

SHARK ATTACKS: DATA CLEANING & WRANGLING - PROJECT

Alexandra Rojas
Monica Duarte

datetime (2022, 05, 21)



Hipótesis

"Los tiburones atacan
mayormente a adolescentes"



Limpieza



- Nulos
- Duplicados
- Relevancia
- Formato



Data Original

	Case Number	Date	Year	Type	Country	Area	Location	Activity	Name	Sex	...	Species	Investigator or Source	pdf		
	0	2018.06.25	25-Jun-2018	2018.0	Boating	USA	California	Oceanside, San Diego County	Paddling	Julie Wolfe	F	...	White shark	R. Collier, GSAF	2018.06.25-Wolfe.pdf	h
	1	2018.06.18	18-Jun-2018	2018.0	Unprovoked	USA	Georgia	St. Simon Island, Glynn County	Standing	Adyson McNeely	F	...	NaN	K.McMurray, TrackingSharks.com	2018.06.18-McNeely.pdf	h
	2	2018.06.09	09-Jun-2018	2018.0	Invalid	USA	Hawaii	Habush, Oahu	Surfing	John Denges	M	...	NaN	K.McMurray, TrackingSharks.com	2018.06.09-Denges.pdf	h
	3	2018.06.08	08-Jun-2018	2018.0	Unprovoked	AUSTRALIA	New South Wales	Arrawarra Headland	Surfing	male	M	...	2 m shark	B. Myatt, GSAF	2018.06.08-Arrawarra.pdf	h
	4	2018.06.04	04-Jun-2018	2018.0	Provoked	MEXICO	Colima	La Ticla	Free diving	Gustavo Ramos	M	...	Tiger shark, 3m	A. Kipper	2018.06.04-Ramos.pdf	h
	
	25718	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	
	25719	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	
	25720	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	
	25721	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	
	25722	xx	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	

25723 rows × 24 columns

Data Original

	Case Number	Date	Year	Type	Country	Area	Location	Activity	Name	Sex	...	Species	Investigator or Source	pdf
0	2018.06.25	25-Jun-2018	2018.0	Boating	USA	California	Oceanside, San Diego County	Paddling	Julie Wolfe	F	...	White shark	R. Collier, GSAF	2018.06.25-Wolfe.pdf
1	2018.06.18	18-Jun-2018	2018.0	Unprovoked	USA	Georgia	St. Simon Island, Glynn County	Standing	Adyson McNeely	F	...	NaN	K.McMurray, TrackingSharks.com	2018.06.18-McNeely.pdf
2	2018.06.09	09-Jun-2018	2018.0	Invalid	USA	Hawaii	Habush, Oahu	Surfing	John Denges	M	...	NaN	K.McMurray, TrackingSharks.com	2018.06.09-Denges.pdf
3	2018.06.08	08-Jun-2018	2018.0	Unprovoked	AUSTRALIA	New South Wales	Arrawarra Headland	Surfing	male	M	...	2 m shark	B. Myatt, GSAF	2018.06.08-Arrawarra.pdf
4	2018.06.04	04-Jun-2018	2018.0	Provoked	MEXICO	Colima	La Ticla	Free diving	Gustavo Ramos	M	...	Tiger shark, 3m	A .Kipper	2018.06.04-Ramos.pdf
...
25718	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
25719	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
25720	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
25721	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
25722	xx	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN

25723 rows × 24 columns

Limpieza

- Limpieza de columnas con pocos valores (1 o 2)
 - Eliminamos filas con valores nulos y filas con 1 solo valor
 - Eliminamos columnas repetidas (Case number, Case number 1 y 2, href)
 - Convertimos a Datos numéricos las columnas de Age y Year
 - Convertimos a “datetime” las columnas Date y Time
 - Creamos un nuevo índice
 - Eliminamos columnas irrelevantes para el análisis: 'Investigator or Source', 'pdf','original order'
 - Arreglamos los nombres de las columnas
 - Normalizamos y corregimos los valores de las columnas: Type, Date, Name, Country, Activity, Sex y Species
-
- Pasamos de: 25723 rows × 24 columns a 2663 rows × 19 columns
 - Añadimos: Hour Month Decade Age range

