Alex Balló i Albert Gómez

Case alcoholic and non-alcoholic

Qüestions

3 Step: Data acquisition

How to load the data?

To load the data we are downloaded the data, and then we use the gzip library for decompress the files and then read the different files using the os library of each directory.

Another important concern is how to save the data once read it?

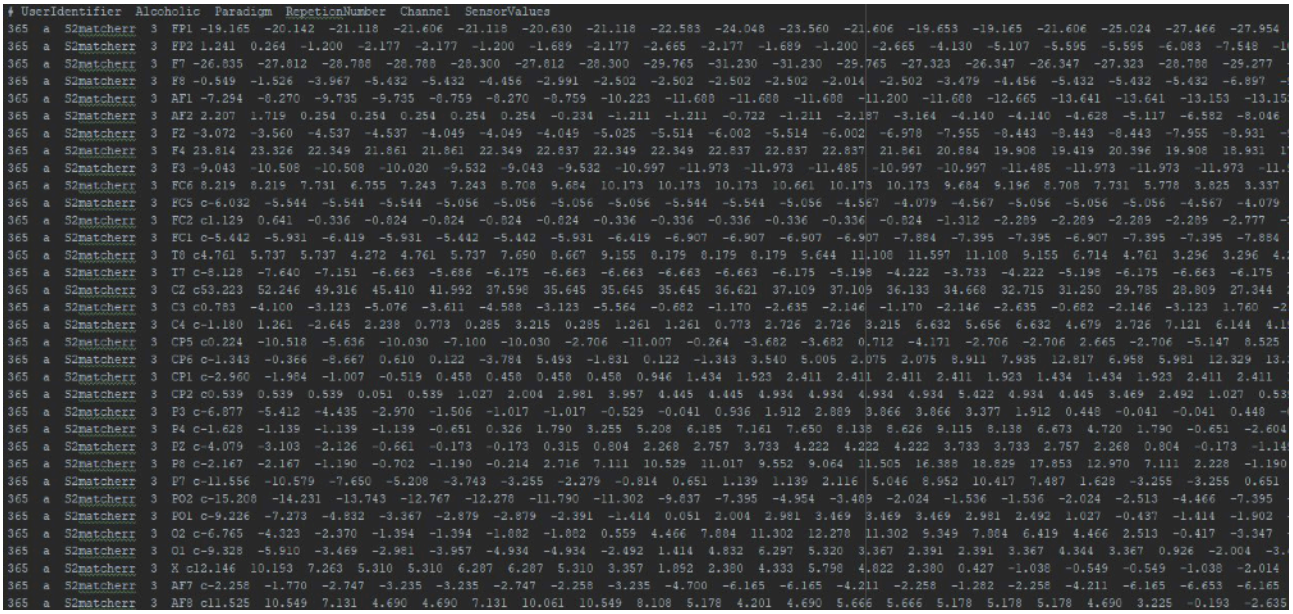
When the data are extracted and readed, we writte it in a new document with all the information that we need (user_id, alcoholic, paradigm, repetition_number, channel) .

In which format the data needs to be kept?

We save the data in a text file format. We didn't find necessary to care about the format.

How many data we want to deal with?

We process all data, because we need to see the differences better.



