

Foundations of dbt

INTRODUCTION TO DBT



Mike Metzger

Data Engineering Consultant

Course goals

- Understand core dbt concepts
- dbt practical applications
- Why it's essential for data professionals
- Navigate a dbt project
- Write & execute basic dbt models
- Gain familiarity with an extremely popular data engineering tool



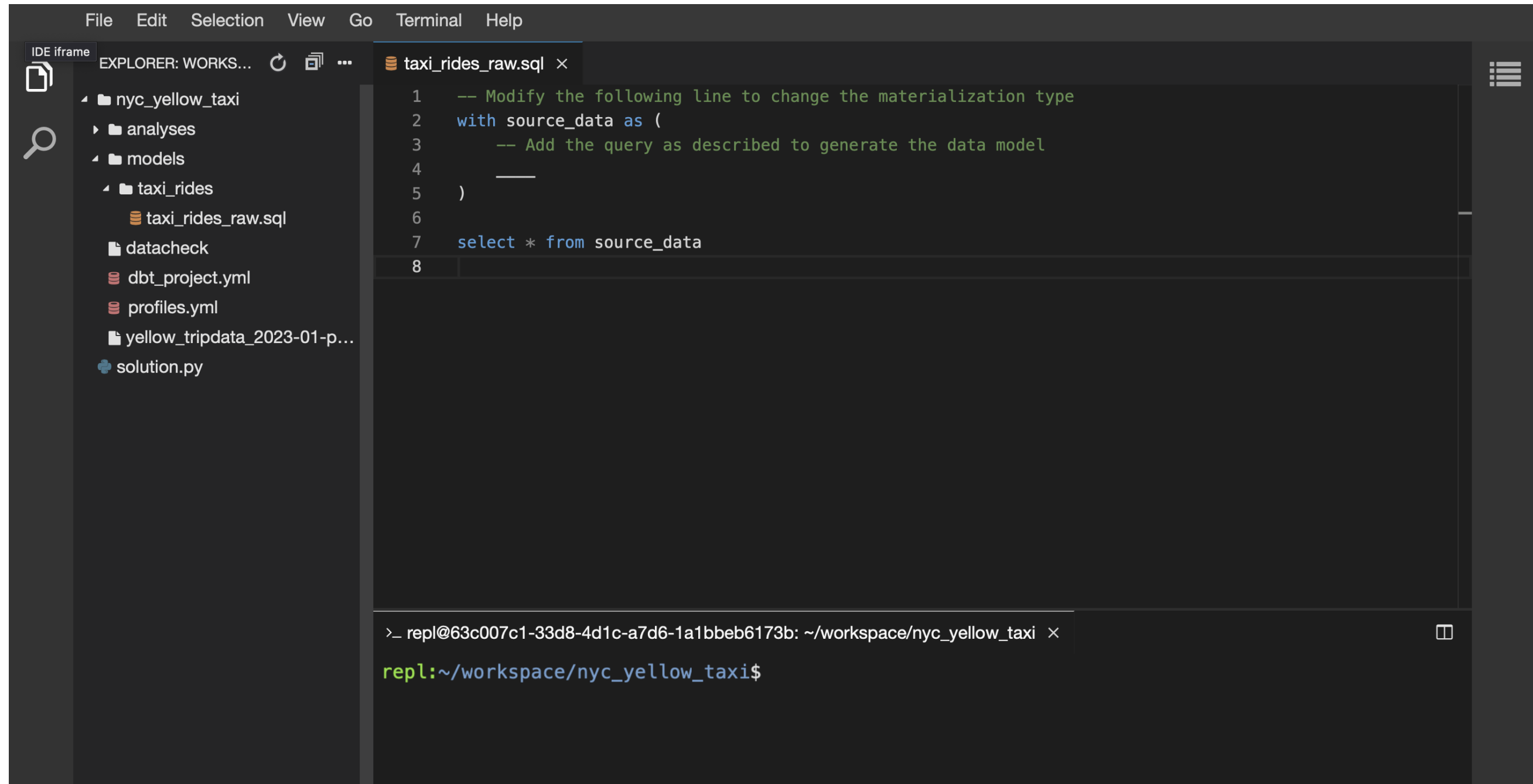
Prerequisite knowledge

- SQL experience
- Command-line knowledge
- Data pipelines / data modeling

The image displays three course cards from DataCamp, each with a dark blue background and white text. Each card includes a 'Continue' button in green and a 'Bookmark' button in white. The first card is for 'Joining Data in SQL', a beginner-level course with 4 hours, 14 videos, 47 exercises, 175,185 participants, and 3950 XP. The second card is for 'Introduction to Shell', also a beginner-level course with 4 hours, 0 videos, 55 exercises, 130,762 participants, and 4650 XP. The third card is for 'Data Warehousing Concepts', an intermediate-level course with 4 hours, 16 videos, 57 exercises, 20,765 participants, and 3450 XP.

Course Title	Level	Duration	Videos	Exercises	Participants	XP
