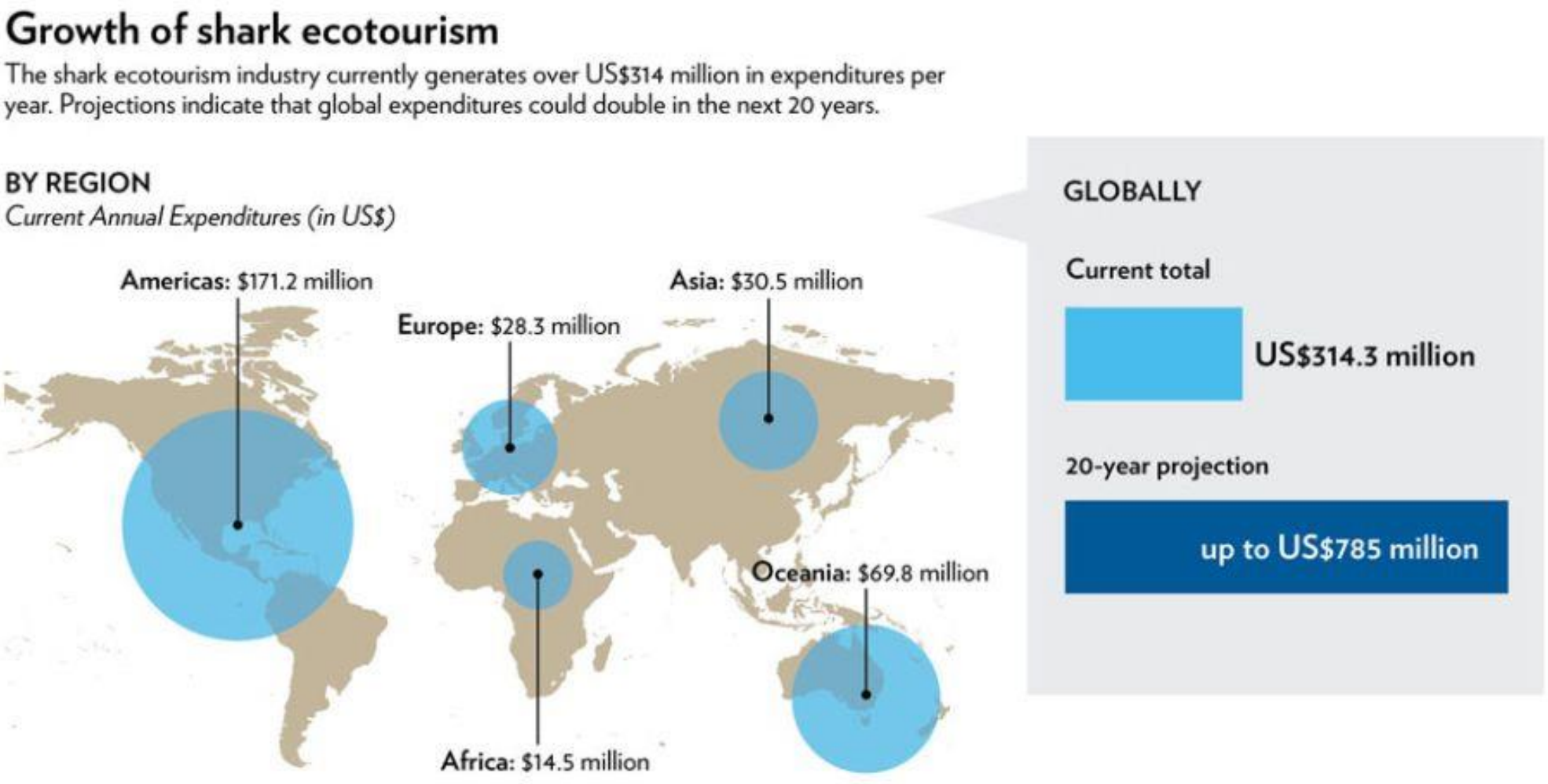

SHARK PROJECT

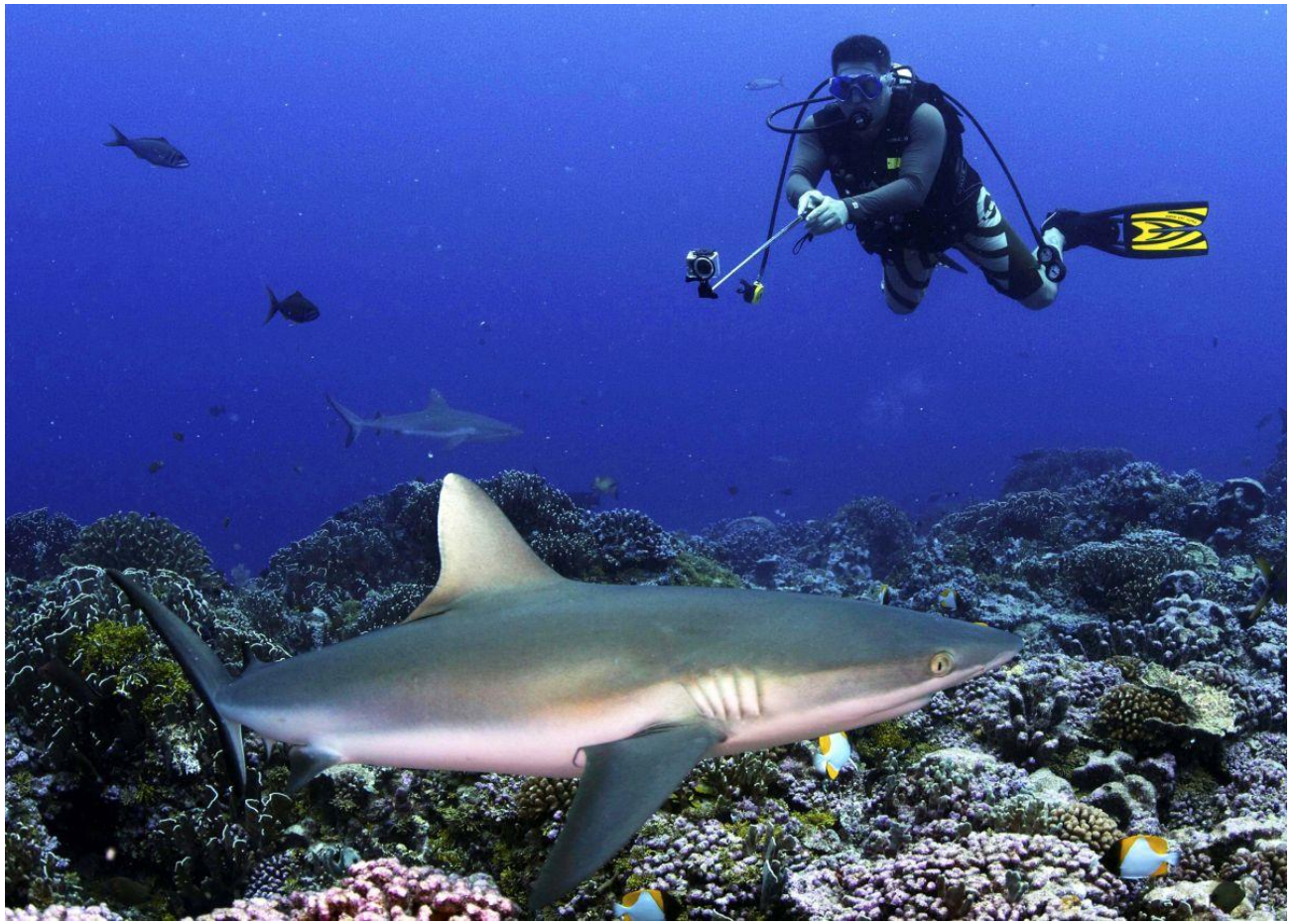
Answer a ‘business’ question!

Achar um lugar adequado na Oceania para estabelecer uma empresa de ecoturismo do tubarão.

Está amplamente demonstrado que um tubarão vivo, reporta muito mais dinheiro que um tubarão morto.

Currently, shark ecotourism brings in $314 million annually worldwide, and this sector is expected to continue growing.





- Arrumando o nome das colunas

```
In [58]: df.columns= df.columns.str.lower()
         df.columns = df.columns.str.replace(' ', '_')
```

```
In [59]: df.head()
```

Out[59]:

	case_number	date	year	type	country	area	location	activity	name	sex_	age	injury	fatal_(y/n)	time	species_
0	2018.06.25	25-Jun-2018	2018.0	Boating	USA	California	Oceanside, San Diego County	Paddling	Julie Wolfe	F	57	No injury to occupant, outrigger canoe and pad...	N	18h00	White shark
							St. Simon								

```
In [7]: df = df.rename(columns={'sex_': 'sex', 'fatal_(y/n)': 'fatal', 'species_': 'species'})
```

```
In [8]: df.info()
```

```
 0   case_number             8702 non-null   object
 1   date                    6302 non-null   object
 2   year                    6300 non-null   float64
 3   type                    6298 non-null   object
 4   country                 6252 non-null   object
 5   area                    5847 non-null   object
 6   location                5762 non-null   object
 7   activity                5758 non-null   object
 8   name                    6092 non-null   object
 9   sex                     5737 non-null   object
 10  age                     3471 non-null   object
 11  injury                  6274 non-null   object
 12  fatal                   5763 non-null   object
 13  time                    2948 non-null   object
 14  species                 3464 non-null   object
 15  investigator_or_source  6285 non-null   object
```