# UserIdentifier	Alcoholic	Paradigm	RepetionNumber	Channel	SensorValues
365	a	S2matcherr	3	FP1	-19.165 -20.142 -21.118 -21.606 -21.118 -20.630 -21.118 -22.583 -24.048 -23.560 -21.606 -19.653 -19.165 -21.606 -25.024 -27.466 -27.954
365	a	S2matcherr	3	FP2	1.241 0.264 -1.200 -2.177 -2.177 -1.200 -1.689 -2.177 -2.665 -2.177 -1.689 -1.200 -2.665 -4.130 -5.107 -5.595 -5.595 -6.083 -7.548 -1
365	a	S2matcherr	3	F7	-26.835 -27.812 -28.788 -28.788 -28.300 -27.812 -28.300 -29.765 -31.230 -31.230 -29.765 -27.323 -26.347 -26.347 -27.323 -28.788 -29.277
365	a	S2matcherr	3	F8	-0.549 -1.526 -3.967 -5.432 -5.432 -4.456 -2.991 -2.502 -2.502 -2.502 -2.502 -2.014 -2.502 -3.479 -4.456 -5.432 -5.432 -5.432 -6.897
365	a	S2matcherr	3	AF1	-7.294 -8.270 -9.735 -9.735 -8.759 -8.270 -8.759 -10.223 -11.688 -11.688 -11.688 -11.200 -11.688 -12.665 -13.641 -13.641 -13.153 -13.15
365	a	S2matcherr	3	AF2	2.207 1.719 0.254 0.254 0.254 0.254 0.254 -0.234 -1.211 -1.211 -0.722 -1.211 -2.187 -3.164 -4.140 -4.140 -4.628 -5.117 -6.582 -8.046
365	a	S2matcherr	3	FZ	-3.072 -3.560 -4.537 -4.537 -4.049 -4.049 -4.049 -5.025 -5.514 -6.002 -5.514 -6.002 -6.978 -7.955 -8.443 -8.443 -8.443 -7.955 -8.931
365	a	S2matcherr	3	F4	23.814 23.326 22.349 21.861 21.861 22.349 22.837 22.349 22.349 22.837 22.837 22.837 21.861 20.884 19.908 19.419 20.396 19.908 18.931 1
365	a	S2matcherr	3	F3	-9.043 -10.508 -10.508 -10.020 -9.532 -9.043 -9.532 -10.997 -11.973 -11.973 -11.485 -10.997 -10.997 -11.485 -11.973 -11.973 -11.973 -11.
365	a	S2matcherr	3	FC6	8.219 8.219 7.731 6.755 7.243 7.243 8.708 9.684 10.173 10.173 10.173 10.661 10.173 10.173 9.684 9.196 8.708 7.731 5.778 3.825 3.337
365	a	S2matcherr	3	FC5	c-6.032 -5.544 -5.544 -5.544 -5.056 -5.056 -5.056 -5.056 -5.544 -5.544 -5.056 -4.567 -4.079 -4.567 -5.056 -5.056 -5.056 -4.567 -4.079
365	a	S2matcherr	3	FC2	c1.129 0.641 -0.336 -0.824 -0.824 -0.824 -0.824 -0.336 -0.336 -0.336 -0.336 -0.336 -0.824 -1.312 -2.289 -2.289 -2.289 -2.289 -2.777
365	a	S2matcherr	3	FC1	c-5.442 -5.931 -6.419 -5.931 -5.442 -5.442 -5.931 -6.419 -6.907 -6.907 -6.907 -6.907 -7.884 -7.395 -7.395 -6.907 -7.395 -7.395 -7.884
365	a	S2matcherr	3	T8	c4.761 5.737 5.737 4.272 4.761 5.737 7.690 8.667 9.155 8.179 8.179 8.179 9.644 11.108 11.597 11.108 9.155 6.714 4.761 3.296 3.296 4.
365	a	S2matcherr	3	T7	c-8.128 -7.640 -7.151 -6.663 -5.686 -6.175 -6.663 -6.663 -6.663 -6.663 -6.175 -5.198 -4.222 -3.733 -4.222 -5.198 -6.175 -6.663 -6.175
365	a	S2matcherr	3	CZ	c53.223 52.246 49.316 45.410 41.992 37.598 35.645 35.645 35.645 36.621 37.109 37.109 36.133 34.668 32.715 31.250 29.785 28.809 27.344
365	a	S2matcherr	3	C3	c0.783 -4.100 -3.123 -5.076 -3.611 -4.588 -3.123 -5.564 -0.682 -1.170 -2.635 -2.146 -1.170 -2.146 -2.635 -0.682 -2.146 -3.123 1.760 -2
365	a	S2matcherr	3	C4	c-1.180 1.261 -2.645 2.238 0.773 0.285 3.215 0.285 1.261 1.261 0.773 2.726 2.726 3.215 6.632 5.656 6.632 4.679 2.726 7.121 6.144 4.1
365	a	S2matcherr	3	CP5	c0.224 -10.518 -5.636 -10.030 -7.100 -10.030 -2.706 -11.007 -0.264 -3.682 -3.682 0.712 -4.171 -2.706 -2.706 2.665 -2.706 -5.147 8.525
365	a	S2matcherr	3	CP6	c-1.343 -0.366 -8.667 0.610 0.122 -3.784 5.493 -1.831 0.122 -1.343 3.540 5.005 2.075 2.075 8.911 7.935 12.817 6.958 5.981 12.329 13.
365	a	S2matcherr	3	CP1	c-2.960 -1.984 -1.007 -0.519 0.458 0.458 0.458 0.458 0.946 1.434 1.923 2.411 2.411 2.411 2.411 1.923 1.434 1.434 1.923 2.411 2.411
365	a	S2matcherr	3	CP2	c0.539 0.539 0.539 0.051 0.539 1.027 2.004 2.981 3.957 4.445 4.445 4.934 4.934 4.934 4.934 5.422 4.934 4.445 3.469 2.492 1.027 0.53
365	a	S2matcherr	3	P3	c-6.877 -5.412 -4.435 -2.970 -1.506 -1.017 -1.017 -0.529 -0.041 0.936 1.912 2.889 3.866 3.866 3.377 1.912 0.448 -0.041 -0.041 0.448
365	a	S2matcherr	3	P4	c-1.628 -1.139 -1.139 -1.139 -0.651 0.326 1.790 3.255 5.208 6.185 7.161 7.650 8.138 8.626 9.115 8.138 6.673 4.720 1.790 -0.651 -2.604
365	a	S2matcherr	3	PZ	c-4.079 -3.103 -2.126 -0.661 -0.173 -0.173 0.315 0.804 2.268 2.757 3.733 4.222 4.222 4.222 3.733 3.733 2.757 2.268 0.804 -0.173 -1.14
365	a	S2matcherr	3	P8	c-2.167 -2.167 -1.190 -0.702 -1.190 -0.214 2.716 7.111 10.529 11.017 9.552 9.064 11.505 16.388 18.829 17.853 12.970 7.111 2.228 -1.190
365	a	S2matcherr	3	P7	c-11.556 -10.579 -7.650 -5.208 -3.743 -3.255 -2.279 -0.814 0.651 1.139 1.139 2.116 5.046 8.952 10.417 7.487 1.628 -3.255 -3.255 0.651
365	a	S2matcherr	3	PO2	c-15.208 -14.231 -13.743 -12.767 -12.278 -11.790 -11.302 -9.837 -7.395 -4.954 -3.489 -2.024 -1.536 -1.536 -2.024 -2.513 -4.466 -7.395
365	a	S2matcherr	3	PO1	c-9.226 -7.273 -4.832 -3.367 -2.879 -2.879 -2.391 -1.414 0.051 2.004 2.981 3.469 3.469 3.469 2.981 2.492 1.027 -0.437 -1.414 -1.902
365	a	S2matcherr	3	O2	c-6.765 -4.323 -2.370 -1.394 -1.394 -1.882 -1.882 0.559 4.466 7.884 11.302 12.278 11.302 9.349 7.884 6.419 4.466 2.513 -0.417 -3.347
365	a	S2matcherr	3	O1	c-9.328 -5.910 -3.469 -2.981 -3.957 -4.934 -4.934 -2.492 1.414 4.832 6.297 5.320 3.367 2.391 2.391 3.367 4.344 3.367 0.926 -2.004 -3.
365	a	S2matcherr	3	X	c12.146 10.193 7.263 5.310 5.310 6.287 6.287 5.310 3.357 1.892 2.380 4.333 5.798 4.822 2.380 0.427 -1.038 -0.549 -0.549 -1.038 -2.014
365	a	S2matcherr	3	AF7	c-2.258 -1.770 -2.747 -3.235 -3.235 -2.747 -2.258 -3.235 -4.700 -6.165 -6.165 -4.211 -2.258 -1.282 -2.258 -4.211 -6.165 -6.653 -6.165
365	a	S2matcherr	3	AF8	c11.525 10.549 7.131 4.690 4.690 7.131 10.061 10.549 8.108 5.178 4.201 4.690 5.666 5.666 5.178 5.178 5.178 4.690 3.225 -0.193 -2.635

