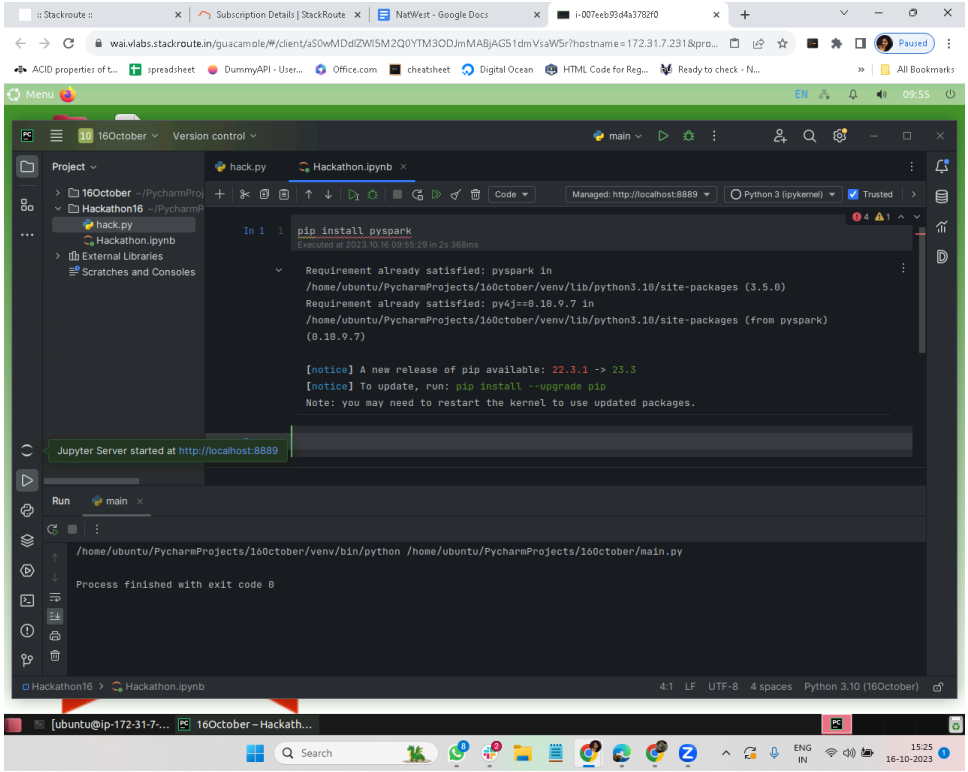
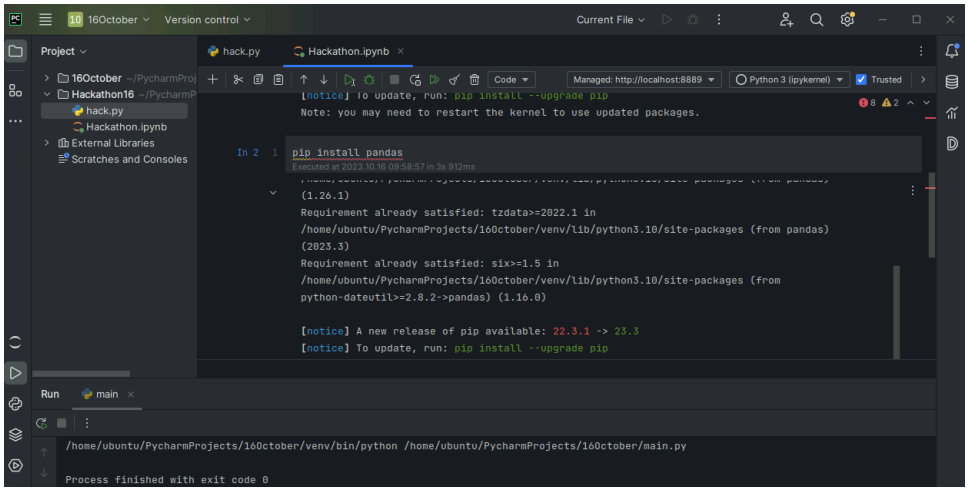
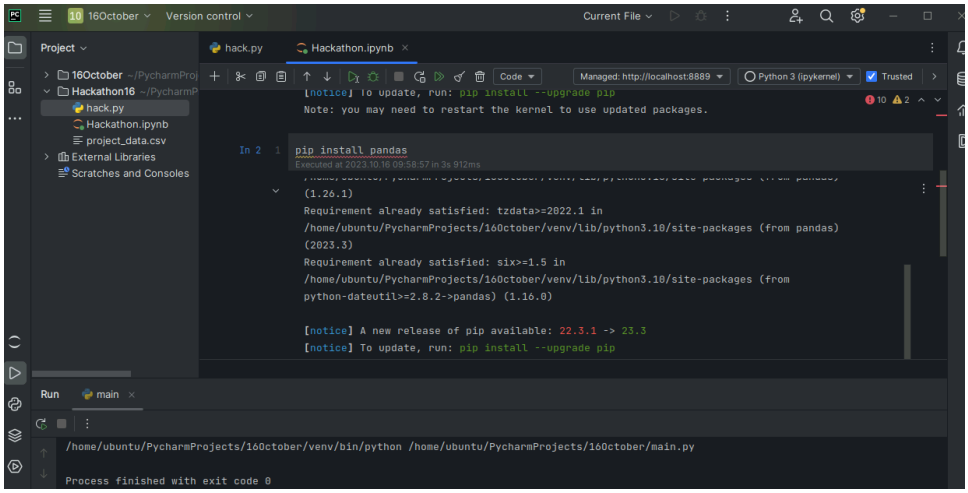


