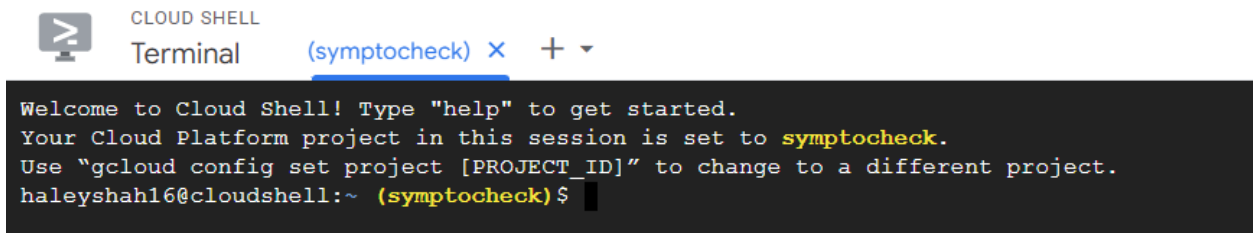


GCP Connection:



```
Cloud Shell
Terminal (symptochek)
Welcome to Cloud Shell! Type "help" to get started.
Your Cloud Platform project in this session is set to symptochek.
Use "gcloud config set project [PROJECT_ID]" to change to a different project.
haleyshah16@cloudshell:~ (symptochek)$
```

DDL Commands:

```
CREATE TABLE Records(
    user_id VARCHAR(255),
    age INT,
    sex VARCHAR(10),
    country VARCHAR(10),
    checkin_date DATETIME,
    trackable_id VARCHAR(255),
    trackable_type VARCHAR(255),
    trackable_name VARCHAR(255),
    trackable_value INT
);

CREATE TABLE Symptoms (
    SymptomId INT AUTO_INCREMENT PRIMARY KEY,
    SymptomName VARCHAR(255) NOT NULL,
    Severity INT
);

CREATE TABLE Treatment (
    TreatmentId INT AUTO_INCREMENT PRIMARY KEY,
    Medicine VARCHAR(255) NOT NULL,
    Dosage VARCHAR(30)
);

CREATE TABLE Illness(
    IllnessId INT AUTO_INCREMENT PRIMARY KEY,
    IllnessName VARCHAR(255) NOT NULL
);

CREATE TABLE Has_Symptom(
    Illness_Symptom_Id INT AUTO_INCREMENT PRIMARY KEY,
    IllnessName VARCHAR(255),
    SymptomName VARCHAR(255),
    Severity INT
);
```

```
CREATE TABLE Has_Treatment(
    Illness_Treatment_Id INT AUTO_INCREMENT PRIMARY KEY,
    IllnessName VARCHAR(255),
    Medicine VARCHAR(255) NOT NULL,
    Dosage VARCHAR(30)
);
```

Insert Into Tables:

```
mysql> SELECT COUNT(user_id) FROM Records;
+-----+
| COUNT(user_id) |
+-----+
|          7976223 |
+-----+
1 row in set (18.99 sec)
```

```
mysql> SELECT COUNT(SymptomId) FROM Symptoms;
+-----+
| COUNT(SymptomId) |
+-----+
|           63142 |
+-----+
1 row in set (0.33 sec)
```

```
mysql> SELECT COUNT(TreatmentId) FROM Treatment;
+-----+
| COUNT(TreatmentId) |
+-----+
|           19557 |
+-----+
1 row in set (0.24 sec)
```

```
mysql> SELECT COUNT(IllnessId) FROM Illness;
+-----+
| COUNT(IllnessId) |
+-----+
|           9203 |
+-----+
1 row in set (0.22 sec)
```

```
mysql> SELECT COUNT(Illness_Symptom_Id) FROM Has_Symptom;
+-----+
| COUNT(Illness_Symptom_Id) |
+-----+
|                14468863 |
+-----+
1 row in set (2.38 sec)
```

```
mysql> SELECT COUNT(Illness_Treatment_Id) FROM Has_Treatment;
+-----+
| COUNT(Illness_Treatment_Id) |
+-----+
|                297965 |
+-----+
1 row in set (0.44 sec)
```

Advanced Queries:

1. Query to get the top conditions/illnesses in which anxiety is a symptom.

```
SELECT r2.trackable_name,COUNT(r2.user_id) AS `count_user_id`  
  
FROM Records r1 JOIN Records r2 ON r1.user_id = r2.user_id  
  
WHERE r1.trackable_type = 'Symptom' AND r1.trackable_name = 'Anxiety' AND  
r2.trackable_type = 'Condition' AND r1.checkin_date = r2.checkin_date  
  
GROUP BY r2.trackable_name  
  
ORDER BY `count_user_id` DESC  
  
LIMIT 15;
```

trackable_name	count_user_id
Anxiety	12982
Fibromyalgia	12650
Depression	11673
Chronic fatigue syndrome	6651
Migraine	5921
IBS	3917
Generalized anxiety disorder	3828
Ehlers-Danlos syndrome	3170
Irritable bowel syndrome	3159
Lyme disease	3120
POTS	2835
Headaches	2663
Asthma	2509
Fatigue	2380
Allergies	2314

