

## HOMEWORK – 7

### 1. Create a dbt project with snowflake connector (1 pt)

```
Administrator: Command Prompt
https://community.getdbt.com/

Happy modeling!

22:49:55 Setting up your profile.
Which database would you like to use?
[1] snowflake

(Don't see the one you want? https://docs.getdbt.com/docs/available-adapters)

Enter a number: 1
D:\VEDA\Python-Installation\Lib\site-packages\snowflake\connector\options.py:108: UserWarning: You have an incompatible
version of 'pyarrow' installed (19.0.1), please install a version that adheres to: 'pyarrow<19.0.0; extra == "pandas"'
  warn_incompatible_dep(
account (https://<this_value>.snowflakecomputing.com): sfedu02-ksb65579
user (dev username): CRICKET
[1] password
[2] keypair
[3] sso
Desired authentication type option (enter a number): 1
password (dev password):
role (dev role): TRAINING_ROLE
warehouse (warehouse name): CRICKET_QUERY_WH
database (default database that dbt will build objects in): USER_DB_CRICKET
schema (default schema that dbt will build objects in): analytics
threads (1 or more) [1]:
22:53:53 Profile build_mau written to C:\Users\VEDA\.dbt\profiles.yml using target's profile_template.yml and your supp
lied values. Run 'dbt debug' to validate the connection.

C:\Windows\System32>
```

### 2. Set up Input models (2pt)

#### 1. Input tables should be built as CTE

```
C:\Windows\System32\build_mau\models>dbt run
23:39:18 Running with dbt=1.9.4
D:\VEDA\Python-Installation\Lib\site-packages\snowflake\connector\options.py:108: UserWarning: You have an incompatible
version of 'pyarrow' installed (19.0.1), please install a version that adheres to: 'pyarrow<19.0.0; extra == "pandas"'
  warn_incompatible_dep(
23:39:19 Registered adapter: snowflake=1.9.2
23:39:20 Found 2 models, 3 sources, 474 macros
23:39:20
23:39:20 Concurrency: 1 threads (target='dev')
23:39:20
23:39:22 1 of 2 START sql view model analytics.session_timestamp ..... [RUN]
23:39:23 1 of 2 OK created sql view model analytics.session_timestamp ..... [SUCCESS 1 in 0.63s]
23:39:23 2 of 2 START sql view model analytics.user_session_channel ..... [RUN]
23:39:23 2 of 2 OK created sql view model analytics.user_session_channel ..... [SUCCESS 1 in 0.30s]
23:39:23
23:39:23 Finished running 2 view models in 0 hours 0 minutes and 3.33 seconds (3.33s).
23:39:23
23:39:23 Completed successfully
23:39:23
23:39:23 Done. PASS=2 WARN=0 ERROR=0 SKIP=0 TOTAL=2

C:\Windows\System32\build_mau\models>
```

2. Show contents of models/input/user\_session\_channel.sql and input/session\_timestamp.sql

```
C: > Windows > System32 > build_mau > models > input > session_timestamp.sql
1  SELECT
2      sessionId,
3      ts
4  FROM {{ source('raw', 'session_timestamp') }}
5  WHERE sessionId IS NOT NULL
```

```
C: > Windows > System32 > build_mau > models > input > user_session_channel.sql
1  SELECT
2      userId,
3      sessionId,
4      channel
5  FROM {{ source('raw', 'user_session_channel') }}
6  WHERE sessionId IS NOT NULL
```

```
11
12 | SELECT * FROM USER_SESSION_CHANNEL
```

Results Chart

	# USERID	A SESSIONID	A CHANNEL
1	184	c41dd99a69df04044aa4e33ece9c9249	Naver
2	80	fdc0eb412a84fa549afe68373d9087e9	Organic
3	251	0a54b19a13b6712dc04d1b49215423d8	Facebook
4	264	a914ecef9c12ffdb9bede64bb703d877	Google
5	744	05ae14d7ae387b93370d142d82220f1b	Facebook
6	229	eddeb82df22554fa67c641e3f8a25566	Google
7	265	4c4ea5258ef3fb3fb1fc48fae9b4408c	Naver

```
14 | SELECT * FROM SESSION_TIMESTAMP
```

