Weekly report: July 31, 2015

Ivan Rios S.

July 31, 2015

1 Goals for the week

- Improve the trip detection algorithm
- Include data from the winter time as part of the ground truth used to evaluate the algorithm

2 Activities

- Analyze accuracy of the algorithm on the winter data from IMEI 3469
- Determine reasons why algorithm is not detecting all the trips during the winter.
- Analyze GPS data for each one of the trips and confirm potential problems
- Implement changes in the algorithm to increase the accuracy
- Determine accuracy of the trips with the changes
- Repeat analysis of trips with the new results of the algorithm
- Populate the DB with the results of the new algorithm

3 What I learnt

Initially, I worked by comparing the results of the current algorithm with the results obtained with the algorithm developed by Mikhail. These were the results:

Month	Mikhail's algorithm	Current Algorithm
January	25	24
February	10	8
March	24	2
April	10	2

The current algorithm was not detecting a considerable amount of trips. For this reason, I modified the parameters and found that the value of charging

current was cutting out many trips. I modified this value and also noticed that many trips were being filtered out because of the manner in which the algorithm was merging trips together when they had overlapping times or the time between trips was short enough to be joined to the previous trip.

I modified the algorithm by saving all the trips and processing them at the end, making sure there are within the defined thresholds (average speed, duration, and distance). In this way, many small trips can be joined to bigger trips at the end instead of filtering them out once they are initially identified. With this change, we not only identify more trips but also increase the accuracy of start and end times of each trip.

Using this new algorithm, I obtained a total of 2350 trips which represents an increment of 75% in the total of trips found in the database. After getting these new results, I implemented the GPS tracking part of the algorithm in order to increase the accuracy; however, the results were not positive since the GPS data is not reliable. I looked into the problems causing this analysis to fail, and I realized that most of the trips (approximately 40%) do not contain enough GPS data to accurately identify movements. Using the GPS data to identify additional trips gave a vast number of false positives due to random measurements of the position of the bike. Additionally, there were many contradictory results obtained from the gyro/linear acceleration and the GPS since their sensitivity is different. For these reasons, I decided not to include this part in the final algorithm. I confirmed the way Mikhail implemented this idea, and basically he determined the earliestStartTime and LatestEndTime using the gyroscope data, and the LatestStartTime and EarliestEndTime using the GPS which I proved not to be accurate. I checked my logged initial/ending times, and the data in the DB to confirm that obtaining average of the "Latest" and "Earliest"start or end times does not estimate accurate values.

Finally, after taking in count the low reliability of the GPS data, I changed the parameters that relied on the distance calculated by the algorithm. With these observations and some fixes according to the feedback obtained from the analysis, I reached the following results:

IMEI 3410:

Date	Logged trips	Identified trips
April 27	3	3
April 29	3	3
May 1	3	2
May 6	2	3
May 7	3	2
May 8	2	2
May 11	2	3
May 13	2	2
May 14	2	2
May 19	3	2
May 22	2	4
May 23	2	1
May 25	2	2
June 1	2	2
June 2	3	2
June 5	3	3
June 8	3	5
TOTAL	42	43

IMEI 5233:

Date	Logged trips	Identified trips
May 25	2	3
May 26	2	2
May 29	2	3
June 1	5	5
June 5	2	2
June 6	5	5
June 7	2	2
June 9	3	3
June 16	2	2
June 20	2	2
June 23	2	2
June 24	3	2
June 27	3	3
July 9	2	0
July 13	3	3
July 14	1	0
TOTAL	41	39

IMEI 3469 (summary):

Month	Logged trips	Identified trips	Comments
Jan	25	27	All trips identified (one missing data)
Feb	10	11	All trips identified
March	24	27	All trips identified
April	10	16	All trips identified

In a similar way as before, most of the trips have been detected; some of the trips are still being divided because they have long pauses in between. However, several additional trips have been eliminated because they were joined together with smaller trips that were deleted before and now have been used as a way to merging bigger trips.