1.PIP install pyspark

2. pip install pandas



3: upload project_data.csv



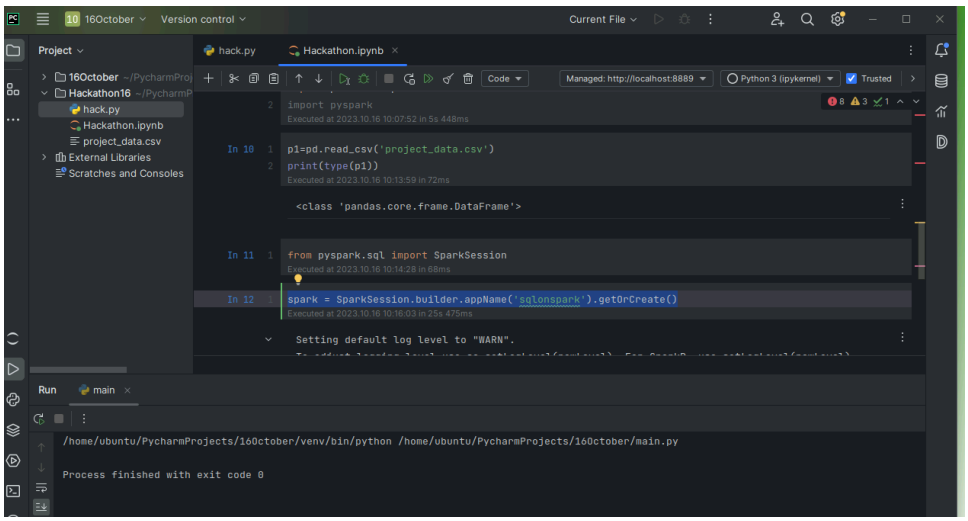
p1=pd.read_csv('project_data.csv')

