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.

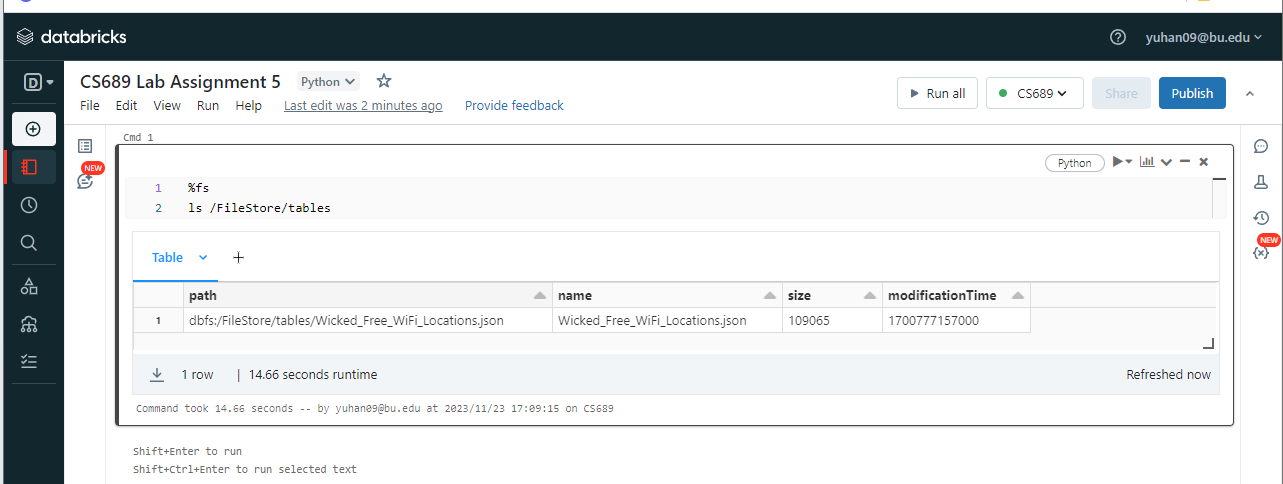
**Name**: Yuhan Xu

**Date: 11/23/2023**

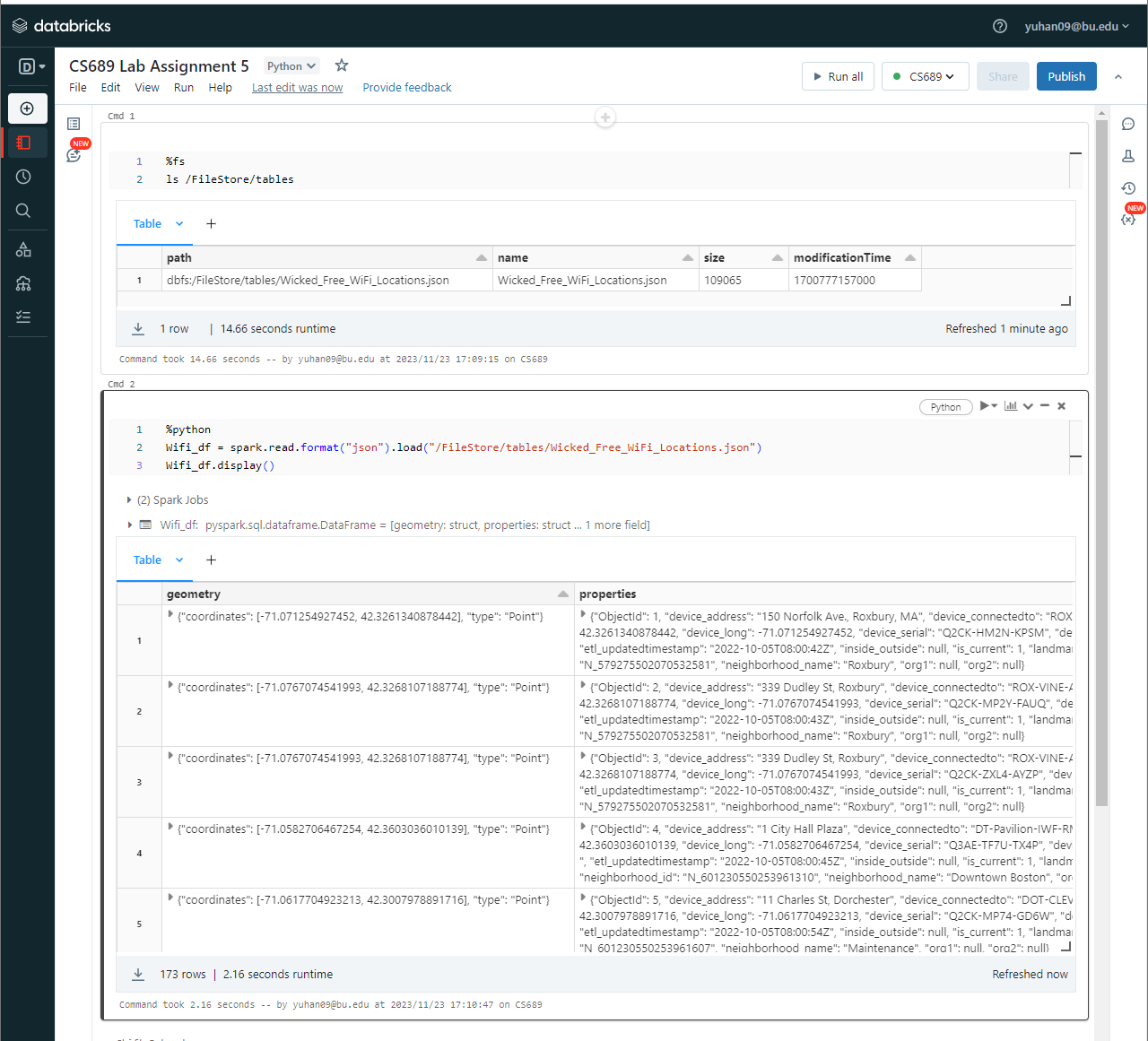
