

```
# Name : Adwait Purao
# UID : 2021300101
# Batch : B
# Class : Comps B
```

```
!pip install pyspark
```

```
Collecting pyspark
  Downloading pyspark-3.5.1.tar.gz (317.0 MB)
    317.0/317.0 MB 2.4 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Requirement already satisfied: py4j==0.10.9.7 in /usr/local/lib/python3.10/dist-packages (from pyspark) (0.10.9.7)
Building wheels for collected packages: pyspark
  Building wheel for pyspark (setup.py) ... done
  Created wheel for pyspark: filename=pyspark-3.5.1-py2.py3-none-any.whl size=317488491 sha256=e2dfd28c5a7441bdc346
  Stored in directory: /root/.cache/pip/wheels/80/1d/60/2c256ed38ddce2fdd93be545214a63e02fbd8d74fb0b7f3a6
Successfully built pyspark
Installing collected packages: pyspark
Successfully installed pyspark-3.5.1
```

```
from pyspark.sql import SparkSession
```

```
# Create SparkSession
spark = SparkSession.builder.master("local").appName('WordCount').getOrCreate()
```

```
# Get the SparkContext
sc = spark.sparkContext
```

```
# Read the text file
text_file = sc.textFile("/content/BDA_10_Text_Input.txt")
```

```
# Perform word count
word_counts = text_file.flatMap(lambda line: line.split(" ")) \
    .map(lambda word: (word, 1)) \
    .reduceByKey(lambda x, y: x + y)
```

```
# Display word counts
for word, count in word_counts.collect():
    print(f"{word}: {count}")
```

```
In: 2
the: 29
heart: 1
of: 20
bustling: 1
city,: 1
where: 4
neon: 1
lights: 2
flicker: 1
and: 13
sounds: 1
traffic: 1
never: 1
cease,: 1
there: 2
lies: 1
a: 10
quaint: 1
little: 1
cafe.: 1
This: 1
cafe,: 2
with: 3
its: 1
cozy: 1
atmosphere: 1
aromatic: 1
brews,: 1
is: 2
favorite: 2
haunt: 1
```

```
for: 2
locals: 1
tourists: 1
alike.: 1
: 6
Every: 1
morning,: 1
aroma: 1
freshly: 1
ground: 1
coffee: 2
fills: 1
air: 1
as: 5
customers: 2
trickle: 1
in,: 1
seeking: 1
refuge: 1
from: 1
chaos: 1
outside: 1
world.: 1
The: 1
barista,: 1
smile: 1
```