**Tópicos   
Ny Post - Ny Times**

Tópicos Ny Times:

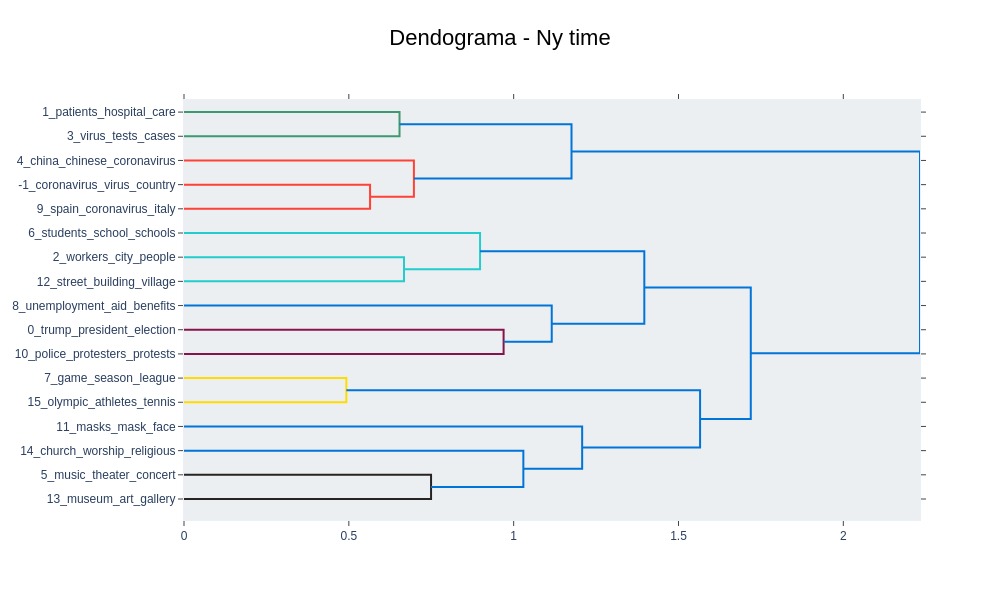
Cuadro de tópicos para Ny Times:

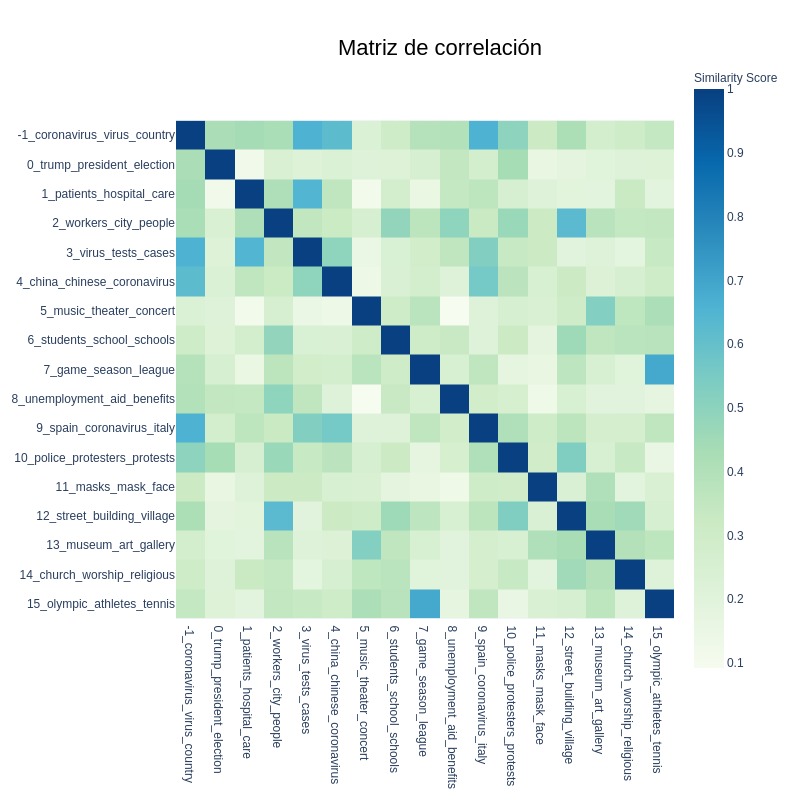
|  | **Topic** | **Count** | **Name** |
| --- | --- | --- | --- |
| 0 | -1 | 349 | -1\_coronavirus\_virus\_country\_pandemic |
| **1** | 0 | 121 | 0\_trump\_president\_election\_campaign |
| **2** | 1 | 82 | 1\_patients\_hospital\_care\_coronavirus |
| **3** | 2 | 69 | 2\_workers\_city\_people\_employees |
| **4** | 3 | 62 | 3\_virus\_tests\_cases\_coronavirus |
| **5** | 4 | 44 | 4\_china\_chinese\_coronavirus\_outbreak |
| **6** | 5 | 42 | 5\_music\_theater\_concert\_show |
| **7** | 6 | 35 | 6\_students\_school\_schools\_children |
| **8** | 7 | 29 | 7\_game\_season\_league\_baseball |
| **9** | 8 | 28 | 8\_unemployment\_aid\_benefits\_money |
| **10** | 9 | 24 | 9\_spain\_coronavirus\_italy\_madrid |
| **11** | 10 | 23 | 10\_police\_protesters\_protests\_city |
| **12** | 11 | 16 | 11\_masks\_mask\_face\_wear |
| **13** | 12 | 14 | 12\_street\_building\_village\_wales |
| **14** | 13 | 14 | 13\_museum\_art\_gallery\_auction |
| **15** | 14 | 13 | 14\_church\_worship\_religious\_rabbi |
| **16** | 15 | 13 | 15\_olympic\_athletes\_tennis\_players |

