3 What are the topics talked about in each hashtag?

- ❖ How the topics in the hashtags change over time?

Methodologies

- ❖ Python data visualization libraries matplotlib and seaborn to perform exploratory data analysis and visualize data.
- ❖ Tweepy, a python package to communicate with Twitter V2 API to extract tweets.
- ❖ Topic modelling methodology to discover latent topics of discussion in the extracted tweets.
- ❖ Spatio – Temporal Clustering methodology to discover trends in topics discussed over time.



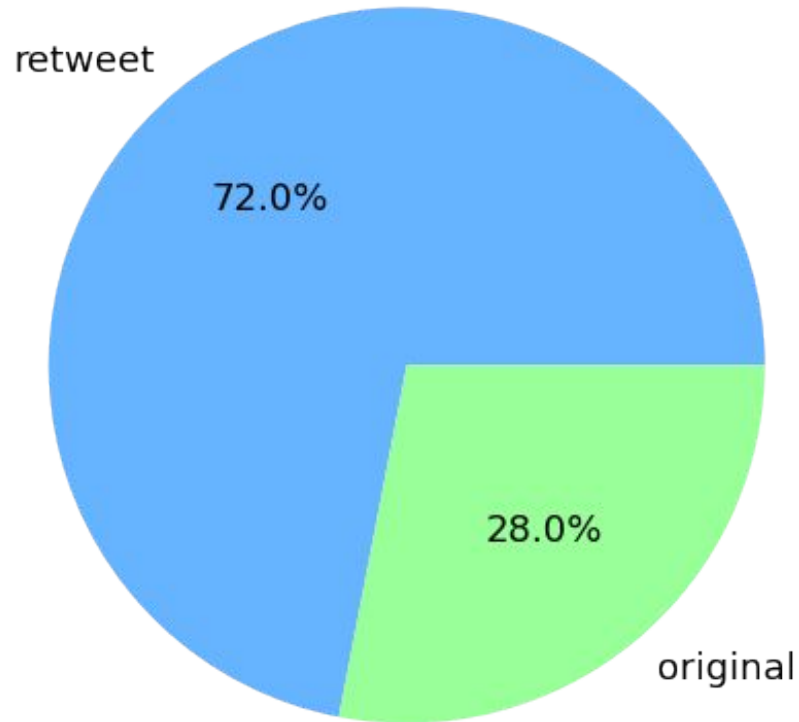
matplotlib

seaborn

Exploratory Data Analysis based on hashtags



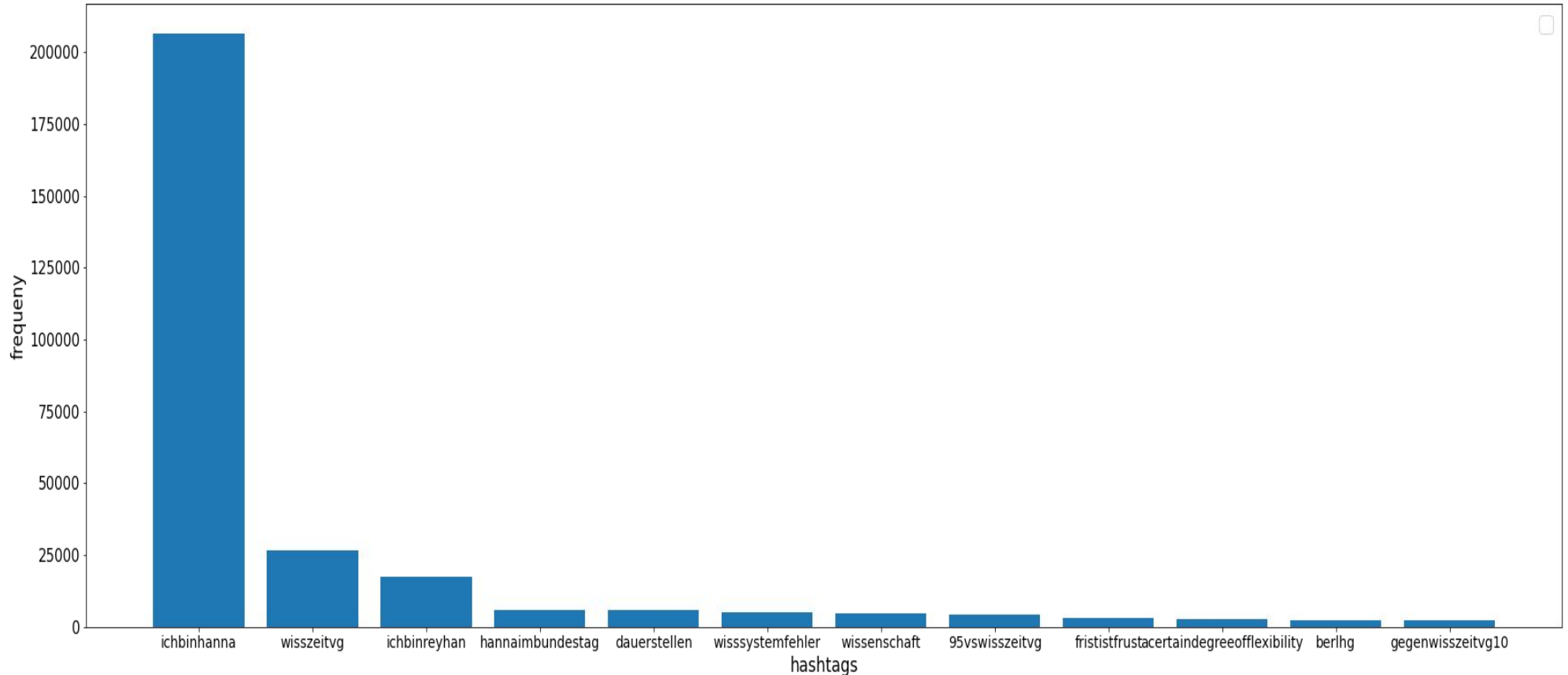
Original Tweets and Retweets for hashtags