- Escolhendo as colunas que usarei

```
In [537]: df2 = df[['date', 'year', 'country', 'area', 'location', 'activity', 'fatal', 'species']]
          df2.head()
```

Out[537]:

	date	year	country	area	location	activity	fatal	species
0	25-Jun-2018	2018.0	USA	California	Oceanside, San Diego County	Paddling	N	White shark
1	18-Jun-2018	2018.0	USA	Georgia	St. Simon Island, Glynn County	Standing	N	NaN
2	09-Jun-2018	2018.0	USA	Hawaii	Habush, Oahu	Surfing	N	NaN
3	08-Jun-2018	2018.0	AUSTRALIA	New South Wales	Arrawarra Headland	Surfing	N	2 m shark
4	04-Jun-2018	2018.0	MEXICO	Colima	La Ticla	Free diving	N	Tiger shark, 3m

- Limpando valores NaN

```
In [540]: df3 = df2.dropna(axis = 0, how = 'all')
```

```
In [541]: df3.shape
```

```
Out[541]: (6302, 8)
```

```
In [542]: df3.tail()
```

Out[542]:

	date	year	country	area	location	activity	fatal	species
6297	Before 1903	0.0	AUSTRALIA	Western Australia	Roebuck Bay	Diving	Y	NaN
6298	Before 1903	0.0	AUSTRALIA	Western Australia	NaN	Pearl diving	Y	NaN
6299	1900-1905	0.0	USA	North Carolina	Ocracoke Inlet	Swimming	Y	NaN
6300	1883-1889	0.0	PANAMA	NaN	Panama Bay 8°N, 79°W	NaN	Y	NaN
6301	1845-1853	0.0	CEYLON (SRI LANKA)	Eastern Province	Below the English fort, Trincomalee	Swimming	Y	NaN

- Separando os 10 países da Oceania com mais ataques de tubarão

```
In [543]: df3['country'].value_counts().head(50)
```

```
Out[543]: USA                    2229
          AUSTRALIA              1338
          SOUTH AFRICA            579
          PAPUA NEW GUINEA        134
          NEW ZEALAND             128
          BRAZIL                  112
          BAHAMAS                 109
          MEXICO                   89
          ITALY                    71
          FIJI                     62
          PHILIPPINES              61
          REUNION                  60
          NEW CALEDONIA            53
          CUBA                     46
          MOZAMBIQUE               45
          SPAIN                    44
          INDIA                    40
          EGYPT                    38
          CROATIA                  34
          JAPAN                    34
          PANAMA                   32
          SOLOMON ISLANDS          30
```

```
In [611]: df4 = df3.drop(df3[(df3.country != 'AUSTRALIA') & (df3.country != 'PAPUA NEW GUINEA') & (df3.country != 'NEW ZEALAND')
                              & (df3.country != 'FIJI') &(df3.country != 'NEW CALEDONIA') & (df3.country != 'SOLOMON ISLANDS')
                              & (df3.country != 'FRENCH POLYNESIA') & (df3.country != 'TONGA') & (df3.country != 'VANUATU')
                              & (df3.country != 'MARSHALL ISLANDS')].index)
```

- Procurando o país da Oceania com mais ataques de tubarão por km de costa

In [614]:
```
coastline = [34218, 5150, 15000, 1129, 2254, 5313, 2525, 419, 2528, 370]

oceania['coastline_km'] = coastline
```

In [615]:
```
oceania
```

Out[615]:

	country	attacks	coastline_km
0	AUSTRALIA	1338	34218
1	PAPUA NEW GUINEA	134	5150
2	NEW ZEALAND	128	15000
3	FIJI	62	1129
4	NEW CALEDONIA	53	2254
5	SOLOMON ISLANDS	30	5313
6	FRENCH POLYNESIA	25	2525
7	TONGA	15	419
8	VANUATU	14	2528
9	MARSHALL ISLANDS	13	370

In [619]:
```
oceania['attacks/km'] = oceania['attacks'] / oceania['coastline_km']
```

In [620]:
```
oceania
```

Out[620]:

	country	attacks	coastline_km	attacks/km
0	AUSTRALIA	1338.0	34218.0	0.039102
1	PAPUA NEW GUINEA	134.0	5150.0	0.026019
2	NEW ZEALAND	128.0	15000.0	0.008533
3	FIJI	62.0	1129.0	0.054916
4	NEW CALEDONIA	53.0	2254.0	0.023514
5	SOLOMON ISLANDS	30.0	5313.0	0.005647
6	FRENCH POLYNESIA	25.0	2525.0	0.009901
7	TONGA	15.0	419.0	0.035800
8	VANUATU	14.0	2528.0	0.005538
9	MARSHALL ISLANDS	13.0	370.0	0.035135

```
In [619]: oceania['attacks/km'] = oceania['attacks'] / oceania['coastline_km']
```

```
In [620]: oceania
```

