

PROJET BIG DATA - THOMAS RIQUELME - 22/05/2020

ANALYSE DE L'IMPACT DE L'ÉPIDÉMIE COVID-19 SUR L'ÉCONOMIE FRANÇAISE

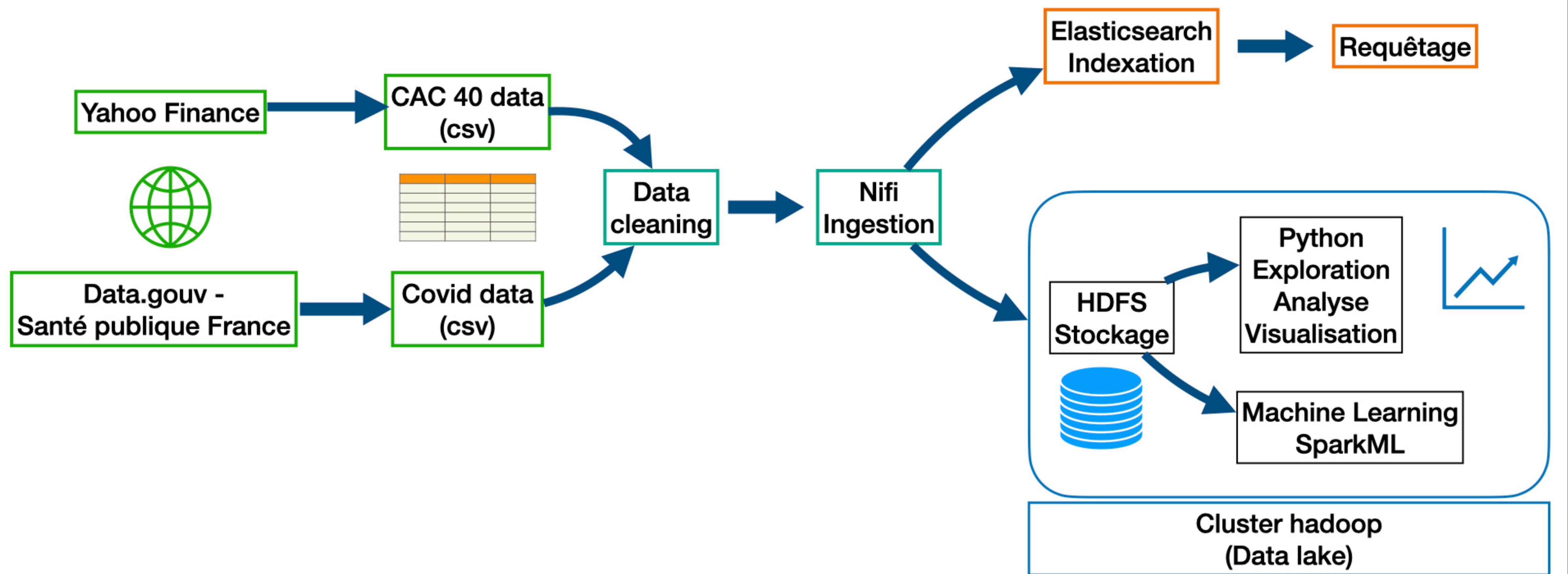
SOMMAIRE

- ▶ Problématique
- ▶ Architecture
- ▶ Data extraction
- ▶ Data ingestion
- ▶ Data storage
- ▶ Data indexation
- ▶ Data exploration, analysis and visualization
- ▶ Machine learning analysis

PROBLÉMATIQUE

- ▶ Quel est l'impact de l'épidémie Covid-19 sur l'économie française ?
- ▶ Sources de données:
 - ▶ nombre de personnes hospitalisés pour le Covid
 - ▶ le cours du CAC 40 = indicateur de l'économie des grandes entreprises françaises

ARCHITECTURE



DATA EXTRACTION

- ▶ Identification de sites open data :
 - ▶ data.gouv.fr: données hospitalières relatives à l'épidémie de COVID-19
 - ▶ fr.finance.yahoo.com: cours du CAC 40

DATA EXTRACTION : COVID-19



data.gouv.fr

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

Plateforme ouverte des données publiques françaises

Données

Réutilisations

Organisations

Tableau de bord

Documentation

🇫🇷

Connexion / Inscription

🔍

Recherche

Données hospitalières relatives à l'épidémie de COVID-19

Ce jeu de données provient d'un service public certifié

COVID-19

Les actions de Santé publique France

Santé publique France a pour mission d'améliorer et de protéger la santé des populations. Durant la crise sanitaire liée à l'épidémie du COVID-19, Santé publique France se charge de surveiller et comprendre la dynamique de l'épidémie, d'anticiper les différents scénarii et de mettre en place des actions pour prévenir et limiter la transmission de ce virus sur le territoire national.

Description du jeu de données

Le présent jeu de données renseigne sur la situation hospitalières concernant l'épidémie de COVID-19.

Quatre fichiers sont proposés :

- Les données hospitalières relatives à l'épidémie du COVID-19 par département et sexe du patient : nombre de patients hospitalisés, nombre de personnes actuellement en réanimation ou soins intensifs,

Producteur



Santé publique France

🏢

VOIR LE PROFIL

✉️

CONTACTER

PRODUCTEUR CERTIFIÉ

Santé publique France est l'agence nationale de santé publique. Créée en mai 2016 par ordonnance et décret, c'est un établissement public administratif sous tutelle du ministère...

Ressources

Voir aussi : ressources communautaires

Fichier principal

donnees-hospitalieres-covid19-2020-05-21-19h00.csv

📄 csv (624.0Ko) 📶 15192 Disponible

PRÉVISUALISER

TÉLÉCHARGER

📄

donnees-hospitalieres-nouveaux-covid19-2020-05-21-19h00.csv

📄 csv (150.7Ko) 📶 11009 Disponible

PRÉVISUALISER

TÉLÉCHARGER

📄

donnees-hospitalieres-classe-age-covid19-2020-05-21-19h00.csv

📄 csv (382.0Ko) 📶 5149 Disponible

PRÉVISUALISER

TÉLÉCHARGER

📄

donnees-hospitalieres-etablissements-covid19-2020-05-21-19h00.csv

📄 csv (137.3Ko) 📶 4775 Disponible

PRÉVISUALISER

TÉLÉCHARGER

📄


Documentation

metadonnees-donnees-hospitalieres-covid19.csv


DATA EXTRACTION : COVID-19 – FORMAT CSV

dep	↕ sexe	↕ jour	↕ hosp	↕ rea	↕ rad	↕ dc	↕
01	0	2020-03-18	2	0	1	0	
01	1	2020-03-18	1	0	1	0	
01	2	2020-03-18	1	0	0	0	
02	0	2020-03-18	41	10	18	11	
02	1	2020-03-18	19	4	11	6	
02	2	2020-03-18	22	6	7	5	
03	0	2020-03-18	4	0	1	0	
03	1	2020-03-18	1	0	0	0	
03	2	2020-03-18	3	0	1	0	
04	0	2020-03-18	3	1	2	0	
04	1	2020-03-18	3	1	0	0	
04	2	2020-03-18	0	0	2	0	
05	0	2020-03-18	8	1	9	0	
05	1	2020-03-18	1	0	7	0	
05	2	2020-03-18	7	1	2	0	