Type	Date	Year	Time	Name	Sex	Age	Country	Area	Location	Activity	Injury	Fatal (Y/N)	Species	Hour	Month	Decade	Age range
Boat	2018-06-25	2018	18:00:00	Julie Wolfe	F	57	USA	California	Oceanside, San Diego County	paddle	No injury to occupant, outrigger canoe and pad...	N	white shark	18	6	2010's	Adult
provoked	2018-06-18	2018	00:00:00	Adyson McNeely	F	11	USA	Georgia	St. Simon Island, Glynn County	stand	Minor injury to left thigh	N	unknown	0	6	2010's	Child
invalid	2018-06-09	2018	07:45:00	John Denges	M	48	USA	Hawaii	Habush, Oahu	surf	Injury to left lower leg from surfboard skeg	N	unknown	7	6	2010's	Adult
provoked	2018-06-03	2018	15:00:00	Jose Ernesto da Silva	M	18	BRAZIL	Pernambuco	Piedade Beach, Recife	swim	FATAL	Y	tiger shark	15	6	2010's	Teenager
provoked	2018-05-27	2018	00:00:00	Unknown	M	52	USA	Florida	Lighthouse Point Park, Ponce Inlet, Volusia County	fish	Minor injury to foot. PROVOKED INCIDENT	N	lemon shark, 3'	0	5	2010's	Adult

Análisis

Hipótesis: “Los tiburones atacan mayormente a los adolescentes”.

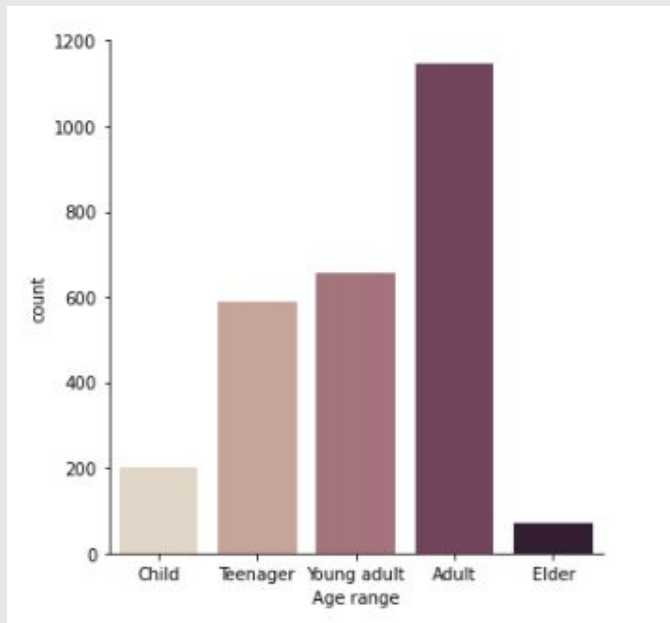
1. Crear “Rango de edades para comparar datos: Trabajar con categorías

Child	0-12
Teenagers	12-18
Young adults	18-25
Adults	25-60
Elder	60-99

Análisis

Para llegar a conclusiones específicas, preguntamos a los datos información respecto a los adolescentes:

¿Son el rango de edad con más ataques?



Adult	42.99%
Young adult	24.67%
Teenager	22.15%
Child	7.47%
Elder	2.70%

Análisis

¿Son los ataques a adolescentes más fatales?