Out[620]:

	country	attacks	coastline_km	attacks/km
0	AUSTRALIA	1338.0	34218.0	0.039102
1	PAPUA NEW GUINEA	134.0	5150.0	0.026019
2	NEW ZEALAND	128.0	15000.0	0.008533
3	FIJI	62.0	1129.0	0.054916
4	NEW CALEDONIA	53.0	2254.0	0.023514
5	SOLOMON ISLANDS	30.0	5313.0	0.005647
6	FRENCH POLYNESIA	25.0	2525.0	0.009901
7	TONGA	15.0	419.0	0.035800
8	VANUATU	14.0	2528.0	0.005538
9	MARSHALL ISLANDS	13.0	370.0	0.035135

Mesmo que o país com mais ataques por km de costa seja Fiji, escolhi Australia (que es o segundo neste ranking) por população e turismo.

- Escolhendo a área da Australia para o negôcio

```
In [623]: australia['area'].value_counts()
```

```
Out[623]: New South Wales                        486
          Queensland                             311
          Western Australia                      189
          South Australia                        104
          Victoria                                90
          Torres Strait                           70
          Tasmania                                41
          Northern Territory                      23
          Westerm Australia                        3
          Victoria                                 2
          Torres Strait                            2
          Territory of Cocos (Keeling) Islands     1
          Queensland                               1
          Norfolk Island                           1
          Name: area, dtype: int64
```

```
In [624]: nsw = australia.drop(australia[(australia.area != 'New South Wales')].index)
```

- Refinando o lugar de New South Wales

In [695]:
```
nsw['location'].value_counts().head(15)
```

Out[695]:
```
Sydney                      10
Sydney Harbor               10
Bondi                        6
Wollongong                   6
Coogee                       6
Newcastle                    5
Cronulla                     4
Lake Macquarie               4
Parramatta River             4
Byron Bay                    4
Marineland, Sydney           4
Avoca Beach                  4
Stockton Beach, Newcastle    4
Lennox Head                  4
Evans Head                   4
Name: location, dtype: int64
```

In [696]:
```
nsw.loc[nsw['location'].str.contains('Sydney') == True, 'location'] = 'Sydney'
```

In [697]:
```
nsw.loc[nsw['location'].str.contains('Newcastle') == True, 'location'] = 'Newcastle'
```

In [698]:
```
nsw.loc[nsw['location'].str.contains('Bondi') == True, 'location'] = 'Sydney'
```

In [699]:
```
nsw.loc[nsw['location'].str.contains('Coogee') == True, 'location'] = 'Sydney'
```

In [700]:
```
nsw.loc[nsw['location'].str.contains('Cronulla') == True, 'location'] = 'Sydney'
```

In [701]:
```
nsw.loc[nsw['location'].str.contains('Lake Macquarie') == True, 'location'] = 'Newcastle'
```

AND THE WINNER IS:

In [702]:
```
nsw['location'].value_counts().head()
```

Out[702]:
```
Sydney              130
Newcastle            38
Wollongong            6
Byron Bay             4
Parramatta River      4
Name: location, dtype: int64
```

In [703]:
```
nsw.head()
```

Out[703]:

	date	year	country	area	location	activity	fatal	species
3	08-Jun-2018	2018.0	AUSTRALIA	New South Wales	Arrawarra Headland	Surfing	N	2 m shark
5	03-Jun-2018	2018.0	AUSTRALIA	New South Wales	Flat Rock, Ballina	Kite surfing	N	NaN
16	09-May-2018	2018.0	AUSTRALIA	New South Wales	Sharpes Beach, Ballina	Surfing	N	Shark involvement not confirmed
20	25-Apr-2018	2018.0	AUSTRALIA	New South Wales	Lennox Head	Surfing	N	Questionable
40	23-Feb-2018	2018.0	AUSTRALIA	New South Wales	Little Congwong Beach, La Perouse	Swimming	N	Juvenile white shark, 2.7 to 3.2 m

In [704]:
```
sydney_newcastle = nsw.drop(nsw[(nsw.location != 'Sydney') & (nsw.location != 'Newcastle')].index)
```

In [705]:
```
sydney_newcastle.shape
```

Out[705]:
```
(168, 8)
```

AND THE WINNER IS: SYDNEY AND NEWCASTLE

In [702]:
```
nsw['location'].value_counts().head()
```

Out[702]:
```
Sydney              130
Newcastle            38
Wollongong            6
Byron Bay             4
Parramatta River      4
Name: location, dtype: int64
```