**Section Two**

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

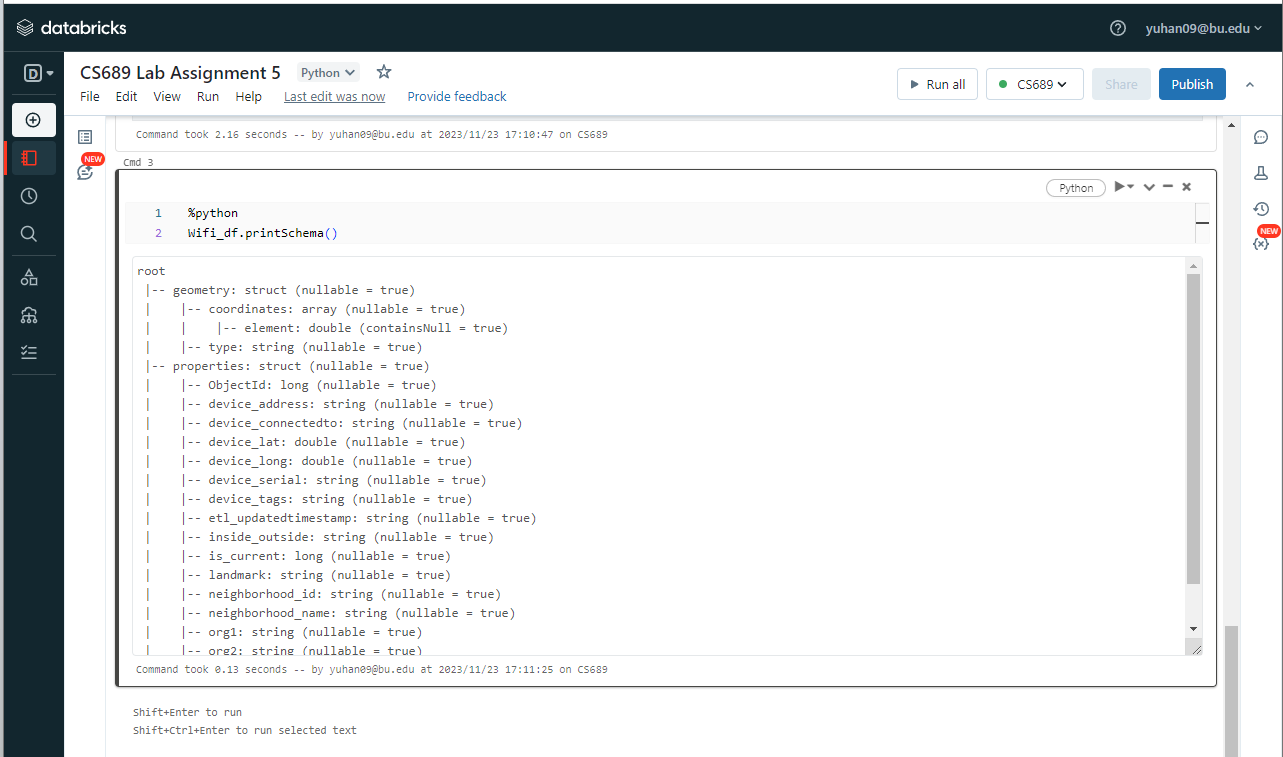
15. Screenshot of the loaded file.



