This submission template is a convenient document for you to provide the screenshots and explanations for Assignment 5.0. This submission template is intended to be used in conjunction with the Assignment 5.0 Instructions document. The instructions document illustrates how to correctly execute each SQL construct, explains important theoretical and practical details, and contains the complete set of instructions on how to complete this lab.

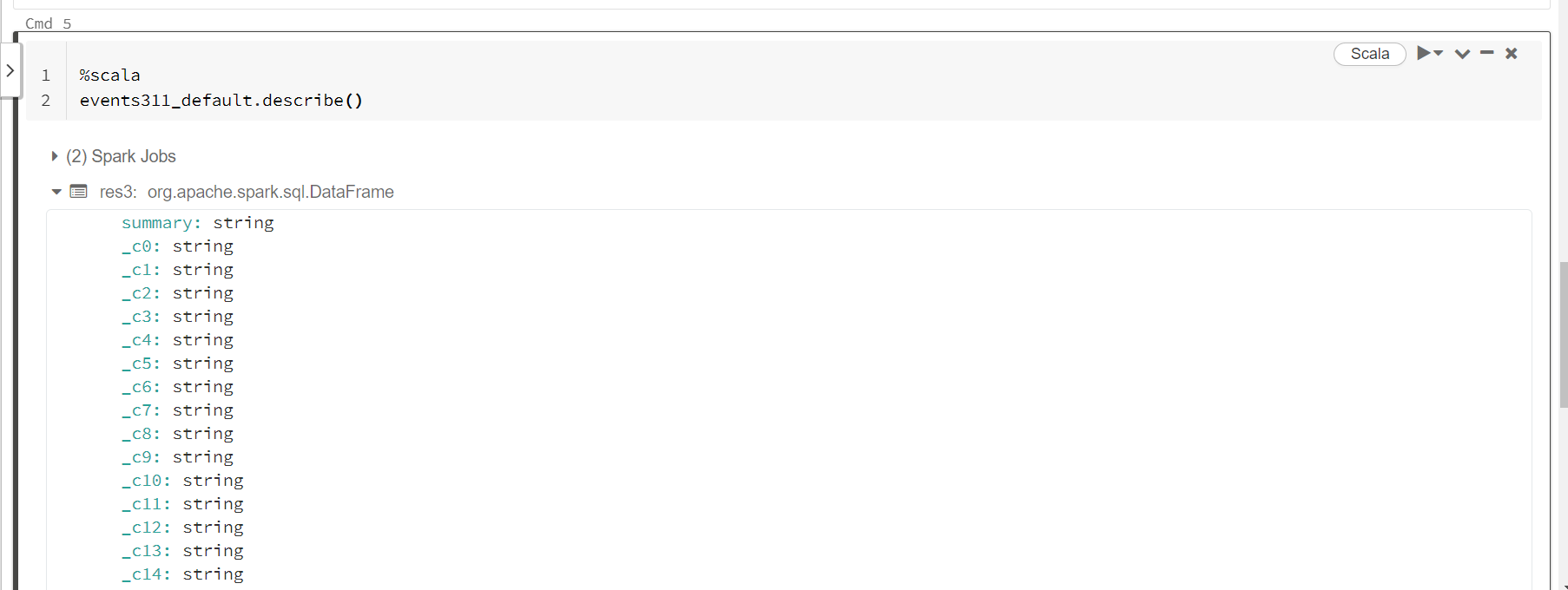
**Name**:

**Date:**

**Section Two**

The screenshots needs to show your user name and the date loaded.