In [703]:
```
nsw.head()
```

Out[703]:

	date	year	country	area	location	activity	fatal	species
3	08-Jun-2018	2018.0	AUSTRALIA	New South Wales	Arrawarra Headland	Surfing	N	2 m shark
5	03-Jun-2018	2018.0	AUSTRALIA	New South Wales	Flat Rock, Ballina	Kite surfing	N	NaN
16	09-May-2018	2018.0	AUSTRALIA	New South Wales	Sharpes Beach, Ballina	Surfing	N	Shark involvement not confirmed
20	25-Apr-2018	2018.0	AUSTRALIA	New South Wales	Lennox Head	Surfing	N	Questionable
40	23-Feb-2018	2018.0	AUSTRALIA	New South Wales	Little Congwong Beach, La Perouse	Swimming	N	Juvenile white shark, 2.7 to 3.2 m

In [704]:
```
sydney_newcastle = nsw.drop(nsw[(nsw.location != 'Sydney') & (nsw.location != 'Newcastle')].index)
```

In [705]:
```
sydney_newcastle.shape
```

Out[705]:
```
(168, 8)
```

AND THE WINNER IS: SYDNEY AND NEWCASTLE

In [702]:
```
nsw['location'].value_counts().head()
```

Out[702]:
```
Sydney              130
Newcastle            38
Wollongong            6
Byron Bay             4
Parramatta River      4
Name: location, dtype: int64
```

In [703]:
```
nsw.head()
```

Out[703]:

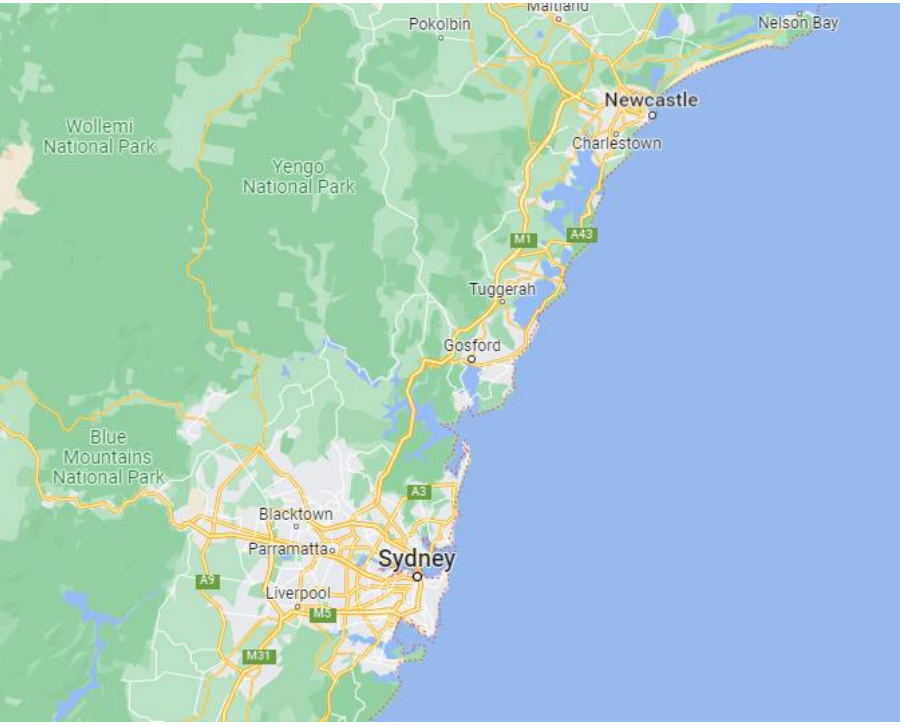
	date	year	country	area	location	acti
3	08-Jun-2018	2018.0	AUSTRALIA	New South Wales	Arrawarra Headland	Sur
5	03-Jun-2018	2018.0	AUSTRALIA	New South Wales	Flat Rock, Ballina	Kite sur
16	09-May-2018	2018.0	AUSTRALIA	New South Wales	Sharpes Beach, Ballina	Sur
20	25-Apr-2018	2018.0	AUSTRALIA	New South Wales	Lennox Head	Sur
40	23-Feb-2018	2018.0	AUSTRALIA	New South Wales	Little Congwong Beach, La Perouse	Swimm

In [704]:
```
sydney_newcastle = nsw.drop(nsw[(nsw.location != 'Sydney') & (nsw.location !=
```