16. Screenshot of the loaded data frame.



17. Provide the query command and the resulting data set



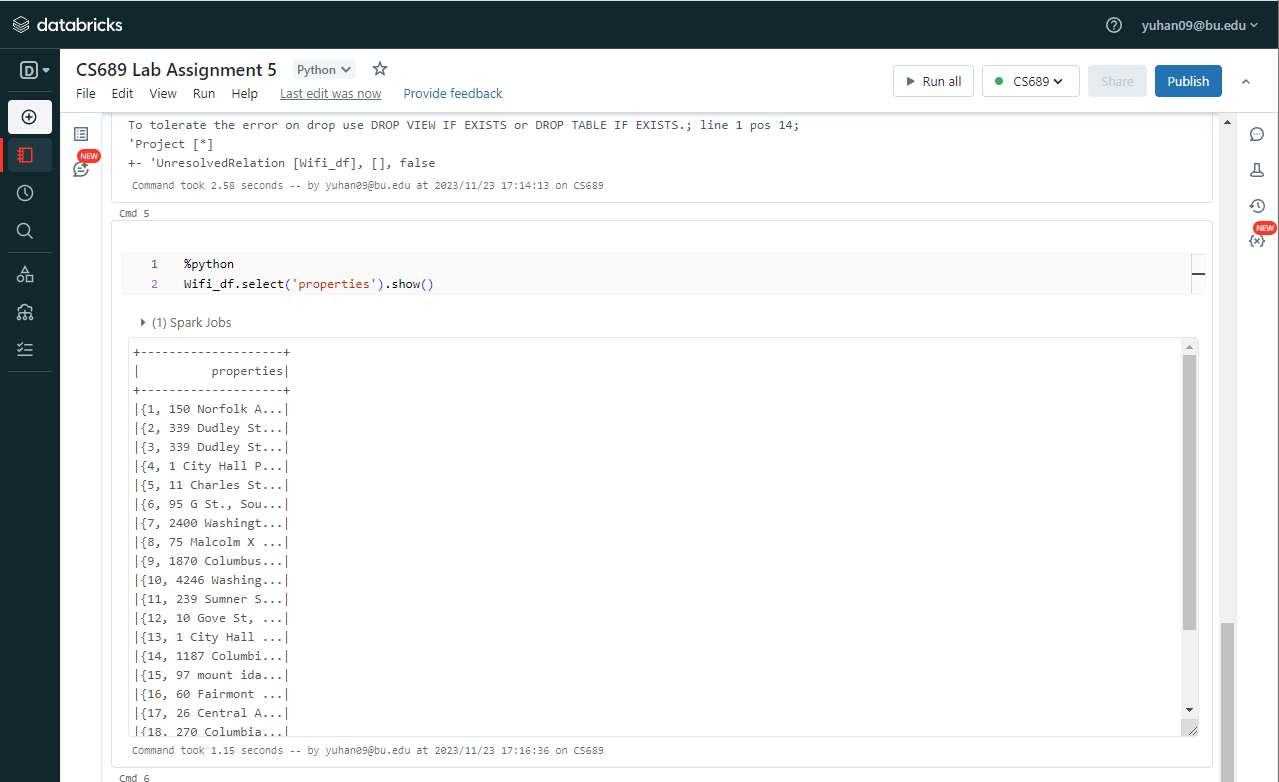
18. Briefly describe the structure of the data frame.

The data frame structure consists of two main parts: **geometry** and **properties**.

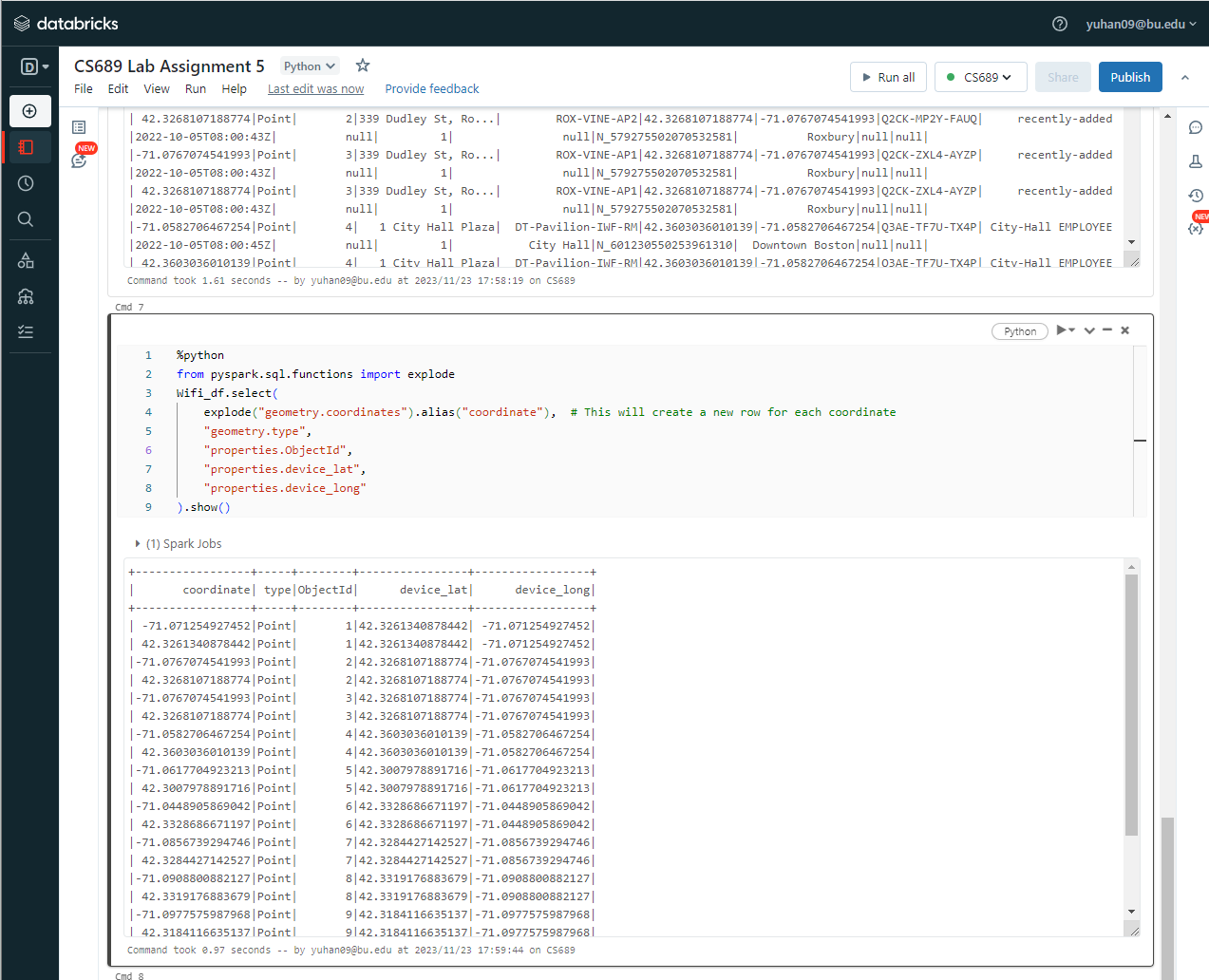
**geometry** includes geographical coordinates and a type descriptor.

