

Data Analytics - Comisión: 32535

Proyecto final

Terrorismo: un análisis global

Autores

Luca Barbini - Lucas López - Leandro Serrano

Institución

CODER HOUSE

ÍNDICE

INTRODUCCIÓN	3
DESCRIPCIÓN Y ALCANCE DE LA TEMÁTICA	3
HIPÓTESIS	3
MARCO METODOLÓGICO	4
DATA SET	4
VARIABLES	5
DIAGRAMA ENTIDAD RELACIÓN	5
DESCRIPCIÓN DE TABLAS	7
RESULTADOS	1
RESULTADOS PRELIMINARES	0
VISUALIZACIÓN DE LOS DATOS	1
CONCLUSIÓN FINAL	0
	1
	0
	1
	7

INTRODUCCIÓN

Este trabajo se propone generar un dashboard a partir de la recopilación de datos sobre ataques terroristas y generar visualizaciones que permitan la comprensión de las principales dimensiones de este temático.

DESCRIPCIÓN Y ALCANCE DE LA TEMÁTICA

El tema principal elegido son los ataques terroristas a nivel internacional desde el año 1970 hasta el año 2017, a partir de un dataset que reúne los atentados y las principales características de los mismos, presentando información relevante como ser: el tipo de atentado, las armas utilizadas, el objetivo principal del atentado, país donde se realizó, entre otras.

De esta forma, se puede observar la evolución espacial y temporal de los ataques terroristas, dimensionando por las principales características de los ataques. A partir de esto, se propone un análisis descriptivo.

HIPÓTESIS

El trabajo se propone como objetivo principal encontrar y visualizar los principales países afectados por ataques terroristas. A su vez, se intenta descubrir cuáles fueron los ataques y armas más utilizados, así como los principales objetivos, permitiendo desagregar estas dimensiones a partir de la evolución temporal, desagregar a nivel regional o nacional y permitiendo interactuar entre las distintas dimensiones consideradas para un análisis más específico. Por lo tanto, se pretende:

- Conocer cuáles fueron los tipos de ataques más utilizados en los atentados.
- Descubrir las regiones más propensas a sufrir atentados terroristas.
- Identificar los objetivos más frecuentes de todos los ataques.
- Averiguar el periodo en el que más casos se hayan registrado.
- Encontrar las organizaciones terroristas que más ataques han realizado.

MARCO METODOLÓGICO

DATASET

Debajo encontrarán el link a los datasets utilizados para acceder a donde se encuentran almacenados, como así también el enlace a su fuente original.

[Kaggle: Fuente original del dataset](#)

Dataset Final

El dataset original seleccionado se extrajo de Kaggle, tal como se detalló en el párrafo anterior. Nos encontramos con un CSV de 181.691 registros y 135 columnas, demostrando ser una matriz con muchas aristas para analizar.

El primer trabajo de limpieza de datos consistió en revisar las 135 variables en Excel y verificar el significado de cada una, pensando ya en un futuro análisis de datos con su respectiva representación

gráfica. Esto nos llevó a realizar un filtro importante de información, ya que mucha data resultaba redundante o poco relevante para contrastar las hipótesis de este trabajo. El dataset pasó a ser de 17 columnas, siendo la cantidad de variables más que suficiente para comenzar a analizarlas una por una.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	eventid	Date	country	country_txt	region	region_txt	success	attacktype	attacktype1_txt	targettype	targettype1_txt	gname	weaptype	weaptype1_txt	nkill	nwound	INT_ANY
2	1	7/2/1	58	Dominican Re	2	Central Amr	1	1	Assassination	14	Private Citizens	MANO-D	13	Unknown	1	0	0
3	2	1/1/1	130	Mexico	1	North Amer	1	6	Hostage Taking (Ki	7	Government (D	23rd of Se	13	Unknown	0	0	1
4	3	1/1/1	160	Philippines	5	Southeast A	1	1	Assassination	10	Journalists & M	Unknown	13	Unknown	1	0	1
5	4	1/1/1	78	Greece	8	Western Eu	1	3	Bombing/Explosic	7	Government (D	Unknown	6	Explosives	Unknown	Unknown	1
6	5	1/1/1	101	Japan	4	East Asia	1	7	Facility/Infrastruct	7	Government (D	Unknown	8	Incendiary	Unknown	Unknown	1
7	6	1/1/1	217	United States	1	North Amer	1	2	Armed Assault	3	Police	Black Natl	5	Firearms	0	0	-9

En ciertas variables, como por ejemplo “nkill” (número de muertos) y “nwound” (número de heridos), nos encontramos con campos vacíos que nos resultaba dificultoso a la hora de analizarlo. La dificultad venía porque no tener datos en ciertos registros, no significa que no haya habido muertes, sino que no se sabe qué cantidad de muertos o heridos hubo en determinados atentados. Para no dejar el campo vacío y que preste a la confusión, decidimos agregar la palabra “Unknown” en cada campo vacío de cada registro, dejando la información clara para el usuario y sencilla para el Data Analyst.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	event	Date	country	country_txt	region	region_txt	success	attacktype	attacktype1_txt	targettype	targettype1_txt	gname	weaptype	weaptype1_txt	nkill	nwound	INT_ANY
5	4	####	78	Greece	8	Western Eu	1	3	Bombing/Explosic	7	Government (C	Unknown	6	Explosives	Unknown	Unknown	1
6	5	####	101	Japan	4	East Asia	1	7	Facility/Infrastruct	7	Government (C	Unknown	8	Incendiary	Unknown	Unknown	1
17	16	####	499	East Germany	9	Eastern Eur	1	3	Bombing/Explosic	2	Government (C	Commune	6	Explosives	Unknown	Unknown	1
36	35	####	499	East Germany	9	Eastern Eur	1	7	Facility/Infrastruct	3	Police	Unknown	8	Incendiary	Unknown	Unknown	1
97	96	####	102	Jordan	10	Middle East	1	2	Armed Assault	18	Tourists	Unknown	5	Firearms	Unknown	Unknown	-9
178	177	####	11	Argentina	3	South Amer	1	2	Armed Assault	4	Military	Argentine	5	Firearms	Unknown	Unknown	0
243	242	####	499	East Germany	9	Eastern Eur	1	7	Facility/Infrastruct	2	Government (C	Unknown	8	Incendiary	Unknown	Unknown	1
279	278	####	199	Switzerland	8	Western Eu	1	3	Bombing/Explosic	6	Airports & Ai	1st of May	6	Explosives	Unknown	Unknown	1
290	289	####	362	West German	8	Western Eu	1	7	Facility/Infrastruct	1	Business	Unknown	8	Incendiary	Unknown	Unknown	1
292	291	####	362	West German	8	Western Eu	1	2	Armed Assault	1	Business	Unknown	5	Firearms	Unknown	Unknown	-9
326	325	####	362	West German	8	Western Eu	1	2	Armed Assault	1	Business	Extraparli	5	Firearms	Unknown	Unknown	1
327	326	####	362	West German	8	Western Eu	1	2	Armed Assault	2	Government (C	Extraparli	5	Firearms	Unknown	Unknown	1

