Report

Assignment 2 - MySQL

Group: 39
Students: Folke Jernbert

Introduction

This assignment was about, first, inserting a dataset into a MySQL database and then querying it for interesting patterns using MySQL-connector. The dataset was a GPS trajectory dataset recording time stamps of people doing activites in and around Beijing from April 2007 to August 2012. I developed the code for the assignment locally and then ssh-ed into a virtual machine provided by IDI for the implementation.

The *set-up* required first that I configured login credentials for the database. Then I used the provided template `DbConnector.py` for connecting to the database. I used this with little changes other than adding environment variables.

The logic for *inserting* the data was done in `insert_data.py`. Three tables, 'User', 'Activity', and 'TrackPoint' were created. In inserting the tables, some data cleaning had to be done such as replacing altitudes set to -777 with NULL-values and ignoring non-valid transportation modes.

Once the database was filled with the dataset, I could start *querying.* Tasks 1 to 6 were performed simply by querying using SQL in the terminal. These queries were relatively simple, generally taking averages and counting occurences. For tasks 7 to 10 Python was used with the connector. The implementation of these queries are provided as attachments with this hand-in. The final, most complicated task, 11, was solved using a complex SQL-query.

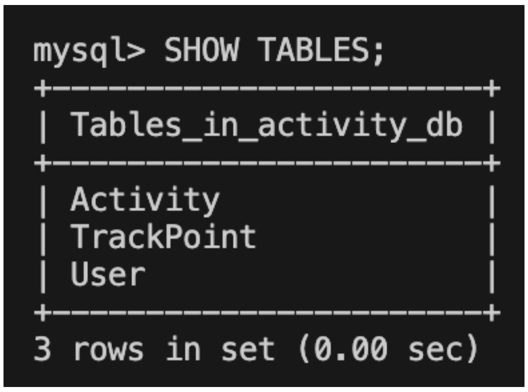
Overall I found the assignment quite enjoyable. A side effect of the project was to learn a couple of tricks in the terminal, like `history`, `!...` and generally how powerful a VM can be when used in combination with git - this enabled me to develop locally and quickly push/pull when I wanted to run scripts. This was a very nice experience. The tasks themselves also gave a feeling of how to approach a dataset during exploratory analysis (EDA), finding counts and averages, checking for invalid points, interesting patterns related to features (for example altitude), and how to formulate interesting inquiries. I have done quite some EDA in Jupyter Notebooks before, using Pandas, but taking such a huge dataset, split up into a folder structure and with labels in a separate text-file, and inserting it into a database before querying it, seems a more, for lack of better words, "production-level solution". It seems the kind of thing data engineers are tasked with before handing over the cleaned database to analysts and data scientists. Getting an experience with such a task, was fun and good learning.

AI-technology like ChatGPT and CoPilot have become integrated and ingrained in my workflow, from project approach, code generation to evaluation. Besides this sparring partner, I worked alone with this project.

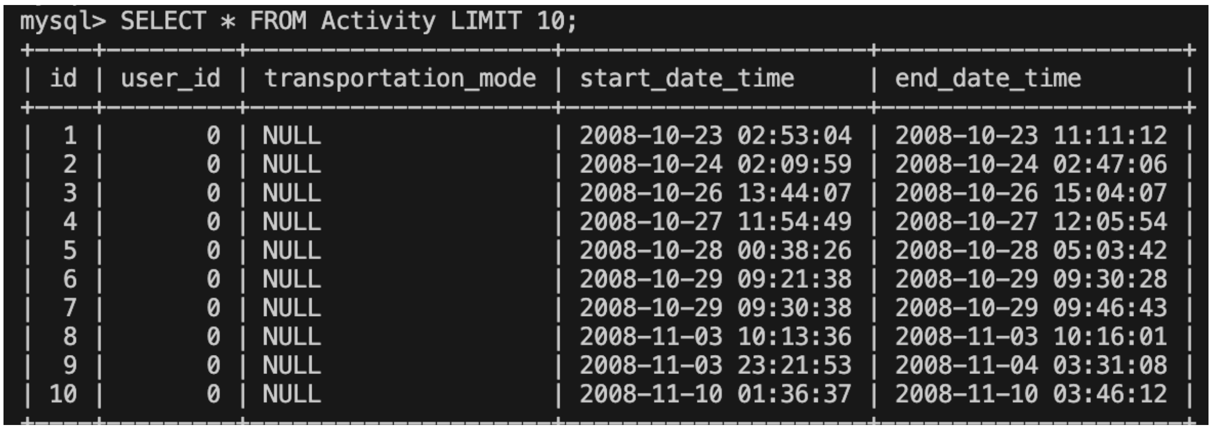
Results

Part 1

Showing tables in the database and top 10 entries for each of the tables.

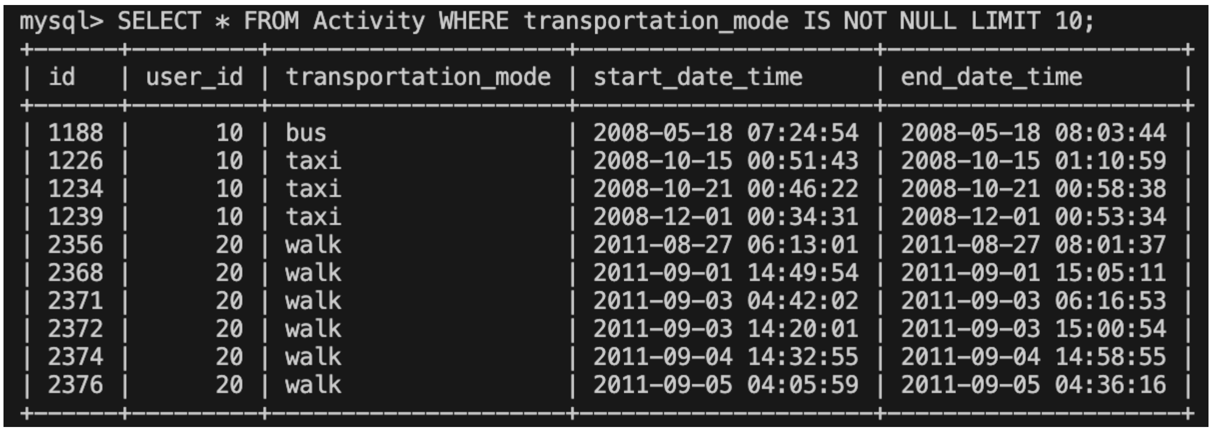


```
mysql> SHOW TABLES;
+------------------------+
| Tables_in_activity_db |
+------------------------+
| Activity               |
| TrackPoint             |
| User                   |
+------------------------+
3 rows in set (0.00 sec)
```

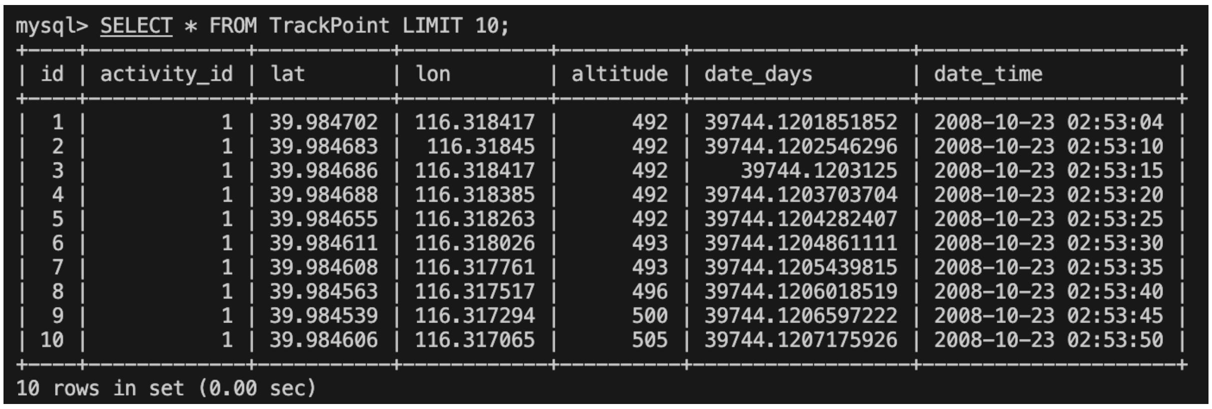


```
mysql> SELECT * FROM Activity LIMIT 10;
+----+---------+---------------------+---------------------+---------------------+
| id | user_id | transportation_mode | start_date_time     | end_date_time       |
+----+---------+---------------------+---------------------+---------------------+
|  1 |       0 | NULL                | 2008-10-23 02:53:04 | 2008-10-23 11:11:12 |
|  2 |       0 | NULL                | 2008-10-24 02:09:59 | 2008-10-24 02:47:06 |
|  3 |       0 | NULL                | 2008-10-26 13:44:07 | 2008-10-26 15:04:07 |
|  4 |       0 | NULL                | 2008-10-27 11:54:49 | 2008-10-27 12:05:54 |
|  5 |       0 | NULL                | 2008-10-28 00:38:26 | 2008-10-28 05:03:42 |
|  6 |       0 | NULL                | 2008-10-29 09:21:38 | 2008-10-29 09:30:28 |
|  7 |       0 | NULL                | 2008-10-29 09:30:38 | 2008-10-29 09:46:43 |
|  8 |       0 | NULL                | 2008-11-03 10:13:36 | 2008-11-03 10:16:01 |
|  9 |       0 | NULL                | 2008-11-03 23:21:53 | 2008-11-04 03:31:08 |
| 10 |       0 | NULL                | 2008-11-10 01:36:37 | 2008-11-10 03:46:12 |
+----+---------+---------------------+---------------------+---------------------+
```

