

## Background and Motivation

- Social media is important to Purdue Athletics because it allows each sports account to grow and engage with their fan bases, create brands for themselves, and interact with other companies and sports teams via sponsorships and reposting.
- The purpose of this project is to collect social media data and conduct data analysis to accomplish the previously outlined relevant social media goals of Purdue Athletics.
- The accounts we scrape are for Volleyball, Women's Basketball, Men's Basketball, Men's Golf, Wrestling, Soccer, Men's Swim/Dive, Women's Tennis, Baseball, Track and Field, PurdueSports, Softball, Football, Men's Tennis, Women's Swim/Dive, and Women's Golf
- The social media platforms we scrape for are Facebook, Instagram and Twitter
- Purdue Athletics was previously collecting and presenting data manually, so help automate this process as much as possible and conduct analysis on the data using tools such as R, Python, and Tableau that we found useful to implement
- Our data analysis involves scraping the social media data (date/time, type of post, likes, and comments) from respective platforms, calculating growth metrics (engagement rate, follower growth) and summarizing the data in reports through visualization,



## SCRAPING AND AUTOMATION

### Instagram:

- Rather than scraping the data by hand every week or two weeks, we decided to speed up the process and make it more efficient
- We are using a CRON job to automatically scrape the data from all the accounts and input it into an Excel file
- A CRON job is a job scheduler that runs a script, the scraper in this case, at a given time every day

### Twitter:

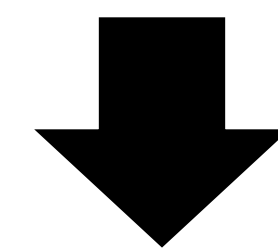
- Using a CRON job, we were able to automate the follower account scraping for engagement rate
- Since we have limited scraping capabilities on Twitter developer account, we still have to manually scrape the tweets

### Facebook:

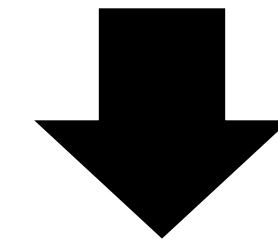
- Manual scraping for Facebook occurs about every 2-3 weeks
- Automation is difficult because of how often bans come from Facebook, but once there is a solution to accounts not getting banned, automation can be achieved



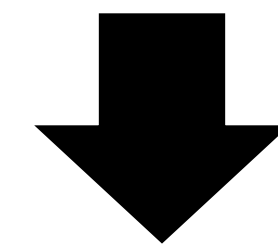
## Scrape the Data



## Categorize the Post



## Clean the Files



## Generate the Graphics

### Categorizing the Data

- After meeting with Dr. David Pierce of Indiana University-Purdue University Indianapolis about his work in categorizing social media posts, the team added categories as a part of our analytics