Finalizado el trabajo de limpieza, logramos tener un dataset conservando los 181.691 registros y reduciéndolo a 17 variables, convirtiéndolo en algo claro, prolijo y fácil de analizar.

VARIABLES

A continuación se detallan las variables consideradas para el modelo:

Nombre de variable	Tipo de variable	Descripción
Event_ID	Int	Identificador individual para cada atentado
Gname	Text	Nombre del grupo terrorista que llevo adelante el atentado
Attack Type	Text	Clasificación categórica del tipo de atentado llevado a cabo
AttackType_ID	Int	Identificador numérico para cada tipo de atentado
TargetType	Text	Clasificación categórica según el objetivo del atentado
TargetType_ID	Int	Identificador numérico para cada tipo de objetivo
WeaponType	Text	Clasifica categóricamente el atentado según el tipo de arma utilizada
WeaponType_ID	Int	Identificador numérico para cada tipo de arma

DIAGRAMA ENTIDAD RELACIÓN

Debido a la naturaleza del dataset, siendo una tabla única con todos los datos, decidimos utilizar un diagrama entidad-relación del tipo “Estrella”. Esto significó para nosotros tener que crear más tablas, dividiendo el dataset en tablas de dimensiones para completar el modelo y así establecer las relaciones correctas para continuar con el análisis de datos.

La creación de tablas se realizó en Excel, mediante el uso de Pivot Tables, Filters y Vlookup 's. Se

crearon

6 tablas haciendo un total de 7, siendo “Atentados” la principal, con la variable “Event_ID” como Primary Key. Las variables “_ID” de cada una de las demás tablas satélites se generaron como Primary Key

para

que funcionen como mapeo a la tabla “Atentados” y Power Bi pueda relacionarlas de la mejor manera.

Se creó especialmente la tabla “Date” con la variable “Date_ID” como primary key. Esto significó

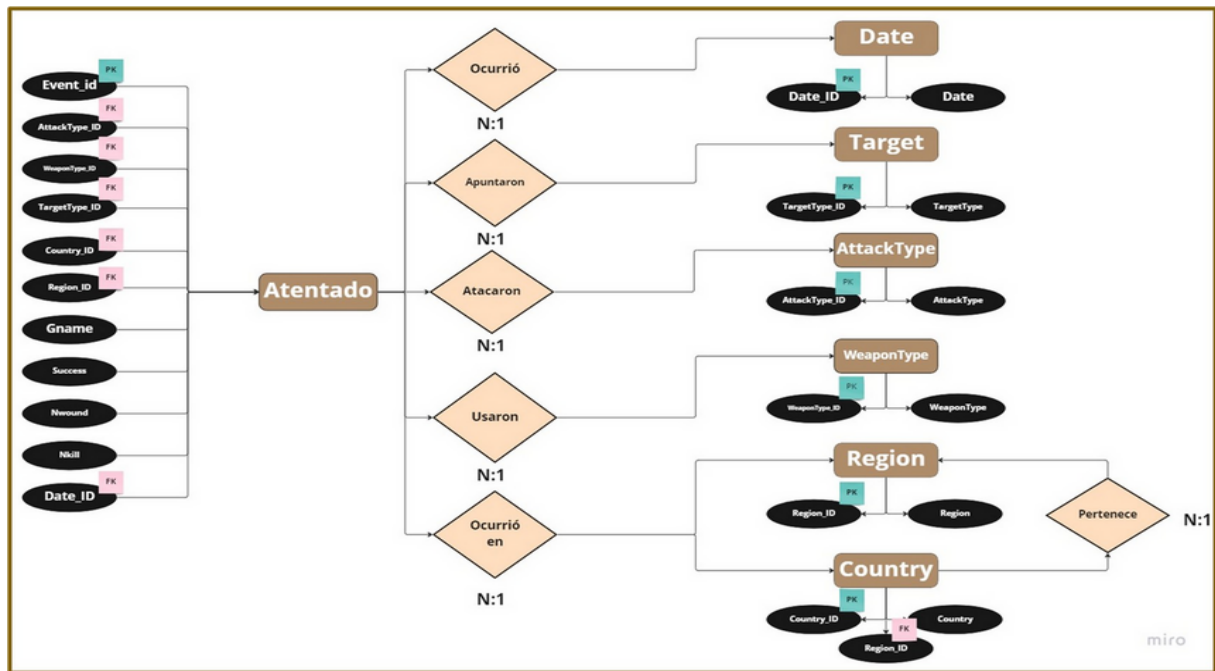
asignar

un número específico a cada fecha de manera individual, buscando que la misma fecha no tenga números de ID diferentes. La tabla “Date” no sólo funcionará como tabla calendario, sino como nexos temporal a todos los análisis gráficos que se realicen en el presente trabajo.