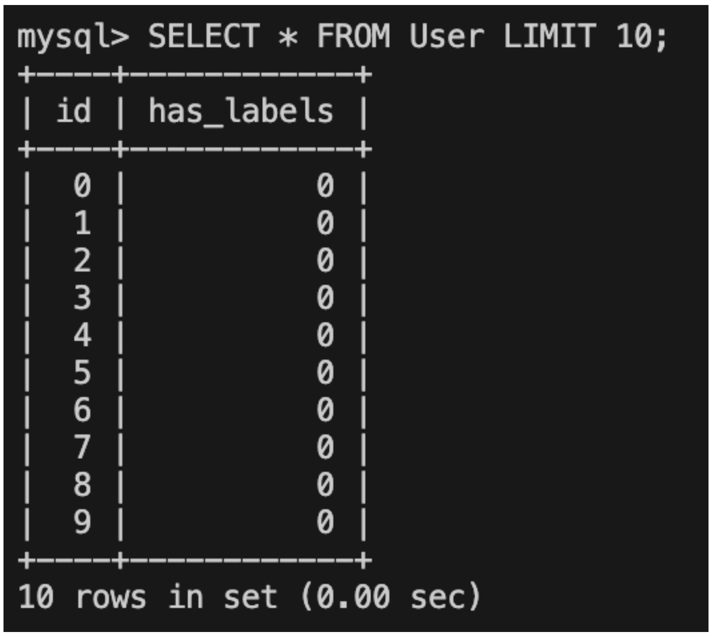
Checking that there in fact are activites with transportation mode not null:



```
mysql> SELECT * FROM Activity WHERE transportation_mode IS NOT NULL LIMIT 10;
+------+---------+---------------------+---------------------+---------------------+
| id   | user_id | transportation_mode | start_date_time     | end_date_time       |
+------+---------+---------------------+---------------------+---------------------+
| 1188 |      10 | bus                 | 2008-05-18 07:24:54 | 2008-05-18 08:03:44 |
| 1226 |      10 | taxi                | 2008-10-15 00:51:43 | 2008-10-15 01:10:59 |
| 1234 |      10 | taxi                | 2008-10-21 00:46:22 | 2008-10-21 00:58:38 |
| 1239 |      10 | taxi                | 2008-12-01 00:34:31 | 2008-12-01 00:53:34 |
| 2356 |      20 | walk                | 2011-08-27 06:13:01 | 2011-08-27 08:01:37 |
| 2368 |      20 | walk                | 2011-09-01 14:49:54 | 2011-09-01 15:05:11 |
| 2371 |      20 | walk                | 2011-09-03 04:42:02 | 2011-09-03 06:16:53 |
| 2372 |      20 | walk                | 2011-09-03 14:20:01 | 2011-09-03 15:00:54 |
| 2374 |      20 | walk                | 2011-09-04 14:32:55 | 2011-09-04 14:58:55 |
| 2376 |      20 | walk                | 2011-09-05 04:05:59 | 2011-09-05 04:36:16 |
+------+---------+---------------------+---------------------+---------------------+
```

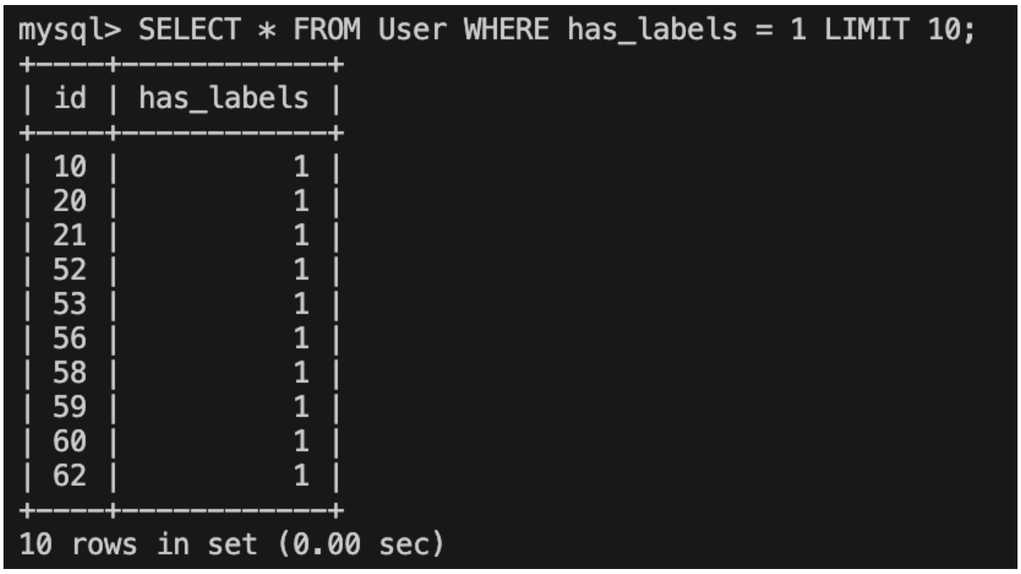


```
mysql> SELECT * FROM TrackPoint LIMIT 10;
+----+-------------+-----------+------------+----------+------------------+---------------------+
| id | activity_id | lat       | lon        | altitude | date_days        | date_time           |
+----+-------------+-----------+------------+----------+------------------+---------------------+
|  1 |           1 | 39.984702 | 116.318417 |      492 | 39744.1201851852 | 2008-10-23 02:53:04 |
|  2 |           1 | 39.984683 |  116.31845 |      492 | 39744.1202546296 | 2008-10-23 02:53:10 |
|  3 |           1 | 39.984686 | 116.318417 |      492 |    39744.1203125 | 2008-10-23 02:53:15 |
|  4 |           1 | 39.984688 | 116.318385 |      492 | 39744.1203703704 | 2008-10-23 02:53:20 |
|  5 |           1 | 39.984655 | 116.318263 |      492 | 39744.1204282407 | 2008-10-23 02:53:25 |
|  6 |           1 | 39.984611 | 116.318026 |      493 | 39744.1204861111 | 2008-10-23 02:53:30 |
|  7 |           1 | 39.984608 | 116.317761 |      493 | 39744.1205439815 | 2008-10-23 02:53:35 |
|  8 |           1 | 39.984563 | 116.317517 |      496 | 39744.1206018519 | 2008-10-23 02:53:40 |
|  9 |           1 | 39.984539 | 116.317294 |      500 | 39744.1206597222 | 2008-10-23 02:53:45 |
| 10 |           1 | 39.984606 | 116.317065 |      505 | 39744.1207175926 | 2008-10-23 02:53:50 |
+----+-------------+-----------+------------+----------+------------------+---------------------+
10 rows in set (0.00 sec)
```



```
mysql> SELECT * FROM User LIMIT 10;
+----+------------+
| id | has_labels |
+----+------------+
|  0 |          0 |
|  1 |          0 |
|  2 |          0 |
|  3 |          0 |
|  4 |          0 |
|  5 |          0 |
|  6 |          0 |
|  7 |          0 |
|  8 |          0 |
|  9 |          0 |
+----+------------+
10 rows in set (0.00 sec)
```

Checking that some ids in fact has labels:

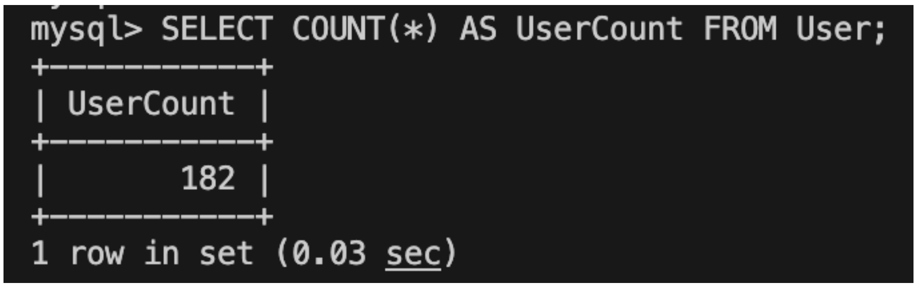


```
mysql> SELECT * FROM User WHERE has_labels = 1 LIMIT 10;
+----+------------+
| id | has_labels |
+----+------------+
| 10 |          1 |
| 20 |          1 |
| 21 |          1 |
| 52 |          1 |
| 53 |          1 |
| 56 |          1 |
| 58 |          1 |
| 59 |          1 |
| 60 |          1 |
| 62 |          1 |
+----+------------+
10 rows in set (0.00 sec)
```

