2013 IEEE International Conference on Healthcare Informatics

Characterizing the Performance and Behaviors of Runners Using Twitter

Qian He, Emmanuel Agu
Department of Computer Science
Worcester Polytechnic Institute
Worcester, MA, United States
qhe, emmanuel@cs.wpi.edu

Diane Strong, Bengisu Tulu
School of Business
Worcester Polytechnic Institute
Worcester, MA, United States
dstrong, bengisu@wpi.edu

Peder Pedersen
Department of Electrical &
Computer Engineering
Worcester Polytechnic Institute
Worcester, MA, United States
pedersen@ece.wpi.edu

Abstract—**Running is a popular physical activity that improves physical and mental wellbeing. Unfortunately, up-to-date information about runners' performance and psychological wellbeing is limited. Many questions remain unanswered, such as how far and how fast runners typically run, their preferred running times and frequencies, how long new runners persist before dropping out, and what factors cause runners to quit. Without hard data, establishing patterns of runner behavior and mitigating challenges they face are difficult. Collecting data manually from large numbers of runners for research studies is costly and time consuming. Emerging Social Networking Services (SNS) and fitness tracking devices make tracking and sharing personal physical activity information easier than before. By monitoring the tweets of a runner group on *Twitter* (SNS) over a 3-month period, we collected 929,825 messages (tweets), in which runners used *Nike+* fitness trackers while running. We found that (1) fitness trackers were most popular in North America (2) one third of runners dropped out after one run (3) Over 95% of runners ran for at least 10 minutes per session (4) less than 2% of runners consistently ran for at least 150 minutes a week, which is the level of physical activity recommended by the CDC (5) 5K was the most popular distance.**

Keywords—social network services; Twitter; physical activity; running; information retrieval; data analysis

I. Introduction

Maintaining regular physical activity is one of the most important ways to stay healthy and reduce the risk of diseases such as cardiovascular disease, type 2 diabetes, and metabolic syndrome. Physical activity not only can help people control weight, strengthen bones and muscles, but also can improve mental health and mood. Research shows that regular physical activity can even reduce the risk of some cancers and increase the chances of living longer [1]. Numerous health benefits of regular physical activity have been confirmed by research results [2].

The Centers for Disease Control and Prevention (CDC) currently recommends that the minimum amount of physical activity required to stay healthy is a total of 150 minutes per week with each session lasting at least 10 minutes [3]. Running is one of the most popular physical exercises all over the world. Unlike other sports, which may require specially designed fields, running only requires a safe sidewalk, a park, or a treadmill. Most runners only need a pair of running shoes as running equipment. A more comprehensive set of equipment may include weather-appropriate running clothes, a music player, pedometers, fitness trackers, or smartphone applications to log running data.

Up-to-date data on runners is important for helping trainers and health professionals to improve runners' performance and reduce attrition. City planners can also use this data for deciding where to place parks and roads and how to make cities more runner-friendly. Finally, programs and apps that coach runners can utilize patterns observed in similar runners to provide targeted advice. Unfortunately, up-to-date performance data on runners is scarce. Questions such as how far casual runners usually run, their running pace, preferred running times, and running frequency are not readily available. Additionally, information about the emotional and psychological state of runners is even less available. Questions about how long new runners persist before they drop out, factors causing runners to drop out, psychological obstacles they face and various factors that affect them when they run remain unanswered. These questions are hard to answer because collecting data manually from large numbers of runners for research studies is costly and time consuming. Without hard data, establishing patterns exhibited by runners and mitigating specific difficulties that runners face are difficult.

Emerging fitness tracking devices such as the *Nike+* series [4] and *Fitbit* [5] have quickly become popular with casual runners. These devices track runners' performance and automatically upload information about their runs (time, location, pace, distance) and their comments to social networking sites such as Twitter, where they can share their experiences with friends instantly.

Our goal was to gather a large number of runners' messages (tweets) on *Twitter*, a popular Social Network Service (SNS), and develop tools to analyze these messages in order to answer some of the questions posed above. *Twitter* publishes an API that facilitates the collection of a large number of tweets on various topics including fitness and running. By monitoring the messages of runners on *Twitter* over a 3-month period, we were able to collect 929,825 tweets from a runner group, in which people used sensors and trackers while running. We processed these runners' tweets and analyzed, resulting in a better understanding of current running

978-0-7695-5089-3/13 $26.00 © 2013 IEEE
DOI 10.1109/ICHI.2013.56

patterns of a large number of diverse runners. We found that (1) these fitness trackers were most popular in North America (2) one third of the runners dropped out after their first runs, (3) Over 95% of runners ran for at least 10 minutes per session which exceeds the CDC recommendation for maintaining good health but (4) less than 2% of runners consistently ran for at least 150 minutes on each week during our 3-month study (5) holiday season did affect number of runs negatively. We confirmed the popularity of the 5K run, morning run, and afternoon run. To the best of our knowledge, our work is the first large-scale characterization of runners using social networks.

II. Background and related work

A. Physical Activity Tracking

Tracking physical activity is important in order to quantify physical health and wellbeing. In the context of physical activity, examples of tracked information include heart rate, respiratory rate, blood glucose level, blood oxygen level, blood pressure, physical activity location, and acceleration. Tracking technologies and measurement devices for personal use are now mature and affordable. For instance, a decade ago, wearable triaxial accelerometers for tracking physical activity were just starting to be used in measurement studies [6–9]. Today, such triaxial accelerometers are standard build-in sensors on many widely owned smartphones.

As interest in personal physical health increased, user interest in quantifying their "health" has also grown. Companies such as *Nike*, *Fitbit*, *Withings*, and *BodyMedia* have developed devices that can measure the distance walked by users, number of stairs or building floors climbed, calories burnt, weight changes, and blood pressure fluctuations. Some of these devices can also synchronize users' measurements with applications on their smartphones or upload the data to their personal health records stored in the cloud. Because more and more smartphones are now equipped with GPS and accelerometers, fitness tracking applications can continuously monitor the user, turning their smartphone into a fitness tracker.

Nike was one of the first companies to make fitness tracking devices available to consumers. The *Nike+ iPod Sensor* was the first device for tracking physical activity in the *Nike+* product line [4]. It is a small sensor that can be placed in selected models of *Nike* running shoes to track runner performance. The *Nike+ iPod Sensor* can be used in several ways. First, *Nike+ iPod*, a factory-installed application on the *iPod touch* and *iPhone* mobile devices, can read runners' data wirelessly from the radio-frequency transmitter (Nordic Semiconductor nFR2402) onboard the *Nike+ iPod Sensor*. This application also uploads the runner's data to the user's account on *Nike+*'s website. Second, *Nike+ SportBand* is a tracking bracelet that can communicate with *Nike+ iPod Sensor*. Users can connect it to a computer through a USB and upload running records.

Following the success of the *Nike+ iPod Sensor*, *Nike* developed two devices that do not have to work with *Nike+ iPod Sensor* and *Nike* running shoes: *Nike+ SportWatch GPS* and *Nike+ FuelBand*. *Nike+ SportWatch GPS* tracks the running distance using GPS. Data upload can be done by connecting it to a computer's USB interface. *Nike+ FuelBand* is an activity tracker that can be worn on the wrist. It has an embedded accelerometer which calculates the number of steps taken and calories burnt. It can communicate with a personal computer through its USB interface or with *iPod touch* and *iPhone* through Bluetooth for uploading running records.