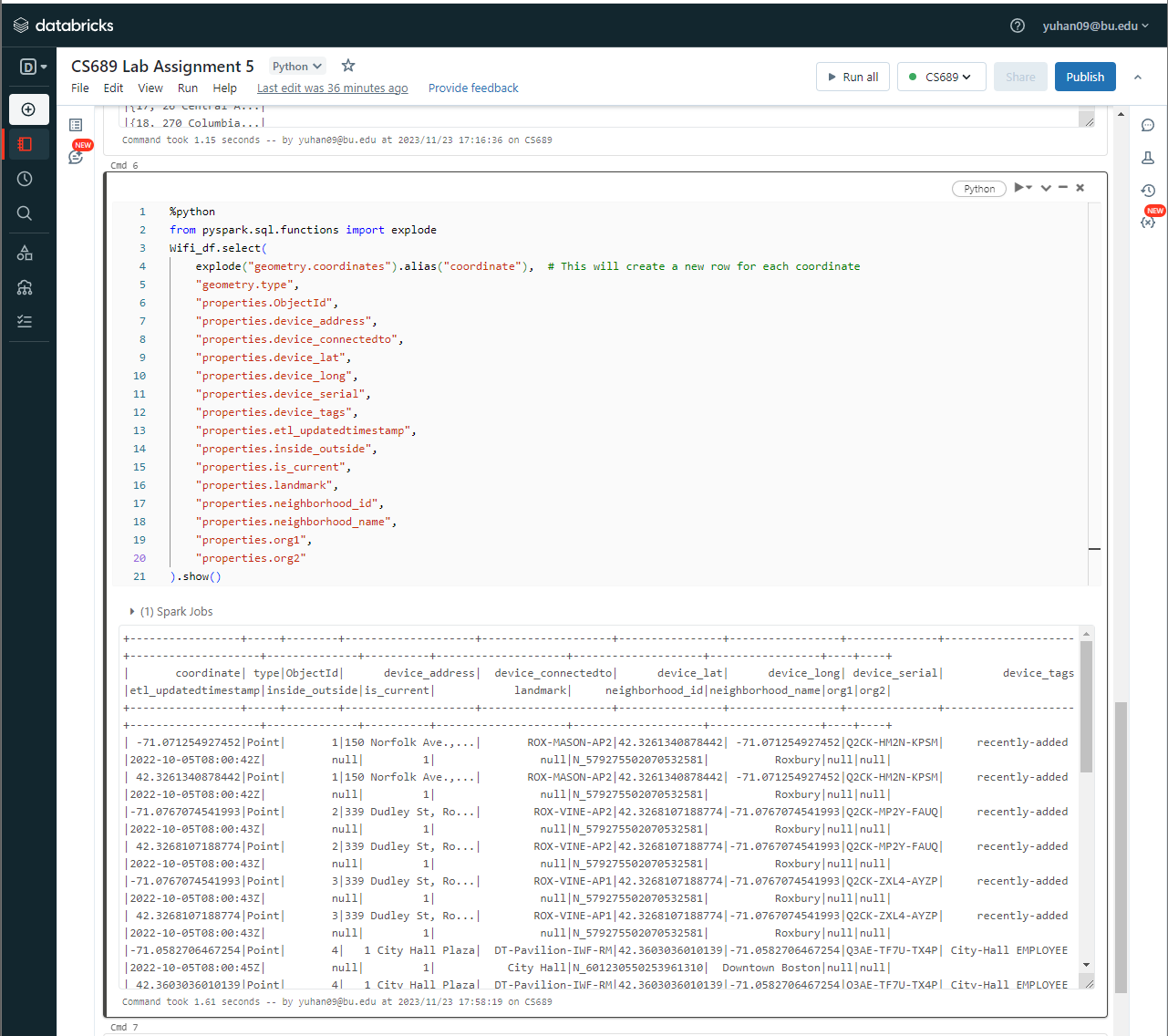
**properties** contains various attributes related to a device, like its location, identification details, and organizational information.

