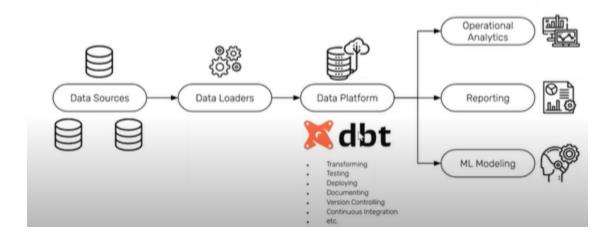
# **Analytics Engineering**

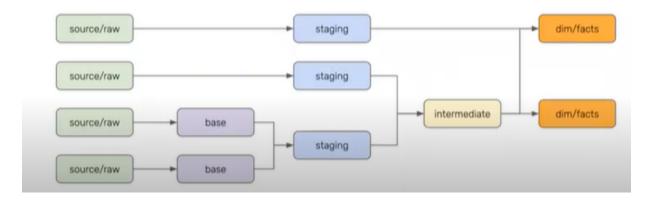
- What is analytics engineering?
- What, exactly, is dbt?

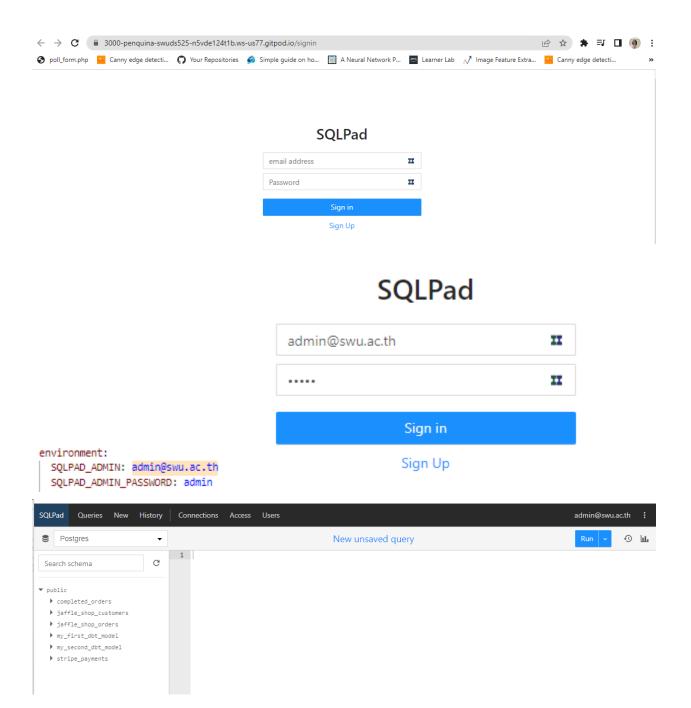
### **Project**

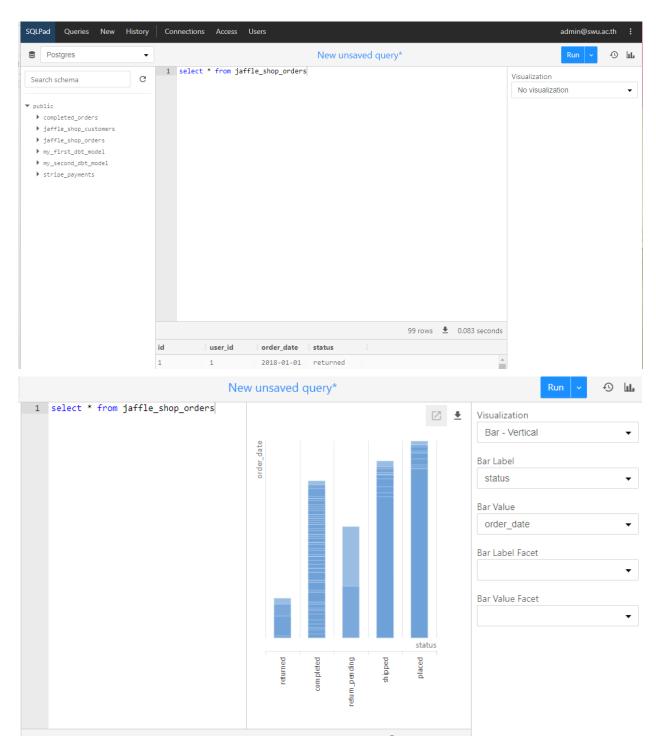
Project: สร้าง Analytics Workflow ด้วย dbt และการนำเอาแนวคิด Analytics Engineering มาใช้

