

Assignment 3, Part 1 - Report

Sample output of 'topic1':

```
regmckie@LAPTOP-VQ2C8KLP: ~$ bin/kafka-console-consumer.sh --topic topic1 --from-beginning --bootstrap-server localhost:9092
{"title": "Drag Race Thailand Season 3 Sneak Peek", "author": "galaxystar1", "text": "", "url": "https://v.redd.it/3he0qymkdd1"}
{"title": "The truth about who Monet Kai Kai\u0026039s with has been revealed", "author": "Kadetm93", "text": "", "url": "https://v.redd.it/tvdlgrgpkdd1"}
{"title": "RuPaul Surprising Season 16 Winner Nymphia Wind at her Dragcon Booth + Queens talking about Drag Con!", "author": "ayanakamura", "text": "", "url": "https://www.youtube.com/watch?v=MnBpGDRJA8ab_channel=OnTheRedCarpet"}
{"title": "Met La Grande Dame", "author": "Ok-Transportation141", "text": "Highlight of my first day at the con ^^ \n\niGD is stun that is all.", "url": "https://i.redd.it/ncldq2ufykdd1.jpeg"}
{"title": "Two musicals enter a cage fight, who walks out alive?", "author": "TravellerofInfinity", "text": "", "url": "https://www.reddit.com/gallery/1e7l1qc"}
{"title": "Nymphia and Plastique", "author": "Broomsgate_01", "text": "", "url": "https://www.reddit.com/gallery/1e7l1qc"}
{"title": "\u0026039s Slumber Party\u0026039s premiered on Dropout today!", "author": "redgold51", "text": "Mon\u0026039s is the host of this show that\u0026039s part \u0026039s Playhouse\u0026039s part talk show, and it\u0026039s SO MUCH FUN!! 10/10 recommend!!!\n\n\u0026039s guest stars for the first episode include Jujubee and Alaska, and Naomi Smalls, Bob, and Kim Chi will be on future episodes!", "url": "https://www.reddit.com/gallery/1e7l1qc"}
{"title": "Bob vs. Monet Lipsync", "author": "BillingsBarana", "text": "This is a high thought but I wanna see Bob and Monet battle to \u0026039s Freaky Nasty\u0026039s by Megan Thee Stallion. Idk why, I just think if anyone would be able to deli ver and nail the rap it would be them. \u0026039s \u0026039s", "url": "https://www.reddit.com/r/rupaulsdragrace/comments/1e7l1qc/bob_vs_monet_lipsync/" }
{"title": "Track Records. CanVIM2, EP1.", "author": "ScottyHasTaste", "text": "", "url": "https://i.redd.it/ewo86id97l1dd1.jpeg"}
{"title": "Does anyone like the new Dragon Queens Walk?", "author": "ElectricChameleon", "text": "I wish they would announce the queens\u0026039s names or do the edit of the walk like they did in the past. While I do like the interview with Michelle Visage, it was hard to watch that while trying to figure out who was who. Anyone else feels this way?\n\nEDIT: DragCon", "url": "https://www.reddit.com/r/rupaulsdragrace/comments/1e7l1qc/does_anyone_like_the_new_dragon_queens_wa lk/" }
{"title": "Jujubee saying Protect Average Drag", "author": "kemes7", "text": "", "url": "https://v.redd.it/s5jqhkvbld1" }
```

Sample output of 'topic2':

```
regmckie@LAPTOP-VQ2C8KLP: ~$ bin/kafka-console-consumer.sh --topic topic2 --from-beginning --bootstrap-server localhost:9092
Drag Race Thailand Season: 1
3 Sneak: 1
Idk: 1
Drag Con: 1
Kadetm93: 1
16: 1
DragCon: 1
Monet: 1
one: 1
Drag Race Thailand Season: 1
Michelle Visage: 1
Megan Thee Stallion: 1
https://www.reddit.com/gallery/1e7m91c: 1
10/10: 1
Dragon Queens Walk: 1
Nymphia: 1
queens\u0026039s: 1
Bob: 3
Alaska: 1
Plastique: 1
Kim Chi: 1
it\u0026039s: 1
Track Records: 1
Monet Lipsync: 1
Two: 1
```

***Note:** Both screenshots are only a fraction of each respective topic's output. There was a lot more to each topic's output, but for the sake of brevity, I only showed you a piece of each of them.*

topic1 is the Kafka topic that stores the incoming real-time subreddit information. The 'read_in_subreddit.py' file uses PRAW (the Python Reddit API Wrapper) to read in submissions from a particular subreddit and then stores that information in the Kafka topic 'topic1.'

topic2 is the Kafka topic that stores the word count of the named entities from topic1. The 'wordcount_named_entities.py' file reads in the subreddit submission information from topic1, extracts the named entities from that text, and performs word count on those named entities, and then stores that information in the Kafka topic 'topic2.'

The data source/subreddit I chose was the *RuPaul's Drag Race* subreddit. *RuPaul's Drag Race* is a competitive reality TV show where drag queens from across the country

Assignment 3, Part 1 - Report

compete to become “America’s next drag superstar.” There is no particular reason why I chose this subreddit other than the fact that I love the show.

When it comes to the results of the named entity word count, it’s no surprise that many of the named entities present were drag queens from the show. Names like Vanjie, Eureka, Bob, and Raven all represent famous drag queens from the show, and they were all mentioned multiple times throughout the various submissions. Likewise, many of the submissions also mentioned different franchises across *RuPaul’s Drag Race*. For example, the phrase “*Canada’s Drag Race*” has 2 mentions. This also isn’t a surprise, especially since a new season of *Canada’s Drag Race* is airing right now, so it’s natural for a subreddit dedicated to the show to have multiple mentions of the new season. Other named entities that received multiple word counts aside from drag queen and franchise names were judges’ names, song and artist names, and names of Drag Race-related events.