

# A study on Channel Popularity in Twitch

Ha Le, Junming Wu, Louis Yu and Melissa Lynn

Gustavus Adolphus College

## Introduction

- 🎮 Twitch is the most popular streaming platform made for competitive streaming and eSport.
- 📢 We want to identify factors that contribute to channels' popularity and growth on Twitch.
- ⚙️ Based on the factors found, we use a machine learning model to predict the growth rate of channels.

## Data Collection

**Step 1:** Generate 10 000 users that were actively streaming on Twitch.

**Step 2:** Monitoring 4000 users per week and repeat the process for a month.

- > Gathering data from each users every 30 minutes.
- > If the user was actively streaming, we recorded the timestamps, # of current followers/viewers, streaming content, language.
- > If the user was not streaming, we only recorded the timestamps and # of current followers.

**Step 3:** Additional data were collected, such as the frequency of videos and clips posted after the streams.

## Average Followers Gained During Streams

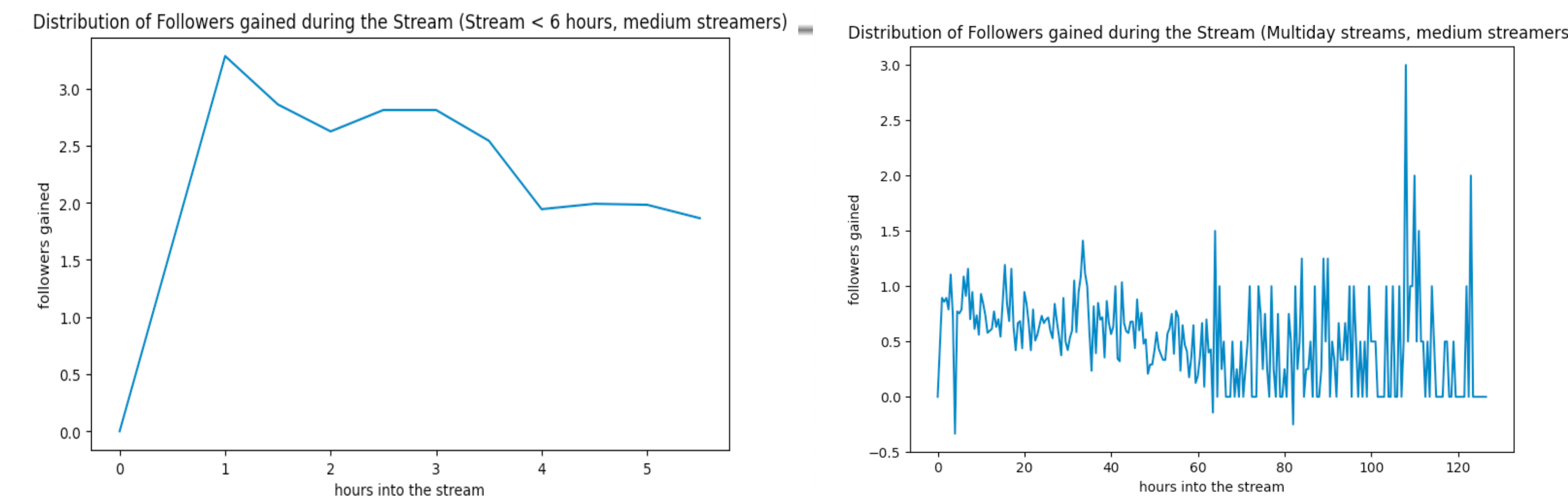


Figure 1: Streams < 6 hours

Figure 2: All Streams

## Distribution of the Length of Streams

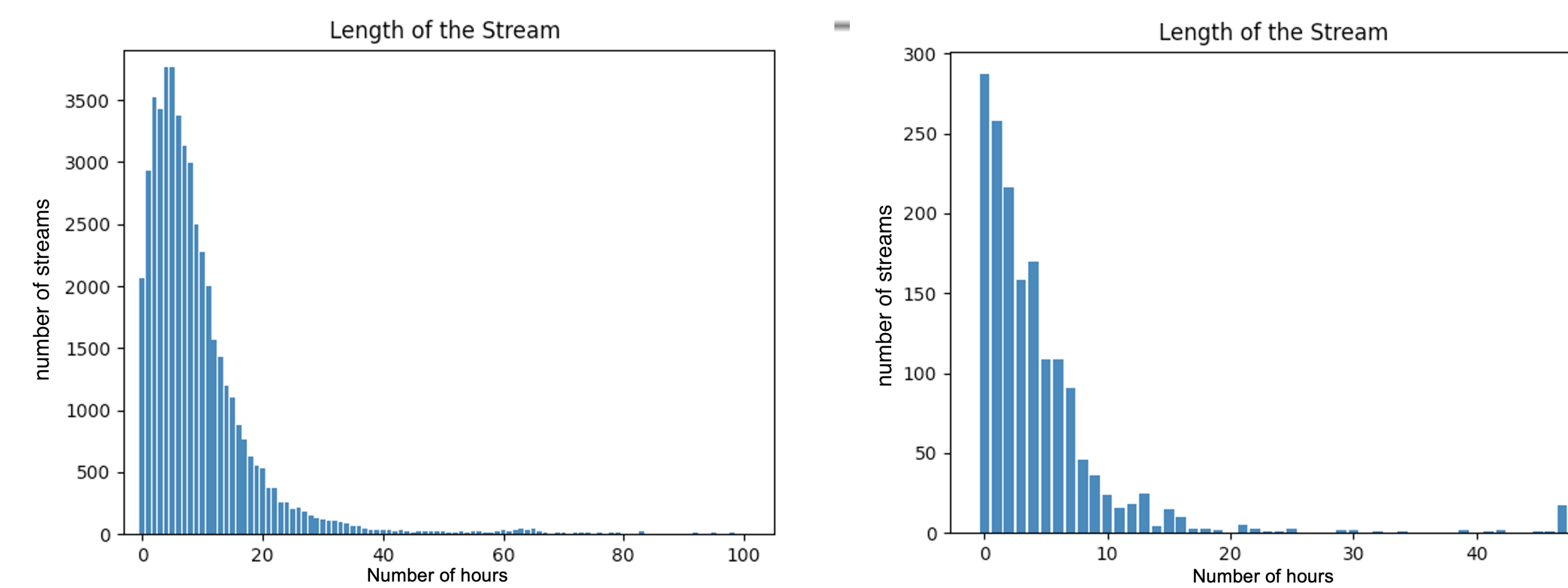


Figure 3: Stream of Entire Dataset

Figure 4: Streams of outliers