Joining Data in SQL	Beginner	4 hours	14 videos	47 exercises	175,185 participants	3950 XP
Introduction to Shell	Beginner	4 hours	0 videos	55 exercises	130,762 participants	4650 XP
Data Warehousing Concepts	Intermediate	4 hours	16 videos	57 exercises	20,765 participants	3450 XP

Interactive learning



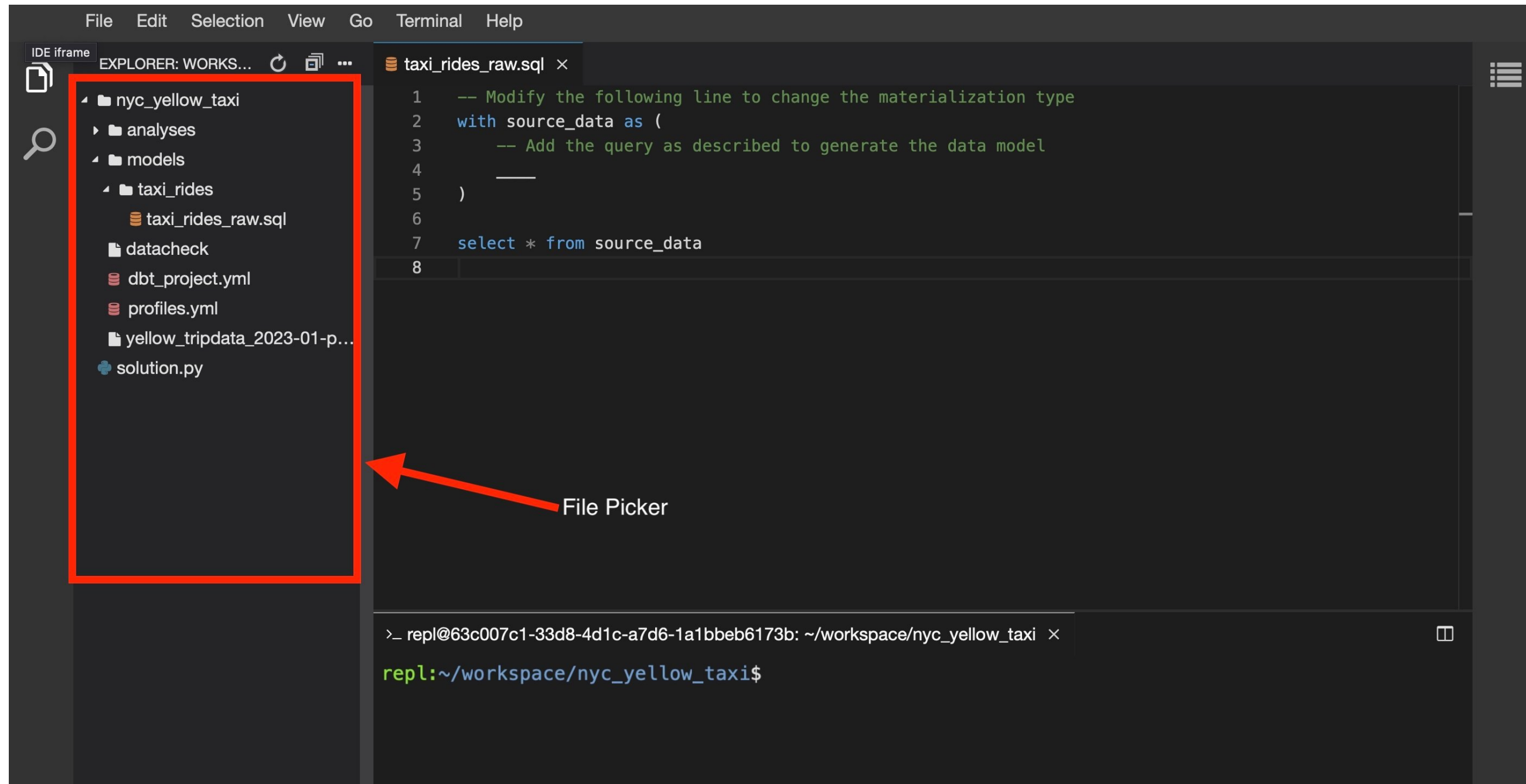
The screenshot displays a DBT IDE interface with a dark theme. The top menu bar includes File, Edit, Selection, View, Go, Terminal, and Help. On the left, the 'EXPLORER: WORKS...' panel shows a file tree with folders 'nyc_yellow_taxi', 'analyses', and 'models'. Inside 'models', there is a sub-folder 'taxi_rides' containing the file 'taxi_rides_raw.sql'. Other files listed are 'datacheck', 'dbt_project.yml', 'profiles.yml', 'yellow_tripdata_2023-01-p...', and 'solution.py'. The main editor window, titled 'taxi_rides_raw.sql', contains the following SQL code:

```
1  -- Modify the following line to change the materialization type
2  with source_data as (
3      -- Add the query as described to generate the data model
4      _____
5  )
6
7  select * from source_data
8
```