Project: สร้าง Analytics Workflow ด้วย dbt และการนำเอาแนวคิด Analytics Engineering มาใช้ สามารถใช้งาน dbt และสร้าง data modeling ของตัวเองจากข้อมูล Greenery ได้









#### Configuration:

profiles.yml file [ERROR not found]
dbt\_project.yml file [OK found and valid]

```
home > gitpod > .dbt > ! profiles.yml > () jaffle > ™ target
      jaffle:
  2
        outputs:
  3
  4
          dev:
  5
            type: postgres
  6
            threads: 1
  7
           host: localhost
  8
            port: 5432
  9
            user: postgres
 10
            pass: postgres
 11
            dbname: postgres
 12
          schema: public
 13
          prod:
 14
 15
           type: postgres
 16
            threads: 1
           host: localhost
 17
 18
            port: 5432
 19
           user: postgres
 20
            pass: postgres
 21
            dbname: postgres
 22
            schema: prod
 23
 24
       target: dev
```

```
drwxr-xr-x 2 gitpod gitpod 91 Nov 6 08:08 tests
drwxr-xr-x 2 gitpod gitpod 22 Nov 6 08:08 snapshots
drwxr-xr-x 2 gitpod gitpod 22 Nov 6 08:08 snapshots
drwxr-xr-x 2 gitpod gitpod 22 Nov 6 08:08 seeds
-rw-r--r-- 1 gitpod gitpod 571 Nov 6 08:08 README.md
drwxr-xr-x 2 gitpod gitpod 51 Nov 6 08:08 macros
drwxr-xr-x 2 gitpod gitpod 51 Nov 6 08:08 macros
drwxr-xr-x 2 gitpod gitpod 22 Nov 6 08:08 analyses
drwxr-xr-x 5 gitpod gitpod 74 Nov 12 10:26 ENV
-rw-r---- 1 gitpod gitpod 1321 Nov 12 10:26 ENV
-rw-r---- 1 gitpod gitpod 76 Nov 12 11:16 models

• (ENV) gitpod /workspace/swu-ds525/06-analytics-engineering/jaffle (main) $ mkdir -p ~/.dbt

• (ENV) gitpod /workspace/swu-ds525/06-analytics-engineering/jaffle (main) $ code ~/.dbt/profiles.yml

• (ENV) gitpod /workspace/swu-ds525/06-analytics-engineering/jaffle (main) $ [
```

```
4
     ```sh
 5
 6 dbt init
 7
 8
9 Edit the dbt profiles
10
11 ```sh
12 code ~/.dbt/profiles.yml
13
14
    ```yml
15
16
     jaffle:
17
      ··outputs:
18
       ···dev:
19
       type: postgres threads: 1
20
21
22
       · · · · · host: ·localhost
       port: 5432
user: postgres
pass: postgres
dbname: postgres
23
24
25
26
27
       · · · schema: public
28
       ····prod:
····type: postgres
····threads: 1
29
30
31
       ····host: ·localhost
32
       ····user: postgres
33
34
35
       ····pass: postgres
····dbname: postgres
36
37
       ···schema: prod
38
      · target: dev
39
40
```

```
    (ENV) gitpod /workspace/swu-ds525/06-analytics-engineering/jaffle (main) $ code ~/.dbt/profiles.yml

    (ENV) gitpod /workspace/swu-ds525/06-analytics-engineering/jaffle (main) $ dbt debug

  16:42:57 Running with dbt=1.2.0
 dbt version: 1.2.0
 python version: 3.8.13
 python path: /workspace/swu-ds525/06-analytics-engineering/ENV/bin/python
 os info: Linux-5.15.0-47-generic-x86_64-with-glibc2.29
 Using profiles.yml file at /home/gitpod/.dbt/profiles.yml
 Using dbt_project.yml file at /workspace/swu-ds525/06-analytics-engineering/jaffle/dbt_project.yml
 Configuration:
    profiles.yml file [OK found and valid]
    dbt_project.yml file [OK found and valid]
  Required dependencies:
  - git [OK found]
  Connection:
    host: localhost
    port: 5432
   user: postgres
   database: postgres
    schema: public
    search_path: None
    keepalives idle: 0
    sslmode: None
    Connection test: [OK connection ok]
  All checks passed!

    (ENV) gitpod /workspace/swu-ds525/06-analytics-engineering/jaffle (main) $ [

∨ 06-analytics-engineering
                              18
                                     restart: always
 > ENV
                              19
                                     depends on:
 ∨ jaffle
                                       - postgres
                              20
  analyses
                                     SQLPAD_ADMIN: admin@swu.ac.th
                              22
  > dbt_packages
                                     SQLPAD_ADMINIPSSNORD: admin
SQLPAD_CONNECTIONS_pgdb_name: Postgres
SQLPAD_CONNECTIONS_pgdb_driver: postgres
SQLPAD_CONNECTIONS_pgdb_dost: postgres
SQLPAD_CONNECTIONS_pgdb_database: postgres
SQLPAD_CONNECTIONS_pgdb_username: postgres
SQLPAD_CONNECTIONS_pgdb_password: postgres
SQLPAD_CONNECTIONS_pgdb_password: postgres
                              23
  > ENV
  > logs
 > macros
                              26
  ∨ models
                              27
                              28
  > example
  marts
                                    ports:
                              30
   > staging
                              31
                                    - 3000:3000
  my_simple_model.sql
                              32
  > seeds
  > snapshots
                              34 postgres-data-volume:
  > target
                            PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
  gitignore
  ! dbt project.vml
                            16:44:19 Found 5 models, 8 tests, 0 snapshots, 0 analyses, 257 macros, 0 operations, 0 seed files, 3 sources, 0 exposures, 0 metrics

 README.md

                             16:44:19
 logs
                             16:44:19 Concurrency: 1 threads (target='dev')
                             16:44:19
                             docker-compose.yml
                       = notes.txt
                            (i) README md
                            16:44:19 2 of 5 CK created table model public.my_first_dbt_model ...... [SELECT 1 in 0.06s]
 = requirements.txt
Frequirements.tot 16:44:19 3 of 5 START view model public.my_simple_model [RUN]

V 07-Data Visualization and DashboardP... 16:44:19 3 of 5 CK created view model public.my_simple_model [CREATE VIEW in 0.035]
 ∨ data
                            16:44:19 5 of 5 OK created view model public.my_second_dbt_model ...... [CREATE VIEW in 0.03s]
{} github_events_01.json
                             16:44:19
{} github_events_02.json
                            16:44:19 Finished running 4 view models, 1 table model in 0 hours 0 minutes and 0.44 seconds (0.44s).
```