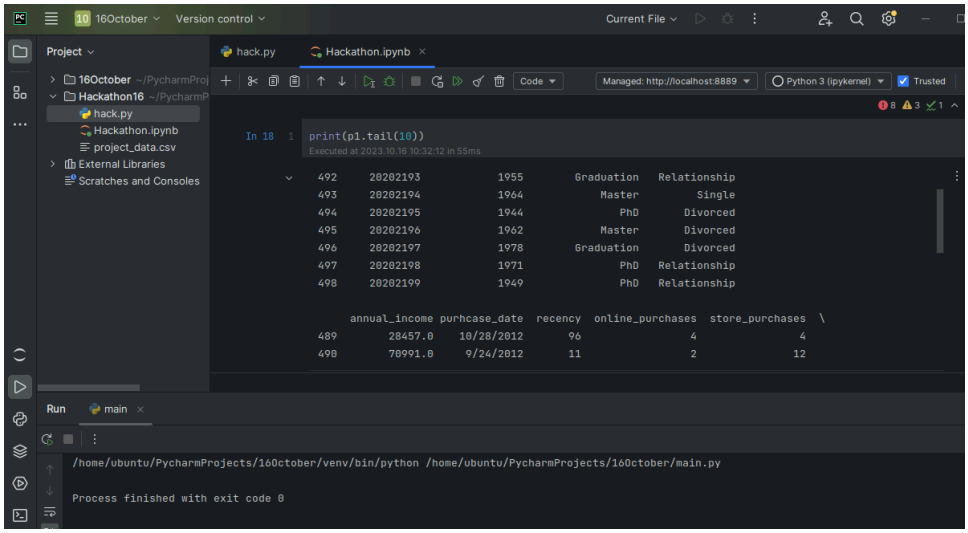
print(type(p1))

print(p1.tail(10))

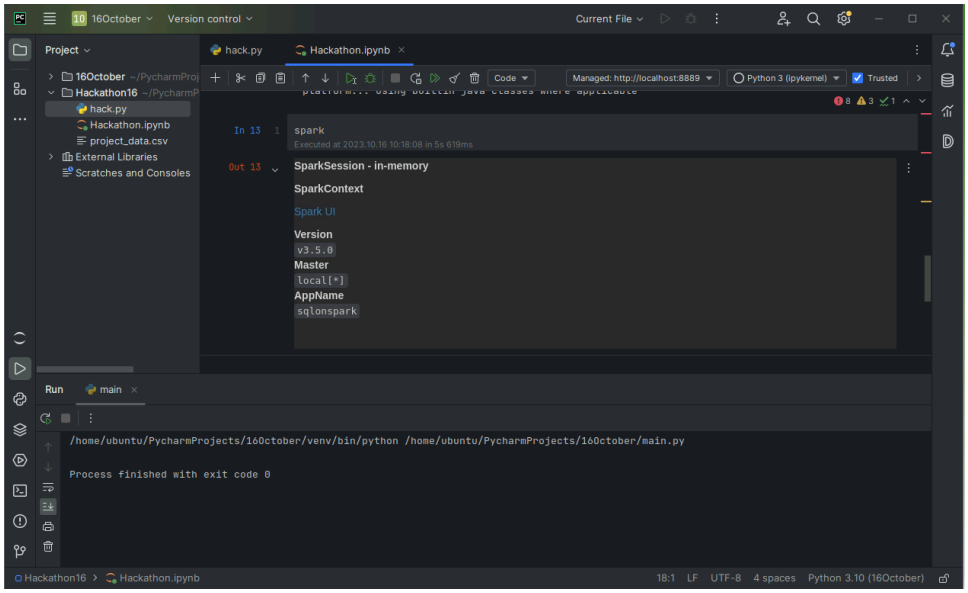
from pyspark.sql import SparkSession

spark = SparkSession.builder.appName('sqlonspark').getOrCreate()

