ASSIGNMENT------SOLUTION SUBMISSION
ON
AZURE ANALYTICS
BY

NAME : SAI KIRAN ANCHE
ROLL NO: DXC262AB12021
BATCH:DXC-262-ANALYTICS-B12-AZURE
COMPANY – DXC TECHNOLOGY
TRAINING UNDER : MANIPAL PRO LEARN
TRAINER NAME – MR. AJAY KUMAR

DATE OF SUBMISSION : 15-06-2022
NO OF QUESTIONS :06
EMPLOYEE DOMAIN - AZURE ANALYTICS

Assignement - 15th June 2022:

1.Using archive1.zip file - please ingest data into databricks DBFS path & query the data,
redesign columns accordingly using dafarame commands - display with notebooks accordingly

2.Using archive2.zip file - please ingest data into databricks DBFS path & query the data
redesign columns accordingly using dafarame commands - display with notebooks accordingly

3.Using archive3.zip file - please ingest data into databricks DBFS path & query the data
redesign columns accordingly using dafarame commands - display with notebooks accordingly

4.Using archive4.zip file - please ingest data into databricks DBFS path & query the data
redesign columns accordingly using dafarame commands - display with notebooks accordingly

5.Using archive5.zip file - please ingest data into databricks DBFS path & query the data
redesign columns accordingly using dafarame commands - display with notebooks accordingly

6.Using archive6.zip file - please ingest data into databricks DBFS path & query the data
redesign columns accordingly using dafarame commands - display with notebooks accordingly

Please create a word / pdf document, and send it to : avyuktitraining1@gmail.com

INTRODUCTION

This Assignment is given by manipal pro learn team on the basis of the training done in the forenoon session of this morning. The main objective behind this assignment is to master the theory and enhance knowledge over creating the data bricks and performing the analytics part.

There are 6 questions and they are of same level. All the questions have been focused on what the trainer taught in the earlier session. All the demonstrations have been done successfully and documented .This assignment gave me immense confidence in mastering the domain that has been assigned to me. Special thanks to Unext team for providing the lab access.

1.Using archive1.zip file - please ingest data into databricks DBFS path & query the data,
redesign columns accordingly using dafarame commands - display with notebooks accordingly

A:

CMD1:

```
from pyspark.sql.types import StructType, StructField, IntegerType, StringType, DateType
```

CMD2:

```
country_codes_schema = StructType(fields=[StructField("FIFA", StringType(),False),
                                          StructField("Dial", StringType(),True),
                                          StructField("ISO3166-1-Alpha-3",StringType(),True),
                                          StructField("MARC", StringType(),True),
                                          StructField("is_independent", StringType(),True),
                                          StructField("ISO3166-1-numeric",IntegerType(),True),
                                          StructField("GAUL", IntegerType(),True),
                                          StructField("FIPS", StringType(),True),
                                          StructField("WMO", StringType(),True),
                                          StructField("ISO3166-1-Alpha-2",StringType(),True),
                                          StructField("ITU", StringType(),True),
                                          StructField("IOC", StringType(),True),
                                          StructField("DS", StringType(),True),
```

```
StructField("UNTERM Spanish Formal", StringType(),True),
StructField("Global Code",StringType(),True),
StructField("Intermediate Region Code",IntegerType(),True),
StructField("official_name_fr",StringType(),True),
StructField("UNTERM French Short",StringType(),True),
StructField("ISO4217-currency_name",StringType(),True),
StructField("Developed / DevelopingCountries", StringType(),                                    True),
StructField("UNTERM Russian Formal",StringType(),True),
StructField("UNTERM English Short",StringType(),True),
StructField("ISO4217-currency_alphabetic_code",StringType(),True),
StructField("Small Island Developing States (SIDS)",StringType(),True),
StructField("UNTERM Spanish Short",StringType(),True),
StructField("ISO4217-currency_numeric_code",IntegerType(),True),
StructField("UNTERM Chinese Formal",StringType(),True),
StructField("UNTERM French Formal",StringType(),True),
StructField("UNTERM Russian Short",StringType(),True),
StructField("M49",IntegerType(),True),
StructField("Sub-region Code",IntegerType(),True),
StructField("Region Code",IntegerType(),True),
StructField("official_name_ar",StringType(),True),
StructField("ISO4217-currency_minor_unit",IntegerType(),True),
```

StructField("UNTERM Arabic Formal",StringType(),True),

StructField("UNTERM Chinese Short",StringType(),True),

StructField("Land Locked Developing Countries (LLDC)",StringType(),True),

StructField("Intermediate Region Name",StringType(),True),

StructField("official_name_es",StringType(),True),

StructField("UNTERM English Formal",StringType(),True),

StructField("official_name_cn",StringType(),True),

StructField("official_name_en",StringType(),True),

StructField("ISO4217-currency_country_name",StringType(),True),

StructField("Least Developed Countries (LDC)",StringType(),True),

StructField("Region Name",StringType(),True),

StructField("UNTERM Arabic Short",StringType(),True),

StructField("Sub-region Name",StringType(),True),

StructField("official_name_ru",StringType(),True),

StructField("Global Name",StringType(),True),

StructField("Capital",StringType(),True),

StructField("Continent",StringType(),True),

StructField("TLD",StringType(),True),

StructField("Languages",StringType(),True),

StructField("Geoname ID",IntegerType(),True),

StructField("CLDR display name",StringType(),True),

```
        StructField("EDGAR",StringType(),True),

    ])
```

CMD3:

```
country_codes_df = spark.read \
.option("header" , True) \
.schema(country_codes_schema) \
.csv("/FileStore/tables/country_codes.csv")
```

CMD4:

```
from pyspark.sql.functions import current_timestamp, to_timestamp, concat, col, lit
```

CMD5:

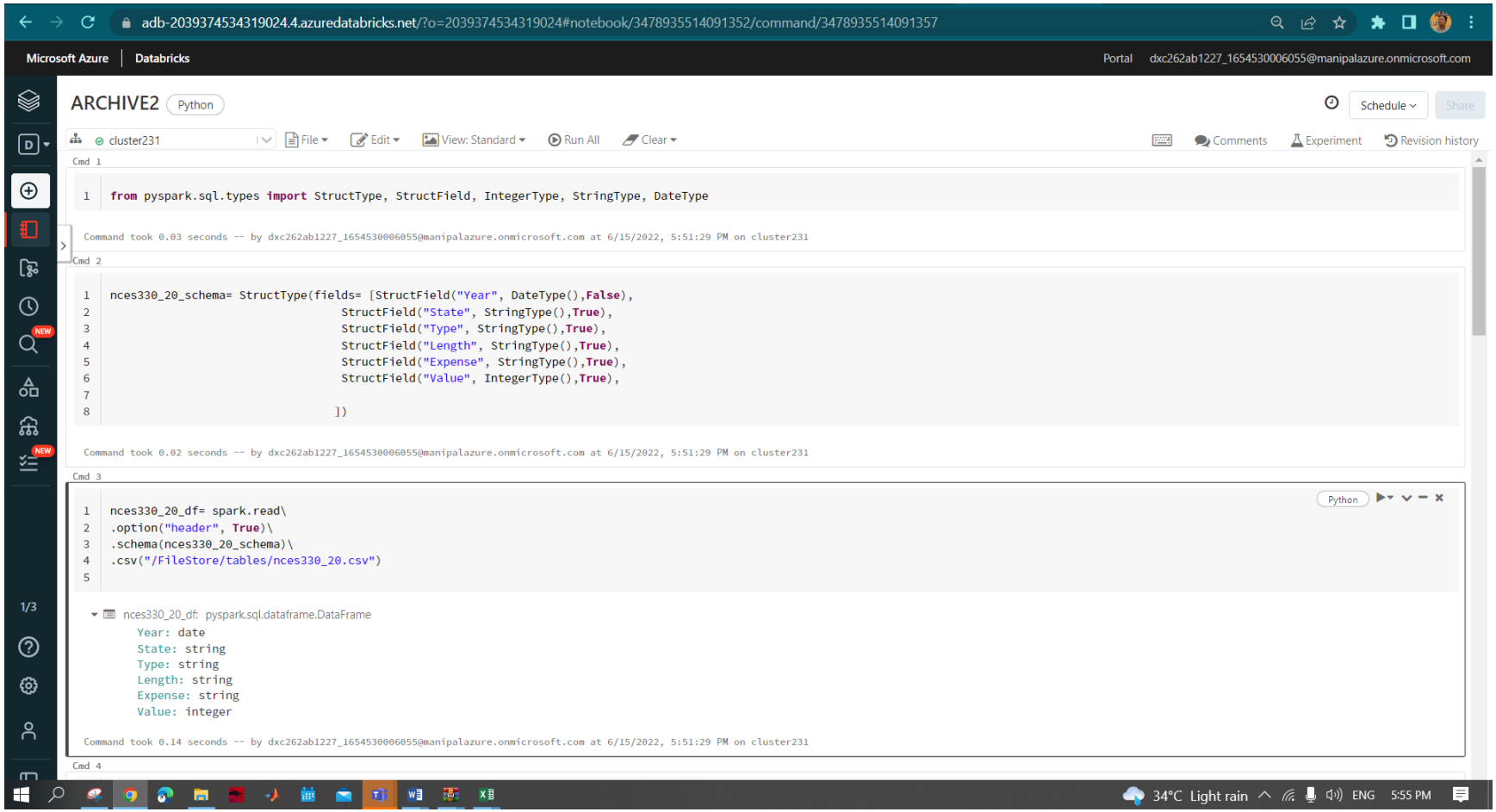
```
country_codes_selected_df = country_codes_df.select(col('FIFA'),
                col('Dial'),col('Developed / Developing Countries').alias('D/UD'),col('UNTERM Chinese Short').alias('Unterm_Chinese_Short'),col('Land Locked Developing Countries (LLDC)').alias('LLDC'),col('official_name_es'),col('Region Name'),col('EDGAR'))
```

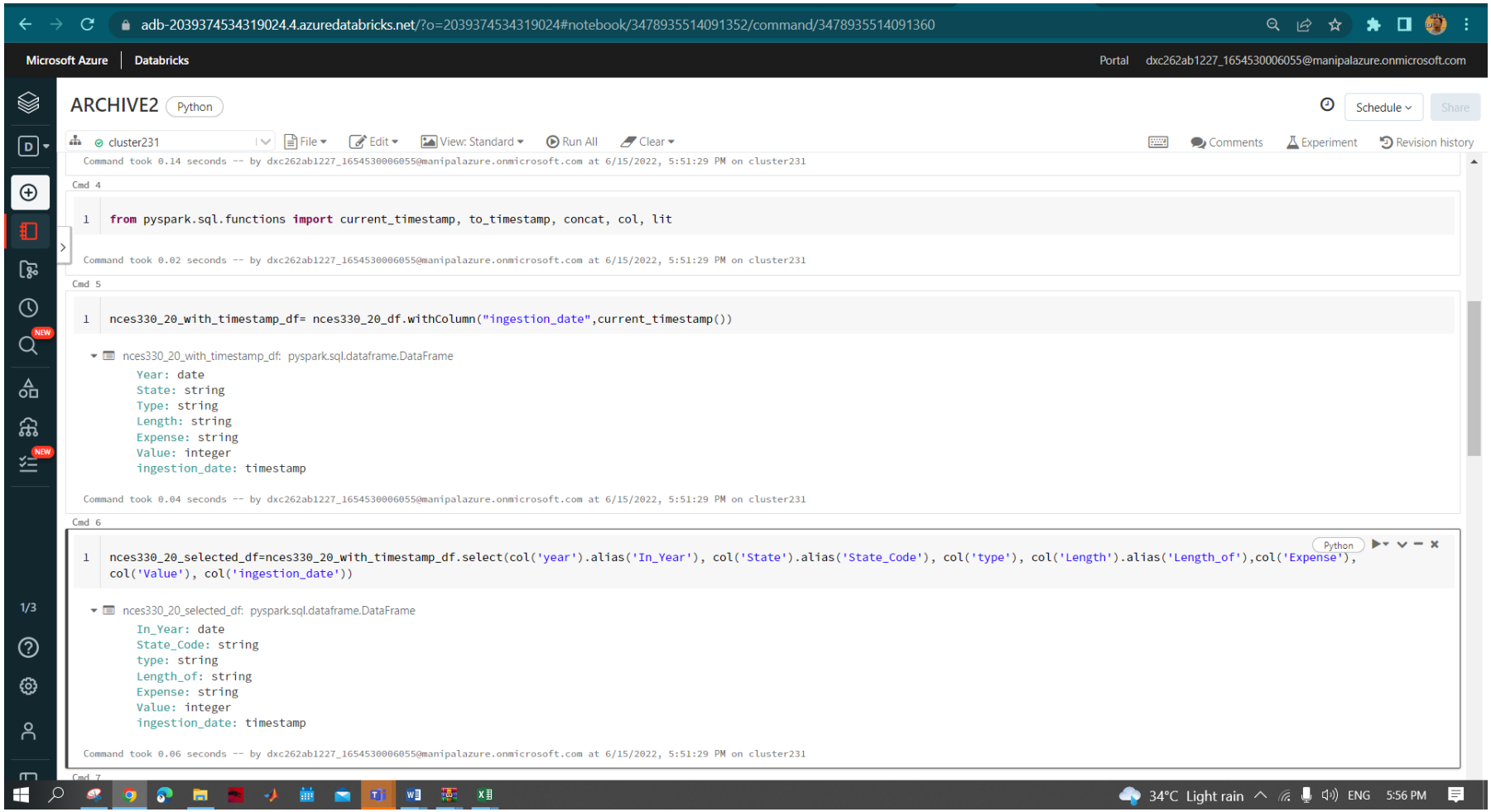
CMD6:

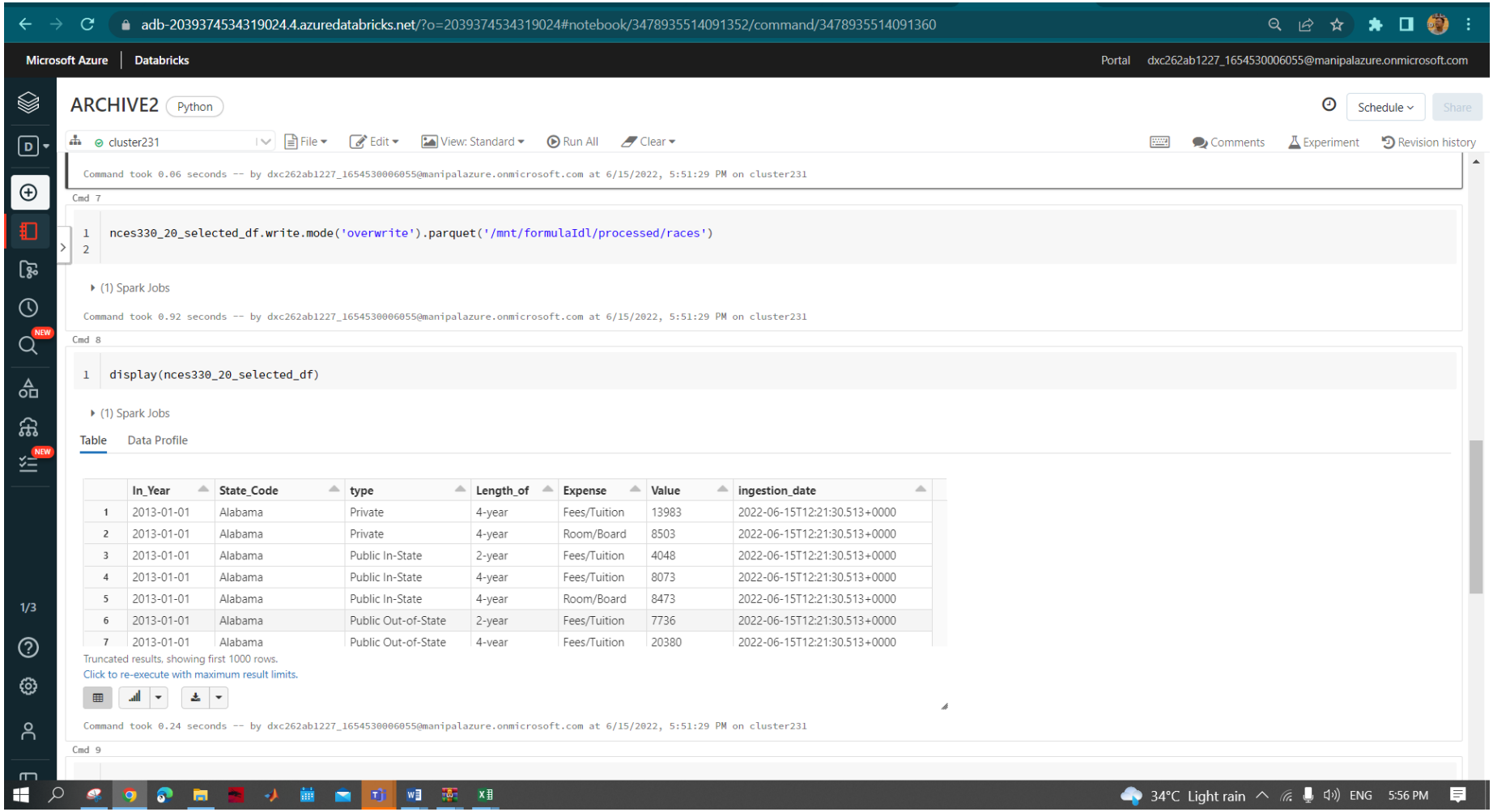
```
display(country_codes_selected_df)
```