16:44:19

16:44:19

16:44:19 Completed successfully

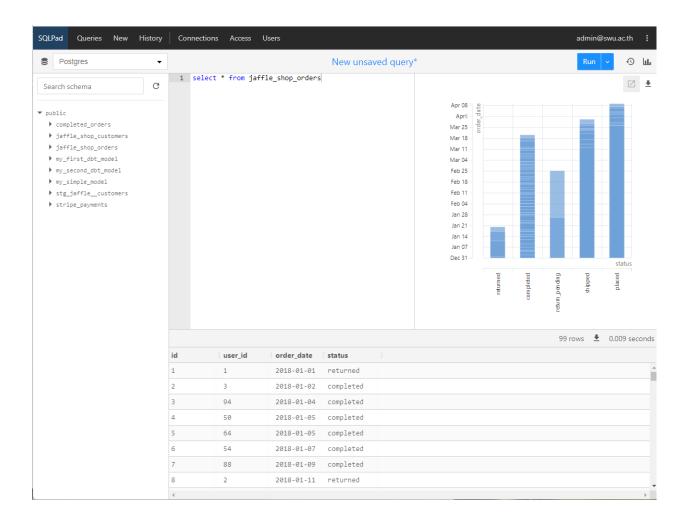
16:44:19 Done, PASS=5 WARN=0 ERROR=0 SKIP=0 TOTAL=5

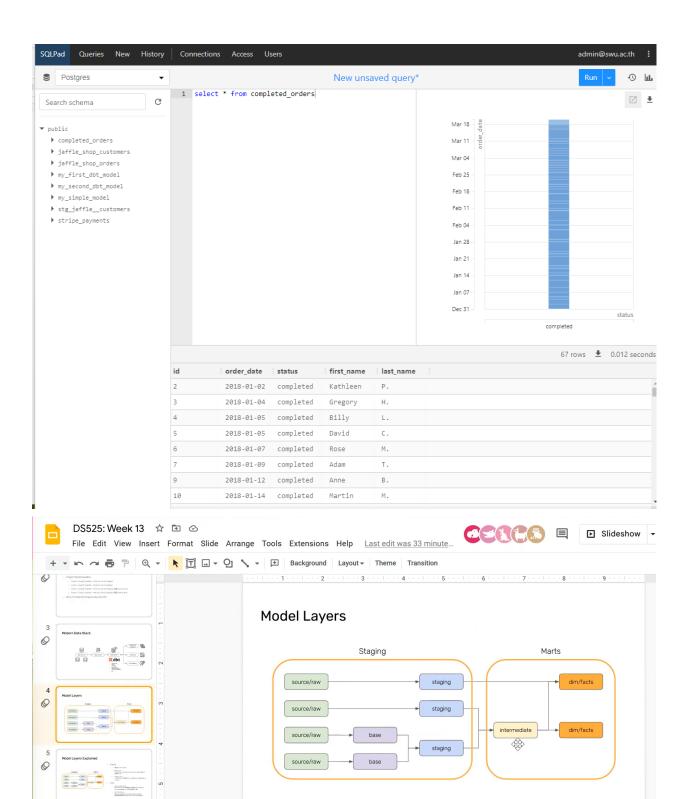
○ (ENV) gitpod /workspace/swu-ds525/06-analytics-engineering/jaffle (main) \$ [