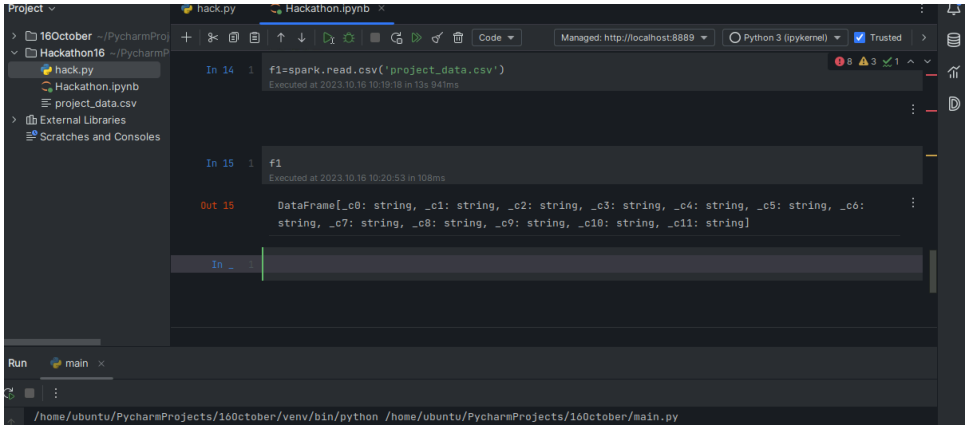




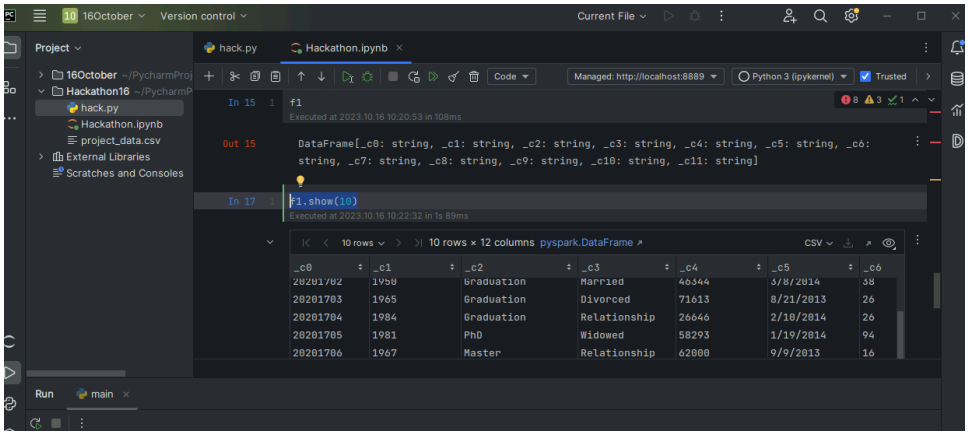
Spark



F1



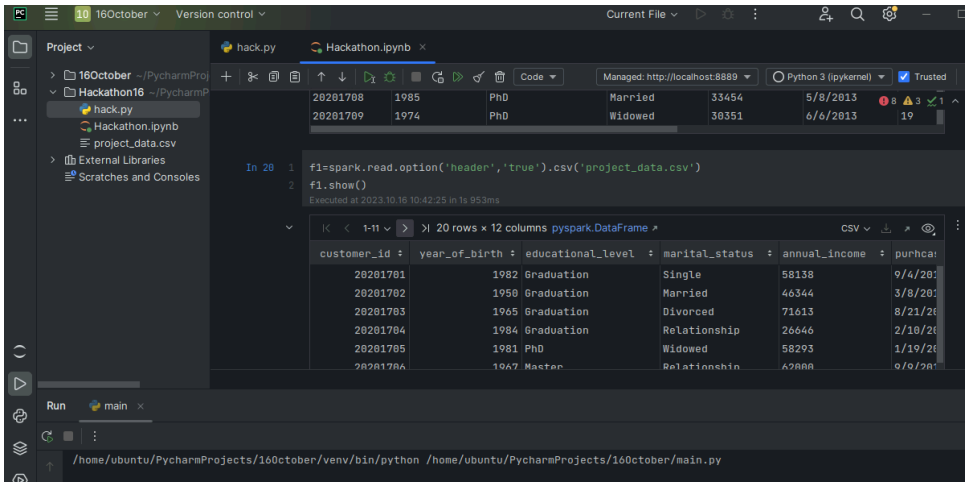
f1.show(10)



To create data frame.

