

BDAT1007- Assignment 1

First, I made sure that I have a local repository set up in RapidMiner. I did not create new repository. On the left side of the interface as I shown in **Figure 1**, I clicked on **Local repository** to create a new connection. Clicked on it and created a connection with the name "**Assign1**." Enter my Twitter credentials and test the connection to ensure it is successful. Once the connection is established, proceed to create a new process.

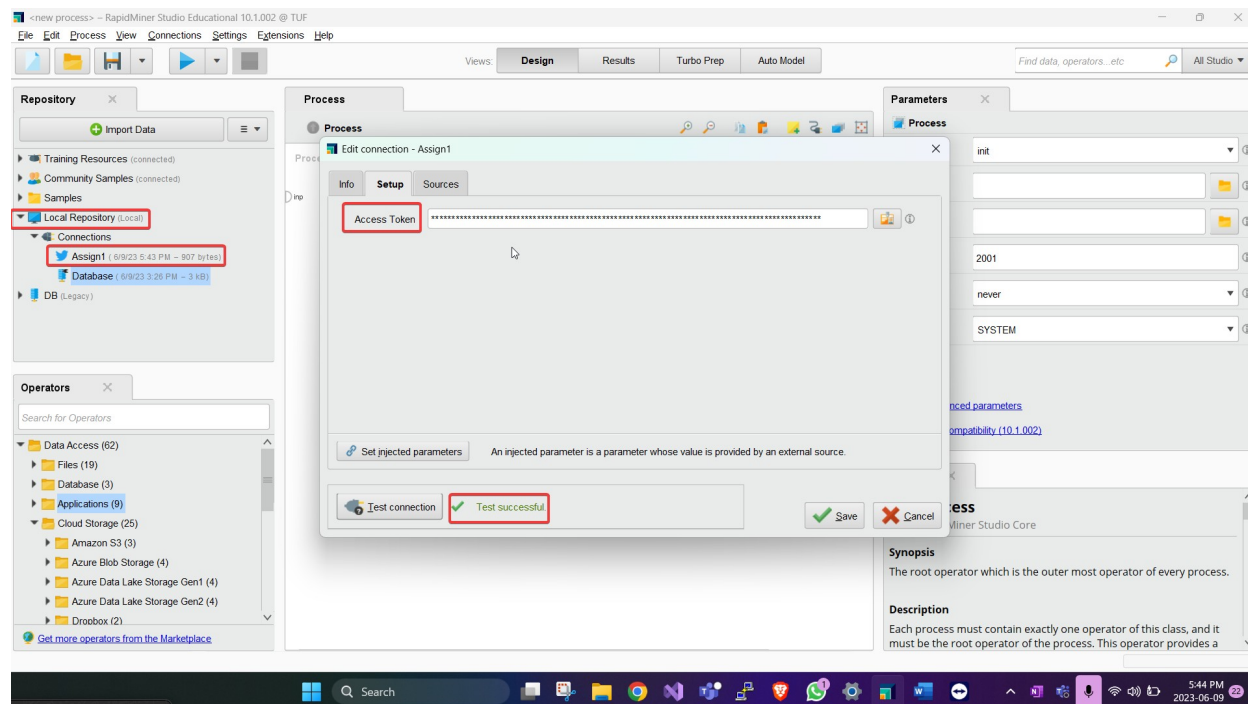


Figure 1 Test successful.

I used the "**Twitter Search**" operator in **RapidMiner** to gather tweets related to the query "**Gaming**." This operator allowed me to collect a dataset of tweets that can be used for sentiment analysis and other types of analysis. By searching for tweets on a specific topic, I can analyze public opinion and track trends.

To configure the "**Twitter Search**" operator in RapidMiner, I accessed the "**Trendmap.com**" website, where I found interesting trends related to gaming. Then, on the right side of the interface, I set up the "Twitter Search" operator by providing the necessary connection and access token. This allows RapidMiner to use the Twitter extension and access the Twitter API.

In the configuration of the operator, I specified the connection entry and query. I selected "**Gaming**" as the query term, set a limit of **1000 tweets to retrieve**, specified the language as **English**, and chose the result type as "recent." These settings ensured that I obtained a dataset of recent tweets related to gaming for further analysis in RapidMiner **as shown in Figure 2**

