

Clustering Creative Writing

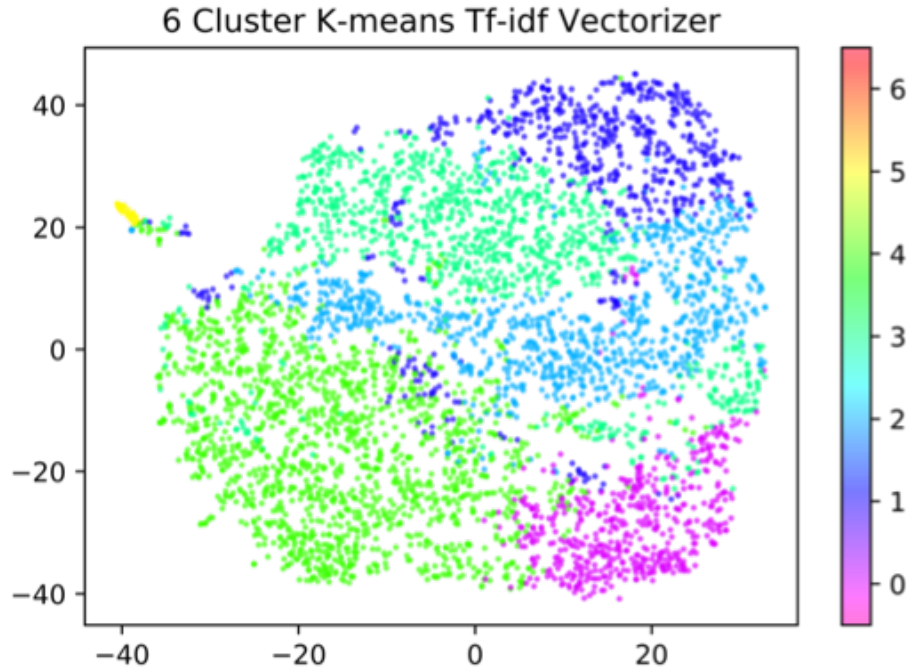
Harmeet Hora

Metis Summer 2018 Cohort

The Data

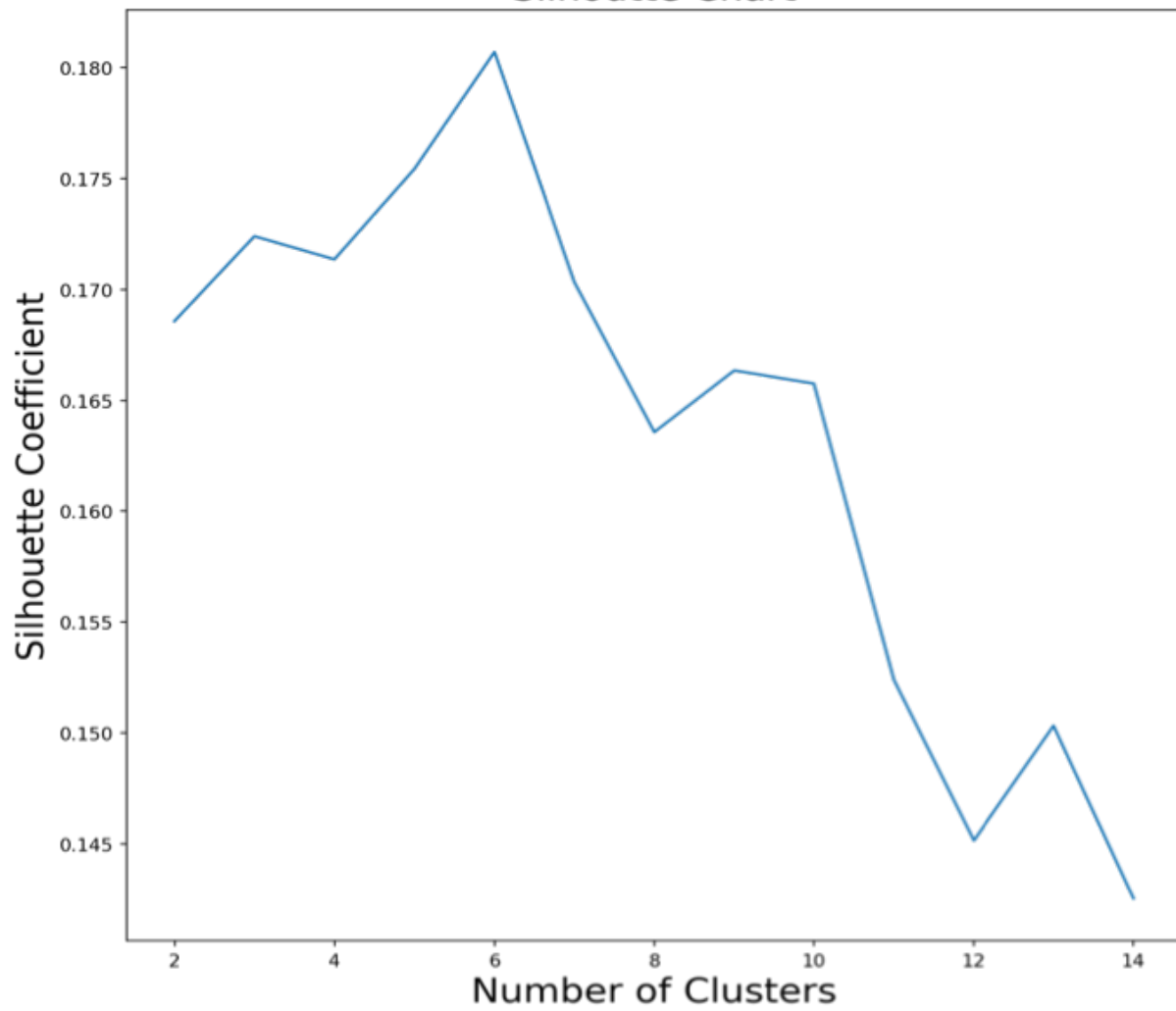
- Scraped using the Python Reddit API Wrapper
- Exclusively /r/writingprompts Subreddit
- Main focus on prompt responses
- Example: The year is 1910. Adolf Hitler, a struggling artist, has fought off dozens of assassination attempts by well meaning time travelers, but this one is different. This traveler doesn't want to kill Hitler, he wants to teach him to paint. He pulls off his hood to reveal the frizzy afro of Bob Ross.