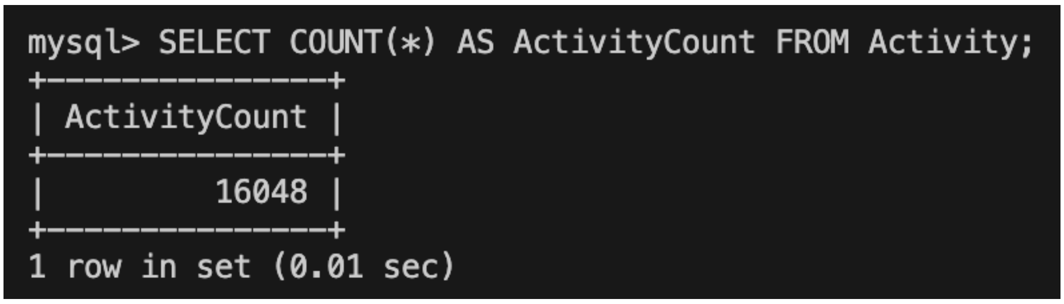
Part 2

Task 1

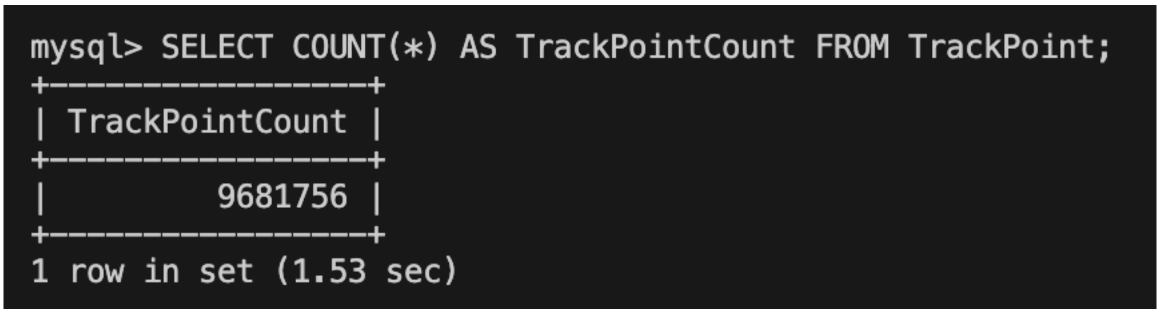
The number of users, activities and trackpoints there are in the dataset.



```
mysql> SELECT COUNT(*) AS UserCount FROM User;
+-----------+
| UserCount |
+-----------+
|       182 |
+-----------+
1 row in set (0.03 sec)
```



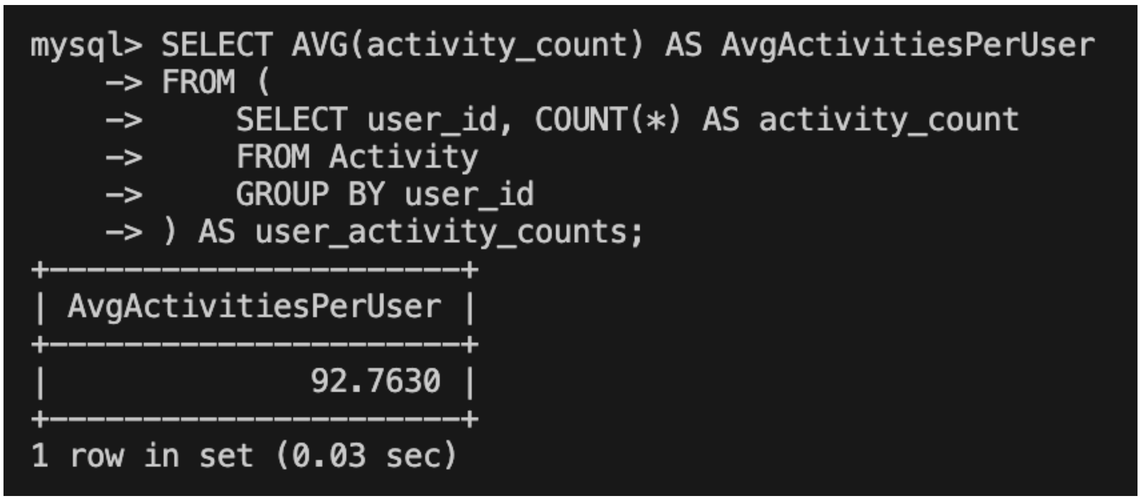
```
mysql> SELECT COUNT(*) AS ActivityCount FROM Activity;
+---------------+
| ActivityCount |
+---------------+
|         16048 |
+---------------+
1 row in set (0.01 sec)
```



```
mysql> SELECT COUNT(*) AS TrackPointCount FROM TrackPoint;
+-----------------+
| TrackPointCount |
+-----------------+
|         9681756 |
+-----------------+
1 row in set (1.53 sec)
```

Task 2

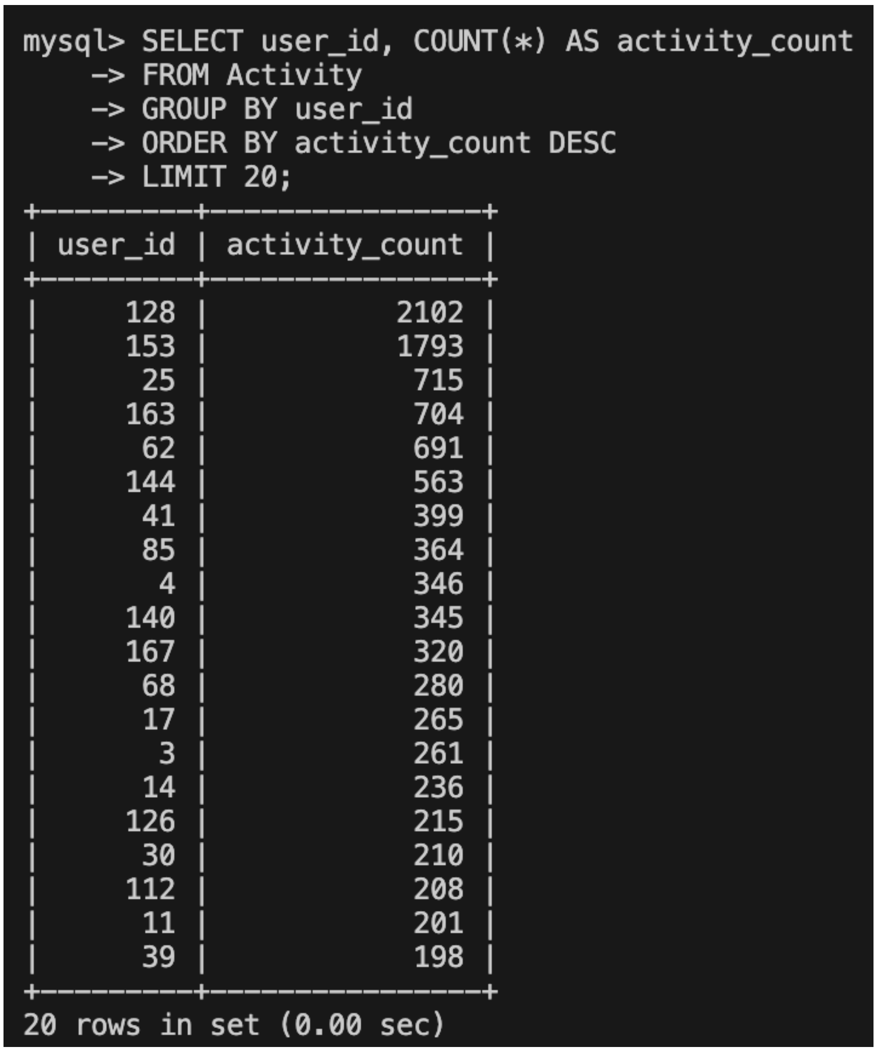
The average number of activities per user.



```
mysql> SELECT AVG(activity_count) AS AvgActivitiesPerUser
    -> FROM (
    ->     SELECT user_id, COUNT(*) AS activity_count
    ->     FROM Activity
    ->     GROUP BY user_id
    -> ) AS user_activity_counts;
+----------------------+
| AvgActivitiesPerUser |
+----------------------+
|              92.7630 |
+----------------------+
1 row in set (0.03 sec)
```