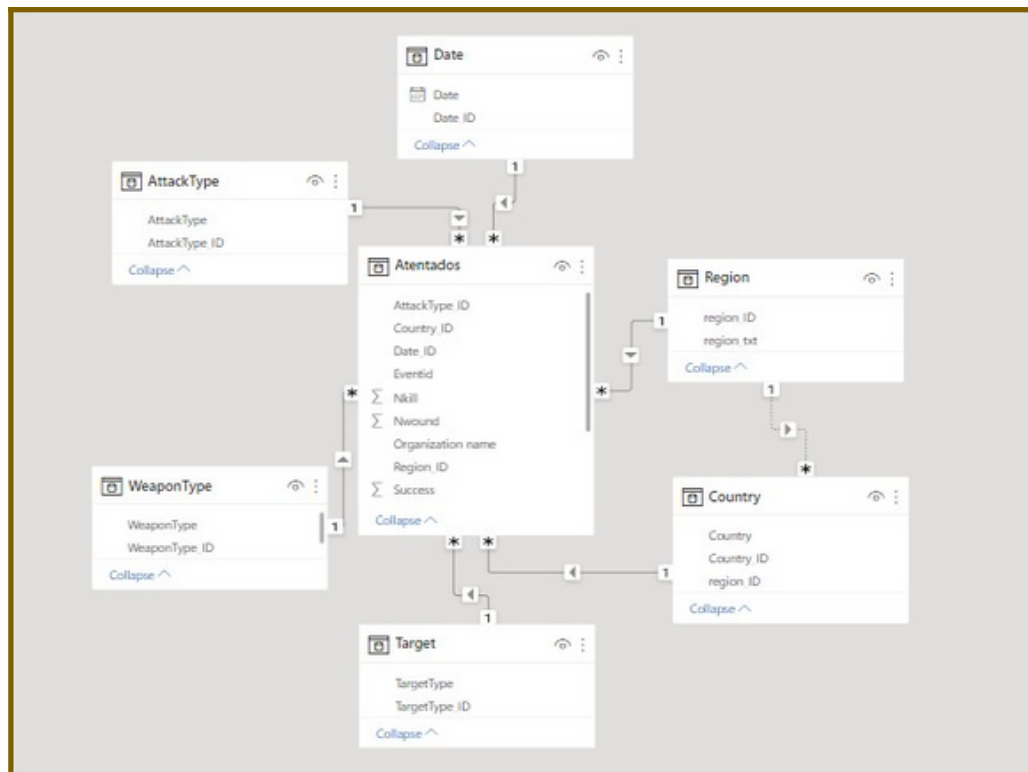
Se dividió la base en tablas referentes a temas específicos y todas con una relación de “N” a “1”,

como se

muestra a continuación.



En Power BI, el diagrama relacional se aplicó de la siguiente manera.



DESCRIPCIÓN DE TABLAS

Date

Tipo de clave	Campo	Tipo de campo
	Date	Datetime
PK	Date_ID	Int

La tabla Date se constituye de la fecha del atentado y un identificador para cada observación.

Atentado

Tipo de clave	Campo	Tipo de campo
PK	Event_ID	Int
FK	AttackType_ID	Int
FK	WeaponType_ID	Int
FK	TargetType_ID	Int
FK	Country_ID	Int
FK	Region_ID	Int
	Gname	Text
	Success	Int
	Nwound	Varchar
	Nkill	Varchar
FK	Date_ID	Int

La entidad Atentado es, como se desprende del modelo relacional, una tabla principal que guarda relación directa con el resto de las tablas del modelo, a partir de las foreign keys que incluye. Entre las variables que se componen se encuentran: un identificador para cada atentado, el identificador del tipo de ataque, de arma y de objetivo, así como el identificador de país, región, el nombre del grupo terrorista al que se le atribuye el atentado, si fue exitoso en su intención, la cantidad de muertes a raíz del atentado así como la cantidad de heridos y un identificador de atentado para las distintas fechas.

Target

Tipo de clave	Campo	Tipo de campo
	TargetType	Text
PK	TargetType_ID	Int

La tabla Target se constituye del tipo de ataque (texto) y un identificador para cada tipo de ataque.

Country

Tipo de clave	Campo	Tipo de campo
	Country	Text
PK	Country_ID	int
FK	Region_ID	int

La tabla Country se constituye del país donde sucedió el atentado (texto), así como el identificador de dicho país y de la región a la que pertenece.

Region

Tipo de clave	Campo	Tipo de campo
	Region	Text
PK	Region_ID	Int

La tabla Región se constituye de la región donde aconteció el ataque terrorista (texto), así como el identificador de dicha región.

CÁLCULOS

Medidas calculadas creadas

Top weapon Type

- Se utilizó esta medida con el fin de identificar el tipo de arma más empleada en los ataques terroristas, la información fue utilizada para visualizar el tipo de arma utilizada por los terroristas en los distintos países del mundo. La medida está compuesta de la siguiente manera:

Top Weapon Type = CALCULATE(
FIRSTNONBLANK(WeaponType[WeaponType], 1),
TOPN(1, all (WeaponType[WeaponType]), [Count eventID], desc))

Average success

- Se utilizó esta medida con el fin de identificar él tasa de éxito en los distintos atentados terroristas. La medida está compuesta de la siguiente manera:

Average Success = DIVIDE(SUM(Atentados[Success]),
COUNTA(Atentados[Eventid]))

Count EventID

- Se utilizó esta medida con el fin de identificar el total de los ataques terroristas. La medida está compuesta de la siguiente manera:

Count eventID = COUNT(Atentados[Eventid])

Count EventID YoY

- Se utilizó esta medida con el fin de hallar el porcentaje de ataques terroristas año tras año. De esta forma se puede interpretar la evolución de los atentados a lo largo del tiempo. La medida (En este caso fue empleada una medida rápida) está compuesta de la siguiente manera:

Count of Eventid YoY% =

IF(

ISFILTERED('Date'[Date]),

ERROR("Time intelligence quick measures can only be grouped or filtered by the Power BI-
provided date hierarchy or primary date column."),

VAR __PREV_YEAR =

CALCULATE(

COUNTA('Atentados'[Eventid]),

DATEADD('Date'[Date].[Date], -1, YEAR)

)

RETURN

DIVIDE(COUNTA('Atentados'[Eventid]) - __PREV_YEAR, __PREV_YEAR)

)

RESULTADOS

RESULTADOS PRELIMINARES

A continuación se detallan los principales resultados derivados de las distintas visualizaciones propuestas. En primer lugar, se desarrolla una pestaña de resumen con los principales datos de interés, posteriormente una desagregación geográfica con el fin de desagregar la información a nivel regional y nacional. Por último, se propone un mayor nivel de detalle, considerando la interacción entre las principales variables consideradas, como ser el tipo de ataque, el arma principal utilizada, el objetivo del atentado y la organización a la que se adjudica el mismo.