The following table presents the new summary of the trips detected in the database:

Parameter	Value
Total number of bicycles	31
Total number of trips (ground truth)	154
Total number of detected trips	3167
Number of additional trips detected	7 (only IMEI 3410 and 5233)
Number of trips not detected	5
Accuracy	˜97%

Both of the previous tables present the final results obtained with the latest version of the algorithm. Since this is the highest accuracy obtained, the aggregated analysis was performed on the trips detected with this algorithm. The following table presents the total a summary of the results obtained from the analysis of all the database:

Parameter	Value
Total number of trips detected	3167
Average number of trips per participant	106
Number of different days of trips	1344
Number days with odd number of trips	562
Number of days with 1 trip	292

As in the previous report, the following tables describe a summary of the trips detected:

The following data was obtained considering the gender of each participant (including the ISS4E team):

Gender	# of Participants	# of Trips	Average
Male	18	1777	99
Female	13	1390	107

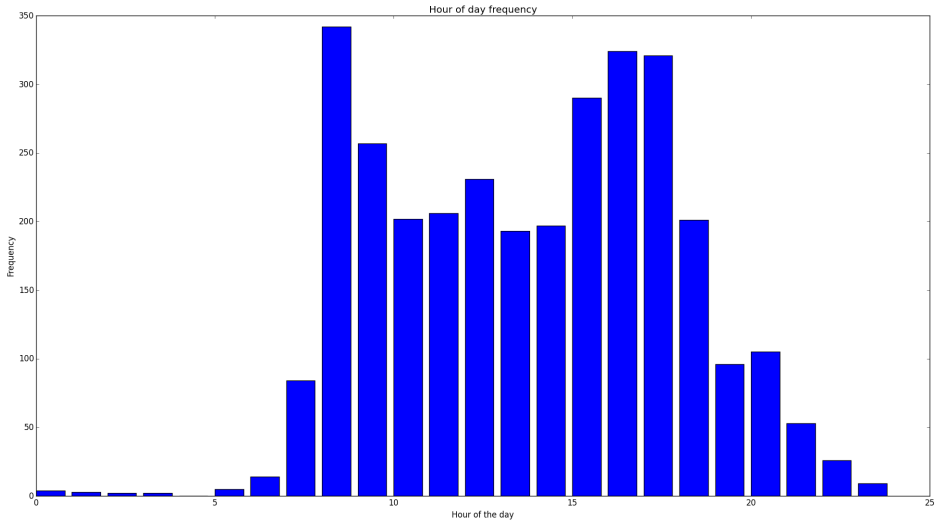
Also, the following data was obtained considering the profession of each participant (without ISS4E team):

Profession	# of Participants	# of Trips	Average
Staff/Faculty	13	1085	84
Student	12	1576	131

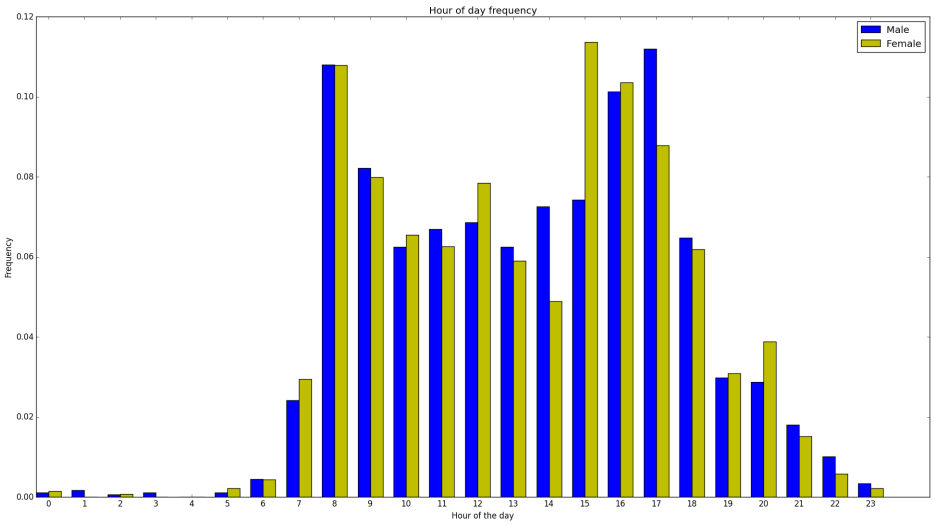
The following are the results obtained from the aggregated analysis of all the trips:

Important note: all the graphs where there is an analysis divided by gender or profession have been modified to use percentages instead of the actual values. This was done because the ratios of male/female and staff/students is not 1.

