Case Study: NYC Green Taxis – Optimizing Revenue

## **📘 Business Context**

Green taxis in NYC serve outer boroughs and airports. The business team wants to analyze **monthly revenue patterns**, **pickup/drop-off hotspots**, and **driver performance** to uncover areas of optimization.

The **executive sponsor** is especially interested in:

* Revenue differences between July and August
* Trends in tipping behavior
* Popular zones by trip count and revenue
* Anomalies or patterns in trip distances and durations

They now have **two months of data** (July and August 2021) and want your team to build a model that scales and supports **month-over-month analysis**.

The zone lookup has also been updated as new trips have been under taken so you would update the seed using the new file.

Note:

* Refer to Lab Instructions for further details if you get stuck at any point

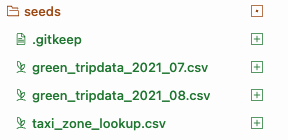
🔎 **Steps**

**All requisite resources can be found** [**at this link**](https://stratanconsulting-my.sharepoint.com/:f:/p/asim/EqR2Kf7EJDlJiSwIyoBfcsEBzVr1uM2Jtp9maFmycqtSPg?e=bXWfNz)**.**

#### **1. Add Additional Seed/Updated Lookup Data**

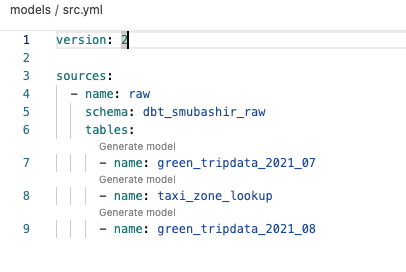
* Add green\_tripdata\_2021\_08.csv to the seeds/ folder.
  + Notes:
    - if it says storage is full >> Go back and save changes
    - Recommend opening csv in notepad before copying into DBT
* Update seed/taxi\_zone\_lookup.csv in DBT with new data from taxi\_zone\_lookup1.csv
* Run dbt seed to load all three datasets into the raw layer.
* Confirm in Neon that a new table is created and data is updated in the taxi\_zone\_lookup table.

This is what your seed folder will look like in DBT.



#### **2. Create a New Staging Model**

* **Create Empty File:** models/staging/stg\_green\_tripdata\_combined.sql
* **Add Code:** 2.staging\_model.sql
* Remember to update src.yml
* Run dbt build
  + Note a failure in not\_null\_my\_first\_dbt\_model\_id is not necessarily a failure if the rest of the build is successful

This is what your src.yml file may look like.

#### **3. Create a New Mart Model**

* **Create Empty File:** models/business/fact\_combined\_trips.sql
* **Add Code:** 3.mart\_model.sql
* Run dbt build

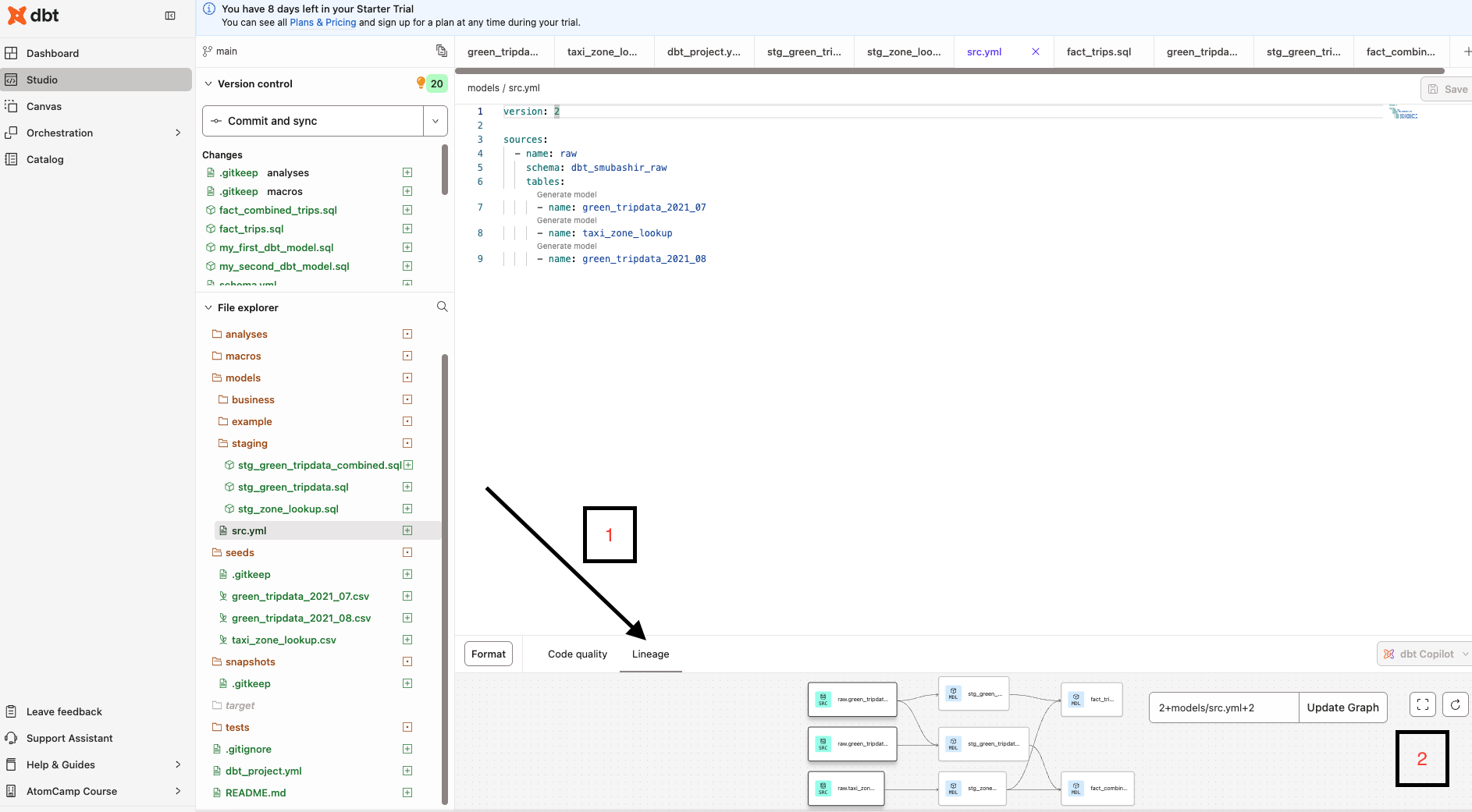
## **📊 Business Questions to Answer**

1. How does revenue compare between July and August?
2. What are the top 10 pickup zones by revenue in each month?
3. Which zones show the largest increase/decrease in revenue between months?
4. How do average tips vary by drop-off zone and month?
5. What’s the average trip distance and fare per zone per month?
6. Do certain pickup zones consistently generate higher tips or longer rides?
7. Is there a correlation between trip distance and tip amount across months?
8. What zones saw the largest drop in average revenue per trip month-over-month?

## **🧩 Deliverables**

1. Set up a DBT project and Neon connection (if you’ve already done so in the lab you can use the same one).
2. Load data using dbt seed.
3. Create staging models for each source table (you can add to existing tables if staging table already created for lab).
4. Create a fact\_combined\_trips model in the mart schema.
5. Write SQL queries to answer each business question.
6. Create a report that contains:
   1. The Lineage Diagram of your models from DBT (see Sample 1 below)
   2. A screenshot from Neon of your mart schema and the fact\_combined\_trips with data (see sample 2 below)
   3. The answers to the business questions above

Sample 1



Sample 2

