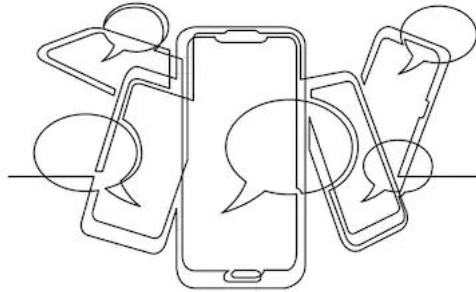


**Product Analysis  
&  
Recommendations**

## Real Talk Mission

**“Help kids know that they’re not alone”**

- *Curated crowdsourced* authentic teen stories, paired with high quality online resources

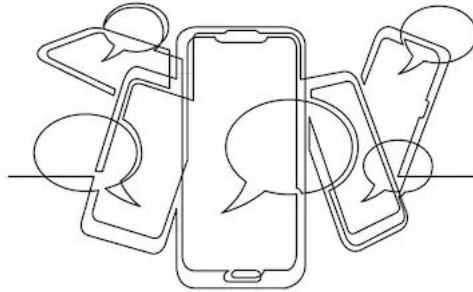


<https://image.shutterstock.com/image-vector/continuous-line-drawing-social-media-260nw-771035026.jpg>

## Real Talk Mission

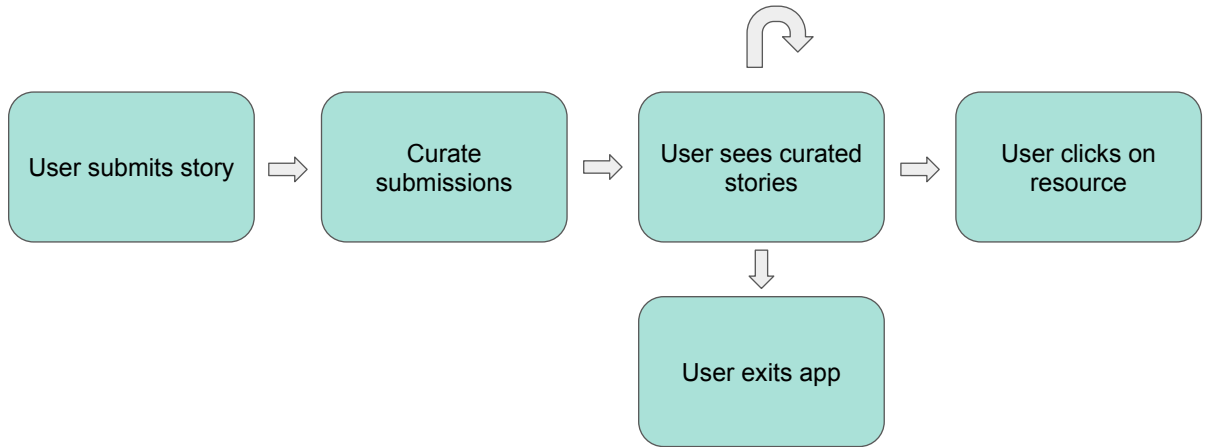
“Help kids know that they’re **not alone**”

- **Curated crowdsourced** authentic teen stories, paired with **high quality online resources**