{} aithub events 03.ison

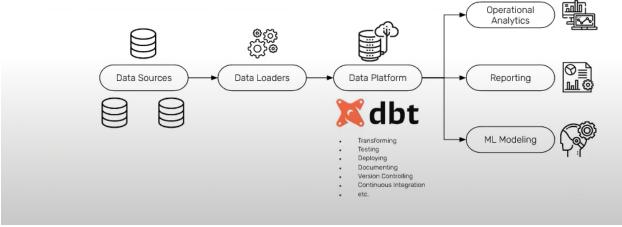
{} github\_events\_04.json

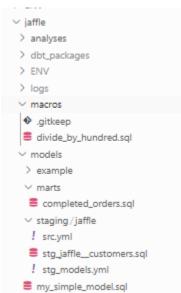
OUTLINE



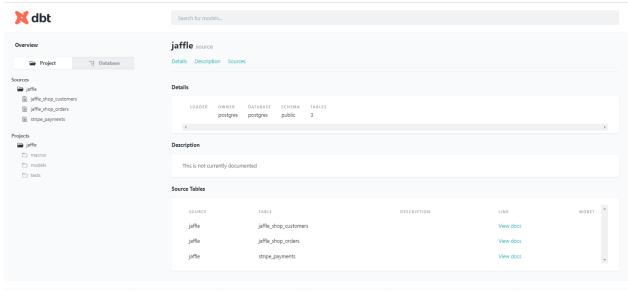


## Modern Data Stack





• (EWW) gitpod /workspace/swu-ds525/06-analytics-engineering/jaffle (main) \$ dbt docs generate
17:21:43 Running with dbt=1.2.0
17:21:43 Found 5 models, 8 tests, 0 snapshots, 0 analyses, 257 macros, 0 operations, 0 seed files, 3 sources, 0 exposures, 0 metrics
17:21:43
17:21:43 Concurrency: 1 threads (target='dev')
17:21:43
17:21:44 Done.
17:21:44 Done.
17:21:44 Building catalog
17:21:44 Catalog written to /workspace/swu-ds525/06-analytics-engineering/jaffle/target/catalog.json
(EWW) gitpod /workspace/swu-ds525/06-analytics-engineering/jaffle (main) \$ dbt docs serve



06-analytics-engineering > jaffle > models > staging > jaffle > ! stg\_models.yml > [ ] models > () 0 > [ ] columns > () 1 > @ name

```
version: 2

models:
- name: stg_jaffle_customers
description: Staging model for customers
columns:
- name: id
- name: name
- name: name
```

```
▼ public
   ▶ completed_orders
   ▼ jaffle_shop_customers
       id int4
       first_name text
       last_name text
   ▼ jaffle_shop_orders
       id int4
       user_id int4
       order_date date
       status text
   ▼ my_first_dbt_model
       id int4
   ▼ my_second_dbt_model
       id int4
   ▼ my_simple_model
       a numeric
       b numeric
       c numeric
   ▼ stg jaffle customers
       id int4
       name text
  ▼ stripe_payments
       id int4
       order_id int4
       payment_method text
       status text
```

