Steps to execute Spark programs:

1. Create a new file named **Bash.sh** inside your **spark-3.3.0-bin-hadoop3** folder by typing **vi Bash.sh** in terminal

2. Copy below code and paste inside **Bash.sh** file

   ```
   export JAVA_HOME=$(readlink -f $(which javac) | awk 'BEGIN {FS="/bin"} {print $1}')
   if ! command -v spark-shell --version &> /dev/null
   then
   export PATH=$(echo $PATH):$(pwd)/bin
   fi
   ```

3. In terminal execute Bash.sh file using **source Bash.sh**

4. Create your program directory and write program in <program_name>.py (names may not be the same for the files),

5. Execute Program using below commnad
   spark-submit <python_file_name>.py <input_file> <output_folder_name>

6. To see output
   cat <output_folder_name>/*