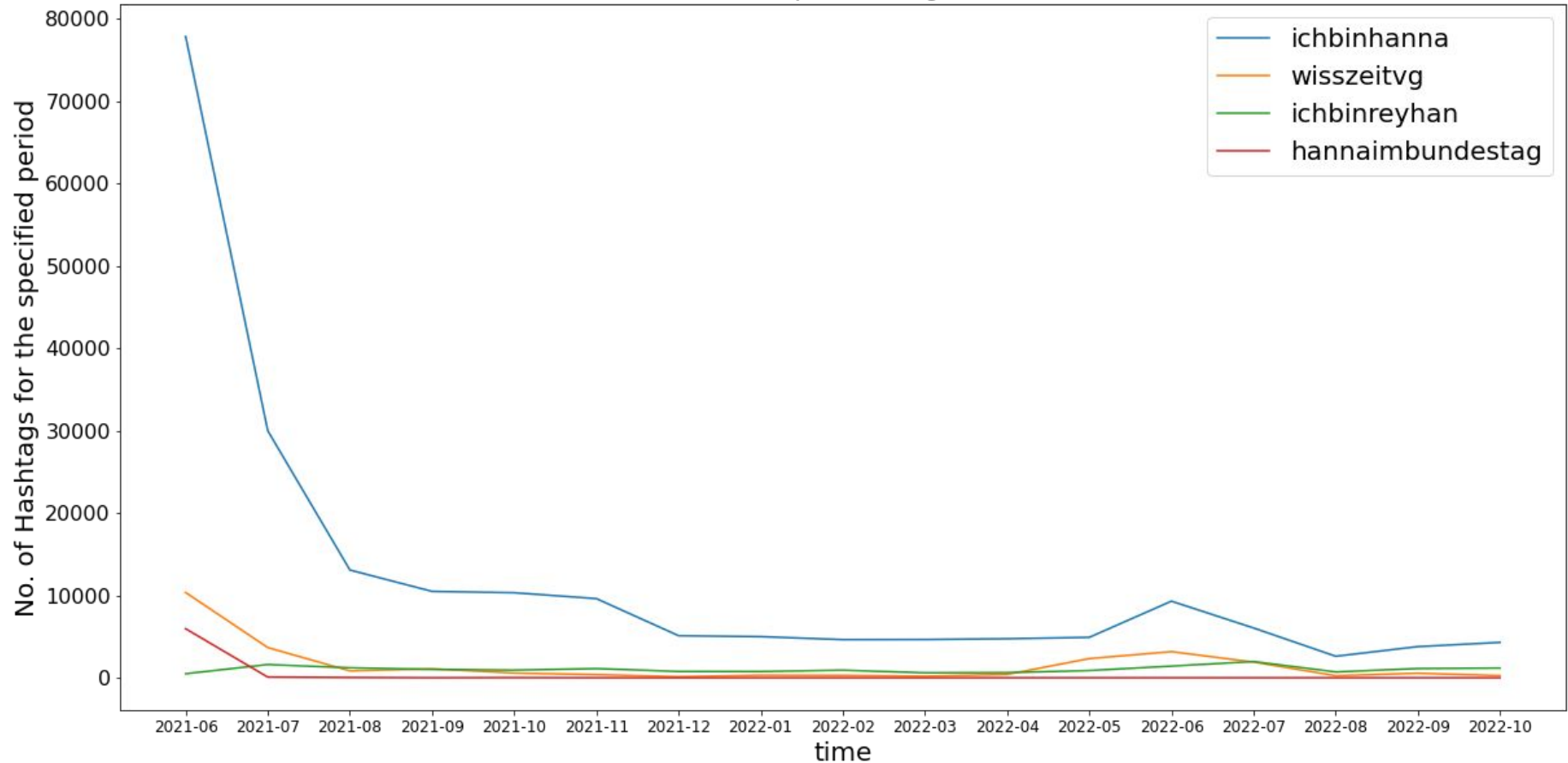
Most frequent Hashtags (Threshold: 2000 Tweets)