En el pasado cuadro pudimos ver la totalidad de tópicos que el algoritmo ha encontrado para las noticias seleccionadas del Ny Times. La columna “Topic” los enumera. Un detalle es que el tópico -1 hace referencia a aquellos términos considerados “outliers”, es decir, los que no se han podido englobar en los tópicos generados. Si bien es cierto que los mismos términos puede aparece posteriormente en los tópicos posteriores, esto puede responder a una lógica de agrupamiento particular de cada tópico, que no se da en el -1. La columna ‘Count’ hace referencia a la cantidad de veces que aparecería cada tópicos en los textos. La tabla los ordena de mayor a menor en número de apariciones. Por último, en la columna ‘Name’ tenemos los tópicos con los términos más representativos de cada uno.

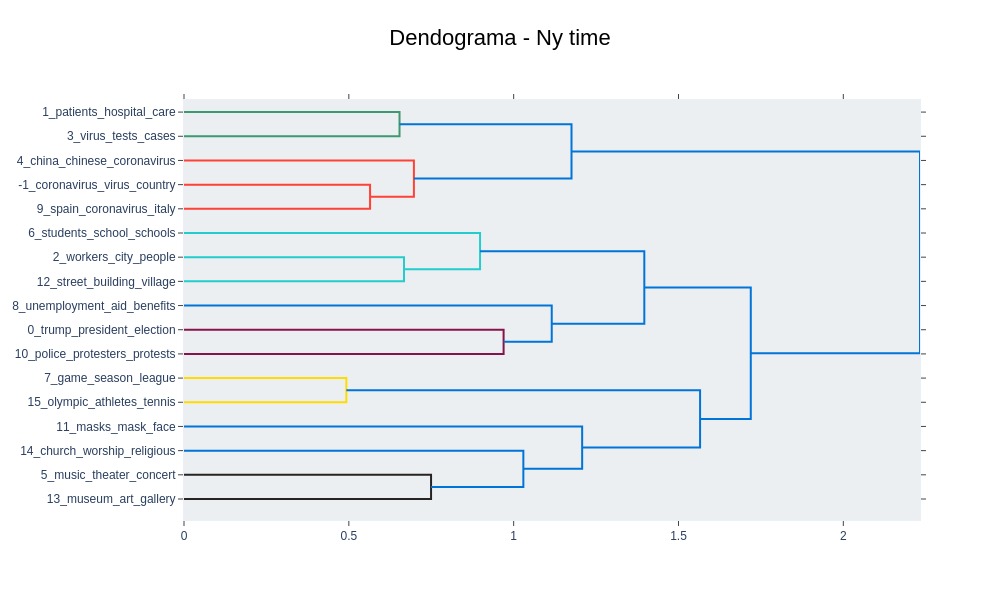
**Representación gráfica de términos más frecuentes en los Tópicos.**

Los siguientes gráficos son la representación de los principales términos por tópico y qué tan significativos son en términos de probabilidad.