15 rows in set (56.83 sec)

EXPLAIN ANALYZE for Original Result:

- > Limit: 15 row(s) (actual time=56052.569..56052.572 rows=15 loops=1)
- > Sort: count_user_id DESC, limit input to 15 row(s) per chunk (actual time=56052.567..56052.569 rows=15 loops=1)
- > Table scan on <temporary> (actual time=0.003..0.669 rows=2252 loops=1)
- > Aggregate using temporary table (actual time=56051.460..56052.282 rows=2252 loops=1)

```
| EXPLAIN |
```

```
-> Limit: 15 row(s)   (actual time=52148.384..52148.388 rows=15 loops=1)
```

- > Sort: count_user_id DESC, limit input to 15 row(s) per chunk (actual time=52148.382..52148.385 rows=15 loops=1)
 - > Table scan on <temporary> (actual time=0.002..0.683 rows=2252 loops=1)
 - > Aggregate using temporary table (actual time=52147.081..52147.938 rows=2252 loops=1)
 - > Inner hash join (r2.user_id = r1.user_id), (r2.checkin_date = r1.checkin_date) (cost=598843837.71 rows=619634) (actual time=25153.772..51908.223 rows=224441 loops=1)
- > Filter: (r2.trackable_type = 'Condition') (cost=6809.60rows=7872) (actual time=0.981..24165.746 rows=111517 loops=1)
 - > Table scan on r2 (cost=6809.60 rows=7871685) (actual time=0.974..23149.448 rows=7976223 loops=1)
 - > Hash
 - > Filter: ((r1.trackable_type = 'Symptom') and (r1.trackable_name = 'Anxiety')) (cost=851039.50 rows=78717) (actual time=0.095..24993.066 rows=61602 loops=1)
 - > Table scan on r1 (cost=851039.50 rows=7871685) (actual time=0.072..23647.995 rows=7976223 loops=1)

```
|
```

```
1 row in set (52.18 sec)
```

Indexing #2

```
mysql> CREATE INDEX trackable_id_idx on Records(trackable_id(10));
Query OK, 0 rows affected (49.95 sec)
Records: 0  Duplicates: 0  Warnings: 0
```

```
mysql> EXPLAIN ANALYZE SELECT r2.trackable_name,COUNT(r2.user_id) AS 'count_user_id' FROM Records r1 JOIN Records r2 ON r1.user_id = r2.user_id WHERE r1.trackable_type = 'Symptom' AND r1.trackable_name = 'Anxiety' AND r2.trackable_type = 'Condition' AND r1.checkin_date = r2.checkin_date GROUP BY r2.trackable_name ORDER BY 'count_user_id' DESC LIMIT 15;
```

[illegible]

Indexing #3

```
mysql> CREATE INDEX user_id_idx on Records(user_id(10));
Query OK, 0 rows affected (53.96 sec)
Records: 0  Duplicates: 0  Warnings: 0
```

```
mysql> EXPLAIN ANALYZE SELECT r2.trackable_name,COUNT(r2.user_id) AS `count_user_id` FROM Records r1 JOIN Records r2 ON r1.user_id = r2.user_id WHERE r1.trackable_type = 'Symptom' AND r1.trackable_name = 'Anxiety' AND r2.trackable_type = 'Condition' AND r1.checkin_date = r2.checkin_date GROUP BY r2.trackable_name ORDER BY `count_user_id` DESC LIMIT 15;
```

2. Query to get the number of females and males who have each condition/illness.

```
SELECT f.Illness, f.Num Female, m.Num Male
```

FROM

```

(SELECT trackable_name AS Illness, COUNT(user_id) as Num_Female
FROM Records
WHERE trackable_type = 'Condition' AND sex = 'female'
GROUP BY Illness) f
JOIN
(SELECT trackable_name AS Illness, COUNT(user_id) as Num_Male
FROM Records
WHERE trackable_type = 'Condition' AND sex = 'male'
GROUP BY Illness) m
ON f.Illness = m.Illness
ORDER BY f.Num_Female DESC, m.Num_Male DESC
LIMIT 15;

```

Illness	Num_Female	Num_Male
Fibromyalgia	47641	1981
Depression	38090	3369
Anxiety	36835	2483
Chronic fatigue syndrome	23115	1772
Migraine	21477	1108
IBS	13905	1368
Fatigue	12325	1143
Endometriosis	11887	199
Asthma	11191	806
Ehlers-Danlos syndrome	11053	701
Irritable bowel syndrome	9445	493
Chronic Migraines	8845	115
Headaches	8570	694
Generalized anxiety disorder	8135	563
POTS	7805	126

15 rows in set (56.76 sec)