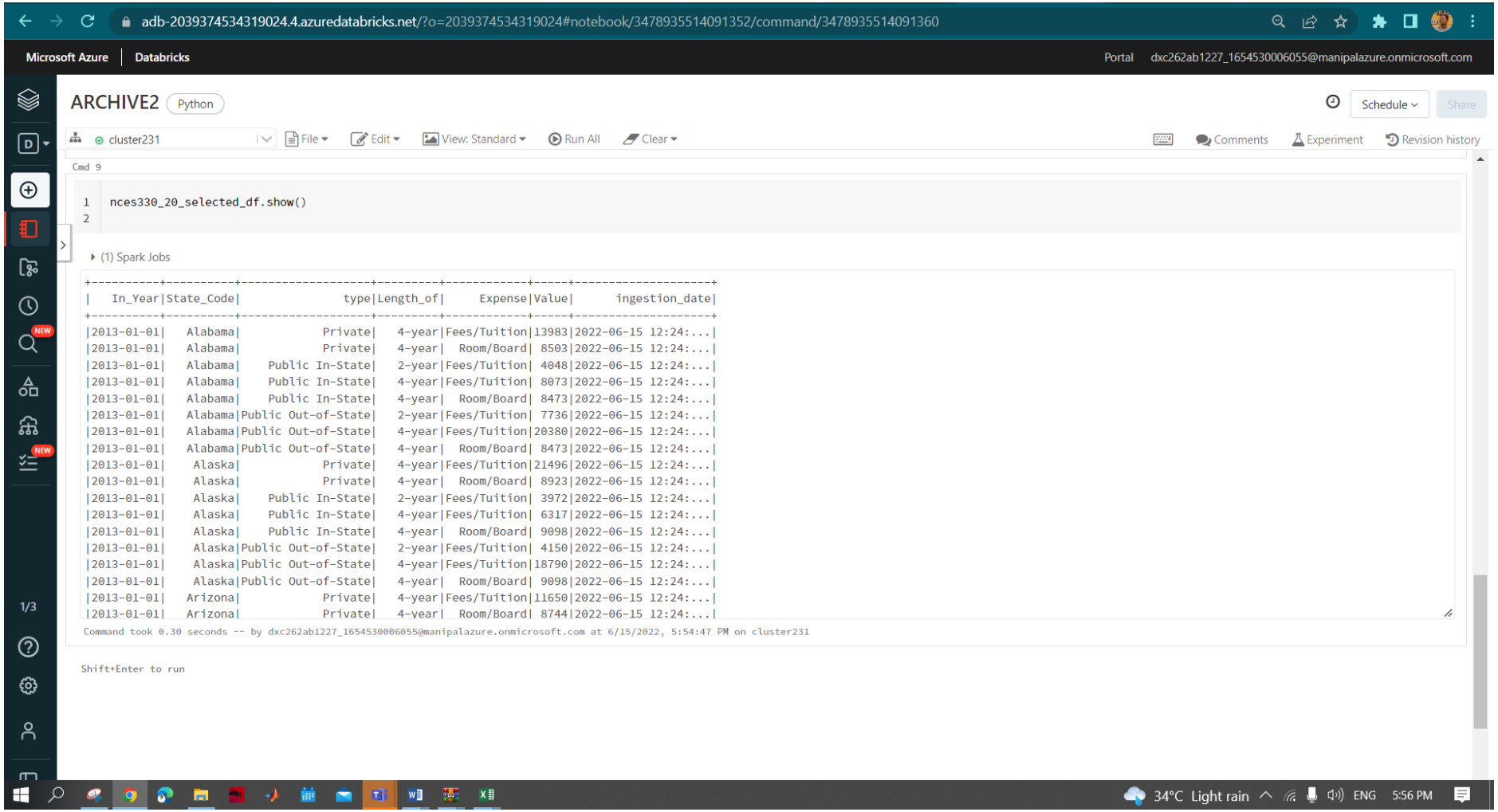
2.Using archive2.zip file - please ingest data into databricks DBFS path & query the data
redesign columns accordingly using dafarame commands - display with notebooks accordingly

A:



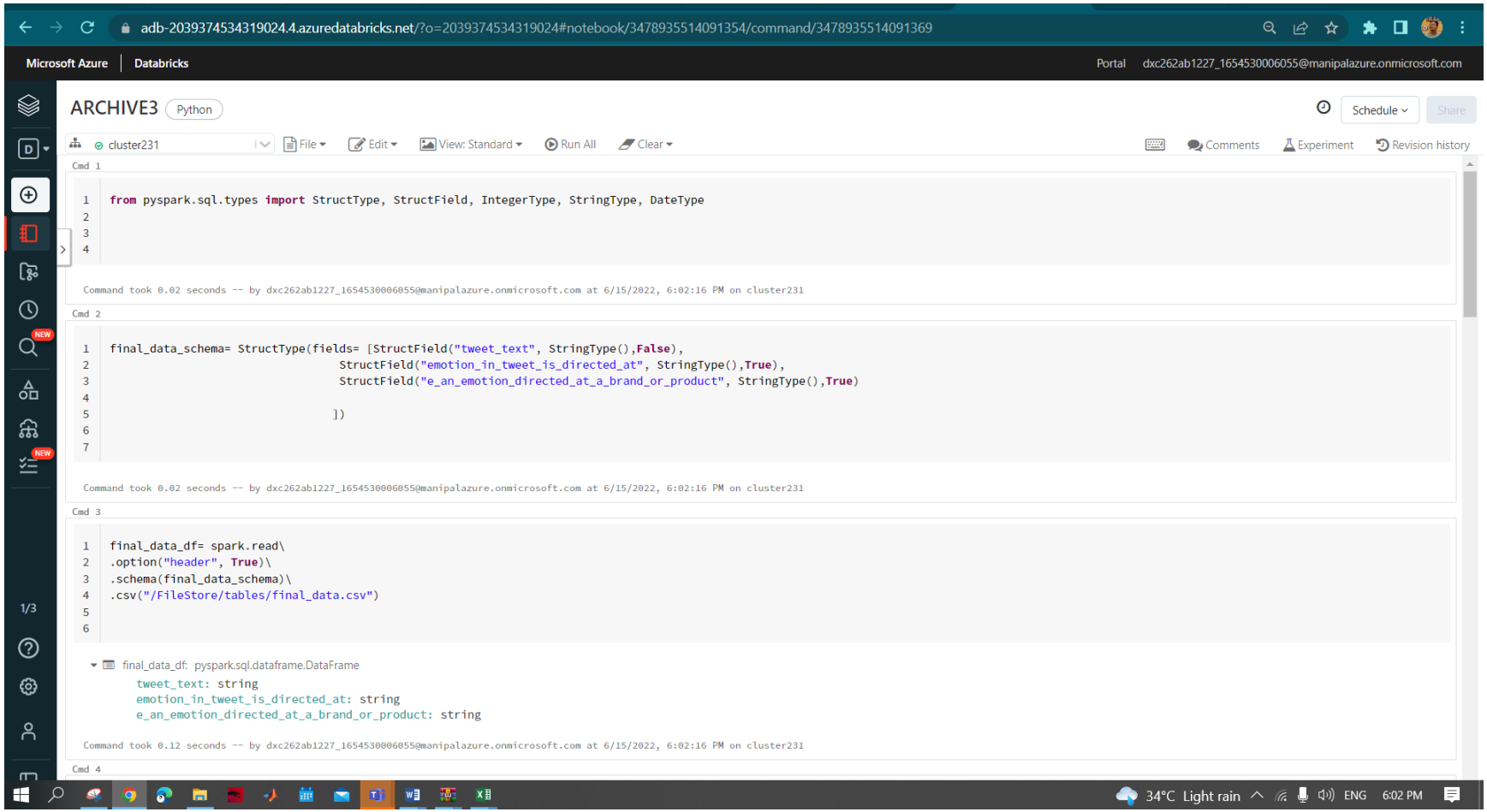

adb-2039374534319024.4.azuredatabricks.net/?o=2039374534319024#notebook/3478935514091352/command/3478935514091360
Microsoft Azure | Databricks
Portal
dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com
ARCHIVE2
Python
Schedule
Share
cluster231
File
Edit
View: Standard
Run All
Clear
Comments
Experiment
Revision history
Command took 0.14 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 5:51:29 PM on cluster231
Cmd 4
1 from pyspark.sql.functions import current_timestamp, to_timestamp, concat, col, lit
Command took 0.02 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 5:51:29 PM on cluster231
Cmd 5
1 nces330_20_with_timestamp_df= nces330_20_df.withColumn("ingestion_date",current_timestamp())
nces330_20_with_timestamp_df: pyspark.sql.dataframe.DataFrame
Year: date
State: string
Type: string
Length: string
Expense: string
Value: integer
ingestion_date: timestamp
Command took 0.04 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 5:51:29 PM on cluster231
Cmd 6
Python
1 nces330_20_selected_df=nces330_20_with_timestamp_df.select(col('year').alias('In_Year'), col('State').alias('State_Code'), col('type'), col('Length').alias('Length_of'),col('Expense'),
col('Value'), col('ingestion_date'))
nces330_20_selected_df: pyspark.sql.dataframe.DataFrame
In_Year: date
State_Code: string
type: string
Length_of: string
Expense: string
Value: integer
ingestion_date: timestamp
Command took 0.06 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 5:51:29 PM on cluster231
Cmd 7
1/3
34°C Light rain
ENG
5:56 PM

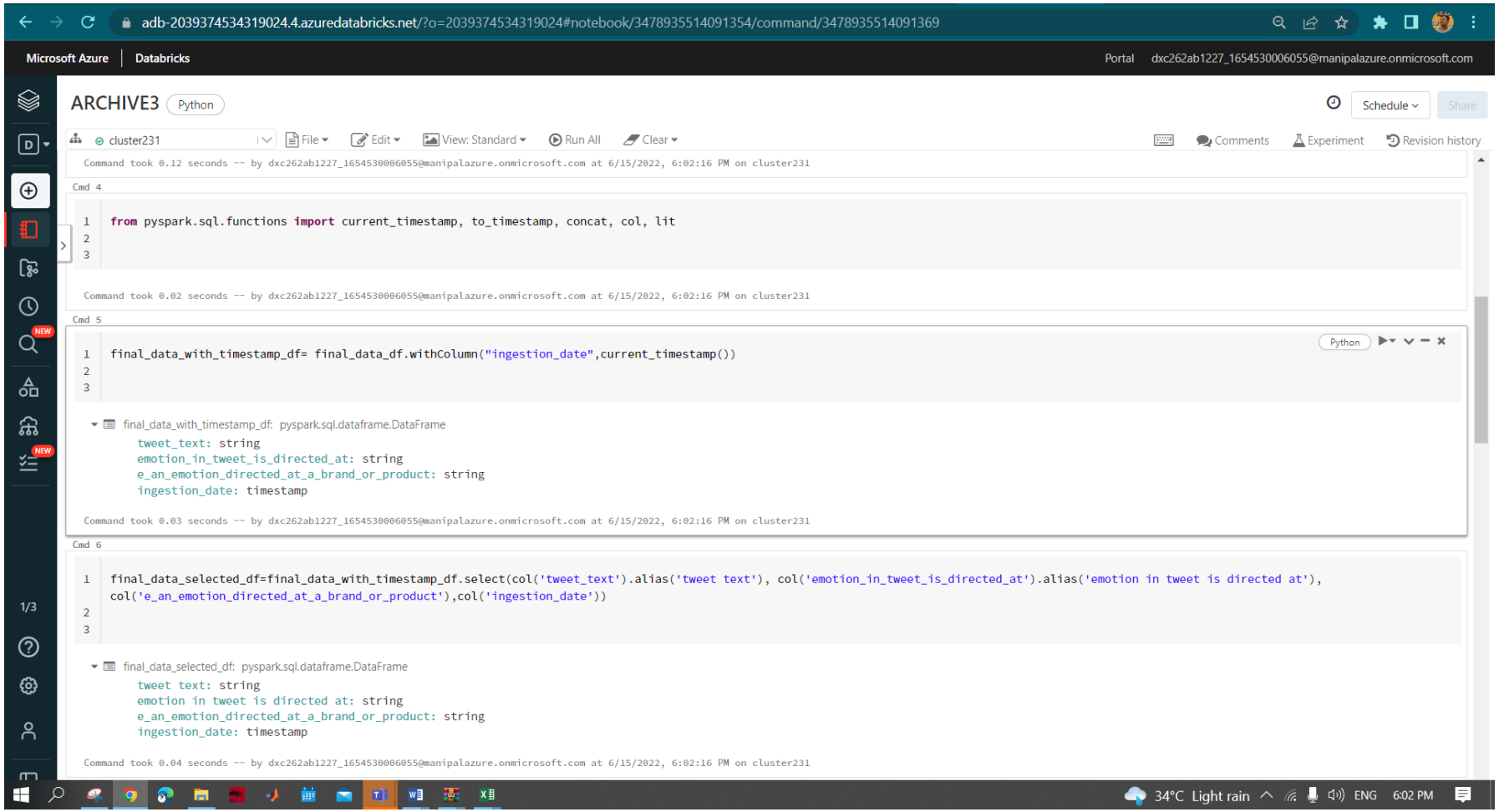

adb-2039374534319024.4.azuredatabricks.net/?o=2039374534319024#notebook/3478935514091352/command/3478935514091360
Microsoft Azure | Databricks
Portal dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com
ARCHIVE2 Python
Schedule
Share
cluster231
File
Edit
View: Standard
Run All
Clear
Comments
Experiment
Revision history
Command took 0.06 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 5:51:29 PM on cluster231
Cmd 7
1 nces330_20_selected_df.write.mode('overwrite').parquet('/mnt/formulaIdl/processed/races')
2
(1) Spark Jobs
Command took 0.92 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 5:51:29 PM on cluster231
Cmd 8
1 display(nces330_20_selected_df)
(1) Spark Jobs
Table
Data Profile
In_Year State_Code type Length_of Expense Value ingestion_date
1 2013-01-01 Alabama Private 4-year Fees/Tuition 13983 2022-06-15T12:21:30.513+0000
2 2013-01-01 Alabama Private 4-year Room/Board 8503 2022-06-15T12:21:30.513+0000
3 2013-01-01 Alabama Public In-State 2-year Fees/Tuition 4048 2022-06-15T12:21:30.513+0000
4 2013-01-01 Alabama Public In-State 4-year Fees/Tuition 8073 2022-06-15T12:21:30.513+0000
5 2013-01-01 Alabama Public In-State 4-year Room/Board 8473 2022-06-15T12:21:30.513+0000
6 2013-01-01 Alabama Public Out-of-State 2-year Fees/Tuition 7736 2022-06-15T12:21:30.513+0000
7 2013-01-01 Alabama Public Out-of-State 4-year Fees/Tuition 20380 2022-06-15T12:21:30.513+0000
Truncated results, showing first 1000 rows.
Click to re-execute with maximum result limits.
Command took 0.24 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 5:51:29 PM on cluster231
Cmd 9
34°C Light rain
ENG
5:56 PM

