There are multiple ways to create notebook, I Initially launched an EMR cluster.

A screenshot of a computer

Description automatically generated

In the next step we need to create an studio and then create a workspace (Notebooks).

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Launching jupyter from the notebook.

A screenshot of a computer

Description automatically generated

Pyspark practice in jupyter notebook.

A screenshot of a computer

Description automatically generated

Running pyspark commands on putty :

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Join examples:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generated

Reading data from csv s3 ..

A screenshot of a computer

Description automatically generated

Reading a csv file from s3 bucket and displaying data in it and performing few operations.

A screenshot of a computer

Description automatically generated

Showing the male candidates ..

A screenshot of a computer

Description automatically generated

Female candidates...

A screenshot of a computer

Description automatically generated

Performing union operation , combining 2 csv files.

A screenshot of a computer

Description automatically generated

Checking for duplicates if any after combining the 2 files.

A computer screen shot of a black screen

Description automatically generated

Write back to the s3 bucket----

A computer screen shot of a black background

Description automatically generated

A screenshot of a computer

Description automatically generated

Reading a txt file from s3 ---

A black background with text

Description automatically generated

Searching for a particular word from the txt file and checking how many times it has repeated.

A computer screen with white text

Description automatically generated

Reading a Json file from s3 ..

A screenshot of a computer

Description automatically generated

In the below example we are reading a car sales report and performing some analysis ..

A screenshot of a computer screen

Description automatically generated

Using FILTER extracted only the BMW from the Brand column.

A screenshot of a computer program

Description automatically generated

Using group by command we displayed all the 7 brands and their average price of their cars.

A screenshot of a computer

Description automatically generated

Using COUNT checking how many cars were sold by each brand.

A screenshot of a computer program

Description automatically generated

Adding a new column as updated date

A screen shot of a computer

Description automatically generated

Writing back to the S3 bucket.

A computer screen with white text

Description automatically generatedA screenshot of a computer

Description automatically generated

Count distinct is used to count the no of car companies in the brand column.

A screenshot of a computer

Description automatically generated