

# TDT4225 Report

## Assignment 2 - MySQL

**Group: 11**

**Students:** Elias Wendt, Marianne Hernholm and Taheera Ahmed

### Introduction

In this assignment we were tasked with implementing a database using MySQL. We first had to insert data from a dataset given an ER-diagram. Our solution for this can be found in **part1.py**. Afterwards we had to query the database, this part can be found in **part2.py**.

As a group we worked together mainly through remote channels. We organized work-sessions via zoom and utilized tools such as visual studio's live-share mechanism. We made a common github repository which can be found at [https://github.com/taheeraahmed/tdt4225/tree/main/Exercise\\_2](https://github.com/taheeraahmed/tdt4225/tree/main/Exercise_2).

### Results

#### Part 1:

```
mysql> SELECT * FROM User LIMIT 10;
```

id	has_labels
000	0
001	0
002	0
003	0
004	0
005	0
006	0
007	0
008	0
009	0

```
10 rows in set (0.00 sec)
```

```
mysql> SELECT * FROM Activity LIMIT 10;
```

id	user_id	transportation_mode	start_date_time	end_date_time
1	135	NULL	2009-01-03 01:21:34	2009-01-03 05:40:31
2	135	NULL	2009-01-02 04:31:27	2009-01-02 04:41:05
3	135	NULL	2009-01-27 03:00:04	2009-01-27 04:50:32
4	135	NULL	2009-01-10 01:19:47	2009-01-10 04:42:47
5	135	NULL	2008-12-27 07:26:04	2008-12-27 12:28:24
6	135	NULL	2009-01-01 02:44:58	2009-01-01 08:25:58
7	135	NULL	2009-01-14 12:17:57	2009-01-14 12:30:53
8	135	NULL	2009-01-12 01:41:22	2009-01-12 02:14:01
9	135	NULL	2009-01-24 06:51:03	2009-01-25 06:01:20
10	135	NULL	2009-01-26 02:26:33	2009-01-26 15:54:36

```
10 rows in set (0.00 sec)
```

```
mysql> SELECT * FROM TrackPoint LIMIT 10;
```

id	activity_id	lat	lon	altitude	date_days	date_time
1	1	39.974292	116.399592	492	39816.0566550926	2009-01-03 01:21:35
2	1	39.974309	116.399523	492	39816.0566666667	2009-01-03 01:21:36
3	1	39.97432	116.399588	492	39816.0566898148	2009-01-03 01:21:38
4	1	39.974365	116.39973	491	39816.0567013889	2009-01-03 01:21:39
5	1	39.974391	116.399782	491	39816.0567361111	2009-01-03 01:21:42
6	1	39.974426	116.399735	491	39816.0567824074	2009-01-03 01:21:46
7	1	39.974458	116.3997	491	39816.0568402778	2009-01-03 01:21:51
8	1	39.974491	116.399732	490	39816.0568981481	2009-01-03 01:21:56
9	1	39.97453	116.399758	489	39816.0569560185	2009-01-03 01:22:01
10	1	39.974577	116.399745	490	39816.0569907407	2009-01-03 01:22:04

```
10 rows in set (0.00 sec)
```

## Part 2:

### Query 1

```

+---+---+---+---+ +---+
|Q|u|e|r|y| |1|
+---+---+---+---+ +---+

#No. of users: 182
#No. of activities: 18669
#No of trackpoints: 9698243

```

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

### Query 2

```

+---+---+---+---+ +---+
|Q|u|e|r|y| |2|
+---+---+---+---+ +---+

Average: 102.5769
Minimum: 1
Max: 2153

```

The screenshot shows the result of the average, minimum and maximum number of activities per user.

### Query 3

```

-----
+---+---+---+---+ +---+
|Q|u|e|r|y| |3|
+---+---+---+---+ +---+

Rank 1: user_id: 128 #No. activities: 2153
Rank 2: user_id: 153 #No. activities: 2024
Rank 3: user_id: 163 #No. activities: 809
Rank 4: user_id: 025 #No. activities: 757
Rank 5: user_id: 062 #No. activities: 706
Rank 6: user_id: 144 #No. activities: 610
Rank 7: user_id: 041 #No. activities: 557
Rank 8: user_id: 085 #No. activities: 427
Rank 9: user_id: 068 #No. activities: 408
Rank 10: user_id: 004 #No. activities: 395
-----

```

The screenshot shows the 10 users with the highest numbers of activities ordered by.