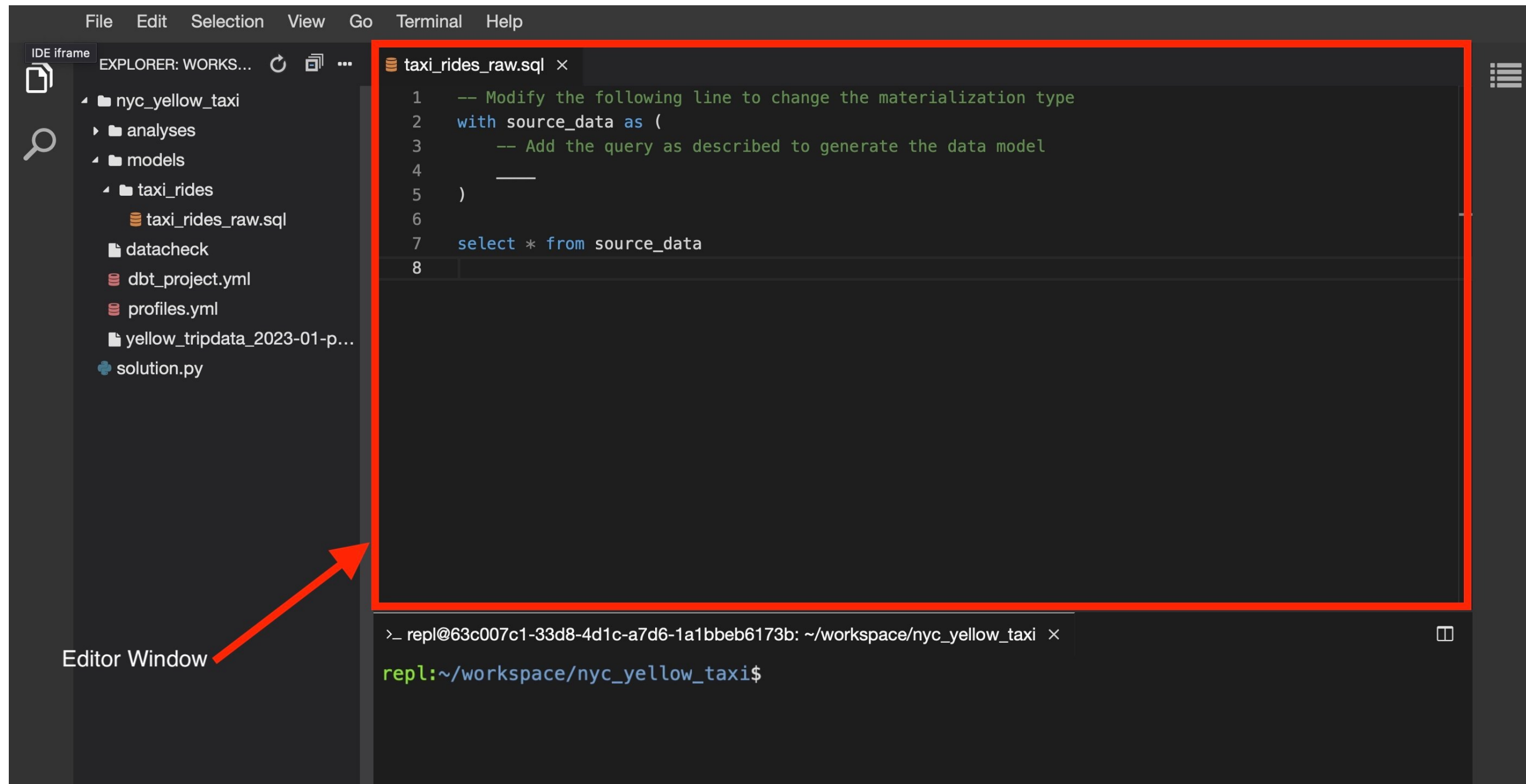
At the bottom, a terminal window shows the command prompt for a REPL session:

```
>_ repl@63c007c1-33d8-4d1c-a7d6-1a1bbeb6173b: ~/workspace/nyc_yellow_taxi ×
repl:~/workspace/nyc_yellow_taxi$
```