The data goes on but we have not been able to represent all of it in screenshot.

4 Step: Data Exploration

Do you detect any problem with your data?

Yes,the problem we detect is that there are some users that have the same ID, only changing if they are alcoholic or non-alcoholic, a good data need to mark the difference between ids.

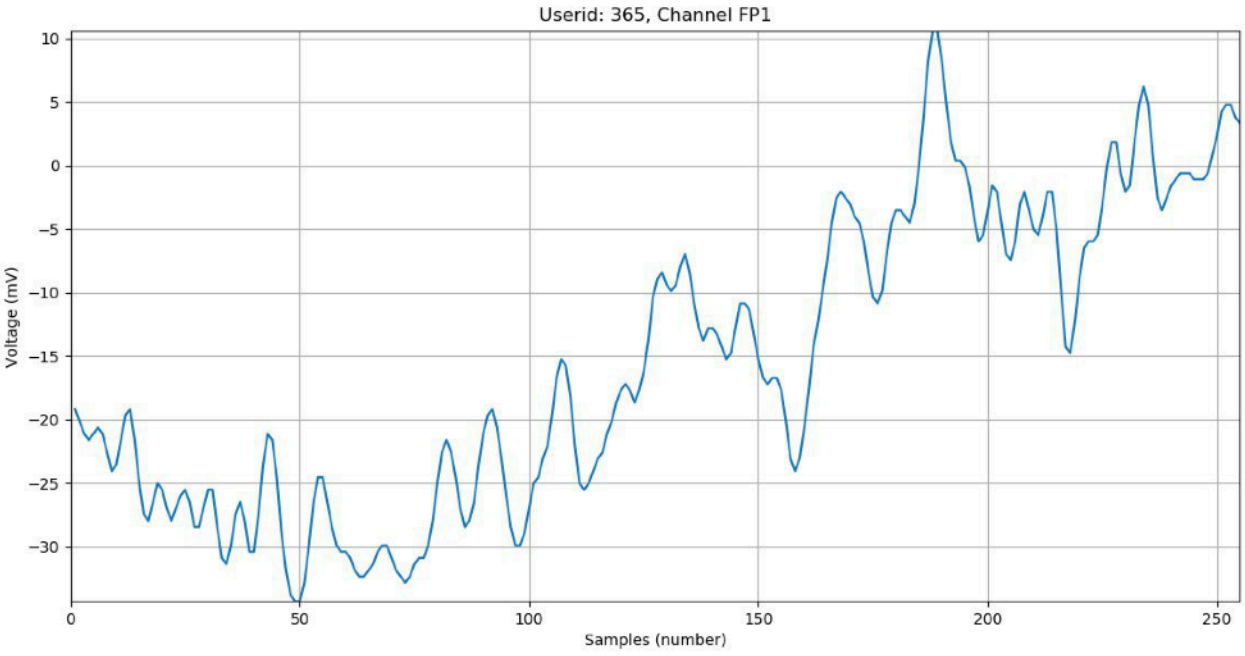
Are there any outliers?