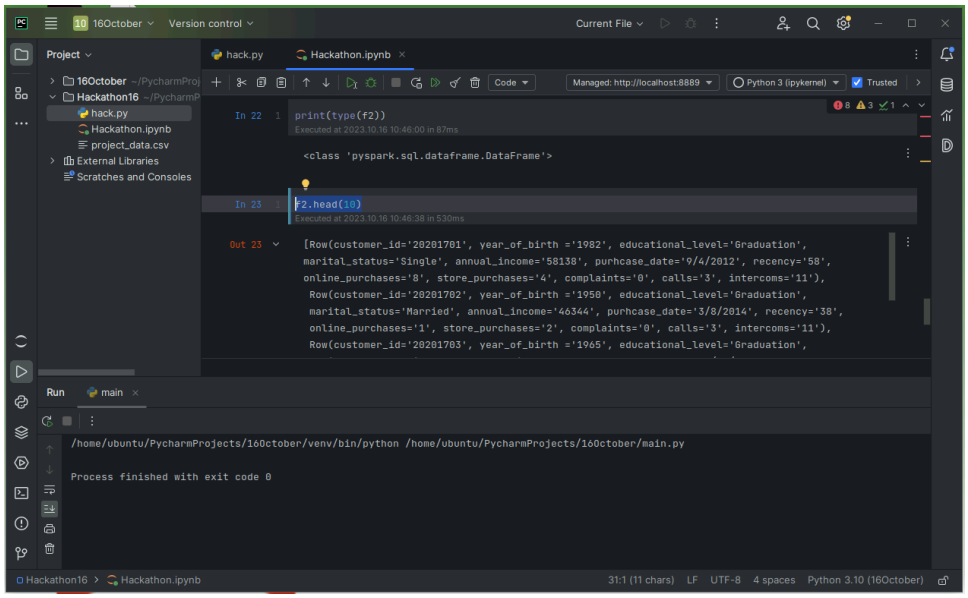
f2=spark.read.option('header','true').csv('project_data.csv')

F2.show()



print(type(f2))

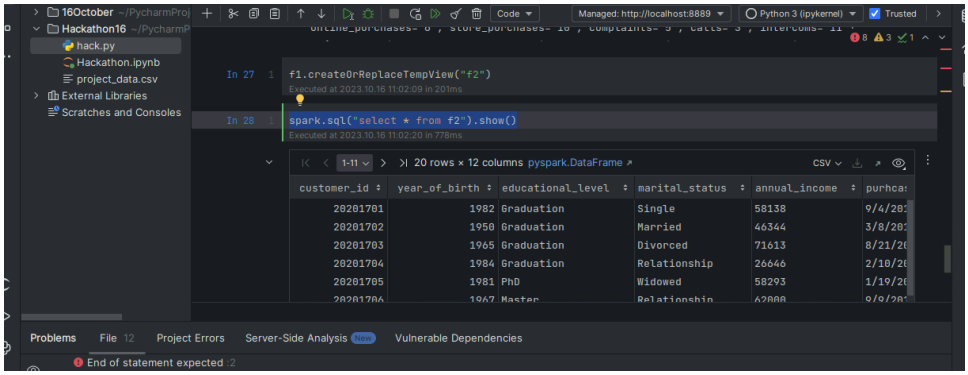
f2.head(10)



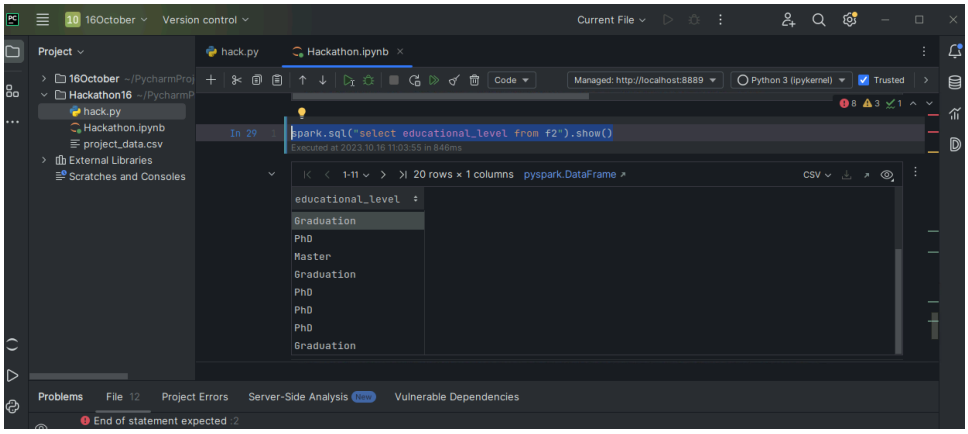
Spark sql operation

f1.createOrReplaceTempView("f2")