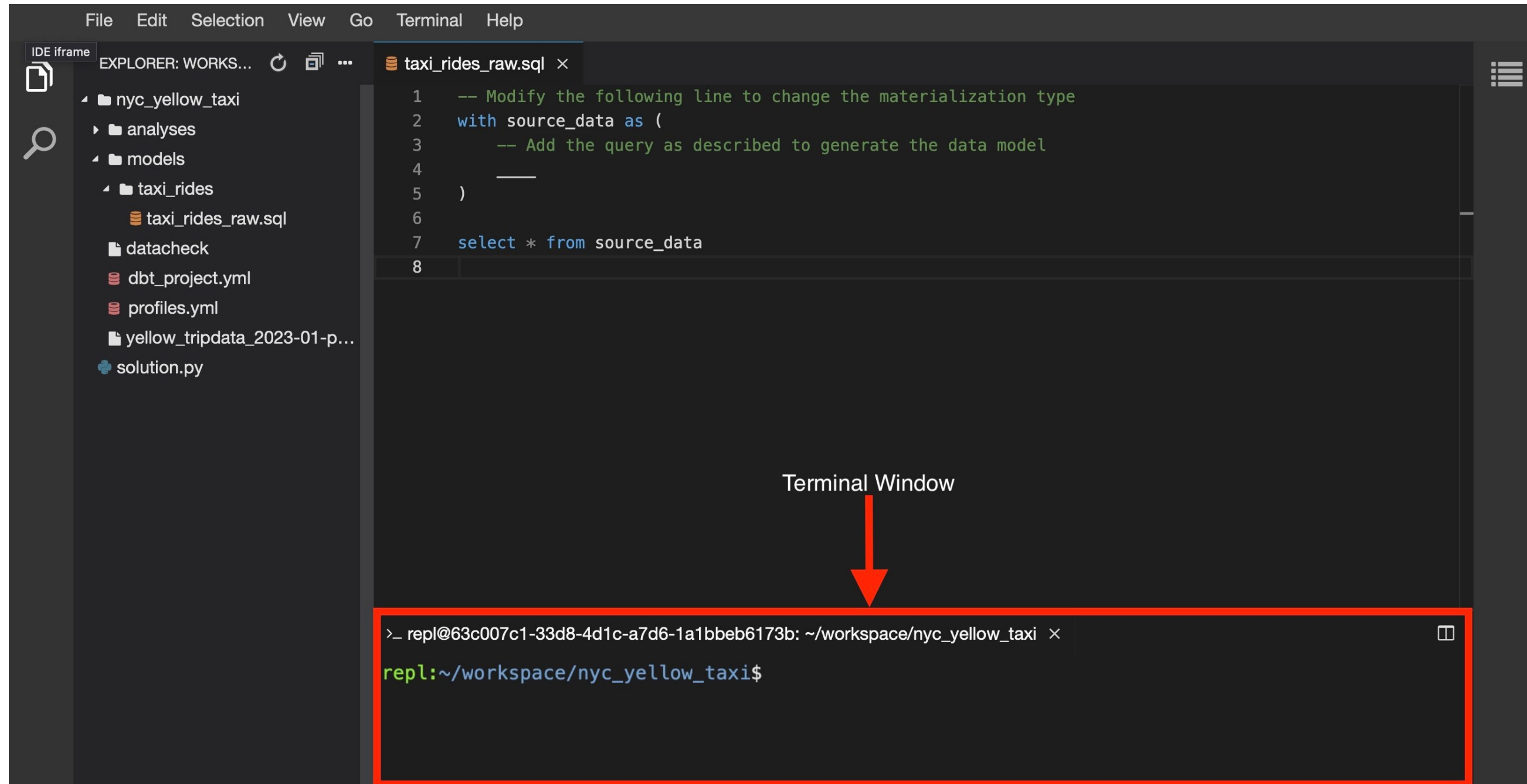
Interactive learning - File picker



Interactive learning - Code Editor



Interactive learning - Terminal



Let's practice!

INTRODUCTION TO DBT

Introducing dbt

INTRODUCTION TO DBT



Mike Metzger
Data Engineer

dbt defined

- "data build tool"
- Primarily handles the T in ELT (sometimes ETL)
- Allows easy switching between data warehouses
- Provides SQL collaboration
- Translates between SQL dialects



What does dbt do?

- Primarily defines data models and transformations using SQL
 - eCommerce, mobile app interactions, etc
- Can define relationships between data models
 - Customers, Addresses, Orders
- Runs the data transformation process(es)
 - Converting raw data from log files to database tables
- Can test for data quality requirements

What does dbt look like?

- Also known as `dbt-core`, open-source
 - Command-line tool available for Mac, Windows, Linux
- dbt Cloud is proprietary
- Main command `dbt`
 - `dbt`
 - `dbt --version`
 - `dbt -h`



Who is dbt for?

dbt is designed for any users that need to transform data:

- Data Engineers
- Analytics Engineers
- Data Analysts



¹ Photo by Christina @ wocintechchat.com on Unsplash

Let's practice!

