Reddit Shaming Karens

By: Brian Bell, Raymond Cosgrove, and Andrew McGrail
Outline

- Background Information
- Pre-planning
- Plans + Overview
- Our Implementation
- Ending + Acknowledgements
Background Information
The main question
The question our project aims to help answer is...

How has the usage of the word *Karen* changed over the course of the global pandemic?

- Dr. Zach
What exactly is a Karen?
Karen is a pejorative slang term for an obnoxious, angry, entitled, and often racist middle-aged white woman who uses her privilege to get her way or police other people’s behaviors.

- dictionary.com
Pre-Planning
The Team

Name: Andrew McGrail
Tasks: Data Collection, Parsing and Caching

Name: Raymond Cosgrove
Tasks: Natural Language Processing, Contact Person

Name: Brian Bell
Tasks: Data Analysis, Communication
The Timeline
## Timeline

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Planning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base Scripts (Parsing Data Col.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial Statistics/Trends (D.A.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic Modeling, Testing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interim Report</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integration Testing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final Testing, Deployment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final Presentation, Report</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Plans + Overview
Our Approach

- Restrict to 18 months before and after the pandemic
  - Multiple subreddit data sets.

- Find significant data trends
  - Use frequency, depth vs. breadth, context
  - Compare keyword usages between subreddits

- Utilize the power of NLP
  - Topic model posts to find trends
Our Motivation

- Usage spike during pandemic (starting March 2020)
- Determine usage triggers
- Has keyword changed over time
Our Motivation Cont.
Our Implementation
Technical Details
The Architecture
### Example Input File

<table>
<thead>
<tr>
<th>#command</th>
<th>subreddit</th>
<th>keyword</th>
<th>start_date</th>
<th>end_date</th>
<th>post_count(optional)</th>
</tr>
</thead>
<tbody>
<tr>
<td>download</td>
<td>postmates</td>
<td>karen</td>
<td>2018-09-01</td>
<td>2021-09-01</td>
<td>500</td>
</tr>
<tr>
<td>download</td>
<td>doordash</td>
<td>karen</td>
<td>2018-09-01</td>
<td>2021-09-01</td>
<td>500</td>
</tr>
<tr>
<td>download</td>
<td>grubhubdrivers</td>
<td>karen</td>
<td>2018-09-01</td>
<td>2021-09-01</td>
<td>500</td>
</tr>
<tr>
<td>download</td>
<td>doordash_drivers</td>
<td>karen</td>
<td>2018-09-01</td>
<td>2021-09-01</td>
<td>500</td>
</tr>
<tr>
<td>download</td>
<td>UberEATS</td>
<td>karen</td>
<td>2018-09-01</td>
<td>2021-09-01</td>
<td>500</td>
</tr>
<tr>
<td>download</td>
<td>deliverooos</td>
<td>karen</td>
<td>2018-09-01</td>
<td>2021-09-01</td>
<td>500</td>
</tr>
</tbody>
</table>

### Group 2: Food employees in general and firm specific (if existent)

| download         | bartenders       | karen   | 2018-09-01   | 2021-09-01 | 500                  |
| download         | serverlife       | karen   | 2018-09-01   | 2021-09-01 | 500                  |
| download         | TalesFromYourServer | karen   | 2018-09-01   | 2021-09-01 | 500                  |
| download         | KitchenConfidential | karen   | 2018-09-01   | 2021-09-01 | 500                  |
| download         | TalesFromTheKitchen | karen   | 2018-09-01   | 2021-09-01 | 500                  |
| download         | dishwashers      | karen   | 2018-09-01   | 2021-09-01 | 500                  |
| download         | McDonaldsEmployees | karen   | 2018-09-01   | 2021-09-01 | 500                  |
| download         | StarbucksBaristas | karen   | 2018-09-01   | 2021-09-01 | 500                  |
| download         | chickFILAworkers | karen   | 2018-09-01   | 2021-09-01 | 500                  |