In [705]:
```
sydney_newcastle.shape
```

Out[705]:
```
(168, 8)
```

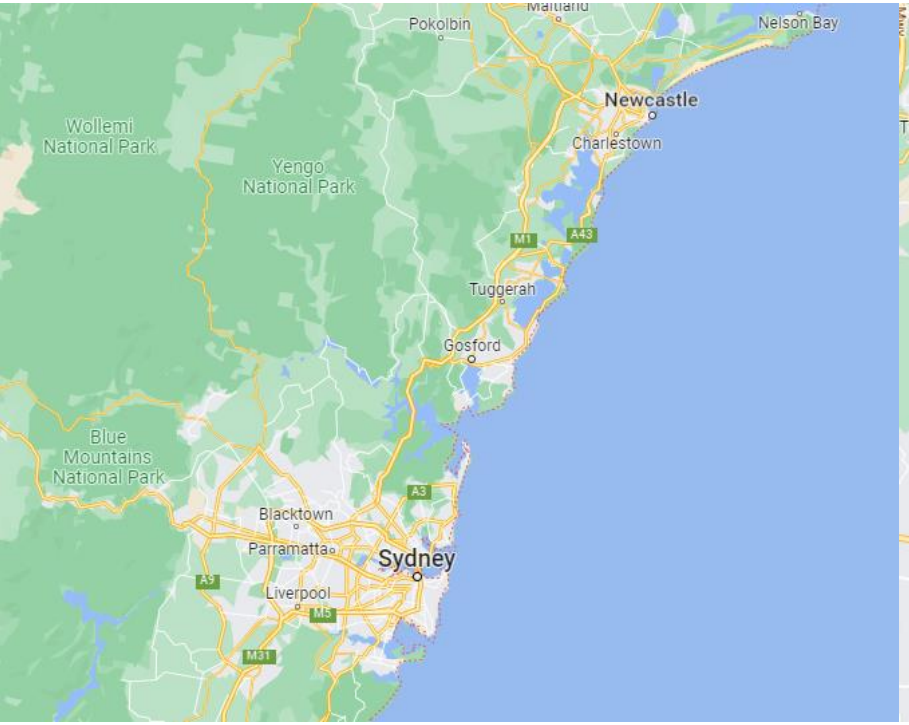
Sydney e Newcastle são duas cidades que se encontram a 150km de distancia, assim nos estabelecendo num ponto médio entre estas duas cidades, com uma lancha com velocidade de cruzeiro de 20 nós, em duas horas e meia conseguimos abarcar uma grande área para oferecer diferentes pontos para mergulhar com tubarões.

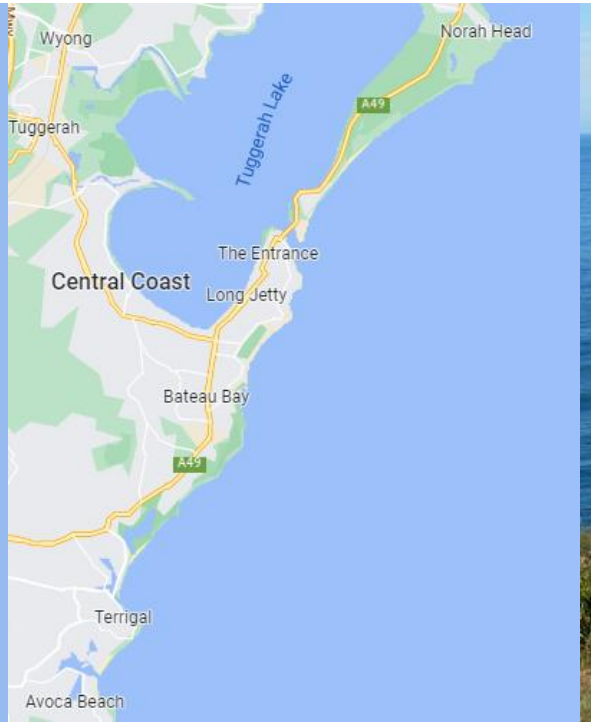


Sydney e Newcastle são duas cidades que se encontram a 150km de distancia, assim nos estabelecendo num ponto médio entre estas duas cidades, com uma lancha com velocidade de cruzeiro de 20 nós, em duas horas e meia conseguimos abarcar uma grande área para oferecer diferentes pontos para mergulhar com tubarões.

BATEAU BAY

Sydney e Newcastle são duas cidades que se encontram a 150km de distancia, assim nos estabelecendo num ponto médio entre estas duas cidades, com uma lancha com velocidade de cruzeiro de 20 nós, em duas horas e meia conseguimos abarcar uma grande área para oferecer diferentes pontos para mergulhar com tubarões.