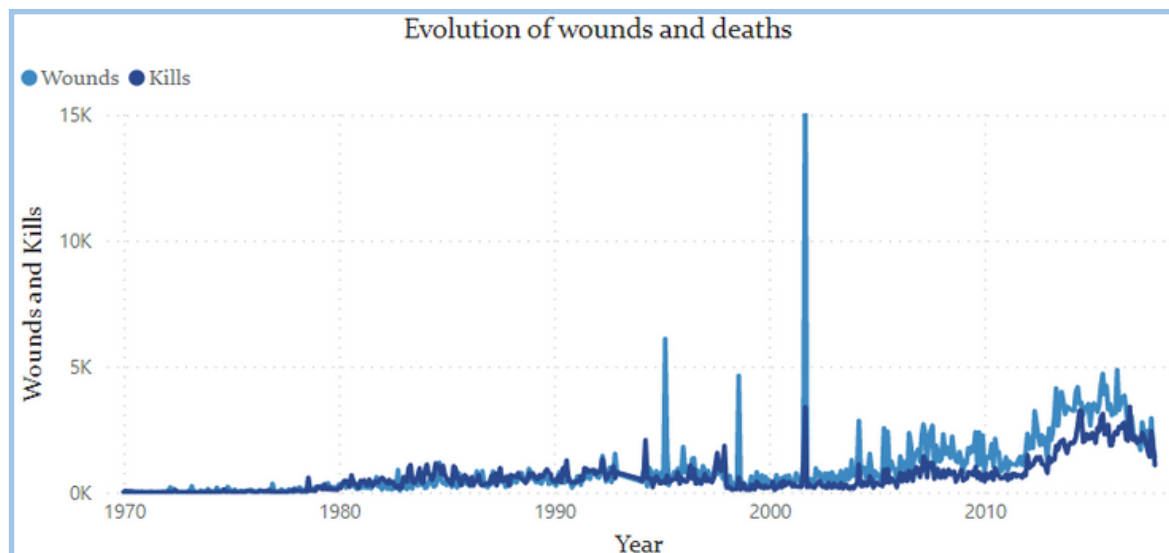
VISUALIZACIÓN DE LOS DATOS

SUMMARY

En este apartado se procura desentrañar las principales características de los atentados terroristas para todo el período, pero también permitiendo desagregar la información a partir de períodos y regiones.

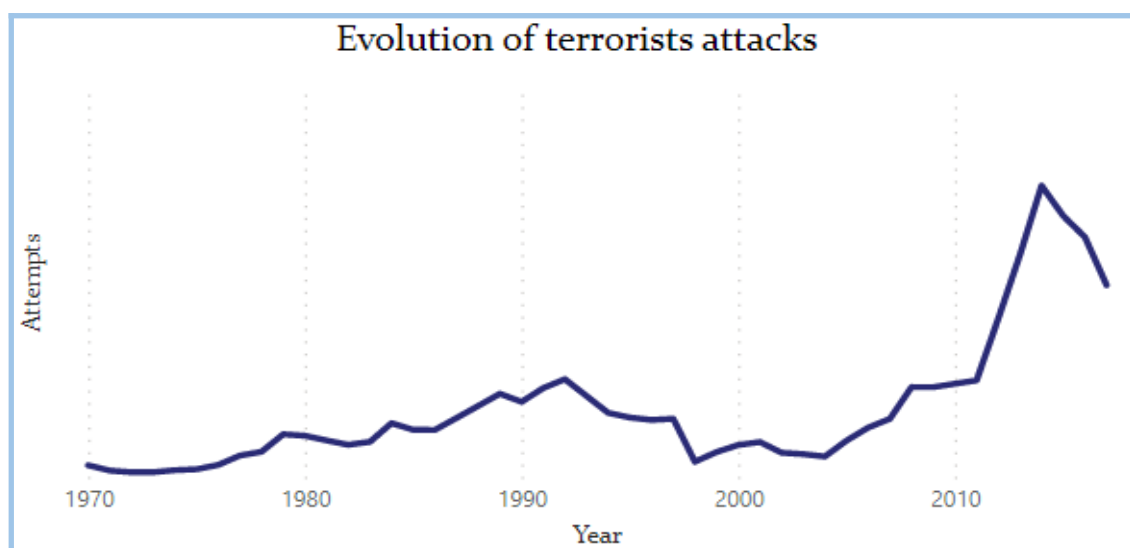
Se observa que el máximo de heridos y muertos se da en el año 2001, alcanzando un total de 6.890 muertes y más de 22.700 heridos para un total de 1.797 ataques terroristas, explicado principalmente por el atentado a WTC con cifras récord en cantidad de víctimas.

Se observan otros años con cifras muy superiores al promedio de víctimas, como ser el segundo trimestre de 1995, donde existió un ataque terrorista en Oklahoma, con un total de 168 muertos, y un ataque gas sarín en Japón, con un saldo de más de 6.000 heridos.