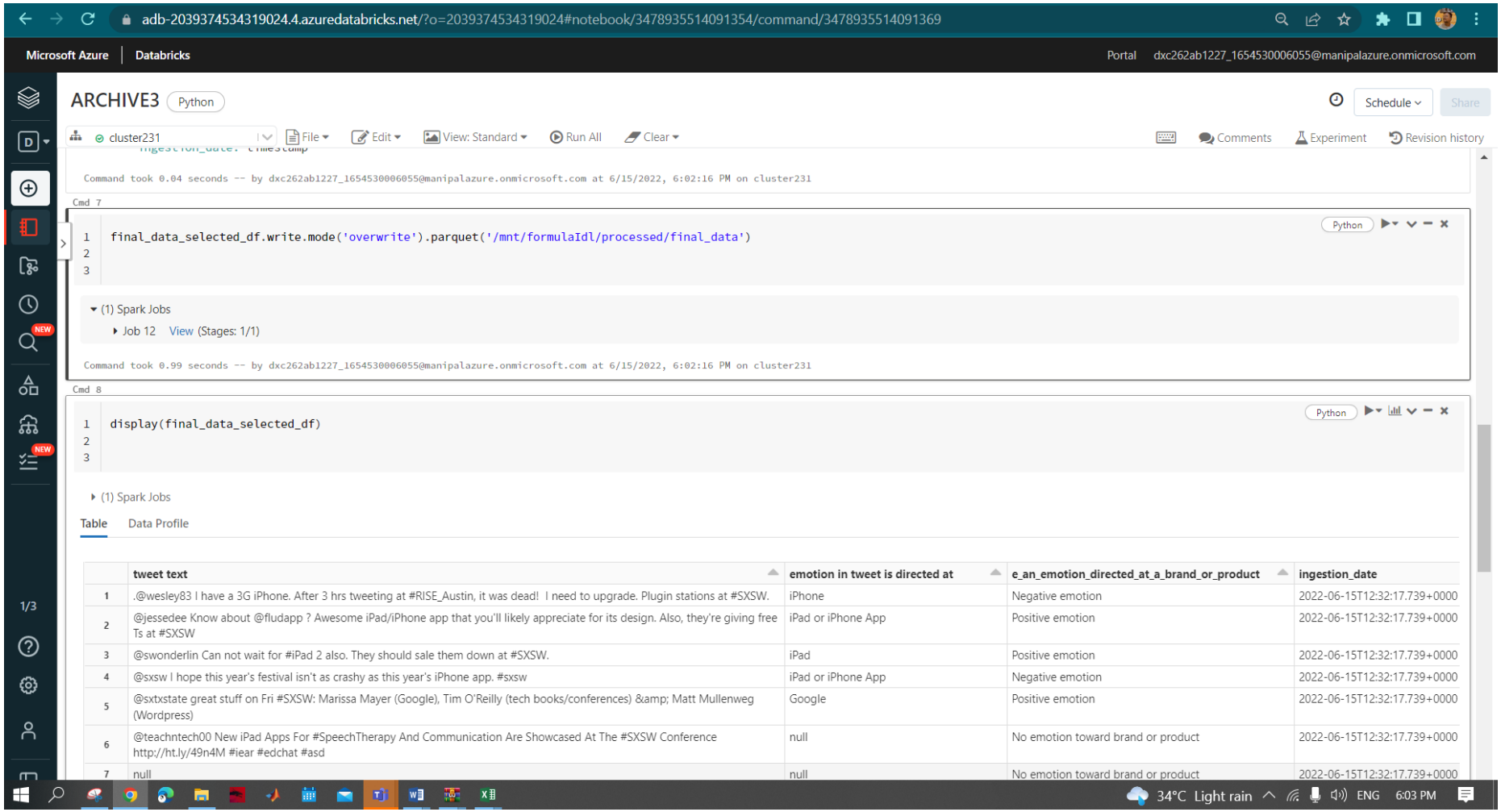

adb-2039374534319024.4.azuredatabricks.net/?o=2039374534319024#notebook/3478935514091352/command/3478935514091360
Microsoft Azure | Databricks
Portal
dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com
ARCHIVE2
Python
Schedule
Share
cluster231
File
Edit
View: Standard
Run All
Clear
Comments
Experiment
Revision history
Cmd 9
1 nces330_20_selected_df.show()
2
(1) Spark Jobs
+----------+----------+--------------------+---------+------------+-----+--------------------+
| In_Year|State_Code| type|Length_of| Expense|Value| ingestion_date|
+----------+----------+--------------------+---------+------------+-----+--------------------+
|2013-01-01| Alabama| Private| 4-year|Fees/Tuition|13983|2022-06-15 12:24:...|
|2013-01-01| Alabama| Private| 4-year| Room/Board| 8503|2022-06-15 12:24:...|
|2013-01-01| Alabama| Public In-State| 2-year|Fees/Tuition| 4048|2022-06-15 12:24:...|
|2013-01-01| Alabama| Public In-State| 4-year|Fees/Tuition| 8073|2022-06-15 12:24:...|
|2013-01-01| Alabama| Public In-State| 4-year| Room/Board| 8473|2022-06-15 12:24:...|
|2013-01-01| Alabama|Public Out-of-State| 2-year|Fees/Tuition| 7736|2022-06-15 12:24:...|
|2013-01-01| Alabama|Public Out-of-State| 4-year|Fees/Tuition|20380|2022-06-15 12:24:...|
|2013-01-01| Alabama|Public Out-of-State| 4-year| Room/Board| 8473|2022-06-15 12:24:...|
|2013-01-01| Alaska| Private| 4-year|Fees/Tuition|21496|2022-06-15 12:24:...|
|2013-01-01| Alaska| Private| 4-year| Room/Board| 8923|2022-06-15 12:24:...|
|2013-01-01| Alaska| Public In-State| 2-year|Fees/Tuition| 3972|2022-06-15 12:24:...|
|2013-01-01| Alaska| Public In-State| 4-year|Fees/Tuition| 6317|2022-06-15 12:24:...|
|2013-01-01| Alaska| Public In-State| 4-year| Room/Board| 9098|2022-06-15 12:24:...|
|2013-01-01| Alaska|Public Out-of-State| 2-year|Fees/Tuition| 4150|2022-06-15 12:24:...|
|2013-01-01| Alaska|Public Out-of-State| 4-year|Fees/Tuition|18790|2022-06-15 12:24:...|
|2013-01-01| Alaska|Public Out-of-State| 4-year| Room/Board| 9098|2022-06-15 12:24:...|
|2013-01-01| Arizona| Private| 4-year|Fees/Tuition|11650|2022-06-15 12:24:...|
|2013-01-01| Arizona| Private| 4-year| Room/Board| 8744|2022-06-15 12:24:...|
Command took 0.30 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 5:54:47 PM on cluster231
Shift+Enter to run
1/3
34°C Light rain
ENG
5:56 PM

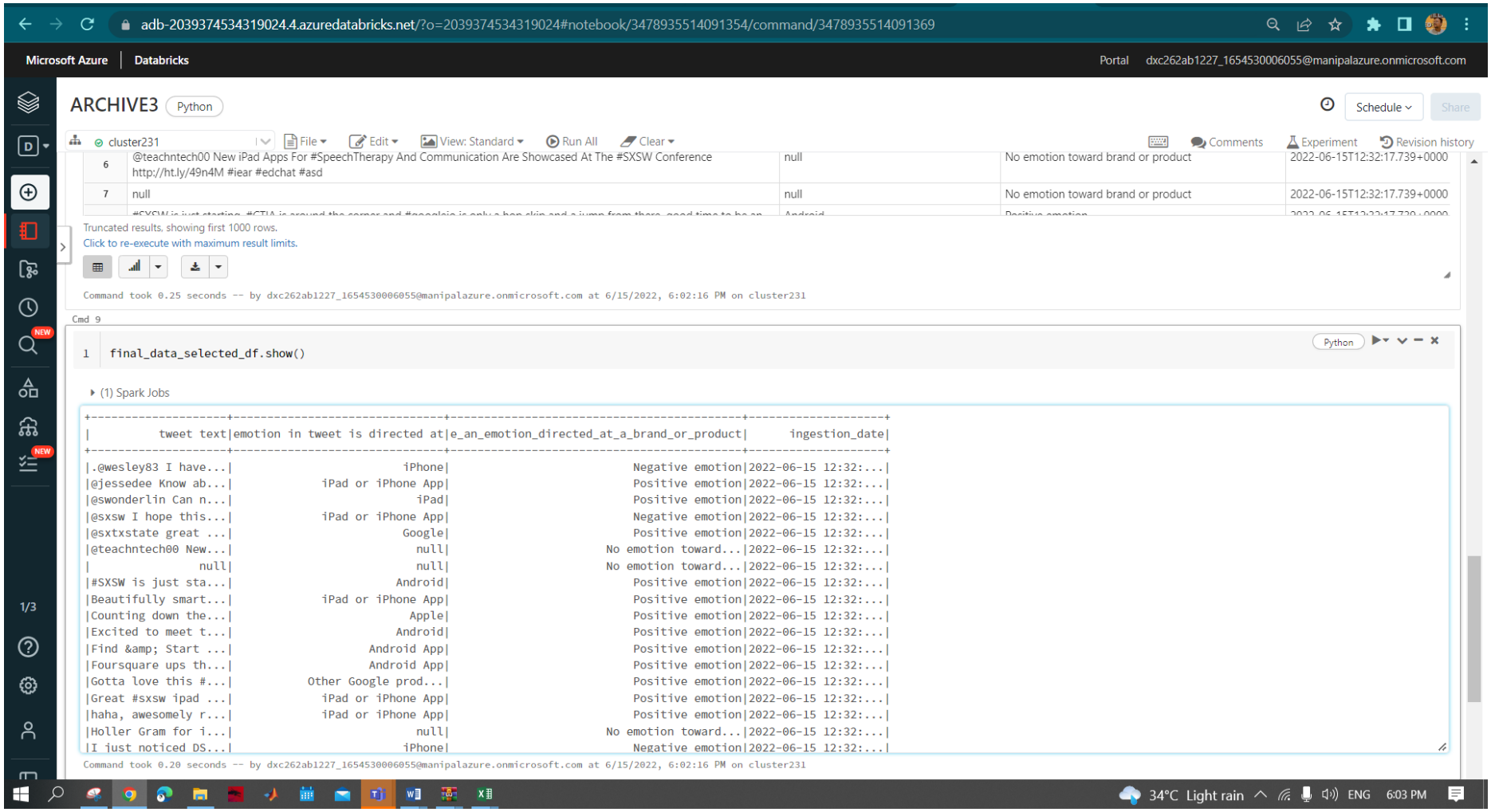
3.Using archive3.zip file - please ingest data into databricks DBFS path & query the data
redesign columns accordingly using dafarame commands - display with notebooks accordingly

A:



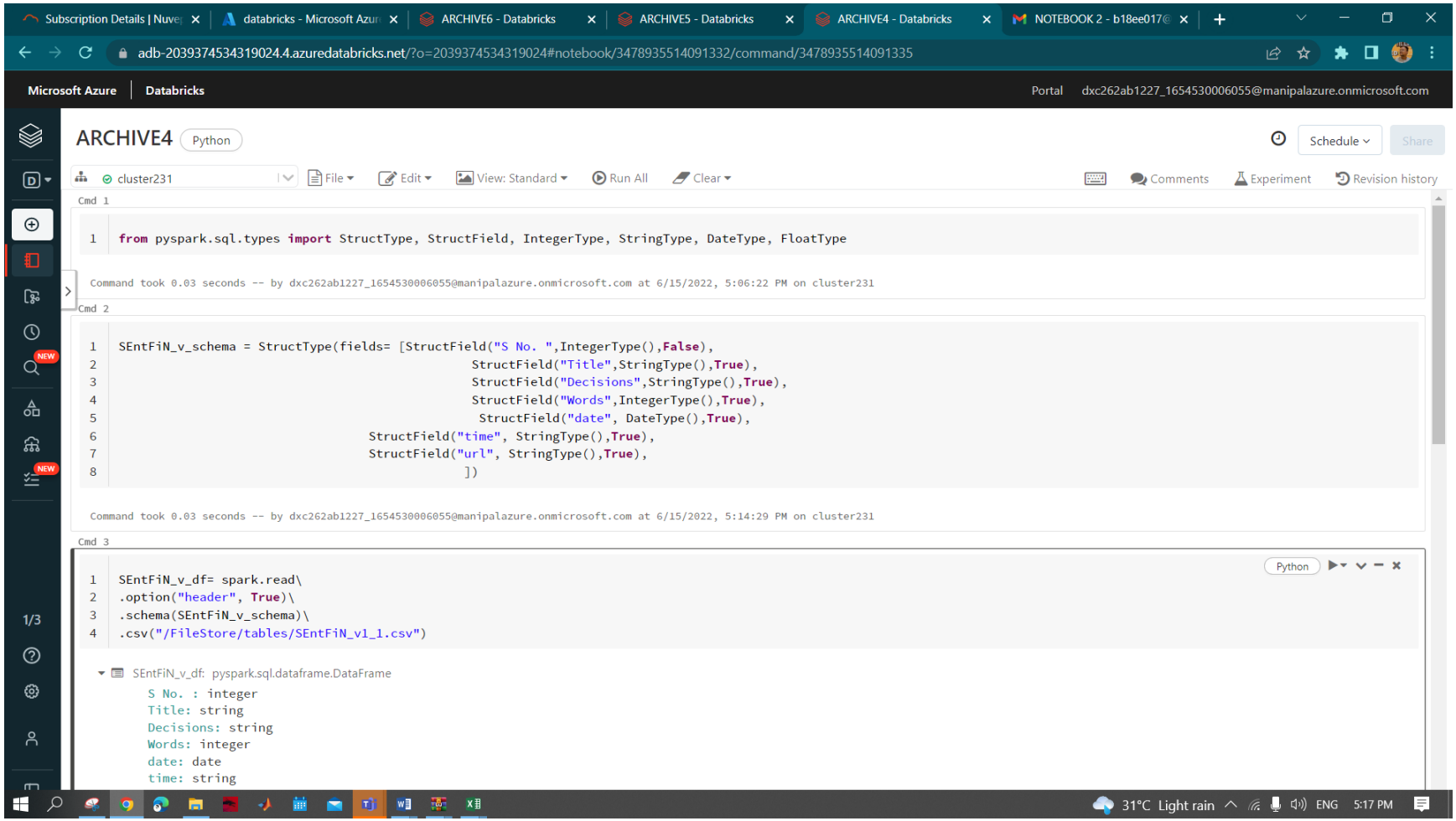

adb-2039374534319024.4.azuredatabricks.net/?o=2039374534319024#notebook/3478935514091354/command/3478935514091369
Microsoft Azure | Databricks
Portal dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com
ARCHIVE3 Python
Schedule
Share
cluster231
File
Edit
View: Standard
Run All
Clear
Comments
Experiment
Revision history
Command took 0.12 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 6:02:16 PM on cluster231
Cmd 4
from pyspark.sql.functions import current_timestamp, to_timestamp, concat, col, lit
Command took 0.02 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 6:02:16 PM on cluster231
Cmd 5
final_data_with_timestamp_df= final_data_df.withColumn("ingestion_date",current_timestamp())
Python
final_data_with_timestamp_df: pyspark.sql.dataframe.DataFrame
tweet_text: string
emotion_in_tweet_is_directed_at: string
e_an_emotion_directed_at_a_brand_or_product: string
ingestion_date: timestamp
Command took 0.03 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 6:02:16 PM on cluster231
Cmd 6
final_data_selected_df=final_data_with_timestamp_df.select(col('tweet_text').alias('tweet text'), col('emotion_in_tweet_is_directed_at').alias('emotion in tweet is directed at'),
col('e_an_emotion_directed_at_a_brand_or_product'),col('ingestion_date'))
final_data_selected_df: pyspark.sql.dataframe.DataFrame
tweet text: string
emotion in tweet is directed at: string
e_an_emotion_directed_at_a_brand_or_product: string
ingestion_date: timestamp
Command took 0.04 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 6:02:16 PM on cluster231
1/3
34°C Light rain
ENG
6:02 PM