21. Provide the query command and the resulting data set

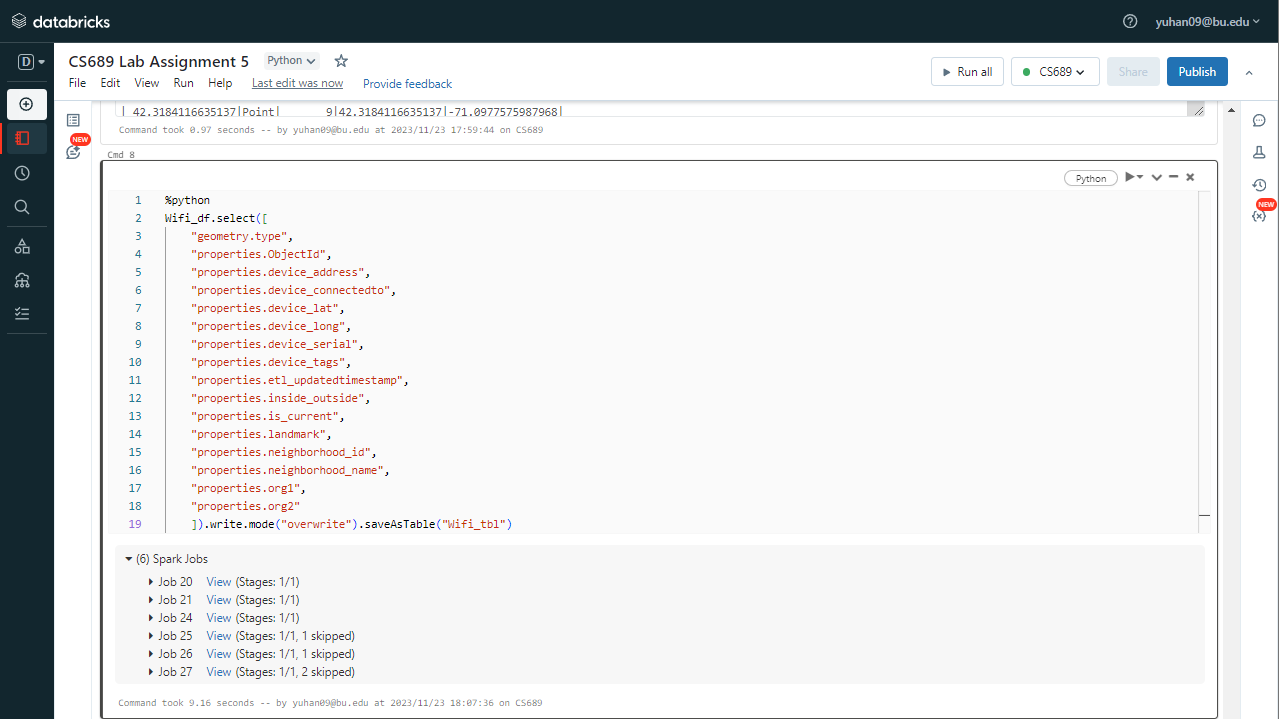
、

Extra credit

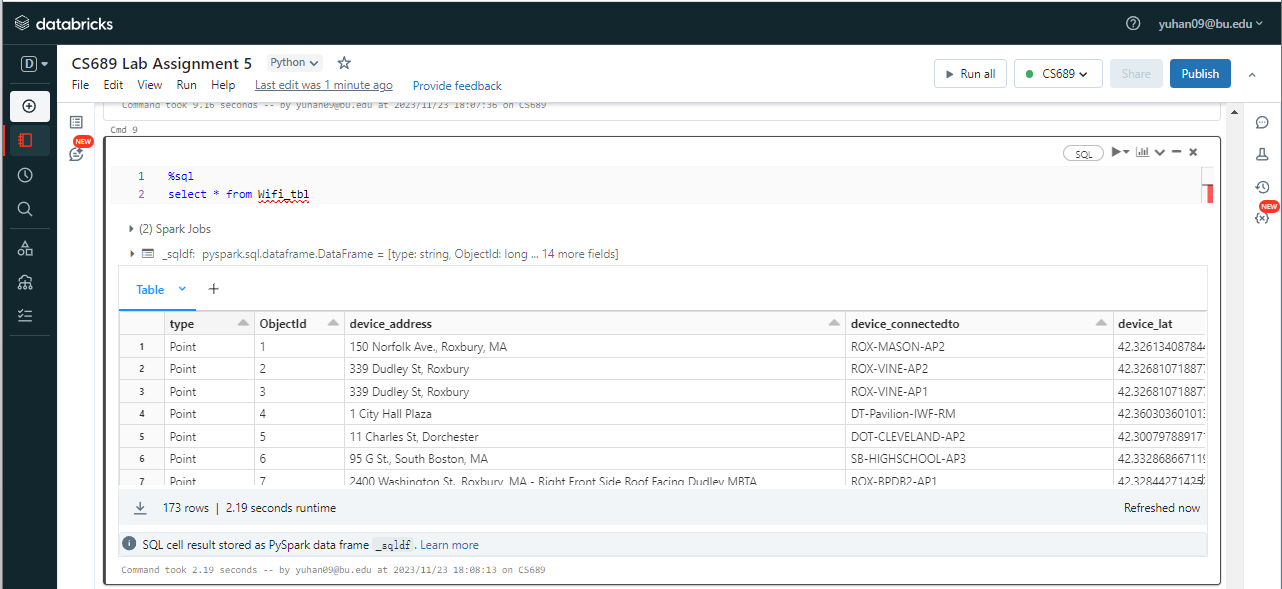
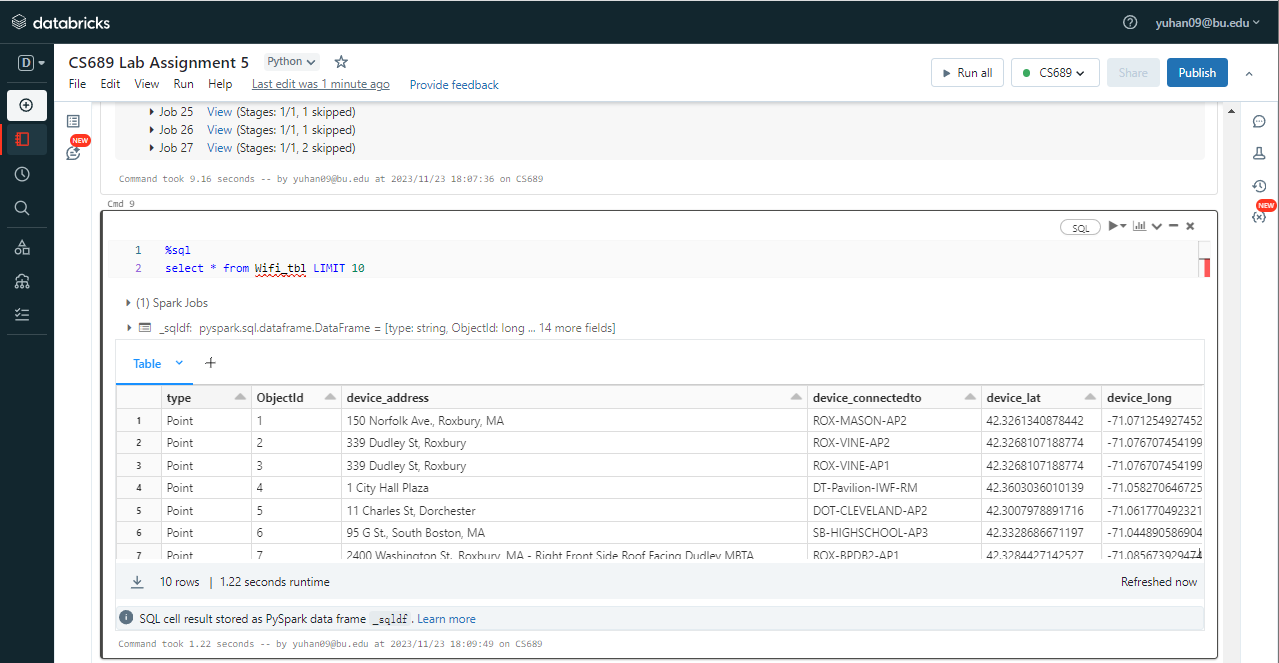