Age range		Fatal (Y/N)
Child	N	92.592593
	Y	7.407407
Teenager	N	82.241379
	Y	17.758621
Young adult	N	81.446541
	Y	18.396226
	UNKNOWN	0.157233
Adult	N	85.405405
	Y	14.504505
	UNKNOWN	0.090090
Elder	N	73.134328
	Y	26.865672

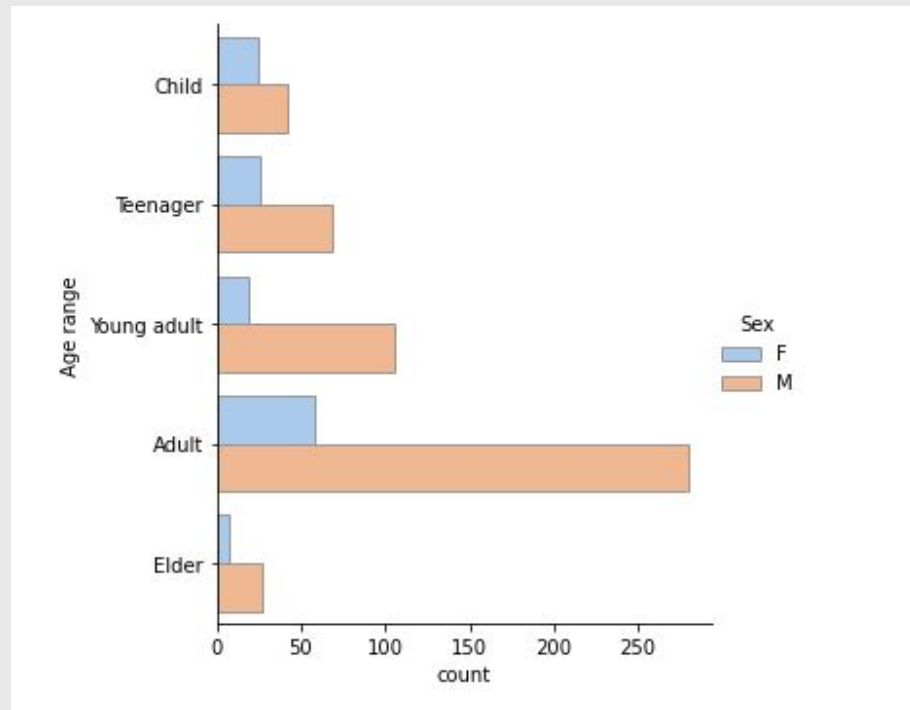
A pesar de que los adolescentes no son el grupo más atacado, su fatalidad es la tercera más alta (17.75%)

Análisis

¿Los ataques en adolescentes se distribuyen igual entre hombres y mujeres?

Age range	Sex	
Child	M	71.356784
	F	28.140704
	Genderless	0.502513
Teenager	M	83.050847
	F	16.610169
	Genderless	0.338983
Young adult	M	90.258752
	F	9.589041
	Genderless	0.152207
Adult	M	87.161572
	F	12.489083
	Genderless	0.349345
Elder	M	76.388889
	F	23.611111