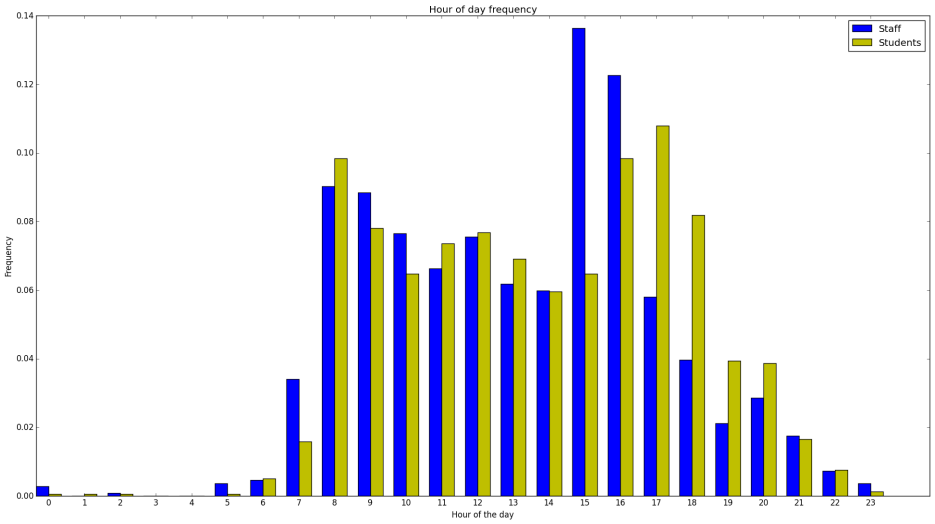
FREQUENCY OF TRIPS PER HOUR (Data includes all the participants and the ISS4E team):



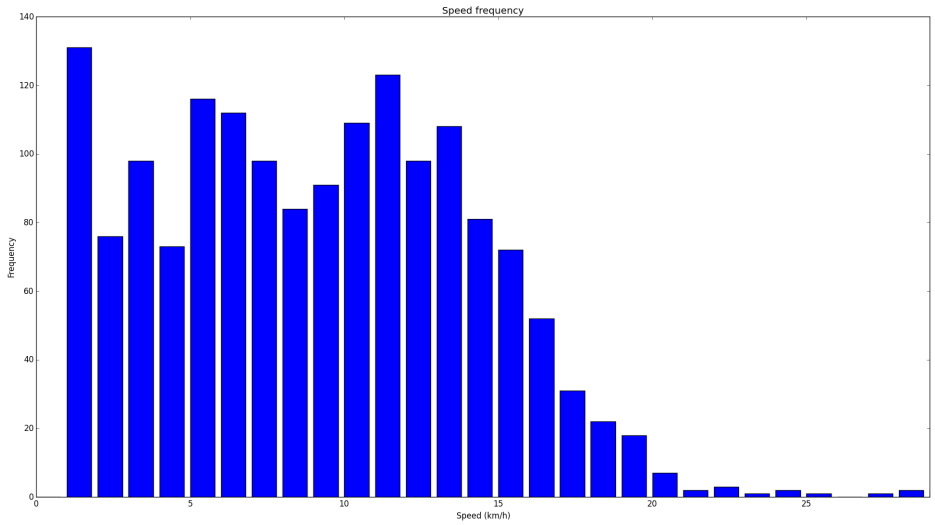
FREQUENCY OF TRIPS PER HOUR DIVIDED BY GENDER (Data includes all the participants and the ISS4E team):



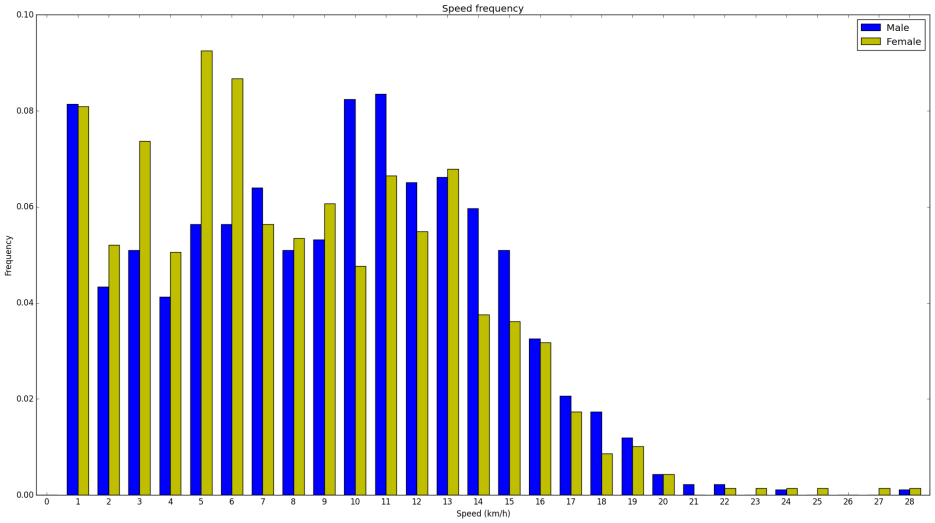
FREQUENCY OF TRIPS PER HOUR DIVIDED BY PROFESSION (Data does not include the ISS4E team):