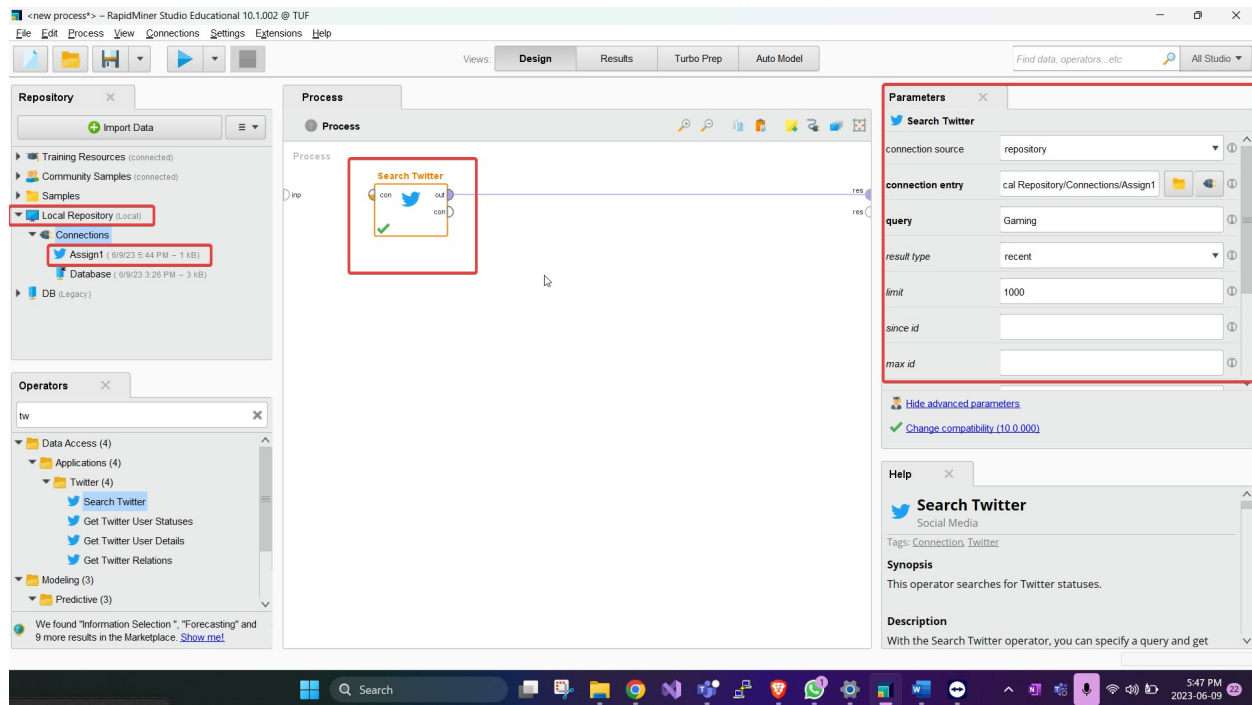


Figure 2 Search Twitter

The image shows the RapidMiner Studio Educational 101.002 interface. At the top, the title bar reads "<new process> - RapidMiner Studio Educational 101.002 @ TUF". The main menu includes File, Edit, Process, View, Connections, Settings, Extensions, and Help. Below the menu is a toolbar with icons for file operations and a 'Views' section with buttons for Design, Results, Turbo Prep, and Auto Model. The 'Results' view is active, displaying a table titled 'ExampleSet (Search Twitter)'. The table has columns: User-Id, To-User, To-User-Id, Lan..., and Text. The 'Text' column contains various tweets, some of which are highlighted in yellow. To the right of the table is a 'Filter (820 / 820 examples)' dropdown. On the far right, a 'Repository' panel shows a tree structure with 'Local Repository (Local)' expanded, revealing 'Connections' and 'Database' sub-items. The bottom status bar indicates 'ExampleSet (820 examples, 1 special attribute, 11 regular attributes)'. The Windows taskbar at the bottom shows the time as 5:48 PM on 2023-06-09.

In Figure 4, the "Write Excel" operator is utilized to export the data to an Excel file. On the right side of the interface, I provided the path for the Excel file by specifying the desired location.

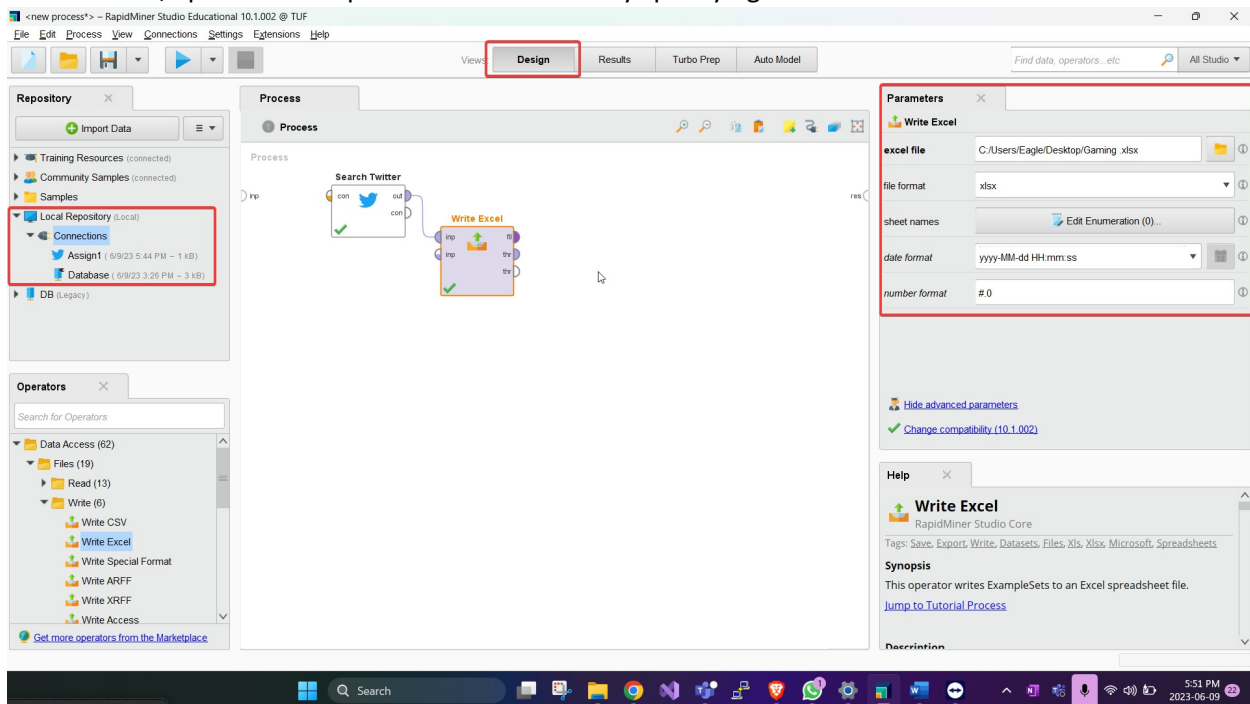


Figure 4 Export the data to an Excel file .

In Figure 5, the "Read Excel" operator is employed to read an Excel file. This operator allowed me to retrieve the data from the specified Excel file and use it for further analysis or processing within RapidMiner.

