1. The question: Bot or not?

Earlier this year Facebook conceded to investors that somewhere between 60 and 80 million fake profiles currently exist of the social network's 2 billion users. This is a significant increase from 2016, when it last reported a mere 18 million counterfeit profiles.
In July, Twitter announced that it would be suspending 70 million accounts, an aggressive action prompted by continuous allegations that the social media platform exercises undue influence on news outlets, such as the U.S.'s most recent presidential elections or the caravan migration of workers from South America.
With populations becoming increasingly dependent on social media platforms as a news source, and with a growing number of false endorsements (in the form os retweets, facebook likes, etc), it is critical that we get familiar with what spambots are made of and auto-clean this mess!
In this study I seek to build a model which will take a user's information, such as their name, gender, age and sample tweets, and identify whether or not a profile is genuine or produced via an automated account creator.

2. My Data

This data was compiled by Crowdflower and is comprised of genuine, traditional, and social spambot Twitter accounts. The initial gathering of data was done on behalf of Twitter Italia and their pursuit of research in the same field. There are a total 248,533 rows and 34 columns that include unique id numbers, screen_name, count of friends, count of favorites, personal urls, language, origin location, and sample tweets.

3. My Minimum Viable Product (MVP)

- Determine common words or phrases used in bot language and that of human vernacular for comparison
- Discover trends in topics, products or markets where bot language is most prevalent
- Provide interpretable visuals that accurately capture these relationships