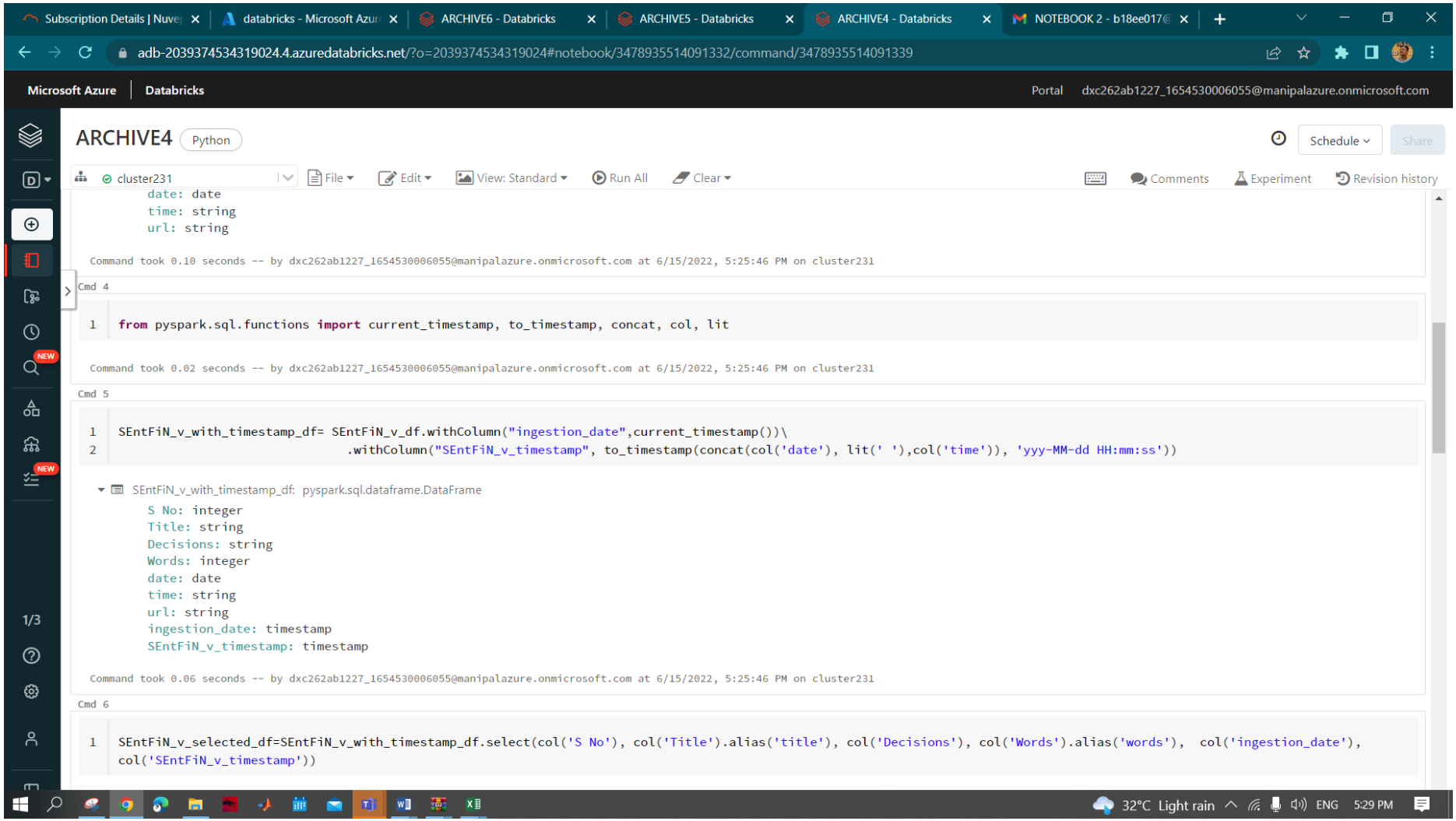

adb-2039374534319024.4.azuredatabricks.net/?o=2039374534319024#notebook/3478935514091354/command/3478935514091369
Microsoft Azure | Databricks
Portal dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com
ARCHIVE3 Python
Schedule
Share
cluster231
File
Edit
View: Standard
Run All
Clear
Comments
Experiment
Revision history
Command took 0.04 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 6:02:16 PM on cluster231
Cmd 7
Python
final_data_selected_df.write.mode('overwrite').parquet('/mnt/formulaIdl/processed/final_data')
(1) Spark Jobs
Job 12 View (Stages: 1/1)
Command took 0.99 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 6:02:16 PM on cluster231
Cmd 8
Python
display(final_data_selected_df)
(1) Spark Jobs
Table
Data Profile
tweet text
emotion in tweet is directed at
e_an_emotion_directed_at_a_brand_or_product
ingestion_date
1 .@wesley83 I have a 3G iPhone. After 3 hrs tweeting at #RISE_Austin, it was dead! I need to upgrade. Plugin stations at #SXSW. iPhone Negative emotion 2022-06-15T12:32:17.739+0000
2 @jessedee Know about @fludapp ? Awesome iPad/iPhone app that you'll likely appreciate for its design. Also, they're giving free Ts at #SXSW iPad or iPhone App Positive emotion 2022-06-15T12:32:17.739+0000
3 @swonderlin Can not wait for #iPad 2 also. They should sale them down at #SXSW. iPad Positive emotion 2022-06-15T12:32:17.739+0000
4 @sxsw I hope this year's festival isn't as crashy as this year's iPhone app. #sxsw iPad or iPhone App Negative emotion 2022-06-15T12:32:17.739+0000
5 @sxtxstate great stuff on Fri #SXSW: Marissa Mayer (Google), Tim O'Reilly (tech books/conferences) & Matt Mullenweg (Wordpress) Google Positive emotion 2022-06-15T12:32:17.739+0000
6 @teachntech00 New iPad Apps For #SpeechTherapy And Communication Are Showcased At The #SXSW Conference http://ht.ly/49n4M #iear #edchat #asd null No emotion toward brand or product 2022-06-15T12:32:17.739+0000
7 null null No emotion toward brand or product 2022-06-15T12:32:17.739+0000
1/3
34°C Light rain
ENG
6:03 PM

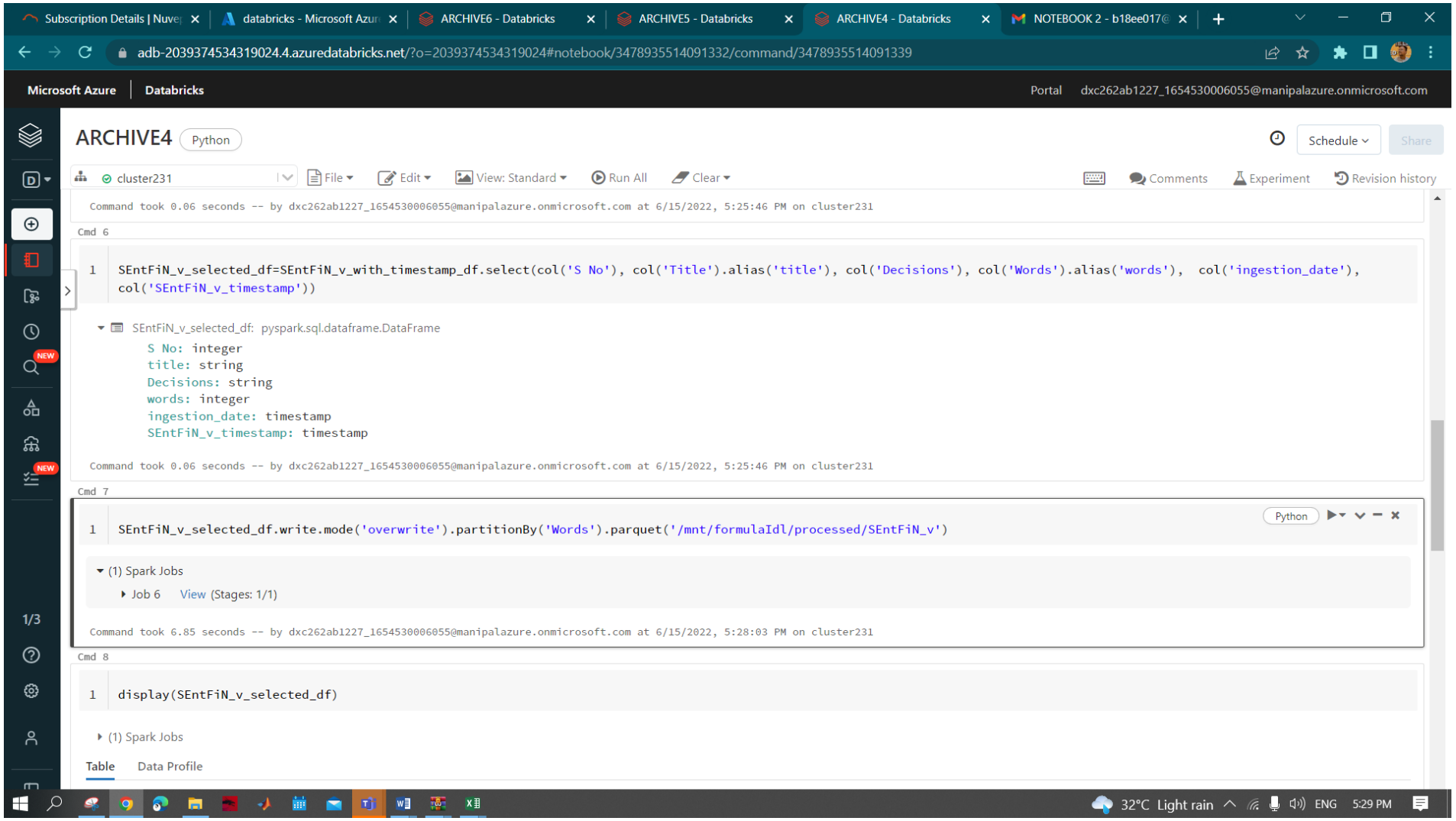

adb-2039374534319024.4.azuredatabricks.net/?o=2039374534319024#notebook/3478935514091354/command/3478935514091369
Microsoft Azure | Databricks
Portal dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com
ARCHIVE3 Python
Schedule
Share
cluster231
File
Edit
View: Standard
Run All
Clear
Comments
Experiment
Revision history
6 @teachntech00 New iPad Apps For #SpeechTherapy And Communication Are Showcased At The #SXSW Conference http://ht.ly/49n4M #iear #edchat #asd null No emotion toward brand or product 2022-06-15T12:32:17.739+0000
7 null null No emotion toward brand or product 2022-06-15T12:32:17.739+0000
Truncated results, showing first 1000 rows.
Click to re-execute with maximum result limits.
Command took 0.25 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 6:02:16 PM on cluster231
Cmd 9
1 final_data_selected_df.show()
Python
(1) Spark Jobs
+----------+----------+----------+----------+
| tweet text|emotion in tweet is directed at|e_an_emotion_directed_at_a_brand_or_product| ingestion_date|
+----------+----------+----------+----------+
|.@wesley83 I have...| iPhone| Negative emotion|2022-06-15 12:32:...|
|@jessedee Know ab...| iPad or iPhone App| Positive emotion|2022-06-15 12:32:...|
|@swonderlin Can n...| iPad| Positive emotion|2022-06-15 12:32:...|
|@sxsw I hope this...| iPad or iPhone App| Negative emotion|2022-06-15 12:32:...|
|@sxtxstate great ...| Google| Positive emotion|2022-06-15 12:32:...|
|@teachntech00 New...| null| No emotion toward...|2022-06-15 12:32:...|
| null| null| No emotion toward...|2022-06-15 12:32:...|
|#SXSW is just sta...| Android| Positive emotion|2022-06-15 12:32:...|
|Beautifully smart...| iPad or iPhone App| Positive emotion|2022-06-15 12:32:...|
|Counting down the...| Apple| Positive emotion|2022-06-15 12:32:...|
|Excited to meet t...| Android| Positive emotion|2022-06-15 12:32:...|
|Find & Start ...| Android App| Positive emotion|2022-06-15 12:32:...|
|Foursquare ups th...| Android App| Positive emotion|2022-06-15 12:32:...|
|Gotta love this #...| Other Google prod...| Positive emotion|2022-06-15 12:32:...|
|Great #sxsw ipad ...| iPad or iPhone App| Positive emotion|2022-06-15 12:32:...|
|haha, awesomely r...| iPad or iPhone App| Positive emotion|2022-06-15 12:32:...|
|Holler Gram for i...| null| No emotion toward...|2022-06-15 12:32:...|
|I just noticed DS...| iPhone| Negative emotion|2022-06-15 12:32:...|
Command took 0.20 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 6:02:16 PM on cluster231
1/3
34°C Light rain
ENG
6:03 PM

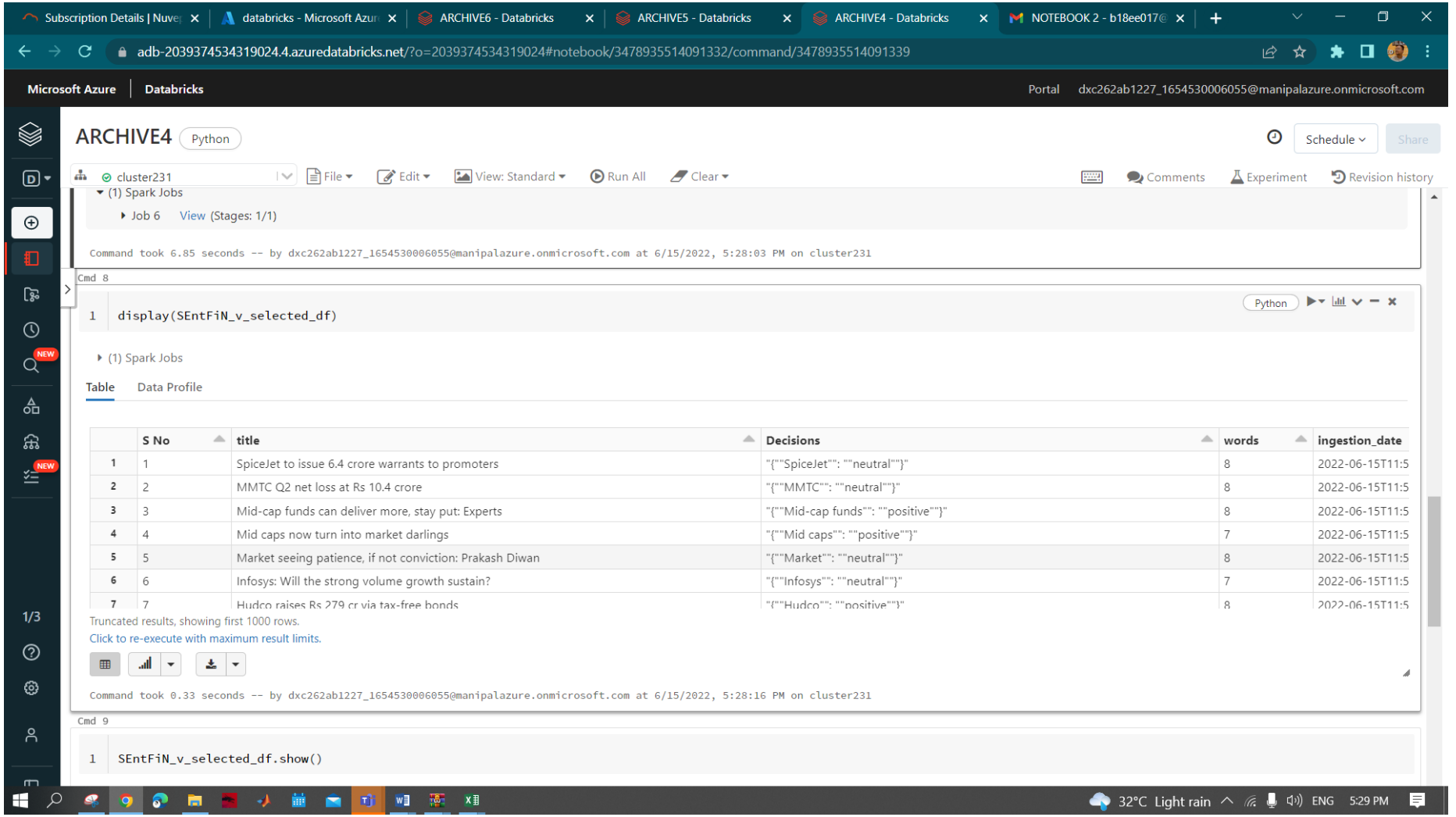
4.Using archive4.zip file - please ingest data into databricks DBFS path & query the data
redesign columns accordingly using dafarame commands - display with notebooks accordingly