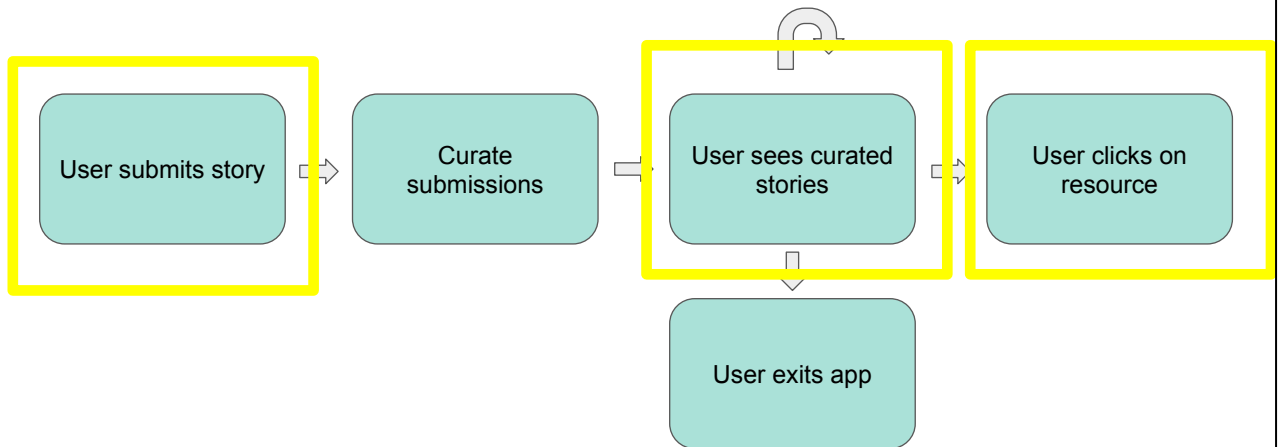


<https://image.shutterstock.com/image-vector/continuous-line-drawing-social-media-260nw-771035026.jpg>

## Principal Actions



## Principal Actions



3 areas that we can focus on to test ways to improve app

1. Top of the funnel for content
2. Surfacing relevant content for retention and engagement
3. Product value of quality of online resources provided for users

## Limitations & Caveats

- Limited Tracking
  - No unique identifier for users to attribute actions
    - Unable to determine causality and retention
  - No impression data available to understand *rates*
    - Ex. click through rate
- Data Quality
  - User-generated content --> unstructured text data
    - No data validation
      - Ex. User-generated identities for Race

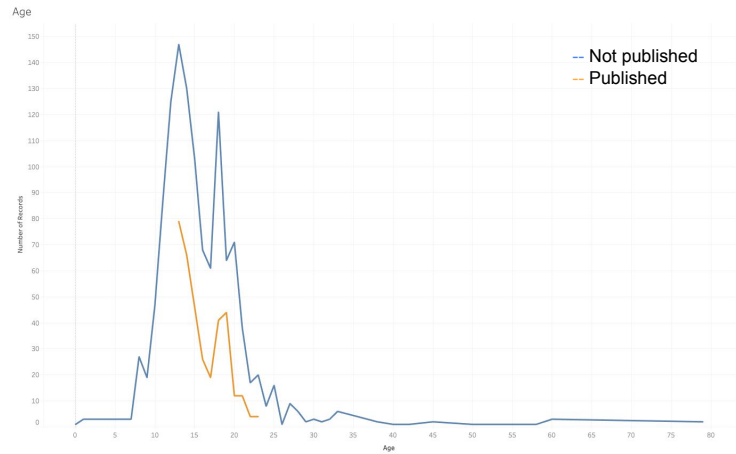
Due to the nature of the problem, it is of utmost importance we take care in tracking data since its sensitive

## Principal Questions and Goals

- Who are our users?
  - User demographics
- What are they interested in?
  - Submissions
  - Resources

# User Demographics

- Average user is in early teens
  - Segment by publishing history reveals core user to be early teens
- Use publishing history as proxy for core users
  - Note, this encodes the curator's bias for what is worth being published

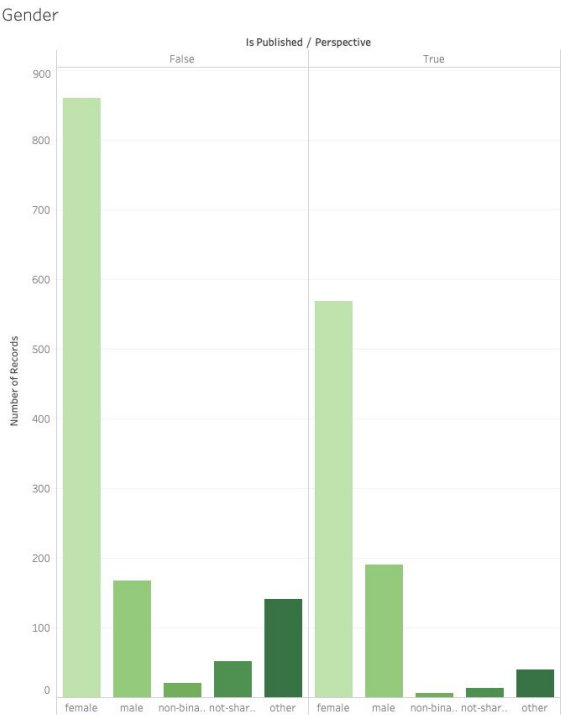


Note that age is a free-text entry so data quality and validation could pose an issue  
ages greater than 100 as outliers for analysis



