

Work on Vehicle Energy Dataset (VED) available here for free :

<https://github.com/gsoh/VED>

This dataset contains GPS trajectories, timeseries data of fuel, energy, speed etc. for different vehicles and drivers.

Each file contains data collected in a specific week.

Each file contains data of different vehicles (VehId) that may have taken several trips (Trip). So, each file contains data collected by different drivers that have taken different trips.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	DayNum	VehId	Trip	Timestamp	Latitude	Longitude	Vehicle Sp	MAF[g/se	Engine RP	Absolute	OAT[DegC	Fuel Rate	Air Con
2	36.49652	5	1443	0	42.2306	-83.6802	47	22.11	1910	58.82353	NaN	NaN	NaN
3	36.49652	5	1443	100	42.2306	-83.6802	50	22.11	1910	58.82353	NaN	NaN	NaN
4	36.49652	5	1443	1200	42.2306	-83.6802	53	22.11	1910	58.82353	NaN	NaN	NaN
5	36.49652	5	1443	1300	42.2306	-83.6802	53	20.73	1803	58.82353	NaN	NaN	NaN
6	36.49652	5	1443	2300	42.2306	-83.6802	55	20.73	1803	58.82353	NaN	NaN	NaN
7	36.49652	5	1443	2900	42.2306	-83.6802	55	20.73	1803	54.90196	NaN	NaN	NaN
8	36.49652	5	1443	3000	42.23057	-83.6808	55	20.73	1803	54.90196	NaN	NaN	NaN
9	36.49652	5	1443	3700	42.23057	-83.6808	60	21.47	1823	54.90196	NaN	NaN	NaN
10	36.49652	5	1443	4800	42.23057	-83.6808	62	21.47	1823	54.90196	NaN	NaN	NaN
11	36.49652	5	1443	5800	42.23057	-83.6808	63	19.7	1766	54.90196	NaN	NaN	NaN
12	36.49652	5	1443	6100	42.23051	-83.6814	63	19.7	1766	54.90196	NaN	NaN	NaN
13	36.49652	5	1443	6900	42.23051	-83.6814	65	19.7	1766	54.90196	NaN	NaN	NaN
14	36.49652	5	1443	7200	42.23051	-83.6814	65	19.7	1766	54.90196	NaN	NaN	NaN
15	36.49652	5	1443	7800	42.23051	-83.6814	65	21.86	1854	54.90196	NaN	NaN	NaN
44272	36.97719	445	668	441200	42.24027	-83.7392	37	NaN	0	0	NaN	NaN	NaN
44273	36.97719	445	668	442900	42.24065	-83.7392	37	NaN	0	0	NaN	NaN	NaN
44274	36.97719	445	668	443500	42.24065	-83.7392	28	NaN	0	0	NaN	NaN	NaN
44275	36.97719	445	668	444700	42.24065	-83.7392	22	NaN	0	0	NaN	NaN	NaN
44276	36.97719	445	668	445900	42.24065	-83.7392	15	NaN	0	0	NaN	NaN	NaN
44277	36.97719	445	668	446000	42.24082	-83.7393	15	NaN	0	0	NaN	NaN	NaN
44278	36.97719	445	668	447200	42.24082	-83.7393	9	NaN	0	0	NaN	NaN	NaN
44279	36.6958	450	947	0	42.2874	-83.719	41	17.42	1733	41.56863	NaN	NaN	NaN
44280	36.6958	450	947	1900	42.2874	-83.719	41	17.25	2080	42.7451	NaN	NaN	NaN
44281	36.6958	450	947	2000	42.2874	-83.719	41	17.25	2080	42.7451	NaN	NaN	NaN
44282	36.6958	450	947	2100	42.2874	-83.719	47	17.25	2080	42.7451	NaN	NaN	NaN

The first task is to cut every trip into individual files and add specific features : address, roadname, roadtype, etc.

Thanks to the GPS location of the vehicle given among the data, and the *Nominatim* API and *Geopy* package, we can retrieve the address on which the vehicle is running.

The *Spacy* package is used to retrieve the roadtypes from addresses or roadnames using NLP.

In the end, we want :

for each individual file (individual trip) : trip timelength, mean speed, mean accel, trip length in km, time spent for each roadtype
and make a unique .csv file gathering all the infos of the trips, having one trip per line

If problem encountered with Spacy :

- be aware of the Numpy version used (not necessarily the latest)
- be sure to be at User to install packages

on Anaconda Prompt :

```
> pip3 install numpy==1.23.0
```

```
> pip install -U spacy
```

The “model-best” model is available on my GitHub folder for upload, and after putting it anywhere locally, please out the right path when `spacy.load(...)`

For the roadtypes, here is what is obtained (the file is for a unique trip) :