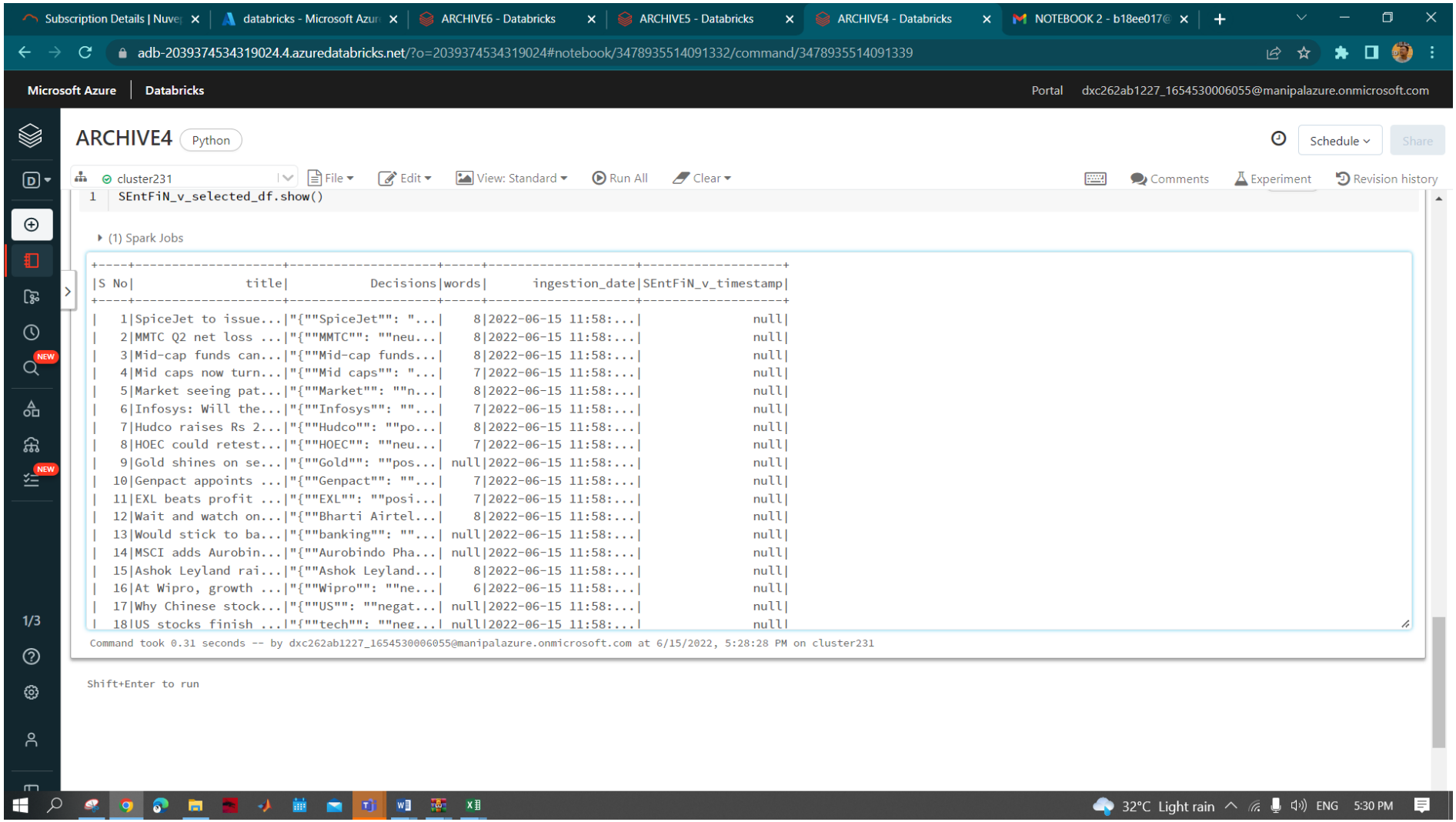
A:




Subscription Details | Nuve
databricks - Microsoft Azur
ARCHIVE6 - Databricks
ARCHIVE5 - Databricks
ARCHIVE4 - Databricks
NOTEBOOK 2 - b18ee017
adb-2039374534319024.4.azuredatabricks.net/?o=2039374534319024#notebook/3478935514091332/command/3478935514091339
Microsoft Azure
Databricks
Portal
dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com
ARCHIVE4
Python
Schedule
Share
cluster231
File
Edit
View: Standard
Run All
Clear
Comments
Experiment
Revision history
date: date
time: string
url: string
Command took 0.10 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 5:25:46 PM on cluster231
Cmd 4
1 from pyspark.sql.functions import current_timestamp, to_timestamp, concat, col, lit
Command took 0.02 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 5:25:46 PM on cluster231
Cmd 5
1 SEntFiN_v_with_timestamp_df= SEntFiN_v_df.withColumn("ingestion_date",current_timestamp())\
2 .withColumn("SEntFiN_v_timestamp", to_timestamp(concat(col('date'), lit(' '),col('time')), 'yyy-MM-dd HH:mm:ss'))
SEntFiN_v_with_timestamp_df: pyspark.sql.dataframe.DataFrame
S No: integer
Title: string
Decisions: string
Words: integer
date: date
time: string
url: string
ingestion_date: timestamp
SEntFiN_v_timestamp: timestamp
Command took 0.06 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 5:25:46 PM on cluster231
Cmd 6
1 SEntFiN_v_selected_df=SEntFiN_v_with_timestamp_df.select(col('S No'), col('Title').alias('title'), col('Decisions'), col('Words').alias('words'), col('ingestion_date'),
col('SEntFiN_v_timestamp'))
1/3
32°C Light rain
ENG
5:29 PM

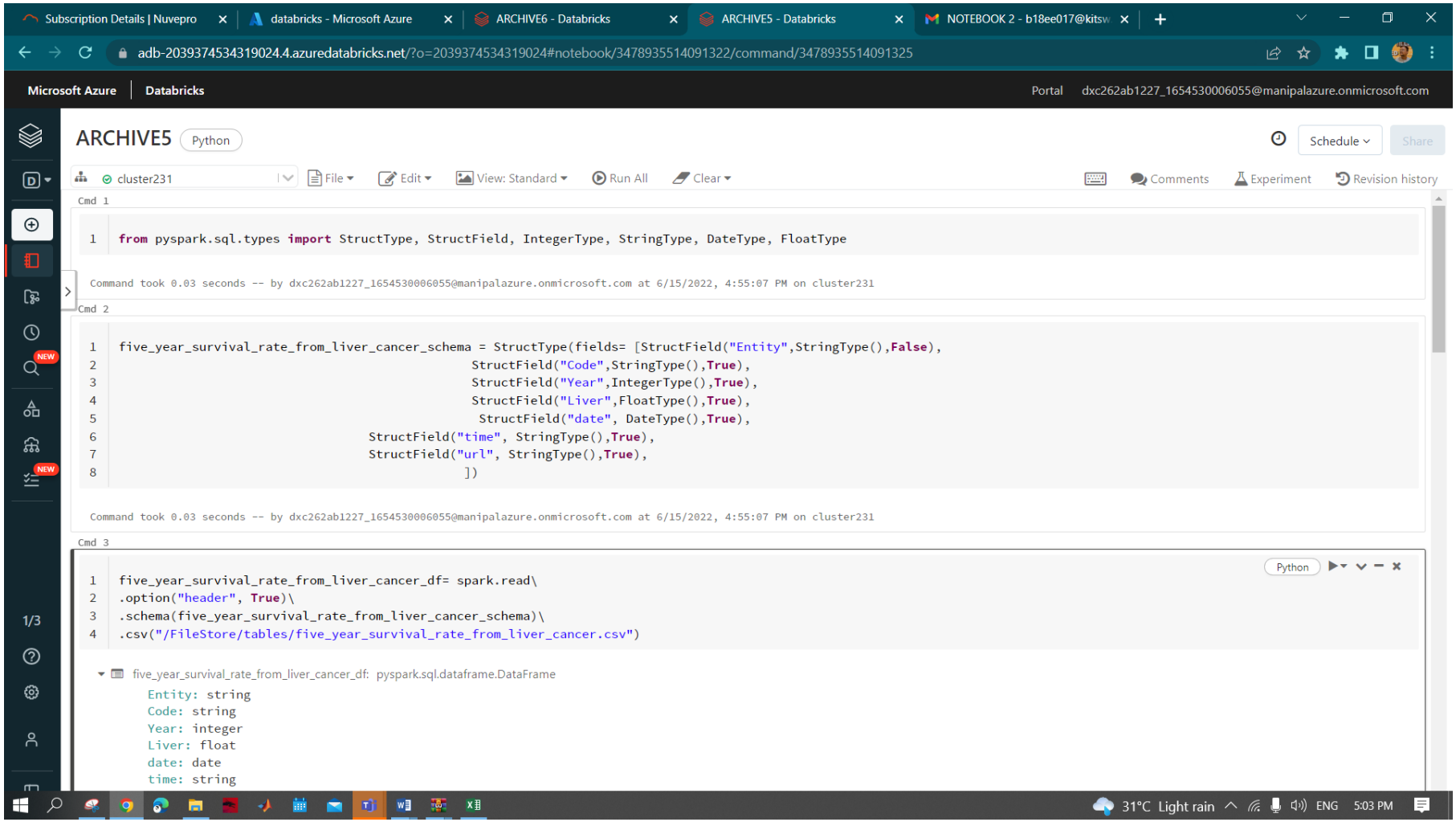

Subscription Details | Nuve
databricks - Microsoft Azur
ARCHIVE6 - Databricks
ARCHIVE5 - Databricks
ARCHIVE4 - Databricks
NOTEBOOK 2 - b18ee017@
adb-2039374534319024.4.azuredatabricks.net/?o=2039374534319024#notebook/3478935514091332/command/3478935514091339
Microsoft Azure | Databricks
Portal
dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com
ARCHIVE4
Python
Schedule
Share
cluster231
File
Edit
View: Standard
Run All
Clear
Comments
Experiment
Revision history
Command took 0.06 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 5:25:46 PM on cluster231
Cmd 6
1 SEntFiN_v_selected_df=SEntFiN_v_with_timestamp_df.select(col('S No'), col('Title').alias('title'), col('Decisions'), col('Words').alias('words'), col('ingestion_date'), col('SEntFiN_v_timestamp'))
SEntFiN_v_selected_df: pyspark.sql.dataframe.DataFrame
S No: integer
title: string
Decisions: string
words: integer
ingestion_date: timestamp
SEntFiN_v_timestamp: timestamp
Command took 0.06 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 5:25:46 PM on cluster231
Cmd 7
1 SEntFiN_v_selected_df.write.mode('overwrite').partitionBy('Words').parquet('/mnt/formulaIdl/processed/SEntFiN_v')
Python
(1) Spark Jobs
Job 6 View (Stages: 1/1)
Command took 6.85 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 5:28:03 PM on cluster231
Cmd 8
1 display(SEntFiN_v_selected_df)
(1) Spark Jobs
Table
Data Profile
1/3
32°C Light rain
ENG
5:29 PM

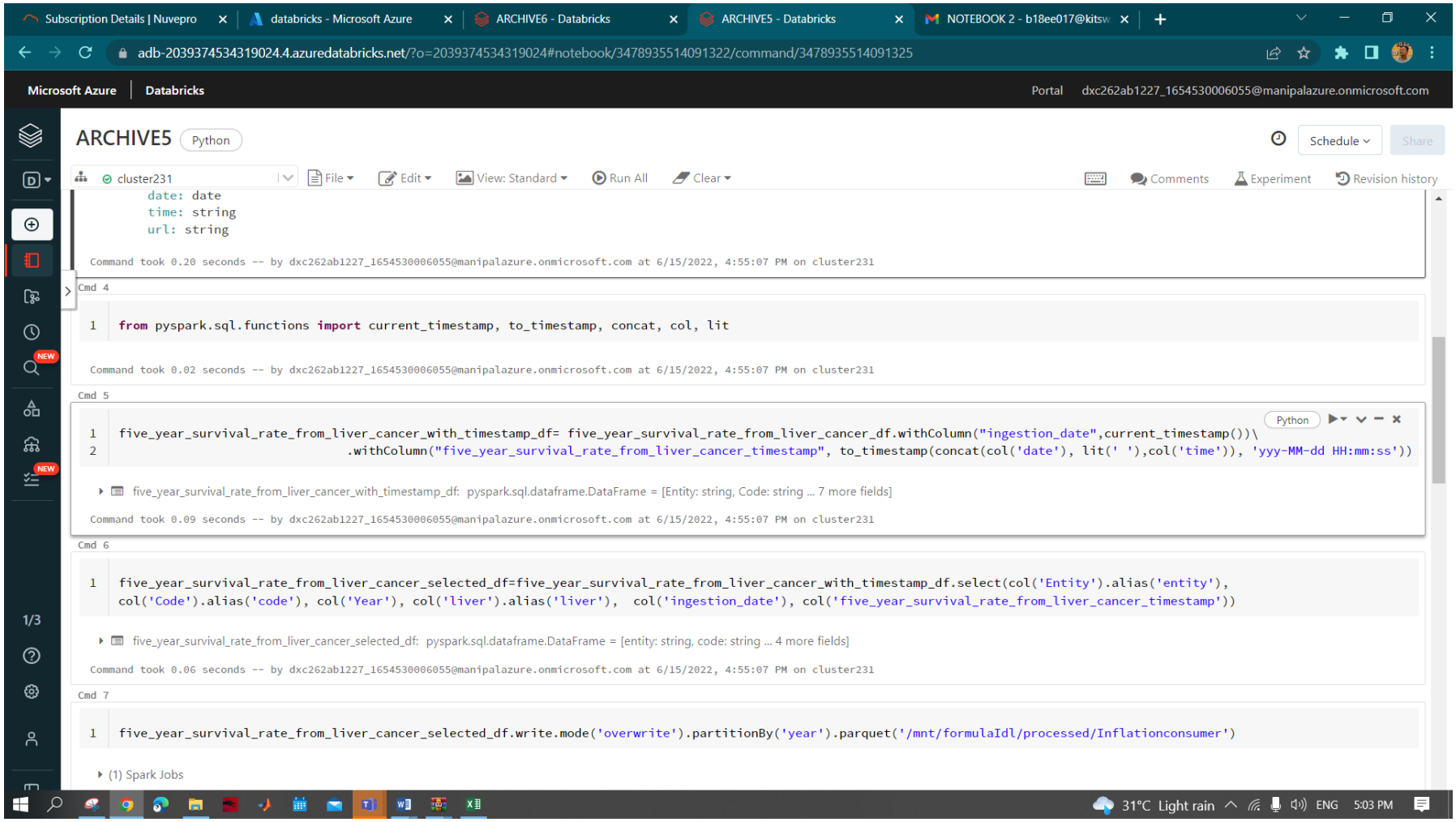

Subscription Details | Nuve
databricks - Microsoft Azur
ARCHIVE6 - Databricks
ARCHIVE5 - Databricks
ARCHIVE4 - Databricks
NOTEBOOK 2 - b18ee017@
adb-2039374534319024.4.azuredatabricks.net/?o=2039374534319024#notebook/3478935514091332/command/3478935514091339
Microsoft Azure
Databricks
Portal
dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com
ARCHIVE4
Python
Schedule
Share
cluster231
File
Edit
View: Standard
Run All
Clear
Comments
Experiment
Revision history
(1) Spark Jobs
Job 6
View (Stages: 1/1)
Command took 6.85 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 5:28:03 PM on cluster231
Cmd 8
1 display(SEntFiN_v_selected_df)
Python
(1) Spark Jobs
Table
Data Profile
S No
title
Decisions
words
ingestion_date
1 1 SpiceJet to issue 6.4 crore warrants to promoters "{""SpiceJet"": ""neutral""}" 8 2022-06-15T11:5
2 2 MMTC Q2 net loss at Rs 10.4 crore "{""MMTC"": ""neutral""}" 8 2022-06-15T11:5
3 3 Mid-cap funds can deliver more, stay put: Experts "{""Mid-cap funds"": ""positive""}" 8 2022-06-15T11:5
4 4 Mid caps now turn into market darlings "{""Mid caps"": ""positive""}" 7 2022-06-15T11:5
5 5 Market seeing patience, if not conviction: Prakash Diwan "{""Market"": ""neutral""}" 8 2022-06-15T11:5
6 6 Infosys: Will the strong volume growth sustain? "{""Infosys"": ""neutral""}" 7 2022-06-15T11:5
7 7 Hudco raises Rs 279 cr via tax-free bonds "{""Hudco"": ""positive""}" 8
Truncated results, showing first 1000 rows.
Click to re-execute with maximum result limits.
Command took 0.33 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 5:28:16 PM on cluster231
Cmd 9
1 SEntFiN_v_selected_df.show()
1/3
32°C Light rain
ENG
5:29 PM