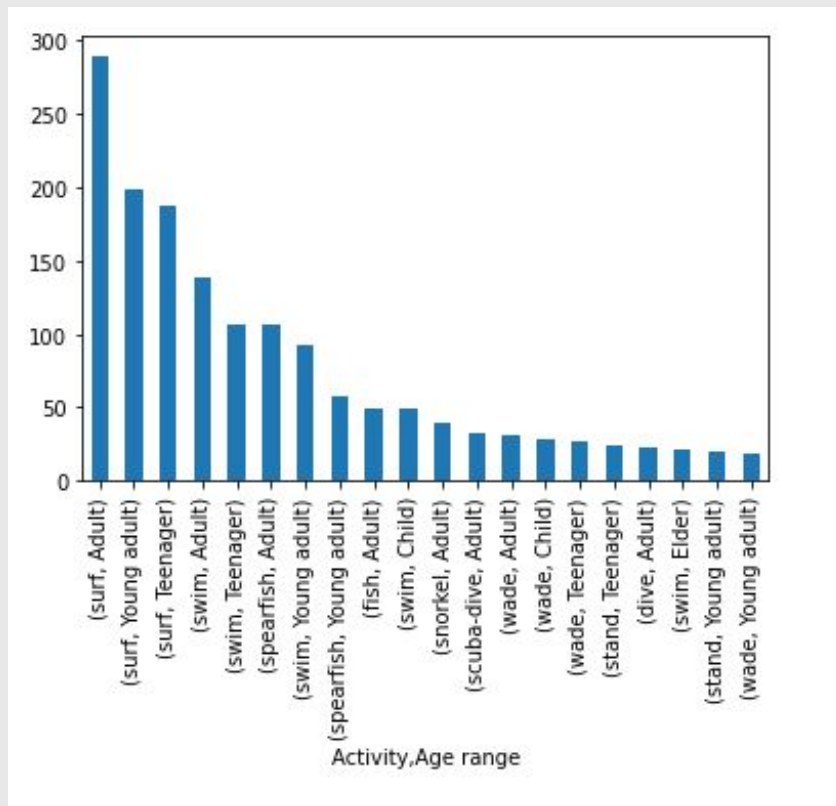
Vemos que en todos los rangos de edad los hombres representan la mayoría de los ataques. Sin embargo, el porcentaje de mujeres atacadas es mayor en niños y ancianos y en tercer lugar las mujeres adolescentes.



Análisis

¿En qué actividad atacan más a los adolescentes?

surf	188
swim	107
wade	27
stand	24
spearfish	19



Análisis

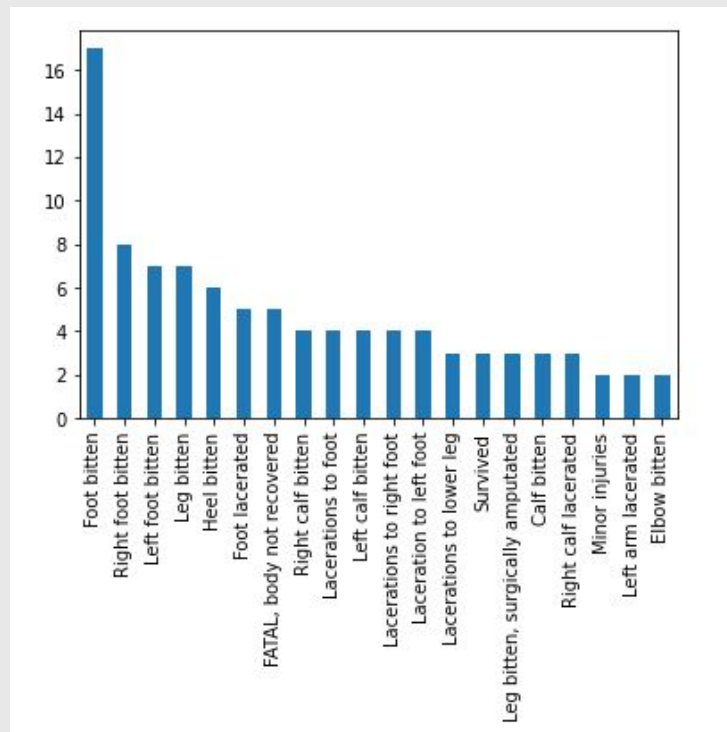
¿Son los adolescentes igual de atacados en todos los países?

	Total	Teens
USA	1270.0	295.0
AUSTRALIA	569.0	132.0
SOUTH AFRICA	323.0	83.0
BAHAMAS	46.0	4.0
BRAZIL	40.0	11.0
NEW ZEALAND	37.0	11.0
REUNION	31.0	6.0
MEXICO	30.0	2.0
MOZAMBIQUE	20.0	3.0
PAPUA NEW GUINEA	20.0	3.0

Vemos que USA, Australia y Suráfrica representan la mayor cantidad de ataques totales y entre los adolescentes.
Sin embargo, Bahamas pasa de ser el 4to al 7mo.