As mentioned, more and more smartphones are now equipped with GPS and accelerometers. *Nike* later developed a GPS tracking application for *iPhone* and *Android* phones (with GPS) called *Nike+ Running App*. If a user is running indoors and no GPS signal can be detected, the app can estimate the distance using the phone's built-in accelerometer.

Hereafter, we shall refer to these *Nike* runner tracking devices and application as "*Nike+* trackers", and we shall call the users of *Nike+* trackers as "*Nike+* users".

B. Data Mining on Public SNS

Social Networking Service (SNS) platforms enable people to share personal interests and news, and to post their current status, increasing access to information about people's daily lives. *Twitter* [10] is one of the most successful Social Networking Services. It was launched on July 15, 2006 and now has more than 517 million users, who send more than 340 million tweets a day in total [11][12]. On *Twitter*, people can join groups by sending tweets with hashtags (#). These groups are based on activities of interest such as running, photography or political views.

In recent years, more and more web/mobile applications, devices and even operating systems have integrated "Share to SNS" and "Auto Sync to SNS" features that automatically post messages to *Twitter*. Consequently, *Twitter* has become a large hub of user-generated information. Additionally, *Twitter* has published APIs that allow posted messages to be retrieved. Researchers can collect data from multiple message streams through *Twitter*, and analyze them to understand people's behaviors. Specific to our work, running devices such as *Nike+* trackers can automatically post SNS messages once runners have completed their runs. These messages typically include runner performance statistics (running time, pace, and distance), as well as comments entered by the runner.

Twitter's message retrieval API is open, well-designed, and RESTful (Representational State Transfer, REST), which has contributed significantly to its success [13]. Using this API, third-party developers can test new ideas using *Twitter*'s infrastructure and researchers have the opportunity to access *Twitter*'s large amounts of data quickly. As such, research into analyzing and mining data on *Twitter* for various purposes has exploded. Previous work has focused on analyzing *Twitter* messages for diverse topics including user classification, geo-location detection and topic trend prediction [14–17]. Our work focuses on analyzing *Twitter* messages to characterize the performance and behavior of runners.

III. Methodology

Nike+ users usually upload their running records to *Nike+* website as to track their performance. For users that have permitted *Nike+* to share their running records on *Twitter*, once they finish a run, a message (called a tweet) is automatically sent to the user's *Twitter* account. Since *Twitter* accounts are public by default, anyone can capture and read these auto-generated runner tweets.

A. Data Source

Twitter provides streaming API that can be used to retrieve batches of messages that have been posted by *Twitter* users. A client application can be programmed to use this API to establish a long-lived HTTP connection with *Twitter*'s server and continuously receive messages, obviating the need to poll the server. *Twitter*'s Streaming API differs significantly from its REST API in the following ways: *(1) No API rate limit:* because once the streaming connection is established, no further API calls are needed; *(2) No missed tweets:* as long as the developer's application server has a very good network connection and runs quickly enough to consume all tweets coming from the pipe, no tweets will be missed.

There are three modes of streaming API supported by the *Twitter* [18], each of which controls what messages are received by a client application:

1) Firehose: streams all public messages on *Twitter*;

2) Sample: streams a small random sample of all public messages;

3) Filter: streams public messages that match one or more specified filter conditions.

In this work, since we analyze only messages that are automatically sent by *Nike+* trackers, we use the *filter* mode to monitor running-related messages posted by these devices.

B. Tool Design

To monitor, analyze, and store running-related messages, we designed a tool with several modules (Fig. 1) including (1) a monitor daemon, for constantly monitoring messages on *Twitter*; (2) an analyzer, for retrieving running information from message text and associated meta information; (3) a database connector, for communicating with the database and backing up tweets; and (4) a report generator, for generating text reports, .CSV files, and .ARFF files.

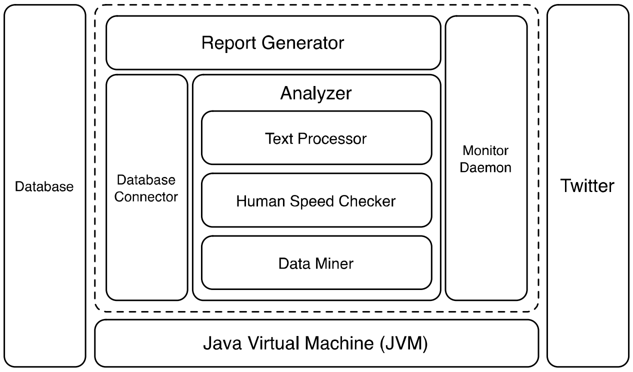

Fig. 1. Architecture of the Tool

C. Tool Implementation

We implemented our tool using the programming language *Scala* and the *H2* database. Several open source libraries were used in our application including: (1) *Twitter4J*, a Java library that wraps *Twitter*'s APIs [19]; (2) *Slick*, a database query and access library for *Scala* [20]; and (3) *Weka*, a library of machine learning algorithms for data mining [21].

Twitter users (or devices) can designate the *Twitter* group(s) on which their messages should appear by embedding the group name prefixed by a "#" symbol in their messages. For instance, messages generated by *Nike+* trackers embed the "#nikeplus" keyword. To retrieve these messages posted to the Nike+ runner's group, we filtered *Twitter* messages using the "#nikeplus" keyword.

In our analysis, we retrieved only messages automatically generated by *Nike+* trackers because they had a more uniform format than human-generated messages, and included meta information such as run time, duration, distance, and location of the run. One issue with filtering messages using *Twitter* group names as keyword is that some tweets sent manually by *Twitter* users (not auto-generated by *Nike+* devices) may also contain this keyword. Therefore, after receiving the messages, we applied a regular expression to filter out the non-auto-sent (human generated) messages. Because different *Nike+* trackers use different message formats, our regular expression was designed to match all *Nike+* trackers:

```
^(.*)I (just )?(finished|crushed) a (\d+.\d+) ?(km|mi) run
?(with a pace of (\d+'\d+).* )?(with a time of (.+) )?with
(.+)\..*$
```

The messages below are examples of messages collected through the process described above. From these messages we could determine the distance, pace, and local time of runs.

- I just finished a 5.72 km run with a pace of 4'56"/km with Nike+ Running. http://t.co/someurl #nikeplus
- I just finished a 2.38 mi run with a pace of 18'44" with Nike+. http://t.co/someurl #nikeplus
- I crushed a 10.2 km run with a pace of 5'50" with Nike+ SportWatch GPS. #nikeplus: http://t.co/nikeplus
- I crushed a 9.0 mi run with Nike+ SportBand. #nikeplus: http://t.co/someurl
- Feel so gooood! I crushed a 6.0 km run with Nike+ iPod. #nikeplus: http://t.co/someurl

IV. DATA ANALYSIS

Our application collected 929,825 tweets containing keyword "nikeplus" in 3 months (from October 10, 2012 to January 9, 2013). We eliminated some categories of tweets for various reasons. We eliminated 338,825 tweets (36.44%) that were not written in English and 45,809 tweets (4.93%) that were written in English but not generated by *Nike+* trackers. We also removed (0.66%, 6,144) tweets that were "retweet"s (messages in which users quoted their friends' messages).

