# Milestone 4: Storytelling

Daniel Baquero

# Description

- Lobbyst4America objective is to provide insights for people who aim to affect legislation. A good strategy and an influential objective is the heart of lobbying. However, selecting the objective can be difficult. By analyzing the tweets data it is possible to find the most influential person  within the congress for specific topics. The project aim is to define who is the most influential person related to the congress in Tweeter.

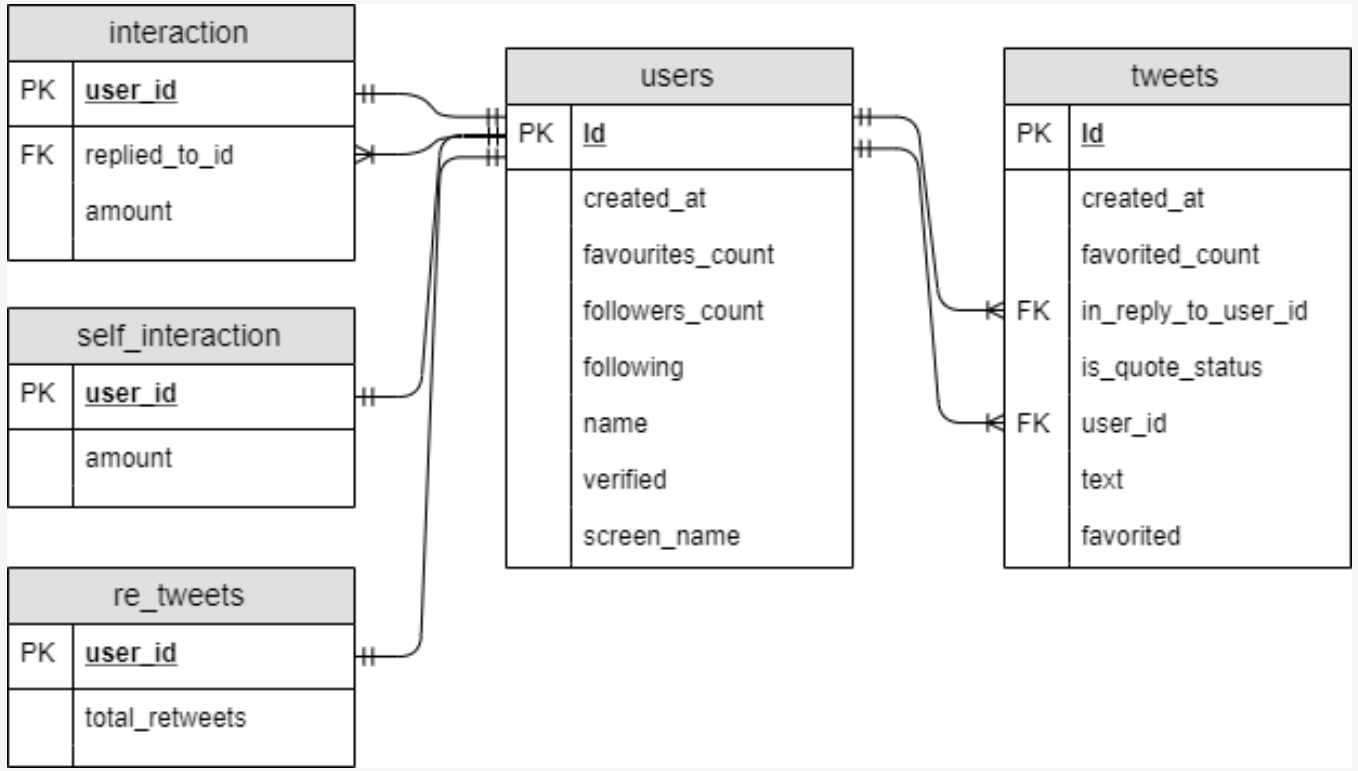
# Client

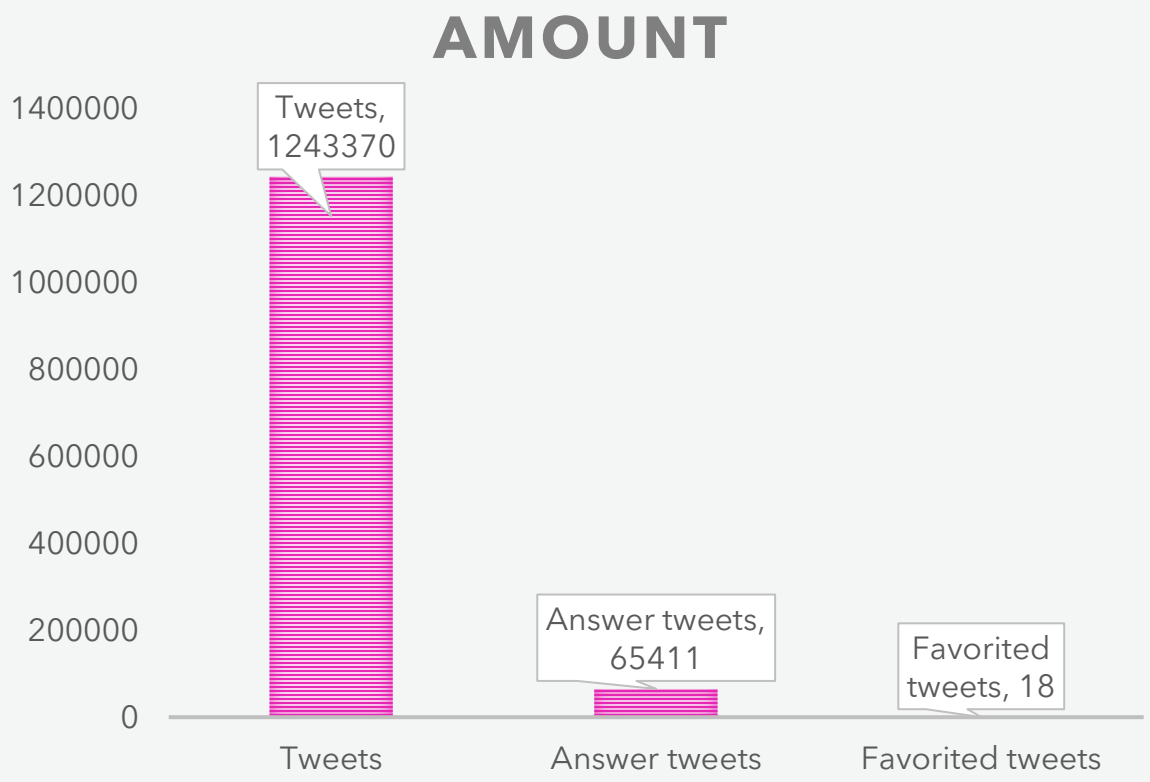
- The client of this presentation is people that might be interested in persuading influential people in the congress. Besides the politics, this people might have specific interests about new legislations.
- The focus would not be in the “data science” aspect of the project.
- It would be a high level, insights and findings, story.

# Approach

- Every question can be answered with existing fields such as:
  - Favorities count
  - Followers count
  - Favortied
  - Total retweets
- Metrics are given as the total number or count of boolean fields.
- By plotting the relation between different metrics the most influential person can be found.
- The aim is to predict who can be influential in the future by comparing similar activity profile with influential people in the present.

# Preparing for the proposal: ER diagram






- 548 Users
- 163433 Average followers
- 414 Average favorited tweets

# Preparing for the proposal: EDA

# Questions & Results

Who has the most followers in tweeter?

Who is the most active person in tweeter?

Is there any unusual activity at the most recent times in tweeter?

Been a verified user has any impact on followers?

# Answers

**Who has the most followers in tweeter?**

**Who is the most active person in tweeter?**

# Is there any unusual activity at the most recent times in tweeter?

Time series analysis couldn't be done due to an unidentified format in the "created at" date and time column.

The format was purely numeric.

There was any match with time stamp format or any other standard date and time format.