Results Chart

	A SESSIONID	⌚ TS
1	7cdace91c487558e27ce54df7cdb299c	2019-05-01 00:13:11.783
2	94f192dee566b018e0acf31e1f99a2d9	2019-05-01 00:49:46.073
3	7ed2d3454c5eea71148b11d0c25104ff	2019-05-01 10:18:43.210
4	f1daf122cde863010844459363cd31db	2019-05-01 13:10:56.413
5	fd0efcca272f704a760c3b61dcc70fd0	2019-05-01 13:45:19.793
6	8804f94e16ba5b680e239a554a08f7d2	2019-05-01 14:23:07.660
7	c5f441cd5f43eb2f2c024e1f8b5d00cd	2019-05-01 15:03:54.650

### 3. Set up Output models (1pt)

#### 1. Show contents of models/output/session\_summary.sql

```
C:\Windows\System32\build_mau\models>dbt run
00:04:39 Running with dbt=1.9.4
D:\VEDA\Python-Installation\Lib\site-packages\snowflake\connector\options.py:108: UserWarning: You have an incompatible
version of 'pyarrow' installed (19.0.1), please install a version that adheres to: 'pyarrow<19.0.0; extra == "pandas"'
  warn_incompatible_dep(
00:04:40 Registered adapter: snowflake=1.9.2
00:04:41 Found 3 models, 3 sources, 474 macros
00:04:41
00:04:41 Concurrency: 1 threads (target='dev')
00:04:41
00:04:42 1 of 3 START sql view model analytics.session_timestamp ..... [RUN]
00:04:42 1 of 3 OK created sql view model analytics.session_timestamp ..... [SUCCESS 1 in 0.60s]
00:04:42 2 of 3 START sql view model analytics.user_session_channel ..... [RUN]
00:04:42 2 of 3 OK created sql view model analytics.user_session_channel ..... [SUCCESS 1 in 0.33s]
00:04:42 3 of 3 START sql view model analytics.session_summary ..... [RUN]
00:04:42 3 of 3 OK created sql view model analytics.session_summary ..... [SUCCESS 1 in 0.36s]
00:04:43
00:04:43 Finished running 3 view models in 0 hours 0 minutes and 2.58 seconds (2.58s).
00:04:43
00:04:43 Completed successfully
00:04:43
00:04:43 Done. PASS=3 WARN=0 ERROR=0 SKIP=0 TOTAL=3
C:\Windows\System32\build_mau\models>
```

```
C: > Windows > System32 > build_mau > models > output > session_summary.sql
1  with _dbt_cte_user_session_channel as (
2      SELECT
3          userId,
4          sessionId,
5          channel
6      FROM USER_DB_CRICKET.raw.user_session_channel
7      WHERE sessionId IS NOT NULL
8  ), _dbt_cte_session_timestamp as (
9      SELECT
10         sessionId,
11         ts
12     FROM USER_DB_CRICKET.raw.session_timestamp
13     WHERE sessionId IS NOT NULL
14 ) SELECT u.userId, u.sessionId, u.channel, st.ts
15 FROM _dbt_cte_user_session_channel u
16 JOIN _dbt_cte_session_timestamp st ON u.sessionId = st.sessionId
```

Results		Chart		
	# USERID	SESSIONID	CHANNEL	TS
1	779	7cdace91c487558e27ce54df7cdb299c	Instagram	2019-05-01 00:13:11.783
2	230	94f192dee566b018e0acf31e1f99a2d9	Naver	2019-05-01 00:49:46.073
3	369	7ed2d3454c5eea71148b11d0c25104ff	Youtube	2019-05-01 10:18:43.210
4	248	f1daf122cde863010844459363cd31db	Naver	2019-05-01 13:10:56.413
5	676	fd0efcca272f704a760c3b61dcc70fd0	Instagram	2019-05-01 13:45:19.793
6	40	8804f94e16ba5b680e239a554a08f7d2	Youtube	2019-05-01 14:23:07.660
7	468	c5f441cd5f43eb2f2c024e1f8b5d00cd	Instagram	2019-05-01 15:03:54.650