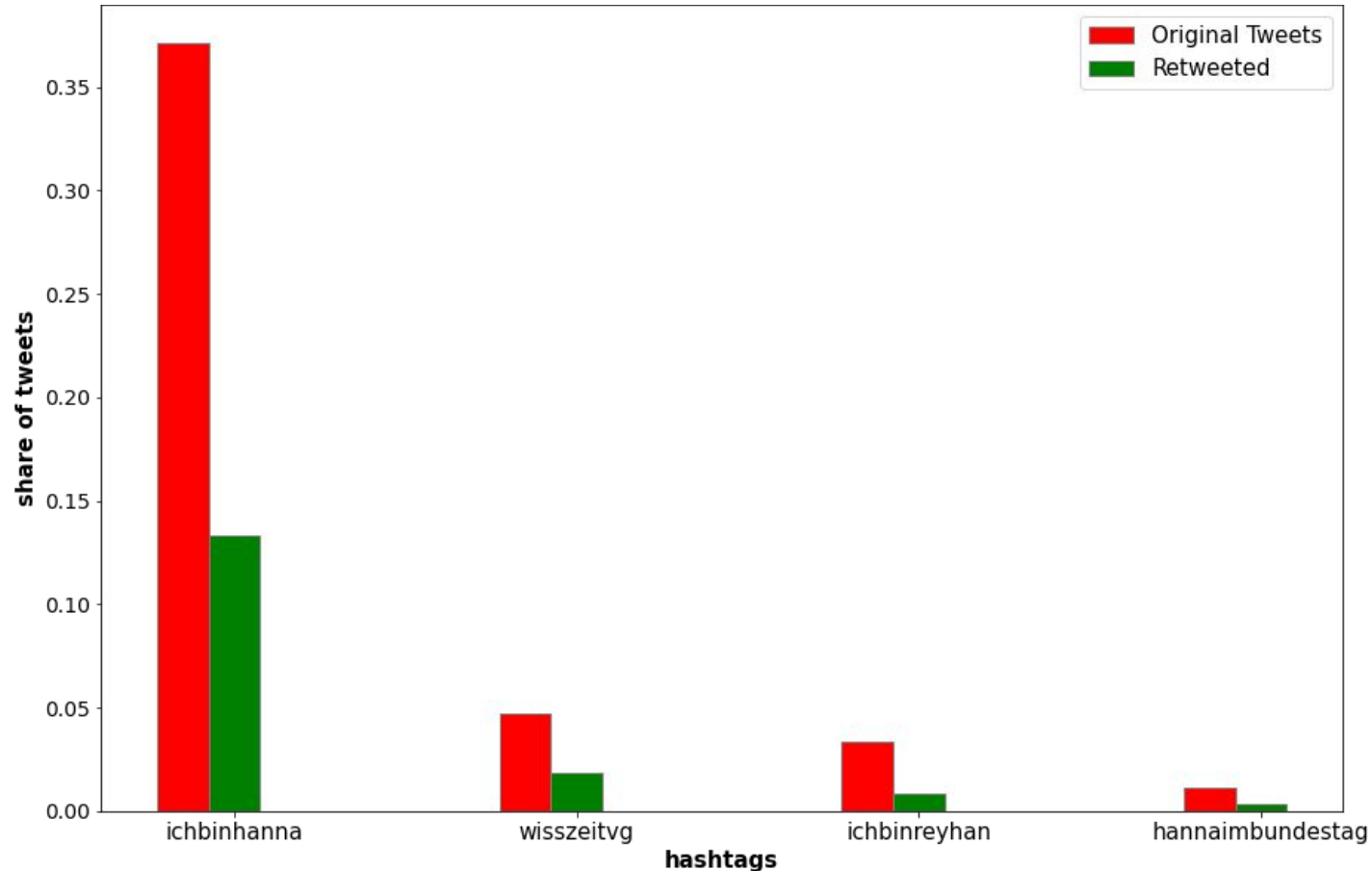
Frequency of most frequent Hashtags