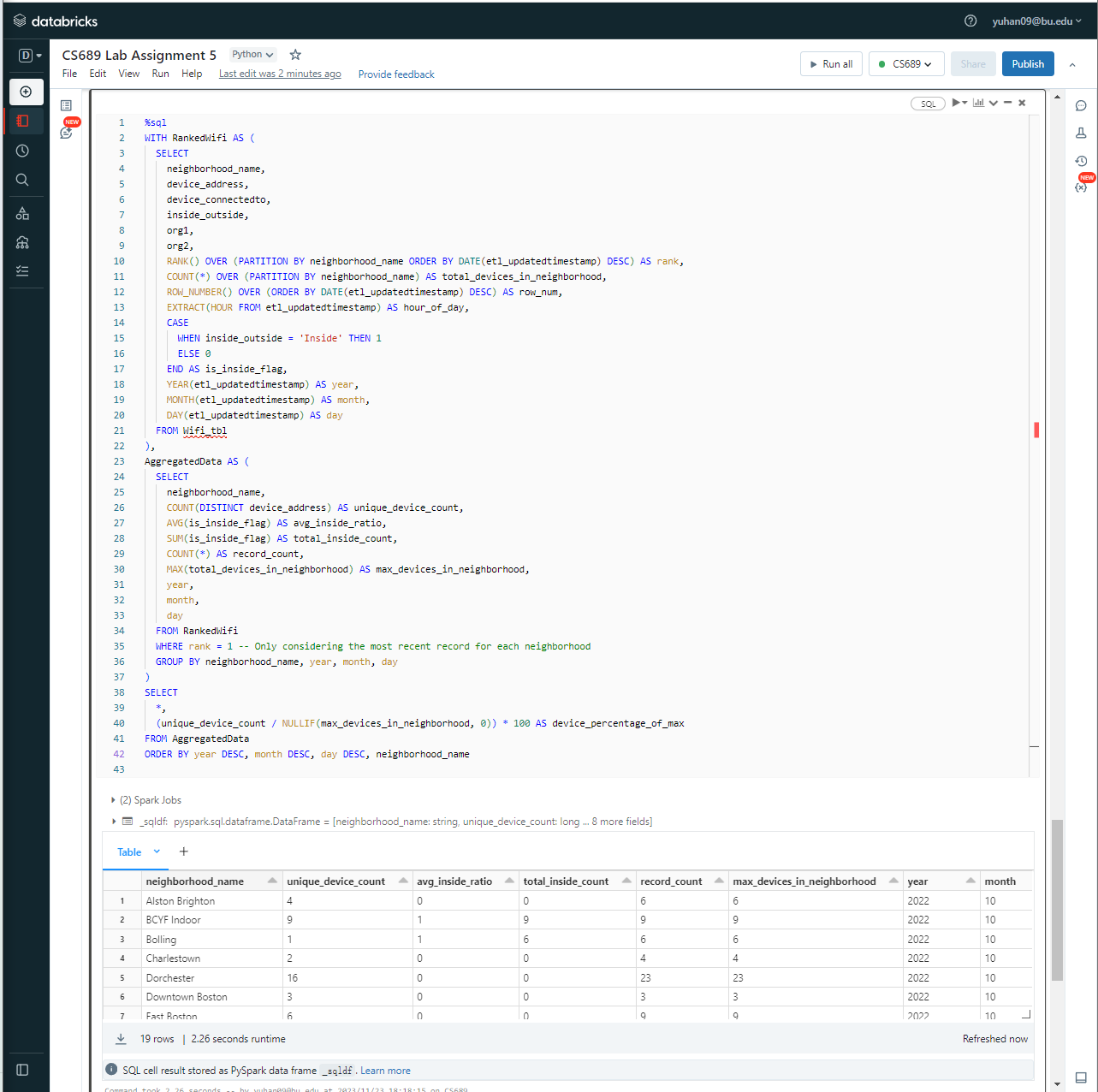
22. Provide the query command and the resulting data set



23. Provide the query command and the resulting data set



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



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

Based on the query results, it seems that the dataset contains information on WiFi devices across various neighborhoods for October 2022. Here are some observations:

**The unique\_device\_count column** suggests that the number of distinct WiFi devices varies by neighborhood. For instance, Charlestown has the highest count with 16, while Allston Brighton has the lowest with 4.