amount float8

amount line

amount point created date

▼ public

completed\_orders

id int4

▼ jaffle\_shop\_customers

first\_name text

last\_name text

▶ jaffle\_shop\_orders

my\_first\_dbt\_model

my\_second\_dbt\_model

stg\_jaffle\_\_customers

order\_id int4

status text

amount line

amount point

created date

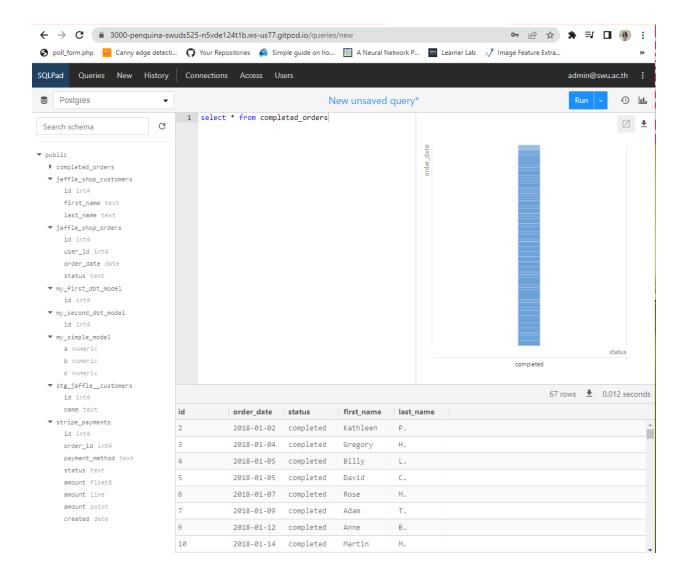
amount float8

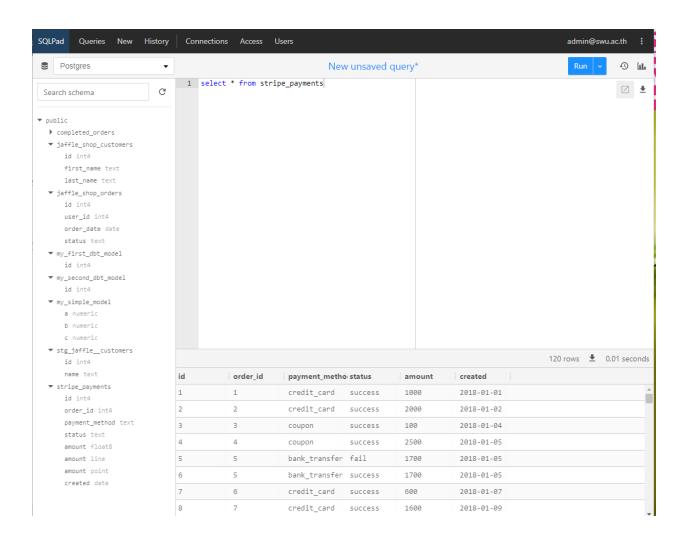
payment\_method text

my\_simple\_model

▼ stripe\_payments

id int4

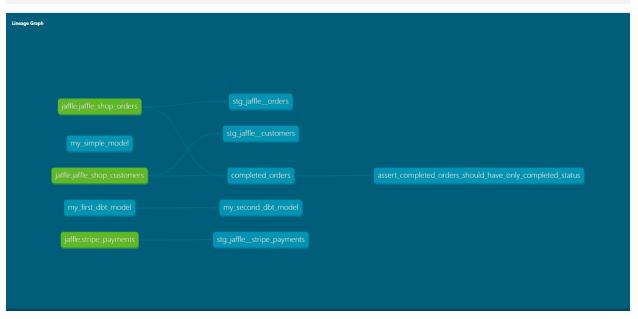


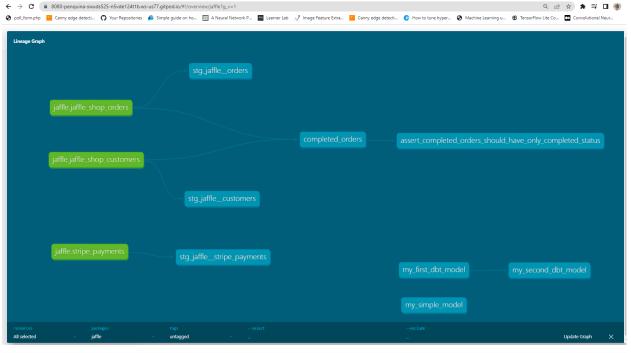


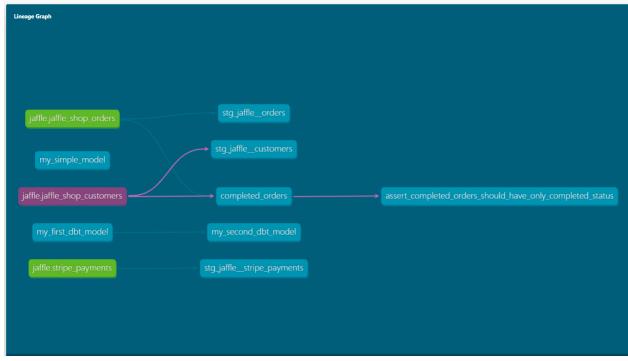
```
06-analytics-engineering > jaffle > models > staging > jaffle > ! stg_models.yml > [ ] models > ( ) 2 > [ ] columns > ( ) 5 > @ name
 1 version: 2
  2
     models:
 3
       - name: stg_jaffle__customers
  4
        description: Staging model for customers
  6
        columns:
  7
           - name: id
 8
            tests:
              - unique
- not_null
  9
 10
 11
        - name: name
 12
        - name: stg_jaffle__orders
 13
 14
        description: Staging model for orders
 15
         columns:
          - name: id
 16
 17
           - name: user_id
 18
            tests:
 19
               - relationships:
                 to: ref('stg_jaffle__customers')
field: id
 20
 21
           name: order_date
 22
 23
           - name: status