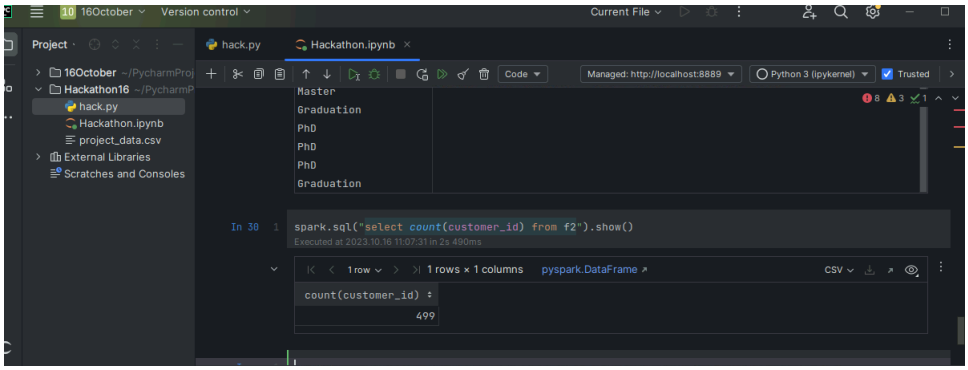
spark.sql("select * from f2").show()



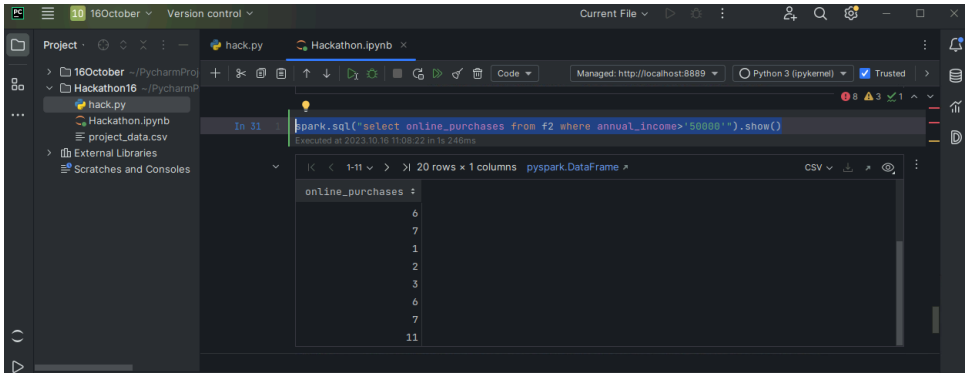
spark.sql("select educational_level from f2").show()



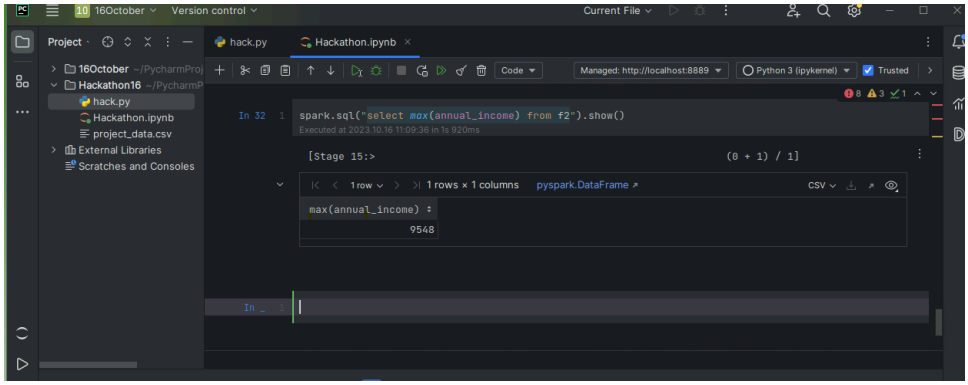
spark.sql("select count(customer_id) from f2").show()