**The avg\_inside\_ratio column** indicates whether devices are predominantly inside or outside, with '0' meaning all devices are outside and '1' meaning all are inside. The data shows that in some neighborhoods, all WiFi devices are inside, while in others, the distribution is mixed.

**The total\_inside\_count mirrors** the avg\_inside\_ratio, giving the total number of devices that are inside.

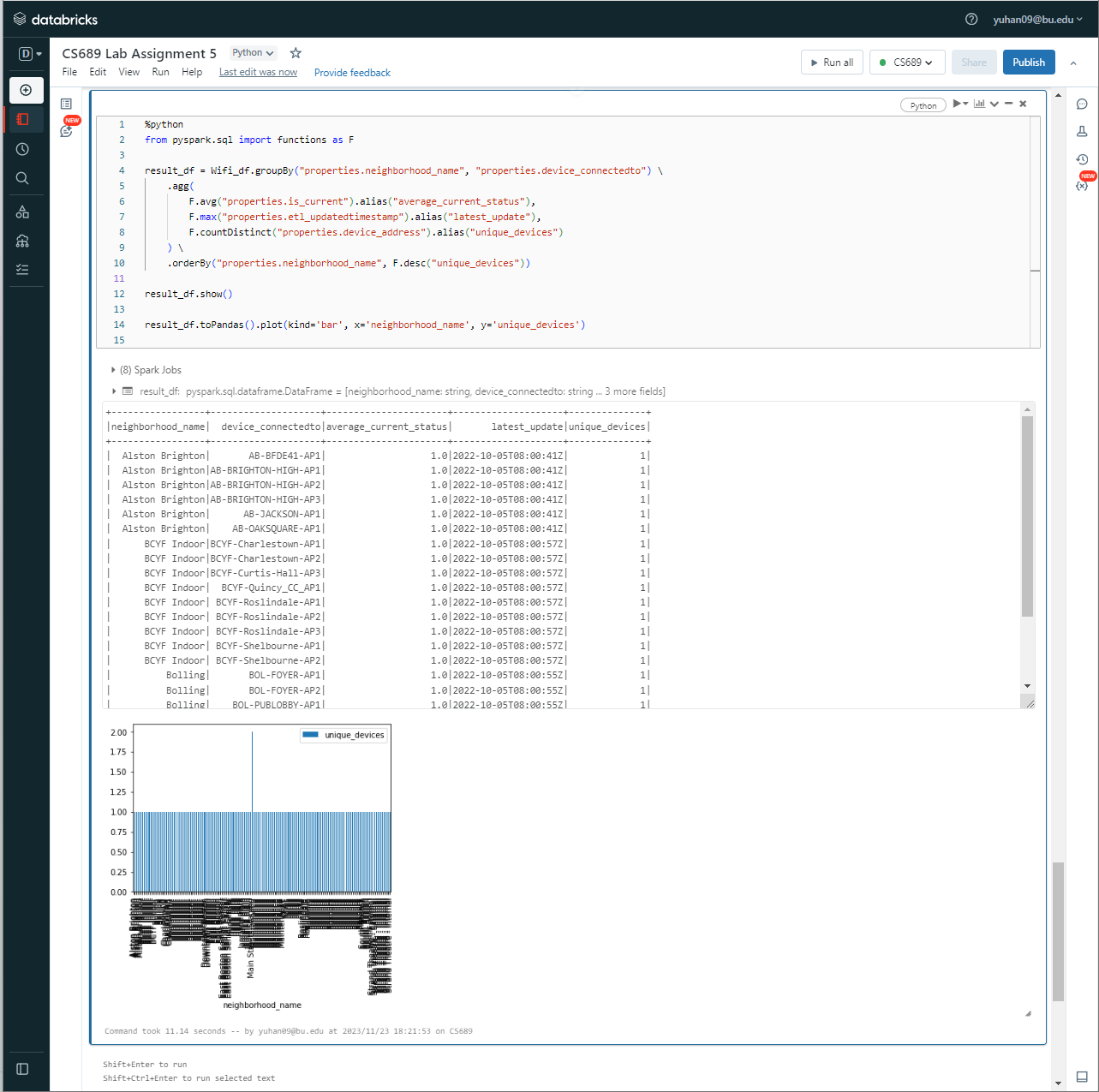
**The record\_count** seems to represent the total number of records for each neighborhood, which coincides with the max\_devices\_in\_neighborhood, suggesting the dataset contains one record per device for the month.

**The year, month, and day columns** show that all the records are from October 2022, with no daily breakdown, implying the data may have been aggregated on a monthly basis.

**The device\_percentage\_of\_max column** is missing in the result, which suggests either an issue with the query or that the column was not included in the final SELECT statement.

This analysis provides insights into the distribution and inside/outside status of WiFi devices across different neighborhoods at a certain point in time.

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



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

Based on the query results, it appears that:

The query has aggregated data by neighborhood\_name and device\_connectedto, which is the identifier of the connection the device is using.

