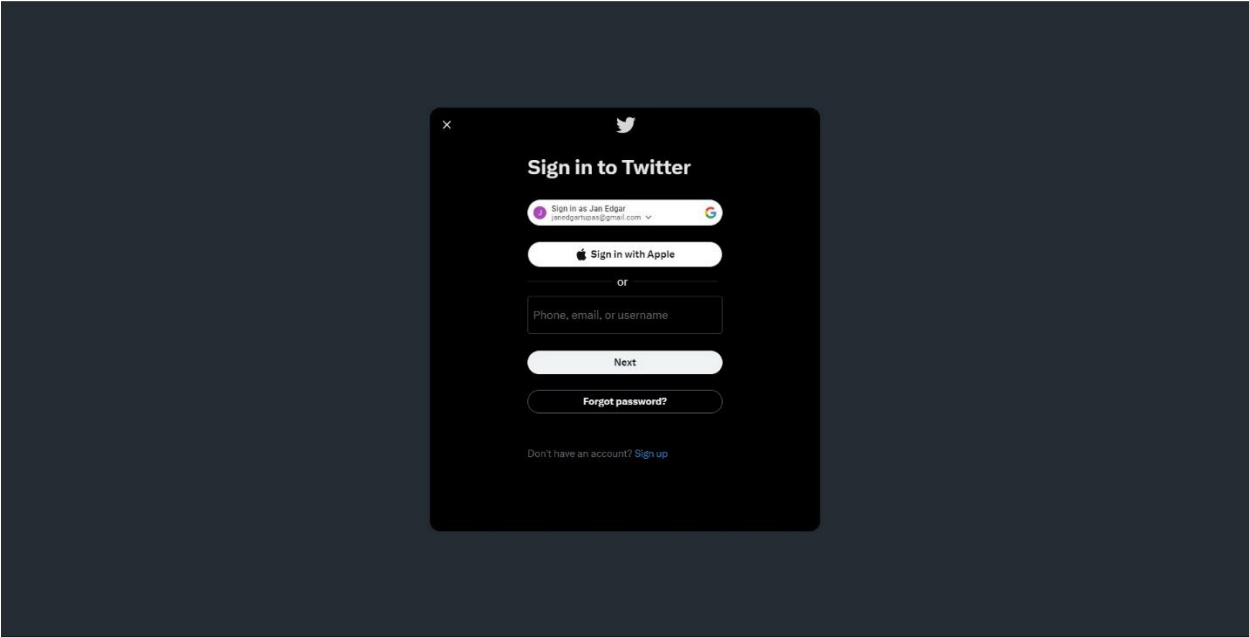# Walkthrough of account preparation for data gathering from Twitter.

1. Account Creation

Navigate to the Twitter website from your browser. You may access it through this link: https://twitter.com/ Create a new twitter account by clicking on **Sign up** or log in to an existing account.
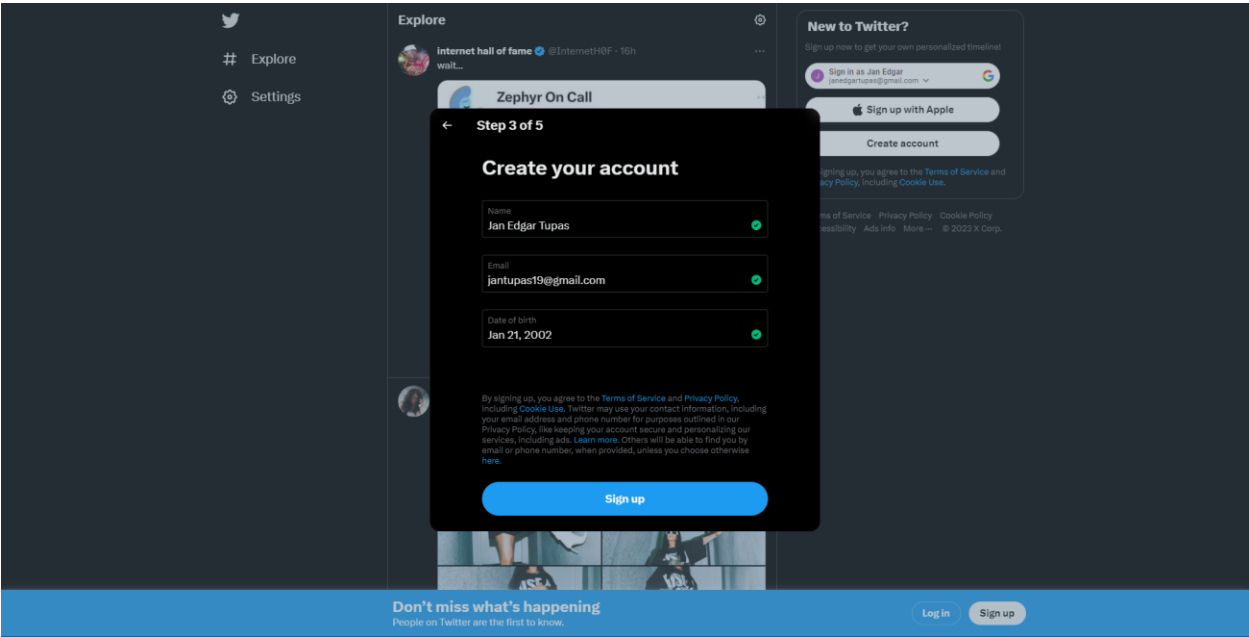

In creating a new twitter account, simply follow the steps shown in the prompts. Enter your name, email address, and date of birth. Afterwards, click **Next.**
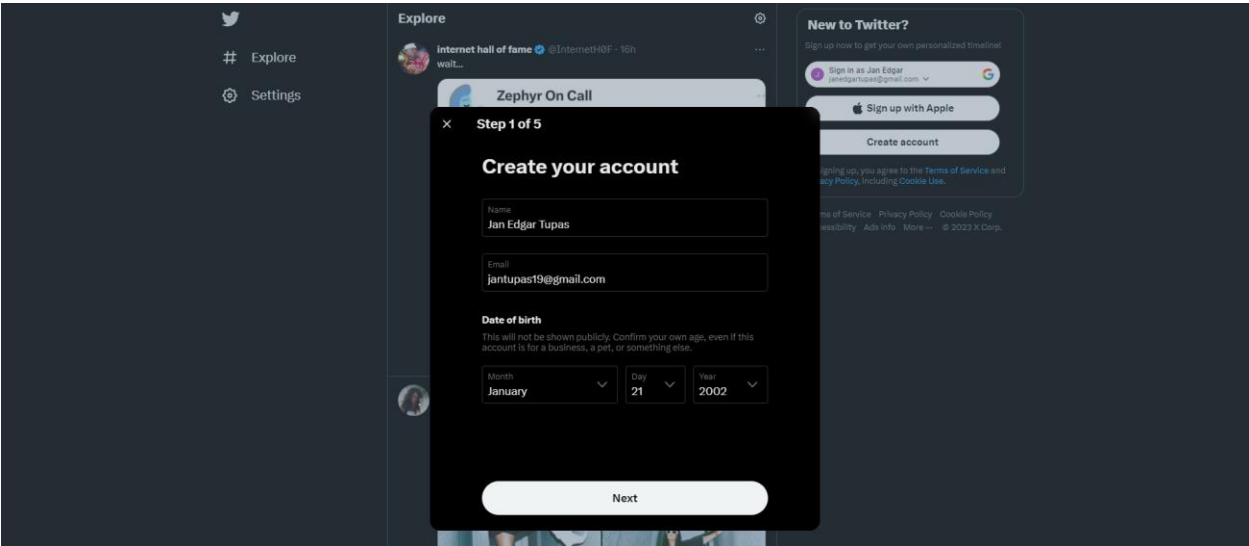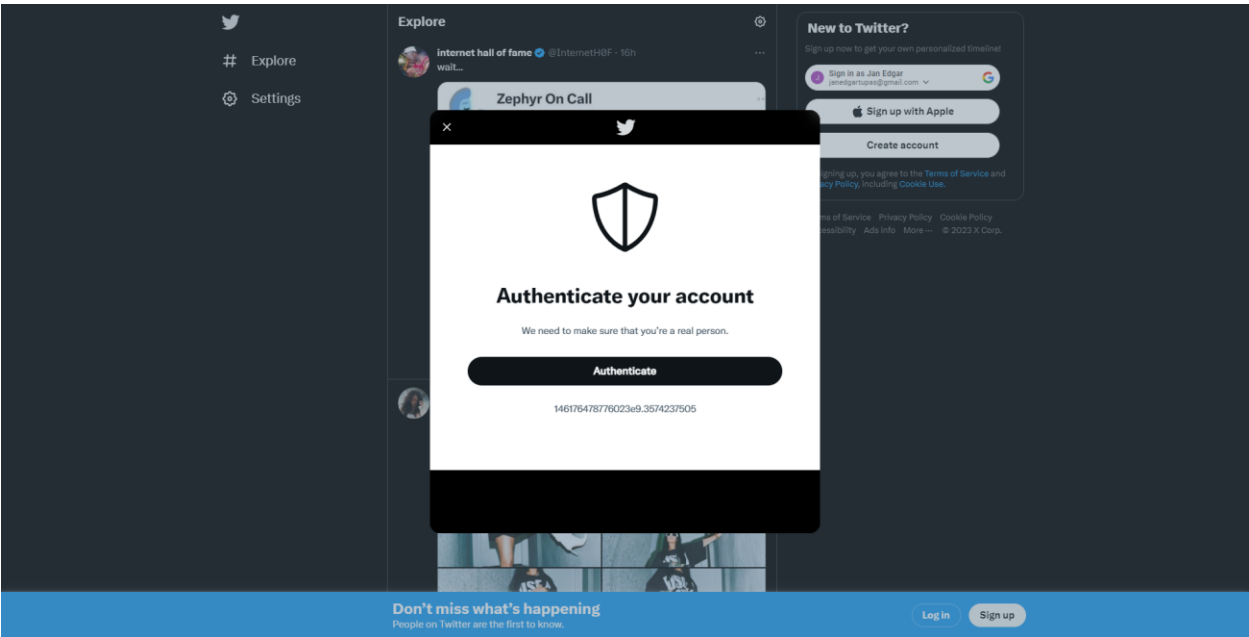