Análisis

Lesiones más comunes en adolescentes



Number of Foot injuries : 448

Number of Leg injuries : 451

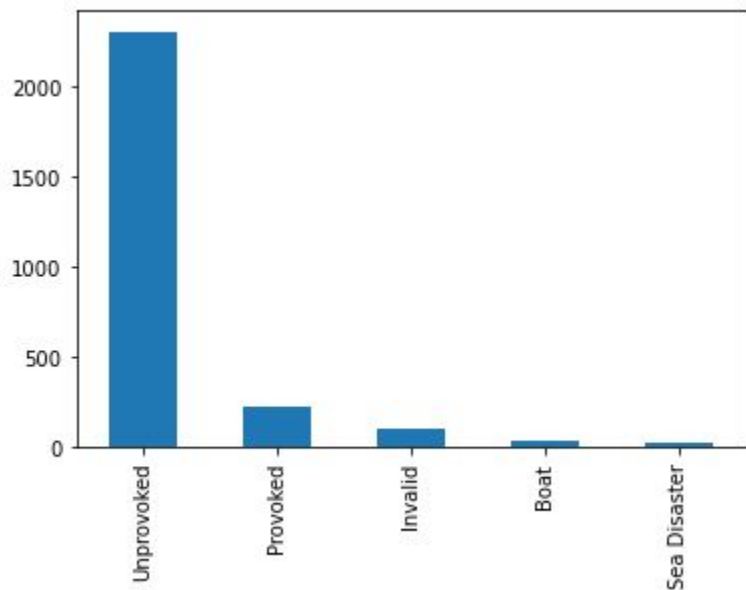
Number of Arm injuries: 286

Number of Hand injuries: 240

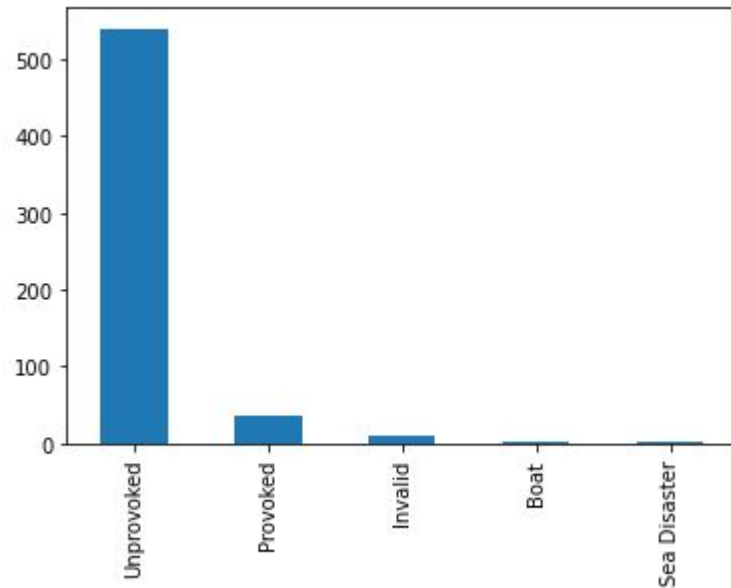
Análisis

Unprovoked	540
Provoked	35
Invalid	11
Boat	3
Sea Disaster	1

¿ Los adolescentes provocan más los Ataques de tiburones?