Task 3

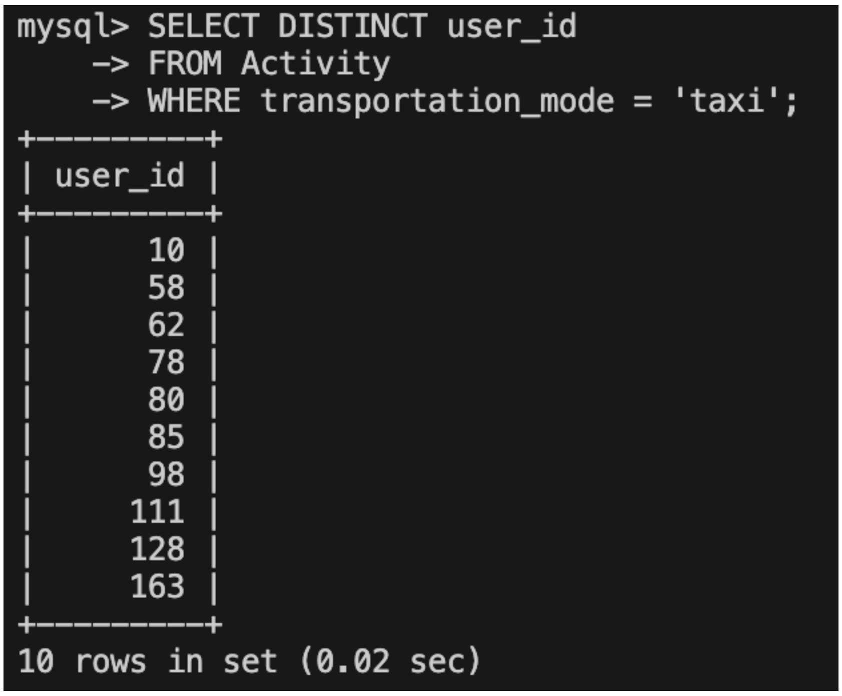
Top 20 users with the highest number of activities.



```
mysql> SELECT user_id, COUNT(*) AS activity_count
    -> FROM Activity
    -> GROUP BY user_id
    -> ORDER BY activity_count DESC
    -> LIMIT 20;
+---------+----------------+
| user_id | activity_count |
+---------+----------------+
|     128 |           2102 |
|     153 |           1793 |
|      25 |            715 |
|     163 |            704 |
|      62 |            691 |
|     144 |            563 |
|      41 |            399 |
|      85 |            364 |
|       4 |            346 |
|     140 |            345 |
|     167 |            320 |
|      68 |            280 |
|      17 |            265 |
|       3 |            261 |
|      14 |            236 |
|     126 |            215 |
|      30 |            210 |
|     112 |            208 |
|      11 |            201 |
|      39 |            198 |
+---------+----------------+
20 rows in set (0.00 sec)
```

Task 4

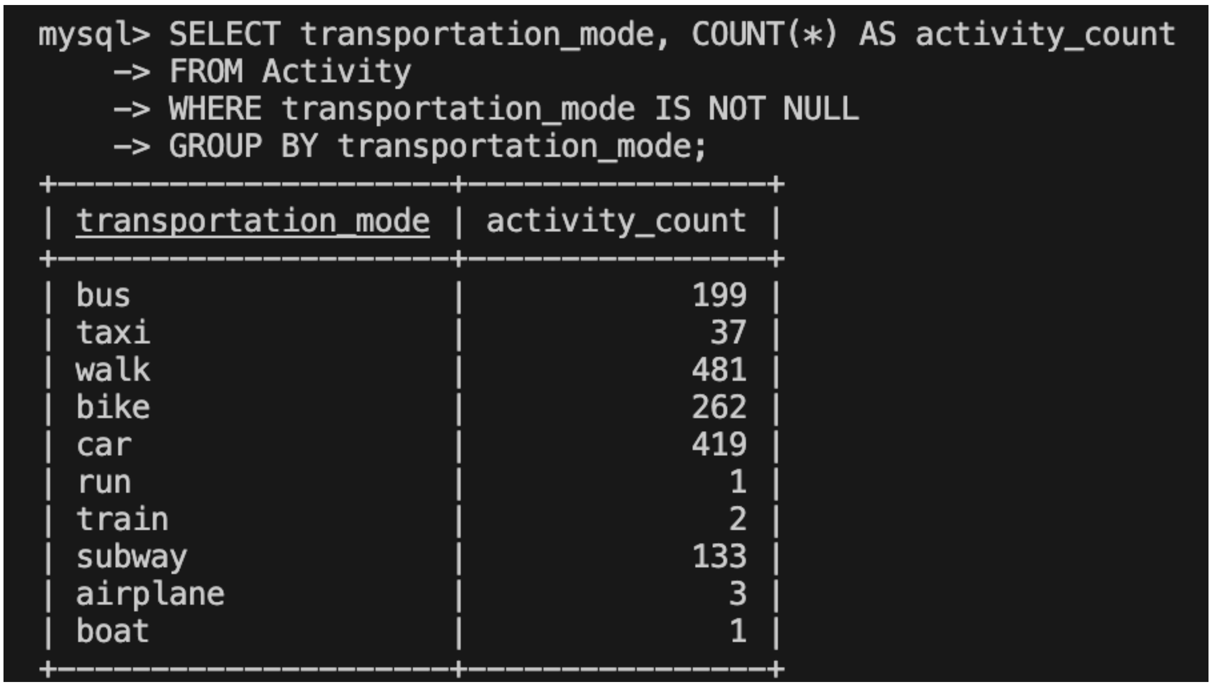
All users who have taken a taxi.



```
mysql> SELECT DISTINCT user_id
    -> FROM Activity
    -> WHERE transportation_mode = 'taxi';
+---------+
| user_id |
+---------+
|      10 |
|      58 |
|      62 |
|      78 |
|      80 |
|      85 |
|      98 |
|     111 |
|     128 |
|     163 |
+---------+
10 rows in set (0.02 sec)
```

Task 5

The activity count of transportation modes.

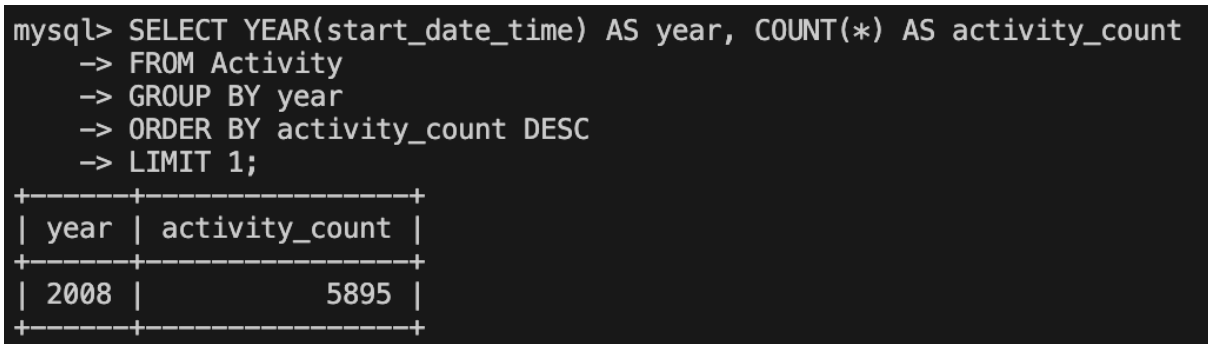


```
mysql> SELECT transportation_mode, COUNT(*) AS activity_count
    -> FROM Activity
    -> WHERE transportation_mode IS NOT NULL
    -> GROUP BY transportation_mode;
+---------------------+----------------+
| transportation_mode | activity_count |
+---------------------+----------------+
| bus                 |            199 |
| taxi                |             37 |
| walk                |            481 |
| bike                |            262 |
| car                 |            419 |
| run                 |              1 |
| train               |              2 |
| subway              |            133 |
| airplane            |              3 |
| boat                |              1 |
+---------------------+----------------+
```

Task 6

a)

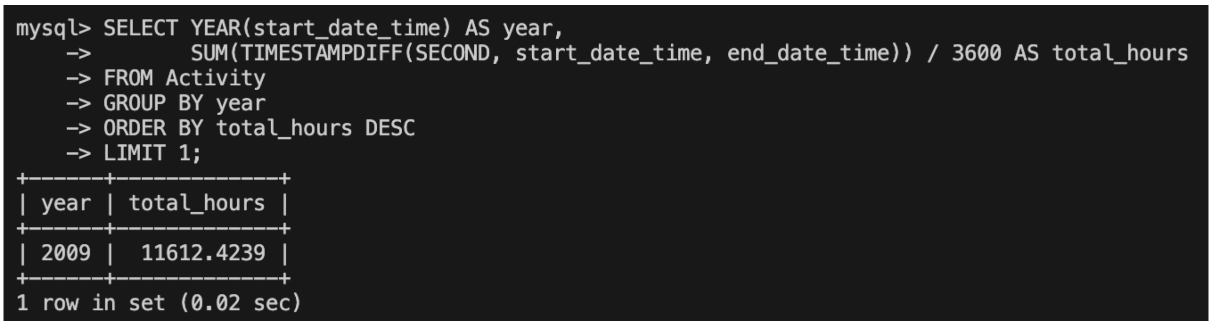
The year with the most activities.



```
mysql> SELECT YEAR(start_date_time) AS year, COUNT(*) AS activity_count
    -> FROM Activity
    -> GROUP BY year
    -> ORDER BY activity_count DESC
    -> LIMIT 1;
+------+----------------+
| year | activity_count |
+------+----------------+
| 2008 |           5895 |
+------+----------------+
```

b)

The year with the most activity recorded hours.

