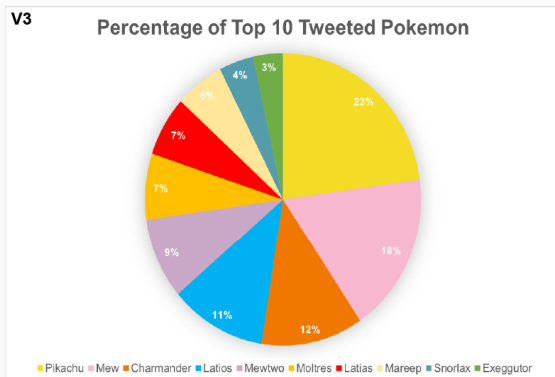# The Hype Behind Pokémon GO

Introduction

The main focus of our project was to do an analysis of Twitter data regarding the use/gameplay of Pokémon Go and create forms of data visualization. We set out to explore the relationship between the number of people playing Pokémon Go and the number of people who are continuously tweeting about playing over the course of a month.

Methods

To gather our Twitter data, we made a Python script that used Twitter's streaming API to stream the data as it arrives on Twitter and copy it straight into a file on our computer that will hold all the tweets collected. We streamed only the tweets using the phrase "#pokemongo". While the data was being streamed in, it would contain a username, tweet context, and a time tag, which was used to help us parse the data. To parse the data we used Python, Bash and R scripts which took advantage of the hashtags used in the data to parse them accurately.

Conclusion

In conclusion, our studies found a relationship between people playing PokemonGo and how they interact with it, such as what they care about, what the most popular Pokemon is and why that is, as well as what times people prefer to play Pokemon Go. Looking at the pie chart and the word cloud we find a correlation with twitter activity and shiny Pokemon as well as Pokemon being featured on Community Day events over the course of a month. There was also an unexpected correlation between PokemonGo and PKG tokens which is a form of crypto currency that can be earned from within the game and the community events.

Data and Models

V1

Popular Tweet Times

V2

V3

Percentage of Top 10 Tweeted Pokemon

Pikachu 23%, Mew 18%, Charmander 12%, Latios 11%, Mewtwo 9%, Moltres 7%, Latias 7%, Mareep 5%, Snorlax 4%, Exeggutor 3%

■ Pikachu ■ Mew ■ Charmander ■ Latios ■ Mewtwo ■ Moltres ■ Latias ■ Mareep ■ Snorlax ■ Exeggutor

V1 acquired data from tweets streamed over the course of a month with the use of a Python script for streaming and a Bash script for saving them in a file and parsing out the relevant time data. From this chart, we can conclude that Pokémon Go Twitter activity is at it's highest during the evening times of day; peaking around 7:00 p.m. (19:00). Presumably due to a majority of Twitter users tweeting about Pokémon Go becoming more active when there is more leisure time.

V2 used tweets streamed from Twitter with the Pokémon Go hashtag. An R script was used to parse the tweets with a stop list and word transformation using the text mining package, which was then made into a word cloud with the wordcloud package. From the word cloud, we found that there is a unexpectedly high amount of Twitter activity surrounding Pokémon Go and a cryptocurrency known as pkg token. We can determined that this cryptocurrency is in high demand even more when events are held. Also, there is a ton of activity regarding capturing legendary shiny Pokemon that are released during Community Day events with Pokémon Go.

V3 used a Bash script that made a list of each word from all the tweets, sorted them based on occurrences with the number of occurrences next to the word, and then picked the top ten most popular Pokémon. From this pie chart, we can conclude that Pikachu is the most tweeted about Pokémon, and this makes sense as it is the franchise's mascot. One of the more surprising things discovered was the popularity of Mareep. This is perhaps tied to a Pokémon Go Community Day event that took place while we were streaming tweets.

Jessica Doan, Pierre Karaffa, Josue Figueroa