            tests:
 24
 25
              - accepted_values:
        values: ['placed', 'shipped', 'return_pending', 'returned', 'completed']
 26
 27
 28
       - name: stg_jaffle__stripe_payments
 29
        description: Staging model for Stripe payments
 30
         columns:
 31
          - name: id
           - name: order_id
 32
           - name: payment_method
 33
           - name: amount
 34
            - name: status
 35
 36
           - name: created
```

```
• (ENV) gitpod /workspace/swu-ds525/06-analytics-engineering/jaffle (main) $ dbt run
17:39:42 Running with dbt=1.2.0
17:39:42 Found 7 models, 12 tests, 0 snapshots, 0 analyses, 257 macros, 0 operations, 0 seed files, 3 sources, 0 exposures, 0 metrics
17:39:42
17:39:42 Concurrency: 1 threads (target='dev')
 17:39:42
17:39:43 4 of 7 OK created view model public.stg_jaffle_customers .................[CREATE VIEW in 0.055]
17:39:43 6 of 7 START view model public.stg jaffle stripe payments ...... [RUN]
17:39:43 6 of 7 CK created view model public.stg_jaffle_stripe_payments ...... [CREATE VIEW in 0.06s]
 17:39:43
17:39:43 Finished running 6 view models, 1 table model in 0 hours 0 minutes and 0.76 seconds (0.76s).
17:39:43
17:39:43 Completed successfully
17:39:43
17:39:43 Done. PASS=7 WARN=0 ERROR=0 SKIP=0 TOTAL=7

    (ENV) gitpod /workspace/swu-ds525/06-analytics-engineering/jaffle (main) $
```