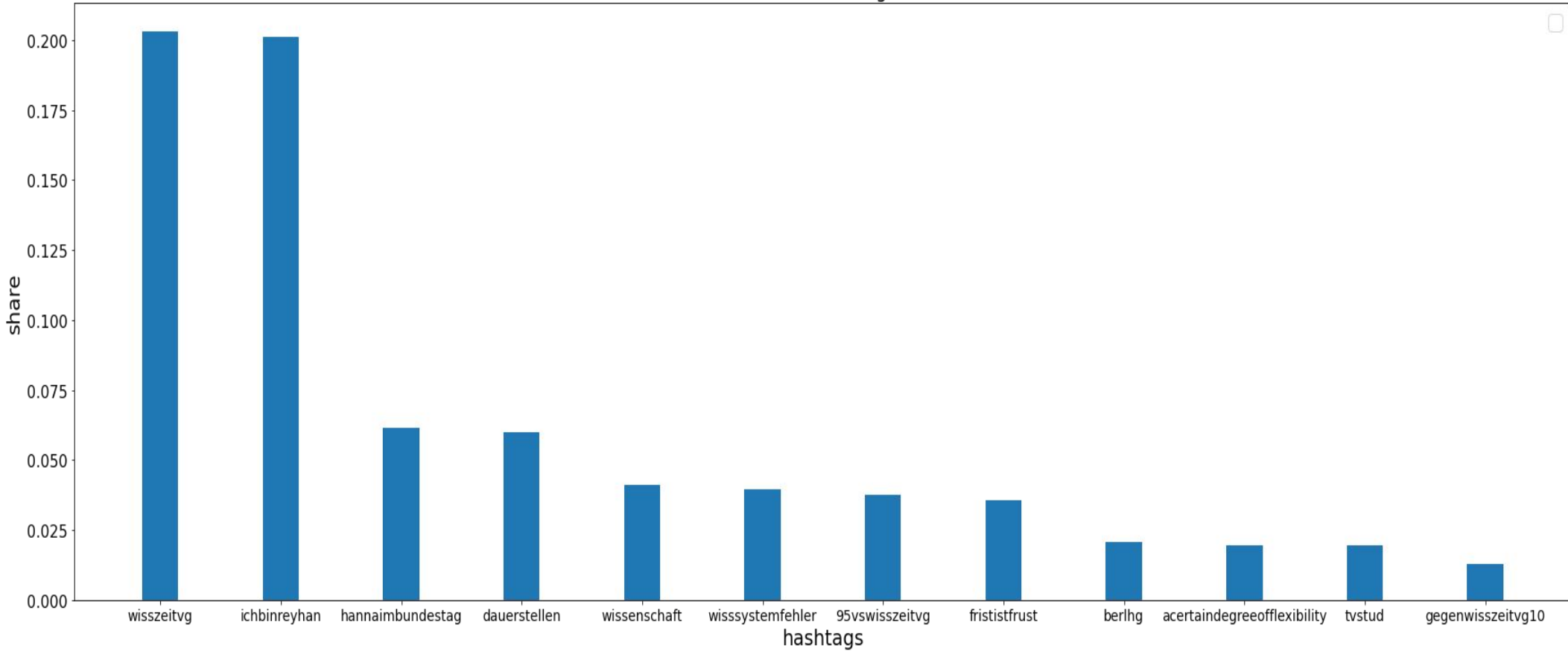
Evolution of frequent hashtags over time