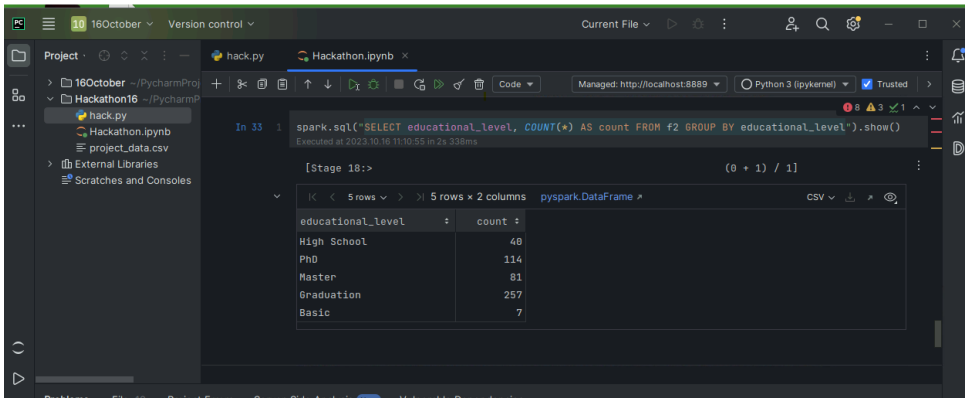
spark.sql("select online_purchases from f2 where annual_income>'50000'").show()



spark.sql("select max(annual_income) from f2").show()



spark.sql("SELECT educational_level, COUNT(*) AS count FROM f2 GROUP BY educational_level").show()