INTRODUCTION TO DBT

Creating a dbt project

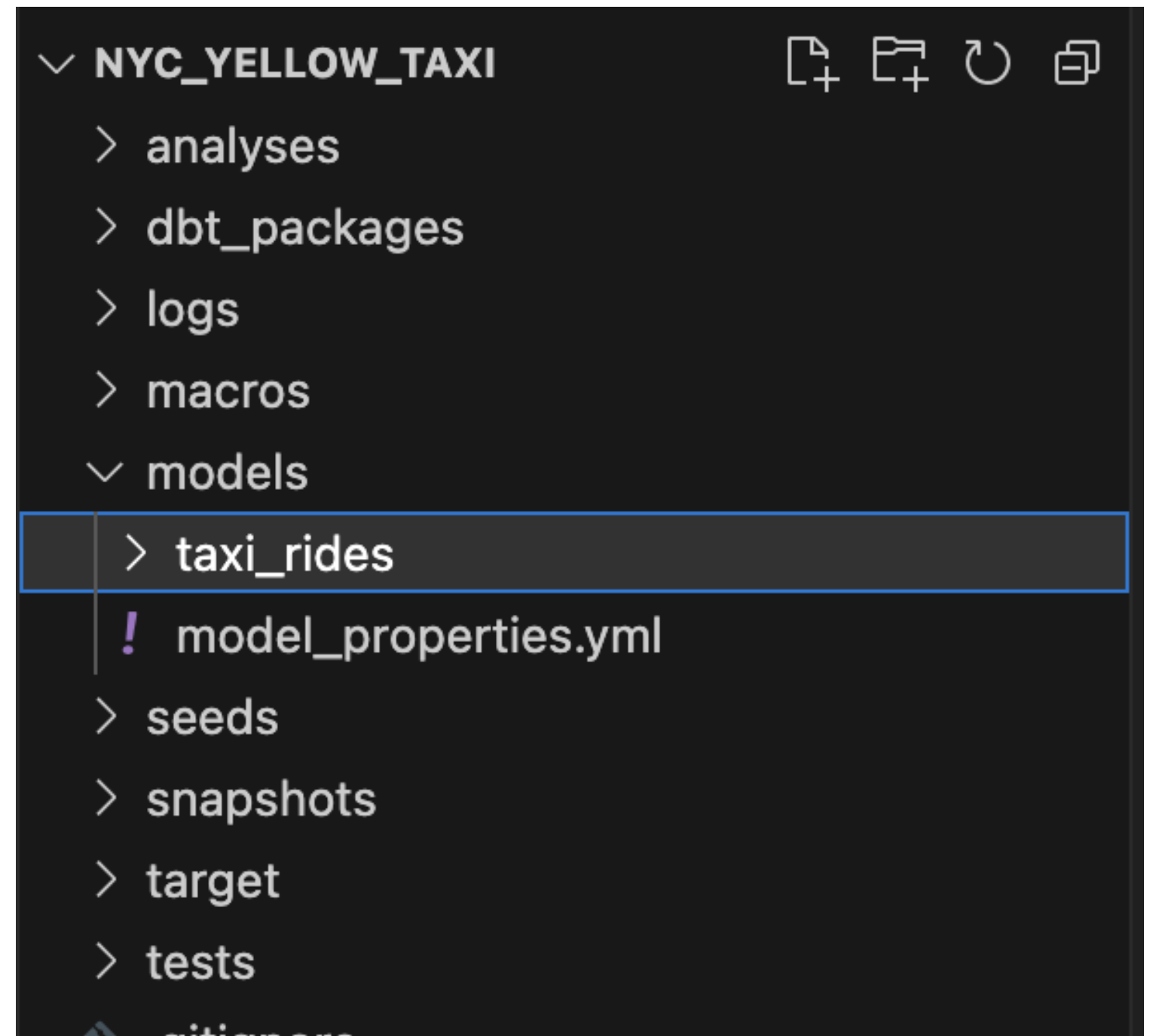
INTRODUCTION TO DBT



Mike Metzger
Data Engineer

What is a dbt project

- Encompass components for working with data in dbt
 - Project configuration
 - Data sources & destinations
 - SQL queries
 - Templates
 - Documentation
- Implemented as a folder structure



How to create a new project

- Use the `dbt init` command
 - Asks the name of the project
 - Asks which database / data warehouse type
- Can consolidate with `dbt init <projectname>`
- Creates the top level project folder and all needed structure

```
repl:~/workspace$ dbt init
16:38:37  Running with dbt=1.8.4
Enter a name for your project (letters, digits, underscore): test_project
Which database would you like to use? [1] duckdb
Enter a number: 1
...
```

Defining configuration with project profiles

- A profile represents a given deployment scenario
