Step-by-step guide for Windows

1. Open Command Prompt or PowerShell

Press **Win + R**, type `cmd`, and hit Enter.

---

2. Create a working directory and input file

```
mkdir C:\wordcount
cd C:\wordcount
echo hello world bye world hello hadoop mapreduce world > input.txt
```

To confirm the file content:

```
type input.txt
```

You should see:

```
hello world bye world hello hadoop mapreduce world
```

---

3. Create the Python script

Create a file named **`wordcount.py`** in the same `C:\wordcount` folder.

You can use **Notepad** or any editor (VS Code, Spyder, etc.), and paste this code:

```
from mrjob.job import MRJob

class MRWordCount(MRJob):

    def mapper(self, _, line):
        for word in line.split():
            yield word, 1

    def reducer(self, word, counts):
        yield word, sum(counts)

if __name__ == '__main__':
    MRWordCount.run()
```

Then save it as:

```
C:\wordcount\wordcount.py
```

---

4. Install MRJob

In your terminal (while Python is installed and accessible), run:

```
pip install mrjob
```

If Python isn't recognized, check that you've added it to PATH — you can test by running:

```
python --version
```

If that fails, you may need to use the full path like:

```
C:\Users\mohammed.fasha\AppData\Local\Programs\Python\Python313\python.exe --
version
```

---

5. Run the MRJob locally

Now execute your script like this:

```
python wordcount.py input.txt
```

You should get output similar to:

```
"bye"        1
"hadoop"     1
"hello"      2
"mapreduce"  1
"world"      3
```

---

6. (Optional) Save output to a file

You can redirect the output to a file:

```
python wordcount.py input.txt > output.txt
```

Then open `output.txt` to see the results.