### Query 4

```

+---+---+---+---+ +---+
|Q|u|e|r|y| |4|
+---+---+---+---+ +---+

#No. of users: 1588
-----

```

The screenshot shows the result of the number of users who have started the activity in one day and finished it the following day.

### Query 5

```

-----
+---+---+---+---+ +---+
|Q|u|e|r|y| |5|
+---+---+---+---+ +---+

There are no rows in the activity table which matches this query
-----

```

The screenshot shows the result of the activities that are registered multiple times. This query will check for duplicate entries/rows given a user and an activity

### Query 6

We weren't able to run this query, because of the crazy runtime.

## Query 7

```

+-----+
|Q|u|e|r|y| |7|
+-----+

```

The following users have never taken a taxi:

```

000, 001, 002, 003, 004, 005, 006, 007, 008, 009, 011, 012, 013, 014, 015, 016, 017, 018, 019, 020, 021, 022, 023, 024, 025, 026, 027, 028, 029, 030, 031, 032, 033, 034, 035, 036, 037, 038, 039, 040, 041, 042, 043, 044, 045, 046, 047, 048, 049, 050, 051, 052, 053, 054, 055, 056, 057, 059, 060, 061, 063, 064, 066, 067, 068, 069, 070, 071, 072, 073, 074, 075, 076, 077, 079, 081, 082, 083, 084, 086, 087, 088, 089, 090, 091, 092, 093, 094, 095, 096, 097, 099, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181,
There are 171 users who have never taken a taxi

```

## Query 8

```

+-----+
|Q|u|e|r|y| |8|
+-----+

```

```

Transportation mode: airplane #No. distinct users: 1
Transportation mode: bike #No. distinct users: 19
Transportation mode: boat #No. distinct users: 1
Transportation mode: bus #No. distinct users: 13
Transportation mode: car #No. distinct users: 9
Transportation mode: None #No. distinct users: 182
Transportation mode: run #No. distinct users: 1
Transportation mode: subway #No. distinct users: 6
Transportation mode: taxi #No. distinct users: 11
Transportation mode: train #No. distinct users: 2
Transportation mode: walk #No. distinct users: 35

```

The screenshot shows the result of the different transportation modes and how many distinct users have used the different transportation modes.

## Query 9

```

+-----+
|Q|u|e|r|y| |9|
+-----+

```

a: Year: 2008 Month: 11 #No. Activities: 1168

b: user\_id with the most activity: 062, n\_activities: 130, recorded hours: 0.1255  
user\_id with the seconds most activity: 014, n\_activities: 81, recorded hours: 21.8310

User with most activities has more recorded hours than user with second most activities: False

This query requires a single point in time to be used for evaluation. In our implementation we decided to use the field **start\_date\_time**.

## Query 10

```

-----
+-+--+--+--+ +-+--+
|Q|u|e|r|y| |1|0|
+-+--+--+--+ +-+--+

```

Total distance walked by user 112 in 2008 was 8860.01449472005 km

## Query 11

Query 11:

```

Rank 1: user_id: 128, gained altitude: 320753928
Rank 2: user_id: 144, gained altitude: 154950971
Rank 3: user_id: 153, gained altitude: 55022457
Rank 4: user_id: 041, gained altitude: 48901441
Rank 5: user_id: 062, gained altitude: 39684129
Rank 6: user_id: 163, gained altitude: 33812450
Rank 7: user_id: 085, gained altitude: 22330487
Rank 8: user_id: 115, gained altitude: 19230688
Rank 9: user_id: 030, gained altitude: 18180406
Rank 10: user_id: 025, gained altitude: 15566490
Rank 11: user_id: 004, gained altitude: 15015898
Rank 12: user_id: 037, gained altitude: 14104188
Rank 13: user_id: 168, gained altitude: 13030684
Rank 14: user_id: 003, gained altitude: 12296961
Rank 15: user_id: 140, gained altitude: 12294186
Rank 16: user_id: 017, gained altitude: 10979265
Rank 17: user_id: 002, gained altitude: 9497397
Rank 18: user_id: 039, gained altitude: 9160365
Rank 19: user_id: 084, gained altitude: 9012634
Rank 20: user_id: 034, gained altitude: 8073763