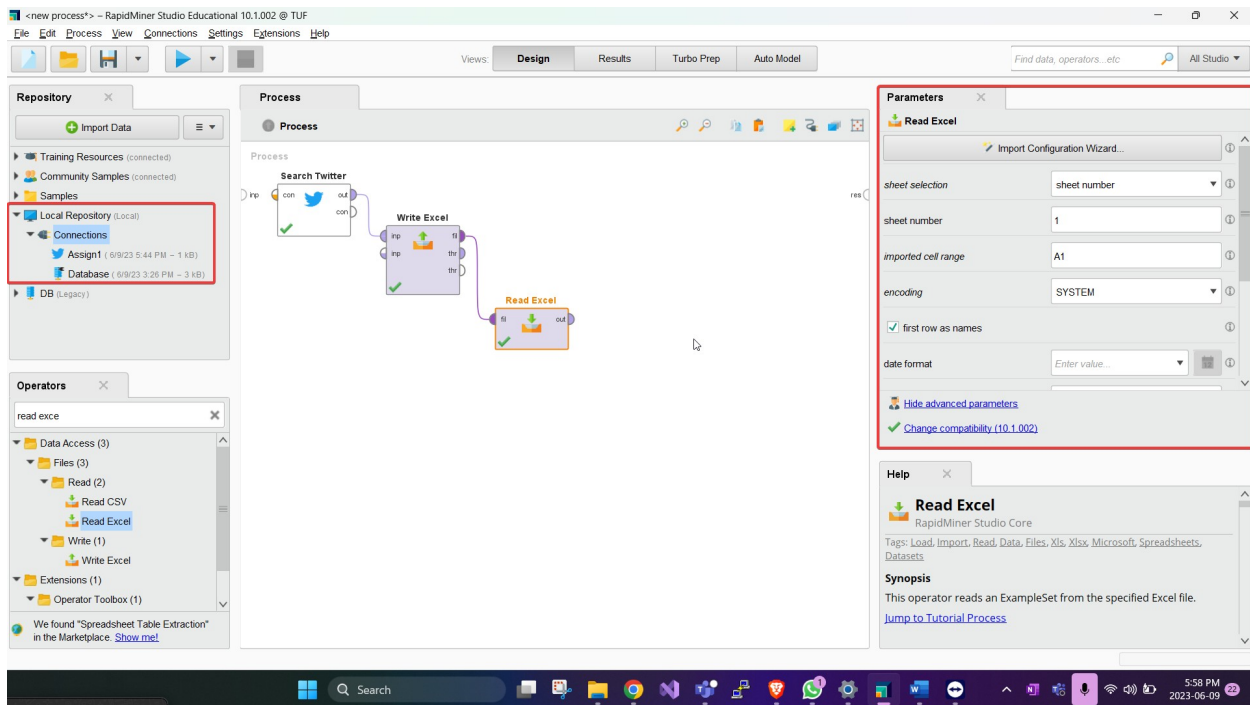


Figure 5 Data Retrieval and Analysis using the "Read Excel" Operator in RapidMiner

In Figure 6, the result of reading the Excel file is displayed. The data from the Excel file is shown, indicating that the file was successfully read, and its contents are available for further analysis or manipulation within RapidMiner.

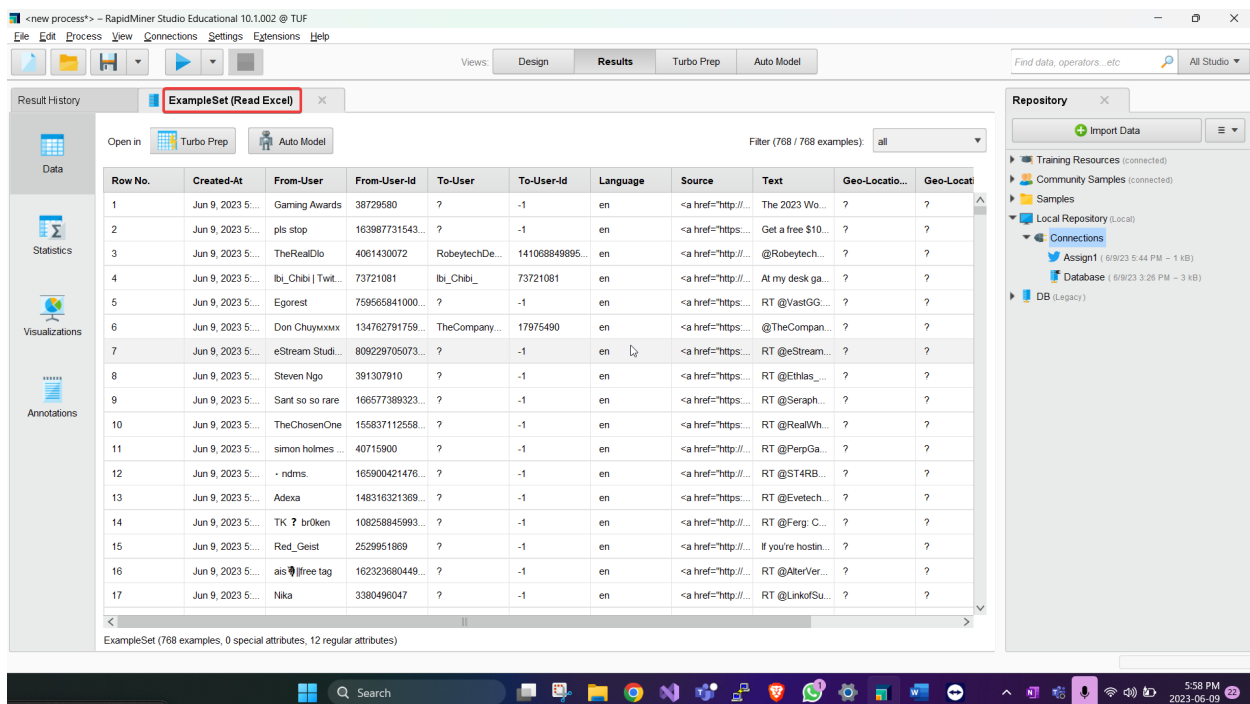


Figure 6 Successful Excel Data Reading for Analysis in RapidMiner

In Figure 7, I applied the "Nominal to Text" operator. By selecting the parameter to filter all attributes and running the operator, I transformed the nominal attributes in the data into text format.

