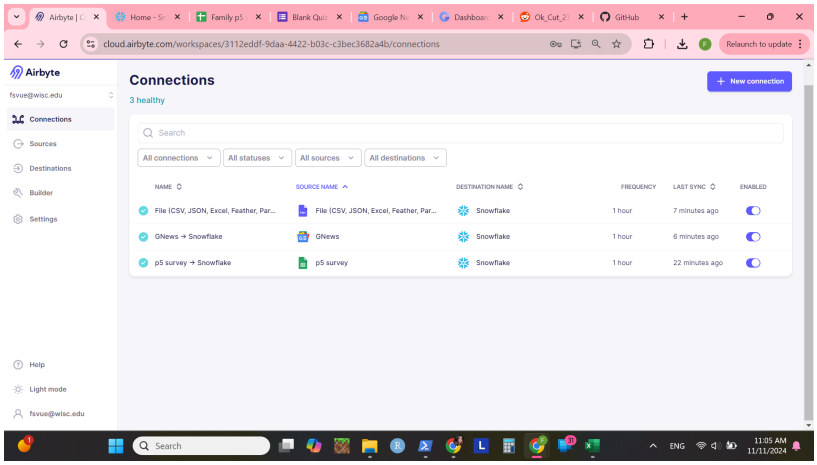
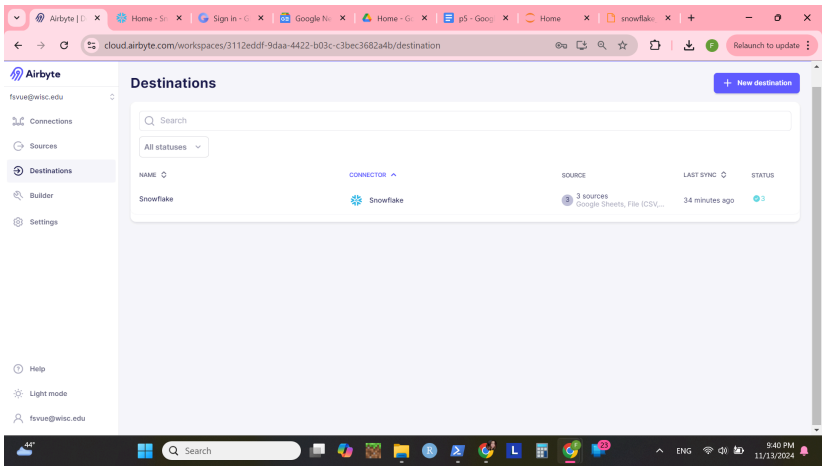
1.2.2 screen shot of Sources section in Airbyte



1.3 Snowflake Setup
https://av13777.us-central1.gcp.snowflakecomputing.com

1.4 Connect Airbyte to Snowflake



2 (USED SQL)

To create this transformation I decided to create a new table called “BOROUGH_TRIP_COUNTS” that found the unique names within taxi_zone_lookup.csv. After this, I decided to make a total count of all the drop off and pick ups in nyc_taxi.csv and have them be listed in the primary key that was the borough names. Once I did this, I decided to use more code to create an output that would find the proportion of each drop off count and pick up count of each borough. I did this by using the count of each borough (depending on drop off or pick up) and divided it by the sum of the pick up or drop off totals. I then named these total_countdo (drop off) and total_coundpu (pick up).

Here is my code :
USE DATABASE AIRBYTE_DATABASE;

```
USE SCHEMA airbyte_schema;

CREATE TABLE Borough_Trip_Counts AS
WITH BOROUGH AS (
   SELECT LOCATIONID, BOROUGH
   FROM AIRBYTE_DATABASE.AIRBYTE_SCHEMA.TAXI_ZONE_LOOKUP
)

SELECT b.BOROUGH,
   COUNT(DISTINCT n.PULOCATIONID) AS PULOCATION_COUNT,
   COUNT(DISTINCT n.DOLOCATIONID) AS DOLOCATION_COUNT
FROM
   AIRBYTE_DATABASE.AIRBYTE_SCHEMA.NYC_TAXI n
LEFT JOIN Borough b
   ON n.PULOCATIONID = b.LOCATIONID
LEFT JOIN borough d
   ON n.DOLOCATIONID = d.LOCATIONID
GROUP BY
   b.BOROUGH;

SELECT DOLOCATION_COUNT/SUM(DOLOCATION_COUNT) OVER () AS total_countdo,
PULOCATION_COUNT/SUM(PULOCATION_COUNT) OVER () AS total_countpu
FROM BOROUGH_TRIP_COUNTS;
```