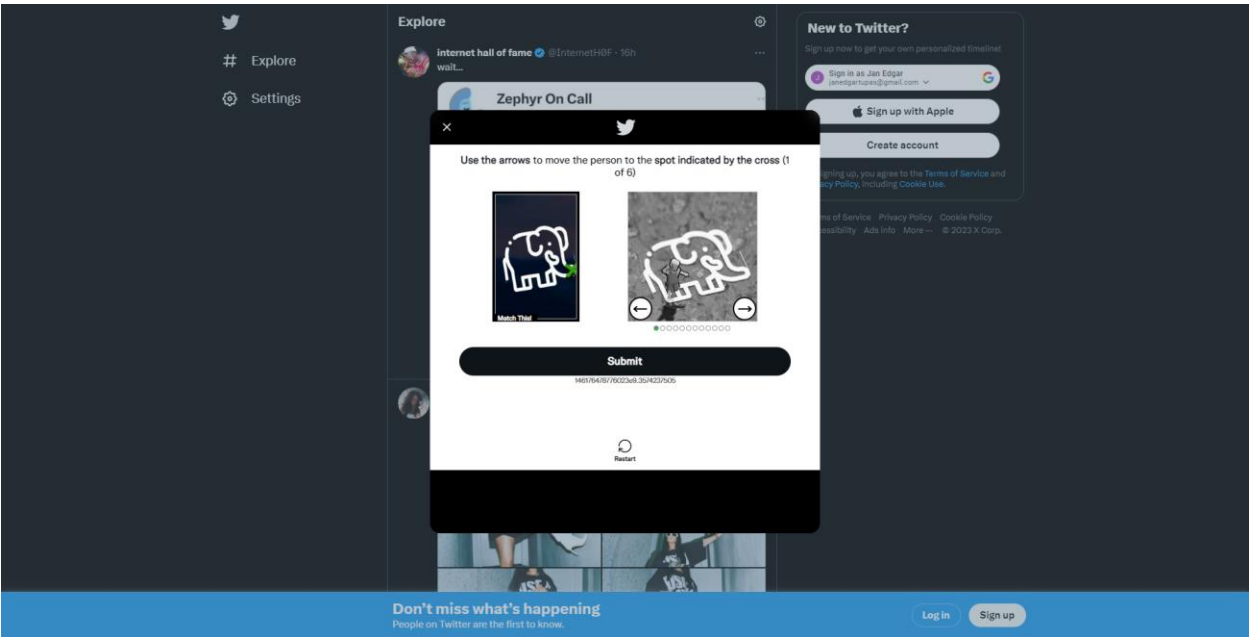You may or may not allow customized experience. Afterwards, click **Next.**
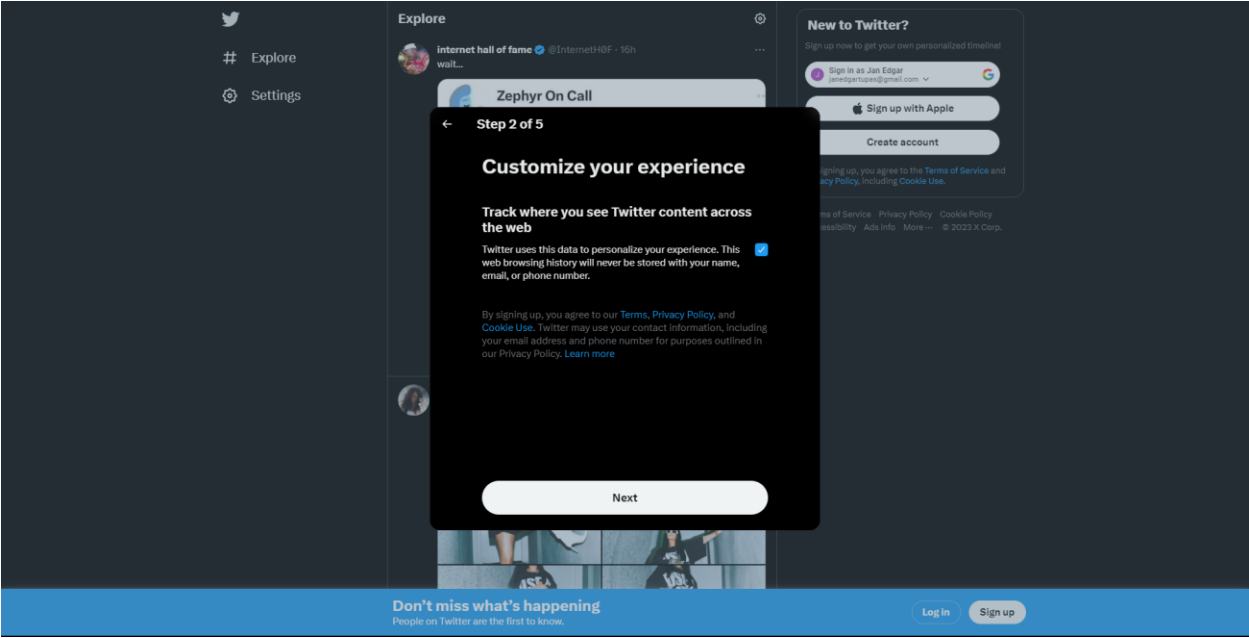
Verify that all information is correct. Click **Sign up.**
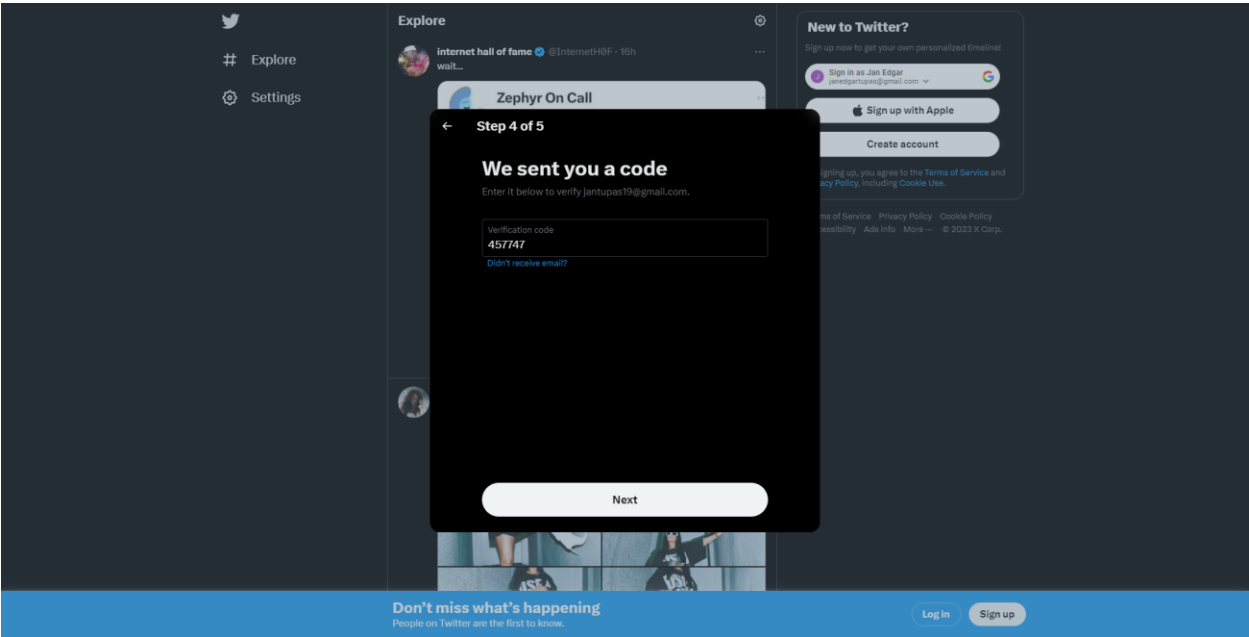

You will be asked to follow the authentication process to verify you are a real person. Click **Authenticate**
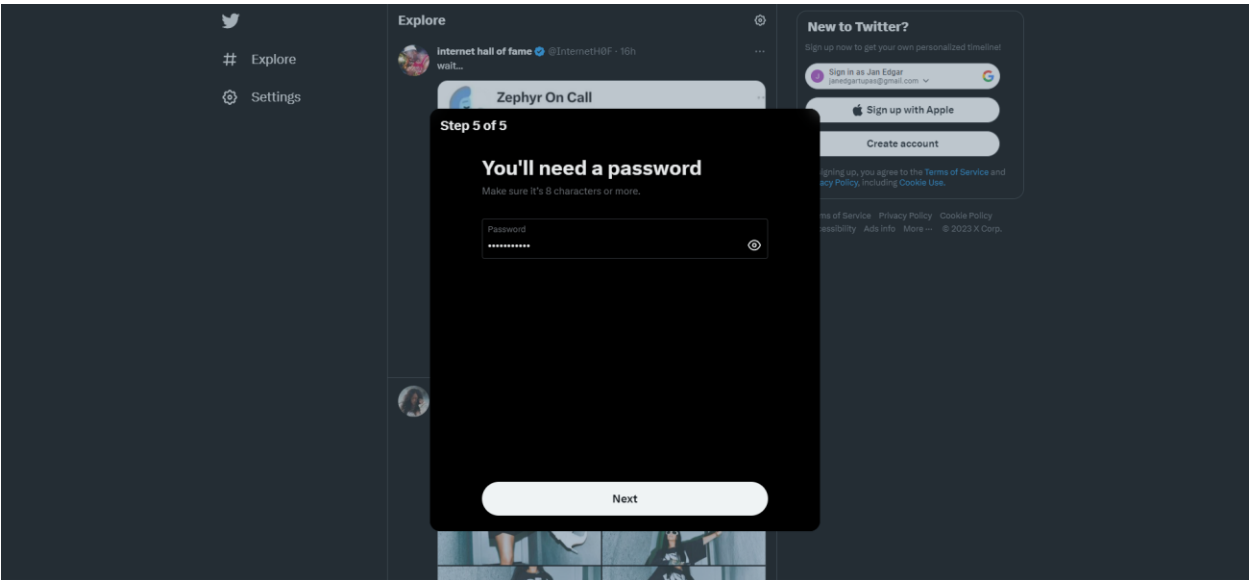

Simply follow the authentication procedures. Click **Submit.**

You will be sent a code to the email you entered. Paste the code in the text box and click **Next.**


Enter your password. This will be the password of your account which you will use for logging in. Click **Next.**

You have successfully created a new Twitter account!

2. Apply for Twitter Developer Account

After creating your account and logging in, navigate to the Twitter Developer Portal website. You may access it through this link: https://developer.twitter.com/en/portal/dashboard The page below will show. Click on **Sign up for Free Account** located at the bottom part.
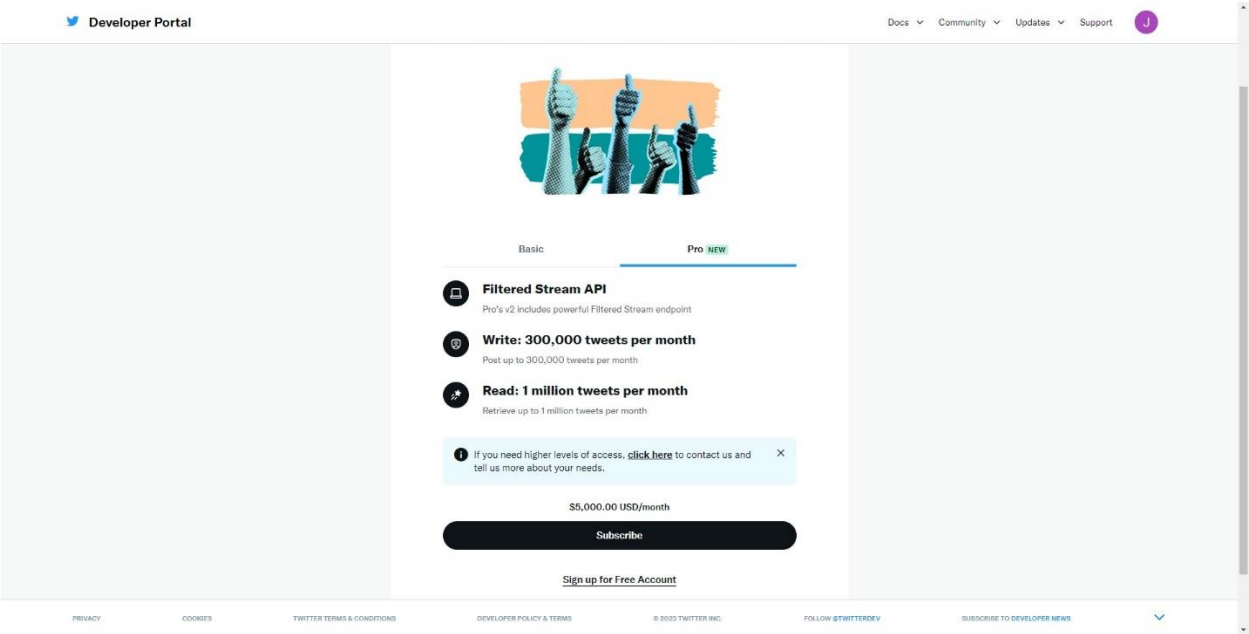
The next page shows the Developer agreement & policy to apply for a developer account. Describe your reasons for applying for a developer account and what you will use Twitter's data and API for. You can state use cases such as for "data science projects" or "school projects."