# User Demographics

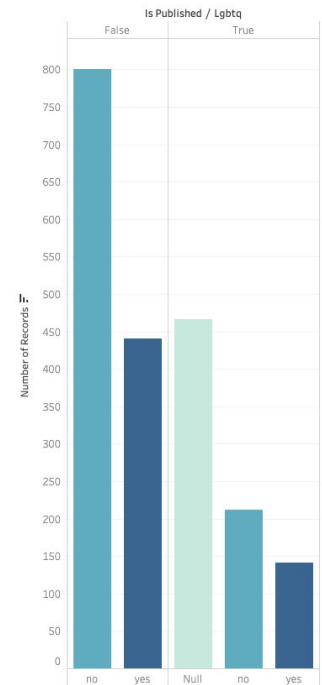
- Predominantly female



# User Demographics

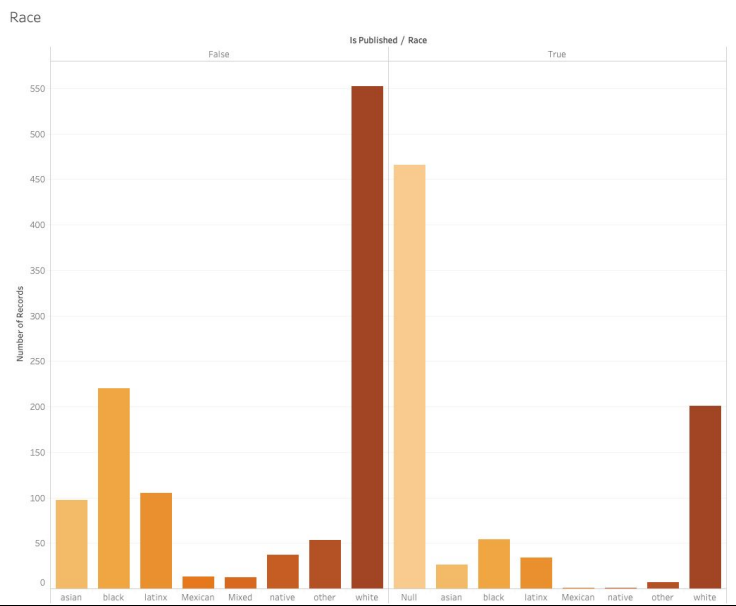
- Most do not identify as queer
  - Segmented by publishing history, many who have published opted to not disclose queer identity when signing up for the app
  - Can we conduct user research and surveys to understand why?

LGBTQ



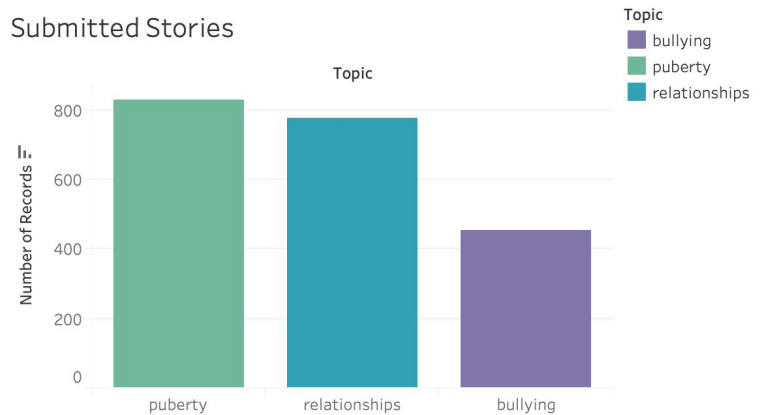
# User Demographics

- Predominantly users who self-identify as white
  - Note data quality issue since racial identities for user profiles are free-text



## User Interests

- Mostly interested in writing about relationships
  - Note: categories provided by app and not filled-in by the user
  - How do we ensure the integrity of the classification of submissions?