We do not detect any more

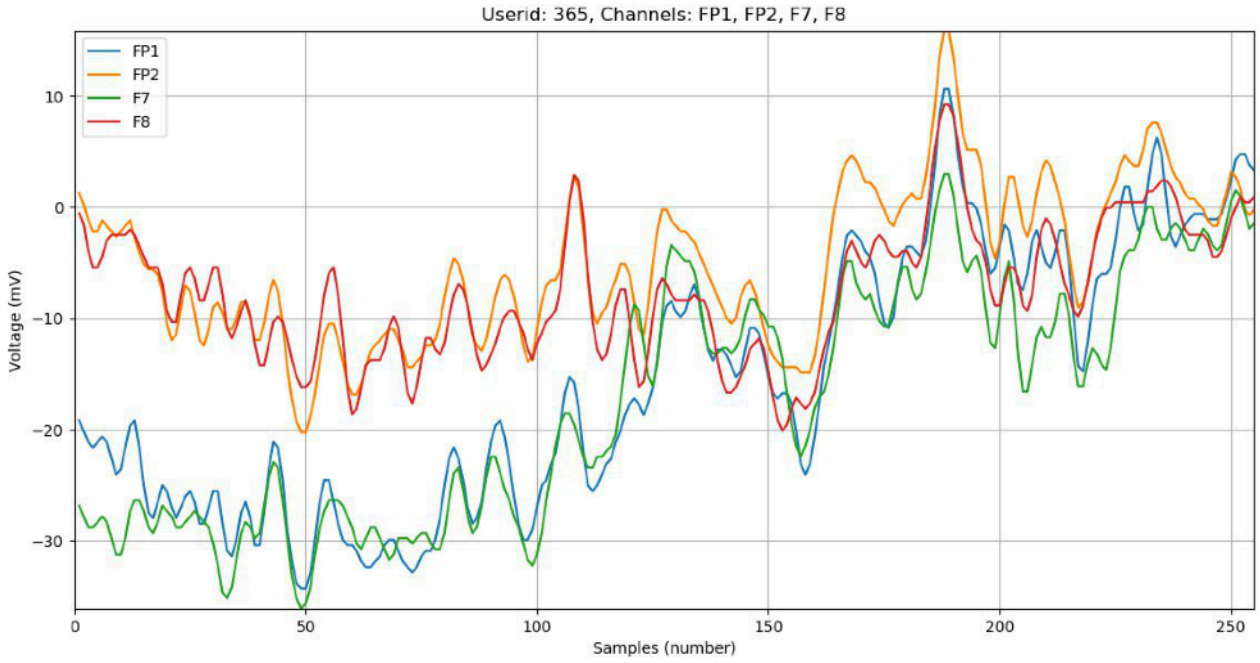
Do the users have the same number of samples?

No more users don’t have the full set of samples

4.1. Exercise 1 Represent the 'FP1' channel (first one). Be sure to correctly specify axis. Hint: we recommend to use something like this



4.2. Exercise 2 Represent the 'FP1' channel (first one) as well as the next 3



4.4. Are there any outliers? First you need to define what is an outlier?

No, we don't find any outlier. An outlier is a data that is much smaller or bigger than the near of the line or data point. More or less, the range of the channels's voltage is between -35mV and 25mV.

5 Step: Data Cleaning, Data Transfomration nd Reporting

Why data can't be Clean at first?

Because we don't have all complete data, invalid values and some values that are outliers are ingored.

Is our data in a tidy format?

At first no, but after processing and comipled the data are tidy. It could be better sorted like the channel number (from 1 to 64) in columns.