Click on all checkboxes and click **Submit.** Once you are granted the account, you will be presented with your dashboard where you can create projects and apps.

You have successfully acquired a Twitter Developer account!

3. Downloading and Installing RapidMiner Studio.

Navigate to the RapidMiner website. You may access it through this link: https://rapidminer.com/platform/ Click on **Account.**
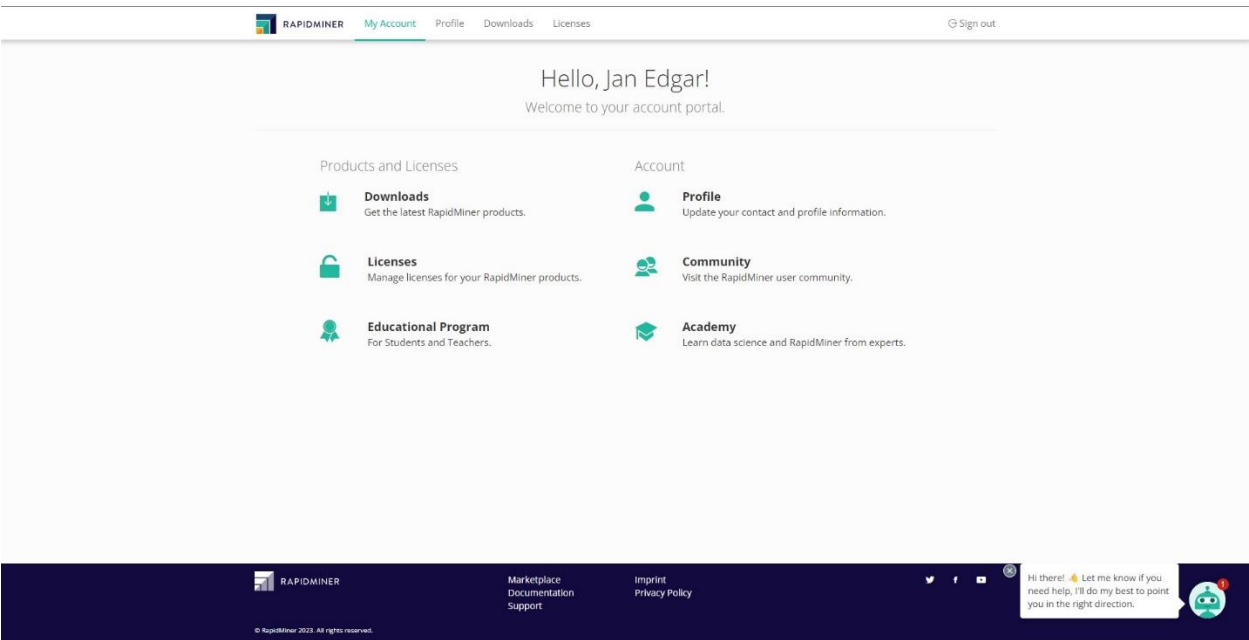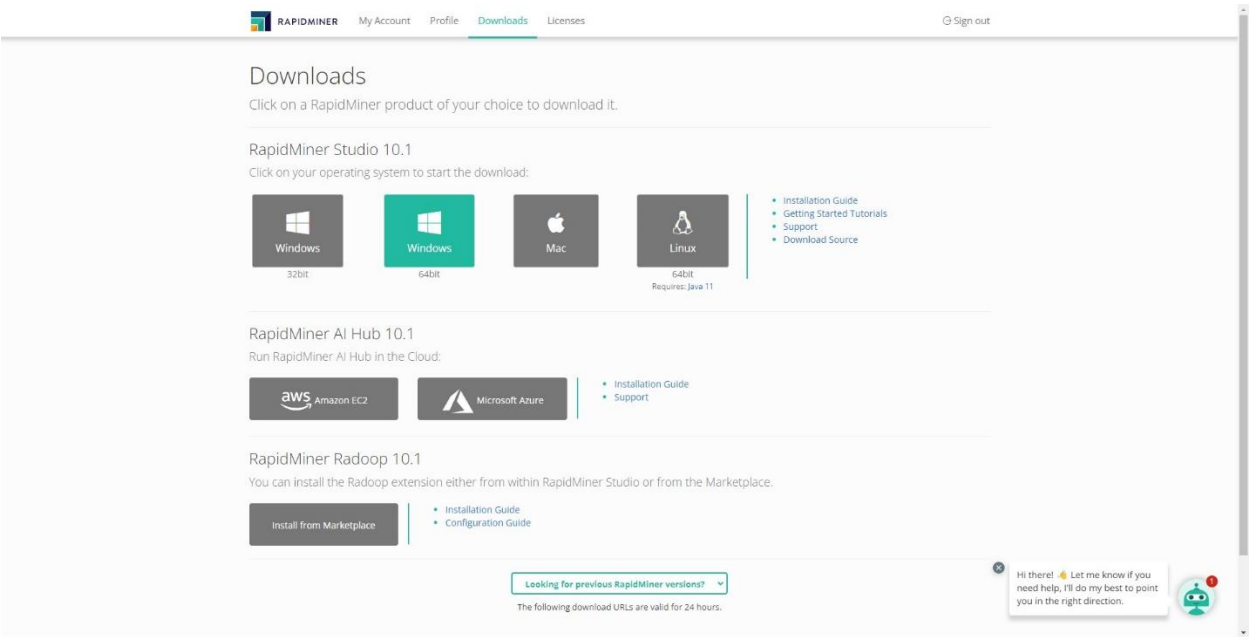
If you already have an existing account, you may simply log in. Otherwise, register a new account and sign in.
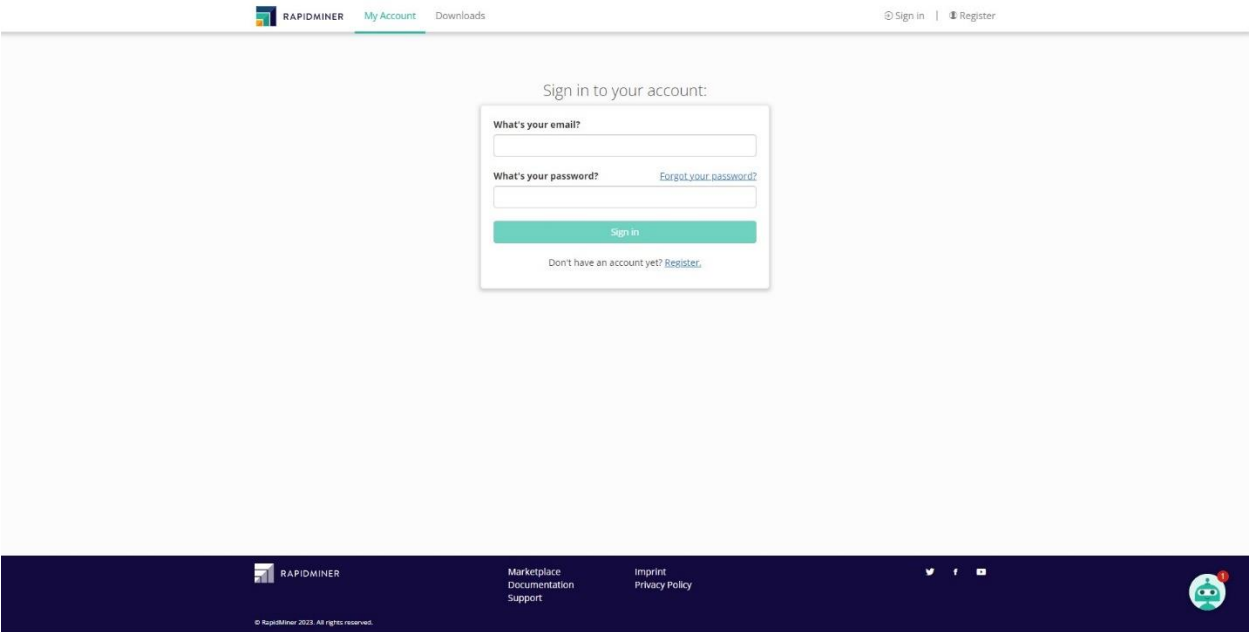
After logging in to your account, you will be presented with your account portal as shown below. Click on **Downloads.**
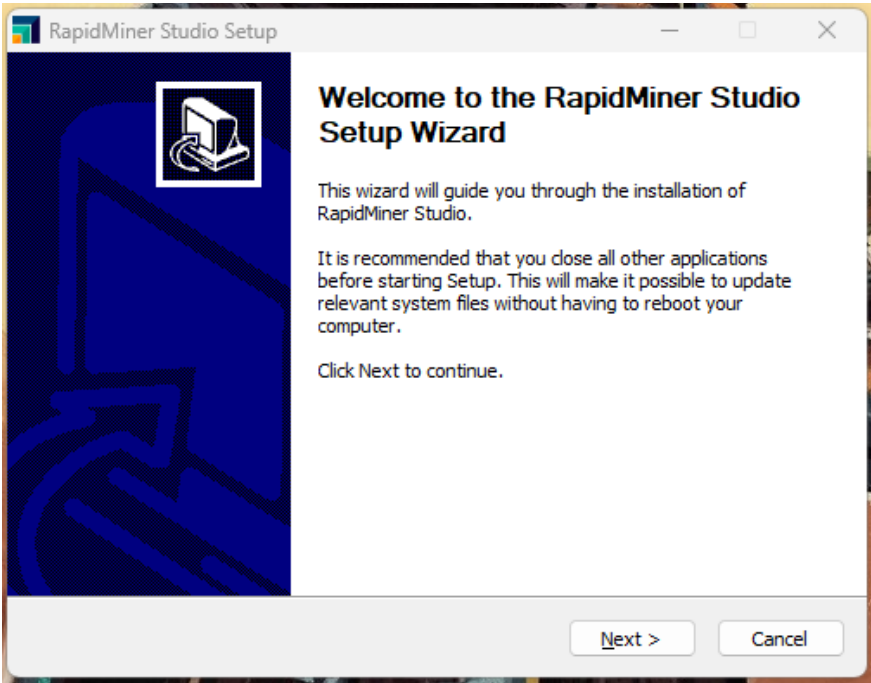