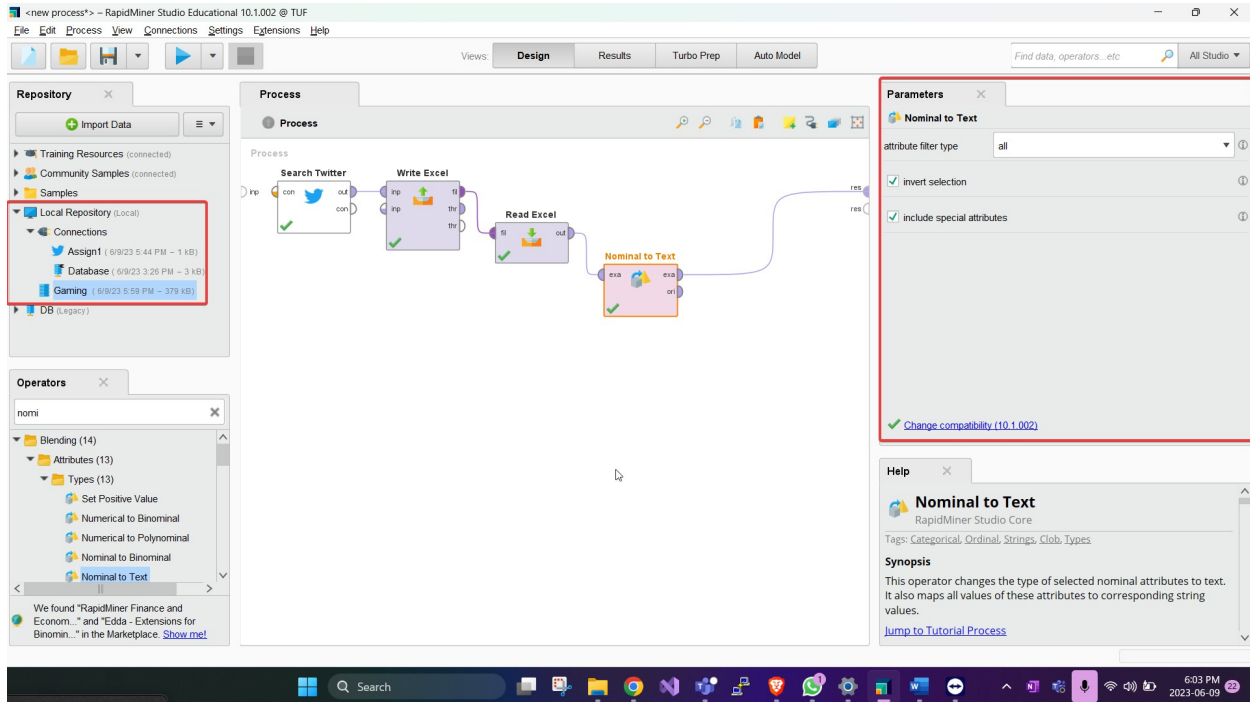


Figure 7 Nominal to Text Conversion for Enhanced Data Analysis

In Figure 8, I utilized the "Select Attribute" operator to choose specific attributes from the dataset. I encountered difficulties while setting the attributes, but eventually, I managed to retrieve the data related to gaming. On the right side of the figure, I configured the operator by setting the parameter type to "include attribute" and selecting a single attribute with the name "gaming." Additionally, I specified that the attribute should contain only text values. Finally, I ran the operator to apply the attribute selection and obtain the desired data subset.