We also found 11,717 tweets (1.26%) that we believed were generated for activities other than running, because the speeds of these runs were faster than humanly possible running speeds. Specifically, runs in which the speeds were faster than the world speed records of shorter or equal distances, were adjudged to contain errors and were eliminated. These abnormal runs were generated by *Nike+ Running App* (when using GPS for tracking) and *Nike+ SportWatch GPS*. We believe that two scenarios may have caused these abnormal speeds: (1) The location where the user ran had a poor or unstable GPS signal; (2) The user used *Nike+ Running App* or *Nike+ SportWatch GPS* while performing other fitness activities such as bicycling or driving.

After removing the above tweets, we were left with 524,330 runner messages (56.39%) that were sent automatically by *Nike+* trackers owned by 83,415 unique *Twitter* users. Hereafter, we shall call these messages as "runs" since a single message is typically auto-generated for each run.

A. *Running Statistics*

All runs contain *distance* information either in "miles" or "kilometers". 278,897 runs (53.19%) have *duration* or *speed*, and 362,384 runs (69.11%) have *UTC offset*, the local timezone's offset from a reference timezone, which can be used to calculate the local time of the run.

1) Location of Runs

Each *Twitter* user has a profile in which they provide additional information about themselves such as their *location* and *timezone*. Since *Twitter* allows users to fill any character into the *location* field, locations sometimes were meaningless. For instance, some users filled in "somewhere near you", "Mars", and "parallel universe" as locations. We also found that multiple cities shared the same name. For example, some users filled in "Worcester" in the location field, making it difficult to tell whether they were in the city of Worcester in Massachusetts (U.S.), Worcester in New York (U.S.), Worcester in England (U.K.), or Worcester in Western Cape (South Africa). Each tweet also has a data field for storing its geographic coordinates. However, only 1231 (0.13%) of the tweets we collected contained this information.

Consequently, we used *timezone* instead to retrieve geo-location information. As we know, there are 40 UTC offsets. The relation between UTC offset and timezone is 1-to-n. For example, timezone "Mexico City" and timezone "Central Standard Time" both are in "UTC−06:00" offset. This character of timezone helps us get better geo-location in terms of timezone area. Finally, we found that runners were in 136 timezones. Colored map [Fig. 2] shows the distribution of runners around world. Table I shows the top 10 timezones in which the captured runs occurred.

Since 46.79% of the analyzed runs were performed in the Eastern Timezone (US & Canada), Central Time (US & Canada), Pacific Time (US & Canada), Mountain Time (US & Canada), and Mexico City, we concluded that *Nike+* fitness trackers were most popular in North America. The only caveat is that this conclusion may be biased by our removal of non-English tweets which may have been generated by *Nike+* trackers. To double-check our conclusion that *Nike+* trackers were most popular in North America, we analyzed the eliminated non-English tweets that were generated from 241,987 runs. Of these runs, 27.49% were from North America, which still made "North America" the top geographical region for *Nike+* trackers.

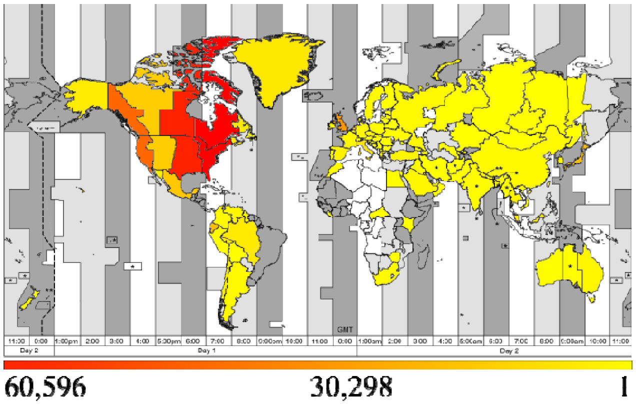

Fig. 2. Distribution of runs around the world

TABLE I. TIME ZONE (TOP 10)

Time Zone	Number of Tweets
Eastern Time (US & Canada)	60,596
Central Time (US & Canada)	52,765
Pacific Time (US & Canada)	35,828
London, United Kingdom	24,749
Quito*	15,846
Tokyo, Japan	15,258
Amsterdam, Netherlands	12,338
Mountain Time (US & Canada)	11,988
Mexico City, Mexico	8,389
Hawaii, USA	6,849

*. Quito, formally San Francisco de Quito, is the capital city of Ecuador

2) Running Speed

Fig. 3 is a histogram showing the number of runs performed at each speed. Fig. 4 shows the number of runs performed for each unique duration. If a run's duration was not automatically embedded by the *Nike+* device, we calculated it using its *speed* and *distance*. Fig. 3 and Fig. 4 have roughly a normal distribution with average speed of 10 km/h and average duration of 35 minutes respectively.

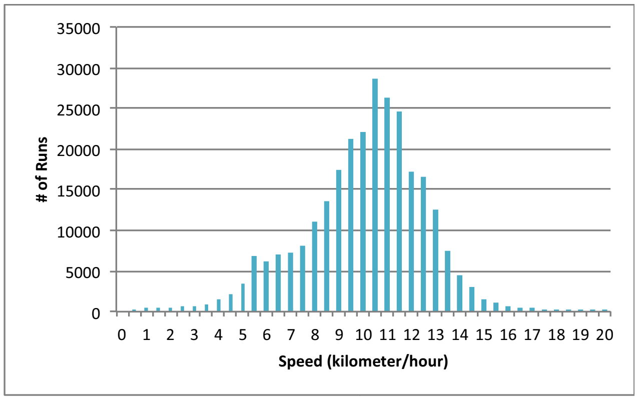

Fig. 3. Number of Runs at a given Speed

3) *Duration*

In Fig. 4, we found 265,797 runs (95.30% of runs with duration information) met the CDC recommended duration for a single physical activity session—at least 10 minutes. However, only a few runners met the other part of the CDC's recommendation—performing at least 150 minutes of physical activity per week to stay healthy [Fig. 5].