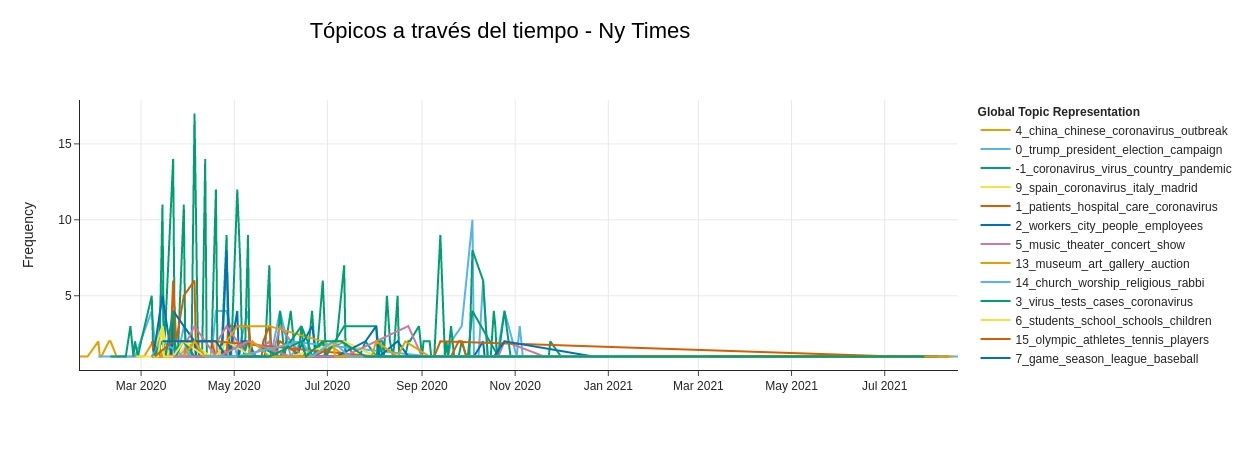


La matriz de correlación expresa qué tan relacionados están los tópicos entre sí. La escala de la derecha, cuyo rango va de 0 a 1, nos indica que cuanto más se correlacionen dos tópicos(el valor se acerque más a 1), más azul será el cuadrado en la matriz principal. En el eje x y el eje y tenemos la totalidad de los tópicos. En la diagonal, cada tópico se enfrenta con si mismo, por eso el color azul más alto.



Un dendrograma es un tipo de representación gráfica o diagrama de datos en forma de árbol) que organiza los datos en subcategorías que se van dividiendo en otros hasta llegar al nivel de detalle deseado (asemejándose a las ramas de un árbol que se van dividiendo en otras sucesivamente). Este tipo de representación permite apreciar claramente las relaciones de agrupación entre los datos e incluso entre grupos de ellos aunque no las relaciones de similitud o cercanía entre categorías. Observando las sucesivas subdivisiones podemos hacernos una idea sobre los criterios de agrupación de los mismos, la distancia entre los datos según las relaciones establecidas, etc

Por ejemplo, si miramos el gráfico anterior, podremos ver cómo el algoritmo ha ido organizando los tópicos de una forma que parece más que coherente. En la parte superior vemos que ha unido el tópico 1 (1\_patients\_hospital\_care\_coronavirus) y el 3 (3\_virus\_tests\_cases\_coronavirus) en la útlima partición, pero si nos vamos más arriba lo une con el 4(4\_china\_chinese\_coronavirus\_outbreak), el -1(1\_coronavirus\_virus\_country\_pandemic) y el 9(9\_spain\_coronavirus\_italy\_madrid). El ejemplo grafica que ests agrupaciones tienen bastante sentido a priori.



El gráfico de tópicos a través del tiempo nos permite observar qué tanto han aparecido los tópicos en los textos, a lo largo de la franja temporal que abarcan los articulos. Tomando este material podremos observar qué tanto enfocaron los diarios sus escritos hacia tal o cuál tópico, dependiendo de la fecha que tomemos.