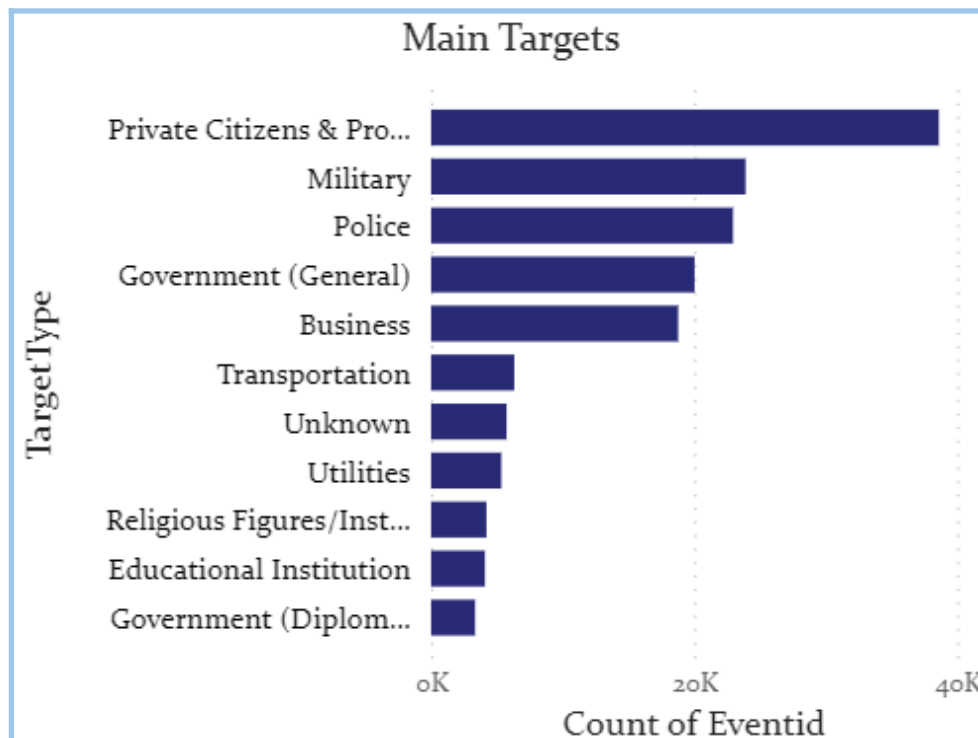


Asimismo, se observa un crecimiento importante tanto en heridos y muertes en promedios por año, particularmente a partir del año 2000, acompañado por un incremento importante en la cantidad de ataques terroristas. A su vez, se observa otro salto hacia el año 2010, superando el incremento para el principio de siglo y, por ende, con el promedio registrado durante el siglo pasado. Se desprende de esto que las formas de ataques terroristas en el último tiempo han mutado hacia más ataques a menor escala.

Si consideramos el panorama general, entre 1970 y 2017 el total de muertes traspasa los 351 mil vidas, mientras que el número de heridos asciende a casi 522 mil.



Con respecto a las armas utilizadas en los ataques terroristas, se observa que las más utilizadas son explosivos (más de 86 mil) y armas de fuego (más de 51 mil) permaneciendo inalterable su posición a lo largo del tiempo. Se observa una reducción en el tiempo del porcentaje de desconocimiento del arma empleada, lo cual puede ser explicado por una mejor recopilación de información o un mejor funcionamiento de las fuerzas de seguridad.

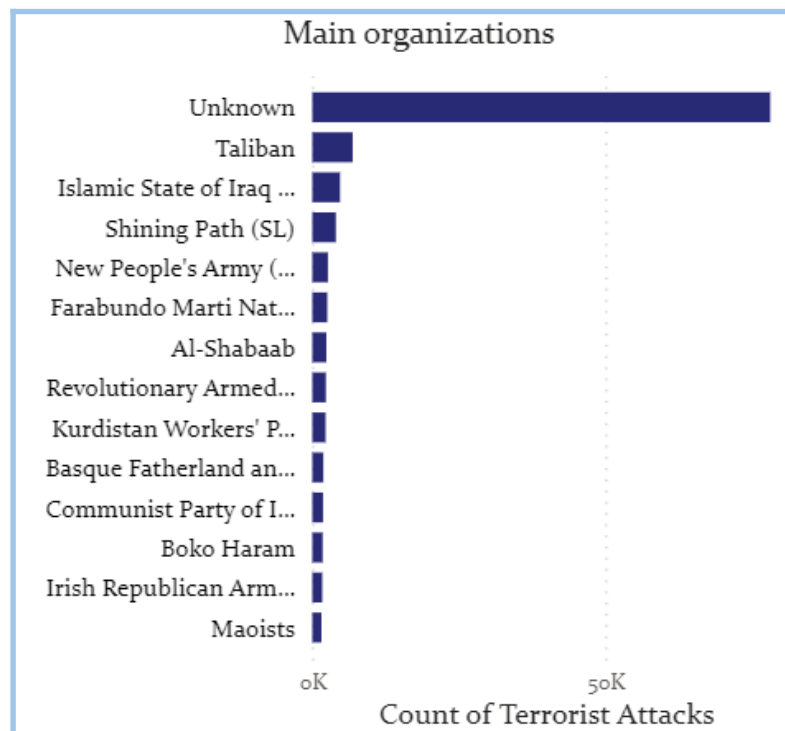


Si analizamos los objetivos principales del atentado, nos encontramos que los principales damnificados son los ciudadanos y la propiedad privada de estos, con casi 40.000 atentados dejando un saldo de más 117 mil fallecidos y 178 mil heridos. En segundo lugar, los militares son el segundo objetivo más atentado con una fuerte concentración en la década del 80 y a partir del año 2010, algo similar a lo que sucede con la policía.

Otro dato importante es que, si bien los explosivos continúan siendo el arma de mayor frecuencia, la proporción de armas de fuego para los atentados cuyo objetivo son las fuerzas de seguridad pública asciende considerablemente.

Otro hecho a señalar es el incremento sustantivo de ataques religiosos en el último siglo, que con menos de 3.000 ataques ha abatido a más de 9.000 personas y ha dejado más de 20.000 heridos.

Posteriormente, se observan ataques más tácticos, con un número inferior de violencia y víctimas, como ser transporte y facilidades, entre otros; mientras que existen otros que atacan a grupos específicos como pueden ser instituciones educativas, periodistas, instituciones religiosas, etc.



A la hora de observar las organizaciones a las que se adjudican los atentados terroristas, se denota que a una buena proporción de los atentados se le desconoce el autor. Si bien esa cifra ha caído con los años y existe cierta desinformación devengada de que las adjudicaciones pueden ser inciertas, aún es difícil encontrar a los artífices de los atentados.

A su vez, las motivaciones y estructuras de las organizaciones terroristas son muy heterogéneas entre sí, su análisis trasciende el objetivo de este trabajo.