28. Screenshot of the loaded data frame.



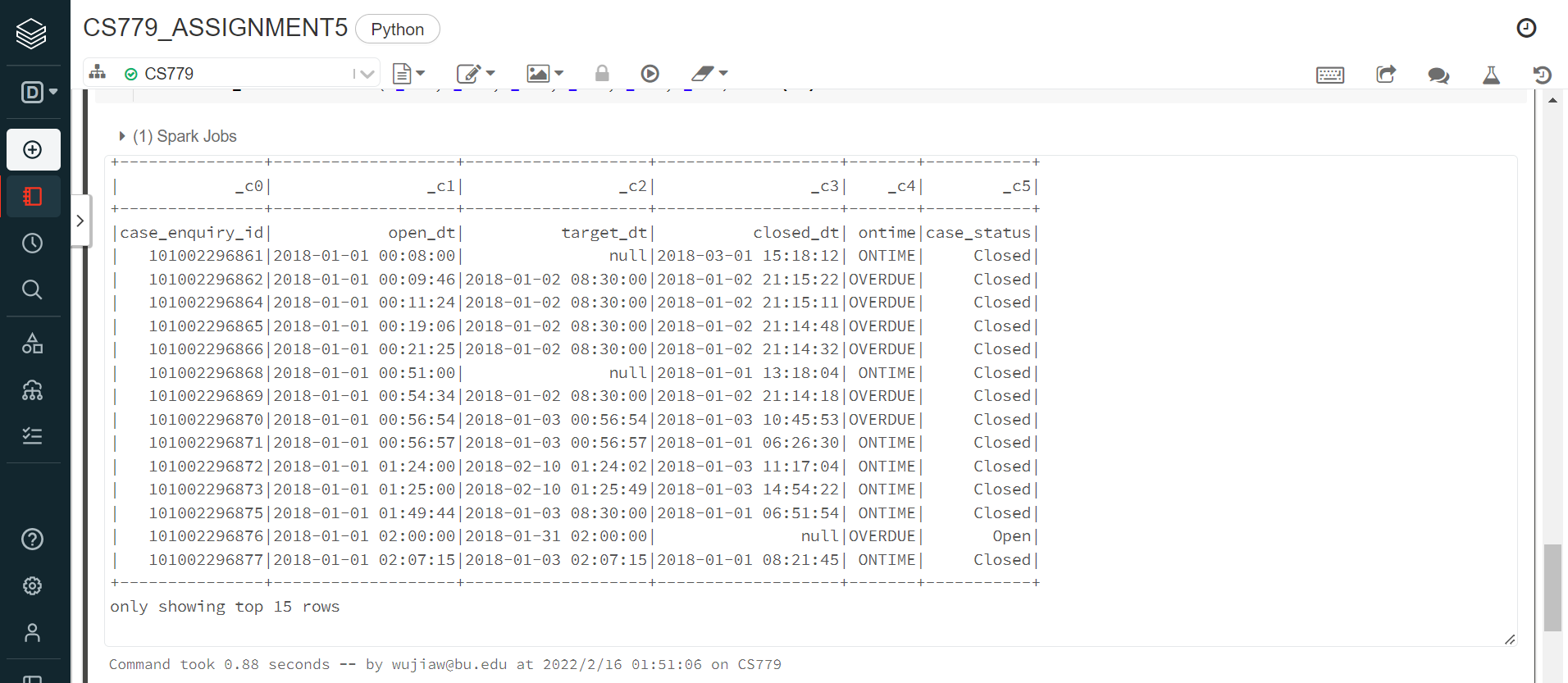


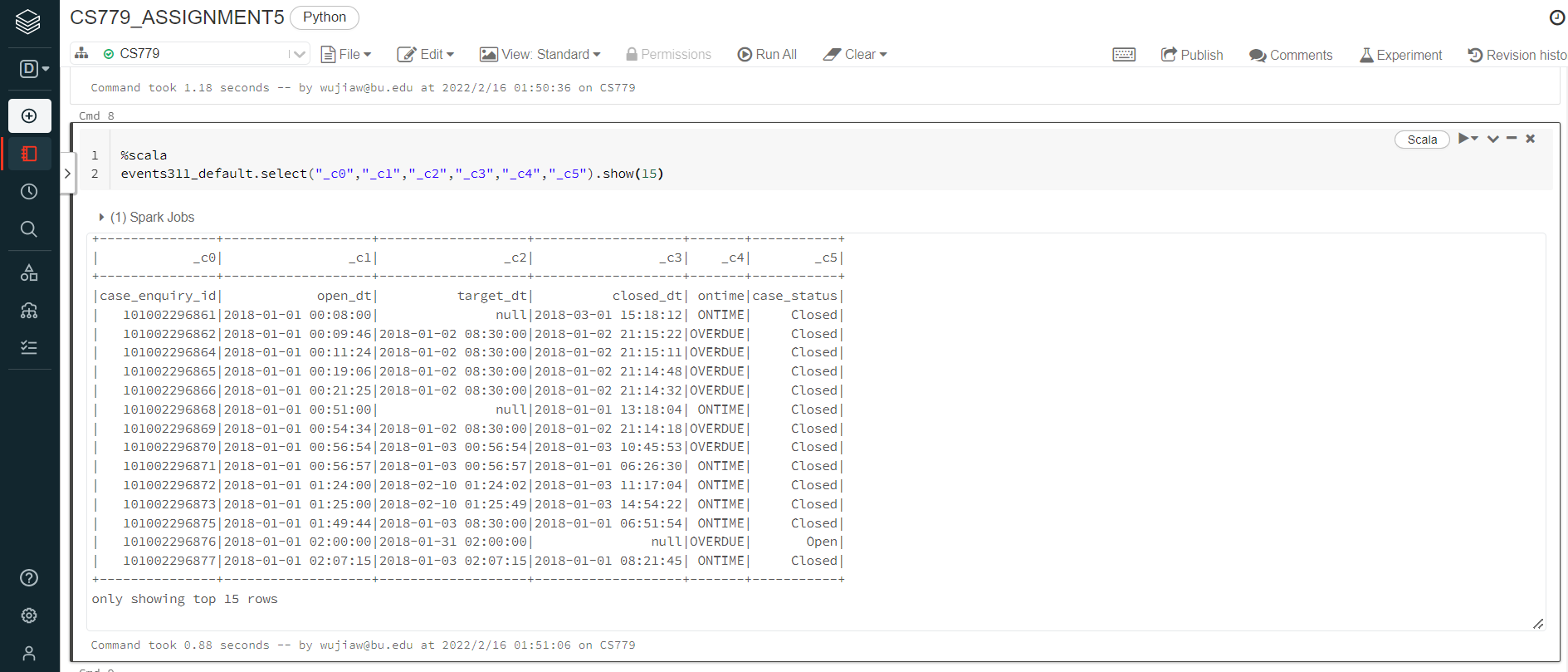
29. How many c fields are listed? \_\_\_29 （from \_c0 to \_c28）\_\_

32. Provide the query command and the resulting data set

%scala

events311\_default.select("\_c0","\_c1","\_c2","\_c3","\_c4","\_c5").show(15)