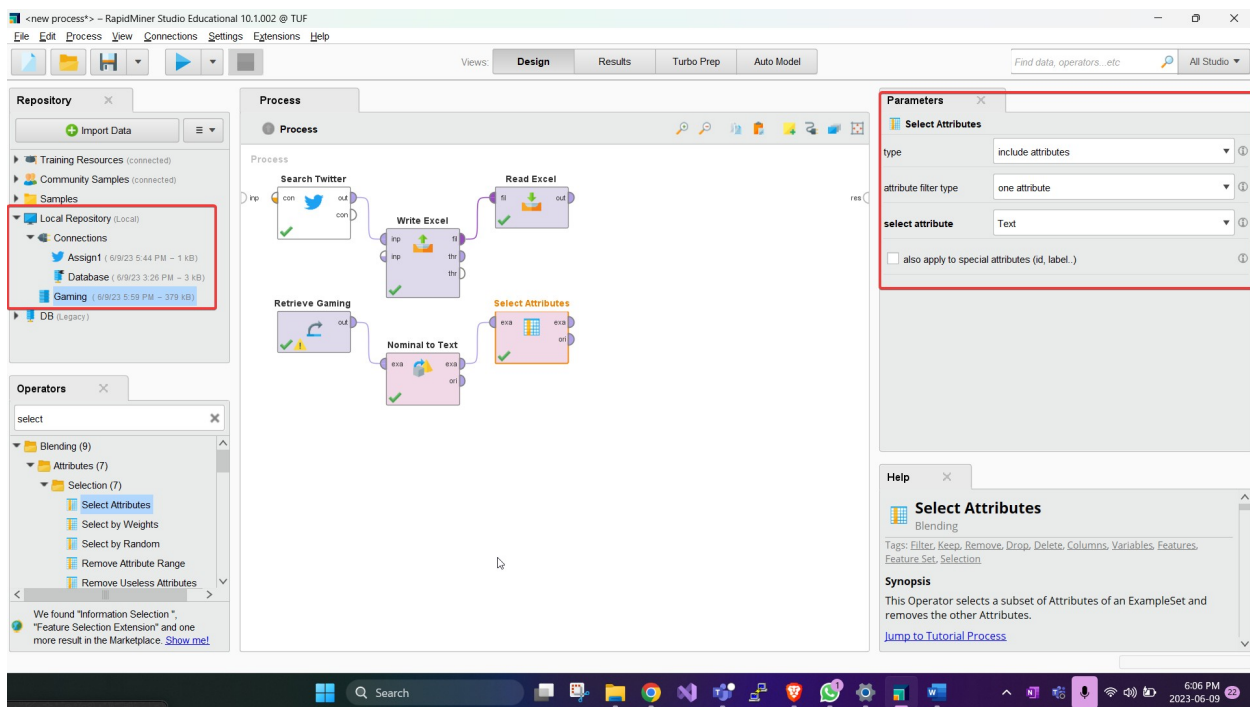


Figure 8 Attribute Selection for Extracting Relevant Gaming Data

Troubleshooting

In Figure 8, I encountered an issue while trying to execute the "Select Attribute" operator. As a troubleshooting step, I decided to retrieve the entire dataset again to address the problem. By doing so, I aimed to ensure that I have the complete dataset available and then proceed to resolve the issue with the "Select Attribute" operator separately.

In Figure 9, the result of the attribute selection process is displayed. It shows the data subset that was obtained by applying the "Select Attribute" operator. The selected attribute(s) are shown, indicating that the operator successfully extracted the desired data based on the specified attribute selection criteria.

ExampleSet (Select Attributes)

Filter (782 / 782 examples): all

Row No.	Text
1	How To Get 5 ...
2	ili Caide...
3	RT @ZLaner...
4	RT @Pikewic...
5	RT @CLXGa...
6	Unlock new ga...
7	RT @ROCC...
8	The rewards s...
9	RT @DemSh...
10	RT @magnu...
11	@GestaltSea...
12	RT @ceccoro...
13	RT @Parallel...
14	@Nutzypoo_t...
15	RT @EndWo...
16	RT @lastofca...
17	ive played Do...
18	RT @Seraph...

ExampleSet (782 examples, 0 special attributes, 1 regular attribute)

Repository

- Import Data
- Training Resources (connected)
- Community Samples (connected)
- Samples
- Local Repository (Local)
- Connections
 - Assign1 (6/9/23 5:44 PM - 1 KB)
 - Database (6/9/23 3:26 PM - 3 KB)
 - Gaming (6/9/23 5:59 PM - 379 KB)
- DB (Legacy)

Figure 9 Successful Data Extraction with Attribute Selection Operator

In Figure 10, the "Replace" operator is utilized to perform a specific text replacement. On the right side of the interface as shown in figure, I set the parameter according to my requirements. I replaced the specified pattern, which is represented by the regular expression `[-a-zA-Z0-9+&@#/%?~_|!.,:]*[-a-zA-Z0-9+&@#/%?~_|]`. The operator successfully applied this replacement as configured, resulting in the desired changes in the text data.

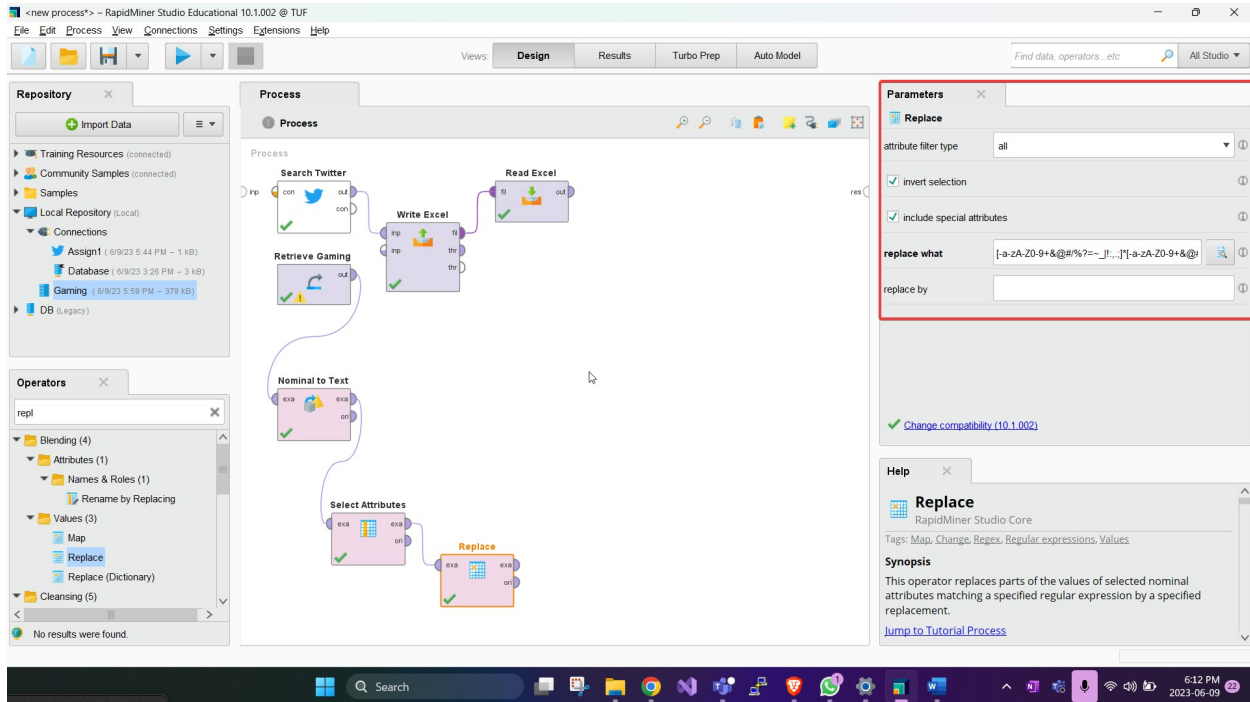


Figure 10 Successful Text Replacement with "Replace" Operator

In Figure 11, a second instance of the "Replace" operator is employed to perform another text replacement. On the right side of the interface, I configured the operator with specific parameters. The replacement pattern used was `RT\s*@[^:]*:\s*[A-Za-z]+`. By applying this

replacement pattern, the operator successfully made the desired changes in the text data, following the specified parameters.