Method - K-Means

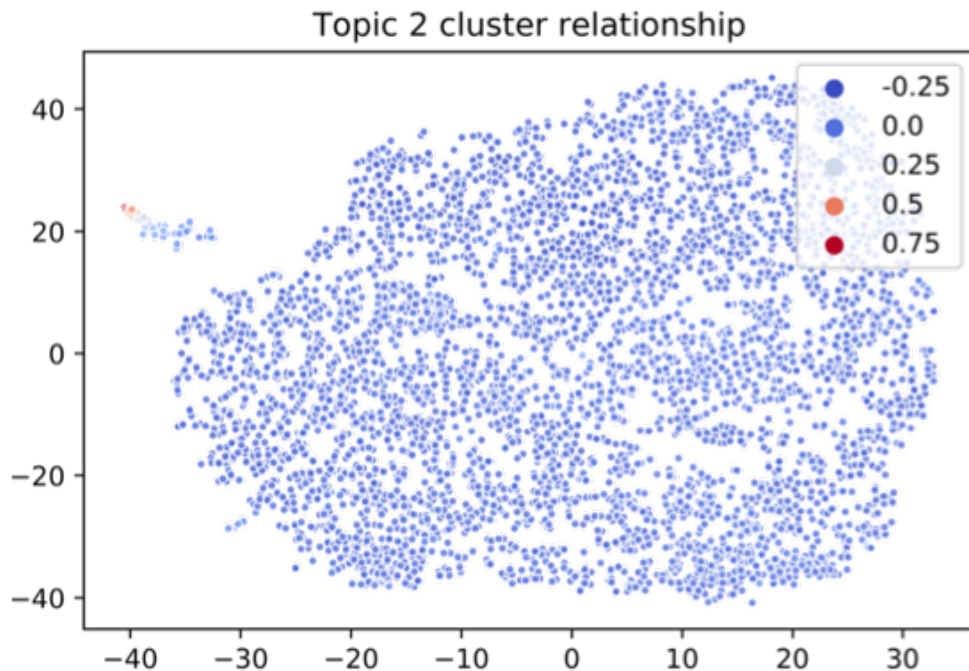


- 0: Immortality, Mortality
- 1: Old Age, slow
- 2: Music
- 3: Working, Jobs
- 4: Evil Forces
- 5: Harry Potter

Silhouette Chart



Cluster Interpretation



- Topic 2: harry, dumbledore, ron, dudley, voldemort, potter, wand, hogwarts, hermione, boy

Next Steps

- Collect more data points
- Analyze prompts
- Incorporate post popularity
- Text generation

**Thank you for
your attention.**