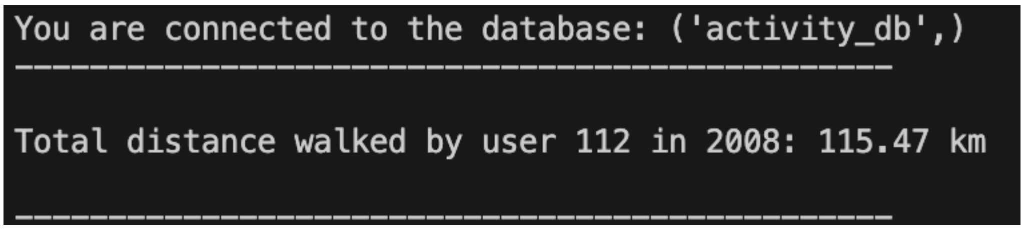


```
mysql> SELECT YEAR(start_date_time) AS year,
    ->        SUM(TIMESTAMPDIFF(SECOND, start_date_time, end_date_time)) / 3600 AS total_hours
    -> FROM Activity
    -> GROUP BY year
    -> ORDER BY total_hours DESC
    -> LIMIT 1;
+------+-------------+
| year | total_hours |
+------+-------------+
| 2009 |  11612.4239 |
+------+-------------+
1 row in set (0.02 sec)
```

Although 2008 is the year with the most activity counts, 2009 is the year with the highest recorded number of hours.

Task 7

Total distance in km walked in 2008 by user 112.
See provided code for implementation (`task7.py`)



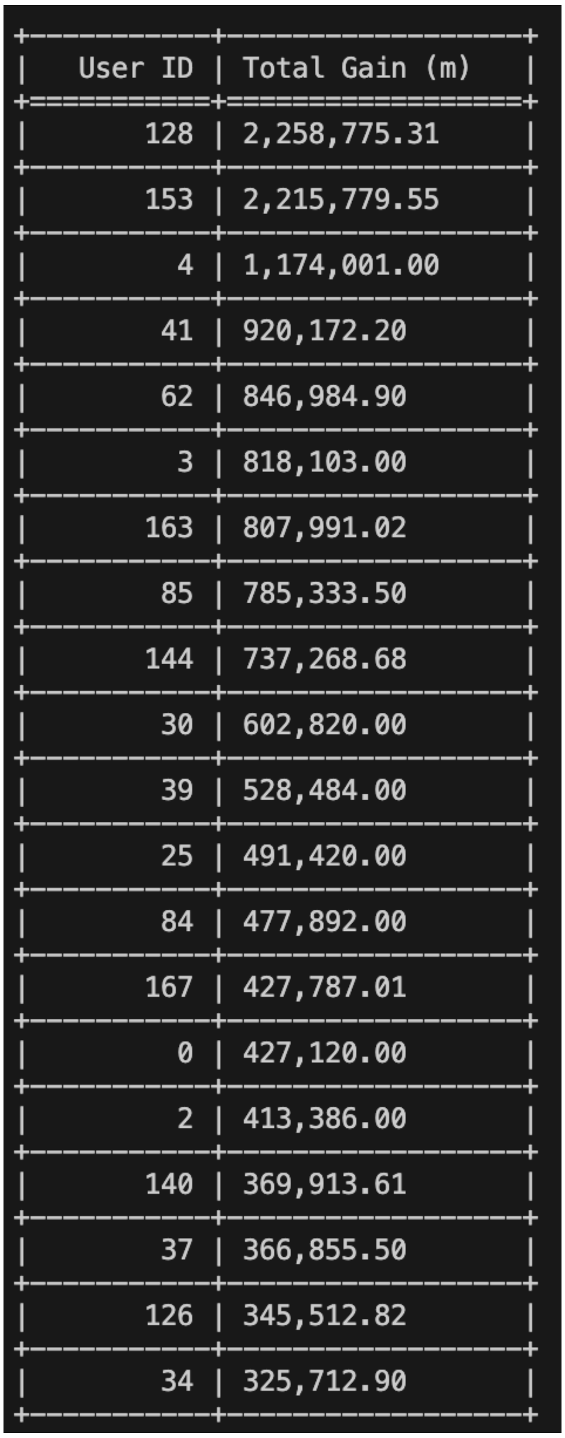
```
You are connected to the database: ('activity_db',)
----------------------------------------------

Total distance walked by user 112 in 2008: 115.47 km

----------------------------------------------
```

Task 8

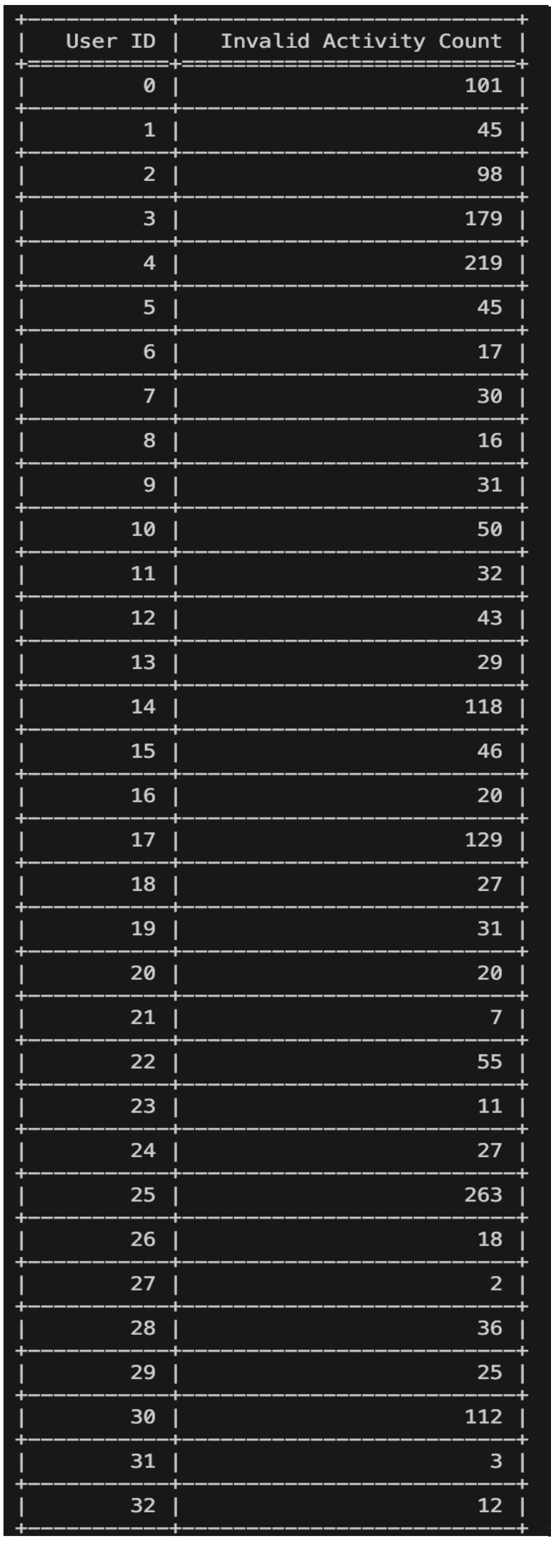
The top 20 users who have gained the most altitude.
See `task8.py` for implementation details.



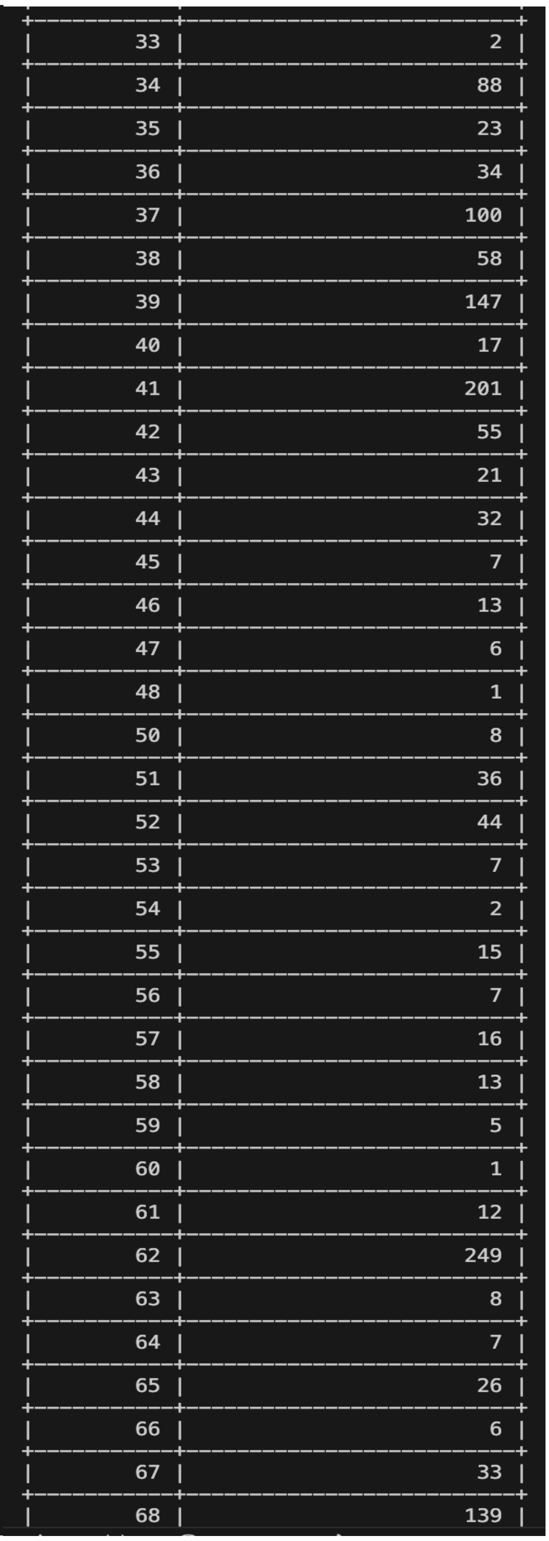
```
+-----------+------------------+
|   User ID | Total Gain (m)   |
+===========+==================+
|       128 | 2,258,775.31     |
+-----------+------------------+
|       153 | 2,215,779.55     |
+-----------+------------------+
|         4 | 1,174,001.00     |
+-----------+------------------+
|        41 | 920,172.20       |
+-----------+------------------+
|        62 | 846,984.90       |
+-----------+------------------+
|         3 | 818,103.00       |
+-----------+------------------+
|       163 | 807,991.02       |
+-----------+------------------+
|        85 | 785,333.50       |
+-----------+------------------+
|       144 | 737,268.68       |
+-----------+------------------+
|        30 | 602,820.00       |
+-----------+------------------+
|        39 | 528,484.00       |
+-----------+------------------+
|        25 | 491,420.00       |
+-----------+------------------+
|        84 | 477,892.00       |
+-----------+------------------+
|       167 | 427,787.01       |
+-----------+------------------+
|         0 | 427,120.00       |
+-----------+------------------+
|         2 | 413,386.00       |
+-----------+------------------+
|       140 | 369,913.61       |
+-----------+------------------+
|        37 | 366,855.50       |
+-----------+------------------+
|       126 | 345,512.82       |
+-----------+------------------+
|        34 | 325,712.90       |
+-----------+------------------+
```