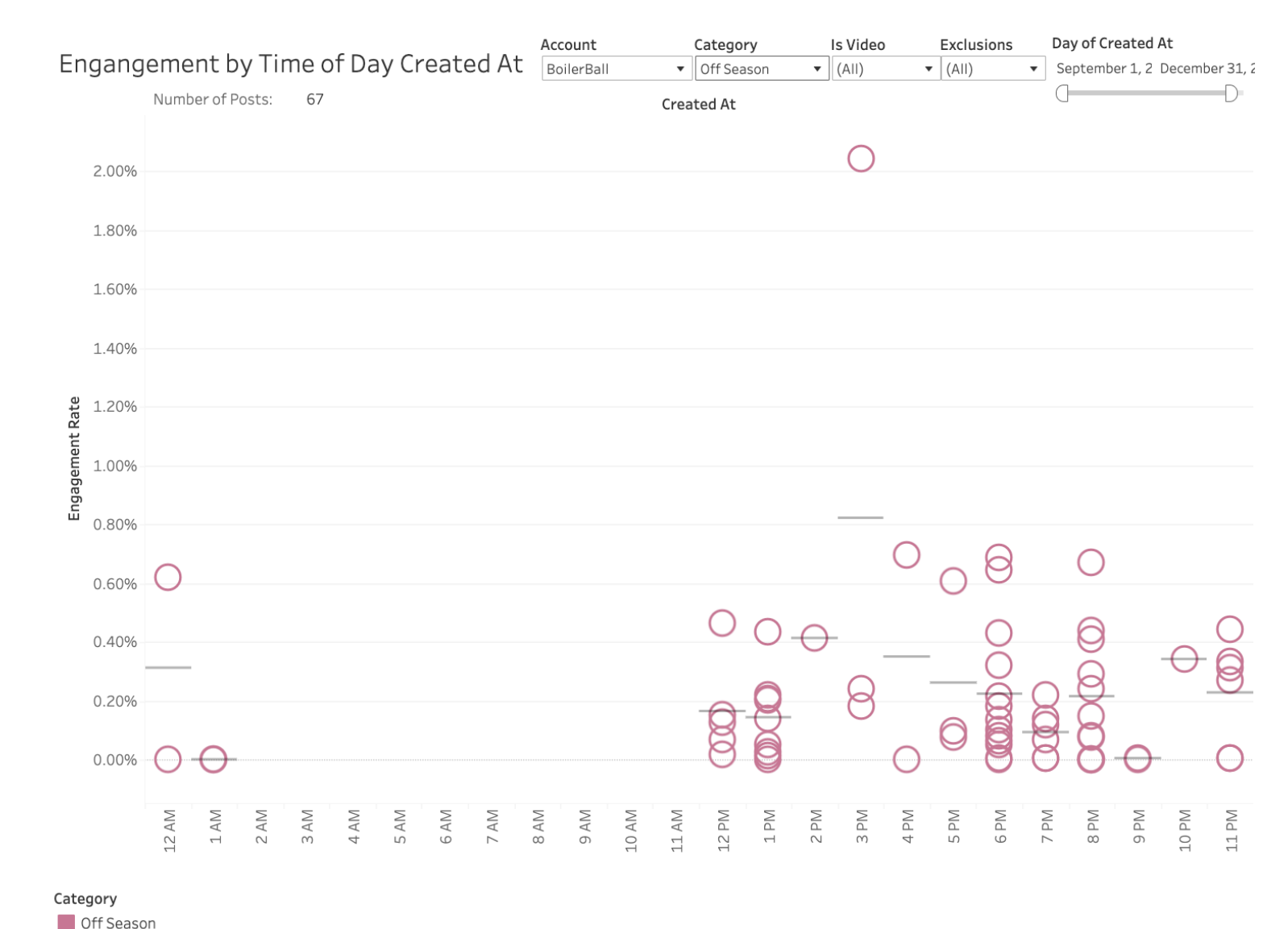
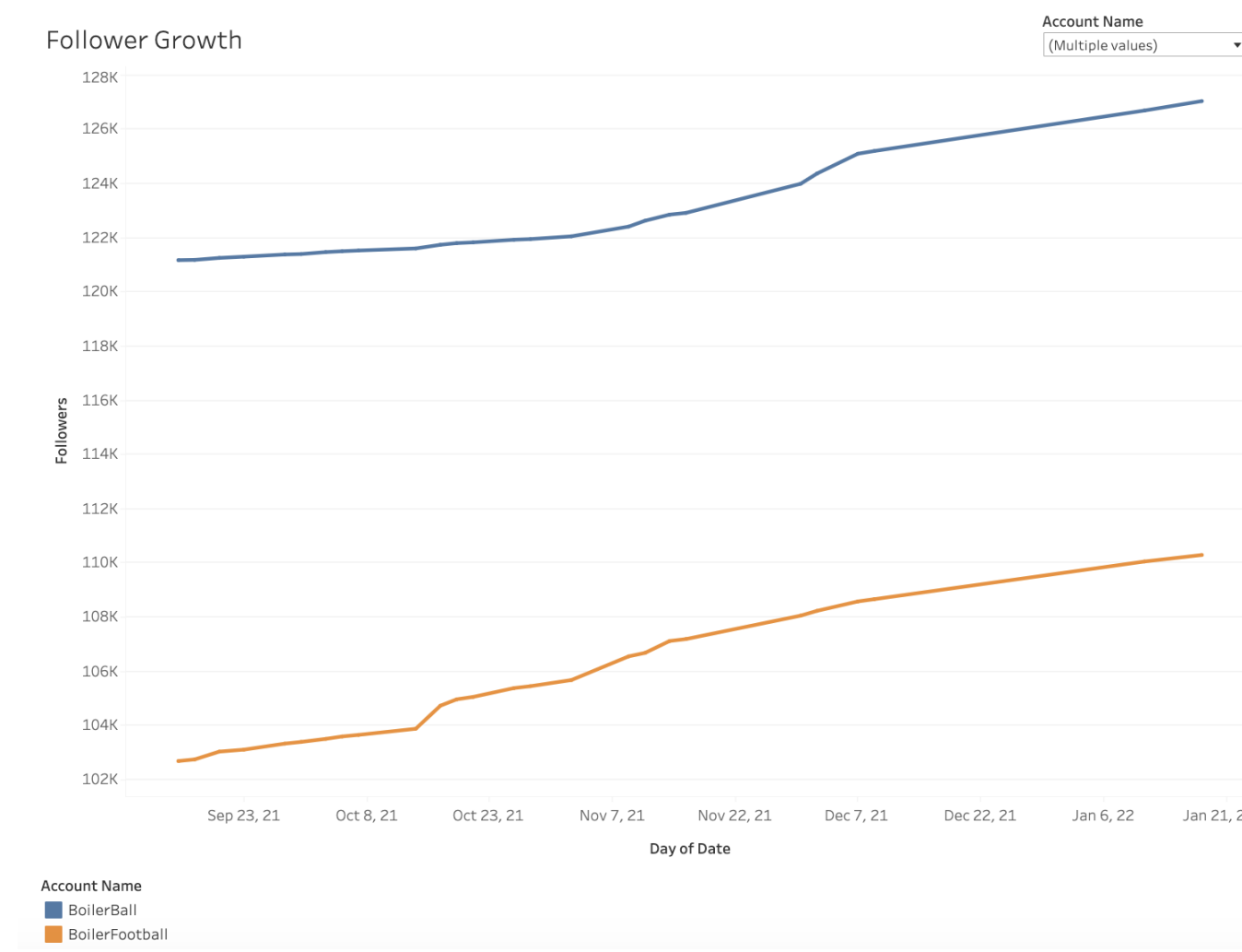