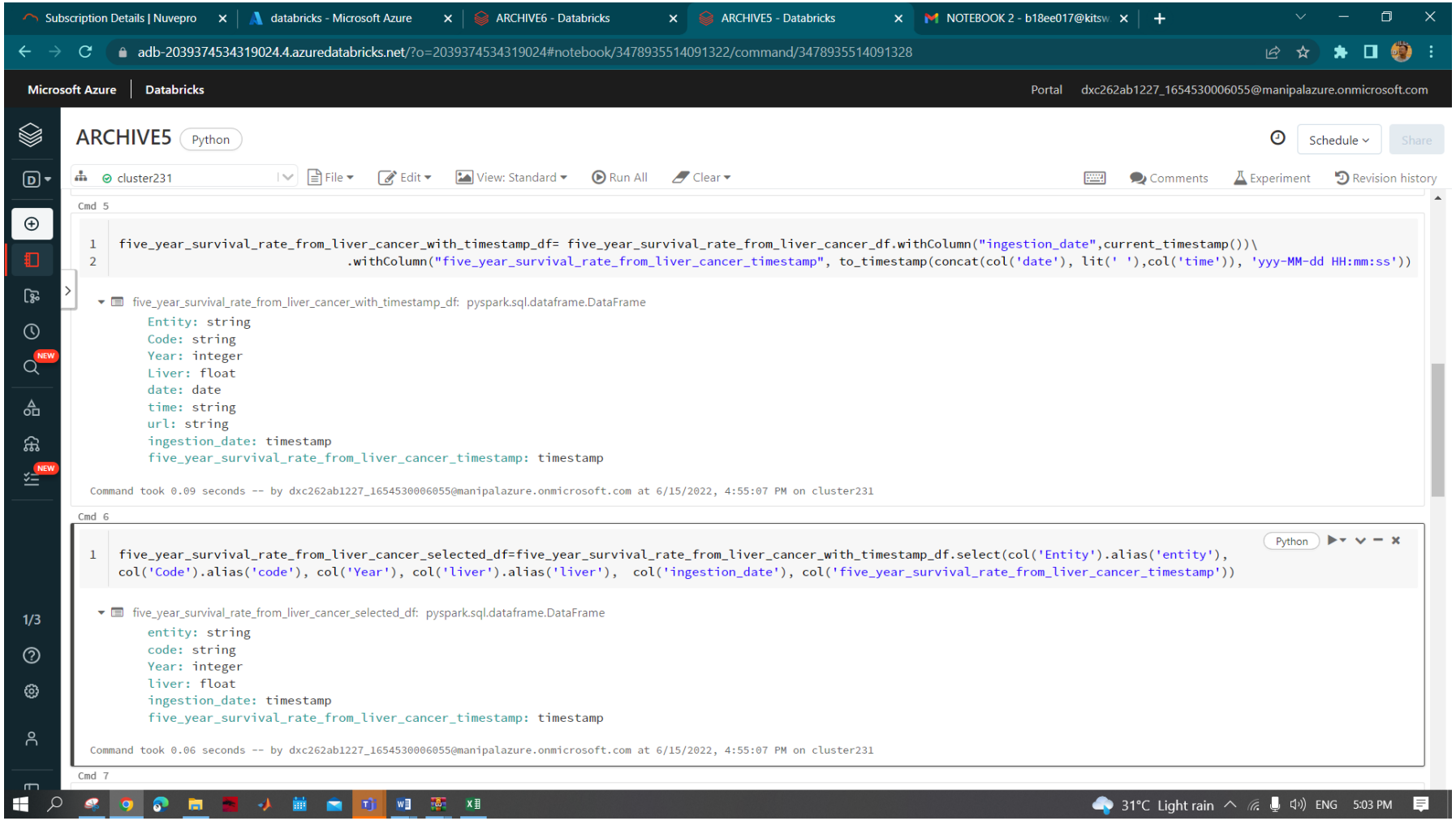

Subscription Details | Nuve
databricks - Microsoft Azur
ARCHIVE6 - Databricks
ARCHIVE5 - Databricks
ARCHIVE4 - Databricks
NOTEBOOK 2 - b18ee017
adb-2039374534319024.4.azuredatabricks.net/?o=2039374534319024#notebook/3478935514091332/command/3478935514091339
Microsoft Azure | Databricks
Portal
dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com
ARCHIVE4
Python
Schedule
Share
cluster231
File
Edit
View: Standard
Run All
Clear
Comments
Experiment
Revision history
1 SEntFiN_v_selected_df.show()
(1) Spark Jobs
+----+--------------------+--------------------+-----+--------------------+--------------------+
|S No| title| Decisions|words| ingestion_date|SEntFiN_v_timestamp|
+----+--------------------+--------------------+-----+--------------------+--------------------+
1	SpiceJet to issue...	"{""SpiceJet"": "...	8	2022-06-15 11:58:...	null
2	MMTC Q2 net loss ...	"{""MMTC"": ""neu...	8	2022-06-15 11:58:...	null
3	Mid-cap funds can...	"{""Mid-cap funds...	8	2022-06-15 11:58:...	null
4	Mid caps now turn...	"{""Mid caps"": "...	7	2022-06-15 11:58:...	null
5	Market seeing pat...	"{""Market"": ""n...	8	2022-06-15 11:58:...	null
6	Infosys: Will the...	"{""Infosys"": ""...	7	2022-06-15 11:58:...	null
7	Hudco raises Rs 2...	"{""Hudco"": ""po...	8	2022-06-15 11:58:...	null
8	HOEC could retest...	"{""HOEC"": ""neu...	7	2022-06-15 11:58:...	null
9	Gold shines on se...	"{""Gold"": ""pos...	null	2022-06-15 11:58:...	null
10	Genpact appoints ...	"{""Genpact"": ""...	7	2022-06-15 11:58:...	null
11	EXL beats profit ...	"{""EXL"": ""posi...	7	2022-06-15 11:58:...	null
12	Wait and watch on...	"{""Bharti Airtel...	8	2022-06-15 11:58:...	null
13	Would stick to ba...	"{""banking"": ""...	null	2022-06-15 11:58:...	null
14	MSCI adds Aurobin...	"{""Aurobindo Pha...	null	2022-06-15 11:58:...	null
15	Ashok Leyland rai...	"{""Ashok Leyland...	8	2022-06-15 11:58:...	null
16	At Wipro, growth ...	"{""Wipro"": ""ne...	6	2022-06-15 11:58:...	null
17	Why Chinese stock...	"{""US"": ""negat...	null	2022-06-15 11:58:...	null
18	US stocks finish ...	"{""tech"": ""neg...	null	2022-06-15 11:58:...	null
Command took 0.31 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 5:28:28 PM on cluster231
Shift+Enter to run
1/3
32°C Light rain
ENG
5:30 PM

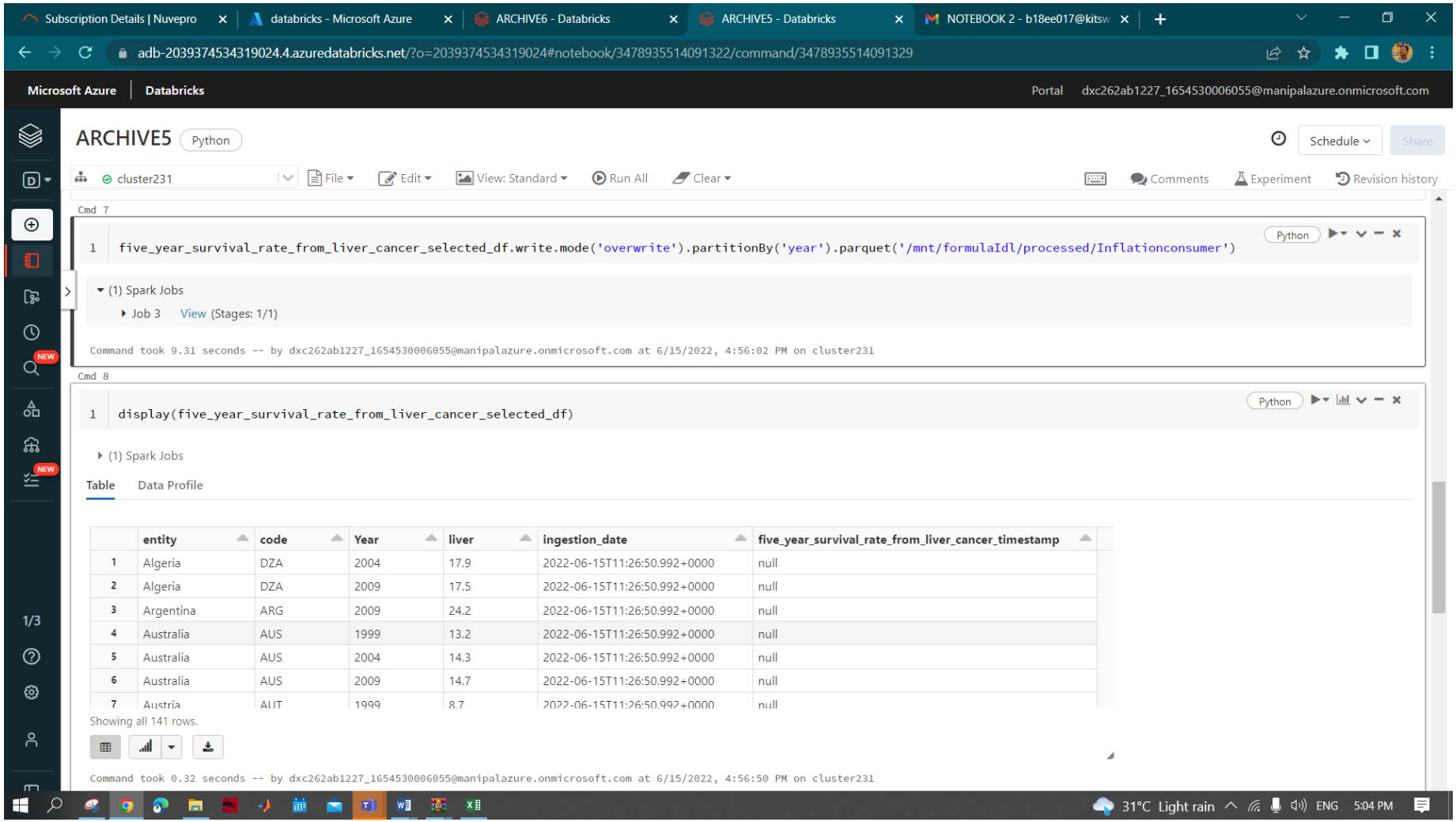
5.Using archive5.zip file - please ingest data into databricks DBFS path & query the data
redesign columns accordingly using dafarame commands - display with notebooks accordingly

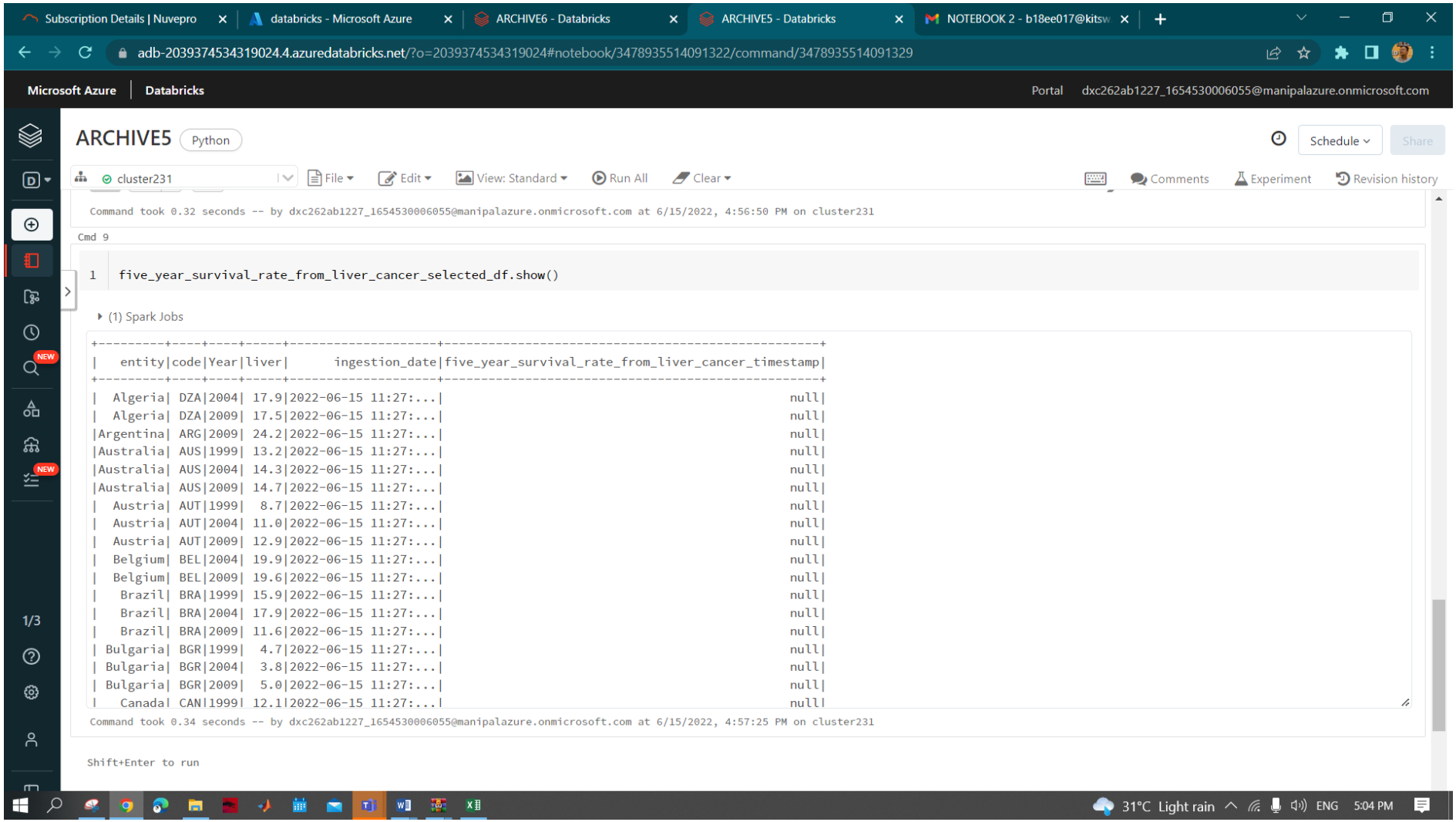
A:




Subscription Details | Nuvepro
databricks - Microsoft Azure
ARCHIVE6 - Databricks
ARCHIVE5 - Databricks
NOTEBOOK 2 - b18ee017@kitsw
adb-2039374534319024.4.azuredatabricks.net/?o=2039374534319024#notebook/3478935514091322/command/3478935514091325
Microsoft Azure | Databricks
Portal
dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com
ARCHIVE5
Python
Schedule
Share
cluster231
File
Edit
View: Standard
Run All
Clear
Comments
Experiment
Revision history
date: date
time: string
url: string
Command took 0.20 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 4:55:07 PM on cluster231
Cmd 4
from pyspark.sql.functions import current_timestamp, to_timestamp, concat, col, lit
Command took 0.02 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 4:55:07 PM on cluster231
Cmd 5
Python
five_year_survival_rate_from_liver_cancer_with_timestamp_df= five_year_survival_rate_from_liver_cancer_df.withColumn("ingestion_date",current_timestamp())\
.withColumn("five_year_survival_rate_from_liver_cancer_timestamp", to_timestamp(concat(col('date'), lit(' '),col('time')), 'yyy-MM-dd HH:mm:ss'))
five_year_survival_rate_from_liver_cancer_with_timestamp_df: pyspark.sql.dataframe.DataFrame = [Entity: string, Code: string ... 7 more fields]
Command took 0.09 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 4:55:07 PM on cluster231
Cmd 6
five_year_survival_rate_from_liver_cancer_selected_df=five_year_survival_rate_from_liver_cancer_with_timestamp_df.select(col('Entity').alias('entity'),
col('Code').alias('code'), col('Year'), col('liver').alias('liver'), col('ingestion_date'), col('five_year_survival_rate_from_liver_cancer_timestamp'))
five_year_survival_rate_from_liver_cancer_selected_df: pyspark.sql.dataframe.DataFrame = [entity: string, code: string ... 4 more fields]
Command took 0.06 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 4:55:07 PM on cluster231
Cmd 7
five_year_survival_rate_from_liver_cancer_selected_df.write.mode('overwrite').partitionBy('year').parquet('/mnt/formulaIdl/processed/Inflationconsumer')
(1) Spark Jobs
1/3
31°C Light rain
ENG
5:03 PM