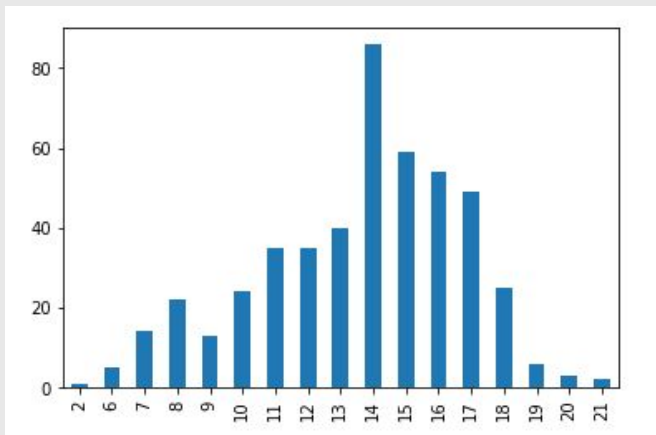
Distribución total



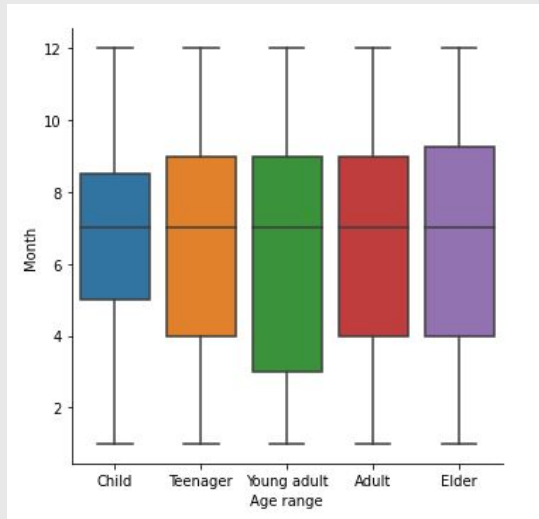
Adolescentes

Análisis

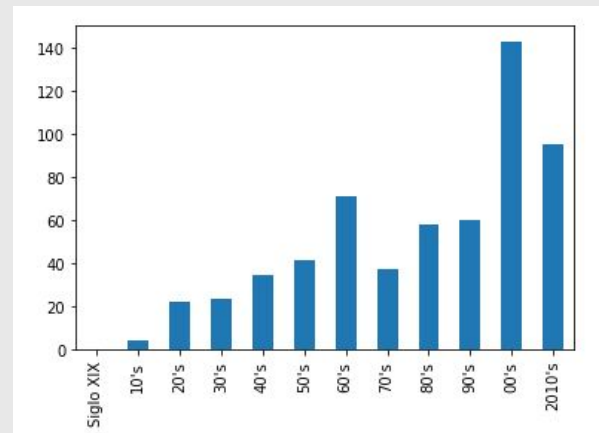
¿Cómo es la distribución temporal de los ataques a adolescentes? ¿Han aumentado los ataques en el tiempo?