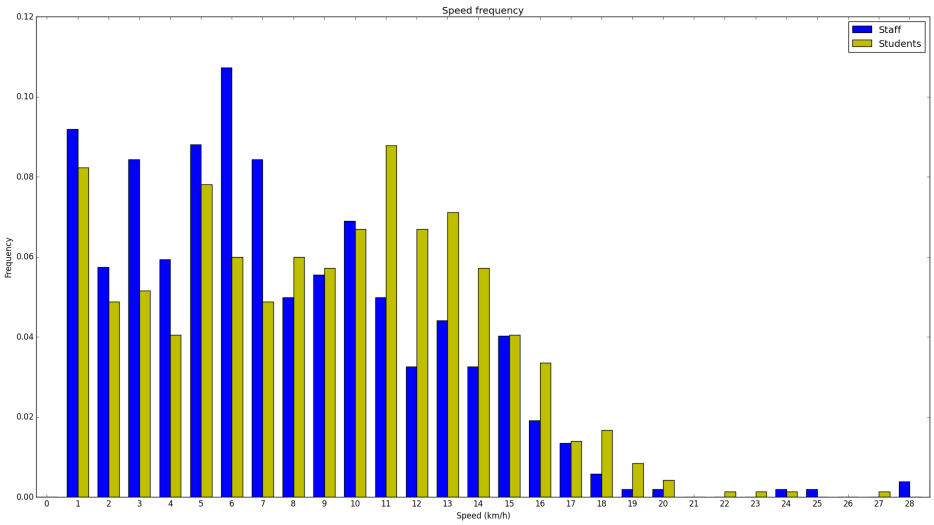
FREQUENCY OF AVERAGE SPEED (KM/H) (Data includes all the participants and the ISS4E team):



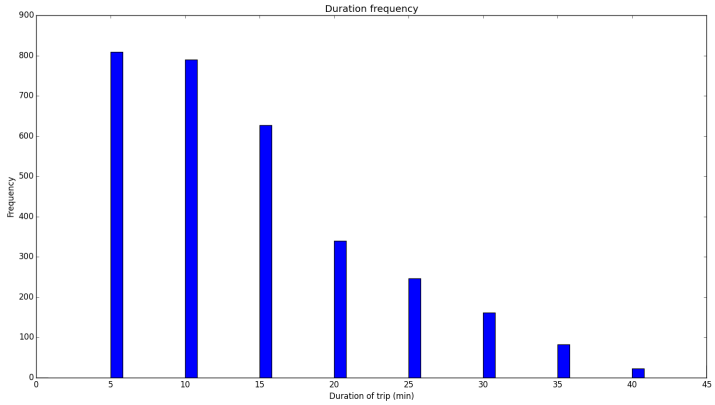
FREQUENCY OF AVERAGE SPEED (KM/H) DIVIDED BY GENDER (Data includes all the participants and the ISS4E team):



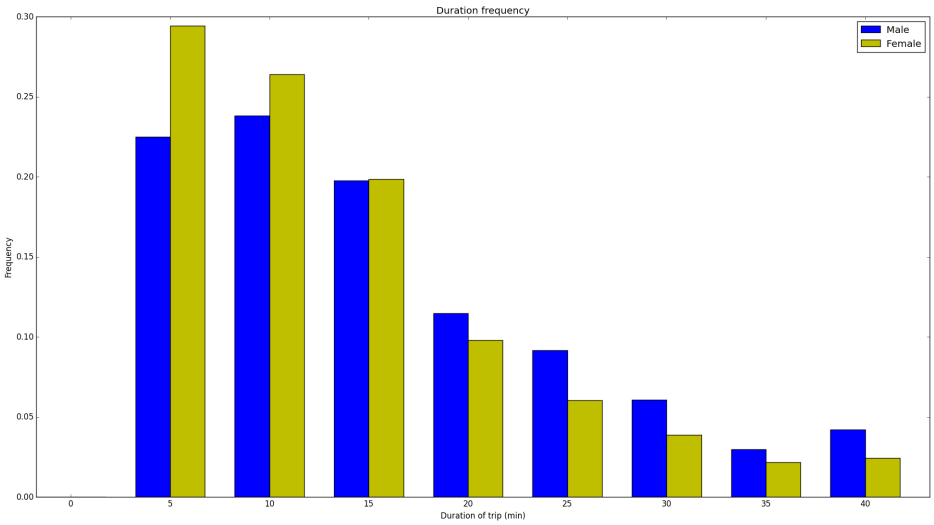
FREQUENCY OF AVERAGE SPEED (KM/H) DIVIDED BY PROFESSION (Data does not include the ISS4E team):



