Exercise: Twitter Data Analysis and Storytelling

Assume you have a custom dataset that contains Twitter data, including tweets, user information, and engagement metrics. The dataset is represented as a list of dictionaries, where each dictionary represents a tweet. The exercise involves analyzing the dataset, filtering tweets based on certain criteria, transforming the data, and generating a story based on the results.

Here are the steps to follow:

1. Create an RDD from the list of dictionaries representing the dataset.
2. Implement a filter operation to select tweets based on specific criteria (e.g., tweets containing a particular hashtag or posted by users with a high number of followers).
3. Transform the filtered tweets by mapping each tweet to the relevant information for storytelling (e.g., tweet text, user handle, engagement metrics).
4. Implement a flatMap operation to obtain a single list of transformed information from the tweets.
5. Generate a story by using the transformed information to create a narrative or insights about the Twitter data.

```
# Input dataset representing Twitter data
dataset = [
    {
        "id": 1,
        "text": "Excited to announce the launch of our new product!
#innovation",
        "user": {
            "handle": "@company",
            "followers": 5000
        },
        "engagement": {
            "likes": 100,
            "retweets": 20
        }
    },
 # Add your own data
]

# Implement the steps
```