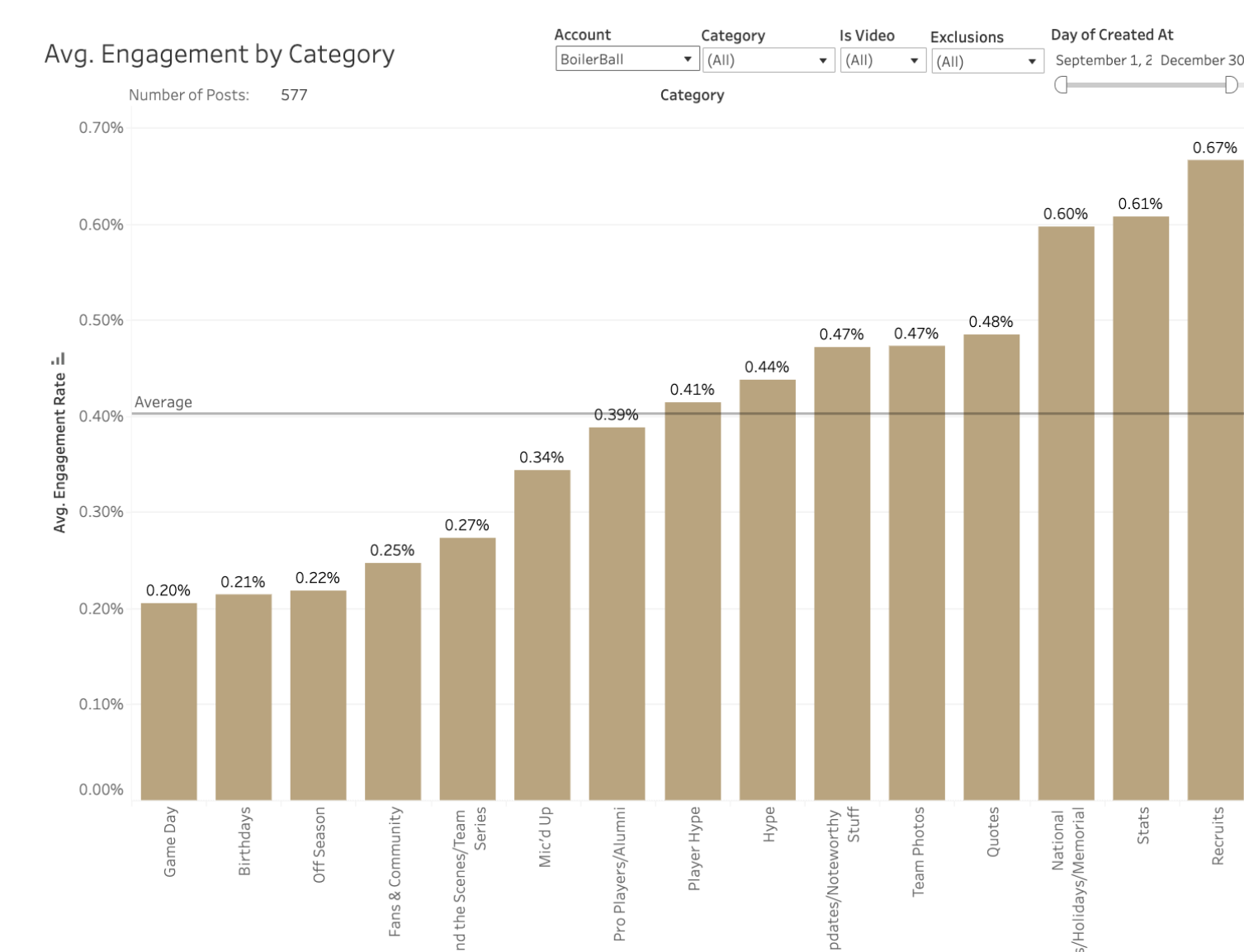
- Originally the categories were based on Dr. Pierce's categories