Task 9

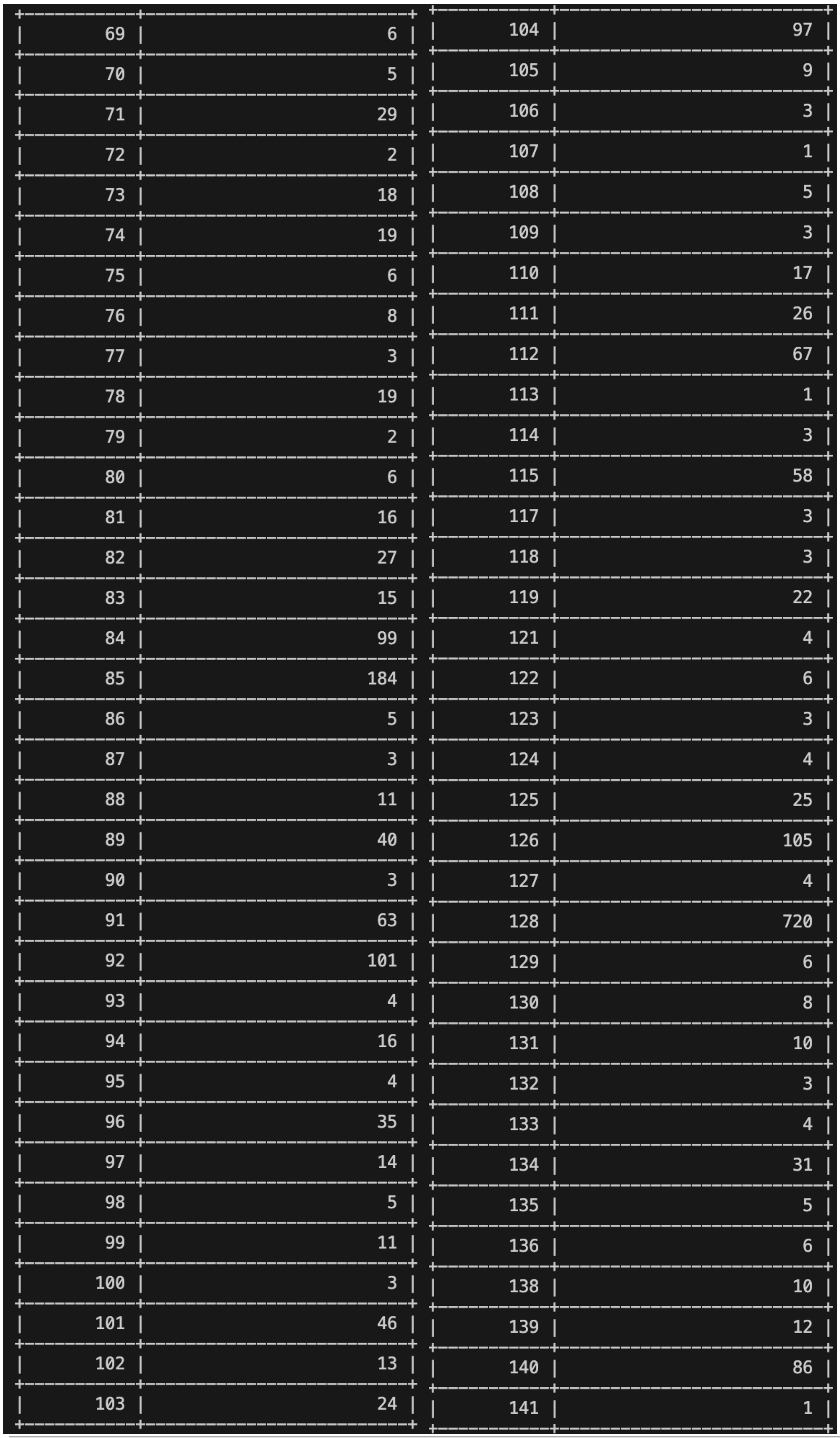
The number of invalid activities per user (listing only users who actually have invalid activities).
Code: `task9.py`



```
+-----------+--------------------------+
|   User ID |   Invalid Activity Count |
+===========+==========================+
|         0 |                      101 |
+-----------+--------------------------+
|         1 |                       45 |
+-----------+--------------------------+
|         2 |                       98 |
+-----------+--------------------------+
|         3 |                      179 |
+-----------+--------------------------+
|         4 |                      219 |
+-----------+--------------------------+
|         5 |                       45 |
+-----------+--------------------------+
|         6 |                       17 |
+-----------+--------------------------+
|         7 |                       30 |
+-----------+--------------------------+
|         8 |                       16 |
+-----------+--------------------------+
|         9 |                       31 |
+-----------+--------------------------+
|        10 |                       50 |
+-----------+--------------------------+
|        11 |                       32 |
+-----------+--------------------------+
|        12 |                       43 |
+-----------+--------------------------+
|        13 |                       29 |
+-----------+--------------------------+
|        14 |                      118 |
+-----------+--------------------------+
|        15 |                       46 |
+-----------+--------------------------+
|        16 |                       20 |
+-----------+--------------------------+
|        17 |                      129 |
+-----------+--------------------------+
|        18 |                       27 |
+-----------+--------------------------+
|        19 |                       31 |
+-----------+--------------------------+
|        20 |                       20 |
+-----------+--------------------------+
|        21 |                        7 |
+-----------+--------------------------+
|        22 |                       55 |
+-----------+--------------------------+
|        23 |                       11 |
+-----------+--------------------------+
|        24 |                       27 |
+-----------+--------------------------+
|        25 |                      263 |
+-----------+--------------------------+
|        26 |                       18 |
+-----------+--------------------------+
|        27 |                        2 |
+-----------+--------------------------+
|        28 |                       36 |
+-----------+--------------------------+
|        29 |                       25 |
+-----------+--------------------------+
|        30 |                      112 |
+-----------+--------------------------+
|        31 |                        3 |
+-----------+--------------------------+
|        32 |                       12 |
+-----------+--------------------------+
```