```



## Query 12

```

|Q|u|e|r|y| |1|2|
+-----+
user_id=135,#invalid_activities=8 - user_id=132,#invalid_activities=3 - user_id=104,#invalid_activities=108 - user_id=103,#invalid_activities=44 - u
ser_id=168,#invalid_activities=80 - user_id=157,#invalid_activities=13 - user_id=150,#invalid_activities=21 - user_id=159,#invalid_activities=7 - us
er_id=166,#invalid_activities=3 - user_id=161,#invalid_activities=13 - user_id=102,#invalid_activities=30 - user_id=105,#invalid_activities=10 - use
r_id=133,#invalid_activities=4 - user_id=134,#invalid_activities=62 - user_id=158,#invalid_activities=12 - user_id=167,#invalid_activities=310 - use
r_id=151,#invalid_activities=1 - user_id=169,#invalid_activities=34 - user_id=024,#invalid_activities=42 - user_id=023,#invalid_activities=16 - use
r_id=015,#invalid_activities=58 - user_id=012,#invalid_activities=67 - user_id=079,#invalid_activities=23 - user_id=046,#invalid_activities=31 - user
_id=041,#invalid_activities=369 - user_id=048,#invalid_activities=1 - user_id=077,#invalid_activities=3 - user_id=083,#invalid_activities=27 - user_
id=084,#invalid_activities=161 - user_id=070,#invalid_activities=6 - user_id=013,#invalid_activities=105 - user_id=014,#invalid_activities=225 - use
r_id=022,#invalid_activities=77 - user_id=025,#invalid_activities=619 - user_id=071,#invalid_activities=64 - user_id=085,#invalid_activities=353 - u
ser_id=082,#invalid_activities=69 - user_id=076,#invalid_activities=11 - user_id=040,#invalid_activities=20 - user_id=078,#invalid_activities=79 - u
ser_id=047,#invalid_activities=10 - user_id=065,#invalid_activities=56 - user_id=091,#invalid_activities=91 - user_id=096,#invalid_activities=85 - u
ser_id=062,#invalid_activities=564 - user_id=054,#invalid_activities=2 - user_id=053,#invalid_activities=8 - user_id=098,#invalid_activities=6 - use
r_id=038,#invalid_activities=67 - user_id=007,#invalid_activities=37 - user_id=000,#invalid_activities=145 - user_id=008,#invalid_activities=37 - us
er_id=036,#invalid_activities=41 - user_id=031,#invalid_activities=3 - user_id=052,#invalid_activities=54 - user_id=099,#invalid_activities=14 - us
r_id=055,#invalid_activities=17 - user_id=063,#invalid_activities=13 - user_id=097,#invalid_activities=31 - user_id=090,#invalid_activities=7 - user
_id=064,#invalid_activities=16 - user_id=030,#invalid_activities=199 - user_id=008,#invalid_activities=23 - user_id=037,#invalid_activities=126 - us
er_id=001,#invalid_activities=55 - user_id=039,#invalid_activities=195 - user_id=006,#invalid_activities=22 - user_id=174,#invalid_activities=67 - u
ser_id=001,#invalid_activities=2 - user_id=173,#invalid_activities=6 - user_id=145,#invalid_activities=5 - user_id=142,#invalid_activities=135 - u
r_id=129,#invalid_activities=6 - user_id=116,#invalid_activities=3 - user_id=111,#invalid_activities=42 - user_id=118,#invalid_activities=4 - user_i
d=127,#invalid_activities=7 - user_id=144,#invalid_activities=431 - user_id=172,#invalid_activities=12 - user_id=181,#invalid_activities=15 - user_i
d=175,#invalid_activities=4 - user_id=121,#invalid_activities=5 - user_id=119,#invalid_activities=31 - user_id=126,#invalid_activities=202 - user_id
=110,#invalid_activities=23 - user_id=128,#invalid_activities=1999 - user_id=117,#invalid_activities=8 - user_id=153,#invalid_activities=1714 - user
_id=154,#invalid_activities=29 - user_id=162,#invalid_activities=10 - user_id=165,#invalid_activities=13 - user_id=131,#invalid_activities=15 - user
_id=136,#invalid_activities=8 - user_id=109,#invalid_activities=4 - user_id=100,#invalid_activities=7 - user_id=107,#invalid_activities=3 - user_id=
138,#invalid_activities=17 - user_id=164,#invalid_activities=7 - user_id=163,#invalid_activities=681 - user_id=155,#invalid_activities=38 - user_id=
152,#invalid_activities=2 - user_id=106,#invalid_activities=3 - user_id=139,#invalid_activities=15 - user_id=101,#invalid_activities=63 - user_id=10