DATA EXTRACTION : CAC 40 – FORMAT CSV



Rechercher des actualités, des symboles ou des entrepri:



Durée: 01 févr. 2020 - 18 mai 2020

Afficher: Cours historiques

Fréquence: Quotidiennement

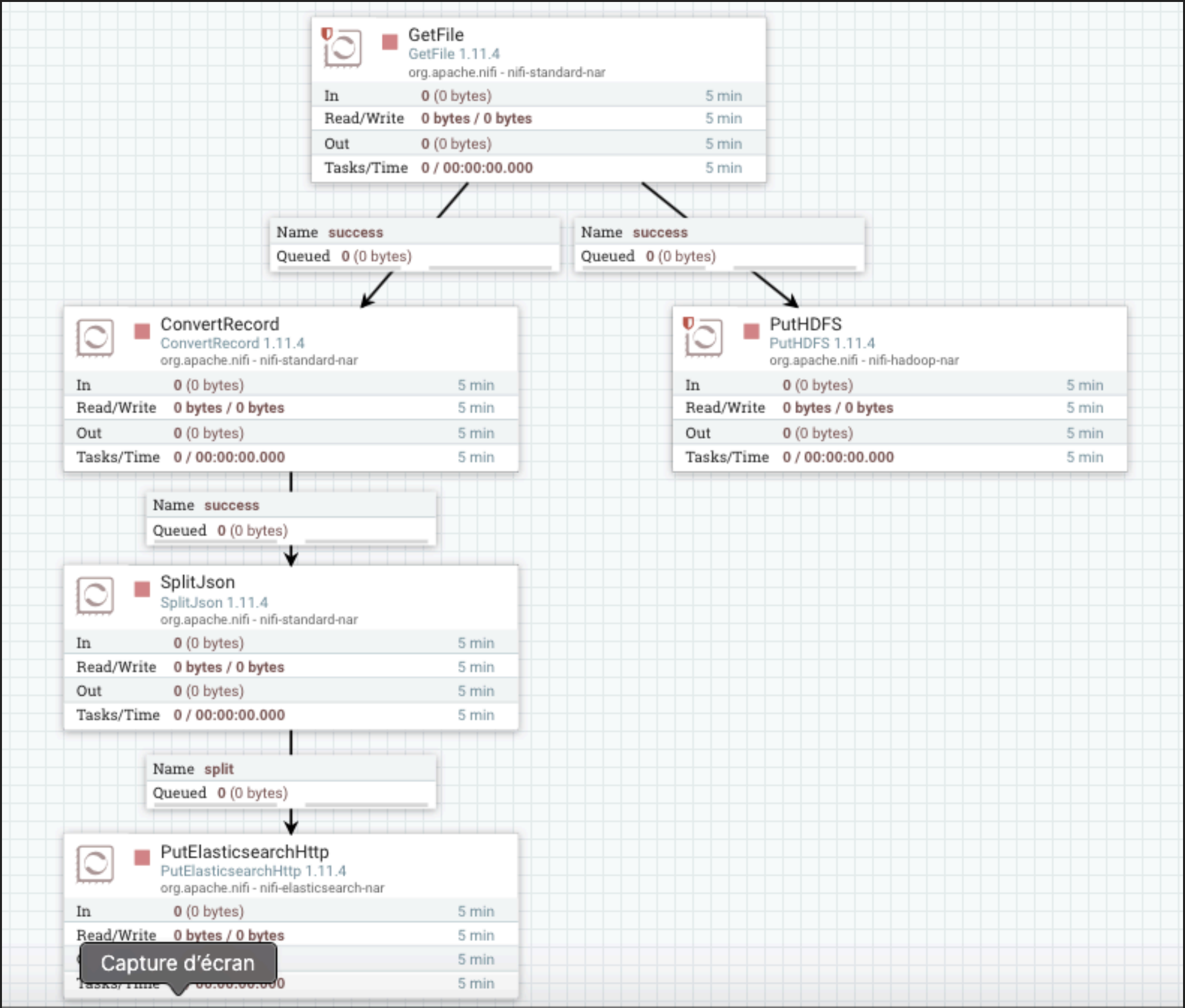
Appliquer

Devise en EUR

Télécharger

Date	Ouverture	Élevé	Faible	Clôture*	Cours de clôture ajusté**	Volume
15 mai 2020	4314,97	4335,77	4255,19	4277,63	4277,63	111236400
14 mai 2020	4294,01	4313,26	4194,58	4273,13	4273,13	127632000
13 mai 2020	4420,50	4422,74	4336,56	4344,95	4344,95	103285300
12 mai 2020	4473,43	4499,39	4453,65	4472,50	4472,50	99269700
11 mai 2020	4561,19	4569,76	4460,24	4490,22	4490,22	83493200
08 mai 2020	4538,09	4560,19	4519,59	4549,64	4549,64	49408600


DATA INGESTION : NIFI



STOCKAGE : HDFS

```
thomas@MacBook-Air-de-Thomas:/usr/local/Cellar/hadoop/3.2.1_1/libexec$ hdfs dfs -ls /project
2020-05-22 08:48:39,554 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform
... using builtin-java classes where applicable
Found 6 items
-rw-r--r--  1 thomas supergroup      5847 2020-05-19 15:30 /project/cac40_2020.csv
-rw-r--r--  1 thomas supergroup      1756 2020-05-21 12:49 /project/cac40_clean_data.csv
-rw-r--r--  1 thomas supergroup      2079 2020-05-21 12:49 /project/covid_clean_data.csv
-rw-r--r--  1 thomas supergroup    608682 2020-05-19 15:30 /project/donnes-hospitalieres-covid19.csv
-rw-r--r--  1 thomas supergroup     36444 2020-05-19 15:30 /project/french_economy_indicators.csv
-rw-r--r--  1 thomas supergroup      1857 2020-05-21 12:49 /project/joined_cac40_covid_data.csv
thomas@MacBook-Air-de-Thomas:/usr/local/Cellar/hadoop/3.2.1_1/libexec$
```