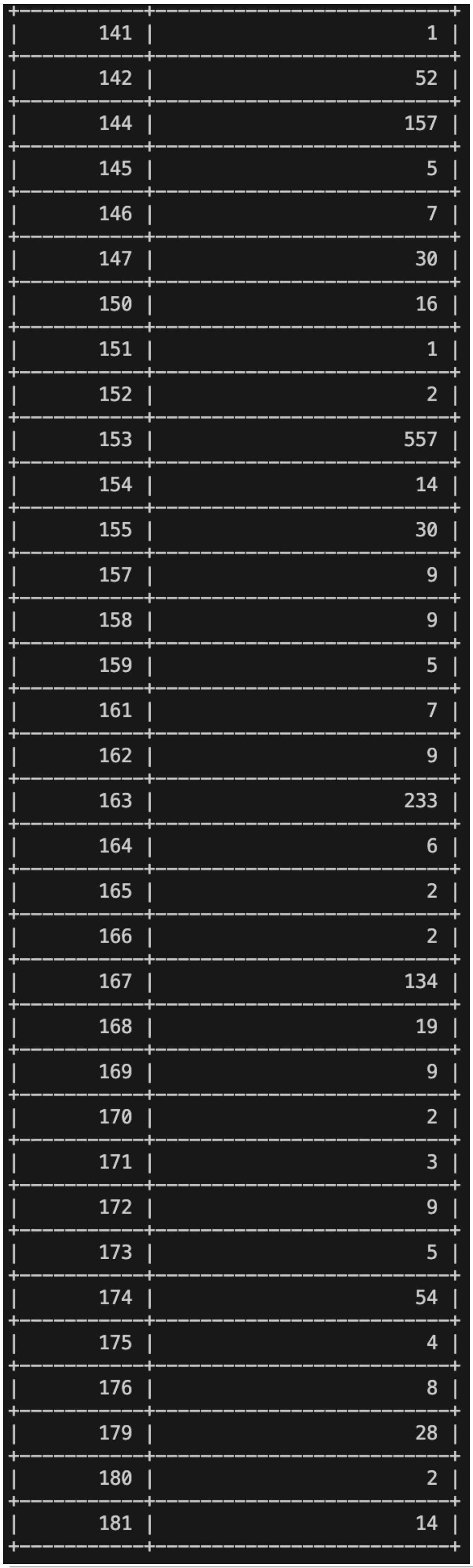


```
+-----------+--------------------------+
|        33 |                        2 |
+-----------+--------------------------+
|        34 |                       88 |
+-----------+--------------------------+
|        35 |                       23 |
+-----------+--------------------------+
|        36 |                       34 |
+-----------+--------------------------+
|        37 |                      100 |
+-----------+--------------------------+
|        38 |                       58 |
+-----------+--------------------------+
|        39 |                      147 |
+-----------+--------------------------+
|        40 |                       17 |
+-----------+--------------------------+
|        41 |                      201 |
+-----------+--------------------------+
|        42 |                       55 |
+-----------+--------------------------+
|        43 |                       21 |
+-----------+--------------------------+
|        44 |                       32 |
+-----------+--------------------------+
|        45 |                        7 |
+-----------+--------------------------+
|        46 |                       13 |
+-----------+--------------------------+
|        47 |                        6 |
+-----------+--------------------------+
|        48 |                        1 |
+-----------+--------------------------+
|        50 |                        8 |
+-----------+--------------------------+
|        51 |                       36 |
+-----------+--------------------------+
|        52 |                       44 |
+-----------+--------------------------+
|        53 |                        7 |
+-----------+--------------------------+
|        54 |                        2 |
+-----------+--------------------------+
|        55 |                       15 |
+-----------+--------------------------+
|        56 |                        7 |
+-----------+--------------------------+
|        57 |                       16 |
+-----------+--------------------------+
|        58 |                       13 |
+-----------+--------------------------+
|        59 |                        5 |
+-----------+--------------------------+
|        60 |                        1 |
+-----------+--------------------------+
|        61 |                       12 |
+-----------+--------------------------+
|        62 |                      249 |
+-----------+--------------------------+
|        63 |                        8 |
+-----------+--------------------------+
|        64 |                        7 |
+-----------+--------------------------+
|        65 |                       26 |
+-----------+--------------------------+
|        66 |                        6 |
+-----------+--------------------------+
|        67 |                       33 |
+-----------+--------------------------+
|        68 |                      139 |
```



```
+-----------+-----------------------------+
|        69 |                           6 |
+-----------+-----------------------------+
|        70 |                           5 |
+-----------+-----------------------------+
|        71 |                          29 |
+-----------+-----------------------------+
|        72 |                           2 |
+-----------+-----------------------------+
|        73 |                          18 |
+-----------+-----------------------------+
|        74 |                          19 |
+-----------+-----------------------------+
|        75 |                           6 |
+-----------+-----------------------------+
|        76 |                           8 |
+-----------+-----------------------------+
|        77 |                           3 |
+-----------+-----------------------------+
|        78 |                          19 |
+-----------+-----------------------------+
|        79 |                           2 |
+-----------+-----------------------------+
|        80 |                           6 |
+-----------+-----------------------------+
|        81 |                          16 |
+-----------+-----------------------------+
|        82 |                          27 |
+-----------+-----------------------------+
|        83 |                          15 |
+-----------+-----------------------------+
|        84 |                          99 |
+-----------+-----------------------------+
|        85 |                         184 |
+-----------+-----------------------------+
|        86 |                           5 |
+-----------+-----------------------------+
|        87 |                           3 |
+-----------+-----------------------------+
|        88 |                          11 |
+-----------+-----------------------------+
|        89 |                          40 |
+-----------+-----------------------------+
|        90 |                           3 |
+-----------+-----------------------------+
|        91 |                          63 |
+-----------+-----------------------------+
|        92 |                         101 |
+-----------+-----------------------------+
|        93 |                           4 |
+-----------+-----------------------------+
|        94 |                          16 |
+-----------+-----------------------------+
|        95 |                           4 |
+-----------+-----------------------------+
|        96 |                          35 |
+-----------+-----------------------------+
|        97 |                          14 |
+-----------+-----------------------------+
|        98 |                           5 |
+-----------+-----------------------------+
|        99 |                          11 |
+-----------+-----------------------------+
|       100 |                           3 |
+-----------+-----------------------------+
|       101 |                          46 |
+-----------+-----------------------------+
|       102 |                          13 |
+-----------+-----------------------------+
|       103 |                          24 |
+-----------+-----------------------------+
```

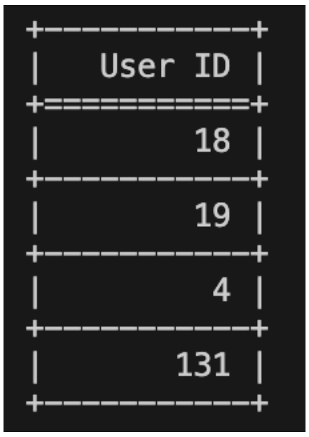
```
+-----------+-----------------------------+
|       104 |                          97 |
+-----------+-----------------------------+
|       105 |                           9 |
+-----------+-----------------------------+
|       106 |                           3 |
+-----------+-----------------------------+
|       107 |                           1 |
+-----------+-----------------------------+
|       108 |                           5 |
+-----------+-----------------------------+
|       109 |                           3 |
+-----------+-----------------------------+
|       110 |                          17 |
+-----------+-----------------------------+
|       111 |                          26 |
+-----------+-----------------------------+
|       112 |                          67 |
+-----------+-----------------------------+
|       113 |                           1 |
+-----------+-----------------------------+
|       114 |                           3 |
+-----------+-----------------------------+
|       115 |                          58 |
+-----------+-----------------------------+
|       117 |                           3 |
+-----------+-----------------------------+
|       118 |                           3 |
+-----------+-----------------------------+
|       119 |                          22 |
+-----------+-----------------------------+
|       121 |                           4 |
+-----------+-----------------------------+
|       122 |                           6 |
+-----------+-----------------------------+
|       123 |                           3 |
+-----------+-----------------------------+
|       124 |                           4 |
+-----------+-----------------------------+
|       125 |                          25 |
+-----------+-----------------------------+
|       126 |                         105 |
+-----------+-----------------------------+
|       127 |                           4 |
+-----------+-----------------------------+
|       128 |                         720 |
+-----------+-----------------------------+
|       129 |                           6 |
+-----------+-----------------------------+
|       130 |                           8 |
+-----------+-----------------------------+
|       131 |                          10 |
+-----------+-----------------------------+
|       132 |                           3 |
+-----------+-----------------------------+
|       133 |                           4 |
+-----------+-----------------------------+
|       134 |                          31 |
+-----------+-----------------------------+
|       135 |                           5 |
+-----------+-----------------------------+
|       136 |                           6 |
+-----------+-----------------------------+
|       138 |                          10 |
+-----------+-----------------------------+
|       139 |                          12 |
+-----------+-----------------------------+
|       140 |                          86 |
+-----------+-----------------------------+
|       141 |                           1 |
+-----------+-----------------------------+
```



```
+------------+------------------------------+
|        141 |                            1 |
+------------+------------------------------+
|        142 |                           52 |
+------------+------------------------------+
|        144 |                          157 |
+------------+------------------------------+
|        145 |                            5 |
+------------+------------------------------+
|        146 |                            7 |
+------------+------------------------------+
|        147 |                           30 |
+------------+------------------------------+
|        150 |                           16 |
+------------+------------------------------+
|        151 |                            1 |
+------------+------------------------------+
|        152 |                            2 |
+------------+------------------------------+
|        153 |                          557 |
+------------+------------------------------+
|        154 |                           14 |
+------------+------------------------------+
|        155 |                           30 |
+------------+------------------------------+
|        157 |                            9 |
+------------+------------------------------+
|        158 |                            9 |
+------------+------------------------------+
|        159 |                            5 |
+------------+------------------------------+
|        161 |                            7 |
+------------+------------------------------+
|        162 |                            9 |
+------------+------------------------------+
|        163 |                          233 |
+------------+------------------------------+
|        164 |                            6 |
+------------+------------------------------+
|        165 |                            2 |
+------------+------------------------------+
|        166 |                            2 |
+------------+------------------------------+
|        167 |                          134 |
+------------+------------------------------+
|        168 |                           19 |
+------------+------------------------------+
|        169 |                            9 |
+------------+------------------------------+
|        170 |                            2 |
+------------+------------------------------+
|        171 |                            3 |
+------------+------------------------------+
|        172 |                            9 |
+------------+------------------------------+
|        173 |                            5 |
+------------+------------------------------+
|        174 |                           54 |
+------------+------------------------------+
|        175 |                            4 |
+------------+------------------------------+
|        176 |                            8 |
+------------+------------------------------+
|        179 |                           28 |
+------------+------------------------------+
|        180 |                            2 |
+------------+------------------------------+
|        181 |                           14 |
+------------+------------------------------+
```