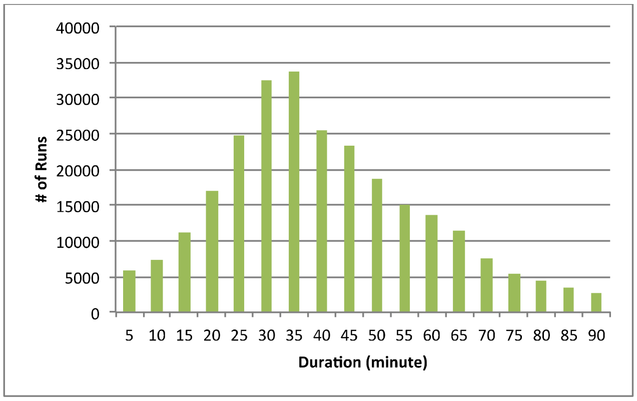

Fig. 4. Number of Runs vs. Duration

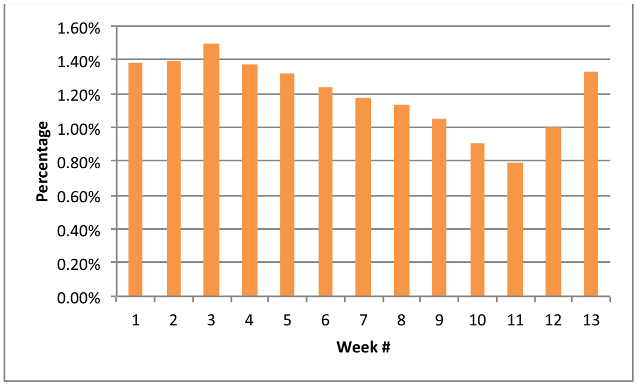

Fig. 5. Percentage of runners who can run 150 or more minutes a week

4) *Distance*

Fig. 6 is a histogram of distances covered by the runners. We found that (1) Runners typically completed integer values of distance as recommended by many training programs. Spikes occured at 2-km, 3.5-km (~2mile), 5-km (~3mile), 6.5-km (~4mile), 8.5-km (5mile), 10-km (6mile), and 16-km (10mile) distances; (2) Runners usually ran slightly longer than these integer distances. We believe these runners kept their *Nike+* trackers on during their "warm up" and "cool down" phases before and after running; (3) The distances covered also had a normal distribution, with 5km as the most popular running distance (known as the 5K run, which is popular with both novice and professional runners).

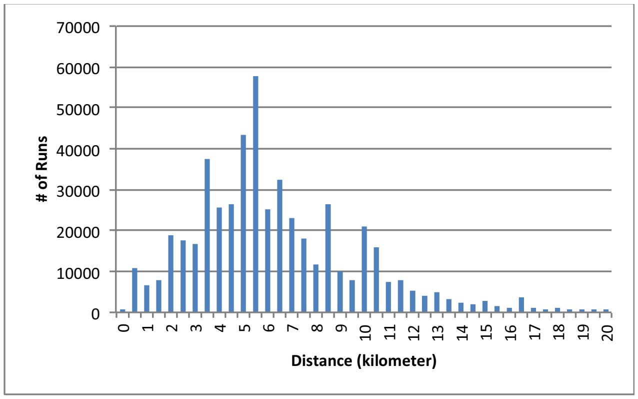

Fig. 6. Number of Runs at a given Distance

B. *Running Patterns*

1) *Local Time of Run*

362,384 runs (69.11%) with *UTC offset* were used in Figures 5-8 to calculate the local time when each run ended. Since running tweets were usually sent immediately or delayed not much after a run was completed, we assume that the message was sent at the end time of each run. The manufacturer's documentation indicates that *Nike+ Running App* (on iPhone and Android phones) will send tweets immediately after each run if the phones have Internet connection, and *Nike+ iPod Sensor*, *Nike+ SportBand*, and *Nike+ FuelBand* will send tweets if they were connected to phones with Bluetooth.

Runners preferred to run in the morning (finished around 10:00 AM) and in the afternoon (finished around 07:00 PM) [Fig. 7].

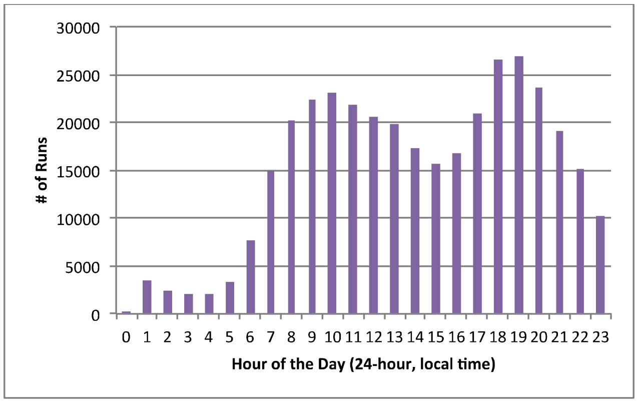

Fig. 7. Number of Runs at a given Hour of the Day (Local time of runner)

We expected that more runs would occur on weekends since people usually have more free time on weekends than on weekdays. Our results, however, surprisingly showed that there was no major difference between weekdays and weekends (Saturdays and Sundays) [Fig. 8]. Friday is the least popular day to run, perhaps because most people are tired from their 5-day work week or they may prefer to spend their Friday evenings for relaxation and entertainment.

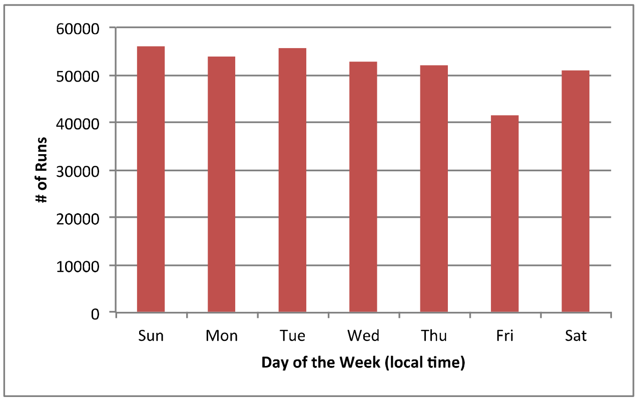

Fig. 8. Number of Runs on each Day of the Week

Fig. 9 shows seasonal patterns including holidays occurring during our analysis period. Similar to weekends, people did not run more on short holidays. However, people ran less during long holidays such as Christmas and New Year. We also saw a spike for the New Year when many runners try to start the year with a resolution to run more.

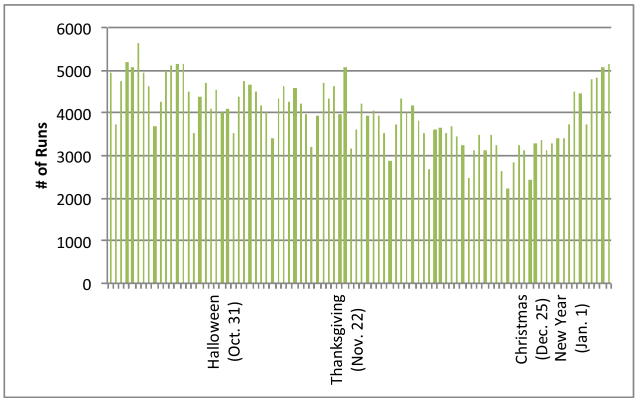

Fig. 9. Number of Runs on Holidays

For each runner, we logged the length of time between consecutive runs. For instance, if a user ran on Monday and Wednesday in the same week, the interval is 2 days. We found 33,428 intervals (7.68%) were less or equal than 0.1 day [Fig. 10]. We believe that was because some users ran multiple times per day.