 - Development
 - Staging / Test
 - Production
- A dbt project can have multiple profiles
- Defined in the `profiles.yml` file

```
marketing_campaign_results:
  outputs:
    dev:
      type: duckdb
      path: dbt.duckdb
    prod:
      type: snowflake
      ...
  target: dev
```

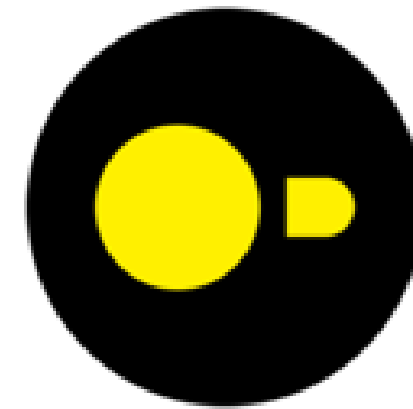
YAML

- Yet Another Markup Language
- Text file, but spacing matters (like Python)
- Used in many development scenarios for configuration
- Rules can be tricky, mainly keep entries lined up as in examples

```
marketing_campaign_results:
  outputs:
    dev:
      type: duckdb
      path: dbt.duckdb
    prod:
      type: snowflake
      ...
  target: dev
```

DuckDB

- Open-source serverless database
 - Similar to sqlite
- Designed for analytics
- Vectorized (meaning *FAST*)
- Easy to use



DuckDB

Let's Practice!

INTRODUCTION TO DBT