8,#invalid_activities=6 - user_id=130,#invalid_activities=14 - user_id=089,#invalid_activities=59 - user_id=042,#invalid_activities=96 - user_id=045
,#invalid_activities=8 - user_id=087,#invalid_activities=7 - user_id=073,#invalid_activities=66 - user_id=074,#invalid_activities=79 - user_id=080,#i
nvalid_activities=9 - user_id=020,#invalid_activities=124 - user_id=027,#invalid_activities=3 - user_id=019,#invalid_activities=39 - user_id=011,#i
nvalid_activities=173 - user_id=016,#invalid_activities=33 - user_id=029,#invalid_activities=39 - user_id=081,#invalid_activities=25 - user_id=075,#i
nvalid_activities=7 - user_id=072,#invalid_activities=2 - user_id=086,#invalid_activities=6 - user_id=044,#invalid_activities=54 - user_id=088,#inv
alid_activities=48 - user_id=043,#invalid_activities=31 - user_id=017,#invalid_activities=252 - user_id=028,#invalid_activities=51 - user_id=010,#in
valid_activities=77 - user_id=026,#invalid_activities=20 - user_id=019,#invalid_activities=68 - user_id=021,#invalid_activities=7 - user_id=003,#inv
alid_activities=253 - user_id=004,#invalid_activities=339 - user_id=032,#invalid_activities=14 - user_id=035,#invalid_activities=24 - user_id=095,#i
nvalid_activities=30 - user_id=061,#invalid_activities=19 - user_id=066,#invalid_activities=7 - user_id=092,#invalid_activities=143 - user_id=059,#i
nvalid_activities=5 - user_id=050,#invalid_activities=6 - user_id=057,#invalid_activities=21 - user_id=068,#invalid_activities=268 - user_id=034,#in
valid_activities=169 - user_id=033,#invalid_activities=3 - user_id=005,#invalid_activities=69 - user_id=002,#invalid_activities=135 - user_id=056,#i
nvalid_activities=26 - user_id=069,#invalid_activities=6 - user_id=051,#invalid_activities=44 - user_id=093,#invalid_activities=14 - user_id=067,#in
valid_activities=80 - user_id=058,#invalid_activities=18 - user_id=060,#invalid_activities=2 - user_id=094,#invalid_activities=21 - user_id=112,#inv
alid_activities=185 - user_id=115,#invalid_activities=168 - user_id=123,#invalid_activities=3 - user_id=124,#invalid_activities=4 - user_id=170,#inv
alid_activities=5 - user_id=141,#invalid_activities=1 - user_id=146,#invalid_activities=10 - user_id=179,#invalid_activities=42 - user_id=125,#inval
id_activities=43 - user_id=122,#invalid_activities=6 - user_id=114,#invalid_activities=23 - user_id=113,#invalid_activities=31 - user_id=147,#invali
d_activities=77 - user_id=178,#invalid_activities=1 - user_id=140,#invalid_activities=317 - user_id=176,#invalid_activities=8 - user_id=171,#invalid
_activities=4 -

```

## Discussion

We came up with our solutions using MySQL and Python. Our solution is trying to compute most of the problems inside the queries, but more complex problems required us to also use Python for preparation and working with the data. We especially struggled with pairwise comparisons of rows inside SQL queries. So we only used SQL to filter and fetch the data and used Python for that.

We did not do anything differently than the assignment-sheet instructed. We followed the instructions both in cleaning and inserting the data and (hopefully) on writing the queries.

Pain points in solving the assignment were query number 6 about the COVID-19 tracking. The run time and the difficulty of this query was very high, so we never got the result of this query.

From this assignment we learned more about MySQL. We got more experience in reading and transforming large datasets, creating tables, and writing queries to achieve the wanted results. We also got more experience in working with MySQL and Python simultaneously. Working in a group gave us more practice in structuring teamwork.

---

## Feedback

Some of the queries could be made more explicit and unambiguous. Our group experienced some confusion in interpreting what some of the queries were actually asking for, but most questions were clarified on Piazza. Other than that, we thought it was an interesting assignment. The size of the groups were also good.