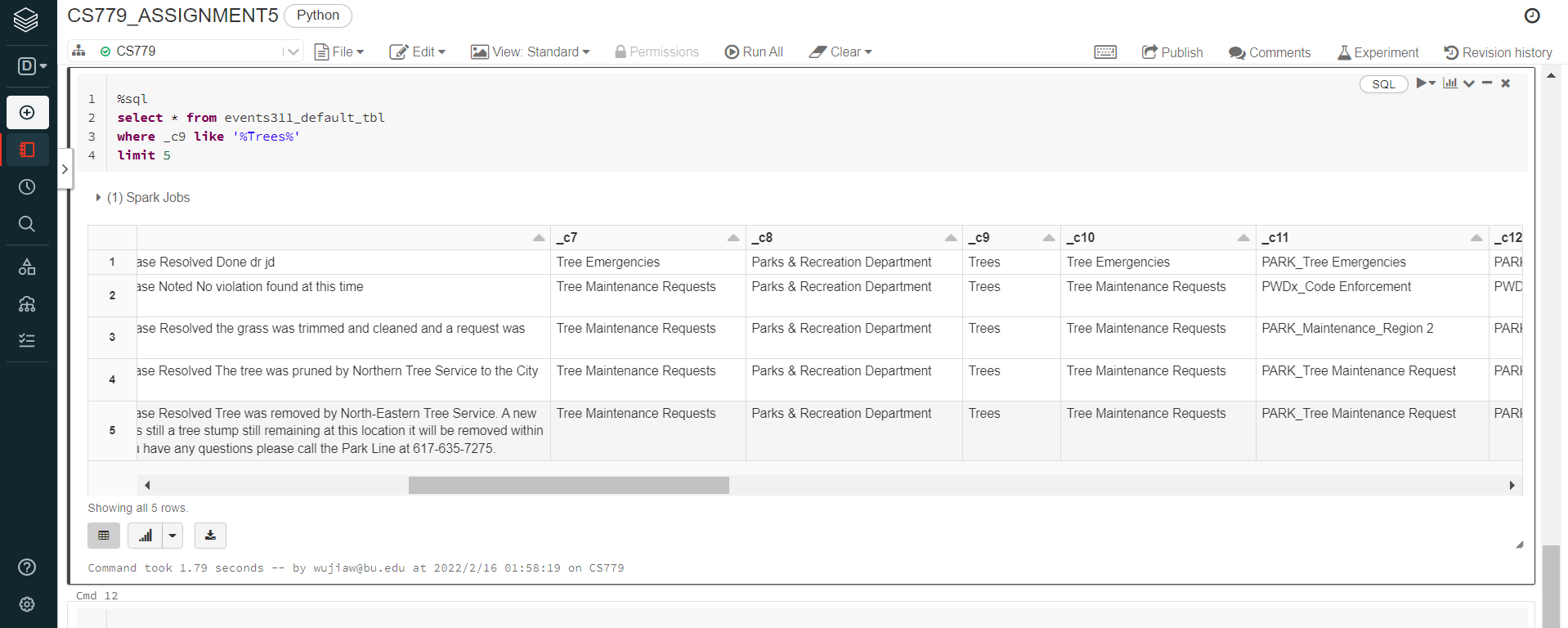
35. Provide the query command and the resulting data set

%sql

select \* from events311\_default\_tbl

where \_c9 like '%Trees%'