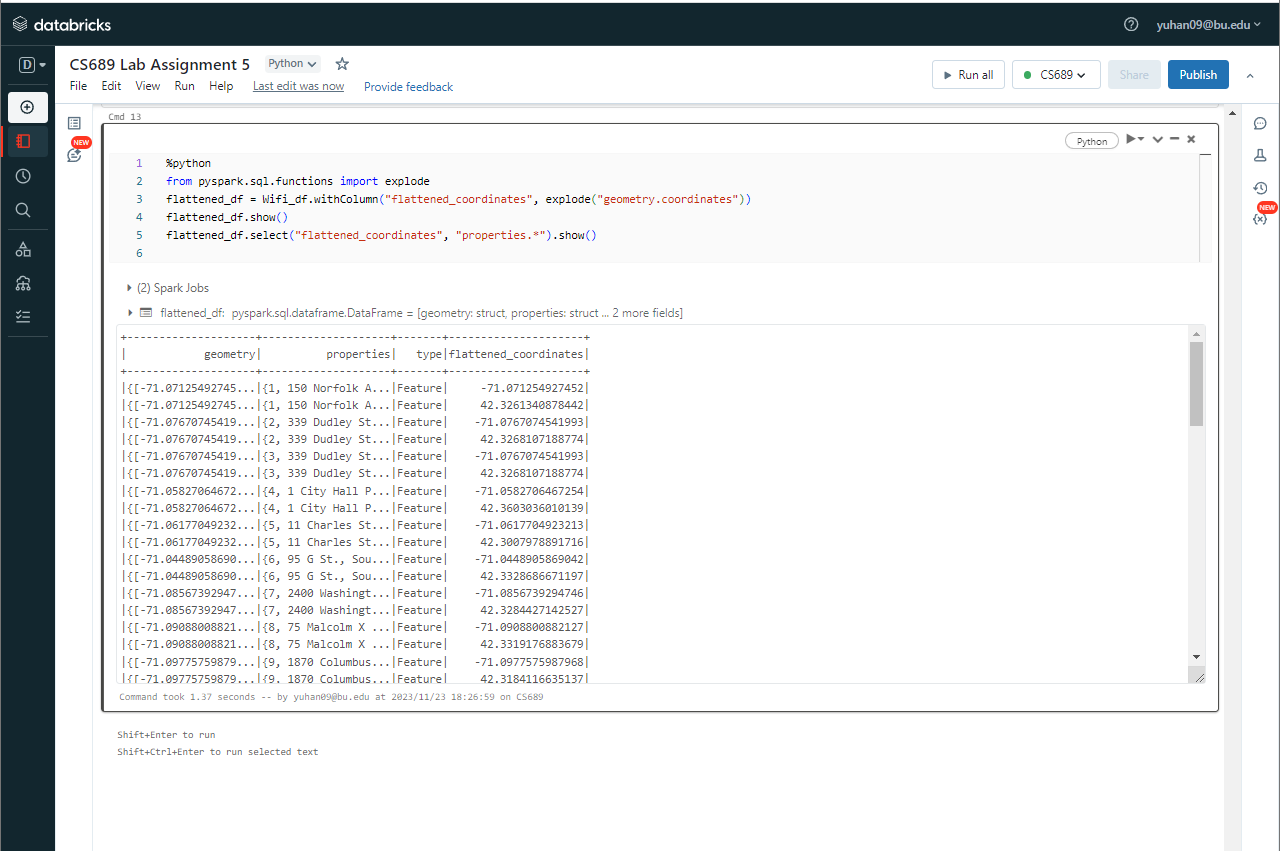
The average\_current\_status seems to be consistently 1 for all the rows in the displayed output, which might indicate that all the devices are currently active or in use, assuming is\_current denotes an active status.

The latest\_update column shows the same timestamp across different entries, suggesting that the data was possibly updated in a batch process at the same time for all devices or that the dataset represents a snapshot at a specific point in time.

The unique\_devices column indicates the number of unique devices for each device\_connectedto within each neighborhood, which is 1 for all the shown records. This suggests there might be a one-to-one relationship between devices and their connection points in this dataset or that the dataset has a granularity that doesn't capture multiple devices connecting to the same point.

The bar chart below the table visualizes the number of unique devices by neighborhood. However, due to the granularity of the data and the way it's presented in the chart, it looks like each device is counted separately for each connection point, resulting in a large number of bars all of the same height, each representing a single device. This visualization may not be particularly helpful as it stands; a more aggregated view (e.g., total devices per neighborhood) might provide clearer insights.

**Extra Credit (2 points):** Note how one of the columns in the original data frame is an array of coordinates. Look to use the explode function to extract those coordinates into a separate flattened data frame.



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.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Criterion | A | B | C | D | F | Letter Grade |
| Correctness and Completeness of Results (70%) | All steps' results are entirely complete and correct | About ¾ of the steps' results are correct and complete | About half of the steps' results are correct and complete | About ¼ of the steps' results are correct and complete | Virtually none of the step's results are correct and complete |  |
| Constitution of SQL/Python and Explanations (30%) | Excellent use and integration of appropriate SQL/Python constructs and supporting explanations | Good use and integration of appropriate SQL/Python constructs and supporting explanations | Mediocre use and integration of appropriate SQL/Python constructs and supporting explanations | Substandard use and integration of appropriate SQL/Python constructs and supporting explanations | Virtually all SQL/Python constructs and supporting explanations are unsuitable or improperly integrated |  |
|  |  |  |  |  | Assignment Grade: |  |