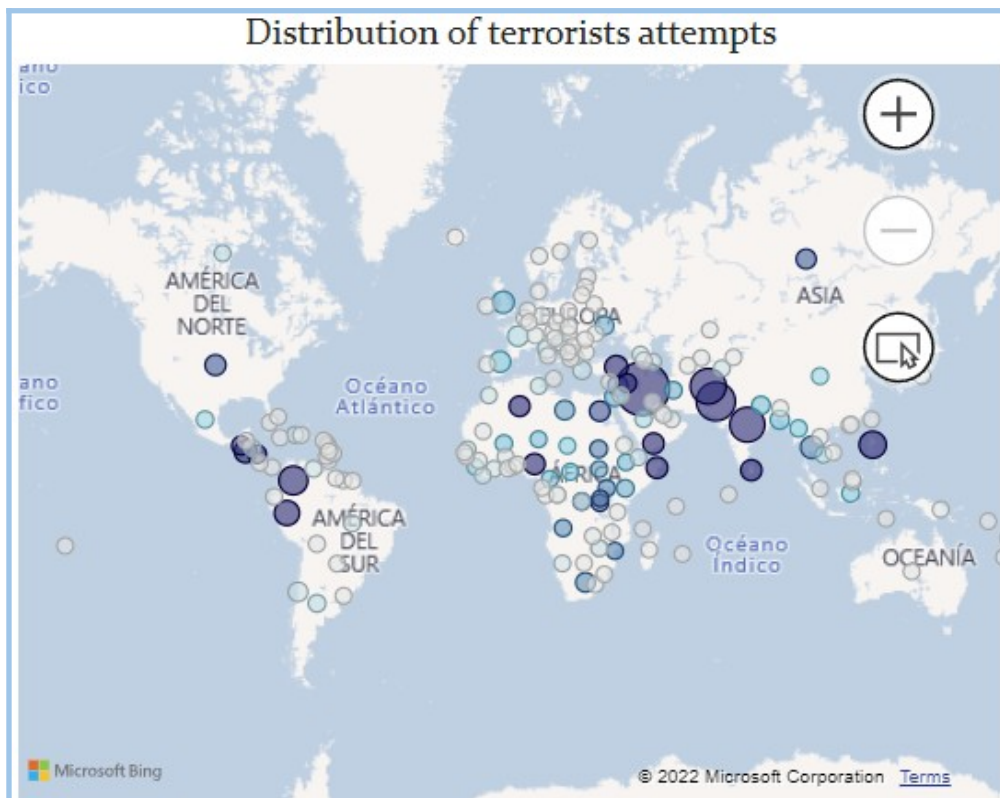
Dicho esto, las principales organizaciones que han desarrollado ataques terroristas en el período considerado son: los talibanes (26,59 mil muertos y 27,87 heridos) con un incremento importante a partir del año 2005, potenciado hacia el año 2011 llegando a realizar 1114 atentados solamente en el año 2015, completando un total de 6.884 atentados desde su inicio; el estado islámico (27,60 mil muertos y 30,62 heridos) que comienza su actividad en el año 2013 y acumula hasta 2017 un total de 4.723 atentados, alcanzando un promedio cercano a mil atentados al año en el período considerado; en tercer lugar aparece Sendero Luminoso, una organización comunista armada de Perú (11,05 muertes y 3,01 heridos) cuyo período de actividad fue esencialmente en los años 80 y alcanzó los 4.010 atentados.

Por último, mencionar que la tasa de éxito de los atentados en todo el período es de 88,23%. Ahora bien, si comenzamos a acercarnos en el tiempo la tasa comienza a decrecer, llegando a un 84,98% si analizamos desde el 2010 al 2017, aunque que la cantidad de ataques parecen ser cada vez mayores y a menor escala, dificultando la creación de políticas y estrategias que puedan prevenir estos sucesos.

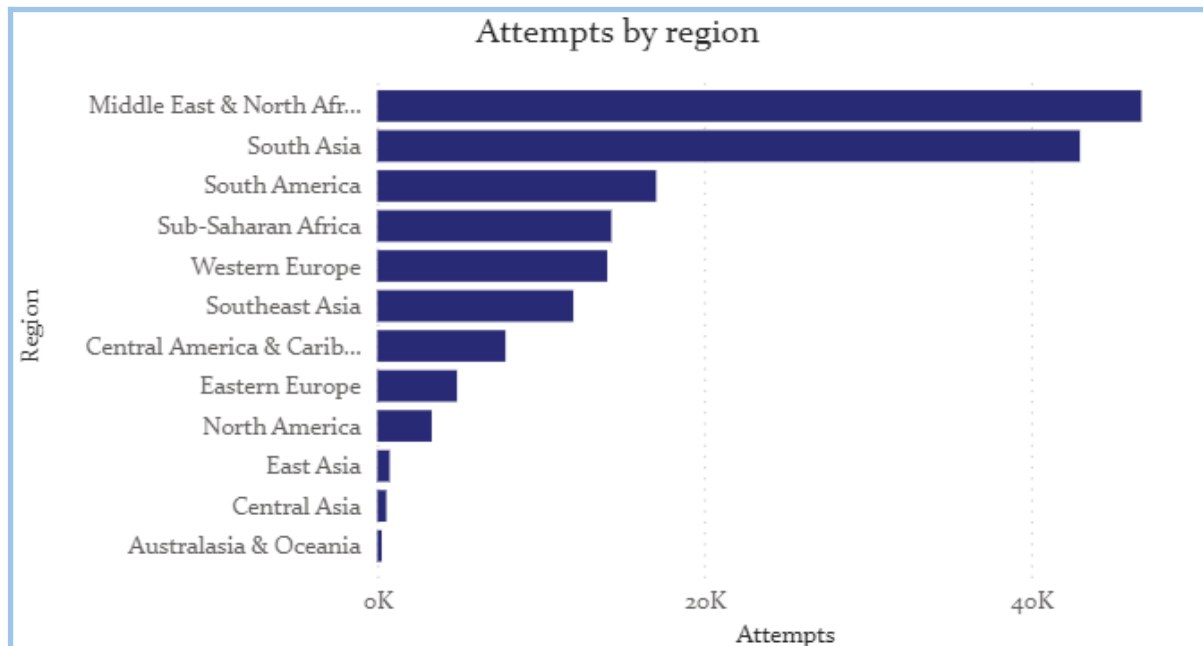
GEO

Este apartado se propone desagregar la información por áreas geográficas, procurando diferenciar la evolución temporal y las principales dimensiones de los ataques explicitadas anteriormente.

En primer lugar se plantea un mapa global con burbujas que permiten obtener información sobre la cantidad de atentados y el número de muertes. El número de atentados se refleja en el tamaño de la burbuja y el número de muertes se refleja en el color de la burbuja, con una escala degradada de verde (hasta 1.000 fallecidos), amarillo (entre 1.000 y 500 muertes) y rojo (superior a 5.000 muertes).



En sintonía con esto se propone un gráfico que permite observar las cantidades de atentados por región, observándose una fuerte concentración en Medio Oriente y Norte de África (46,75 mil atentados) y el Sur de Asia (42,99 mil atentados). Posteriormente aparece América del Sur y Europa Occidental (con principal preponderancia en el siglo pasado), así como la África Sub-sahariana, con una cantidad de atentados entre los 14.000 y 20.000.