## Distribution of the # of Streams Per Week

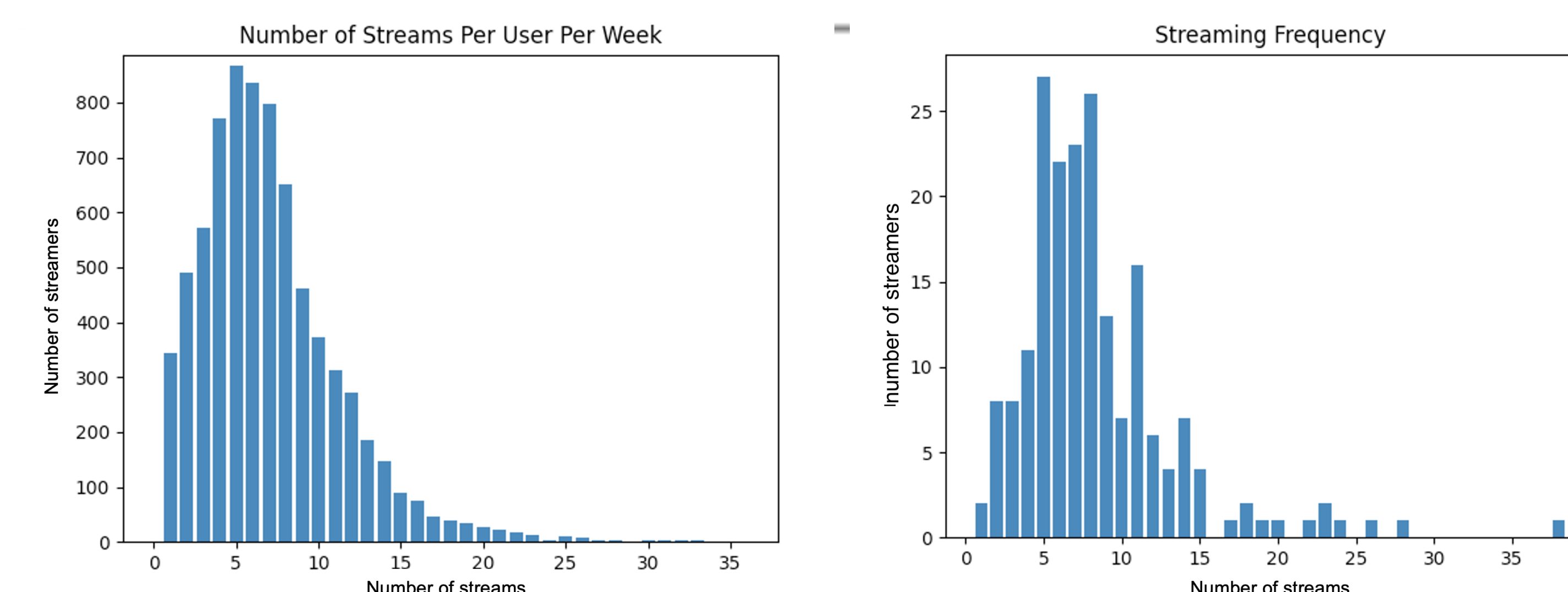


Figure 5: Streams of Entire Dataset

Figure 6: Streams of outliers

## Distribution of Languages and Contents

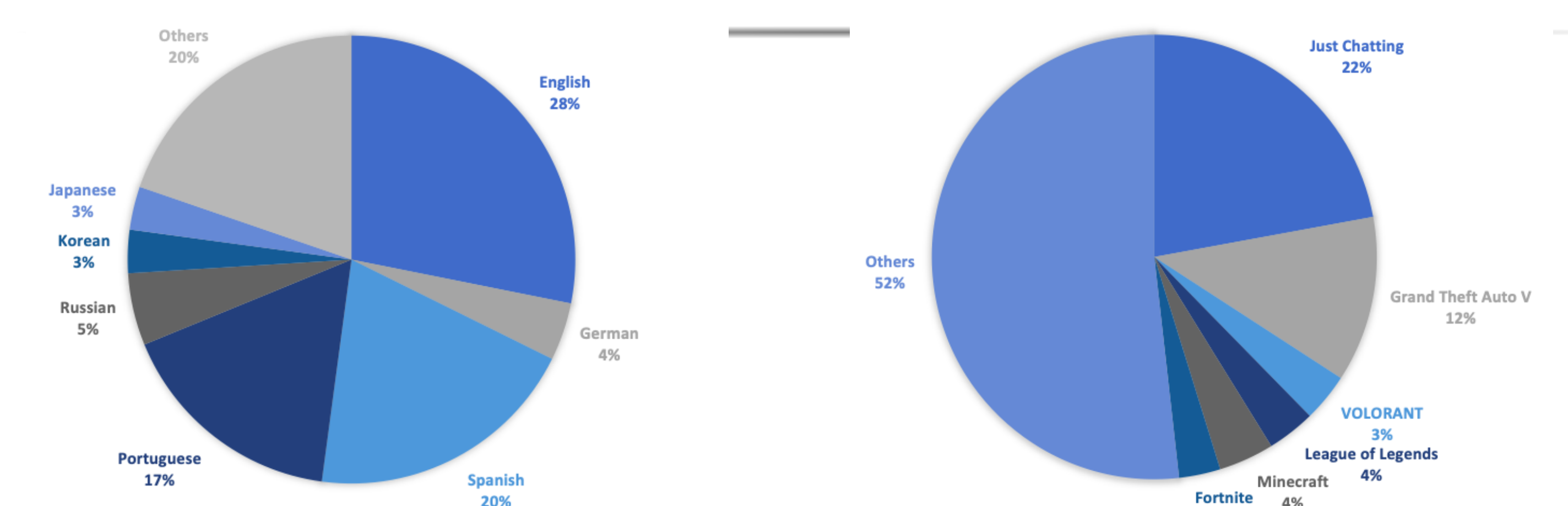


Figure 7: Average viewers

Figure 8: Followers

## Analyzing Outliers

To understand what is the best strategies to gain more follower during streaming periods, we isolates streamers who gained more followers compared to other streamers at the same popularity.

- > We categorize streamers based on the number of their current followers into 20 intervals: 0-1000, 1000-2000, 2000-3000 and so on.
- > In each category, we pick top 5% streamers who gained the most followers in a week.
- > Based on those top 5% streamers' data, we create Figure 1 to Figure 6 and table 2, and presume the effective strategies to be popular.

## Fitting ML Model

**Classifying streamers** into three different groups based on the growth rate:

- > *good* streamers : those who gain more followers than the average followers gained by streamers in the same bracket.
- > *bad* streamers : those who lost followers
- > *average* streamers : the rest

Use **Random Forest Classifier** to predict the growth rate of a channel:

- > Use synthesis minority oversampling technique (SMOTE) to fix the unbalance dataset
- > Parameters: # of initial followers, # of streams per week, contents of the streams, whether they stream popular games
- > Model returns the predicted class of the streamer: 1=good, 0=average, -1=bad

## Categorizing Streamers By Current Followers

We classified the streamers into 4 groups based on their initial number of followers.

Categories	Initial number of followers	Number of users
small	0-5000	1857
medium	5000-10000	1177
big	10000-100000	3383
mega	> 100000	1062

Table 1: Categories of Users

## Gaming vs Non-Gaming Content

	entire dataset(%)	outliers(%)
gaming	70.78	50.00
non-gaming	29.22	50.00

Table 2: Content Distribution

## Results

Based on the outlier analysis, we provide some useful strategies for streamers looking to gain followers:

- > Stream at least 5 times per week
- > Each stream lasts less than 12 hours
- > Mixed non-gaming content with gaming content
- > Stream more than 40 hours per week

	Entire dataset	Streamers w strategy
Average Follower Gained	101	342
Median Follower Gained	29	67

At the maximum tree depth of 10, the resulted machine learning model produced can correctly predict the growth rate of a channel with the accuracy rate of upto 72%.