limit 5

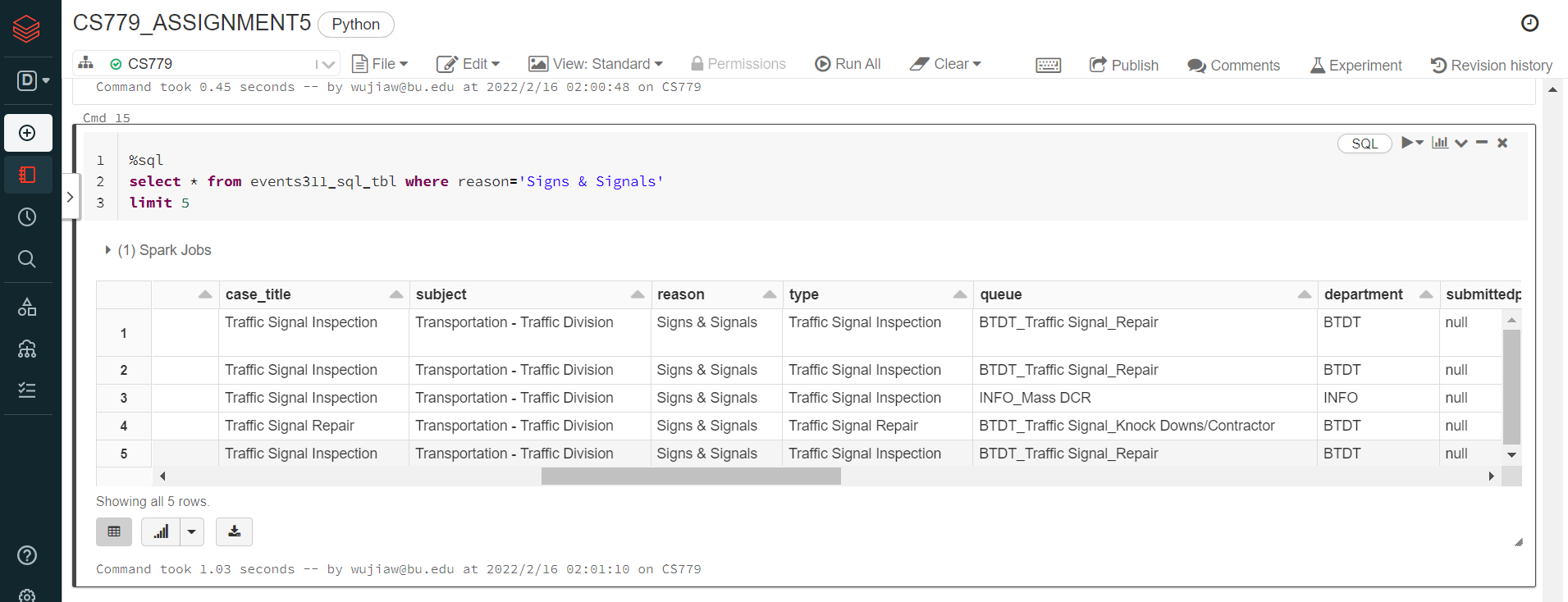


39. Provide the query command and the resulting data set

%sql

select \* from events311\_sql\_tbl where reason='Signs & Signals'