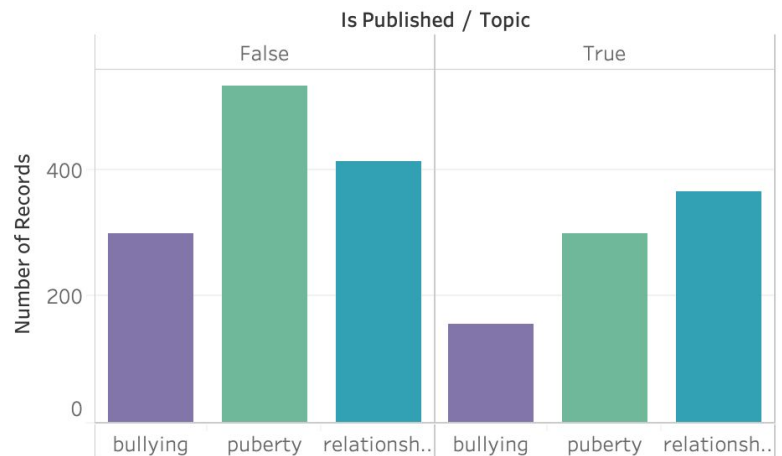


Using submitted stories as a proxy for interest

## User Interests

- Most submissions are about puberty, but most published stories are about relationships
  - Is spam more correlated with puberty and not getting published?
  - Is relationship content better quality?

### Submission and Publishing

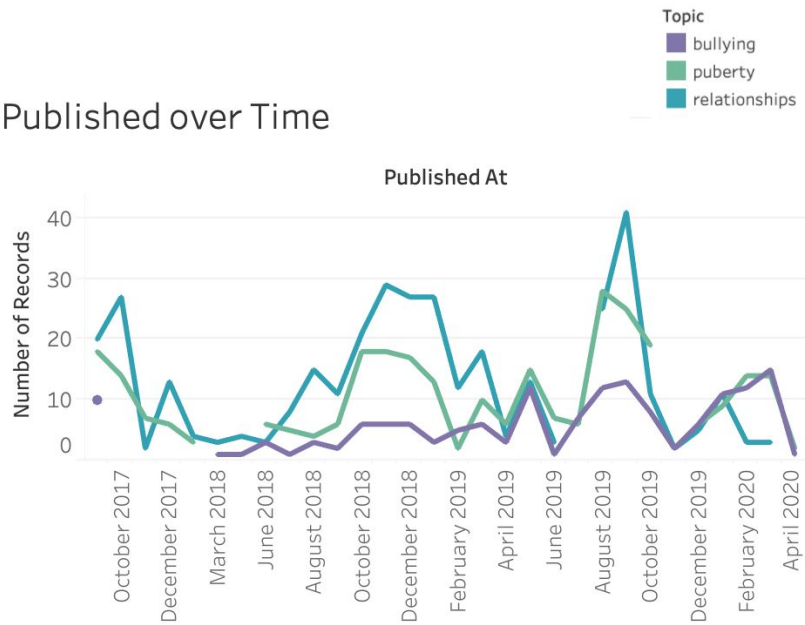


Published stories to encode quality posts and filter out spam to judge authentic stories

# User Interests

- Relationship posts continue to dominant content over time

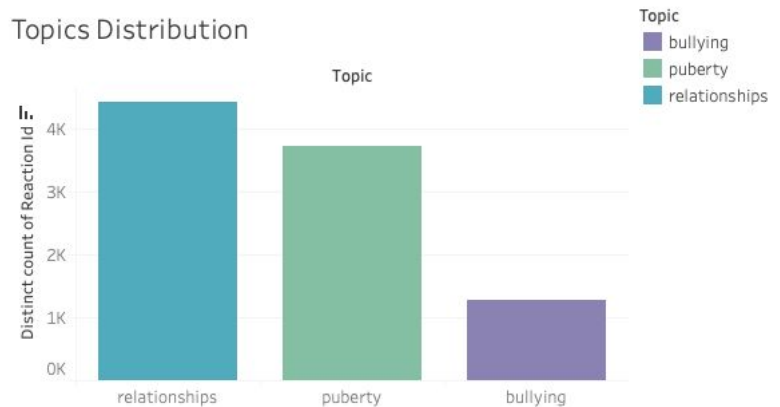
Published over Time



Published over time

## User Interests

- User preference present in reaction trends
  - Relationship topic has more reactions than other topics
  - Similar average number of reactions per story in relationship posts and puberty