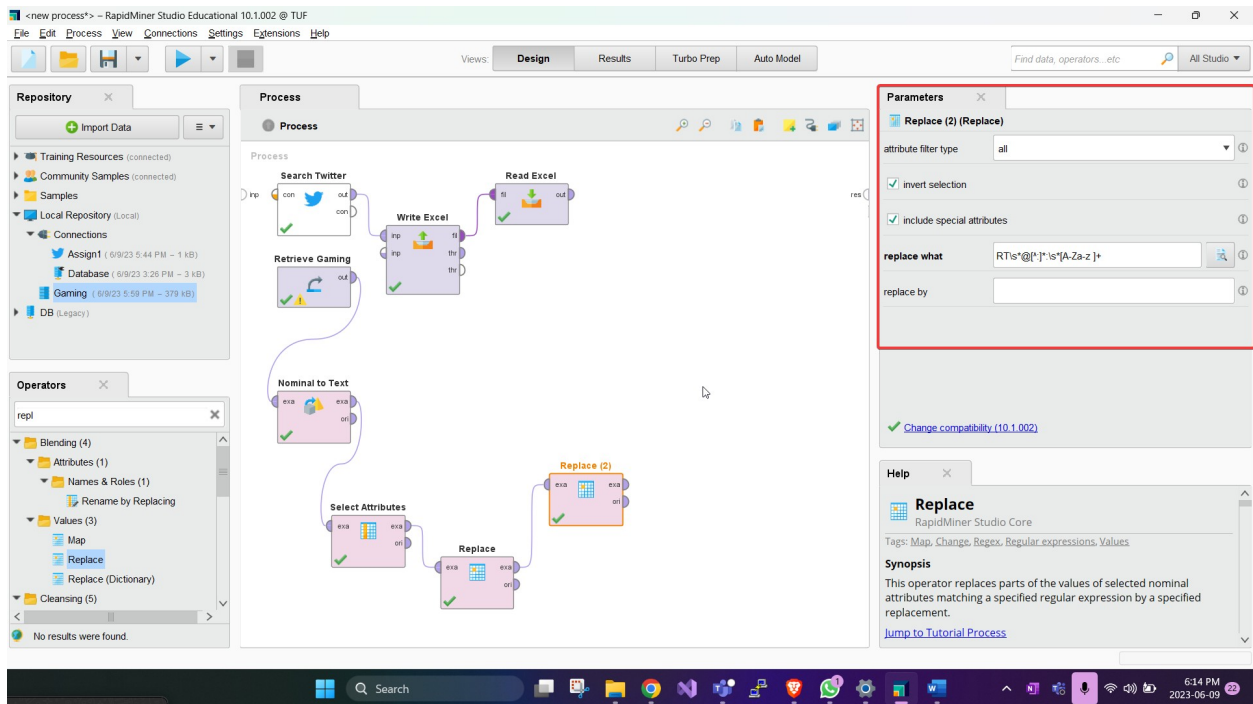


Figure 11 Successful Text Replacement with "Replace" Operator (Second Instance)

In Figure 12, the "Extract Sentiments" operator is applied to perform sentiment analysis on the text data. On the right side of the interface, I configured the operator by selecting the model vendor and specifying the text attribute. After running the operator, it generated the output, which represents the sentiment analysis results based on the chosen model.