Vendo o listado atividades, podemos ver que a grande maioria foi nadando, tomando banho, e praticando surf, pelo que presumivelmente não precisaremos de nos adentrar no mar, e sim procurar lugares de costa com presença habitual destes

```
In [707]: sydney_newcastle['activity'].value_counts().head(10)

Out[707]: Swimming          29
          Bathing           20
          Surfing           14
          Fishing           12
          Surf skiing        5
          Spearfishing       3
          Standing           3
          Sailing            3
          Fell overboard     2
          Rowing             2
          Name: activity, dtype: int64
```

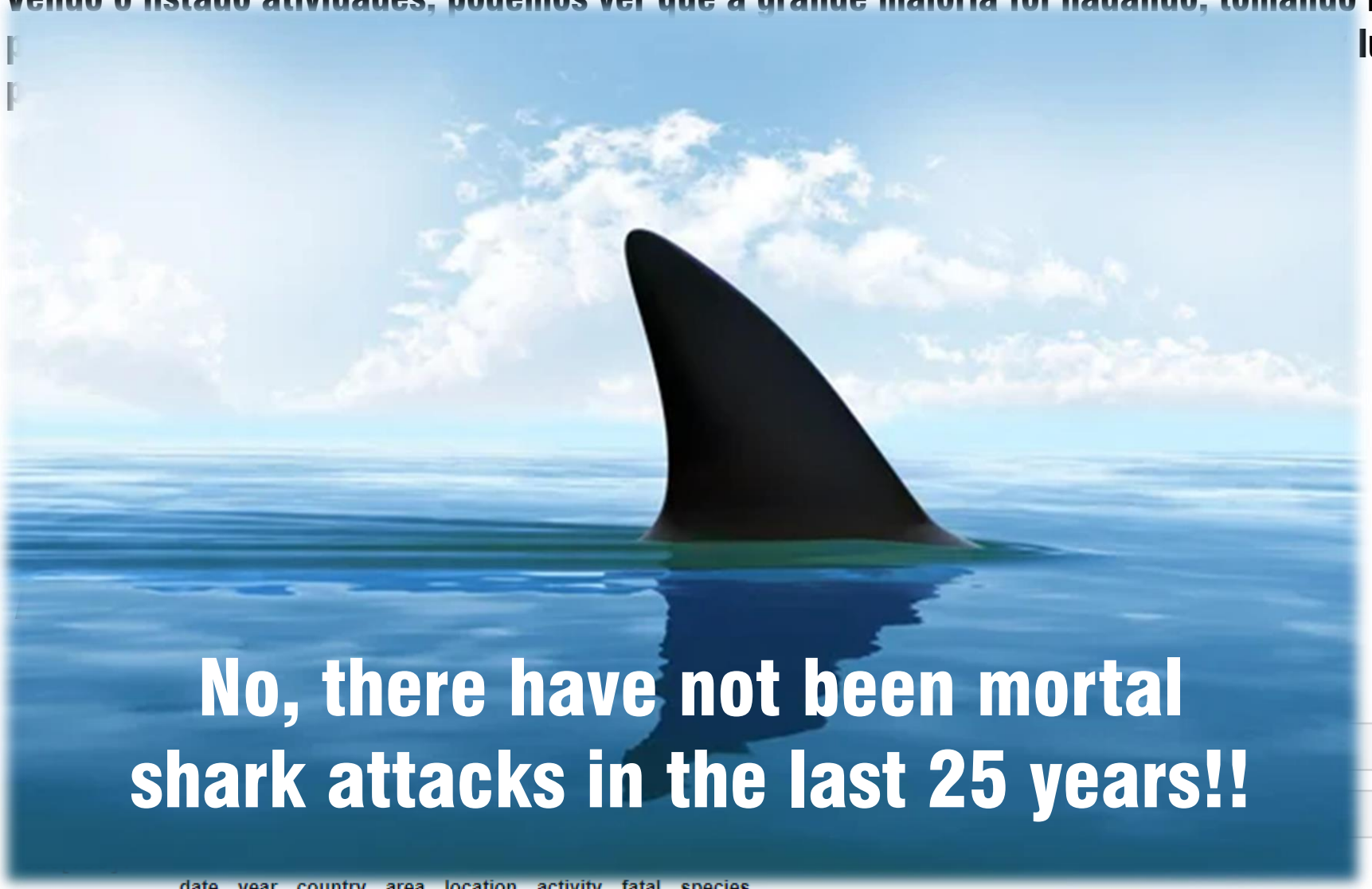
Analisando os ataques mortais nos últimos 25 anos nesta região

```
In [789]: sn_fatal_25 = sydney_newcastle[(sydney_newcastle.fatal == 'Y') & (sydney_newcastle.year > 1997)]

In [790]: sn_fatal_25

Out[790]:
```

date	year	country	area	location	activity	fatal	species

Vendo o listado atividades, podemos ver que a grande maioria foi nadando, tomando banho, e praticando surf,

lugares de costa com



date	year	country	area	location	activity	fatal	species

Analisando os ataques não mortais nos últimos 25 anos nesta região

In [791]:
```
sn_nao_fatal_25 = sydney_newcastle[(sydney_newcastle.fatal == 'N') & (sydney_newcastle.year > 1997)]
```