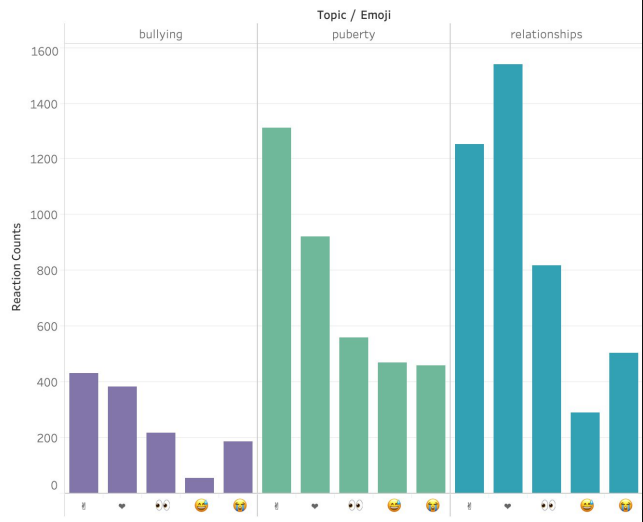


Topics that generate the most reactions from users to get proxy for interests

# User Interests

- Discrepancy in user engagement when slicing by topics and types of reactions (Emojis)
  - ‘Me too’ more prevalent in bullying and puberty topics, but ‘like’ more prevalent in posts about relationships

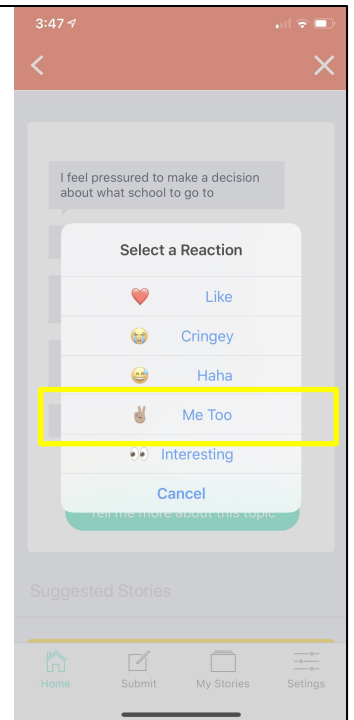
Topics and Emojis





## User Interests

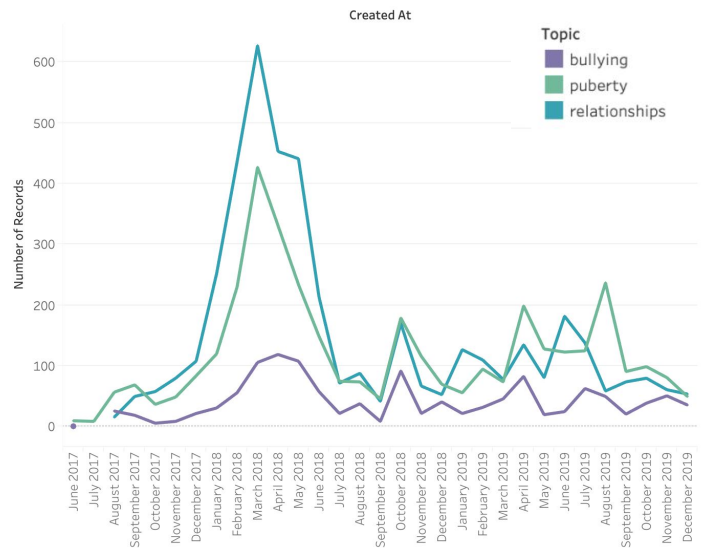
- Users have to hold down the post to trigger the reactions prompt
  - Can we test different ways to display that action to determine if we're inadvertently hiding the prompt?
- Most popular reaction is "Me Too"
  - Typically you would expect the first reaction displayed to be most popular
  - Can we make the conclusion that this is evidence of users resonating with the story, and are relating?



## Product Value

- Do not have impression data, or app-open data
- Use reaction as proxy
  - Spike in activity in March 2018
    - Is this real or a data collection issue?
    - If it's real, what caused the spike, and how do we replicate the effect to increase engagement?