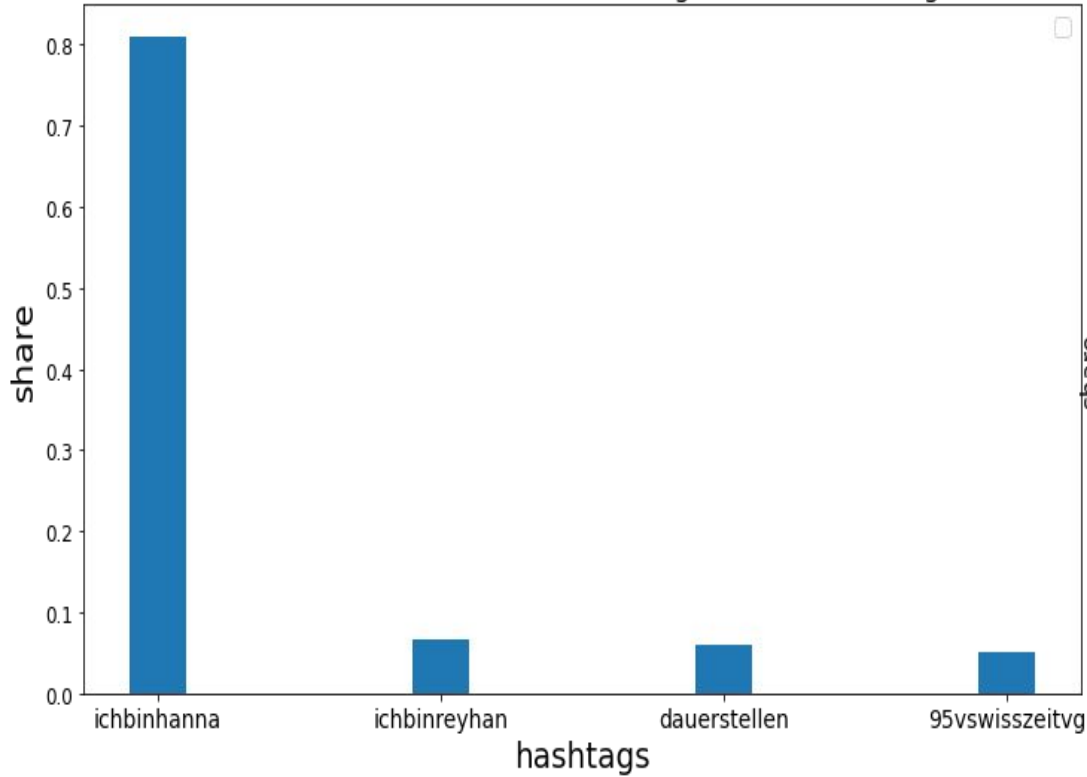
Share of original and retweets with respect to hashtags



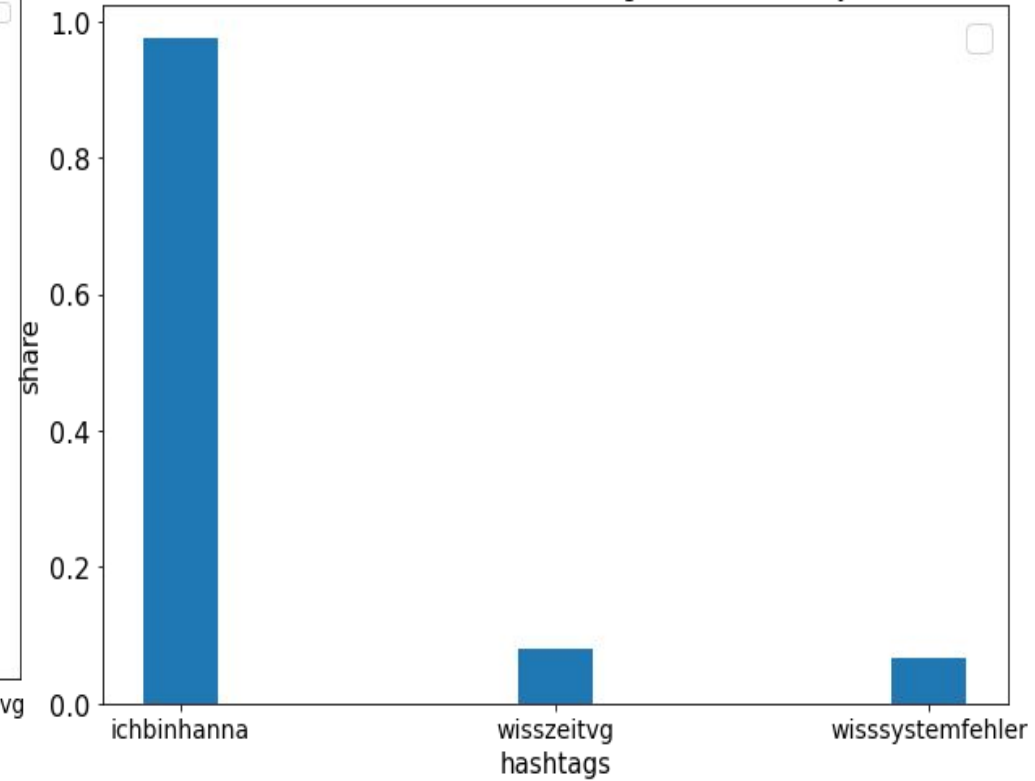
share of most co ocurrent Hashtags with #ichbinhanna



share of most co ocurent Hashtags with #wisszeitvg



share of most co ocurent Hashtags with #ichbinreyhan



THANK YOU
Any Questions ?