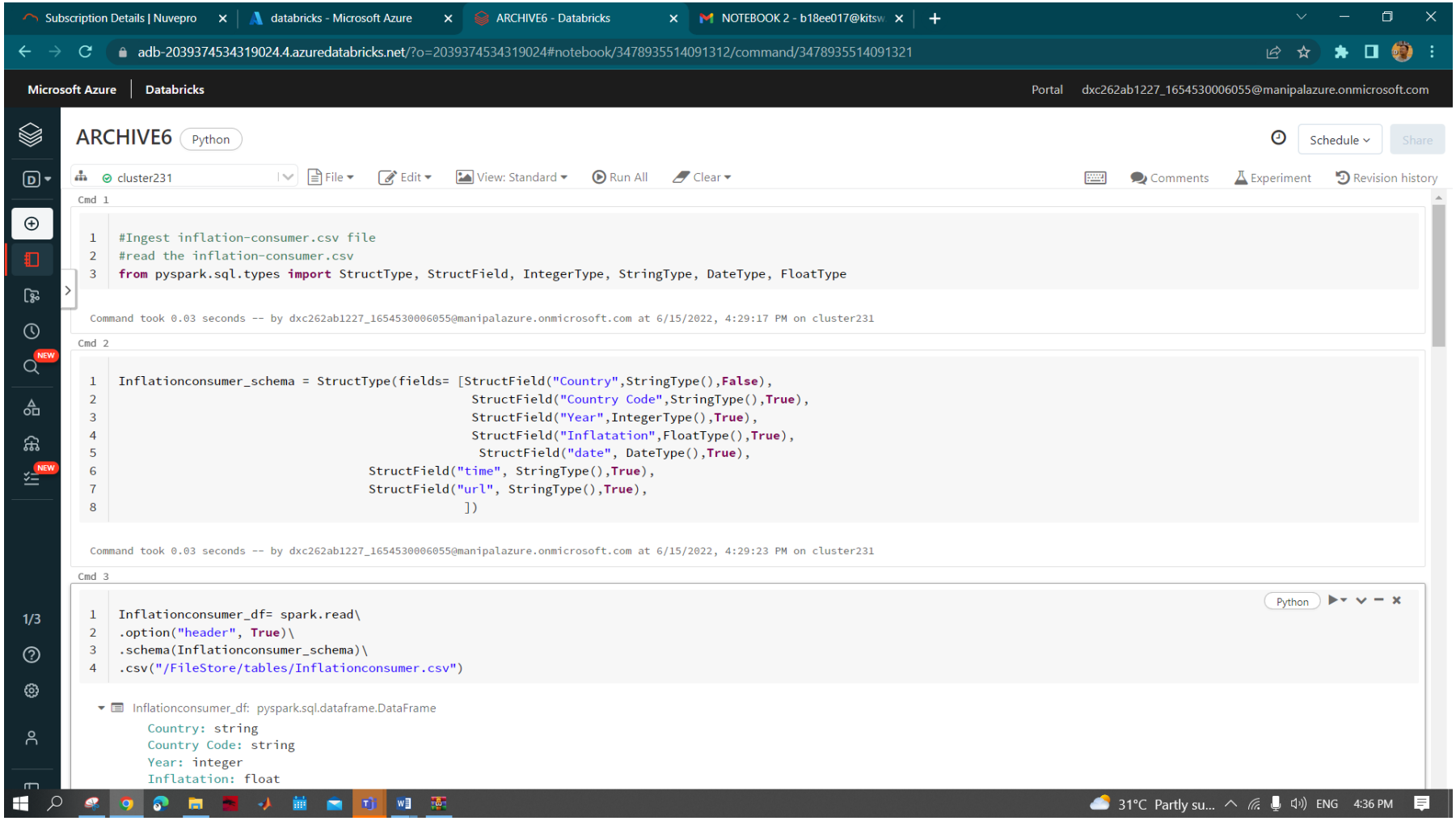

Subscription Details | Nuvepro
databricks - Microsoft Azure
ARCHIVE6 - Databricks
ARCHIVE5 - Databricks
NOTEBOOK 2 - b18ee017@kitsw
adb-2039374534319024.4.azuredatabricks.net/?o=2039374534319024#notebook/3478935514091322/command/3478935514091328
Microsoft Azure
Databricks
Portal
dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com
ARCHIVE5
Python
Schedule
Share
cluster231
File
Edit
View: Standard
Run All
Clear
Comments
Experiment
Revision history
Cmd 5
five_year_survival_rate_from_liver_cancer_with_timestamp_df= five_year_survival_rate_from_liver_cancer_df.withColumn("ingestion_date",current_timestamp())\
.withColumn("five_year_survival_rate_from_liver_cancer_timestamp", to_timestamp(concat(col('date'), lit(' '),col('time')), 'yyy-MM-dd HH:mm:ss'))
five_year_survival_rate_from_liver_cancer_with_timestamp_df: pyspark.sql.dataframe.DataFrame
Entity: string
Code: string
Year: integer
Liver: float
date: date
time: string
url: string
ingestion_date: timestamp
five_year_survival_rate_from_liver_cancer_timestamp: timestamp
Command took 0.09 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 4:55:07 PM on cluster231
Cmd 6
Python
five_year_survival_rate_from_liver_cancer_selected_df=five_year_survival_rate_from_liver_cancer_with_timestamp_df.select(col('Entity').alias('entity'),
col('Code').alias('code'), col('Year'), col('liver').alias('liver'), col('ingestion_date'), col('five_year_survival_rate_from_liver_cancer_timestamp'))
five_year_survival_rate_from_liver_cancer_selected_df: pyspark.sql.dataframe.DataFrame
entity: string
code: string
Year: integer
liver: float
ingestion_date: timestamp
five_year_survival_rate_from_liver_cancer_timestamp: timestamp
Command took 0.06 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 4:55:07 PM on cluster231
Cmd 7
1/3
31°C Light rain
ENG
5:03 PM

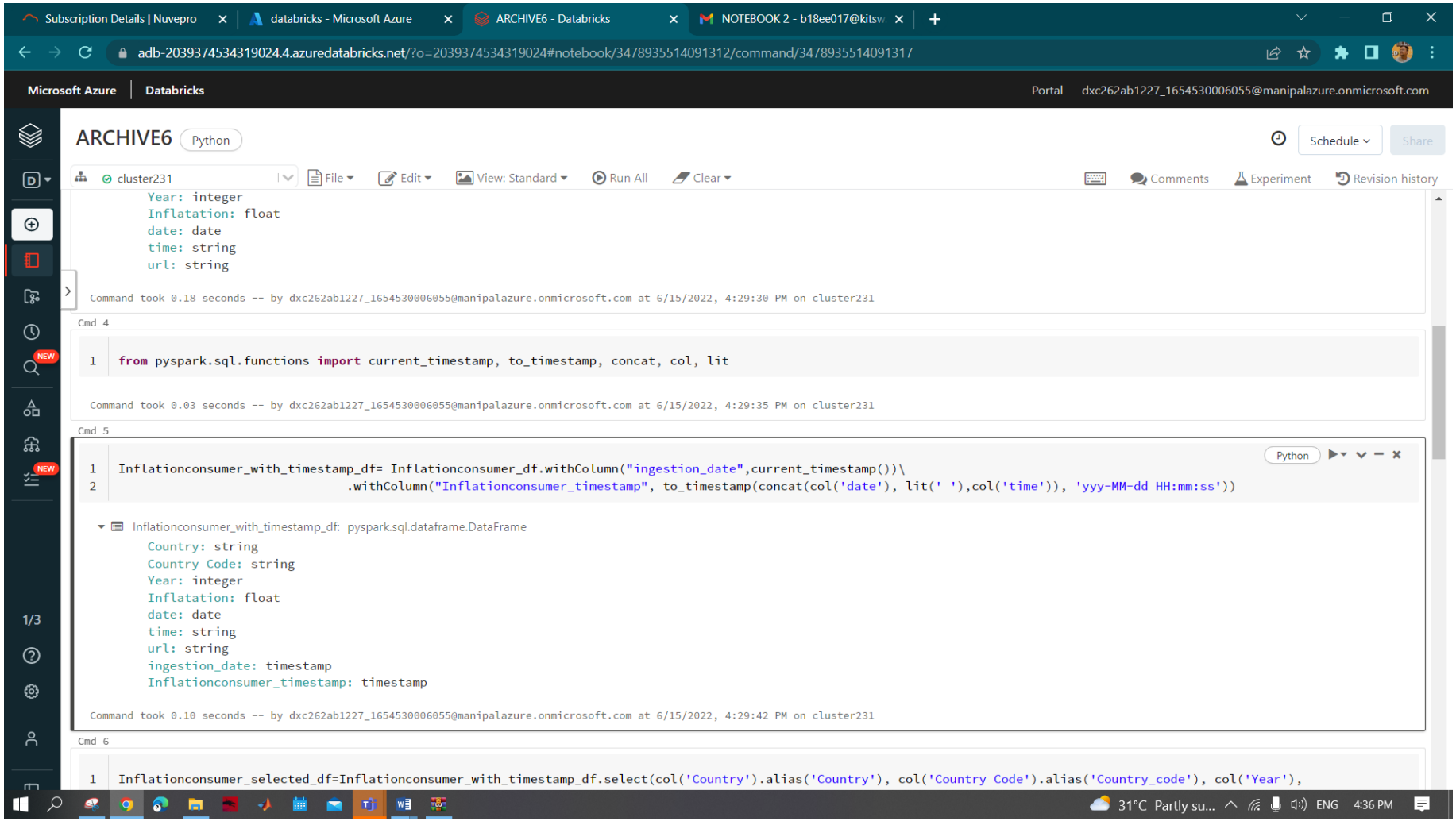

Subscription Details | Nuvepro
databricks - Microsoft Azure
ARCHIVE6 - Databricks
ARCHIVE5 - Databricks
NOTEBOOK 2 - b18ee017@kitsw
adb-2039374534319024.4.azuredatabricks.net/?o=2039374534319024#notebook/3478935514091322/command/3478935514091329
Microsoft Azure | Databricks
Portal dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com
ARCHIVE5 Python
Schedule
Share
cluster231
File
Edit
View: Standard
Run All
Clear
Comments
Experiment
Revision history
Cmd 7
1 five_year_survival_rate_from_liver_cancer_selected_df.write.mode('overwrite').partitionBy('year').parquet('/mnt/formulaIdl/processed/Inflationconsumer')
Python
(1) Spark Jobs
Job 3 View (Stages: 1/1)
Command took 9.31 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 4:56:02 PM on cluster231
Cmd 8
1 display(five_year_survival_rate_from_liver_cancer_selected_df)
Python
(1) Spark Jobs
Table
Data Profile
entity code Year liver ingestion_date five_year_survival_rate_from_liver_cancer_timestamp
1 Algeria DZA 2004 17.9 2022-06-15T11:26:50.992+0000 null
2 Algeria DZA 2009 17.5 2022-06-15T11:26:50.992+0000 null
3 Argentina ARG 2009 24.2 2022-06-15T11:26:50.992+0000 null
4 Australia AUS 1999 13.2 2022-06-15T11:26:50.992+0000 null
5 Australia AUS 2004 14.3 2022-06-15T11:26:50.992+0000 null
6 Australia AUS 2009 14.7 2022-06-15T11:26:50.992+0000 null
7 Austria AUT 1999 8.7 2022-06-15T11:26:50.992+0000 null
Showing all 141 rows.
Command took 0.32 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 4:56:50 PM on cluster231
1/3
31°C Light rain
ENG
5:04 PM

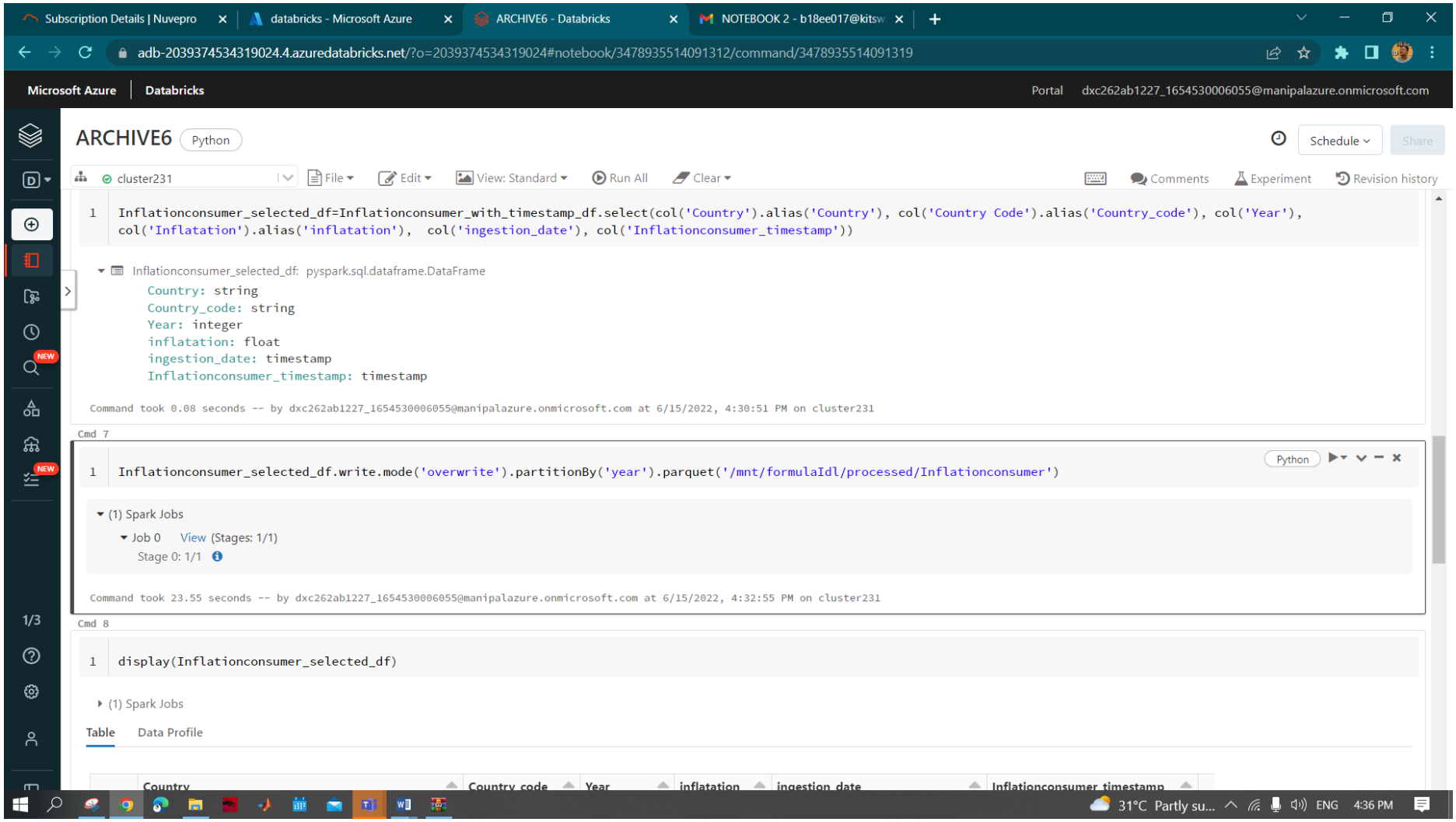

Subscription Details | Nuvepro
databricks - Microsoft Azure
ARCHIVE6 - Databricks
ARCHIVE5 - Databricks
NOTEBOOK 2 - b18ee017@kitsw
adb-2039374534319024.4.azuredatabricks.net/?o=2039374534319024#notebook/3478935514091322/command/3478935514091329
Microsoft Azure | Databricks
Portal
dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com
ARCHIVE5
Python
Schedule
Share
cluster231
File
Edit
View: Standard
Run All
Clear
Comments
Experiment
Revision history
Command took 0.32 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 4:56:50 PM on cluster231
Cmd 9
1 five_year_survival_rate_from_liver_cancer_selected_df.show()
(1) Spark Jobs
+---------+----+----+-----+-------------------+--+
| entity|code|Year|liver| ingestion_date|five_year_survival_rate_from_liver_cancer_timestamp|
+---------+----+----+-----+-------------------+--+
| Algeria| DZA|2004| 17.9|2022-06-15 11:27:...| null|
| Algeria| DZA|2009| 17.5|2022-06-15 11:27:...| null|
|Argentina| ARG|2009| 24.2|2022-06-15 11:27:...| null|
|Australia| AUS|1999| 13.2|2022-06-15 11:27:...| null|
|Australia| AUS|2004| 14.3|2022-06-15 11:27:...| null|
|Australia| AUS|2009| 14.7|2022-06-15 11:27:...| null|
Austria	AUT	1999	8.7	2022-06-15 11:27:...	null
Austria	AUT	2004	11.0	2022-06-15 11:27:...	null
Austria	AUT	2009	12.9	2022-06-15 11:27:...	null
Belgium	BEL	2004	19.9	2022-06-15 11:27:...	null
Belgium	BEL	2009	19.6	2022-06-15 11:27:...	null
Brazil	BRA	1999	15.9	2022-06-15 11:27:...	null
Brazil	BRA	2004	17.9	2022-06-15 11:27:...	null
Brazil	BRA	2009	11.6	2022-06-15 11:27:...	null
Bulgaria	BGR	1999	4.7	2022-06-15 11:27:...	null
Bulgaria	BGR	2004	3.8	2022-06-15 11:27:...	null
Bulgaria	BGR	2009	5.0	2022-06-15 11:27:...	null
Canada	CAN	1999	12.1	2022-06-15 11:27:...	null
Command took 0.34 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 4:57:25 PM on cluster231
Shift+Enter to run
1/3
31°C Light rain
ENG
5:04 PM