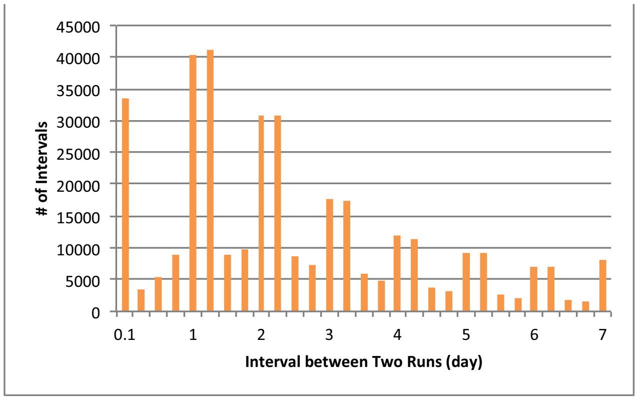

Fig. 10. Interval between Two Runs

2) Frequency

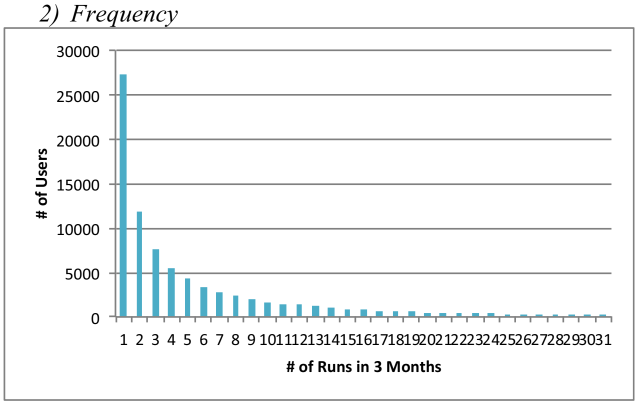

Fig. 11. Number of Runs in 3 Months

Even though our chosen 3-month study period had several holidays and were winter months in North America and Europe, we were still surprised that 27,222 users (32.63%) only ran once in this 3-month period, and then dropped out [Fig. 11]. This may point to the existence of significant obstacles that led to runner attrition. Reasons for attrition shall be studied more in future work.

Fig. 12 shows the frequency *f*, calculated as the average number of runs completed during the running period:

$$f = \frac{n}{t_{n-1} - t_0}$$

where t_0 is the time of the first run captured for a given user, t_{n-1} is the time of the last captured run, and n is the number of runs for that user. Runners captured only once are not shown in [Fig. 12]. During our 3-month observation period, we found most runners (82.08%) ran less than twice a week, and 2.55% runners ran every day.

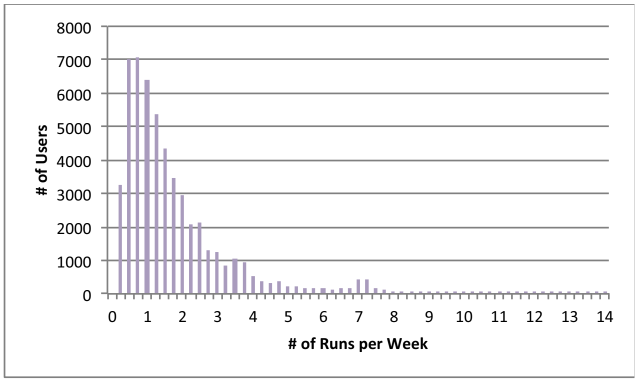

Fig. 12. Number of Runs per Week

3) Temporal Patterns

The figure in Appendix I displays daily and weekly run times on a 2 dimensional grid. We found that on weekends, runners preferred to run in the morning, but on weekdays, they favored afternoons and evenings. On weekends, longer distances were run. Specifically, the number of 10-mile runs and half-marathon runs on weekends was significantly higher than that on weekdays [Appendix II].

We also noticed that runners tended to run either very fast or slow over short distances (bi-modal). For longer distances, most runners reduced their speeds, presumably to complete those longer distances. Essentially, runners showed a more narrow range of running speeds (consistent pace) for longer distances as shown in Appendix III.

C. *Groups of Runs*

In previous sections, we found that runs were not uniformly distributed on any attribute, which implied that some runs were similar to each other and clustered. In order to find clusters of similar runners, we performed a clustering on 10% (random sample) of our dataset. As previously mentioned, *distance*, *speed*, and *hour of the day* approximately had normal distributions. Therefore, the K-means algorithm was a reasonable choice for clustering runs. In order to find the optimal number of clusters, which is an important parameter in K-means clustering; we used the Expectation–Maximization (EM) algorithm during clustering. Four clusters were found by 10-fold cross validation [Table II]. Thirteen iterations were performed. The log likelihood was -8.99851.

TABLE II. GROUPS OF RUNS

Attributes		**Cluster**			
		0 (31%)	***1 (22%)***	***2 (23%)***	***3 (24%)***
Distance	**Mean**	6.3702	8.6622	4.7642	5.1885
	Std. Dev.	3.7232	4.3979	2.3975	2.3264
Speed	**Mean**	9.93	10.3449	9.3904	9.9124
	Std. Dev.	1.8244	1.1287	2.9118	0.3349
Hour of the Day (24-hour)	**Mean**	13.9058	13.641	14.2453	13.8271
	Std. Dev.	4.4543	5.1141	5.4906	5.4476
Day of the Week	**SUN**	1	3638.1254	1366.9369	374.9377
	MON	1	2095.0644	2702.8169	484.1187
	TUE	15973.5037	1385.3826	2222.175	1919.9386
	WEN	1	873.5256	1577.4306	2716.0438
	THU	1	851.8888	1568.7295	2827.3818
	FRI	1	845.866	1227.6719	2018.4621
	SAT	1	1846.3363	1159.5854	1872.0783
	ALL	15979.5037	11536.1892	11825.3462	12212.9609

These clusters are shown in Fig. 13, and have the following characteristics:

1) Cluster 0: Most runs in cluster 0 were done on Tuesdays with a moderate speed (column 3, row 1).

2) Cluster 1: Runs in cluster 1 were performed with a longer distance (column 0) and a moderate speed (column 1, row 0). Saturdays were preferred running days (col 3, row 2).

3) Cluster 2: Runs in cluster 2 had a big variance in speed, covering from the lowest to the fastest (column 1, row 1). The distance of this cluster was shorter, compared to cluster 1 (column 1, row 0). Mondays were preferred running days (col 3, row 2).

4) Cluster 3: The speed of runs in this cluster was moderate and with a very low variance (column 1). Runs were seldom performed on Sundays and Mondays. Wednesdays and Thursdays were the preferred running days (column 3, row 2).

No significant characteristic on “Hour of the Day” was found among these 4 groups of runs.