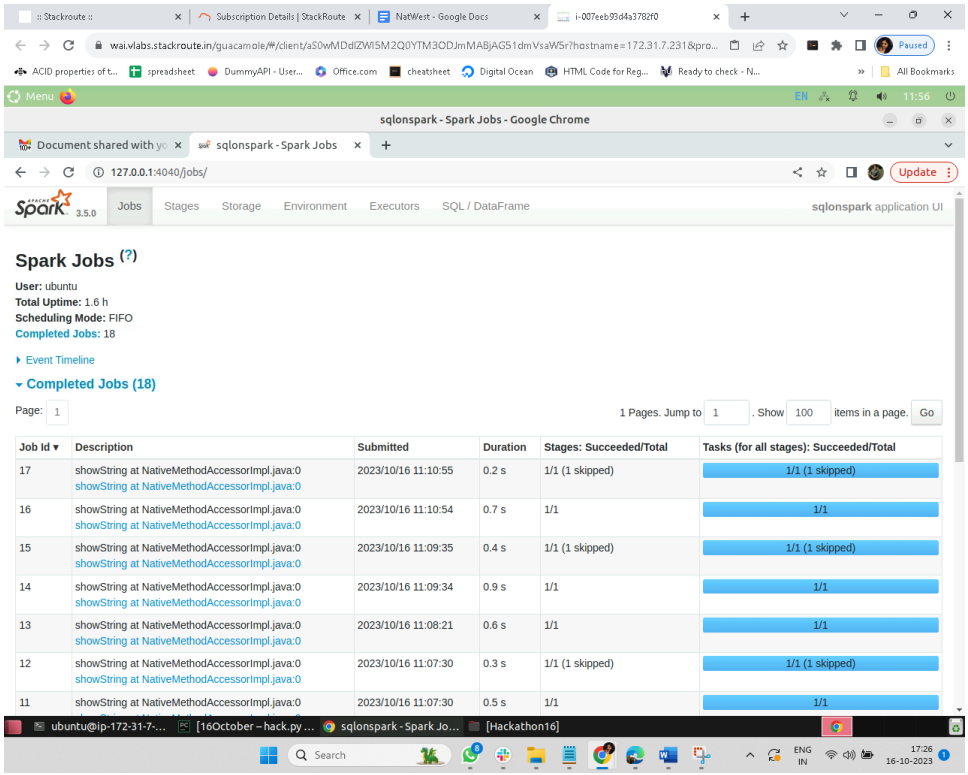
Part 2 with other csv.file

Image_1000.csv

```python
from pyspark.sql import SparkSession
from pyspark.sql.functions import *
if __name__ == "__main__":
    print('Streaming ingestion is started....')
spark=SparkSession.builder.appName('Filesys-stream').getOrCreate()
spark=SparkSession.builder.appName('Scot').getOrCreate()
stream
print(
stream_dataf.printSchema()
```

```
/home/ubuntu/PycharmProjects/16October/venv/bin/python /home/ubuntu/PycharmProjects/Hackathon16/hack.py
Streaming ingestion is started....
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
23/10/16 11:53:12 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicabl
23/10/16 11:53:14 WARN Utils: Service 'SparkUI' could not bind on port 4040. Attempting port 4041.
23/10/16 11:53:17 WARN SparkSession: Using an existing Spark session; only runtime SQL configurations will take effect.
23/10/16 11:53:17 WARN TextSocketSourceProvider: The socket source should not be used for production applications! It does not support recove
True
root
 |-- value: string (nullable = true)

Process finished with exit code 0
```



:: Stackroute :: | Subscription Details | StackRoute | NatWest - Google Docs | i-007eeb93d4a3782f0

wai.vlabs.stackroute.in/guacamole/#/client/aS0wMDdlZWI5M2Q0YTM3ODJmMABjAG51dmVsaW5r?hostname=172.31.7.231&pro...

ACID properties of t... | spreadsheet | DummyAPI - User... | Office.com | cheatsheet | Digital Ocean | HTML Code for Reg... | Ready to check - N... | All Bookmarks

Menu | EN | 11:58

sqlonspark - Spark Jobs - Google Chrome

Document shared with yo | sqlonspark - Spark Jobs

127.0.0.1:4040/jobs/ | Update

	showString at NativeMethodAccessorImpl.java:0				
11	showString at NativeMethodAccessorImpl.java:0 showString at NativeMethodAccessorImpl.java:0	2023/10/16 11:07:30	0.5 s	1/1	1/1
10	showString at NativeMethodAccessorImpl.java:0 showString at NativeMethodAccessorImpl.java:0	2023/10/16 11:03:55	0.3 s	1/1	1/1
9	showString at NativeMethodAccessorImpl.java:0 showString at NativeMethodAccessorImpl.java:0	2023/10/16 11:02:20	0.2 s	1/1	1/1
8	head at /tmp/ipykernel_4344/1773772497.py:1 head at /tmp/ipykernel_4344/1773772497.py:1	2023/10/16 10:46:38	0.2 s	1/1	1/1
7	showString at NativeMethodAccessorImpl.java:0 showString at NativeMethodAccessorImpl.java:0	2023/10/16 10:44:51	0.1 s	1/1	1/1
6	csv at NativeMethodAccessorImpl.java:0 csv at NativeMethodAccessorImpl.java:0	2023/10/16 10:44:50	0.2 s	1/1	1/1
5	showString at NativeMethodAccessorImpl.java:0 showString at NativeMethodAccessorImpl.java:0	2023/10/16 10:42:24	0.3 s	1/1	1/1
4	csv at NativeMethodAccessorImpl.java:0 csv at NativeMethodAccessorImpl.java:0	2023/10/16 10:42:24	0.2 s	1/1	1/1
3	csv at NativeMethodAccessorImpl.java:0 csv at NativeMethodAccessorImpl.java:0	2023/10/16 10:40:49	0.4 s	1/1	1/1
2	showString at NativeMethodAccessorImpl.java:0 showString at NativeMethodAccessorImpl.java:0	2023/10/16 10:22:32	0.2 s	1/1	1/1
1	showString at NativeMethodAccessorImpl.java:0 showString at NativeMethodAccessorImpl.java:0	2023/10/16 10:22:07	0.6 s	1/1	1/1
0	csv at NativeMethodAccessorImpl.java:0 csv at NativeMethodAccessorImpl.java:0	2023/10/16 10:19:16	2 s	1/1	1/1

Page: 1 1 Pages. Jump to 1 . Show 100 items in a page. Go

ubuntu@ip-172-31-7-... | [16October – hack.py ... | sqlonspark - Spark Jo... | [Hackathon16]

Search | ENG IN | 17:28 16-10-2023