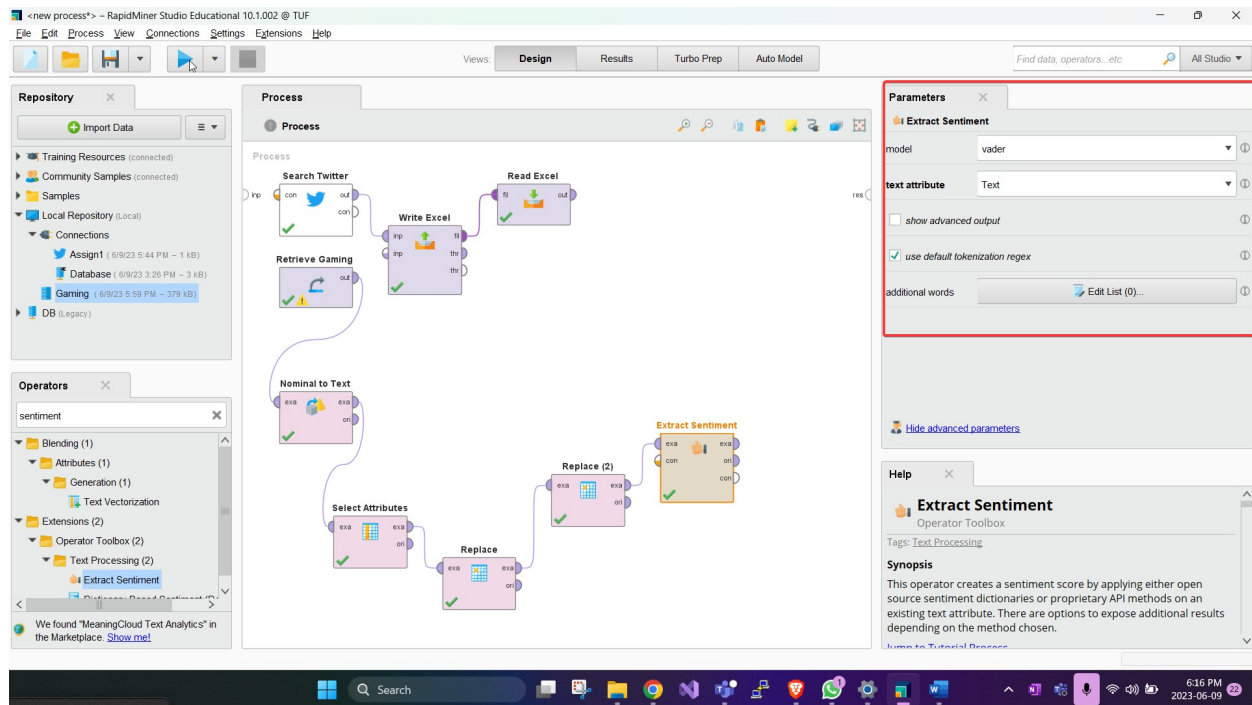
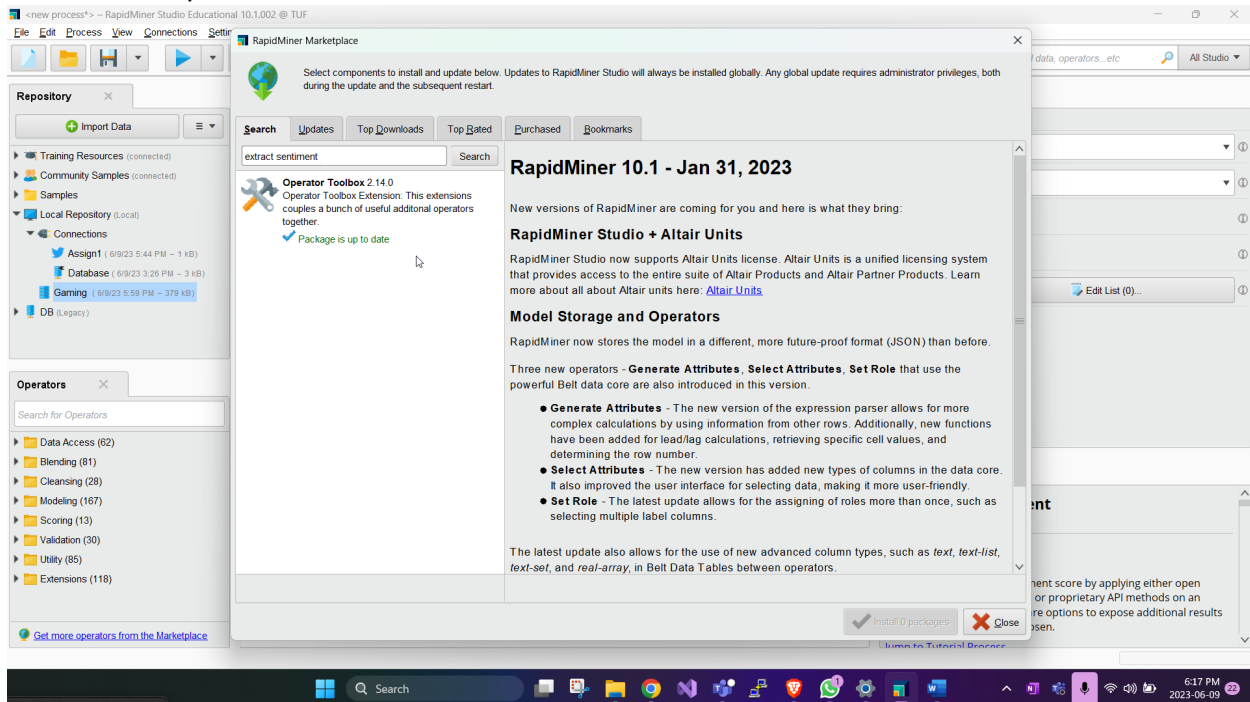


Figure 12 Sentiment Analysis using "Extract Sentiments" Operator

In the subsequent figure, I displayed the results of the sentiment analysis performed using the "Extract Sentiments" operator. To execute this operator, I had to install additional toolbox operators. The figure showcases the successful installation of these toolbox operators, which were required to run the

sentiment analysis and obtain the desired results.



In Figure 13, the "Generate Attribute" operator is utilized to create a new attribute based on certain conditions. The expression used in the operator is "if(Score > 0, "Positive", if (Score < 0, "Negative", if (Score == 0, "Neutral", "")))". This expression checks:

- If the "Score" is greater than 0, the new attribute will be labeled as "Positive".
- If the "Score" is less than 0, the new attribute will be labeled as "Negative".
- If the "Score" is equal to 0, the new attribute will be labeled as "Neutral".
- If none of the above conditions are met, the new attribute will be left empty ("").