Select the appropriate operating system of your computer. This will then begin the download for RapidMiner Studio installer.
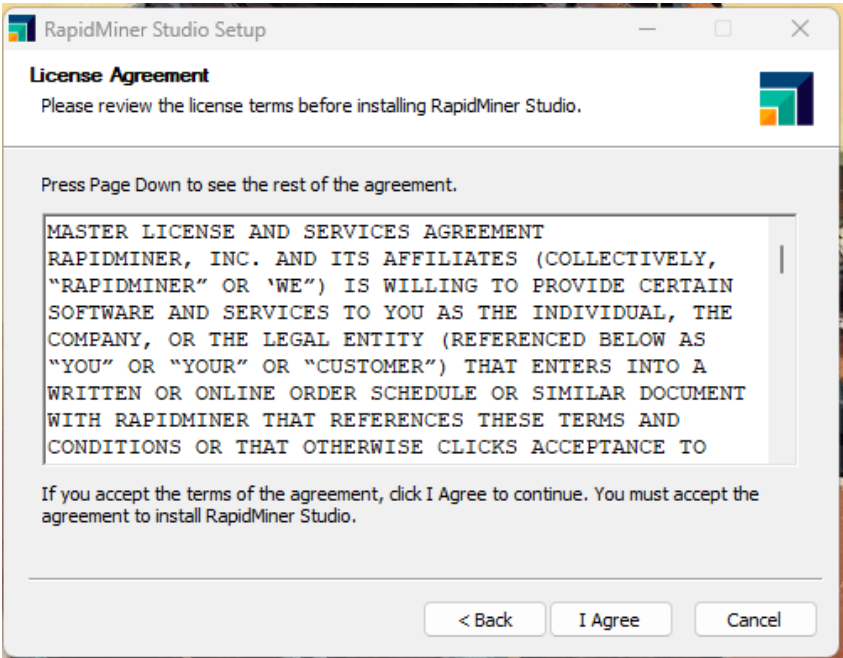
After the download has completed, run the installer. Simply follow the steps in the setup to complete the installation.
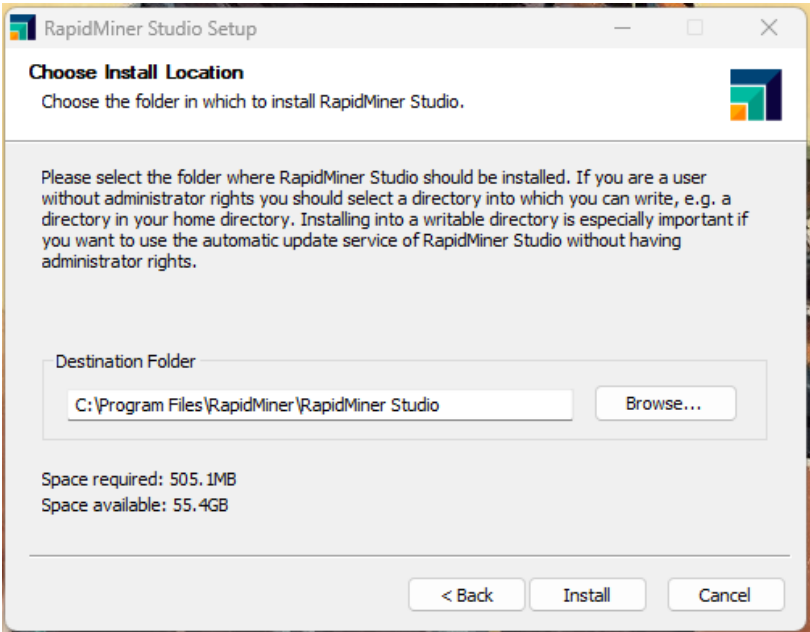
You will be prompted with the **RapidMiner Studio Setup**. Click **Next.**
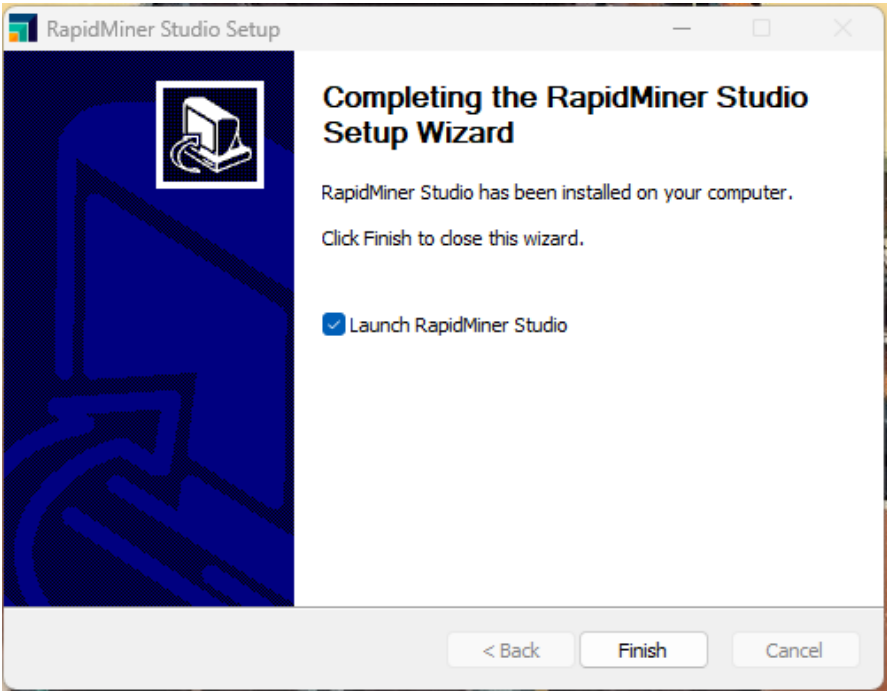

You will be prompted with the **License Agreement**. Click **I Agree.**

Specify the location of RapidMiner Studio's installation. Afterwards, Click **Install.**

After finishing the installation check on **Launch RapidMiner Studio.** Then click **Finish.**

You have successfully installed RapidMiner Studio!

## 4. Connecting RapidMiner to Twitter for Data Gathering.

If it is your first time running RapidMiner Studio, you will be asked to login in to your account and install your licenses. Simply enter your login credentials and click **Login and Install.**
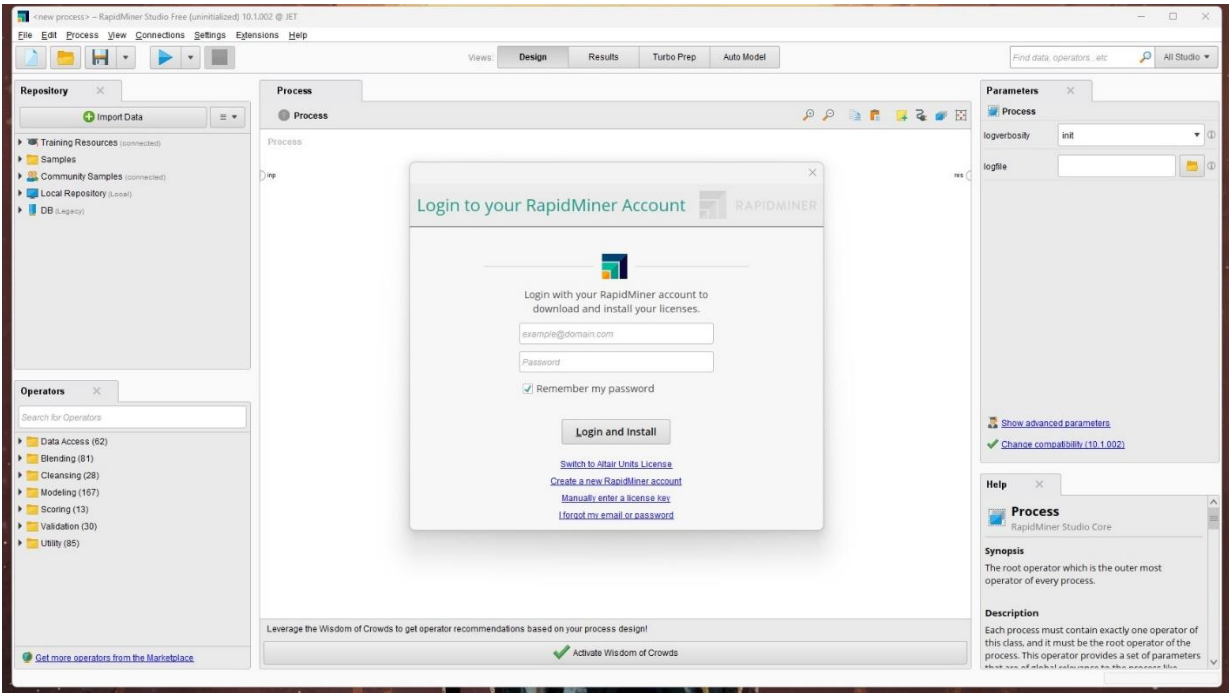
After logging in, you will be presented with the RapidMiner Studio. Navigate to the **Repository** tab located at the left above the **Operations** tabs. Find and right click on **Local Repository**. Click **Create Connection.**
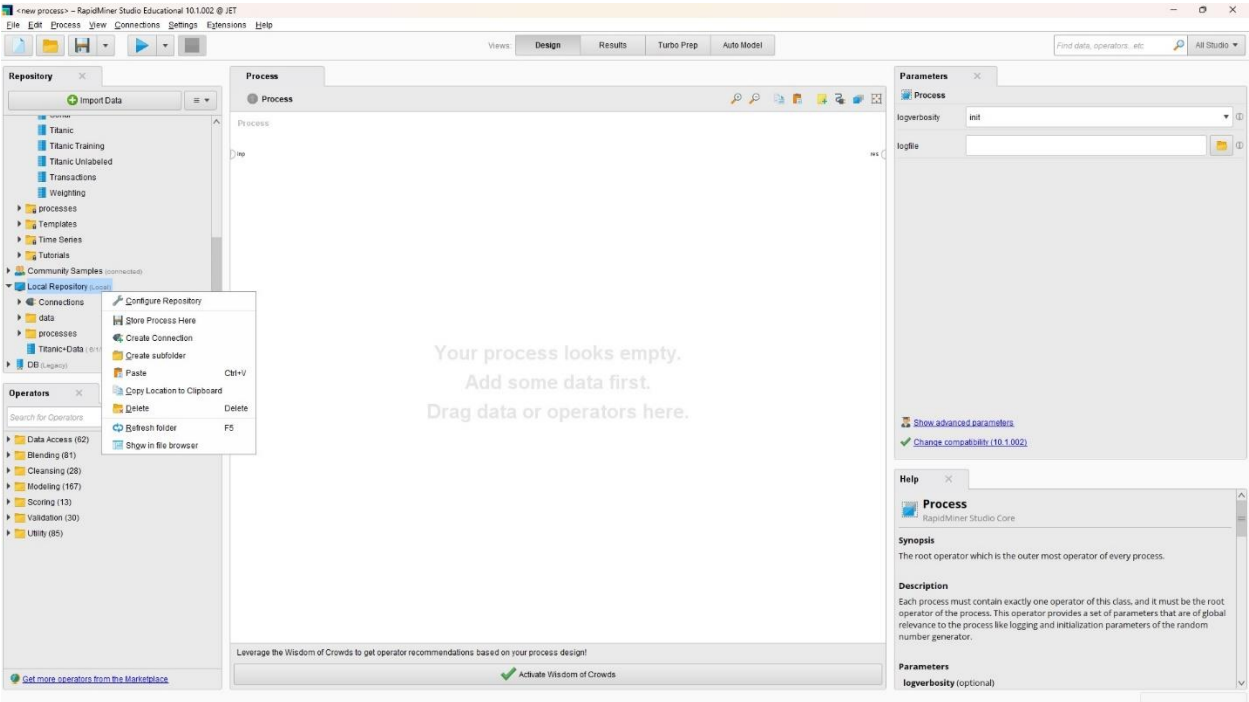