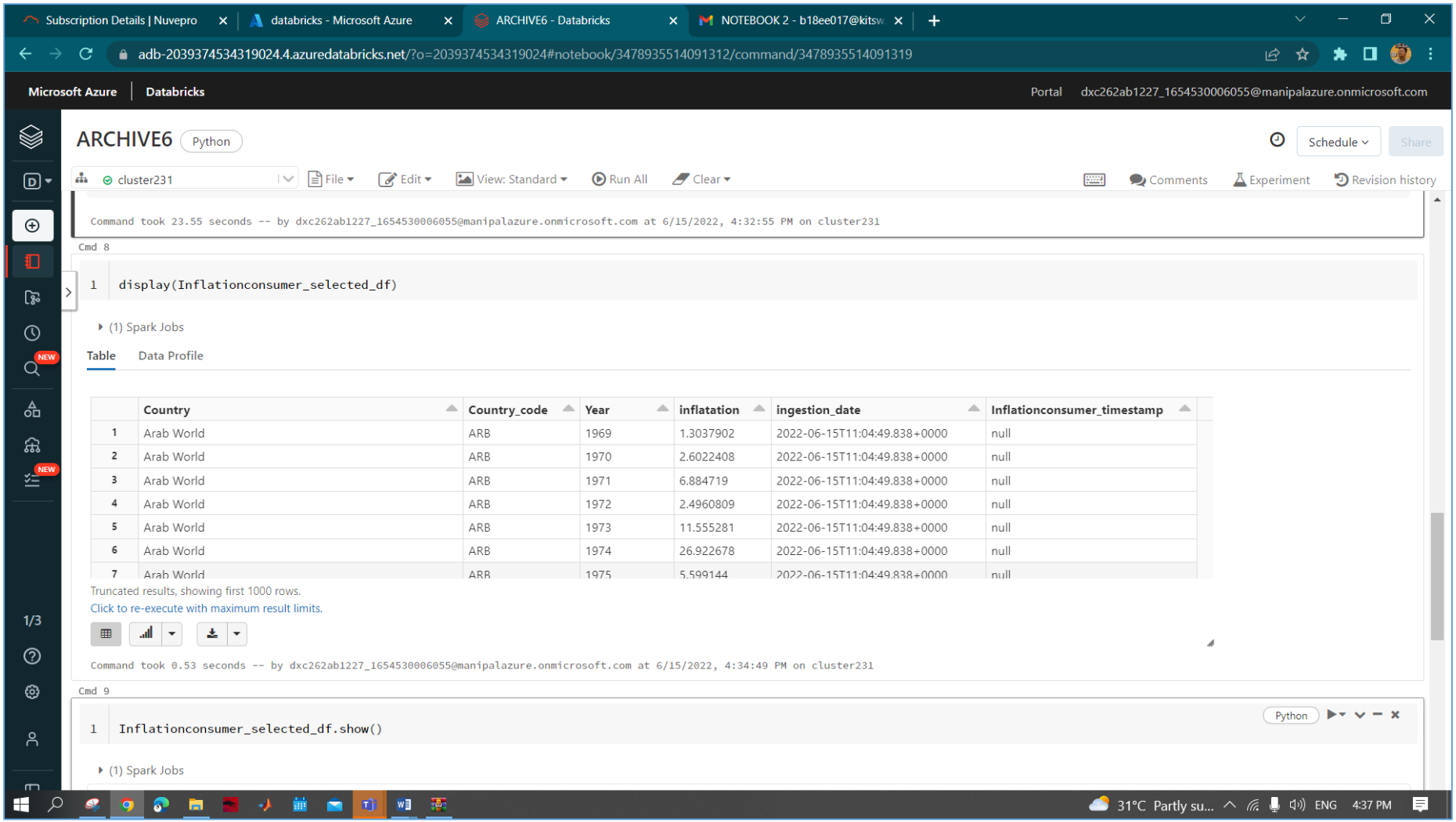
6.Using archive6.zip file - please ingest data into databricks DBFS path & query the data
redesign columns accordingly using dafarame commands - display with notebooks accordingly

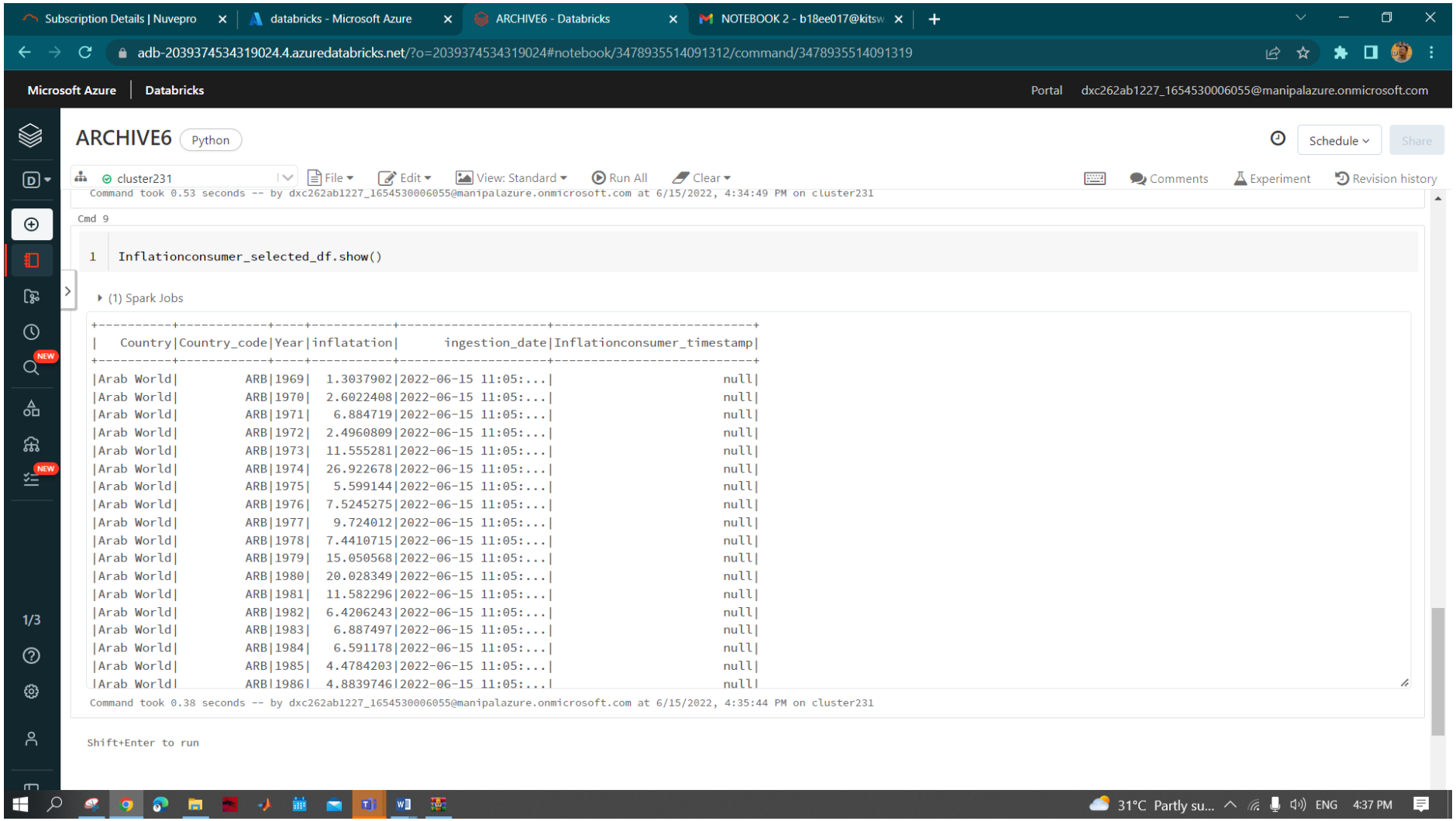
A:


Subscription Details | Nuvepro
databricks - Microsoft Azure
ARCHIVE6 - Databricks
NOTEBOOK 2 - b18ee017@kitsw
adb-2039374534319024.4.azuredatabricks.net/?o=2039374534319024#notebook/3478935514091312/command/3478935514091321
Microsoft Azure | Databricks
Portal
dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com
ARCHIVE6
Python
Schedule
Share
cluster231
File
Edit
View: Standard
Run All
Clear
Comments
Experiment
Revision history
Cmd 1
#Ingest inflation-consumer.csv file
#read the inflation-consumer.csv
from pyspark.sql.types import StructType, StructField, IntegerType, StringType, DateType, FloatType
Command took 0.03 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 4:29:17 PM on cluster231
Cmd 2
Inflationconsumer_schema = StructType(fields= [StructField("Country",StringType(),False),
StructField("Country Code",StringType(),True),
StructField("Year",IntegerType(),True),
StructField("Inflatation",FloatType(),True),
StructField("date", DateType(),True),
StructField("time", StringType(),True),
StructField("url", StringType(),True),
])
Command took 0.03 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 4:29:23 PM on cluster231
Cmd 3
Python
Inflationconsumer_df= spark.read\
.option("header", True)\
.schema(Inflationconsumer_schema)\
.csv("/FileStore/tables/Inflationconsumer.csv")
Inflationconsumer_df: pyspark.sql.dataframe.DataFrame
Country: string
Country Code: string
Year: integer
Inflatation: float
1/3
31°C Partly su...
ENG
4:36 PM


Subscription Details | Nuvepro
databricks - Microsoft Azure
ARCHIVE6 - Databricks
NOTEBOOK 2 - b18ee017@kitsw
adb-2039374534319024.4.azuredatabricks.net/?o=2039374534319024#notebook/3478935514091312/command/3478935514091317
Microsoft Azure | Databricks
Portal
dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com
ARCHIVE6
Python
Schedule
Share
cluster231
File
Edit
View: Standard
Run All
Clear
Comments
Experiment
Revision history
Year: integer
Inflatation: float
date: date
time: string
url: string
Command took 0.18 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 4:29:30 PM on cluster231
Cmd 4
from pyspark.sql.functions import current_timestamp, to_timestamp, concat, col, lit
Command took 0.03 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 4:29:35 PM on cluster231
Cmd 5
Python
Inflationconsumer_with_timestamp_df= Inflationconsumer_df.withColumn("ingestion_date",current_timestamp())\
.withColumn("Inflationconsumer_timestamp", to_timestamp(concat(col('date'), lit(' '),col('time')), 'yyy-MM-dd HH:mm:ss'))
Inflationconsumer_with_timestamp_df: pyspark.sql.dataframe.DataFrame
Country: string
Country Code: string
Year: integer
Inflatation: float
date: date
time: string
url: string
ingestion_date: timestamp
Inflationconsumer_timestamp: timestamp
Command took 0.10 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 4:29:42 PM on cluster231
Cmd 6
Inflationconsumer_selected_df=Inflationconsumer_with_timestamp_df.select(col('Country').alias('Country'), col('Country Code').alias('Country_code'), col('Year'),
1/3
31°C Partly su...
ENG
4:36 PM


Subscription Details | Nuvepro
databricks - Microsoft Azure
ARCHIVE6 - Databricks
NOTEBOOK 2 - b18ee017@kitsw
adb-2039374534319024.4.azuredatabricks.net/?o=2039374534319024#notebook/3478935514091312/command/3478935514091319
Microsoft Azure
Databricks
Portal
dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com
ARCHIVE6
Python
Schedule
Share
cluster231
File
Edit
View: Standard
Run All
Clear
Comments
Experiment
Revision history
1 Inflationconsumer_selected_df=Inflationconsumer_with_timestamp_df.select(col('Country').alias('Country'), col('Country Code').alias('Country_code'), col('Year'), col('Inflatation').alias('inflatation'), col('ingestion_date'), col('Inflationconsumer_timestamp'))
Inflationconsumer_selected_df: pyspark.sql.dataframe.DataFrame
Country: string
Country_code: string
Year: integer
inflatation: float
ingestion_date: timestamp
Inflationconsumer_timestamp: timestamp
Command took 0.08 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 4:30:51 PM on cluster231
Cmd 7
1 Inflationconsumer_selected_df.write.mode('overwrite').partitionBy('year').parquet('/mnt/formulaIdl/processed/Inflationconsumer')
Python
(1) Spark Jobs
Job 0 View (Stages: 1/1)
Stage 0: 1/1
Command took 23.55 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 4:32:55 PM on cluster231
Cmd 8
1 display(Inflationconsumer_selected_df)
(1) Spark Jobs
Table
Data Profile
Country
Country_code
Year
inflatation
ingestion_date
Inflationconsumer_timestamp
1/3
31°C Partly su...
ENG
4:36 PM


Subscription Details | Nuvepro
databricks - Microsoft Azure
ARCHIVE6 - Databricks
NOTEBOOK 2 - b18ee017@kitsw
adb-2039374534319024.4.azuredatabricks.net/?o=2039374534319024#notebook/3478935514091312/command/3478935514091319
Microsoft Azure | Databricks
Portal
dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com
ARCHIVE6
Python
Schedule
Share
cluster231
File
Edit
View: Standard
Run All
Clear
Comments
Experiment
Revision history
Command took 23.55 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 4:32:55 PM on cluster231
Cmd 8
1 display(Inflationconsumer_selected_df)
(1) Spark Jobs
Table
Data Profile
Country
Country_code
Year
inflatation
ingestion_date
Inflationconsumer_timestamp
1 Arab World ARB 1969 1.3037902 2022-06-15T11:04:49.838+0000 null
2 Arab World ARB 1970 2.6022408 2022-06-15T11:04:49.838+0000 null
3 Arab World ARB 1971 6.884719 2022-06-15T11:04:49.838+0000 null
4 Arab World ARB 1972 2.4960809 2022-06-15T11:04:49.838+0000 null
5 Arab World ARB 1973 11.555281 2022-06-15T11:04:49.838+0000 null
6 Arab World ARB 1974 26.922678 2022-06-15T11:04:49.838+0000 null
7 Arab World ARB 1975 5.599144 2022-06-15T11:04:49.838+0000 null
Truncated results, showing first 1000 rows.
Click to re-execute with maximum result limits.
Command took 0.53 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 4:34:49 PM on cluster231
Cmd 9
Python
1 Inflationconsumer_selected_df.show()
(1) Spark Jobs
1/3
31°C Partly su...
ENG
4:37 PM


Subscription Details | Nuvepro
databricks - Microsoft Azure
ARCHIVE6 - Databricks
NOTEBOOK 2 - b18ee017@kitsw
adb-2039374534319024.4.azuredatabricks.net/?o=2039374534319024#notebook/3478935514091312/command/3478935514091319
Microsoft Azure | Databricks
Portal
dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com
ARCHIVE6
Python
Schedule
Share
cluster231
File
Edit
View: Standard
Run All
Clear
Comments
Experiment
Revision history
Command took 0.53 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 4:34:49 PM on cluster231
Cmd 9
1 Inflationconsumer_selected_df.show()
(1) Spark Jobs
+----------+------------+----+-----------+-------------------+--------------------------+
| Country|Country_code|Year|inflatation| ingestion_date|Inflationconsumer_timestamp|
+----------+------------+----+-----------+-------------------+--------------------------+
|Arab World| ARB|1969| 1.3037902|2022-06-15 11:05:...| null|
|Arab World| ARB|1970| 2.6022408|2022-06-15 11:05:...| null|
|Arab World| ARB|1971| 6.884719|2022-06-15 11:05:...| null|
|Arab World| ARB|1972| 2.4960809|2022-06-15 11:05:...| null|
|Arab World| ARB|1973| 11.555281|2022-06-15 11:05:...| null|
|Arab World| ARB|1974| 26.922678|2022-06-15 11:05:...| null|
|Arab World| ARB|1975| 5.599144|2022-06-15 11:05:...| null|
|Arab World| ARB|1976| 7.5245275|2022-06-15 11:05:...| null|
|Arab World| ARB|1977| 9.724012|2022-06-15 11:05:...| null|
|Arab World| ARB|1978| 7.4410715|2022-06-15 11:05:...| null|
|Arab World| ARB|1979| 15.050568|2022-06-15 11:05:...| null|
|Arab World| ARB|1980| 20.028349|2022-06-15 11:05:...| null|
|Arab World| ARB|1981| 11.582296|2022-06-15 11:05:...| null|
|Arab World| ARB|1982| 6.4206243|2022-06-15 11:05:...| null|
|Arab World| ARB|1983| 6.887497|2022-06-15 11:05:...| null|
|Arab World| ARB|1984| 6.591178|2022-06-15 11:05:...| null|
|Arab World| ARB|1985| 4.4784203|2022-06-15 11:05:...| null|
|Arab World| ARB|1986| 4.8839746|2022-06-15 11:05:...| null|
Command took 0.38 seconds -- by dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com at 6/15/2022, 4:35:44 PM on cluster231
Shift+Enter to run
1/3
31°C Partly su...
ENG
4:37 PM

RESULT

Almost all the test cases have been solved and presented successfully in the present document except 1 due to lack of data.

CONCLUSIONS

All the case studies have been solved successfully with all the concepts that have been covered in the training session. It's really a great experience of learning while solving the cases. This case study gave me immense confidence regarding my ability to upskill in new technologies.