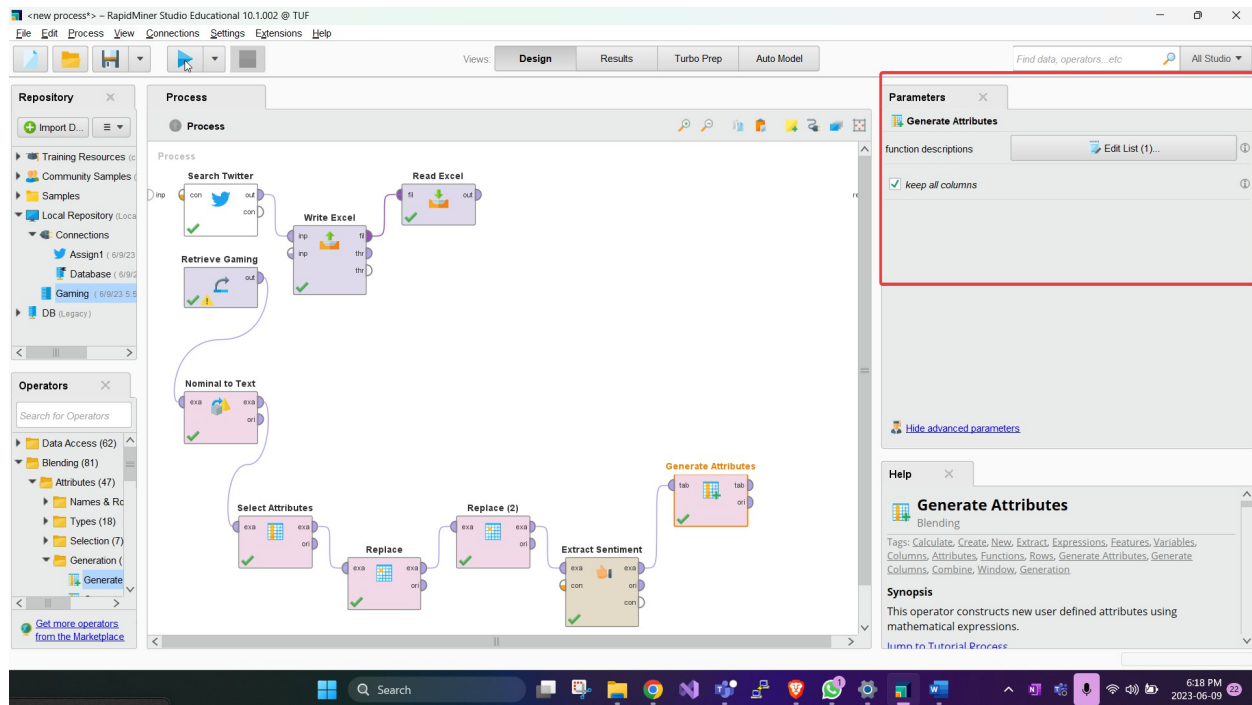


Figure 13 Conditional Attribute Generation with "Generate Attribute" Operator

Towards the end of this workflow, I employed the "Write Excel" operator to export the final result to an Excel file. In Figure 14, I specified the desired path for the Excel file where the final result would be stored. By running the operator, the data was successfully written to the specified location, providing the final output in Excel format.

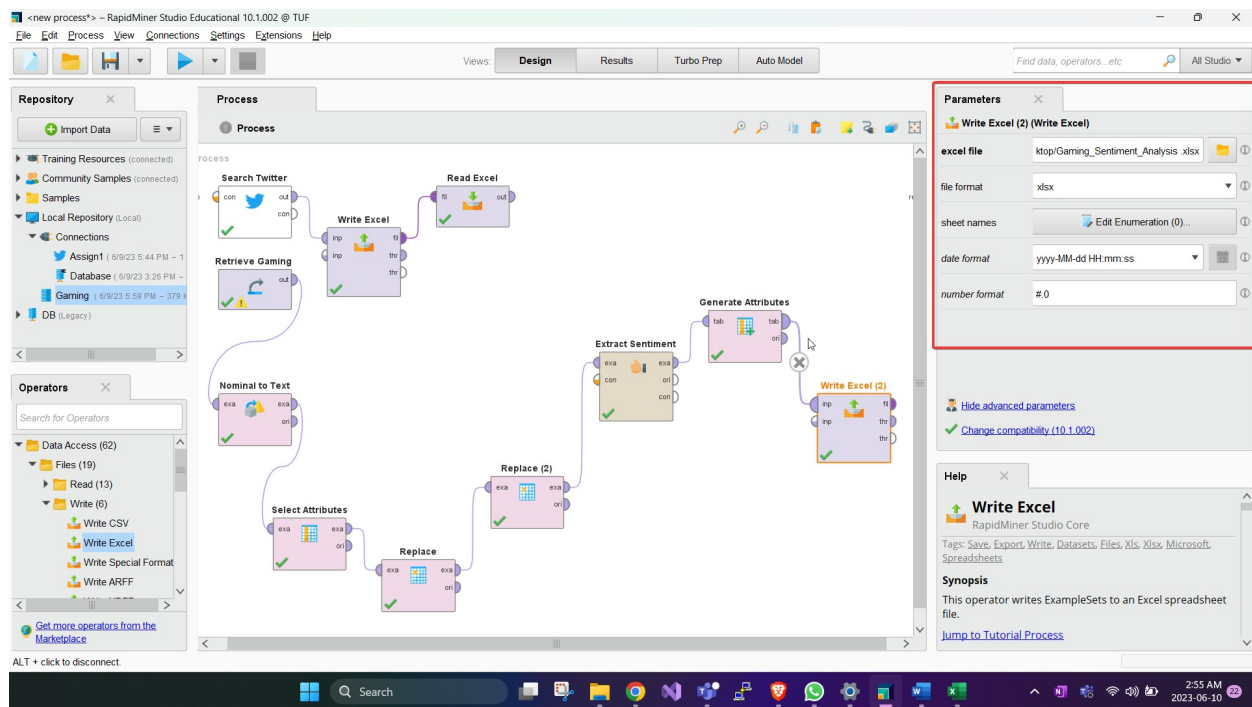


Figure 14 Title: Exporting Final Results to Excel with "Write Excel" Operator

- **Elaboration on why you choose your topic, and what's your findings?**

I chose the topic of gaming for sentiment analysis due to my lifelong interest in the gaming industry. I wanted to gain insights into public sentiments and opinions related to gaming. By collecting tweets using the "Twitter Search" operator in RapidMiner, I gathered a dataset for analysis. The findings allowed me to understand the overall sentiment towards gaming and identify trends within the gaming community. The analysis provided valuable insights into public perceptions and helped me to get deep knowledge of the gaming industry.

- **Elaboration on your feedback of the exercise – is it easy to do? And what have you learned?**

The exercise I completed involved setting up a local repository connection in RapidMiner Studio, using the "Twitter Search" operator to gather tweets, configuring operators for data extraction and manipulation, performing sentiment analysis, generating new attributes, and exporting the results to an Excel file. Through this exercise, I gained practical experience in utilizing RapidMiner Studio's features for data analysis and learned about configuring operators, handling data, and performing sentiment analysis. While the exercise overall was relatively easy, I did encounter some challenges along the way. Despite following all the steps, I struggled with extracting sentiments and choosing the appropriate attribute. This aspect proved to be more difficult for me

to understand. However, I managed to work through the problems and successfully run the operators. The experience with RapidMiner Studio, being a new tool for me, has piqued my interest, and I would like to delve deeper into learning more about its functionalities.