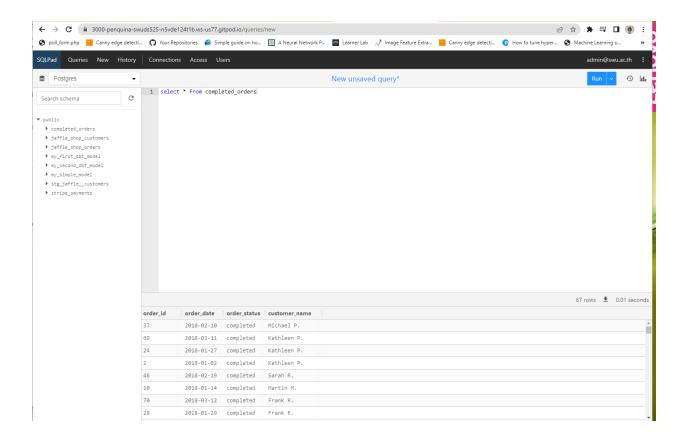




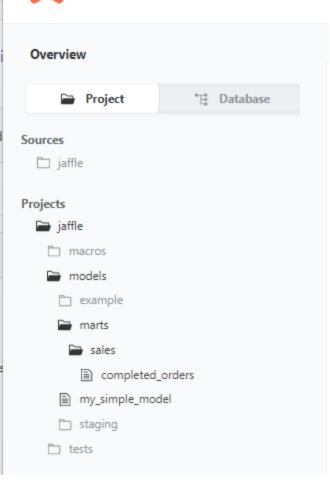


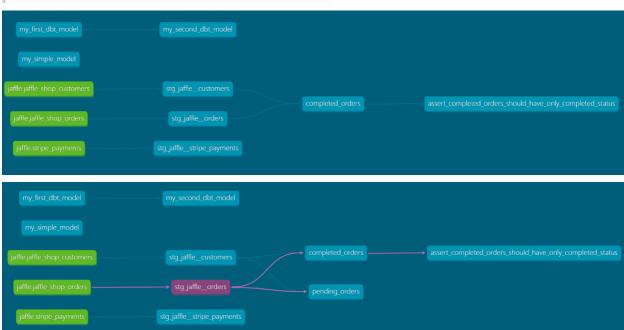
```
06-analytics-engineering > jaffle > models > marts > sales > = completed_orders.sql
  1 select
  2
         o.id
  3
         , o.order_date
          , o.status
  4
         , c.first_name
  5
          , c.last_name
  6
  7
  8  from {{ source('jaffle', 'jaffle_shop_orders') }} as o
9  join {{ source('jaffle', 'jaffle_shop_customers') }} as c
 10 on
 11 | o.user_id = c.id
 where status = 'completed'
```

```
06-analytics-engineering > jaffle > models > marts > sales > = completed_orders.sql
 1 with
 3
    orders as (
 4
 5
      select * from {{ ref('stg_jaffle_orders') }}
 6
 7
 8
 9
    , customers as (
 10
      select * from {{ ref('stg_jaffle_customers') }}
 11
 12
 13
 14
    , final as (
 15
 16
      select
        o.id as order id
 17
        , o.order_date
 18
        , o.status as order_status
 19
         , c.name as customer_name
 20
 21
      from orders as o
 22
      ioin customers as c
 23
 24
        o.user_id = c.id
 25
 26
      where status = 'completed'
 27
 28
   select * from final
 29
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
▶ (ENV) gitpod /workspace/swu-ds525/06-analytics-engineering/jaffle (main) $ dbt run
18:20:43 Running with dbt=1.2.0
18:20:44 Found 7 models, 12 tests, 0 snapshots, 0 analyses, 257 macros, 0 operations, 0 seed files, 3 sources, 0 exposures, 0 metrics
18:20:44
18:20:44 Concurrency: 1 threads (target='dev')
18:20:44
18:20:44 1 of 7 CK created table model public.my_first_dbt_model ...... [SELECT 1 in 0.14s]
18:20:44 5 of 7 CK created view model public.stg_jaffle_stripe_payments ................. [CREATE VIEW in 0.05s]
18:20:44
18:20:44 Finished running 1 table model, 6 view models in 0 hours 0 minutes and 0.61 seconds (0.61s).
18:20:44
18:20:44 Completed successfully
18:20:44
18:20:44 Done, PASS=7 WARN=0 ERROR=0 SKIP=0 TOTAL=7
(ENV) gitpod /workspace/swu-ds525/06-analytics-engineering/jaffle (main) $
```

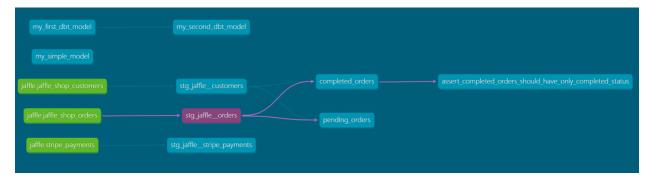




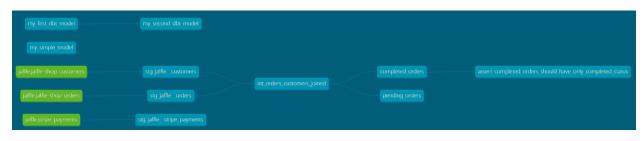




```
06-analytics-engineering > jaffle > models > marts > sales > = pending_orders.sql
  1 with
  2
  3
      orders as (
  4
  5
           select * from {{ ref('stg_jaffle_orders') }}
  6
  7
  8
      , customers as (
  9
 10
          select * from {{ ref('stg_jaffle_customers') }}
 11
 12
 13
      , final as (
 14
 15
          select
 16
             o.id as order_id
 17
 18
              , o.order_date
 19
              , o.status as order_status
 20
              , c.name as customer_name
 21
          from orders as o
 22
 23
           join customers as c
 24
 25
              o.user_id = c.id
           where status = 'pending'
 26
 27
 28
 29
     select * from final
```



```
06-analytics-engineering > jaffle > models > marts > sales > intermediate > = int_orders_customers_joined.sql
  1 with
  2
      orders as (
  3
  4
  5
           select * from {{ ref('stg_jaffle_orders') }}
  6
  7
  8
      , customers as (
  9
 10
 11
          select * from {{ ref('stg_jaffle_customers') }}
 12
 13
 14
 15
      , final as (
 16
 17
          select
             o.id as order_id
 18
              , o.order_date
 19
              , o.status as order_status
 20
              , c.name as customer_name
 21
 22
 23
          from orders as o
 24
          join customers as c
 25
          on
              o.user_id = c.id
 26
 27
 28
 29
 30
     select * from final
```