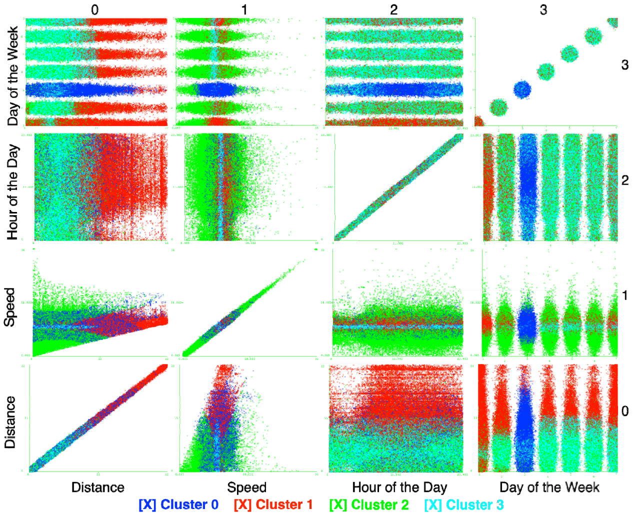

Fig. 13. Clusters shown with attributes

V. DISCUSSION AND FUTURE WORK

During data analysis, we found some interesting phenomena, some of which can be explained from personal experience, but cannot be validated quantitatively. For instance, we do not know if the decrease in runs on Fridays is because runners want to relax or seek entertainment. In future work, we shall analyze text and qualitative comments of runners’ daily tweets to verify our guess.

We saw that 32.6% of runners ran only once, and then dropped out. We would like to understand what happened to these runners. Did the runners choose to to perform other physical activities, stop using Nike+ trackers, or meet some barriers? If they faced some barriers, then what were these barriers? Toscos *et al* [23] identified eighteen unique obstacles to physical activity in an online forum, including illness and injury, lack of willpower, and lack of time. We would like to find out if these eighteen barriers are also common among runners. Analyzing the daily tweets of runners in future work shall also shed more light on these obstacles and reasons for attrition.

The correlations between distance, duration, speed, time of the day, day of the week, and time of the next run were not clear. Identifying the underlying correlations can help us build models for predicting impending runner attrition based on the runner's performance history.

Finally, we would like to speculate on potential uses for the data we have presented in this paper. First, healthcare professionals can use our data as statistical evidence for recommending specific runs to their patients. For instance, since 5K and 30-minute runs are popular, they are reasonable recommendations for the patients who need more physical activity and whose physical conditions allow. Secondly, coaching programs and apps can generate personalized advice and feedback based on a runner's history, their performance and patterns we found in our data. For example, if a user usually ran 5 miles at a speed of 10 mph, an intelligent app may advice the user to slow down to 8.5 mph since observed correlations between distance, speed, and frequency, indicate that "running 5 miles at a speed of 10 mph" significantly increased the chances of an injury, which ultimately resulted in a reduced frequency of running. Third, city planners can use our data to make public facilities in their cities runner-friendly. For instance, when should a city turn on the streetlights? While sunset is a major factor, if city planners take into consideration the safety of runners, they may adjust the time of turning on streetlights since many people run between 5PM and 8PM. Finally, for public health researchers, if other data sources are connected, our collected data and developed tool may help them find correlations between the incidence/prevalence of different diseases and the intensity/frequency of running. For example, running twice a week may reduce the risk of getting the flu.

VI. Conclusion

In this paper, we explored the use of *Twitter,* a popular social networking service, in characterizing the performance and behavior of runners over a 3-month period. We found that (1) these fitness trackers were most popular in North America; (2) one third of the runners dropped out after their first runs; (3) over 95% of runners ran for at least 10 minutes per session which exceeds the CDC recommendation for maintaining good health; but (4) less than 2% of runners consistently ran for at least 150 minutes on each week during our 3-month study; (5) holiday season did affect number of runs negatively. We confirmed the popularity of the 5K run, morning run, and afternoon run.

References

[1] "Physical Activity for Everyone: The Benefits of Physical Activity," 2011. [Online]. Available: http://www.cdc.gov/physicalactivity/everyone/health/index.html.

[2] D. E. R. Warburton, C. W. Nicol, and S. S. D. Bredin, "Health benefits of physical activity: the evidence.," *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*, vol. 174, no. 6, pp. 801–9, Mar. 2006.

[3] United States. Dept. of Health and Human Services., "2008 physical activity guidelines for Americans be active, healthy, and happy!," *ODPHP publication no U0036*. U.S. Dept. of Health and Human Services,, Washington, DC, 2008.

[4] "Nike+." [Online]. Available: http://nikeplus.nike.com/plus/.

[5] "Fitbit." [Online]. Available: http://www.fitbit.com/.

[6] M. J. Mathie, B. G. Celler, N. H. Lovell, and A. C. F. Coster, "Classification of basic daily movements using a triaxial accelerometer," *Medical & Biological Engineering & Computing*, vol. 42, no. 5, pp. 679–687, 2004.

[7] K. Zhang, P. Werner, M. Sun, F. X. Pi-Sunyer, and C. N. Boozer, "Measurement of human daily physical activity.," *Obesity research*, vol. 11, no. 1, pp. 33–40, Jan. 2003.

[8] G. Plasqui and K. R. Westerterp, "Physical activity assessment with accelerometers: an evaluation against doubly labeled water," *Obesity (Silver Spring)*, vol. 15, no. 10, pp. 2371–2379, 2007.

[9] R. P. Troiano, D. Berrigan, K. W. Dodd, L. C. Mâsse, T. Tilert, and M. McDowell, "Physical activity in the United States measured by accelerometer.," *Medicine & Science in Sports & Exercise*, vol. 40, no. 1, pp. 181–188, 2008.

[10] "Twitter." [Online]. Available: https://twitter.com/.

[11] "Twitter Blog: Twitter turns six." [Online]. Available: http://blog.twitter.com/2012/03/twitter-turns-six.html.

[12] "Twitter May Have 500M+ Users But Only 170M Are Active, 75% On Twitter's Own Clients | TechCrunch." [Online]. Available: http://techcrunch.com/2012/07/31/twitter-may-have-500m-users-but-only-170m-are-active-75-on-twitters-own-clients/.

[13] "Twitter's Open Platform Advantage – ReadWrite." [Online]. Available: http://readwrite.com/2007/09/05/twitter_open_platform_advantage.

[14] M. De Choudhury, N. Diakopoulos, and M. Naaman, "Unfolding the event landscape on twitter: classification and exploration of user categories," *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, Seattle, Washington, USA, pp. 241–244, 2012.

[15] M. J. Paul and M. Dredze, "You Are What You Tweet: Analyzing Twitter for Public Health," *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. 2011.

[16] Z. Cheng, J. Caverlee, and K. Lee, "You are where you tweet: a content-based approach to geo-locating twitter users," *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, Toronto, ON, Canada, pp. 759–768, 2010.

[17] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?," in *Proceedings of the 19th international conference on World wide web - WWW '10*, 2010, p. 591.

[18] "Public streams | Twitter Developers." [Online]. Available: https://dev.twitter.com/docs/streaming-apis/streams/public.

[19] "Twitter4J - A Java library for the Twitter API." [Online]. Available: http://twitter4j.org/en/index.html.

[20] "Slick." [Online]. Available: http://slick.typesafe.com/.

[21] "Weka 3 - Data Mining with Open Source Machine Learning Software in Java." [Online]. Available: http://www.cs.waikato.ac.nz/ml/weka/.

[22] "Track and Field best all-time performances." [Online]. Available: http://www.alltime-athletics.com/.