Horas



Mes

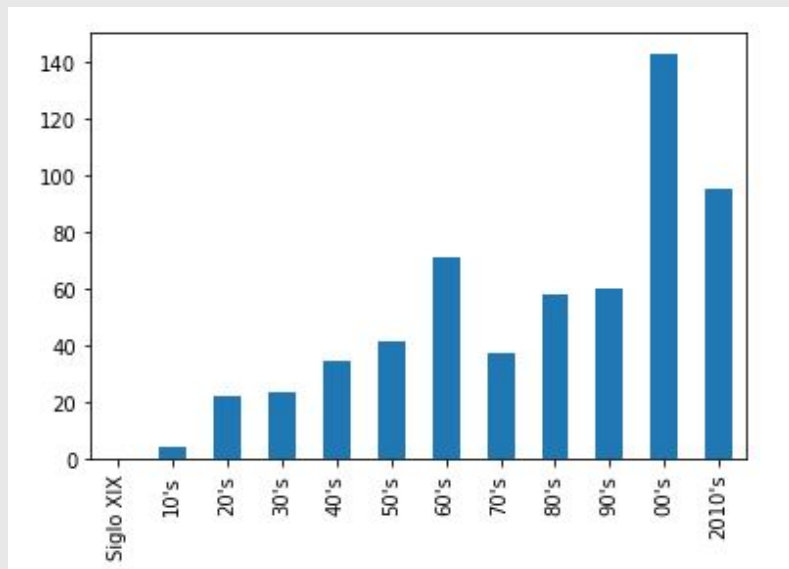


Décadas

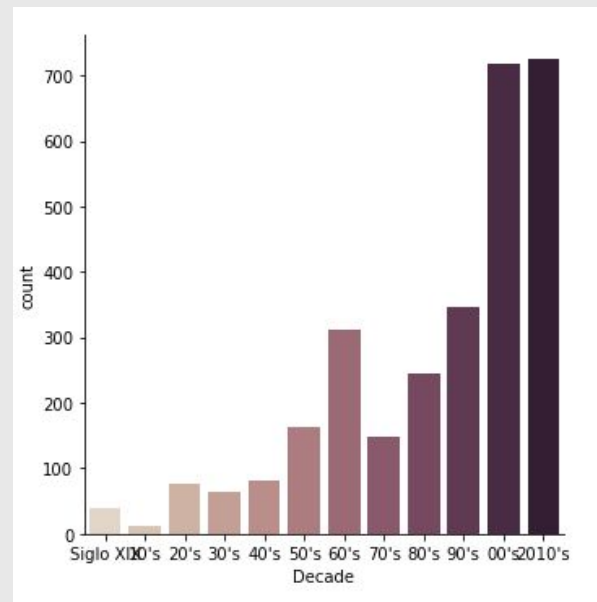
Análisis

2018	42
2017	89
2016	93
2015	107
2014	77

¿Cómo es la distribución temporal de los ataques a adolescentes? ¿Han aumentado los ataques en el tiempo?



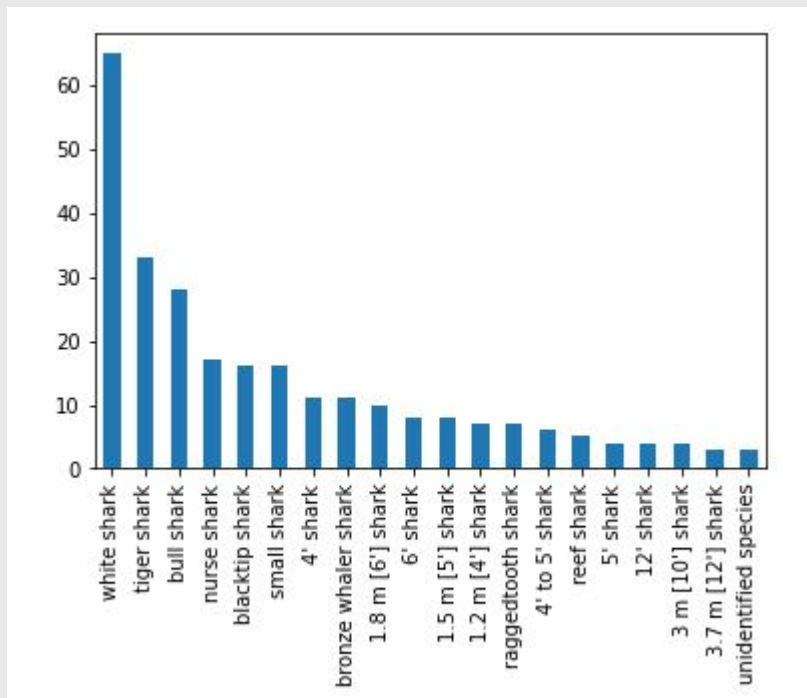
Adolescentes



Total

Análisis

¿Cuáles son las especies que más atacan a los adolescentes?



white shark	65
tiger shark	33
bull shark	28
nurse shark	17
blacktip shark	16

Conclusión

"Los adolescentes no son el grupo de edad más atacado. Y en la mayoría de las variables los datos se distribuyen igual que en la totalidad de ataques"