You will be prompted with the **Create a new connection** tab. Set **Connection Type** to **Twitter**. Check that the selected **Repository** is correct and set a **Connection Name**. Afterwards, click **Create**.
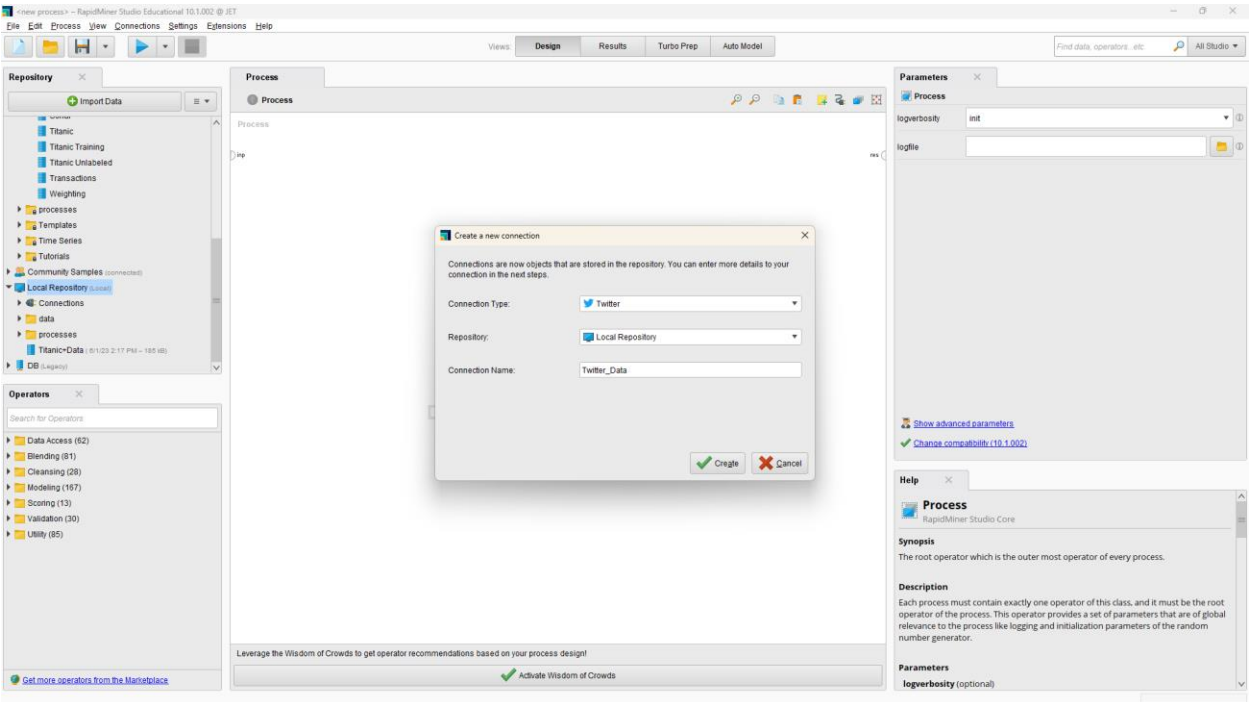
You will then be prompted with the **Edit connection** tab. Click on the **Setup** tab. We then need to request an access token. To do so, click on the 🔳 button.
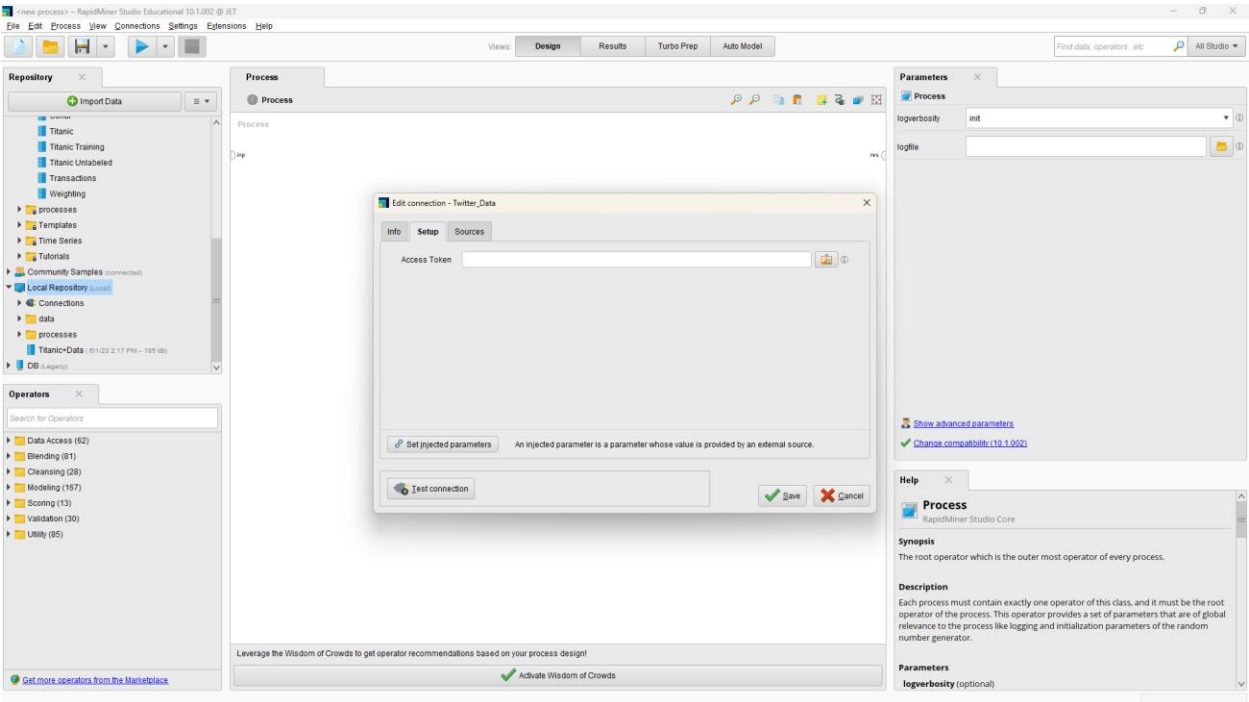
You will then be prompted with the **OAuth** tab. Click on the **Request access token** under **Open URL**. This will open the Twitter website in your browser and ask you to login and authorize RapidMiner Social Media Extension to access your account. Click on **Authorize app**.
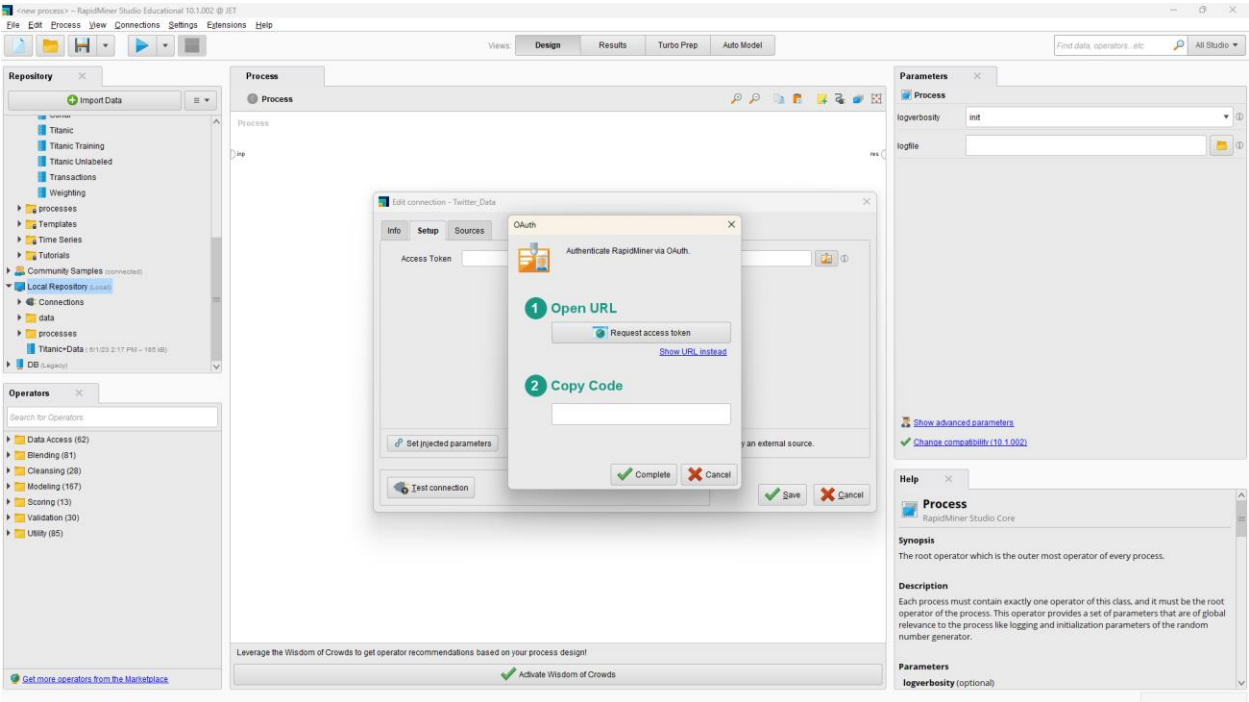
A page will open and show the access token. Copy the access token and navigate back to RapidMiner Studio.