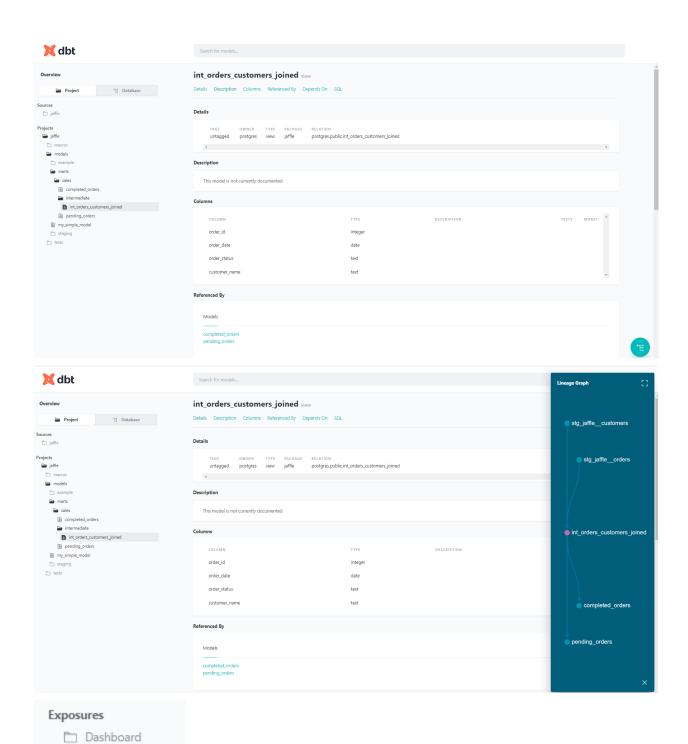
G

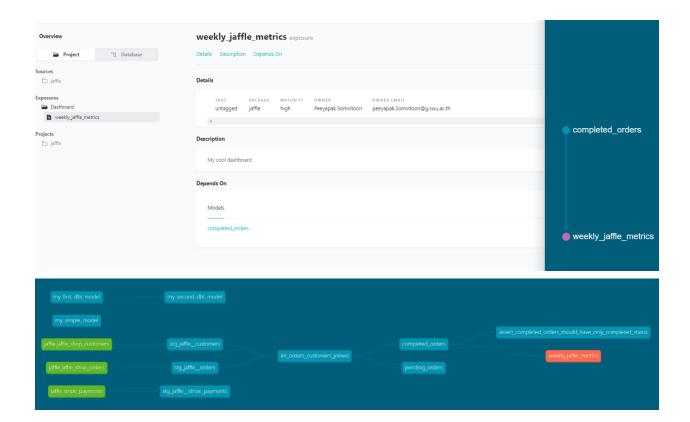
1 select \* from completed\_orders

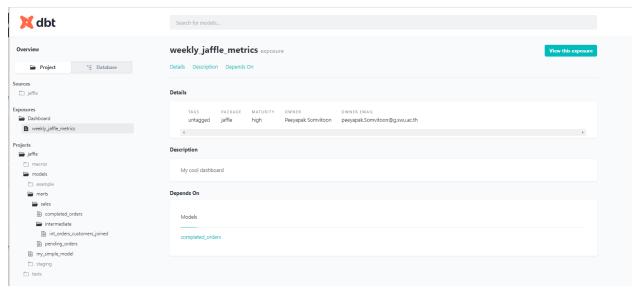
#### **▼** public

- completed\_orders
- ▶ jaffle\_shop\_customers
- jaffle\_shop\_orders
- my\_first\_dbt\_model
- my\_second\_dbt\_model
- my\_simple\_model
- stg\_jaffle\_\_customers
- ▶ stripe\_payments

order_id	order_date	order_status	customer_name
37	2018-02-10	completed	Michael P.
69	2018-03-11	completed	Kathleen P.
24	2018-01-27	completed	Kathleen P.
2	2018-01-02	completed	Kathleen P.
46	2018-02-19	completed	Sarah R.
10	2018-01-14	completed	Martin M.
70	2018-03-12	completed	Frank R.
28	2018-01-29	completed	Frank R.







```
06-analytics-engineering > jaffle > models > marts > sales > ! exposures.yml > [ ] exposures > {} 0 > \exists maturity
 1 version: 2
  2
  3 ∨ exposures:
  4
  5 ~
        - name: weekly_jaffle_metrics
  6
        type: dashboard
  7
         maturity: high
  8
         url: https://bi.tool/dashboards/1
 9 🗸
         description: >
          My cool dashboard
 10
 11
 12 ~
          depends_on:
          - ref('completed_orders')
 13
 14
 15 V
          owner:
           name: Peeyapak Somvitoon
 16
 17
           email: peeyapak.Somvitoon@g.swu.ac.th
```