

# First Program In Pyspark (02-Pyspark-First-Program)

We saw in our walkthrough that we count the number of stock quotes for AAPL using code that looks like this:

```
rdd = sc.textFile("hdfs:///data/stocks-flat/input")
aapl = rdd.filter(lambda line: "AAPL" in line)
aapl.count()
```

Try using that pattern to count the number of lines in the file /data/shakespeare/input that contain the word "king". You'll want to use the toLowerCase function before the contains function to ensure all the text is lower case.

---

One solution looks like this:

```
rdd = sc.textFile("hdfs:///data/shakespeare/input")
kings = rdd.filter(lambda line: "king" in line.lower())
kings.count()
```

You should have gotten 4773 lines that reference a king in all of Shakespeare. That's almost 3% of the 175,376 lines in all of Shakespeare's plays and poems.