

4 Spark

```
wget --no-check-certificate  
→ 'https://docs.google.com/uc?export=download&id=1Kay7TkrEr-3u1Q340b0uEhhibpwAaMPj'  
→ -O spark.csv
```

In this section, you will be using a dataset of restaurants and reviews. The dataset can be found here (Format: [INDEX, Name, Region, Number of reviews, Review Text] . Load the dataset into Spark, and address the following tasks:

1. Determine the number of records the dataset has in total.
2. Find the restaurant with the highest number of reviews.
3. Determine the restaurant with the longest name.
4. Find the number of reviews for each region.
5. Determine the most frequently occurring term in the review column that doesn't include one of the following stop words: "A", "The", "of".