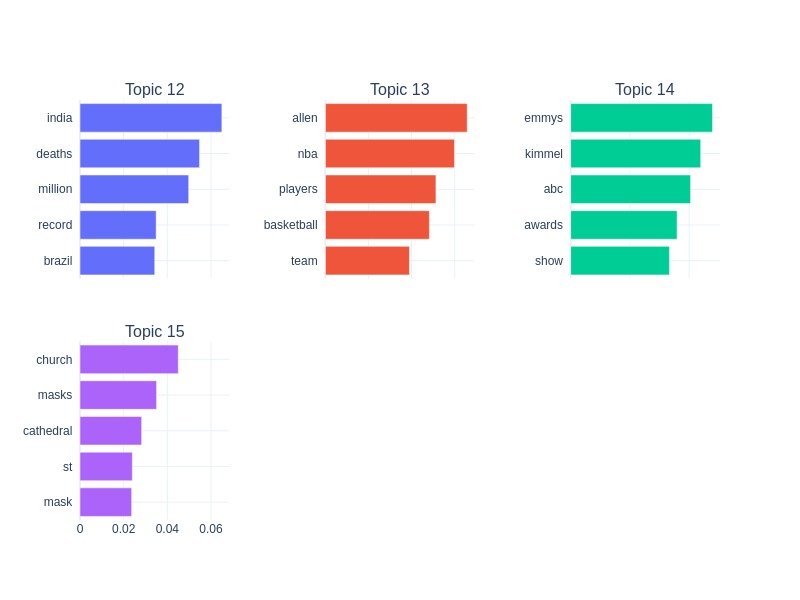
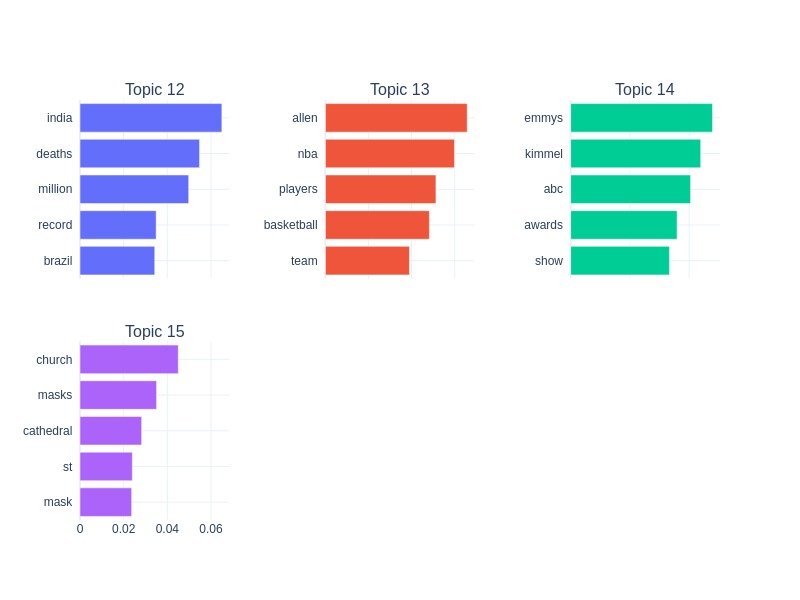
Ny post

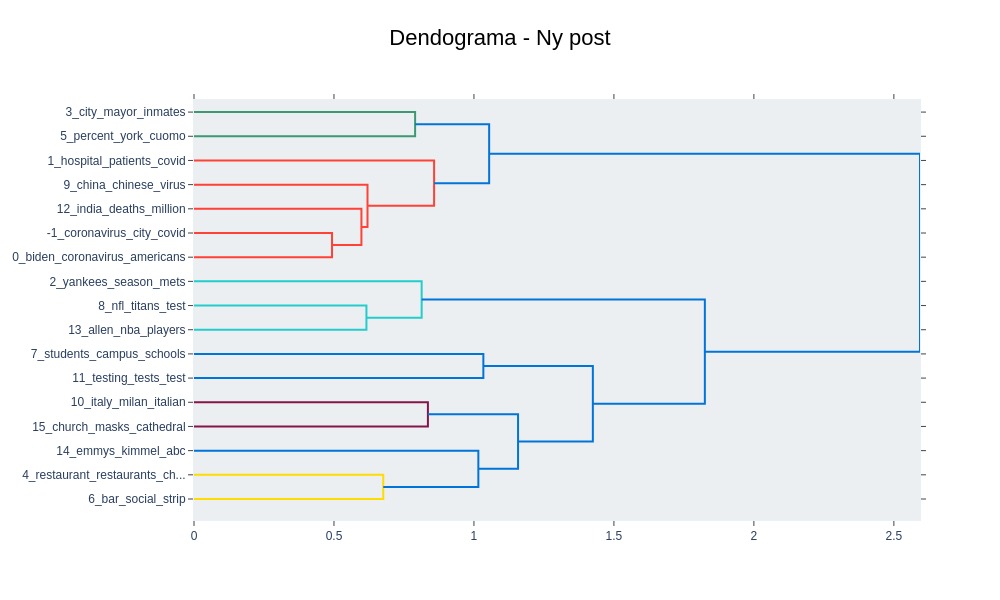
Cuadro de tópicos para Ny Post::

