

**Code:**

```
lines = sc.textFile("/FileStore/tables/demo-1.txt")
lines = lines.map(lambda x: x.replace(" ", "").replace(".", ""))

bigrams = lines.flatMap(lambda s: [(s[i], s[i+1]), 1] for i in range(0, len(s)-1)) bigrams.collect()
bigrams_count = bigrams.reduceByKey(lambda x, y: x+y) bigrams_count.collect()

# Minimum (5 least occurring bi-gram)
min_bigram = bigrams_count.sortBy(lambda x: x[1]).map(lambda t: (f"{t[0][0]}{t[0][1]}", t[1]))
print(min_bigram.take(5))

# Maximum (5 most commonly occurring bi-gram)
max_bigram = bigrams_count.sortBy(lambda x: x[1], False).map(lambda t: (f"{t[0][0]}{t[0][1]}",
t[1])) print(max_bigram.take(5))
```

**Output :****(8) Spark Jobs**

```
[('19', 3), ('51', 3), ('67', 4), ('63', 4), ('66', 5)]
[('72', 19), ('50', 17), ('68', 16), ('85', 16), ('47', 15)]
Command took 1.44 seconds -- by skaur7@stevens.edu at 03/26/2022, 4:33:15 PM on My Cluster
```