INDEXATION : ELASTICSEARCH-KIBANA - CAC 40

 kibana

Discover

Visualize

Dashboard

Timelion

Dev Tools

Management


Dev Tools

Console

1776 - "query": {
1777 - "match_all": {}
1778 - }
1779 - }
1780
1781 GET _cat/indices
1782
1783 GET _cat/indices?h=h,s,i,id,p,r,dc,dd,ss,creation
1784 .date.string
1785 GET cac40_2020/_search
1786 {
1787 "query": {
1788 "match_all": {}
1789 }
1790 }
1791
1792 GET french_economy_indicators/_search
1793 {
1794 "query": {
1795 "match_all": {}
1796 }
1797 }
1798
1799 GET donnees-hospitalieres-covid19/_search
1800 {
1801 "query": {
1802 "match_all": {}
1803 }
1804 }
1805 }
1806 }
1807 }
1808 }
1809 }
1810 }
1811 }
1812 }
1813 }
1814 }
1815 }
1816 }
1817 }
1818 }
1819 }
1820 }
1821 }
1822 }
1823 }
1824 }
1825 }
1826 }
1827 }
1828 }
1829 }
1830 }
1831 }
1832 }
1833 }
1834 }
1835 }
1836 }
1837 }
1838 }
1839 }
1840 }
1841 }
1842 }
1843 }
1844 }
1845 }
1846 }
1847 }
1848 }
1849 }
1850 }
1851 }
1852 }
1853 }
1854 }
1855 }
1856 }
1857 }
1858 }
1859 }
1860 }
1861 }
1862 }
1863 }
1864 }
1865 }
1866 }
1867 }
1868 }
1869 }
1870 }
1871 }
1872 }
1873 }
1874 }
1875 }
1876 }
1877 }
1878 }
1879 }
1880 }
1881 }
1882 }
1883 }
1884 }
1885 }
1886 }
1887 }
1888 }
1889 }
1890 }
1891 }
1892 }
1893 }
1894 }
1895 }
1896 }
1897 }
1898 }
1899 }
1900 }