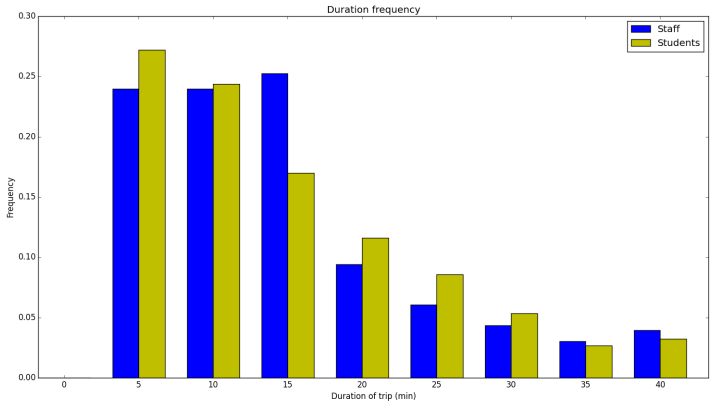
FREQUENCY OF TRIP DURATION (Minutes) (Data includes all the participants and the ISS4E team):



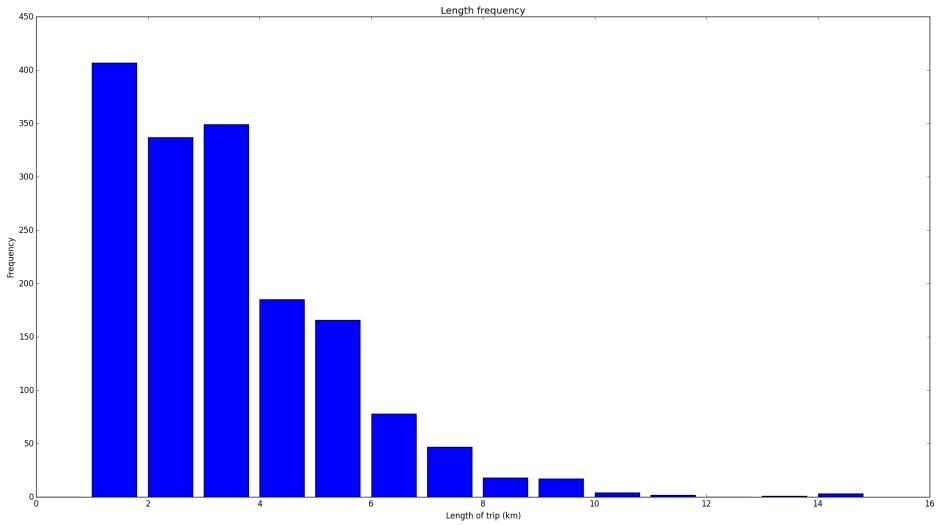
FREQUENCY OF TRIP DURATION (Minutes) DIVIDED BY GENDER (Data includes all the participants and the ISS4E team):



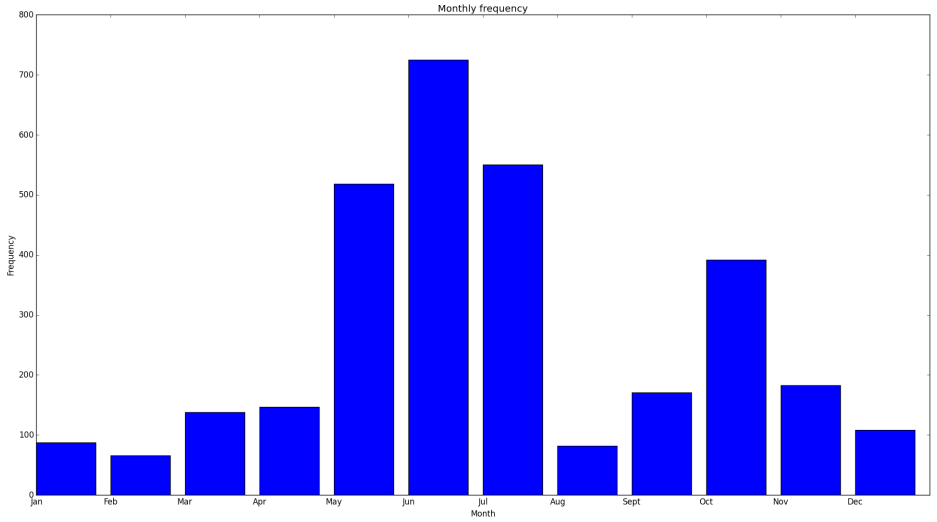
FREQUENCY OF TRIP DURATION (Minutes) DIVIDED BY PROFESSION (Data does not include the ISS4E team):