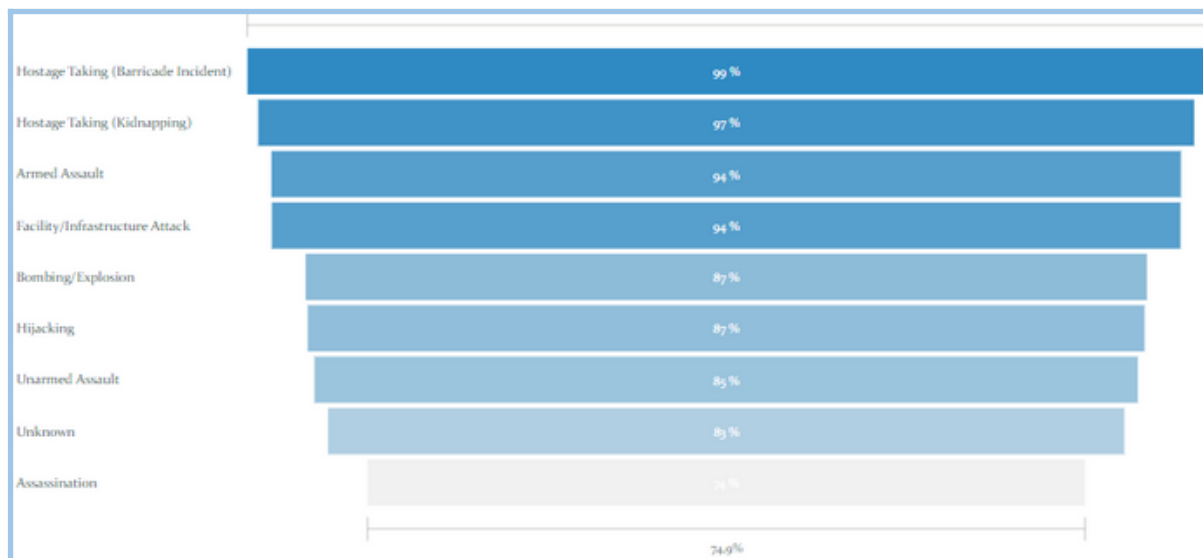
Al analizar por países, se observa que durante el siglo XX los países con mayor ataques terroristas son sudamericanos, entre ellos Colombia y Perú, con una cantidad de 5.794 y 5.319 atentados respectivamente. Le siguen países como El Salvador (3,89 mil), India (3,12 mil), España (2,64 mil) y Gran Bretaña (2,51 mil). A

partir del 2000 comienza a crecer sustantivamente la influencia de los conflictos en Medio Oriente y en el Sur de Asia, particularmente en países como Irak que alcanza para el año 2017 los 23.170 ataques terroristas con un saldo de 67.777 muertos y 133.347 heridos, Pakistán y Afganistán con el surgimiento de los talibanes con 12.182 y 11.830 ataques terroristas respectivamente y, nuevamente India, atentados causados por la ULFA, maoístas y el partido comunista, con un total de 8.539 atentados. Si se consideran exclusivamente estas dos regiones se concluye en un total 73.663 atentados que representan el 71% del total de casos del período analizado (2000-2017).

DETAILS

En primer lugar, se propone una matriz que explicita la cantidad de atentados y víctimas por país, con un pequeño gráfico que sintetiza la evolución en el tiempo y el máximo valor anual. A su vez, se plantea un análisis del éxito de los diferentes tipos de ataque, así como los tipos de ataque más reiterados para cada país, así como las principales armas según organización terrorista. Todo lo anterior se puede desagregar espacial y temporalmente para desentrañar la evolución de estas dimensiones.

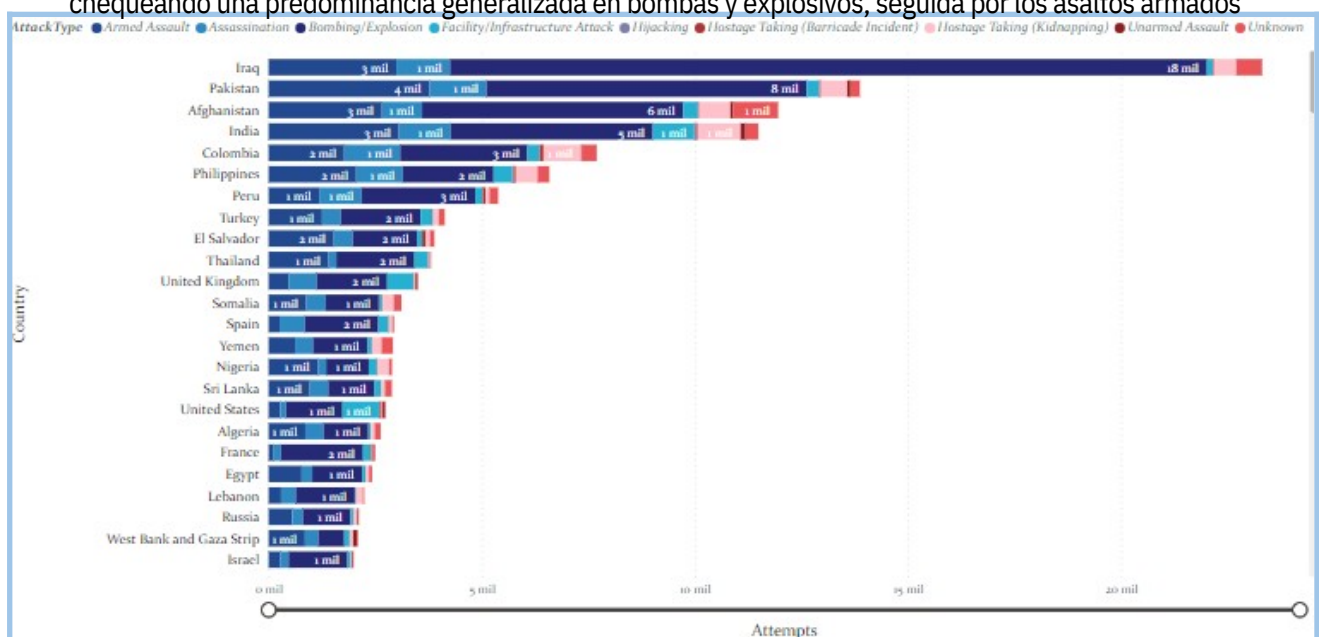
Para comenzar, se presenta la evolución de las tasas de éxito para los diferentes tipos de ataque, liderada por los secuestros, asaltos armados y ataques a edificios, con un porcentaje entre 94% y 99% de efectividad. Si consideramos el ataque más utilizado, es decir, bombas y explosivos, se observa un 87,23%.