1 - {
2 "took" : 6,
3 "timed_out" : false,
4 "_shards" : {
5 "total" : 5,
6 "successful" : 5,
7 "skipped" : 0,
8 "failed" : 0
9 },
10 "hits" : {
11 "total" : 72,
12 "max_score" : 1.0,
13 "hits" : [
14 {
15 "_index" : "cac40_2020",
16 "_type" : "_doc",
17 "_id" : "9BQgLXIBPUCMkXT2gF5",
18 "_score" : 1.0,
19 "_source" : {
20 "Date" : "2020-02-17",
21 "Open" : 6079.53,
22 "High" : 6088.6,
23 "Low" : 6064.92,
24 "Close" : 6085.95,
25 "Adj Close" : 6085.95,
26 "Volume" : 58001900
27 }
28 },
29 {
30 "_index" : "cac40_2020",
31 "_type" : "_doc",
32 "_id" : "9BQgLXIBPUCMkXT2gF5",
33 "_score" : 1.0,
34 "_source" : {
35 "Date" : "2020-02-17",
36 "Open" : 6079.53,
37 "High" : 6088.6,
38 "Low" : 6064.92,
39 "Close" : 6085.95,
40 "Adj Close" : 6085.95,
41 "Volume" : 58001900
42 }
43 },
44 {
45 "_index" : "cac40_2020",
46 "_type" : "_doc",
47 "_id" : "9BQgLXIBPUCMkXT2gF5",
48 "_score" : 1.0,
49 "_source" : {
50 "Date" : "2020-02-17",
51 "Open" : 6079.53,
52 "High" : 6088.6,
53 "Low" : 6064.92,
54 "Close" : 6085.95,
55 "Adj Close" : 6085.95,
56 "Volume" : 58001900
57 }
58 },
59 {
60 "_index" : "cac40_2020",
61 "_type" : "_doc",
62 "_id" : "9BQgLXIBPUCMkXT2gF5",
63 "_score" : 1.0,
64 "_source" : {
65 "Date" : "2020-02-17",
66 "Open" : 6079.53,
67 "High" : 6088.6,
68 "Low" : 6064.92,
69 "Close" : 6085.95,
70 "Adj Close" : 6085.95,
71 "Volume" : 58001900
72 }
73 },
74 {
75 "_index" : "cac40_2020",
76 "_type" : "_doc",
77 "_id" : "9BQgLXIBPUCMkXT2gF5",
78 "_score" : 1.0,
79 "_source" : {
80 "Date" : "2020-02-17",
81 "Open" : 6079.53,
82 "High" : 6088.6,
83 "Low" : 6064.92,
84 "Close" : 6085.95,
85 "Adj Close" : 6085.95,
86 "Volume" : 58001900
87 }
88 },
89 {
90 "_index" : "cac40_2020",
91 "_type" : "_doc",
92 "_id" : "9BQgLXIBPUCMkXT2gF5",
93 "_score" : 1.0,
94 "_source" : {
95 "Date" : "2020-02-17",
96 "Open" : 6079.53,
97 "High" : 6088.6,
98 "Low" : 6064.92,
99 "Close" : 6085.95,
100 "Adj Close" : 6085.95,
101 "Volume" : 58001900
102 }
103 },
104 {
105 "_index" : "cac40_2020",
106 "_type" : "_doc",
107 "_id" : "9BQgLXIBPUCMkXT2gF5",
108 "_score" : 1.0,
109 "_source" : {
110 "Date" : "2020-02-17",
111 "Open" : 6079.53,
112 "High" : 6088.6,
113 "Low" : 6064.92,
114 "Close" : 6085.95,
115 "Adj Close" : 6085.95,
116 "Volume" : 58001900
117 }
118 },
119 {
120 "_index" : "cac40_2020",
121 "_type" : "_doc",
122 "_id" : "9BQgLXIBPUCMkXT2gF5",
123 "_score" : 1.0,
124 "_source" : {
125 "Date" : "2020-02-17",
126 "Open" : 6079.53,
127 "High" : 6088.6,
128 "Low" : 6064.92,
129 "Close" : 6085.95,
130 "Adj Close" : 6085.95,
131 "Volume" : 58001900
132 }
133 },
134 {
135 "_index" : "cac40_2020",
136 "_type" : "_doc",
137 "_id" : "9BQgLXIBPUCMkXT2gF5",
138 "_score" : 1.0,
139 "_source" : {
140 "Date" : "2020-02-17",
141 "Open" : 6079.53,
142 "High" : 6088.6,
143 "Low" : 6064.92,
144 "Close" : 6085.95,
145 "Adj Close" : 6085.95,
146 "Volume" : 58001900
147 }
148 },
149 {
150 "_index" : "cac40_2020",
151 "_type" : "_doc",
152 "_id" : "9BQgLXIBPUCMkXT2gF5",
153 "_score" : 1.0,
154 "_source" : {
155 "Date" : "2020-02-17",
156 "Open" : 6079.53,
157 "High" : 6088.6,
158 "Low" : 6064.92,
159 "Close" : 6085.95,
160 "Adj Close" : 6085.95,
161 "Volume" : 58001900
162 }
163 },
164 {
165 "_index" : "cac40_2020",
166 "_type" : "_doc",
167 "_id" : "9BQgLXIBPUCMkXT2gF5",
168 "_score" : 1.0,
169 "_source" : {
170 "Date" : "2020-02-17",
171 "Open" : 6079.53,
172 "High" : 6088.6,
173 "Low" : 6064.92,
174 "Close" : 6085.95,
175 "Adj Close" : 6085.95,
176 "Volume" : 58001900
177 }
178 },
179 {
180 "_index" : "cac40_2020",
181 "_type" : "_doc",
182 "_id" : "9BQgLXIBPUCMkXT2gF5",
183 "_score" : 1.0,
184 "_source" : {
185 "Date" : "2020-02-17",
186 "Open" : 6079.53,
187 "High" : 6088.6,
188 "Low" : 6064.92,
189 "Close" : 6085.95,
190 "Adj Close" : 6085.95,
191 "Volume" : 58001900
192 }
193 },
194 {
195 "_index" : "cac40_2020",
196 "_type" : "_doc",
197 "_id" : "9BQgLXIBPUCMkXT2gF5",
198 "_score" : 1.0,
199 "_source" : {
200 "Date" : "2020-02-17",
201 "Open" : 6079.53,
202 "High" : 6088.6,
203 "Low" : 6064.92,
204 "Close" : 6085.95,
205 "Adj Close" : 6085.95,
206 "Volume" : 58001900
207 }
208 },
209 {
210 "_index" : "cac40_2020",
211 "_type" : "_doc",
212 "_id" : "9BQgLXIBPUCMkXT2gF5",
213 "_score" : 1.0,
214 "_source" : {
215 "Date" : "2020-02-17",
216 "Open" : 6079.53,
217 "High" : 6088.6,
218 "Low" : 6064.92,
219 "Close" : 6085.95,
220 "Adj Close" : 6085.95,
221 "Volume" : 58001900
222 }
223 },
224 {
225 "_index" : "cac40_2020",
226 "_type" : "_doc",
227 "_id" : "9BQgLXIBPUCMkXT2gF5",
228 "_score" : 1.0,
229 "_source" : {
230 "Date" : "2020-02-17",
231 "Open" : 6079.53,
232 "High" : 6088.6,
233 "Low" : 6064.92,
234 "Close" : 6085.95,
235 "Adj Close" : 6085.95,
236 "Volume" : 58001900
237 }
238 },
239 {
240 "_index" : "cac40_2020",
241 "_type" : "_doc",
242 "_id" : "9BQgLXIBPUCMkXT2gF5",
243 "_score" : 1.0,
244 "_source" : {
245 "Date" : "2020-02-17",
246 "Open" : 6079.53,
247 "High" : 6088.6,
248 "Low" : 6064.92,
249 "Close" : 6085.95,
250 "Adj Close" : 6085.95,
251 "Volume" : 58001900
252 }
253 },
254 {
255 "_index" : "cac40_2020",
256 "_type" : "_doc",
257 "_id" : "9BQgLXIBPUCMkXT2gF5",
258 "_score" : 1.0,
259 "_source" : {
260 "Date" : "2020-02-17",
261 "Open" : 6079.53,
262 "High" : 6088.6,
263 "Low" : 6064.92,
264 "Close" : 6085.95,
265 "Adj Close" : 6085.95,
266 "Volume" : 58001900
267 }
268 },
269 {
270 "_index" : "cac40_2020",
271 "_type" : "_doc",
272 "_id" : "9BQgLXIBPUCMkXT2gF5",
273 "_score" : 1.0,
274 "_source" : {
275 "Date" : "2020-02-17",
276 "Open" : 6079.53,
277 "High" : 6088.6,
278 "Low" : 6064.92,
279 "Close" : 6085.95,
280 "Adj Close" : 6085.95,
281 "Volume" : 58001900
282 }
283 },
284 {
285 "_index" : "cac40_2020",
286 "_type" : "_doc",
287 "_id" : "9BQgLXIBPUCMkXT2gF5",
288 "_score" : 1.0,
289 "_source" : {
290 "Date" : "2020-02-17",
291 "Open" : 6079.53,
292 "High" : 6088.6,