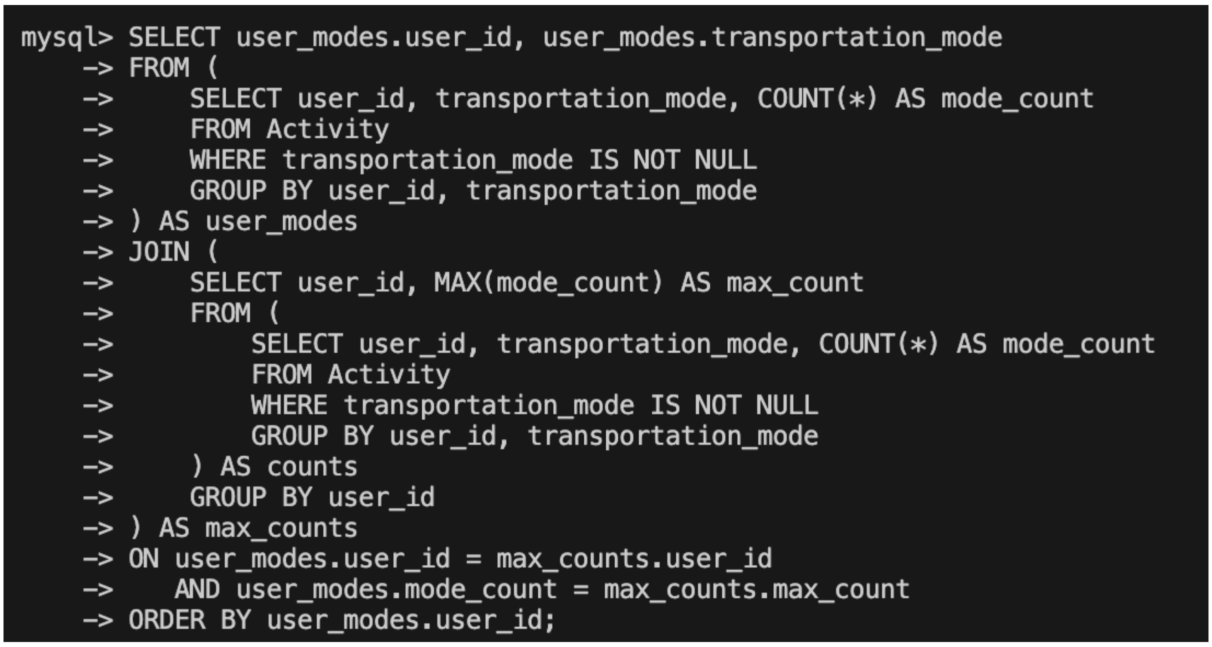
Task 10

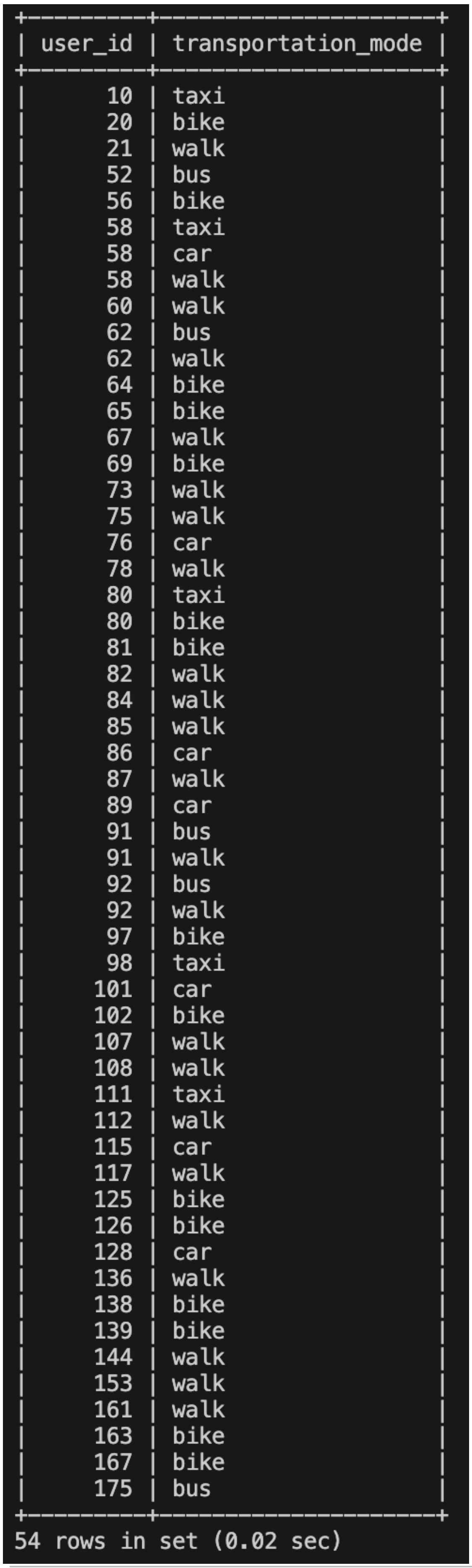
Users who have been in the Forbidden City of Beijing (within a radius of 100 meters of the coordinates lat: 39.916, lon: 116.397).


```
+------------+
|    User ID |
+============+
|         18 |
+------------+
|         19 |
+------------+
|          4 |
+------------+
|        131 |
+------------+
```

Task 11

The most used transportation mode of users who have registered it.


```
mysql> SELECT user_modes.user_id, user_modes.transportation_mode
    -> FROM (
    ->     SELECT user_id, transportation_mode, COUNT(*) AS mode_count
    ->     FROM Activity
    ->     WHERE transportation_mode IS NOT NULL
    ->     GROUP BY user_id, transportation_mode
    -> ) AS user_modes
    -> JOIN (
    ->     SELECT user_id, MAX(mode_count) AS max_count
    ->     FROM (
    ->         SELECT user_id, transportation_mode, COUNT(*) AS mode_count
    ->         FROM Activity
    ->         WHERE transportation_mode IS NOT NULL
    ->         GROUP BY user_id, transportation_mode
    ->     ) AS counts
    ->     GROUP BY user_id
    -> ) AS max_counts
    -> ON user_modes.user_id = max_counts.user_id
    ->    AND user_modes.mode_count = max_counts.max_count
    -> ORDER BY user_modes.user_id;
```



```
+---------+---------------------+
| user_id | transportation_mode |
+---------+---------------------+
|      10 | taxi                |
|      20 | bike                |
|      21 | walk                |
|      52 | bus                 |
|      56 | bike                |
|      58 | taxi                |
|      58 | car                 |
|      58 | walk                |
|      60 | walk                |
|      62 | bus                 |
|      62 | walk                |
|      64 | bike                |
|      65 | bike                |
|      67 | walk                |
|      69 | bike                |
|      73 | walk                |
|      75 | walk                |
|      76 | car                 |
|      78 | walk                |
|      80 | taxi                |
|      80 | bike                |
|      81 | bike                |
|      82 | walk                |
|      84 | walk                |
|      85 | walk                |
|      86 | car                 |
|      87 | walk                |
|      89 | car                 |
|      91 | bus                 |
|      91 | walk                |
|      92 | bus                 |
|      92 | walk                |
|      97 | bike                |
|      98 | taxi                |
|     101 | car                 |
|     102 | bike                |
|     107 | walk                |
|     108 | walk                |
|     111 | taxi                |
|     112 | walk                |
|     115 | car                 |
|     117 | walk                |
|     125 | bike                |
|     126 | bike                |
|     128 | car                 |
|     136 | walk                |
|     138 | bike                |
|     139 | bike                |
|     144 | walk                |
|     153 | walk                |
|     161 | walk                |
|     163 | bike                |
|     167 | bike                |
|     175 | bus                 |
+---------+---------------------+
54 rows in set (0.02 sec)
```

Discussion

In general I find the assignment went along quite smoothly. There were some initial hick-ups with opening the ports in the configuration file of MySQL, but this issue was quickly resolved. Syncing with GitHub was also in the beginning something of a pain point. My quota of Git LFS was used up, and so I couldn't transfer the data files using that solution. Resetting the commit took some time, but eventually I was able to add the datafiles to the .gitignore-file, and I used a file server to transfer the data files to the VM.

As for the tasks, I found it nice to brush up on some SQL querying from this spring. While the basics, like selecting, grouping by, and aggregating, feel quite comfortable, things was more difficult with constructing nested queries such as in task 11.

The concept of the MySQL-connector in Python was not entirely new to me, as I in the database course previous semester used an ORM with Flask in the main project. A difference, however, is that I in this task wrote SQL-queries directly into the Python script.

As briefly alluded to in the introduction, I found the tasks quite relevant for real-world scenarios, where typical tasks of the data handler are data cleaning, inserting data into a database, checking for basic statistics, and finding some basic insights like who has been in the Forbidden City. This is a nice skill to have, and I'm glad I've gotten some practice in that through this project.