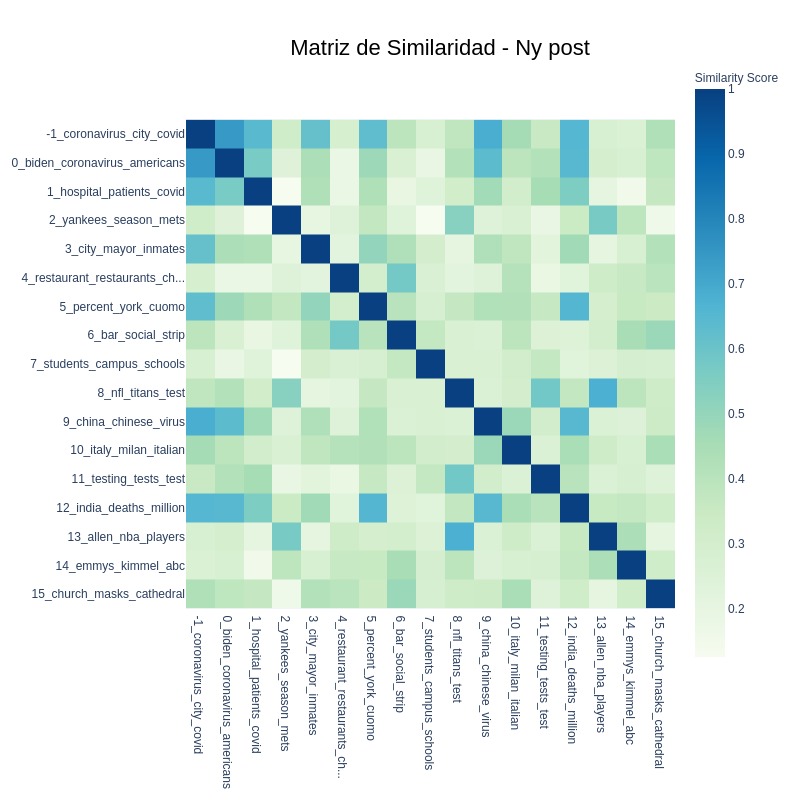
|  | **Topic** | **Count** | **Name** |
| --- | --- | --- | --- |
| **0** | -1 | 254 | -1\_coronavirus\_city\_covid\_virus |
| **1** | 0 | 99 | 0\_biden\_coronavirus\_americans\_pandemic |
| **2** | 1 | 59 | 1\_hospital\_patients\_covid\_virus |
| **3** | 2 | 26 | 2\_yankees\_season\_mets\_baseball |
| **4** | 3 | 25 | 3\_city\_mayor\_inmates\_jails |
| **5** | 4 | 24 | 4\_restaurant\_restaurants\_chefs\_customers |
| **6** | 5 | 24 | 5\_percent\_york\_cuomo\_deaths |
| **7** | 6 | 21 | 6\_bar\_social\_strip\_brooklyn |
| **8** | 7 | 20 | 7\_students\_campus\_schools\_university |
| **9** | 8 | 15 | 8\_nfl\_titans\_test\_players |
| **10** | 9 | 15 | 9\_china\_chinese\_virus\_korea |
| **11** | 10 | 15 | 10\_italy\_milan\_italian\_spain |
| **12** | 11 | 12 | 11\_testing\_tests\_test\_seattle |
| **13** | 12 | 11 | 12\_india\_deaths\_million\_record |
| **14** | 13 | 11 | 13\_allen\_nba\_players\_basketball |
| **15** | 14 | 10 | 14\_emmys\_kimmel\_abc\_awards |
| **16** | 15 | 10 | 15\_church\_masks\_cathedral\_st |