293 "Low" : 6064.92,
294 "Close" : 6085.95,
295 "Adj Close" : 6085.95,
296 "Volume" : 58001900
297 }
298 },
299 {
300 "_index" : "cac40_2020",
301 "_type" : "_doc",
302 "_id" : "9BQgLXIBPUCMkXT2gF5",
303 "_score" : 1.0,
304 "_source" : {
305 "Date" : "2020-02-17",
306 "Open" : 6079.53,
307 "High" : 6088.6,
308 "Low" : 6064.92,
309 "Close" : 6085.95,
310 "Adj Close" : 6085.95,
311 "Volume" : 58001900
312 }
313 },
314 {
315 "_index" : "cac40_2020",
316 "_type" : "_doc",
317 "_id" : "9BQgLXIBPUCMkXT2gF5",
318 "_score" : 1.0,
319 "_source" : {
320 "Date" : "2020-02-17",
321 "Open" : 6079.53,
322 "High" : 6088.6,
323 "Low" : 6064.92,
324 "Close" : 6085.95,
325 "Adj Close" : 6085.95,
326 "Volume" : 58001900
327 }
328 },
329 {
330 "_index" : "cac40_2020",
331 "_type" : "_doc",
332 "_id" : "9BQgLXIBPUCMkXT2gF5",
333 "_score" : 1.0,
334 "_source" : {
335 "Date" : "2020-02-17",
336 "Open" : 6079.53,
337 "High" : 6088.6,
338 "Low" : 6064.92,
339 "Close" : 6085.95,
340 "Adj Close" : 6085.95,
341 "Volume" : 58001900
342 }
343 },
344 {
345 "_index" : "cac40_2020",
346 "_type" : "_doc",
347 "_id" : "9BQgLXIBPUCMkXT2gF5",
348 "_score" : 1.0,
349 "_source" : {
350 "Date" : "2020-02-17",
351 "Open" : 6079.53,
352 "High" : 6088.6,
353 "Low" : 6064.92,
354 "Close" : 6085.95,
355 "Adj Close" : 6085.95,
356 "Volume" : 58001900
357 }
358 },
359 {
360 "_index" : "cac40_2020",
361 "_type" : "_doc",
362 "_id" : "9BQgLXIBPUCMkXT2gF5",
363 "_score" : 1.0,
364 "_source" : {
365 "Date" : "2020-02-17",
366 "Open" : 6079.53,
367 "High" : 6088.6,
368 "Low" : 6064.92,
369 "Close" : 6085.95,
370 "Adj Close" : 6085.95,
371 "Volume" : 58001900
372 }
373 },
374 {
375 "_index" : "cac40_2020",
376 "_type" : "_doc",
377 "_id" : "9BQgLXIBPUCMkXT2gF5",
378 "_score" : 1.0,
379 "_source" : {
380 "Date" : "2020-02-17",
381 "Open" : 6079.53,
382 "High" : 6088.6,
383 "Low" : 6064.92,
384 "Close" : 6085.95,
385 "Adj Close" : 6085.95,
386 "Volume" : 58001900
387 }
388 },
389 {
390 "_index" : "cac40_2020",
391 "_type" : "_doc",
392 "_id" : "9BQgLXIBPUCMkXT2gF5",
393 "_score" : 1.0,
394 "_source" : {
395 "Date" : "2020-02-17",
396 "Open" : 6079.53,
397 "High" : 6088.6,
398 "Low" : 6064.92,
399 "Close" : 6085.95,
400 "Adj Close" : 6085.95,
401 "Volume" : 58001900
402 }
403 },
404 {
405 "_index" : "cac40_2020",
406 "_type" : "_doc",
407 "_id" : "9BQgLXIBPUCMkXT2gF5",
408 "_score" : 1.0,
409 "_source" : {
410 "Date" : "2020-02-17",
411 "Open" : 6079.53,
412 "High" : 6088.6,
413 "Low" : 6064.92,
414 "Close" : 6085.95,
415 "Adj Close" : 6085.95,
416 "Volume" : 58001900
417 }
418 },
419 {
420 "_index" : "cac40_2020",
421 "_type" : "_doc",
422 "_id" : "9BQgLXIBPUCMkXT2gF5",
423 "_score" : 1.0,
424 "_source" : {
425 "Date" : "2020-02-17",
426 "Open" : 6079.53,
427 "High" : 6088.6,
428 "Low" : 6064.92,
429 "Close" : 6085.95,
430 "Adj Close" : 6085.95,
431 "Volume" : 58001900
432 }
433 },
434 {
435 "_index" : "cac40_2020",
436 "_type" : "_doc",
437 "_id" : "9BQgLXIBPUCMkXT2gF5",
438 "_score" : 1.0,
439 "_source" : {
440 "Date" : "2020-02-17",
441 "Open" : 6079.53,
442 "High" : 6088.6,
443 "Low" : 6064.92,
444 "Close" : 6085.95,
445 "Adj Close" : 6085.95,
446 "Volume" : 58001900
447 }
448 },
449 {
450 "_index" : "cac40_2020",
451 "_type" : "_doc",
452 "_id" : "9BQgLXIBPUCMkXT2gF5",
453 "_score" : 1.0,
454 "_source" : {
455 "Date" : "2020-02-17",
456 "Open" : 6079.53,
457 "High" : 6088.6,
458 "Low" : 6064.92,
459 "Close" : 6085.95,
460 "Adj Close" : 6085.95,
461 "Volume" : 58001900
462 }
463 },
464 {
465 "_index" : "cac40_2020",
466 "_type" : "_doc",
467 "_id" : "9BQgLXIBPUCMkXT2gF5",
468 "_score" : 1.0,
469 "_source" : {
470 "Date" : "2020-02-17",
471 "Open" : 6079.53,
472 "High" : 6088.6,
473 "Low" : 6064.92,
474 "Close" : 6085.95,
475 "Adj Close" : 6085.95,
476 "Volume" : 58001900
477 }
478 },
479 {
480 "_index" : "cac40_2020",
481 "_type" : "_doc",
482 "_id" : "9BQgLXIBPUCMkXT2gF5",
483 "_score" : 1.0,
484 "_source" : {
485 "Date" : "2020-02-17",
486 "Open" : 6079.53,
487 "High" : 6088.6,
488 "Low" : 6064.92,
489 "Close" : 6085.95,
490 "Adj Close" : 6085.95,
491 "Volume" : 58001900
492 }
493 },
494 {
495 "_index" : "cac40_2020",
496 "_type" : "_doc",
497 "_id" : "9BQgLXIBPUCMkXT2gF5",
498 "_score" : 1.0,
499 "_source" : {
500 "Date" : "2020-02-17",
501 "Open" : 6079.53,
502 "High" : 6088.6,
503 "Low" : 6064.92,
504 "Close" : 6085.95,
505 "Adj Close" : 6085.95,
506 "Volume" : 58001900
507 }
508 },
509 {
510 "_index" : "cac40_2020",
511 "_type" : "_doc",
512 "_id" : "9BQgLXIBPUCMkXT2gF5",
513 "_score" : 1.0,
514 "_source" : {
515 "Date" : "2020-02-17",
516 "Open" : 6079.53,
517 "High" : 6088.6,
518 "Low" : 6064.92,
519 "Close" : 6085.95,
520 "Adj Close" : 6085.95,
521 "Volume" : 58001900
522 }
523 },
524 {
525 "_index" : "cac40_2020",
526 "_type" : "_doc",
527 "_id" : "9BQgLXIBPUCMkXT2gF5",
528 "_score" : 1.0,
529 "_source" : {
530 "Date" : "2020-02-17",
531 "Open" : 6079.53,
532 "High" : 6088.6,
533 "Low" : 6064.92,
534 "Close" : 6085.95,
535 "Adj Close" : 6085.95,
536 "Volume" : 58001900
537 }
538 },
539 {
540 "_index" : "cac40_2020",
541 "_type" : "_doc",
542 "_id" : "9BQgLXIBPUCMkXT2gF5",
543 "_score" : 1.0,
544 "_source" : {
545 "Date" : "2020-02-17",
546 "Open" : 6079.53,
547 "High" : 6088.6,
548 "Low" : 6064.92,
549 "Close" : 6085.95,
550 "Adj Close" : 6085.95,
551 "Volume" : 58001900
552 }
553 },
554 {
555 "_index" : "cac40_2020",
556 "_type" : "_doc",
557 "_id" : "9BQgLXIBPUCMkXT2gF5",
558 "_score" : 1.0,
559 "_source" : {
560 "Date" : "2020-02-17",
561 "Open" : 6079.53,
562 "High" : 6088.6,
563 "Low" : 6064.92,
564 "Close" : 6085.95,
565 "Adj Close" : 6085.95,
566 "Volume" : 58001900
567 }
568 },
569 {
570 "_index" : "cac40_2020",
571 "_type" : "_doc",
572 "_id" : "9BQgLXIBPUCMkXT2gF5",
573 "_score" : 1.0,
574 "_source" : {
575 "Date" : "2020-02-17",
576 "Open" : 6079.53,
577 "High" : 6088.6,
578 "Low" : 60

