14542

Description:

The URL below has links to monthly reports (e.g.: https://www.e-stat.go.jp/en/stat-search/files?page=1&layout=datalist&lid=000001425362). Can we please get File 1-2: Traffic accidents occurrence by month which looks like this:

1-2@Traffic accidents by month

Month	Traffic accidentsiProvisionalj number	Traffic accidentsiProvisionalj Per day	Traffic accidentsiProvisionalj Change	Traffic accidentsiProvisionalj Component ratio	Fatalities number of people	Fatalities Per day	Fatalities Change	Fatalities Component ratio	InjuriesiProvisionalj number of people	InjuriesiProvisionalj Per day	InjuriesiProvisionalj Change	InjuriesiProvisionalj Component ratio
January	23419	755	-440	-1.8	217	7	34	18.6	27793	897	-402	-1.4
February	23369	835	2178	10.3	170	6.1	-6	-3.4	27555	984	3008	12.3
March	26442	853	1422	5.7	226	7.3	23	11.3	31406	1013	1921	6.5
April	24410	814	822	3.5	184	6.1	-7	-3.7	28994	966	1017	3.6
May	24534	791	547	2.3	208	6.7	-6	-2.8	29435	950	805	2.8
June	24769	826	-573	-2.3	176	5.9	-15	-7.9	29142	971	-667	-2.2
July	24714	797	-156	-0.6	236	7.6	29	14	29293	945	-221	-0.7
August	25450	821	1372	5.7	240	7.7	10	4.3	31026	1001	1875	6.4
September	24840	828	584	2.4	215	7.2	-7	-3.2	29504	983	643	2.2
October	27861	899	607	2.2	252	8.1	-9	-3.4	32781	1057	454	1.4
November	27726	924	356	1.3	254	8.5	2	0.8	32644	1088	437	1.4
December	***	***	***	***	***	***	***	***	***	***	***	***
Total	277534	831	6719	2.5	2378	7.1	48	2.1	329573	987	8870	2.8

We would like to put this in long format i.e. Month, Metric, Unit (i.e. number, per day etc.), period (monthly or annual) and value.

Note, that the December files look different as they have annual data but tab 1-3 has the data we are looking for

List of datasets

			Copy URL	View in list form
Statistics name	Statistics about Road Traffic			Detail
Dataset category0	Traffic accidents situation(December)			
Frequency	Yearly			
Survey date	2022			

Table number	Statistical table	Survey date	Update date	View / Download
	Traffic accidents situation(December)	2022	2023-03-02	⬇ EXCEL Report
	Table of contents	2022	2023-03-02	⬇ EXCEL Report
1 Road traffic accidents				
1-1	Traffic accidents	2022	2023-03-02	⬇ CSV ⬇ EXCEL Report
1-2	Traffic accidents per 100,000 persons	2022	2023-03-02	⬇ CSV ⬇ EXCEL Report
1-3	Trends in traffic accidents by year(1948-2021)	2022	2023-03-02	⬇ CSV ⬇ EXCEL Report
2 Fatalities				

1-3 Trends in traffic accidents by year (1948～2022)
交通事故発生状況の推移（昭和23年～令和４年）

	Traffic accidents	Index	Fatal accidents	Index	Fatalities	Index	Injuries	Index	Fatalities per 100,000 persons	Index
2008	766,394	107	5,079	32	5,209	31	945,703	96	4.08	25
2009	737,637	103	4,837	31	4,979	30	911,215	93	3.90	24
2010	725,924	101	4,808	30	4,948	30	896,297	91	3.88	24
2011	692,084	96	4,560	29	4,691	28	854,613	87	3.66	22
2012	665,157	93	4,307	27	4,438	26	825,392	84	3.47	21
2013	629,033	88	4,293	27	4,388	26	781,492	80	3.44	21
2014	573,842	80	4,013	25	4,113	25	711,374	73	3.23	20
2015	536,899	75	4,028	25	4,117	25	666,023	68	3.24	20
2016	499,201	70	3,790	24	3,904	23	618,853	63	3.07	19
2017	472,165	66	3,630	23	3,694	22	580,850	59	2.91	18
2018	430,601	60	3,449	22	3,532	21	525,846	54	2.79	17
2019	381,237	53	3,133	20	3,215	19	461,775	47	2.54	16
2020	309,178	43	2,784	18	2,839	17	369,476	38	2.25	14
2021	305,196	43	2,583	16	2,636	16	362,131	37	2.09	13
2022	300,839	42	2,550	16	2,610	16	356,601	36	2.08	13

＊1　Per capita data is calculated on the basis of the national census population or the population estimate as of October 1st of the previous year compiled by the Ministry of Internal Affairs and Communications.

　2　Indices are calculated with 1970 as 100.

　3　Data for 1971 and all preceding years does not cover Okinawa Prefecture.

　4　Data for 1965 and all preceding years includes property damage-only accidents.

　5　Data for 1959 and all preceding years does not include minor accidents (injury requiring less than eight days of treatment or property damage of up to 20,000 yen in value).

Root URL:

https://www.npa.go.jp/publications/statistics/koutsuu/toukeihyo_e.html

Job Frequency:

Daily

Output Columns:

Column name	Datatype	Xpath/location on site/number in screenshot	Example value	Comment
scrape_datetime	*datetime*	N/A	datetime.datetime.utcnow().isoformat()	*Ensure this is a datetime in ISO-8601 format, and have one value per run, evaluated at start of the script, rather than at the time of each request*
PERIOD_START	Datetime		2023-11-01	
PERIODICITY	Str		M	
METRIC	Str		Traffic Accidents	
UNIT	Str		Number	
VALUE	Float		27726	

Timeline:

You may complete this job any time and submit any required files to the linked Github repository within one week of accepting the job.

Please submit your code here: repo to be shared

Submission Files:

Sample.csv for sample data

A requirement.txt

scrape/ - containing all of the source code

Main file: scrape.py that will be run with a output $filename.

History.csv

Job Schema/Output Format:

You should save the output csv using these settings from a pandas DataFrame:

```
encoding="utf-8",
 line_terminator="\n",
 quotechar='"',
 quoting=csv.QUOTE_ALL,
 index=False
```

Package preferences and Scraping notes

Selenium is a last resort (only purposeful obfuscation from website owner)

Generally unnecessary to run a full browser

Please provide reasoning if using Selenium

Beautiful soup

Avoid hard codes, references should be relative

Parse json from an api rather than html

Runtime Environment:

Your code will be copied form the root to/usr/src/scrape

You should feel free to modify the requirements as you need. However, you must keep the awscli dependency

You may also upload additional binaries into the repository root and reference them there.

Please do not change the Dockerfile or shell scripts in the repository as this will cause automated test failure.

python scrape.py $filename

Page access limitations (max requests / day):

If you encounter a captcha during your scrape job, please contact the job poster before continuing.

10% of website traffic max