limit 5



40. Provide the query command and the resulting data set including chart

%sql

select ontime, count(1) from events311\_sql\_tbl

group by ontime



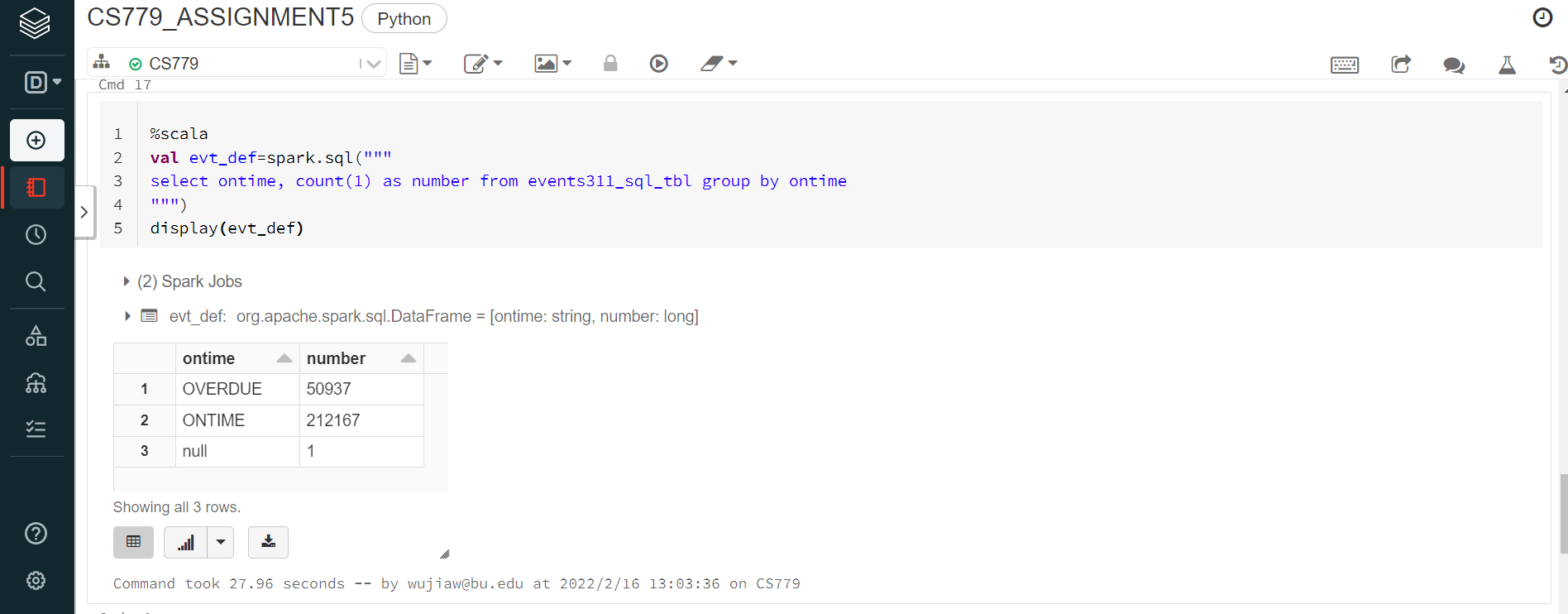
%scala

val evt\_def=spark.sql("""

select ontime, count(1) as number from events311\_sql\_tbl group by ontime