INDEXATION : ELASTICSEARCH-KIBANA - COVID 19

kibana

Discover

Visualize

Dashboard

Timelion

Dev Tools

Management

Dev Tools

Console

1781 GET _cat/indices

1782

1783 GET _cat/indices?h=h,s,i,id,p,r,dc,dd,ss,creation.date.string

1784

1785 GET cac40_2020/_search

1786 {

1787 "query": {

1788 "match_all": {}

1789 }

1790 }

1791

1792 GET french_economy_indicators/_search

1793 {

1794 "query": {

1795 "match_all": {}

1796 }

1797 }

1798

1799 GET donnees-hospitalieres-covid19/_search

1800 {

1801 "query": {

1802 "match_all": {}

1803 }

1804 }

1805

1806 GET donnees-hospitalieres-covid19-nouveaux/_search

1807 {

1- {

2 "took" : 6,

3 "timed_out" : false,

4 "_shards" : {

5 "total" : 5,

6 "successful" : 5,

7 "skipped" : 0,

8 "failed" : 0

9 },

10 "hits" : {

11 "total" : 18120,

12 "max_score" : 1.0,

13 "hits" : [

14 {

15 "_index" : "donnees-hospitalieres-covid19",

16 "_type" : "_doc",

17 "_id" : "thMgLXIBPUCMKmXTvLlu",

18 "_score" : 1.0,

19 "_source" : {

20 "dep" : 1,

21 "sexe" : 2,

22 "jour" : "2020-03-18",

23 "hosp" : 1,

24 "rea" : 0,

25 "rad" : 0,

26 "dc" : 0

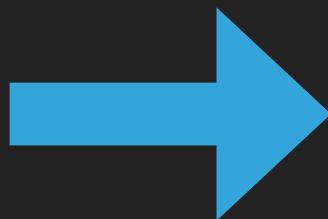
27 }

28 },

ANALYSE DONNÉES COVID 19

	dep	sexe	jour	hosp	rea	rad	dc
0	01	0	2020-03-18	2	0	1	0
1	01	1	2020-03-18	1	0	1	0
2	01	2	2020-03-18	1	0	0	0
3	02	0	2020-03-18	41	10	18	11
4	02	1	2020-03-18	19	4	11	6
...
18481	974	1	2020-05-18	10	4	56	0
18482	974	2	2020-05-18	2	0	59	0
18483	976	0	2020-05-18	55	10	178	18
18484	976	1	2020-05-18	31	4	87	12
18485	976	2	2020-05-18	24	6	88	6

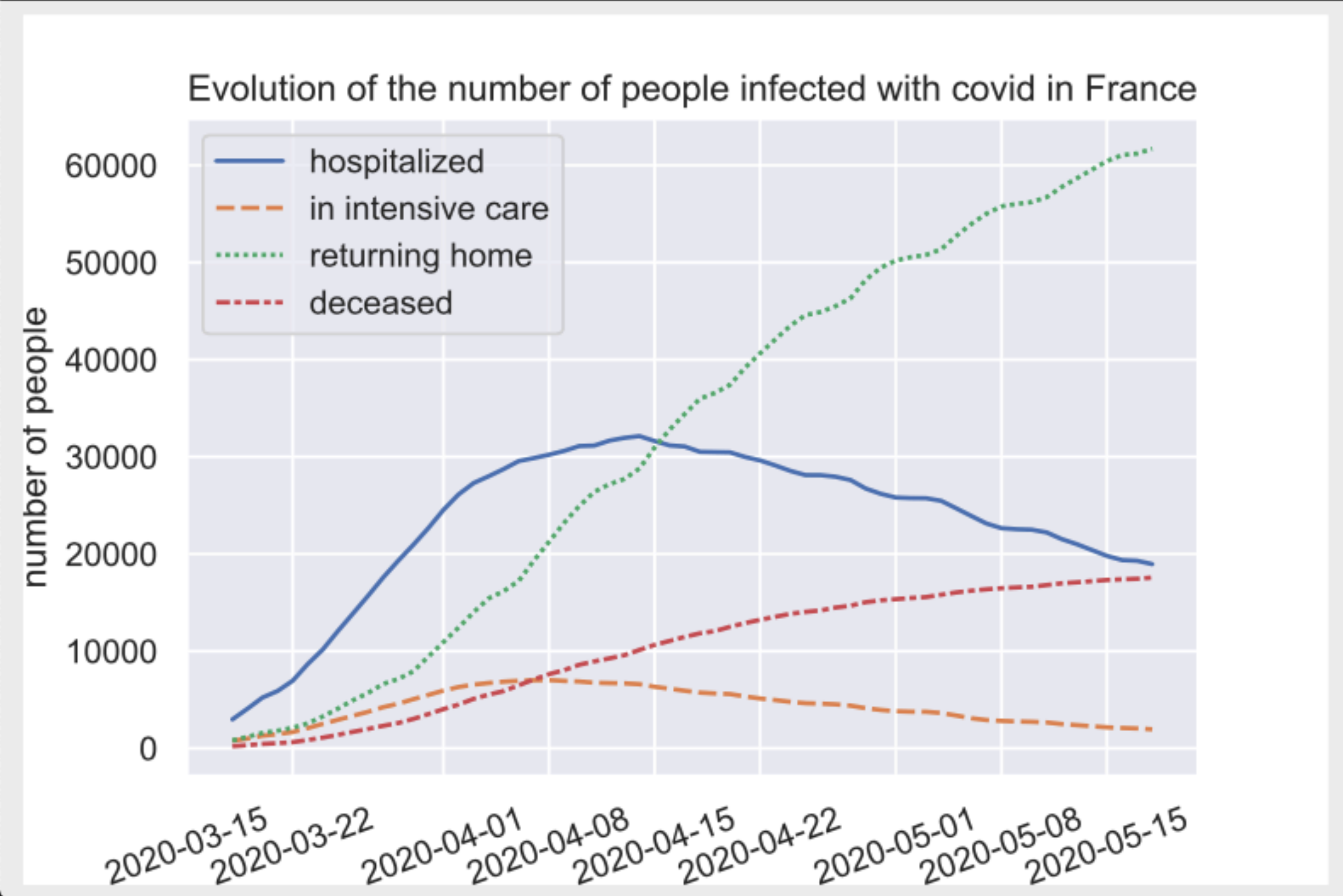
18486 rows x 7 columns



	hospitalized	in intensive care	returning home	deceased
Date				
2020-03-18	2972	771	816	218
2020-03-19	4073	1002	1180	327
2020-03-20	5226	1297	1587	450
2020-03-21	5900	1453	1811	525
2020-03-22	6954	1674	2117	632
...
2020-05-13	21009	2385	58664	17082
2020-05-15	19801	2162	60439	17323
2020-05-16	19372	2091	61057	17393
2020-05-17	19302	2047	61204	17447
2020-05-18	18956	1958	61719	17570

61 rows x 4 columns

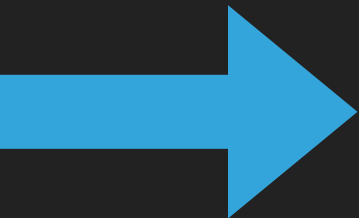
ANALYSE DONNÉES COVID 19



ANALYSE DONNÉES CAC 40

	Open	High	Low	Close	Adj Close	Volume
Date						
2020-02-03	5822.040039	5857.399902	5804.140137	5832.509766	5832.509766	75677800
2020-02-04	5863.359863	5935.049805	5862.819824	5935.049805	5935.049805	85595000
2020-02-05	5919.060059	6004.330078	5912.490234	5985.399902	5985.399902	107839000
2020-02-06	6045.040039	6050.939941	6008.549805	6038.180176	6038.180176	106153100
2020-02-07	6027.540039	6044.970215	5999.950195	6029.750000	6029.750000	82348500
...
2020-05-11	4561.189941	4569.759766	4460.240234	4490.220215	4490.220215	83493200
2020-05-12	4473.430176	4499.390137	4453.649902	4472.500000	4472.500000	99269700
2020-05-13	4420.500000	4422.740234	4336.560059	4344.950195	4344.950195	103285300
2020-05-14	4294.009766	4313.259766	4194.580078	4273.129883	4273.129883	127632000
2020-05-15	4314.970215	4335.770020	4255.189941	4277.629883	4277.629883	111236400