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		DayNum	VehId	Trip	Timestamp	Latitude[d]	Longitude[d]	Vehicle Sp	MAF[g/sec]	Engine RPM	Road_name	Address	Road_type	
2	0	22.90105	5	1369	0	42.26948	-83.7391	48	4.42	1154	Packard St	747, Packa	Street	
3	2	22.90105	5	1369	1000	42.26933	-83.7391	50	15.99	1214	Arch Stree	727, Arch	Street	
4	4	22.90105	5	1369	2300	42.26933	-83.7391	53	9.4	1158	Arch Stree	727, Arch	Street	
5	6	22.90105	5	1369	4400	42.26933	-83.7391	54	4.94	1175	Arch Stree	727, Arch	Street	
6	8	22.90105	5	1369	6000	42.26879	-83.7386	55	4.94	1175	Packard St	821, Packa	Street	
7	10	22.90105	5	1369	6400	42.26879	-83.7386	55	0.87	0	Packard St	821, Packa	Street	
8	12	22.90105	5	1369	7500	42.26879	-83.7386	54	0.87	0	Packard St	821, Packa	Street	
9	14	22.90105	5	1369	10600	42.26893	-83.7376	55	0.87	0	Vaughn Str	913, Vaug	Street	
10	16	22.90105	5	1369	12700	42.26893	-83.7376	54	0.87	0	Vaughn Str	913, Vaug	Street	
11	18	22.90105	5	1369	15100	42.2684	-83.7367	55	0.87	0	East Unive	1204, East	Avenue	
12	20	22.90105	5	1369	20100	42.26731	-83.7369	55	0.87	0	Brookwoo	897, Brook	Place	
13	22	22.90105	5	1369	21700	42.26731	-83.7369	53	0.87	0	Brookwoo	897, Brook	Place	
14	24	22.90105	5	1369	24800	42.26698	-83.7367	52	0.87	0	Packard St	1028, Pack	Street	
15	26	22.90105	5	1369	26100	42.26667	-83.7364	52	0.87	0	Packard St	1102, Pack	Street	
16	28	22.90105	5	1369	30100	42.26626	-83.7359	53	0.87	0	Packard St	1108, Pack	Street	
17	30	22.90105	5	1369	33100	42.26602	-83.7354	54	0.87	0	Packard St	Argus Farn	Street	
18	32	22.90105	5	1369	39100	42.26545	-83.7346	54	0.87	0	Packard St	1209, Pack	Street	
19	34	22.90105	5	1369	41700	42.26545	-83.7346	56	0.87	0	Packard St	1209, Pack	Street	
20	36	22.90105	5	1369	43200	42.265	-83.7342	55	0.87	0	Packard St	Packard M	Street	
21	38	22.90105	5	1369	46100	42.26467	-83.7339	56	0.87	0	Packard St	1315, Pack	Street	
22	40	22.90105	5	1369	49100	42.26434	-83.7335	55	0.87	0	Granger Av	1301, Gran	Avenue	
23	42	22.90105	5	1369	52100	42.26402	-83.7332	54	0.87	0	Packard St	Packard +	Street	
24	44	22.90105	5	1369	58700	42.26354	-83.7328	54	10.7	1146	Packard St	1415, Pack	Street	
25	46	22.90105	5	1369	59300	42.26315	-83.7326	54	10.7	1146	Gardner R	1358, Gard	Road	
26	48	22.90105	5	1369	60300	42.26315	-83.7326	53	10.7	1146	Gardner R	1358, Gard	Road	
27	50	22.90105	5	1369	62000	42.26281	-83.7323	52	9.61	1168	Packard St	1542, Pack	Street	
28	52	22.90105	5	1369	64200	42.26281	-83.7323	50	6.28	1143	Packard St	1542, Pack	Street	
29	54	22.90105	5	1369	65100	42.26246	-83.7321	50	6.28	1143	Packard St	1542, Pack	Street	

Finally, after processing all the created trip files, a summary of information about each trip is created in a single file. This gives information about timelength, average acceleration, and time spent on each roadtype :

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1		VehId	Trip	Vitesse mc	Vitesse mc	Duree tota	Accelerati	Duree par type de route (sec)								
2	0	130	1848	30	8	521	0.04	[('Unknown', 376.0), ('Road', 145.0)]								
3	1	128	1570	43	12	278	-0.09	[('Road', 215.0), ('Avenue', 62.0)]								
4	2	10	3271	41	11	420	-0.02	[('Road', 308.0), ('Trail', 53.0), ('Street', 49.0), ('Drive', 3.0)]								
5	3	10	3263	26	7	1134	-0.01	[('Street', 765.0), ('Road', 106.0), ('Bikeway', 98.0), ('Way', 48.0), ('Trail', 39.0), ('Avenue', 29.0), ('Parkw								
6																

These information can give an idea if the driving was aggressive or not.

Time spent on the types of roads VS mean speeds : it would be odd to see a mean speeds that is high if the driver spent most of his time on streets.

If an average acceleration is too different from zero, it could mean that the driving was aggressive since it turned right/left too frequently, or he breaks too aggressively etc.