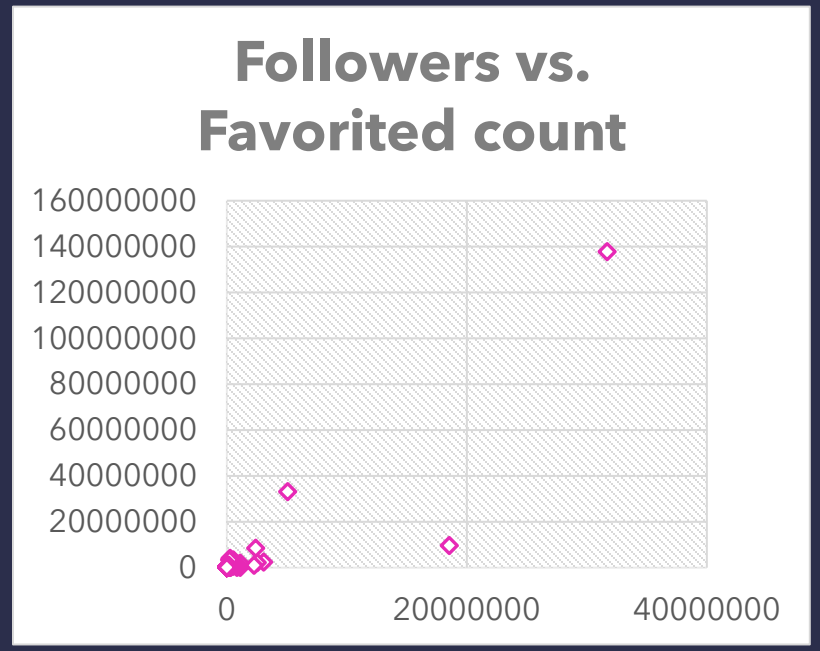
# Been a verified user has any impact on followers?

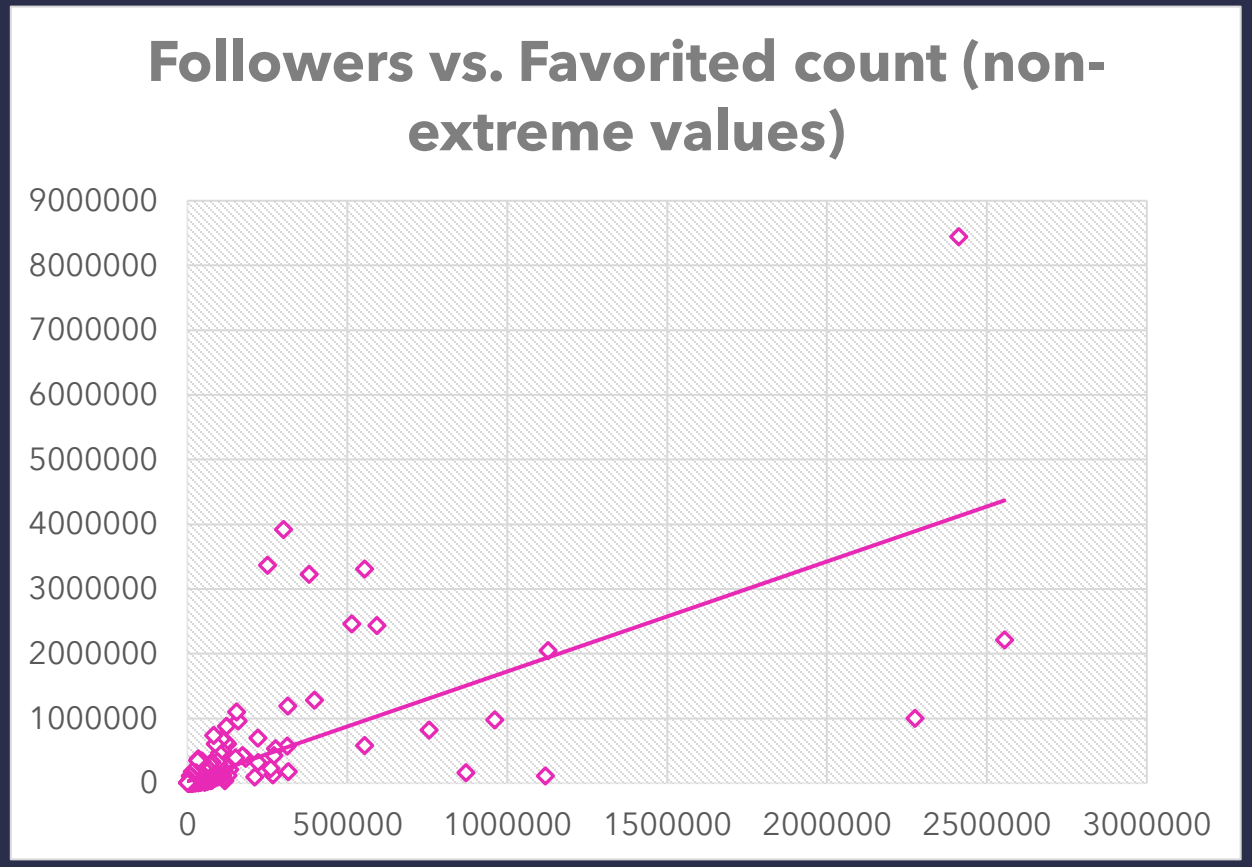
| | Verified | Non-verified |
|---|---|---|
| Total followers | 89483880 | 77902 |
| Amount of users | 530 | 18 |