Paste the access token under **Copy Code**. Afterwards, click on **Complete**.
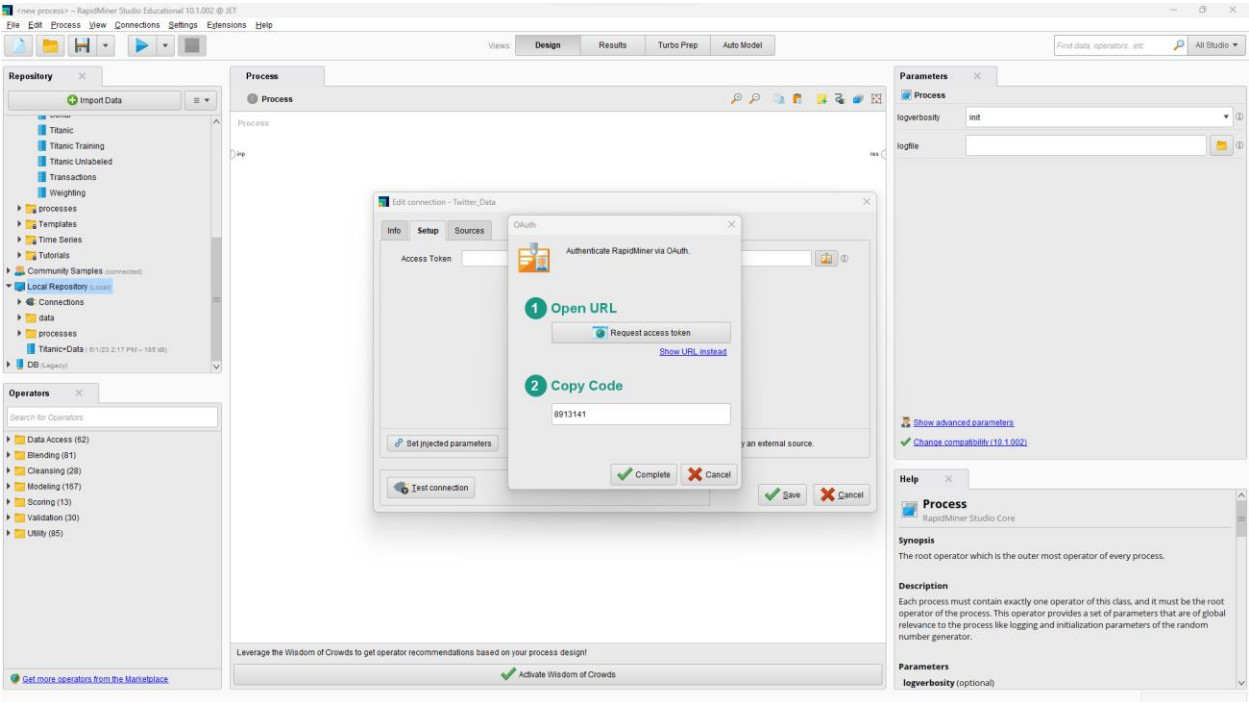
To ensure that the connection was successful, click on **Test connection**. A test successful message should be shown to verify successful connection. Click on **Save** to save the connection and finish the connection process. In case of a failed test, ensure that the access token you copied was correct. Otherwise, repeat the entire process.
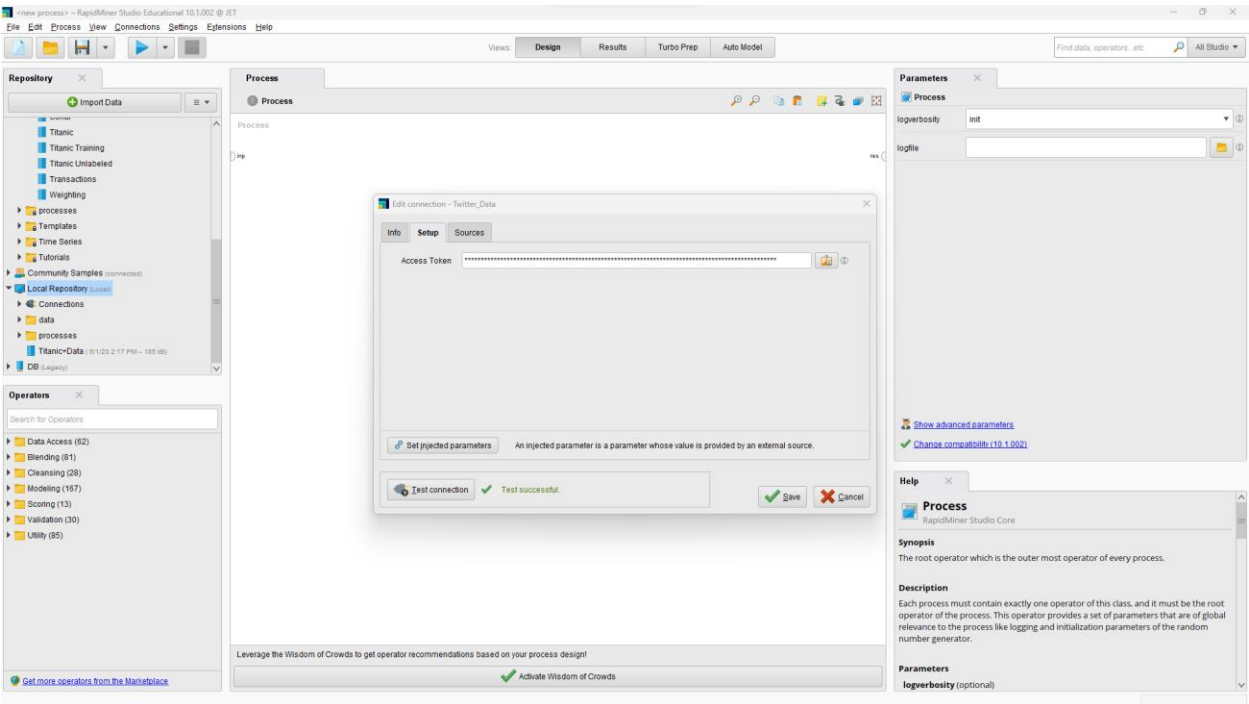

You have now successfully connected Twitter to RapidMiner and can now begin Data Gathering!