""")

display(evt\_def)



%python

import pandas as pd

import matplotlib.pyplot as plt

evt\_def=spark.sql("""

select ontime, count(1) as number from events311\_sql\_tbl group by ontime

""")

x=['OVERDUE','ONTIME','None']

evt\_def.describe()

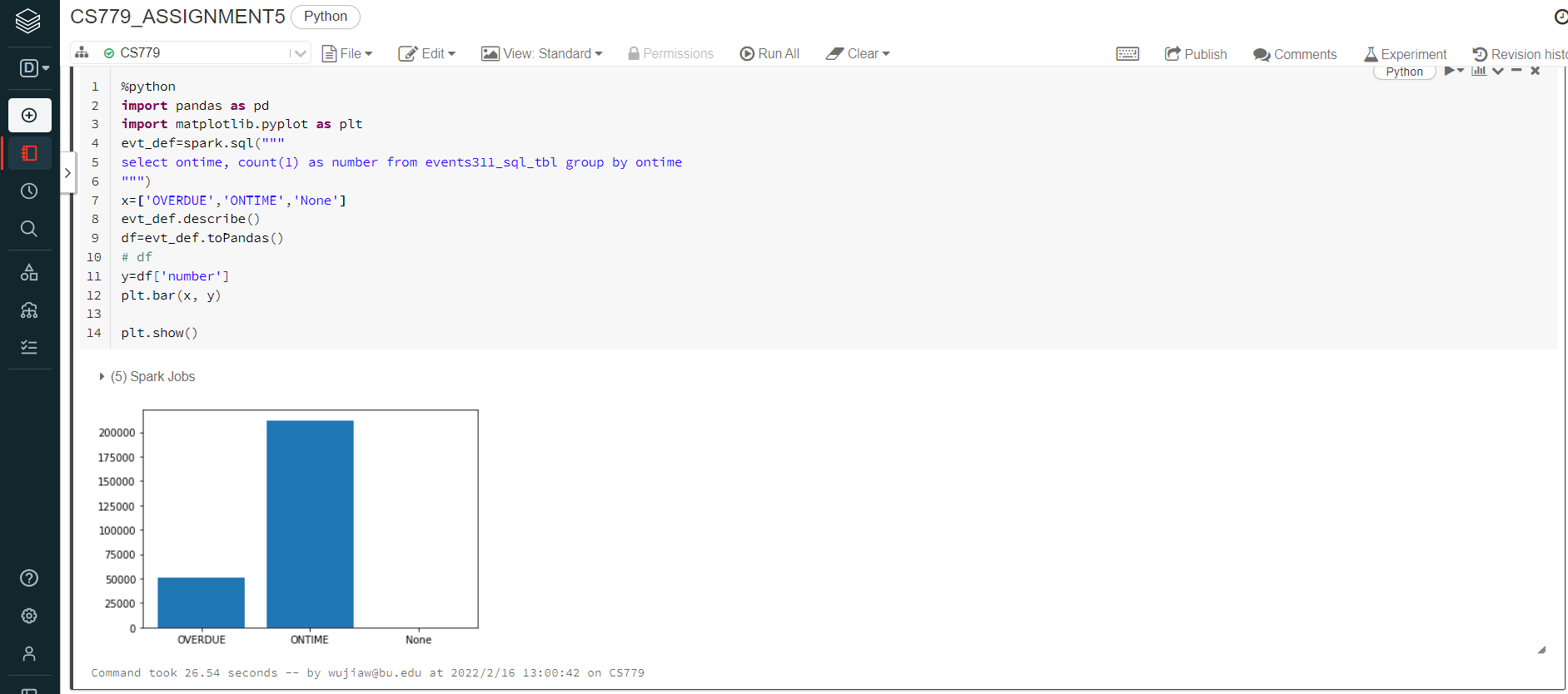
df=evt\_def.toPandas()