# Hypothesis

- People who retweet or answer to themselves don’t have as many followers as people who interact with others.
- People with favorited tweets have the most followers.
- The more tweets activity, the more followers.

# Correlations




**Correlation : 0.662**

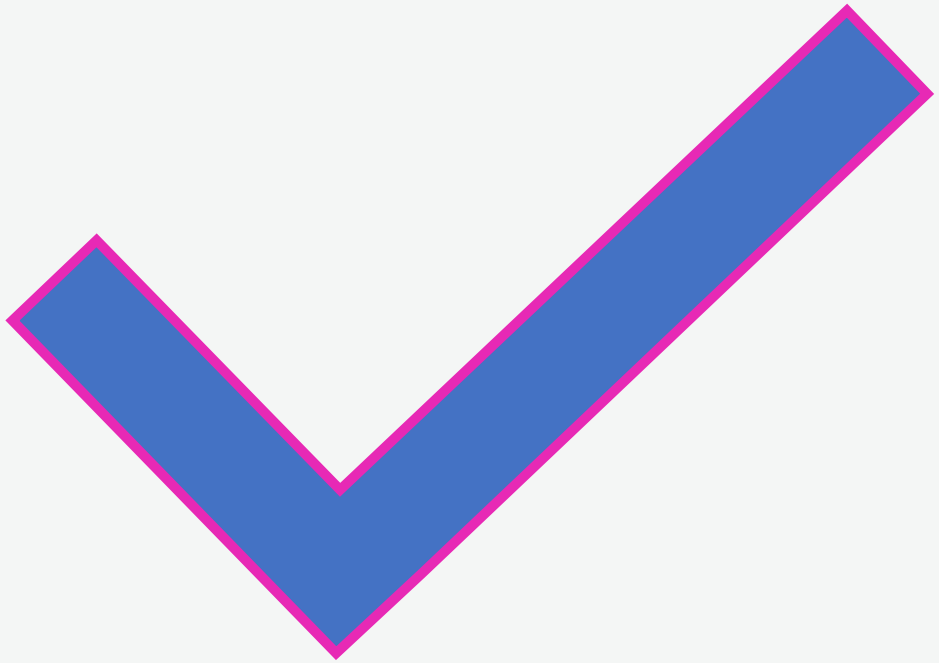
- All following correlations have values near zero and big p-values (There is not a statistical correlation):
  - Followers vs. total tweets.
  - Followers vs. self replies.
  - Followers vs. replies.

# More correlations

# Hypotheses discussion

- People who retweet or answer to themselves don’t have as many followers as people who interact with others.
  - There is no correlation between any of both measures respect to the followers' count.
- People with favorited tweets have the most followers.
  - The correlation is positive demonstrating that having more favorited tweets is a sign of having more followers.
- The more tweets activity, the more followers.
  - This could  not be tested with dates. However, the total number of tweets didn’t represent having more followers.

# Recommendations and Actions



- People who have more favorited tweets are more influential. Persuading this individuals could have a positive outcome.
- Having a verified account in tweeter is a synonym of having more followers.
- The most influential person is the president Donald Trump.
- Text analysis should be performed to unveil the most influential person in specific topics.