Si observamos la tasa de éxito para cada región se observa que las regiones con mayores porcentajes son Latinoamérica y el Caribe y la África Sub-sahariana, con tasas superiores al 90%.

Continuando con el análisis del tipo de ataque, se desagrega según la cantidad de ataques para cada país,

chequeando una predominancia generalizada en bombas y explosivos, seguida por los asaltos armados



Las tendencias son sostenidas para los distintos países. Observándose una mayor predominancia en ataques a infraestructura y edificios, algunos países presentan una mayor participación de desconocimiento en sus datos, pero en líneas generales se concentran los ataques terroristas en asaltos armados y bombas o explosivos.

En el intento por generar un perfil de los ataques terroristas, se detallan los tipos de armas para

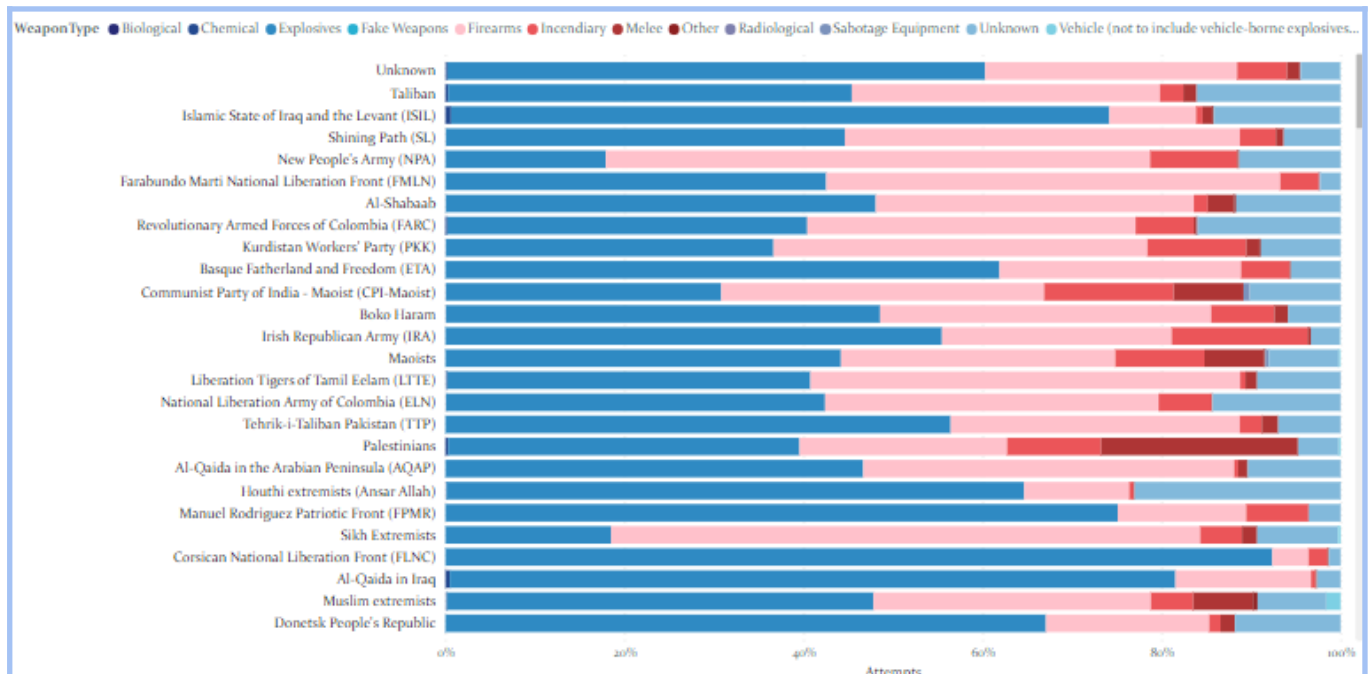
las

diferentes organizaciones terroristas, incluyendo el espectro de los atentados que no tienen adjudicado ningún autor. Se observa la predominancia de explosivos y armas de fuego, independientemente de

la

organización, en sintonía a los tipos de ataques detallados anteriormente. A su vez se observa un

porcentaje significativo de desconocimiento sobre las armas empleados en el general de las diferentes organizaciones. A su vez, cabe mencionar que organizaciones como los talibanes o el Estado Islámico presentan una mayor proporción de ataques con explosivos, mientras que otros grupos más relacionados a grupos políticos es más habitual el uso de armas de fuego.



CONCLUSIÓN FINAL

En conclusiones finales, un proyecto que creímos simple se nos terminó por dar un trabajo impresionante, no solo por el manejo de información que tuvimos, sino por el aprendizaje que conllevo el hacerlo. Durante el trabajo no solo aprendimos manejo de datos, si no de formato y diseño, tanto así como herramientas que seguramente nos ayudarán en el futuro.

La información obtenida a lo largo de la limpieza, y armando visualizaciones, nos indica una creciente de casos terrorista, y lo que es preocupante son la cantidad de vidas y heridos que causa cada uno de estos atentados, y de la misma forma, el aumento que está teniendo estos casos en los últimos años

a comparación de fechas anteriores. Los gráficos no mienten, y países como Irak son los más afectados. El terrorismo es un tema serio, y de mucho debate. No solo por los choques culturales, sino por las acciones y consecuencias que tiene el mismo. Cualquiera que sean las razones por la que cometer tales actos no son justificables, y el terrorismo nunca será la respuesta.

Como parte de la creciente de los atentados terrorista, y la preocupante tasa de éxito de los mismos, también es un indicador del grado de complejidad que tiene los gobiernos responsables del pueblo para

prevenirlos, y en los casos prevenidos, llevo detrás una gran investigación y seguimiento de sospechosos para poder evitar este tipo de catástrofes.

El trabajo final nos dio un resultado limpio, y una agradable visualización de datos, con un acabado algo minimalista, y una paleta de colores que ayuda a reflejar bien la importancia de los números. Se agradece

muchísimo al usuario responsable por la recopilación de datos del dataset, y doy las gracias a

mis

compañeros de trabajo por el buen trabajo logrado. Es importante mencionar que este trabajo no sería posible de no ser por la atención y buena documentación de Coderhouse al momento de la realización del curso.