### Group 3: Restaurants

| download         | Chipotle         | karen   | 2018-09-01   | 2021-09-01 | 500                  |
| download         | Panera           | karen   | 2018-09-01   | 2021-09-01 | 500                  |
| download         | McLouge          | karen   | 2018-09-01   | 2021-09-01 | 500                  |
| download         | Inout            | karen   | 2018-09-01   | 2021-09-01 | 500                  |
| download         | jimmyjohns       | karen   | 2018-09-01   | 2021-09-01 | 500                  |
| download         | Dominos          | karen   | 2018-09-01   | 2021-09-01 | 500                  |
| download         | chickFILA        | karen   | 2018-09-01   | 2021-09-01 | 500                  |
| download         | JerseyMikes      | karen   | 2018-09-01   | 2021-09-01 | 500                  |
| download         | Subway           | karen   | 2018-09-01   | 2021-09-01 | 500                  |
| download         | Dairyqueen       | karen   | 2018-09-01   | 2021-09-01 | 500                  |
| download         | wendys           | karen   | 2018-09-01   | 2021-09-01 | 500                  |
| download         | Popeyes          | karen   | 2018-09-01   | 2021-09-01 | 500                  |
| download         | BurgerKing       | karen   | 2018-09-01   | 2021-09-01 | 500                  |
Example Data File

```json
"comment_limit": 2048,
"comment_sort": "confidence",
"_reddit": "<praw.reddit.Reddit object at 0x7f8a8e736af0>",
"approved_at_utc": null,
"subreddit": "wendys",
"selftext": "<truncated text for slides>",
"author_fullname": "t2_k1383",
"saved": false,
"mod_reason_title": null,
"gilded": 0,
"clicked": false,
"title": "\"with everything\"",
"link_flair_text": [],
"subreddit_name_prefixed": "r/wendys",
"hidden": false,
"pwls": 6,
"link_flair_css_class": "",
"downs": 0,
"thumbnail_height": null,
"top_awarded_type": null,
"hide_score": false,
"name": "t3_myjfhv",
```

Example Statistics

2019 {'posts': 0, 'num_comments': 0, 'score': 0, 'over_18': 0, 'is_video': 0, 'media_only': 0}
Feb
2019 {'posts': 0, 'num_comments': 0, 'score': 0, 'over_18': 0, 'is_video': 0, 'media_only': 0}
Mar
2019 {'posts': 1, 'num_comments': 29, 'score': 25, 'over_18': 0.0, 'is_video': 0.0, 'media_only': 0.0}
Apr
2019 {'posts': 0, 'num_comments': 0, 'score': 0, 'over_18': 0, 'is_video': 0, 'media_only': 0}
May
2019 {'posts': 0, 'num_comments': 0, 'score': 0, 'over_18': 0, 'is_video': 0, 'media_only': 0}
Jun
2019 {'posts': 0, 'num_comments': 0, 'score': 0, 'over_18': 0, 'is_video': 0, 'media_only': 0}
Jul
2019 {'posts': 0, 'num_comments': 0, 'score': 0, 'over_18': 0, 'is_video': 0, 'media_only': 0}
Aug
2019 {'posts': 0, 'num_comments': 0, 'score': 0, 'over_18': 0, 'is_video': 0, 'media_only': 0}
Sep
2019 {'posts': 0, 'num_comments': 0, 'score': 0, 'over_18': 0, 'is_video': 0, 'media_only': 0}
Oct
2019 {'posts': 0, 'num_comments': 0, 'score': 0, 'over_18': 0, 'is_video': 0, 'media_only': 0}
Nov
2019 {'posts': 1, 'num_comments': 63, 'score': 58, 'over_18': 0.0, 'is_video': 0.0, 'media_only': 0.0}
Dec
2019 {'posts': 1, 'num_comments': 19, 'score': 8, 'over_18': 0.0, 'is_video': 0.0, 'media_only': 0.0}
Jan
2020 {'posts': 2, 'num_comments': 96, 'score': 110, 'over_18': 0.0, 'is_video': 0.0, 'media_only': 0.0}
Feb
2020 {'posts': 0, 'num_comments': 0, 'score': 0, 'over_18': 0, 'is_video': 0, 'media_only': 0}
Mar
2020 {'posts': 0, 'num_comments': 0, 'score': 0, 'over_18': 0, 'is_video': 0, 'media_only': 0}
Apr
2020 {'posts': 0, 'num_comments': 0, 'score': 0, 'over_18': 0, 'is_video': 0, 'media_only': 0}
May
2020 {'posts': 5, 'num_comments': 81, 'score': 76, 'over_18': 0.0, 'is_video': 0.0, 'media_only': 0.0}
Jun
Topic Modeling
Deliverables

- Collection of historical Reddit data spanning 36 months
  - 30 subreddits
  - Keyword statistics within each

- Python Executable
  - Historical data collection
  - Data analysis and visualization
  - Natural language processing
Future Plans

- Higher Granularity graphs
  - Graphing options already supported
  - Per subreddit graphs

- Reddit image analysis
  - Image downloading supported
  - OCR to pull text from images to process content

- Extend to different industries
  - Framework provided so similar experiments can be conducted
Ending + Acknowledgements
Acknowledgements

Dr. Florian
Client

Dr. Fox
Professor
Credits

- https://www.dictionary.com/browse/karen
- http://files.pushshift.io/reddit/
Thank you!

Questions?