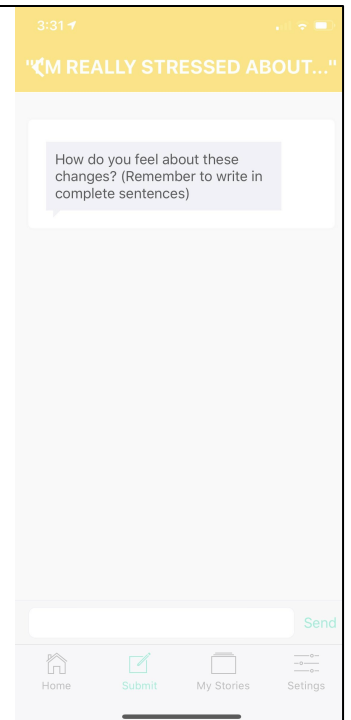
Reactions Over Time



Reactions over time to judge volatility of interests

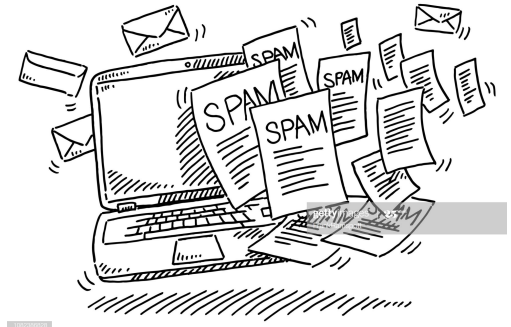
# Submission Quality

- Currently: *manual* review of all submissions to filter quality posts
  - How do we reduce the energy and time investment in curating posts?
  - Predicting submission quality:
    - *Cannot* use age, race, gender, sexual orientation as features as that would encode bias into the models
    - Can we do *text analysis* to engineer new features to predict the quality of a post?



## Submission Quality

- What are some ways we can *encourage* and *identify* quality posts?
  - How can we create a better experience so that we increase the number of quality stories submitted?
    - Deploy changes to product intended to influence the quality of submitted stories



<https://www.gettyimages.com/detail/illustration/abundance-of-spam-email-messages-laptop-royalty-free-illustration/1082386828>

# Resource Quality

- Currently, we do not know the number of users who:
  - See the resource attached
  - Click on the resource attached
  - Engage due to the quality of the resource and become activated users
- Recommendation:
  - Collect data which allow us to understand user behavior and draw insights

## THE DATA SCIENCE HIERARCHY OF NEEDS

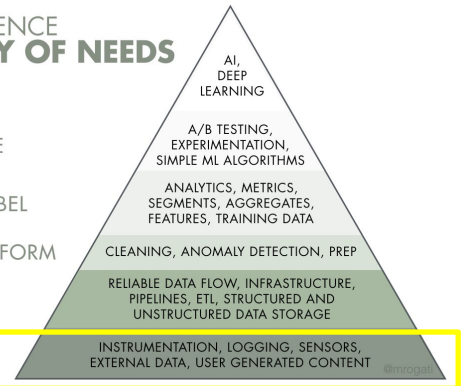
LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT



<https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007>

# Takeaways

- What do we know?
  - Early-teen, female average user
    - Write about puberty, engaged with posts about relationships
    - Do not identify as queer
- What do we not know?
  - Currently, no data logging user actions
    - Understand the value that the product provides for engaged users
      - “Magic moment”
    - Implement initiatives which allow for more users to reach magic moment
  - Understanding the User lifecycle
    - Where are users from?
    - How did they hear about the app/what referrer led them to the app?
    - Where are steep drop-offs where users churn? And why?



<https://image.shutterstock.com/image-vector/global-drawing-earth-day-vector-260nw-562291960.jpg>

## Recommendations

- Improve tracking and logging of user actions to better understand habits and draw insights
  - Needed to be able to measure things to improve
- Data validation step to improve quality of user-generated data to understand demographics and trends
  - Ex. phone number --> location
- Experiments to identify user value and key to engagement
- Text Analysis (Topic Modeling) to draw granular insights from user stories
  - Creates additional dimensions from unstructured text data
- Recommendation Engine to kickstart Engagement Flywheel
- [Tableau Dashboards](#)

<https://public.tableau.com/profile/jenn.w2341#!/vizhome/SubmissionsandPublishing/Dashboard1?publish=yes>

[https://public.tableau.com/profile/jenn.w2341#!/vizhome/UserActivity\\_15866652380440/UserActivity?publish=yes](https://public.tableau.com/profile/jenn.w2341#!/vizhome/UserActivity_15866652380440/UserActivity?publish=yes)