Jan Edgar Tupas CS171/BM1

5. Download a sample data set from Twitter using RapidMiner

In this data set, we retrieve tweets from a specific Twitter user. From **Connection**, drag the twitter connection you created earlier to the process area in the middle of the studio.
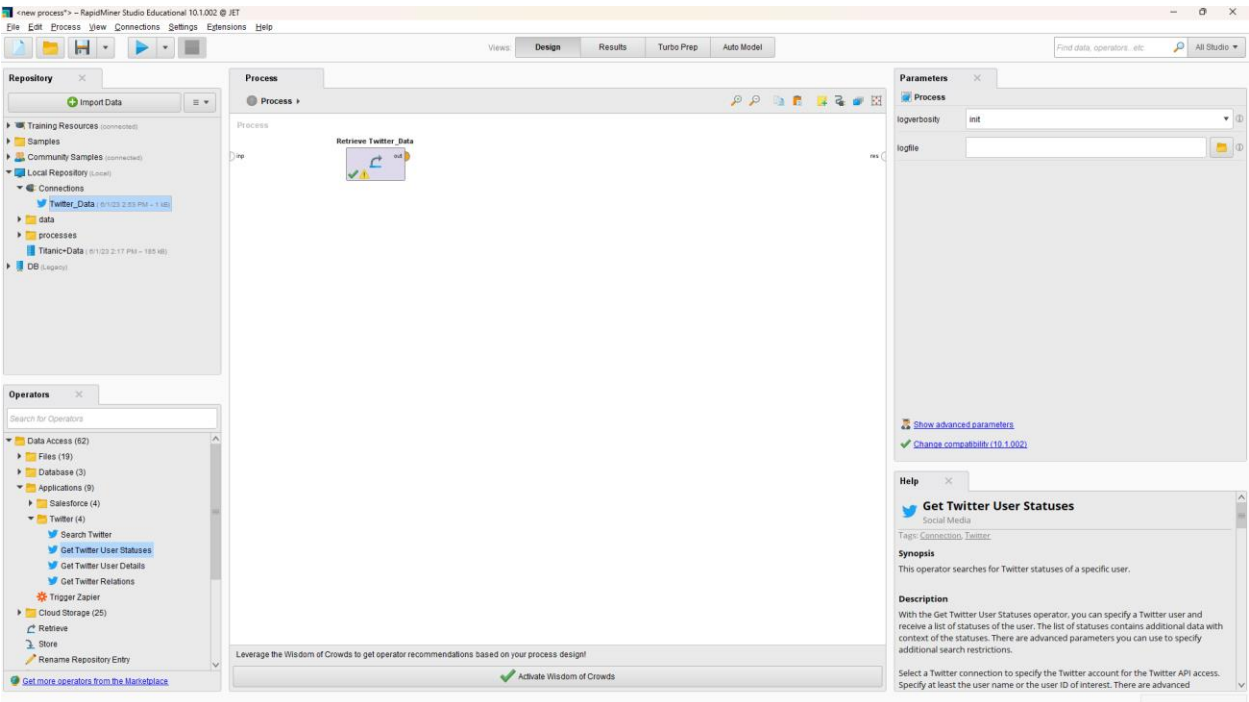
Navigate to the Operators tab located at the left and below the Repository tab. Search for the **Get Twitter User Statuses** operator in the search box. Drag the operator to the process area in the middle of the studio.
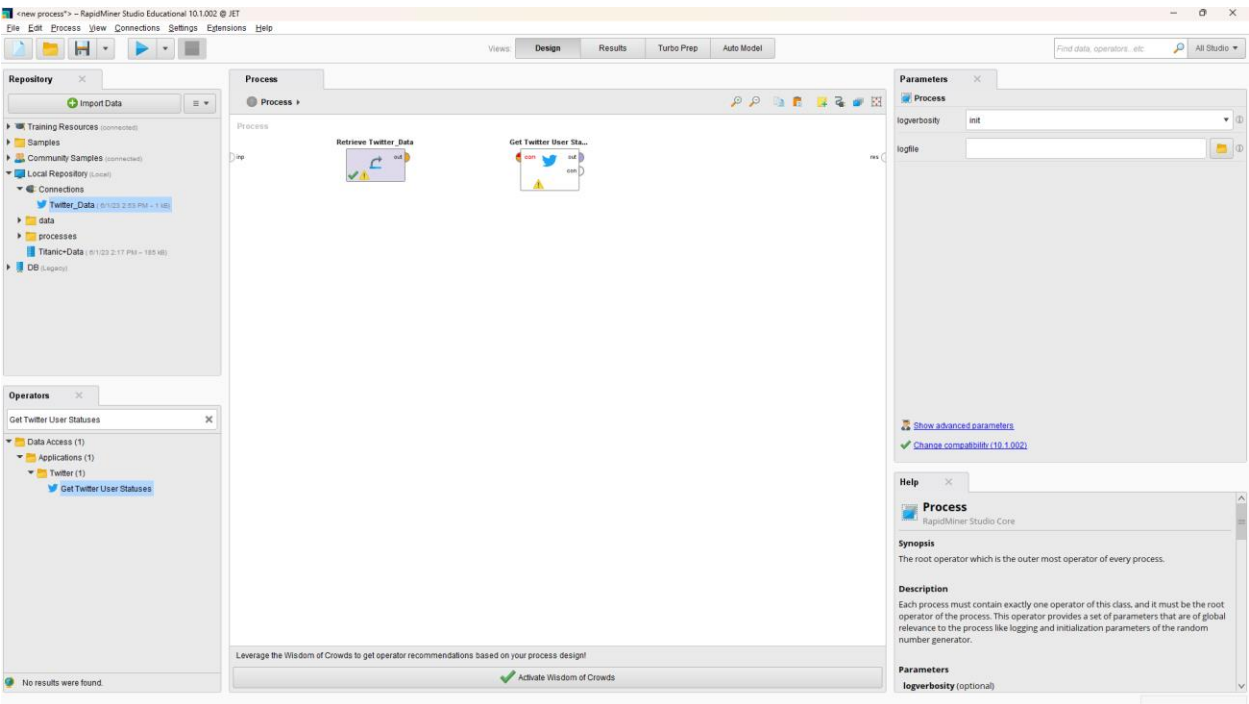
Navigate to the **Parameters** tab on the right. Click on the folder icon along the connection entry line.
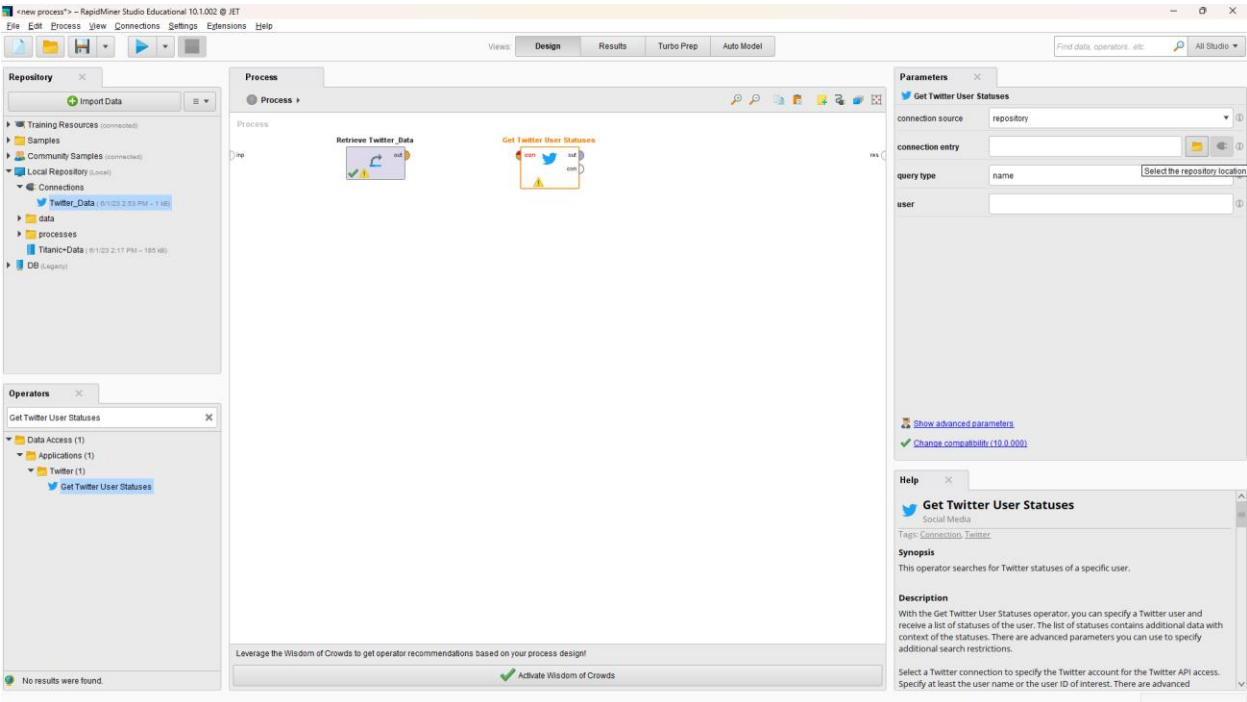
You will be prompted with the **Repository Browser** tab. Select the Twitter connection that was created earlier and click the **OK**.
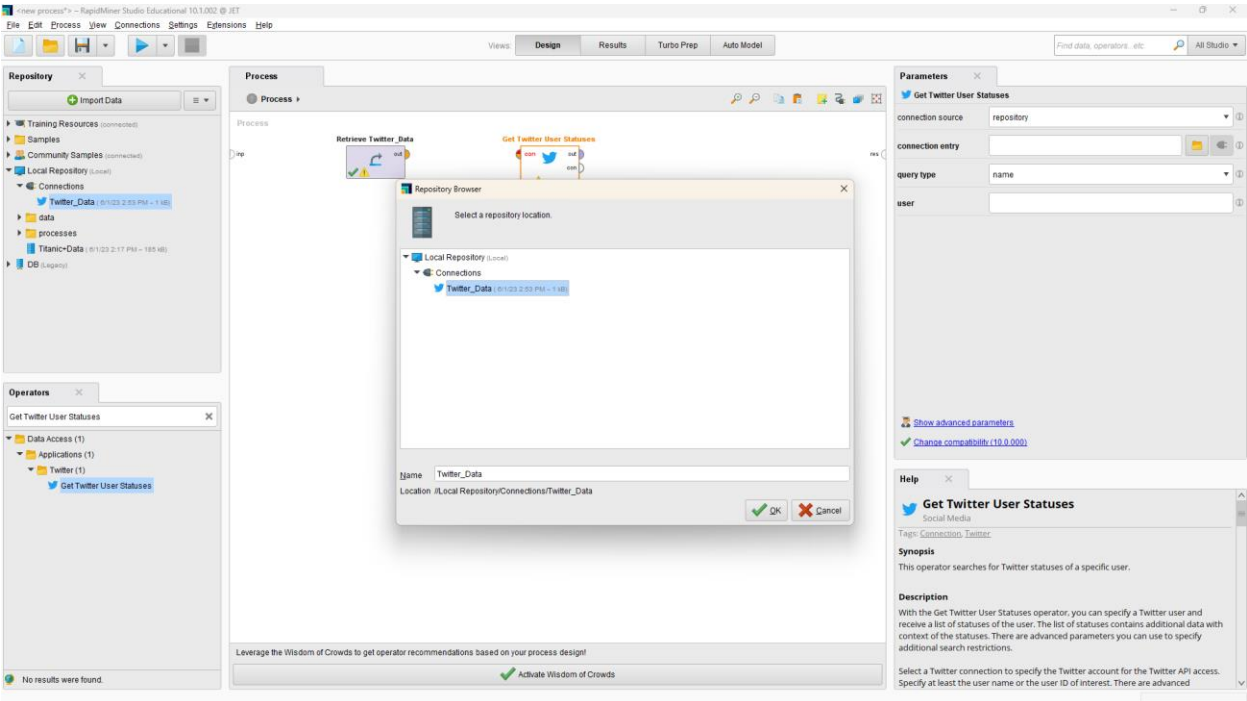
In the **Parameters** tab beside **user**, enter the username of the user you want to retrieve tweets from. You can also change **query type** to id and enter the id of the user you want to retrieve tweets from. Click on **Show advanced parameters.** You can then specify the **limit** of the number of tweets to be shown in the results, and the **since ID** and **max ID** to specify the period of time of the tweets.
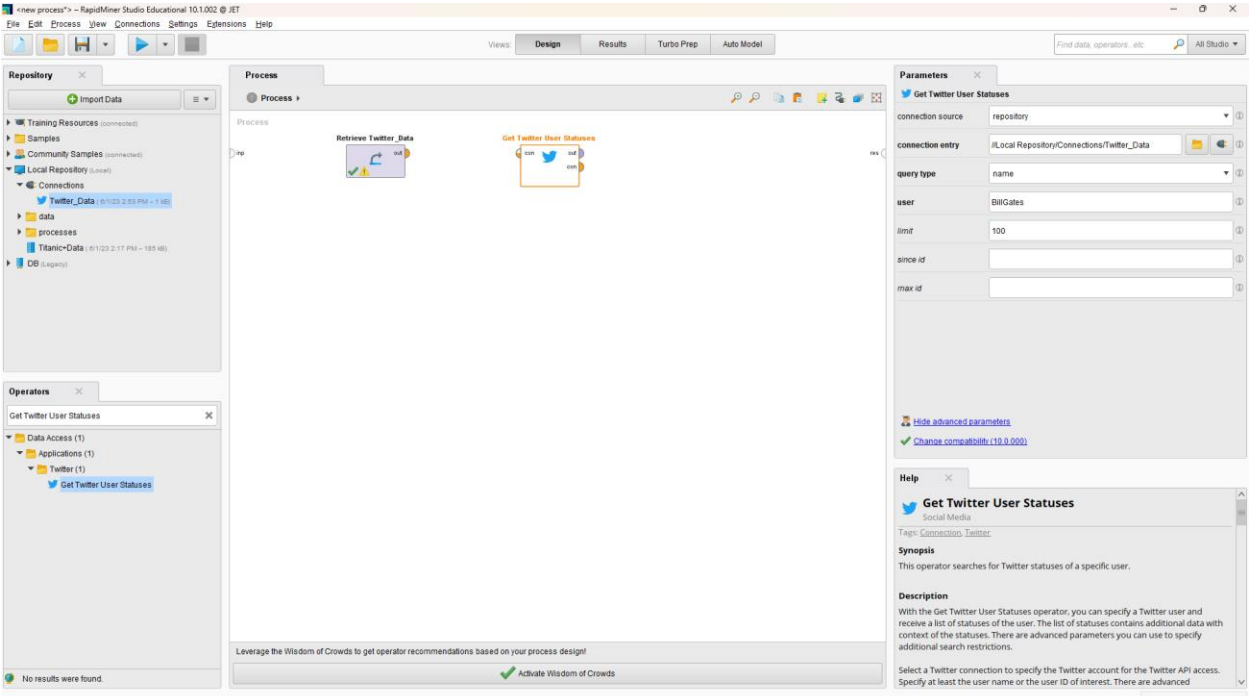
Navigate to the process area located in the middle of the studio. Click **Retrieve Twitter_Data.output** and connect it to the **Get Twitter User Statuses.connection.** click the **Get Twitter User Statuses.output** and connect it to the **process.result**. Click on the run ▶ button.
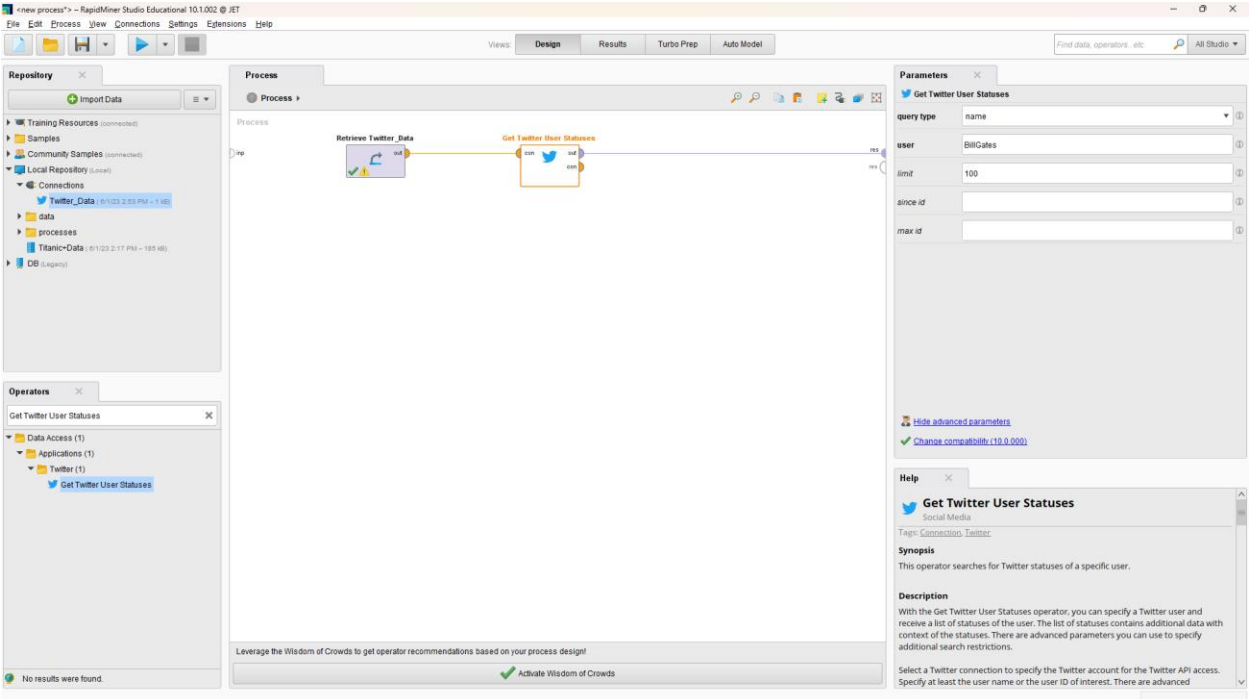
The results are shown where the last 100 tweets of Bill Gates are returned. The data set includes ID, date and time the tweet was created, the user that created the tweet and their ID, the user that the tweet is sent to and their ID, the language of the tweet, the source link of the tweet, the geo-location of the tweet with regards to latitude and longitude, and the number of retweets.