[23] T. Toscos, S. Consolvo, and D. W. McDonald, "Barriers to Physical Activity: A Study of Self-Revelation in an Online Community.," *Journal of Medical Systems*, vol. 35, no. 5, pp. 1225–42, 2011.

Appendices

Appendix I. Day of the Week vs. Hour of the Day

Percentage		Day of the Week (local)								Grand Total
		SUN	MON	TUE	WED	THU	FRI	SAT	(blank)	
Hour of the Day (24-hour, local)	[0-1]	0.09%	0.08%	0.12%	0.11%	0.11%	0.10%	0.08%	0.00%	0.69%
	(1-2]	0.07%	0.05%	0.07%	0.08%	0.07%	0.07%	0.07%	0.00%	0.47%
	(2-3]	0.05%	0.05%	0.06%	0.06%	0.06%	0.05%	0.05%	0.00%	0.39%
	(3-4]	0.05%	0.06%	0.06%	0.06%	0.06%	0.05%	0.05%	0.00%	0.39%
	(4-5]	0.05%	0.10%	0.11%	0.10%	0.11%	0.10%	0.06%	0.00%	0.63%
	(5-6]	0.09%	0.24%	0.29%	0.28%	0.26%	0.22%	0.10%	0.00%	1.48%
	(6-7]	0.20%	0.47%	0.53%	0.51%	0.50%	0.43%	0.25%	0.00%	2.89%
	(7-8]	0.46%	0.54%	0.61%	0.63%	0.60%	0.51%	0.54%	0.00%	3.90%
	(8-9]	0.71%	0.54%	0.55%	0.56%	0.58%	0.51%	0.82%	0.00%	4.27%
	(9-10]	0.91%	0.52%	0.51%	0.48%	0.57%	0.47%	0.91%	0.00%	4.37%
	(10-11]	0.92%	0.47%	0.44%	0.46%	0.51%	0.43%	0.89%	0.00%	4.13%
	(11-12]	0.89%	0.45%	0.46%	0.40%	0.46%	0.41%	0.81%	0.00%	3.89%
	(12-13]	0.81%	0.47%	0.49%	0.42%	0.45%	0.41%	0.67%	0.00%	3.73%
	(13-14]	0.63%	0.43%	0.44%	0.40%	0.41%	0.37%	0.57%	0.00%	3.25%
	(14-15]	0.56%	0.40%	0.39%	0.37%	0.38%	0.34%	0.50%	0.00%	2.96%
	(15-16]	0.58%	0.43%	0.45%	0.43%	0.40%	0.38%	0.52%	0.00%	3.18%
	(16-17]	0.65%	0.60%	0.60%	0.57%	0.55%	0.48%	0.56%	0.00%	4.01%
	(17-18]	0.69%	0.84%	0.83%	0.82%	0.75%	0.57%	0.57%	0.00%	5.09%
	(18-19]	0.57%	0.93%	0.94%	0.90%	0.81%	0.56%	0.44%	0.00%	5.15%
	(19-20]	0.44%	0.86%	0.88%	0.83%	0.74%	0.47%	0.34%	0.00%	4.55%
	(20-21]	0.35%	0.69%	0.73%	0.65%	0.62%	0.37%	0.26%	0.00%	3.68%
	(21-22]	0.28%	0.52%	0.58%	0.53%	0.48%	0.28%	0.21%	0.00%	2.88%
	(22-23]	0.20%	0.34%	0.38%	0.33%	0.32%	0.20%	0.17%	0.00%	1.95%
	(23-24]	0.13%	0.20%	0.21%	0.20%	0.19%	0.14%	0.11%	0.00%	1.18%
	(blank)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	30.91%	30.91%
Grand Total		10.38%	10.29%	10.73%	10.16%	10.01%	7.94%	9.56%	30.91%	100.00%

Appendix II. Day of the Week vs. Distance

Percentage		Day of the Week (local)								Grand Total
		SUN	MON	TUE	WED	THU	FRI	SAT	(blank)	
Distance (kilometer)	[0-0.5]	0.20%	0.22%	0.22%	0.21%	0.21%	0.18%	0.18%	0.75%	2.18%
	(0.5-1]	0.10%	0.13%	0.13%	0.13%	0.13%	0.11%	0.10%	0.47%	1.30%
	(1-1.5]	0.14%	0.16%	0.17%	0.18%	0.15%	0.13%	0.13%	0.45%	1.51%
	(1.5-2]	0.29%	0.43%	0.42%	0.41%	0.39%	0.30%	0.29%	1.10%	3.64%
	(2-2.5]	0.29%	0.40%	0.40%	0.38%	0.37%	0.28%	0.28%	0.98%	3.38%
	(2.5-3]	0.29%	0.40%	0.40%	0.39%	0.36%	0.28%	0.28%	0.86%	3.25%
	(3-3.5]	0.58%	0.84%	0.87%	0.80%	0.79%	0.60%	0.60%	2.16%	7.25%
	(3.5-4]	0.42%	0.55%	0.58%	0.51%	0.50%	0.40%	0.40%	1.59%	4.95%
	(4-4.5]	0.46%	0.59%	0.59%	0.57%	0.53%	0.42%	0.45%	1.55%	5.15%
	(4.5-5]	0.72%	0.92%	0.96%	0.92%	0.89%	0.69%	0.76%	2.57%	8.43%
	(5-5.5]	1.02%	1.19%	1.30%	1.18%	1.19%	0.91%	1.07%	3.33%	11.19%
	(5.5-6]	0.46%	0.52%	0.56%	0.52%	0.51%	0.41%	0.45%	1.50%	4.92%
	(6-6.5]	0.59%	0.65%	0.71%	0.67%	0.64%	0.51%	0.57%	1.95%	6.28%
	(6.5-7]	0.40%	0.46%	0.51%	0.48%	0.48%	0.36%	0.41%	1.39%	4.48%
	(7-7.5]	0.35%	0.35%	0.37%	0.35%	0.36%	0.30%	0.33%	1.10%	3.52%
	(7.5-8]	0.24%	0.22%	0.23%	0.23%	0.23%	0.17%	0.21%	0.73%	2.26%
	(8-8.5]	0.52%	0.50%	0.53%	0.51%	0.51%	0.38%	0.48%	1.67%	5.10%
	(8.5-9]	0.22%	0.19%	0.19%	0.20%	0.18%	0.15%	0.20%	0.63%	1.96%
	(9-9.5]	0.19%	0.13%	0.15%	0.14%	0.15%	0.12%	0.16%	0.48%	1.52%
	(9.5-10]	0.56%	0.35%	0.37%	0.36%	0.38%	0.30%	0.44%	1.27%	4.03%
	(10-10.5]	0.51%	0.26%	0.27%	0.27%	0.28%	0.21%	0.35%	0.95%	3.10%
	(10.5-11]	0.21%	0.12%	0.12%	0.12%	0.12%	0.11%	0.16%	0.44%	1.40%
	(11-11.5]	0.21%	0.13%	0.14%	0.13%	0.13%	0.11%	0.18%	0.46%	1.49%
	(11.5-12]	0.15%	0.09%	0.08%	0.09%	0.09%	0.08%	0.11%	0.34%	1.04%
	(12-12.5]	0.11%	0.07%	0.06%	0.08%	0.07%	0.06%	0.10%	0.26%	0.80%
	(12.5-13]	0.15%	0.07%	0.07%	0.07%	0.07%	0.07%	0.13%	0.30%	0.92%
	(13-13.5]	0.10%	0.05%	0.05%	0.05%	0.05%	0.04%	0.09%	0.22%	0.64%
	(13.5-14]	0.07%	0.03%	0.03%	0.03%	0.03%	0.03%	0.06%	0.15%	0.45%
	(14-14.5]	0.07%	0.03%	0.03%	0.03%	0.02%	0.03%	0.05%	0.12%	0.38%
	(14.5-15]	0.10%	0.04%	0.03%	0.03%	0.03%	0.03%	0.07%	0.17%	0.50%
	(15-15.5]	0.07%	0.03%	0.02%	0.02%	0.02%	0.02%	0.04%	0.10%	0.33%
	(15.5-16]	0.05%	0.01%	0.01%	0.01%	0.02%	0.01%	0.03%	0.08%	0.24%
	(16-16.5]	0.14%	0.05%	0.04%	0.04%	0.04%	0.04%	0.12%	0.23%	0.70%
	(16.5-17]	0.04%	0.02%	0.01%	0.01%	0.01%	0.02%	0.03%	0.07%	0.22%
	(17-17.5]	0.03%	0.01%	0.01%	0.01%	0.01%	0.01%	0.02%	0.04%	0.14%
	(17.5-18]	0.04%	0.01%	0.01%	0.01%	0.01%	0.01%	0.03%	0.07%	0.21%
	(18-18.5]	0.02%	0.01%	0.01%	0.01%	0.01%	0.01%	0.02%	0.03%	0.11%
	(18.5-19]	0.02%	0.01%	0.01%	0.01%	0.01%	0.01%	0.01%	0.03%	0.10%
	(19-19.5]	0.03%	0.01%	0.01%	0.01%	0.01%	0.01%	0.02%	0.05%	0.14%
	(19.5-20]	0.03%	0.01%	0.01%	0.01%	0.01%	0.01%	0.02%	0.03%	0.10%
	(20-20.5]	0.02%	0.01%	0.00%	0.00%	0.01%	0.00%	0.01%	0.03%	0.10%
	(20.5-21]	0.04%	0.01%	0.00%	0.01%	0.01%	0.01%	0.02%	0.04%	0.13%
	(21-21.5]	0.13%	0.02%	0.01%	0.01%	0.01%	0.01%	0.06%	0.12%	0.37%
	(21.5-22]	0.05%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.04%	0.12%
	(blank)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Grand Total		10.38%	10.29%	10.73%	10.16%	10.01%	7.94%	9.56%	30.91%	100.00%