72 rows x 6 columns

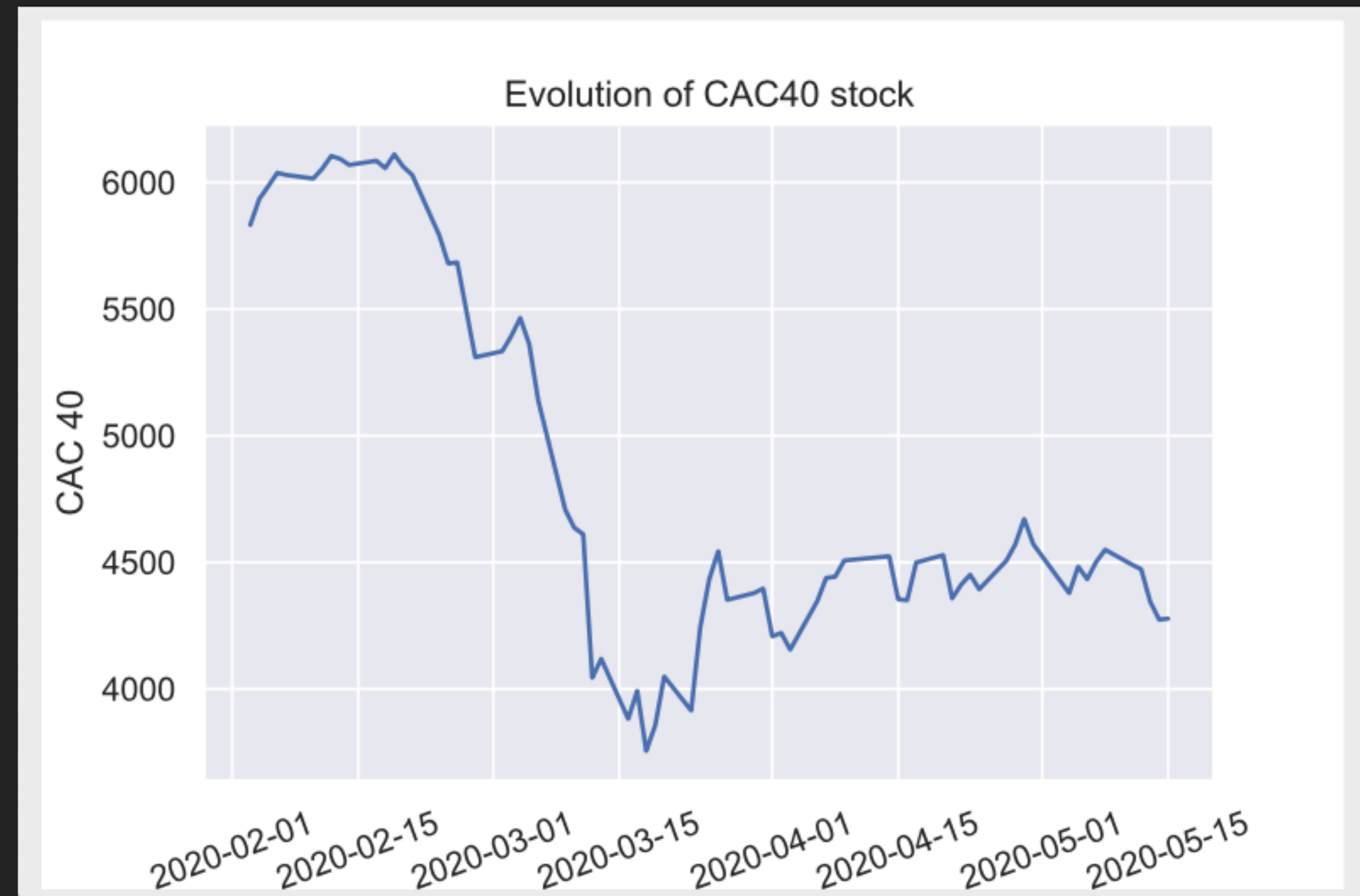


	Adj Close
Date	
2020-02-03	5832.509766
2020-02-04	5935.049805
2020-02-05	5985.399902
2020-02-06	6038.180176
2020-02-07	6029.750000
...	...
2020-05-11	4490.220215
2020-05-12	4472.500000
2020-05-13	4344.950195
2020-05-14	4273.129883
2020-05-15	4277.629883

72 rows x 1 columns

ANALYSE DONNÉES CAC 40

- ▶ On note une baisse du CAC40 à partir du 25-02-2020
- ▶ 25-02-2020: 1er mort du Covid en France et début de la pandémie
- ▶ Cours le plus bas du CAC40 le 18-03-2020
- ▶ 16-03-2020: annonce du début du confinement en France



ANALYSE DONNÉES : CAC 40 + COVID

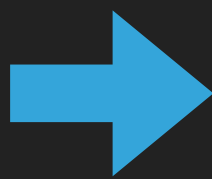
Adj Close	
Date	
2020-02-03	5832.509766
2020-02-04	5935.049805
2020-02-05	5985.399902
2020-02-06	6038.180176
2020-02-07	6029.750000
...	...
2020-05-11	4490.220215
2020-05-12	4472.500000
2020-05-13	4344.950195
2020-05-14	4273.129883
2020-05-15	4277.629883

72 rows x 1 columns

+

hospitalized in intensive care returning home deceased				
Date				
2020-03-18	2972	771	816	218
2020-03-19	4073	1002	1180	327
2020-03-20	5226	1297	1587	450
2020-03-21	5900	1453	1811	525
2020-03-22	6954	1674	2117	632
...
2020-05-13	21009	2385	58664	17082
2020-05-15	19801	2162	60439	17323
2020-05-16	19372	2091	61057	17393
2020-05-17	19302	2047	61204	17447
2020-05-18	18956	1958	61719	17570