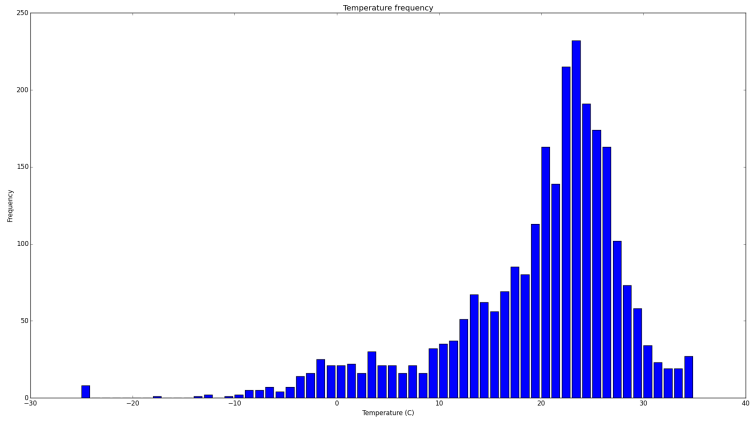
FREQUENCY OF TRIP LENGTH (km) (Data includes all the participants and the ISS4E team):



FREQUENCY OF TRIPS PER MONTH (Data includes all the participants and the ISS4E team):



FREQUENCY OF AVERAGE TEMPERATURE (Data includes all the participants and the ISS4E team):



Finally, here is a summary of the number of trips per IMEI number per month:

IMEI	Total trips	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
1473	355	3	31	45	44	65	68	91	0	0	3	5	0
8664	303	16	1	30	9	28	29	41	2	0	69	56	22
0603	232	39	23	26	7	2	2	11	0	1	32	33	56
3469	194	27	11	27	16	0	1	2	0	34	33	22	21
7303	154	0	0	0	4	37	60	49	0	0	2	2	0
2910	141	0	0	0	0	0	23	41	20	27	30	0	0
0665	133	0	0	0	9	38	45	18	16	0	1	6	0
7459	132	0	0	0	7	22	38	17	0	16	27	5	0
9050	121	0	0	2	0	11	33	45	0	0	23	7	0
4381	117	0	0	1	15	23	35	33	0	3	7	0	0
6904	115	0	0	0	0	77	13	8	0	0	5	12	0
7710	107	0	0	2	4	19	19	19	0	30	13	1	0
7517	88	0	0	0	0	37	22	21	0	0	8	0	0
9407	85	0	0	0	3	19	49	6	6	0	2	0	0
0669	85	1	0	0	0	26	38	5	0	0	10	5	0
9519	73	0	0	0	0	9	38	14	0	0	12	0	0
3410	68	0	0	0	7	25	13	0	0	12	9	0	2
6097	68	0	0	1	4	4	10	15	4	0	15	10	5
8508	67	1	0	4	0	4	12	21	2	0	23	0	0
0636	65	0	0	0	0	8	19	17	15	0	6	0	0
5233	59	0	0	0	0	12	34	4	0	0	9	0	0
6089	52	0	0	0	0	3	7	17	4	4	16	1	0
3215	52	0	0	0	1	1	16	14	3	14	3	0	0
3014	46	0	0	0	4	4	15	0	0	10	13	0	0
8870	44	0	0	0	10	14	15	0	0	0	2	1	2
0657	41	0	0	0	0	2	17	7	5	0	7	3	0
1210	40	0	0	0	0	4	16	12	0	0	0	8	0
6473	36	0	0	0	0	0	14	7	0	0	10	5	0
5432	32	0	0	0	0	8	4	1	0	17	2	0	0
6994	32	0	0	0	3	0	8	12	5	3	0	1	0
0587	30	0	0	0	0	16	12	2	0	0	0	0	0