Appendix III. Speed vs. Distance

Percentage		Speed&(km/h)																																				Grand&Total
		[0#1]	(1#2]	(2#3]	(3#4]	(4#5]	(5#6]	(6#7]	(7#8]	(8#9]	(9#10]	(10#11]	(11#12]	(12#13]	(13#14]	(14#15]	(15#16]	(16#17]	(17#18]	(18#19]	(19#20]	(20#21]	(21#22]	(22#23]	(23#24]	(24#25]	(25#26]	(26#27]	(27#28]	(28#29]	(29#30]	(30#31]	(31#32]	(32#33]	(33#34]	(34#35]	(35#36]	(blank)	
Distance (kilometer)	[0#0.5]	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.2%	2.2%
	(0.5#1]	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.1%	0.0%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.6%	1.3%
	(1#1.5]	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.7%	1.5%
	(1.5#2]	0.0%	0.0%	0.0%	0.1%	0.1%	0.2%	0.1%	0.2%	0.2%	0.2%	0.2%	0.2%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.9%	3.6%
	(2#2.5]	0.0%	0.0%	0.0%	0.0%	0.1%	0.2%	0.1%	0.2%	0.2%	0.2%	0.2%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.8%	3.4%
	(2.5#3]	0.0%	0.0%	0.0%	0.0%	0.1%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.7%	3.2%
	(3#3.5]	0.0%	0.0%	0.0%	0.0%	0.1%	0.2%	0.3%	0.3%	0.5%	0.6%	0.6%	0.3%	0.3%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	3.8%	7.2%
	(3.5#4]	0.0%	0.0%	0.0%	0.0%	0.1%	0.5%	0.2%	0.3%	0.3%	0.3%	0.4%	0.2%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.5%	4.9%
	(4#4.5]	0.0%	0.0%	0.0%	0.0%	0.1%	0.2%	0.2%	0.3%	0.4%	0.4%	0.4%	0.3%	0.2%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.7%	5.1%
	(4.5#5]	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.2%	0.3%	0.5%	0.8%	0.8%	0.6%	0.4%	0.2%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	4.1%	8.4%
	(5#5.5]	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.2%	0.4%	0.7%	1.0%	1.3%	0.9%	0.6%	0.2%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	5.5%	11.2%
	(5.5#6]	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.2%	0.3%	0.4%	0.5%	0.5%	0.2%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.3%	4.9%
	(6#6.5]	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.2%	0.4%	0.6%	0.7%	0.6%	0.4%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.9%	6.3%
	(6.5#7]	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.1%	0.3%	0.4%	0.5%	0.4%	0.3%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.1%	4.5%
	(7#7.5]	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.1%	0.2%	0.3%	0.5%	0.3%	0.2%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.6%	3.5%
	(7.5#8]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.1%	0.2%	0.3%	0.2%	0.2%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.9%	2.3%
	(8#8.5]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.3%	0.4%	0.6%	0.5%	0.4%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.4%	5.1%
	(8.5#9]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.2%	0.2%	0.2%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.9%	2.0%
	(9#9.5]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.2%	0.2%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.6%	1.5%
	(9.5#10]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.2%	0.4%	0.6%	0.5%	0.3%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.6%	4.0%
	(10#10.5]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.2%	0.3%	0.5%	0.4%	0.3%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.2%	3.1%
	(10.5#11]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.2%	0.2%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.5%	1.4%
	(11#11.5]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.2%	0.2%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.6%	1.5%
	(11.5#12]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.2%	0.2%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%	1.0%
	(12#12.5]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	0.8%
	(12.5#13]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%	0.9%
	(13#13.5]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	0.6%
	(13.5#14]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.4%
	(14#14.5]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.4%
	(14.5#15]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.5%
	(15#15.5]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.3%
	(15.5#16]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.2%
	(16#16.5]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	0.7%
	(16.5#17]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.2%
	(17#17.5]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%
	(17.5#18]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.2%
	(18#18.5]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
	(18.5#19]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
	(19#19.5]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%
	(19.5#20]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
	(20#20.5]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
	(20.5#21]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%
	(21#21.5]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.4%
	(21.5#22]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%
	(blank)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Grand&Total		0.1%	0.2%	0.2%	0.4%	1.0%	2.5%	2.7%	3.6%	5.8%	8.1%	10.4%	8.0%	5.6%	2.3%	0.9%	0.3%	0.2%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	47.2%	100.0%