#### 4. Add Snapshot to the output table (session\_summary) (2pt)

##### 1. Show contents of snapshots/snapshot\_session\_summary.sql

```
version of 'pyarrow' installed (19.0.1), please install a version that adheres to: 'pyarrow<19.0.0; extra == "pandas"'
warn_incompatible_dep(
01:03:25 Registered adapter: snowflake=1.9.2
01:03:26 Found 3 models, 1 snapshot, 3 sources, 474 macros
01:03:26
01:03:26 Concurrency: 1 threads (target='dev')
01:03:26
01:03:27 1 of 1 START snapshot snapshot.snapshot_session_summary ..... [RUN]
01:03:31 1 of 1 OK snapshotted snapshot.snapshot_session_summary ..... [SUCCESS 1 in 3.93s]
01:03:31
01:03:31 Finished running 1 snapshot in 0 hours 0 minutes and 5.20 seconds (5.20s).
01:03:31
01:03:31 Completed successfully
01:03:31
01:03:31 Done. PASS=1 WARN=0 ERROR=0 SKIP=0 TOTAL=1
```

```
36
37 SELECT * FROM snapshot_session_summary
38 LIMIT 10
```

Results Chart

	#	USER	SESSIONID	CHANNEL	TS	DBT_SCD_ID
1	779		7cdace91c487558e27ce54df7cdb299c	Instagram	2019-05-01 00:13:11.783	968a20fba603a2f1669abf0e392558
2	230		94f192dee566b018e0acf31e1f99a2d9	Naver	2019-05-01 00:49:46.073	d1484d42374543038d705da14d625
3	369		7ed2d3454c5eea71148b11d0c25104ff	Youtube	2019-05-01 10:18:43.210	8fdecd2c89c86d082fe65756e64c7b8
4	248		f1daf122cde863010844459363cd31db	Naver	2019-05-01 13:10:56.413	3140d8abcb37b1fcf782c3fd72685fe
5	676		fd0efcca272f704a760c3b61dcc70fd0	Instagram	2019-05-01 13:45:19.793	5d6a0aab72232c7f67f4420e2a870b
6	40		8804f94e16ba5b680e239a554a08f7d2	Youtube	2019-05-01 14:23:07.660	c068e54e00def6d8edb661f6ce9e80

#### 5. Add at least 2 tests to sessionId field of session\_summary (2pt)

##### 1. Show contents of models/schema.yml

```
C:\Windows\System32\build_mau>dbt test
23:48:02 Running with dbt=1.9.4
D:\VEDA\Python-Installation\Lib\site-packages\snowflake\connector\options.py:108: UserWarning: You have an incompatible version of 'pyarrow' installed
(19.0.1), please install a version that adheres to: 'pyarrow<19.0.0; extra == "pandas"'
warn_incompatible_dep(
23:48:03 Registered adapter: snowflake=1.9.2
23:48:04 Found 3 models, 1 snapshot, 3 data tests, 3 sources, 474 macros
23:48:04
23:48:04 Concurrency: 1 threads (target='dev')
23:48:04
23:48:05 1 of 3 START test not_null_session_summary_sessionId ..... [RUN]
23:48:06 1 of 3 PASS not_null_session_summary_sessionId ..... [PASS in 1.06s]
23:48:06 2 of 3 START test not_null_session_summary_userId ..... [RUN]
23:48:06 2 of 3 PASS not_null_session_summary_userId ..... [PASS in 0.49s]
23:48:06 3 of 3 START test unique_session_summary_sessionId ..... [RUN]
23:48:08 3 of 3 PASS unique_session_summary_sessionId ..... [PASS in 1.28s]
23:48:08
23:48:08 Finished running 3 data tests in 0 hours 0 minutes and 3.99 seconds (3.99s).
23:48:08
23:48:08 Completed successfully
23:48:08
23:48:08 Done. PASS=3 WARN=0 ERROR=0 SKIP=0 TOTAL=3
```

C: > Windows > System32 > build\_mau > models > ! schema.yml

```
1  version: 2
2
3  models:
4    - name: session_summary
5      description: "Analytics model for session data"
6      columns:
7        - name: userId
8          description: "Unique identifier for a session"
9          tests:
10            - not_null
11
12        - name: sessionId
13          description: "ID of the user associated with the session"
14          tests:
15            - unique
16            - not_null
17
```