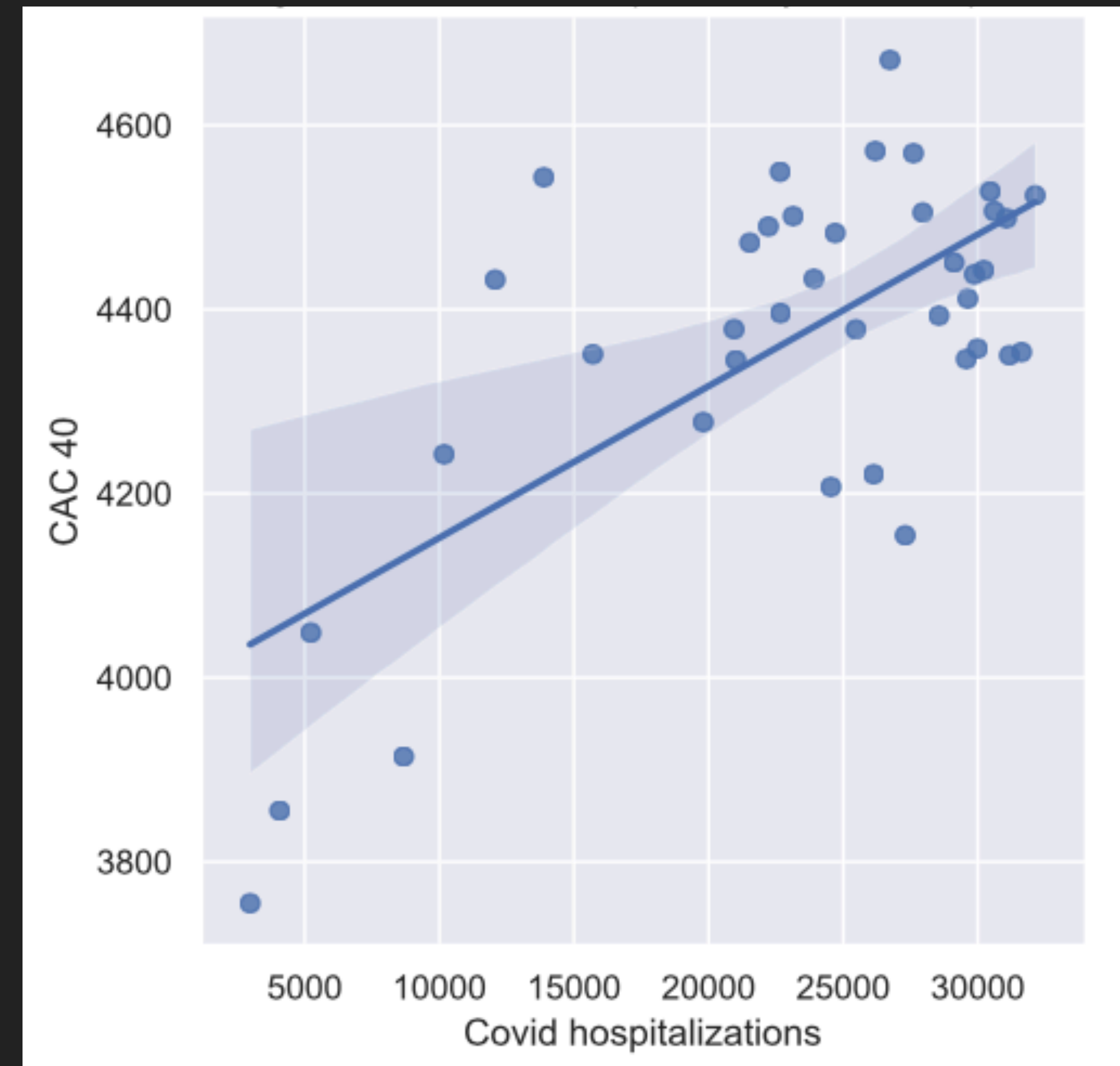
61 rows x 4 columns



	Date	Adj Close	hospitalized	in intensive care	returning home	deceased
0	2020-03-18	3754.840088	2972	771	816	218
1	2020-03-19	3855.500000	4073	1002	1180	327
2	2020-03-20	4048.800049	5226	1297	1587	450
3	2020-03-23	3914.310059	8673	2080	2567	860
4	2020-03-24	4242.700195	10176	2516	3281	1100
5	2020-03-25	4432.299805	12072	2935	4085	1388
6	2020-03-26	4543.580078	13879	3351	4947	1696
7	2020-03-27	4351.490234	15701	3758	5698	1995
8	2020-03-30	4378.509766	20946	5056	7923	3024
9	2020-03-31	4396.120117	22672	5496	9443	3523
10	2020-04-01	4207.240234	24543	5940	10934	4032
11	2020-04-02	4220.959961	26131	6305	12427	4503
39 rows × 6 columns						

REGRESSION LINEAIRE : CAC 40 COVID

- ▶ Quel est la relation entre CAC40 et Covid?
- ▶ Est-ce que la variable CAC40 peut être prédite par la variable Covid ?
- ▶ $Y = a \cdot X + b$
- ▶ $CAC40 = a \cdot Covid + b$
- ▶ On cherche a et b pour minimiser l'écart entre les points et la droite de régression



REGRESSION LINEAIRE : CAC 40 COVID

- ▶ Analyse de régression linéaire réalisée avec 3 bibliothèques différentes
 - ▶ Scikitlearn - python
 - ▶ 1 feature ('hospitalized') ; RMSE: 160.9 ; $r^2 = -1.35$
 - ▶ Pyspark - Spark python
 - ▶ 1 feature ('hospitalized') ; RMSE: 118.7 ; $r^2 = 0.28$
 - ▶ SparkML - Spark scala
 - ▶ 1 feature ('hospitalized') ; RMSE: 174.0 ; $r^2 = -2.86$
 - ▶ 4 features ("hospitalized", "in intensive care", "returning home", "deceased") ; RMSE: 162.32 ; $r^2 = -2.36$

REGRESSION LINEAIRE : CAC 40 COVID

- ▶ Conclusion:
- ▶ Le modèle n'est pas assez bon ($r^2 \sim 1$; $RMSE < 1$)
- ▶ La nombre d'hospitalisation Covid n'est pas un facteur suffisant pour expliquer la valeur du CAC 40
- ▶ Les 4 facteurs (hospitalisation, réanimation, décès, retour domicile) n'expliquent pas mieux la valeur du CAC40

DISCUSSION

- ▶ On remarque que seules les annonces fortes ont un impact sur le cours de la bourse.
- ▶ Par exemple :
 - ▶ L'annonce du premier mort covid en France et des premiers cas infectés
 - ▶ Répercussion : chute du cours du CAC40
 - ▶ Cours le plus bas du CAC40 coïncide avec l'annonce du confinement
- ▶ De même :
 - ▶ L'annonce de résultats encourageants aux tests du vaccin Covid de la société Moderna
 - ▶ Répercussion: augmentation de 25% de l'action Moderna
 - ▶ Répercussion2: augmentation du Dow Jones de 3.5% (hausse la plus importante en une journée depuis début avril)
 - ▶ Source: <https://www.lefigaro.fr/conjoncture/les-espoirs-d-un-vaccin-contre-le-covid-19-font-s-envoler-les-bourses-20200518>

CONCLUSION

- ▶ Extraction de données Covid et CAC40
- ▶ Ingestion dans un datalake Hadoop
- ▶ Indexation dans le moteur de recherche Elasticsearch
- ▶ Exploration, analyse et visualisation de données avec Python
- ▶ Calcul d'un modèle de régression linéaire avec Spark pour prédire le cours du CAC 40 avec le nombre d'hospitalisation Covid
- ▶ Disponibilité du code source : <https://github.com/triquelme/covid-econometrics>