# df

y=df['number']

plt.bar(x, y)

plt.show()



41. Very briefly explain what you have discovered based on your data set from the query above.

Based on the data set from the query above, it can be found that “ontime” records are most, about 4 times of “overdue” records. And only one record that doesn’t know whether if ontime.

Use the **Ask your Facilitator Discussion Board** if you have any questions regarding the how to approach this assignment.

Save your assignment as ***lastnameFirstname\_lassignment5.doc*** and submit it in the *Assignments* section of the course.

For help uploading files please refer to the *Technical Support* page in the syllabus.

Your lab submission will be evaluated according to the following rubric.

|  |  |  |
| --- | --- | --- |
|  | **Letter Grade** | **Qualities Demonstrated by the Lab Submission** |
| **Correctness, completeness, and constitution**  **Measures the correctness and completeness of the results, and the quality of the constitution of the SQL constructs** | A+ 🡺 100 | The results and explanations are entirely complete and correct for all steps. There are absolutely no technical or other errors present. There is no known way to improve the logic and makeup of any of the SQL constructs. |
| A 🡺 96 | One insignificant technical or other error is present, but otherwise the results and explanations are entirely complete and correct for all steps. Excluding the insignificant error, there is no known way to improve the makeup of any of the SQL constructs. |
| A- 🡺 92 | One or two consequential technical or other errors are present, but otherwise the results and explanations are entirely complete and correct for all steps. Excluding the one or two errors, there is no known way to improve the makeup of any of the SQL constructs. |
| B+ 🡺 88 | A few steps have significantly incomplete or incorrect results or explanations. The results and explanations are complete and correct for the remainder of the steps. The logic and makeup of most SQL constructs are sound. |
| B 🡺 85 | A few steps have significantly incomplete or incorrect results or explanations. The results and explanations are mostly complete and correct for the remainder of the steps, with the exception of a few insignificant technical or other errors. The logic and makeup of most SQL constructs are sound. |
| B- 🡺 82 | About ¼ of the steps have significantly incomplete or incorrect results or explanations. The results and explanations are complete and correct for the remainder of the steps. The logic and makeup of at least ¾ of the SQL constructs are sound. |
| C+ 🡺 78 | About ¼ of the steps have significantly incomplete or incorrect results or explanations. The results and explanations are mostly complete and correct for the remainder of the steps, with the exception of a few insignificant technical or other errors. The logic and makeup of at least ¾ of the SQL constructs are sound. |
| C 🡺 75 | About half of the steps have significantly incomplete or incorrect results or explanations. The results and explanations are complete and correct for the remainder of the steps. The logic and makeup of at least half of the SQL constructs are sound. |
| C- 🡺 72 | About half of the steps have significantly incomplete or incorrect results or explanations. The results and explanations are mostly complete and correct for the remainder of the steps, with the exception of a few insignificant technical or other errors. The logic and makeup of at least half of the SQL constructs are sound. |
| D 🡺 67 | About ¾ of the steps have significantly incomplete or incorrect results or explanations. The results and explanations are complete and correct for the remainder of the steps. The logic and makeup of at least ¼ of the SQL constructs are sound |
| F 🡺 0 | All or almost all of the steps have incomplete or incorrect results or explanations. The logic and makeup of all or almost all of the SQL constructs are unsound. |