- Met with our team mentor and the Social Media Team of Purdue Athletics and altered the categories to best fit Purdue Athletics' Social Media needs

- The categories used are Offseason General, Pro Players/Alumni, Statistics, Purdue History, Mic'd Up, Fans & Community, Game Day, Team Series/Behind the Scenes, Player Features, Special Recognition & Initiatives, Team Features, Recruiting/Signees, Facilities, Fundraising Content, Sponsored Content, In-Season General, Game Week, Ticket Information/Sales, and Other

- Team members manually categorize the posts for specific teams across the three platforms

## VISUALIZATION



## Visualizations

Each platform has a dashboard created using Tableau. The dashboards show visualizations that allow the user to see how average engagement compares across categories and time of day posted for each account on each platform as well as follower growth for each account. From these visuals, we are able to draw conclusions about what type of posts are the most successful and at what time of day they should be posted. This helps the athletic department decide what type of posts they should put the most effort into. The visuals can also be filtered down by when they were posted if the user wishes to only look at a specific month or even a specific week. Outlying data points that may skew the data can also be excluded for a more accurate breakdown.

## Conclusion

In conclusion, the focus of our team's project with Purdue Athletics is to help achieve the goals of growth and engagement in social media. We work to achieve these goals by scraping data from the social media platforms of Twitter, Instagram, and Facebook, and using this data to create useful visuals to better analyze the social media data. Automation is also a goal of our team and we have made progress into fully automating the process of data scraping social media platforms.

## Future Goals

- Continue to work with the Shiny App in R
- Digital engagement: using our tool with the Purdue Sports app and website
- Have more consistent scraping with Facebook and not getting our accounts banned
- Continue to automate data scraping