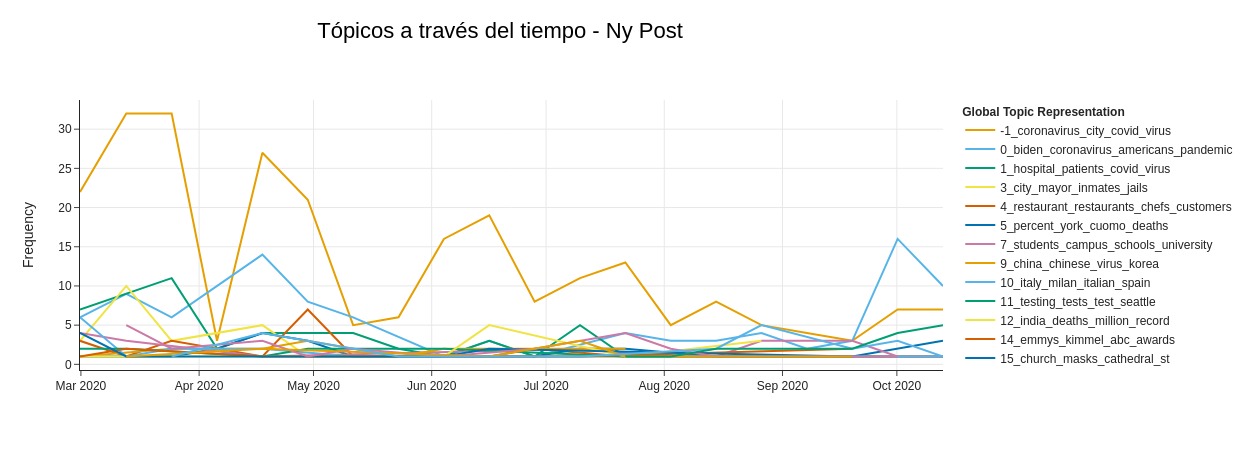
**Representación gráfica de términos más frecuentes en los Tópicos.**











Tópicos Ambos diarios:

Cuadro de tópicos para ambos diarios:

| **Topic** | **Count** | **Name** |  |
| --- | --- | --- | --- |
| **0** | -1 | 548 | -1\_coronavirus\_virus\_pandemic\_state |
| **1** | 0 | 150 | 0\_biden\_election\_campaign\_voters |
| **2** | 1 | 104 | 1\_city\_york\_cuomo\_blasio |
| **3** | 2 | 77 | 2\_music\_museum\_theater\_show |
| **4** | 3 | 73 | 3\_china\_chinese\_outbreak\_virus |
| **5** | 4 | 63 | 4\_students\_school\_schools\_campus |
| **6** | 5 | 52 | 5\_patients\_hospital\_care\_ventilators |
| **7** | 6 | 49 | 6\_spain\_italy\_coronavirus\_europe |
| **8** | 7 | 46 | 7\_workers\_employees\_work\_unemployment |
| **9** | 8 | 42 | 8\_restaurant\_restaurants\_meat\_chefs |
| **10** | 9 | 37 | 9\_cases\_virus\_states\_state |
| **11** | 10 | 35 | 10\_baseball\_mets\_league\_yankees |
| **12** | 11 | 34 | 11\_died\_father\_family\_mr |
| **13** | 12 | 32 | 12\_touchdown\_nfl\_quarterback\_giants |
| **14** | 13 | 29 | 13\_police\_protesters\_protests\_protest |
| **15** | 14 | 28 | 14\_street\_brooklyn\_borough\_neighbors |
| **16** | 15 | 26 | 15\_masks\_mask\_face\_wear |
| **17** | 16 | 25 | 16\_tournament\_olympic\_tennis\_games |
| **18** | 17 | 24 | 17\_tests\_testing\_antibodies\_antibody |
| **19** | 18 | 22 | 18\_democrats\_mnuchin\_senate\_trump |
| **20** | 19 | 21 | 19\_tested\_pence\_trump\_quarantine |
| **21** | 20 | 20 | 20\_church\_francis\_worship\_religious |
| **22** | 21 | 18 | 21\_countries\_virus\_country\_coronavirus |
| **23** | 22 | 16 | 22\_players\_nets\_basketball\_nba |
| **24** | 23 | 13 | 23\_doctors\_baby\_births\_pregnant |
| **25** | 24 | 12 | 24\_ship\_cruise\_quarantine\_ships |
| **26** | 25 | 12 | 25\_inmates\_prison\_prisons\_prisoners |
| **27** | 26 | 11 | 26\_netanyahu\_lebanon\_lebanese\_government |
| **28** | 27 | 10 | 27\_markets\_stock\_fed\_investors |