To export the results, click on **Turbo Prep** located above the data set table. Click on the three dots **...** at the right side beside **HISTORY** and click **EXPORT**.

Jan Edgar Tupas CS171/BM1

You can choose between formats in exporting the data set. Click **CSV.**
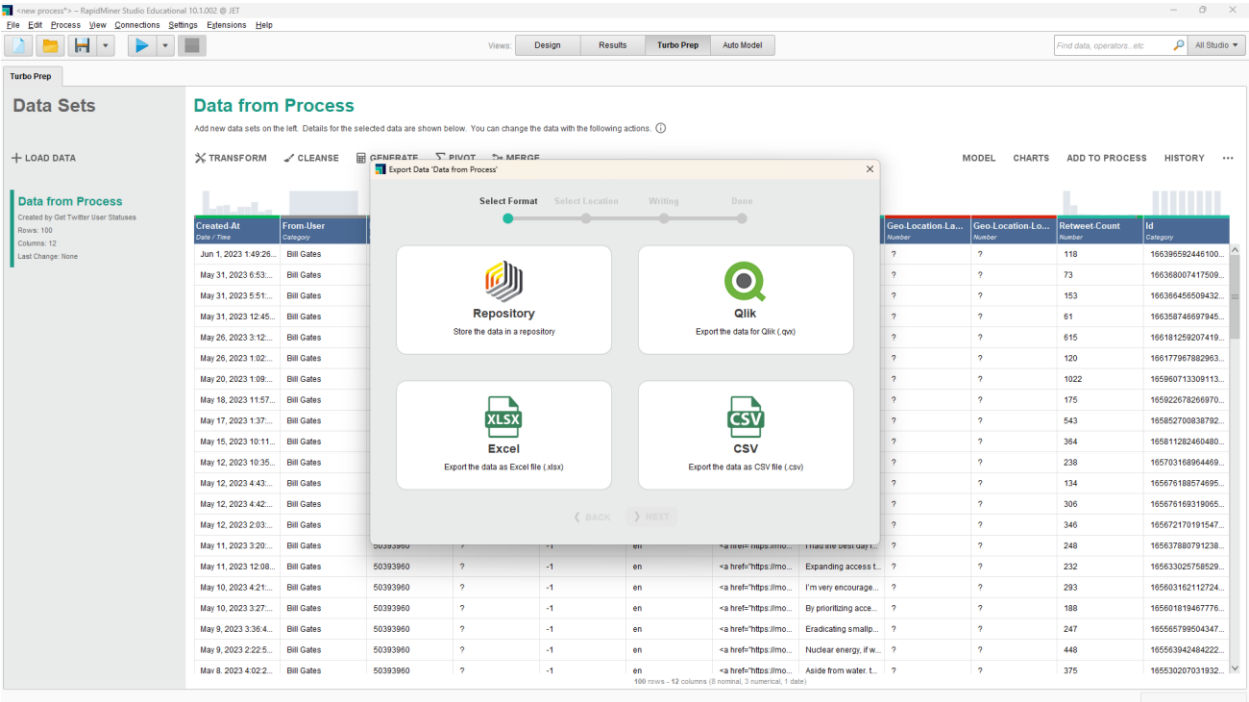
Select the location where you want to export the csv file to. Afterwards, specify the name of the file. Click **Next.**
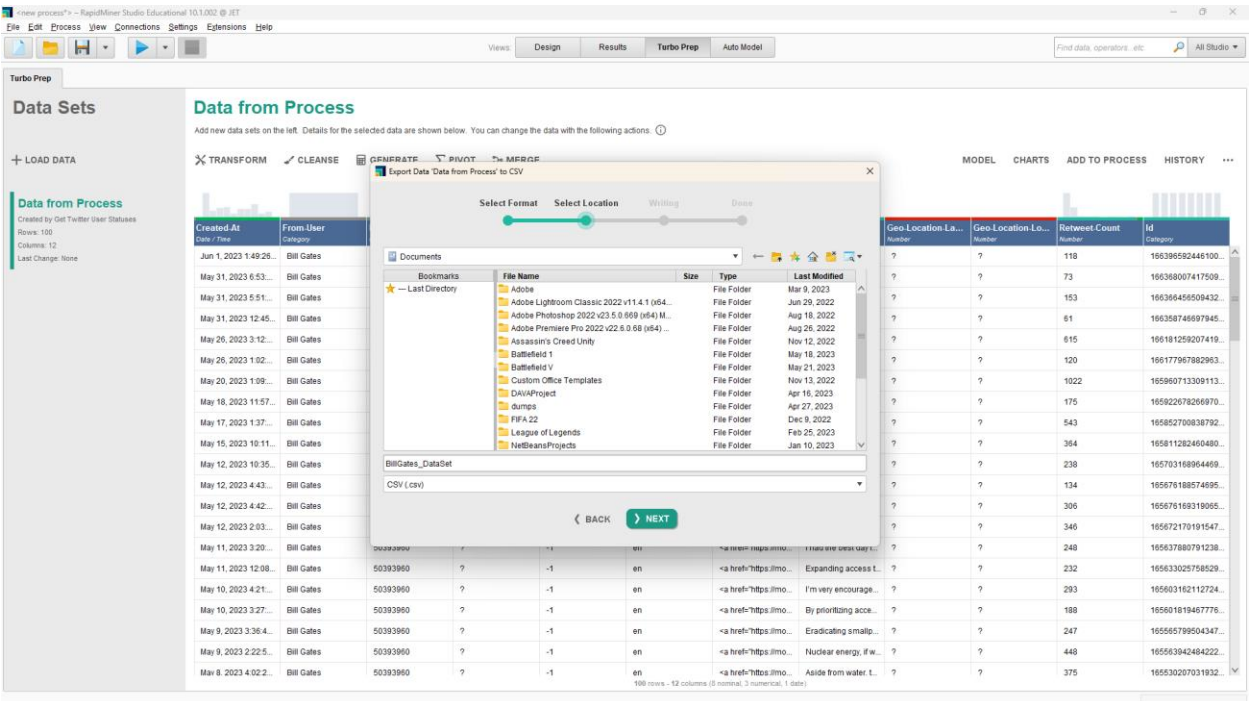
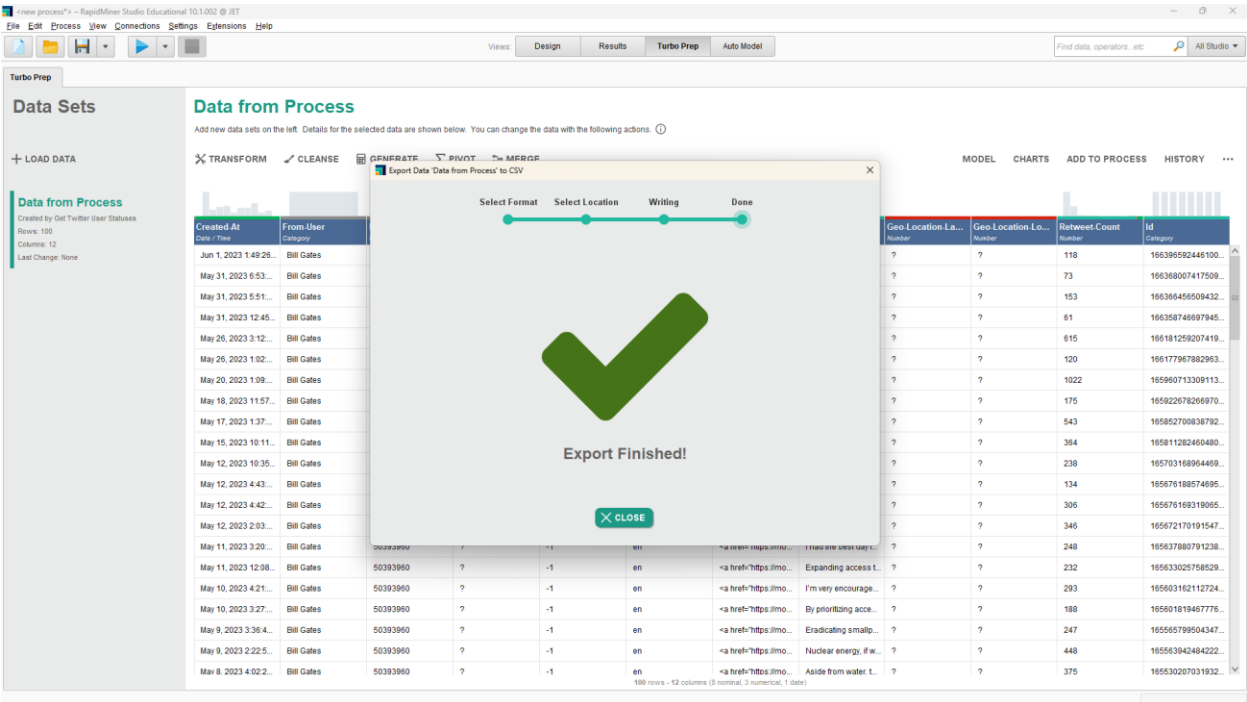You will then be prompted with a tab showing that the export is finished. Click **Close.** Navigate to the location where you stored your file and open it to verify that the data set has been exported.

You have now successfully gathered and downloaded a data set from Twitter using RapidMiner!