In [792]:
```
sn_nao_fatal_25
```

Out[792]:

	date	year	country	area	location	activity	fatal	species
0	28-Mar-2016	2016.0	AUSTRALIA	New South Wales	Sydney	Surfing	N	NaN
1	21-Dec-2015	2015.0	AUSTRALIA	New South Wales	Sydney	Surfing	N	Bronze whaler shark, 6'
2	13-Dec-2015	2015.0	AUSTRALIA	New South Wales	Newcastle	Fishing	N	White shark, 3.5 m
3	03-Jun-2012	2012.0	AUSTRALIA	New South Wales	Newcastle	Surf skiing	N	White shark, 2 m
4	11-Feb-2010	2010.0	AUSTRALIA	New South Wales	Sydney	Surfing	N	Wobbegong shark, 1.6m
5	12-Feb-2009	2009.0	AUSTRALIA	New South Wales	Sydney	Surfing	N	White shark, 2.5m
6	11-Feb-2009	2009.0	AUSTRALIA	New South Wales	Sydney	Diving, but on the surface when bitten by the ...	N	Bull shark, 2.7 m
7	27-Dec-2008	2008.0	AUSTRALIA	New South Wales	Sydney	Kayaking	N	White shark, 4m to 5m
8	11-Oct-2008	2008.0	AUSTRALIA	New South Wales	Newcastle	Surfing	N	NaN
10	11-Apr-2006	2006.0	AUSTRALIA	New South Wales	Newcastle	Surfing	N	Bronze whaler shark, a juvenile
11	15-Mar-2006	2006.0	AUSTRALIA	New South Wales	Sydney	Surfing	N	Bronze whaler shark, 2 m
12	01-Nov-2003	2003.0	AUSTRALIA	New South Wales	Newcastle	Standing	N	NaN
13	11-Feb-2003	2003.0	AUSTRALIA	New South Wales	Sydney	Swimming	N	Thought to involve a 2 m [6.75'] grey nurse shark
15	12-Apr-2002	2002.0	AUSTRALIA	New South Wales	Newcastle	Swimming	N	NaN
16	07-Feb-2002	2002.0	AUSTRALIA	New South Wales	Sydney	Kayaking	N	C. leucas tooth fragment recovered from kayak
18	02-Mar-2000	2000.0	AUSTRALIA	New South Wales	Sydney	Swimming	N	NaN

Apenas 19 ataques não mortais nos últimos 25 anos

Vamos ver as preciosidades que os nossos clientes terão a oportunidade de contemplar

In [864]:
```
sydney_newcastle['species'] = sydney_newcastle['species'].str.lower()
```

In [865]:
```
sydney_newcastle.head()
```

Out[865]:

	date	year	country	area	location	activity	fatal	species
0	28-Mar-2016	2016.0	AUSTRALIA	New South Wales	Sydney	Surfing	N	NaN
1	21-Dec-2015	2015.0	AUSTRALIA	New South Wales	Sydney	Surfing	N	bronze whaler shark, 6'
2	13-Dec-2015	2015.0	AUSTRALIA	New South Wales	Newcastle	Fishing	N	white shark, 3.5 m
3	03-Jun-2012	2012.0	AUSTRALIA	New South Wales	Newcastle	Surf skiing	N	white shark, 2 m
4	11-Feb-2010	2010.0	AUSTRALIA	New South Wales	Sydney	Surfing	N	wobbegong shark, 1.6m

In [866]:
```
sydney_newcastle.loc[sydney_newcastle['species'].str.contains('white') == True, 'species'] = 'white shark'
```

In [867]:
```
sydney_newcastle.loc[sydney_newcastle['species'].str.contains('bull') == True, 'species'] = 'bull shark'
```

In [868]:
```
sydney_newcastle.loc[sydney_newcastle['species'].str.contains('tiger') == True, 'species'] = 'tiger shark'
```

In [869]:
```
sydney_newcastle.loc[sydney_newcastle['species'].str.contains('wobbegong') == True, 'species'] = 'wobbegong shark'
```

In [870]:
```
sydney_newcastle.loc[sydney_newcastle['species'].str.contains('blue pointer') == True, 'species'] = 'blue pointer shark'
```

In [871]:
```
sydney_newcastle.loc[sydney_newcastle['species'].str.contains('whaler') == True, 'species'] = 'bronze whaler shark'
```

In [872]:
```
sydney_newcastle.loc[sydney_newcastle['species'].str.contains('nurse') == True, 'species'] = 'grey nurse shark'
```

In [873]:
```
sydney_newcastle.loc[sydney_newcastle['species'].str.contains('mako') == True, 'species'] = 'mako'
```

In [874]:
```
sydney_newcastle['species'].value_counts().head(10)
```

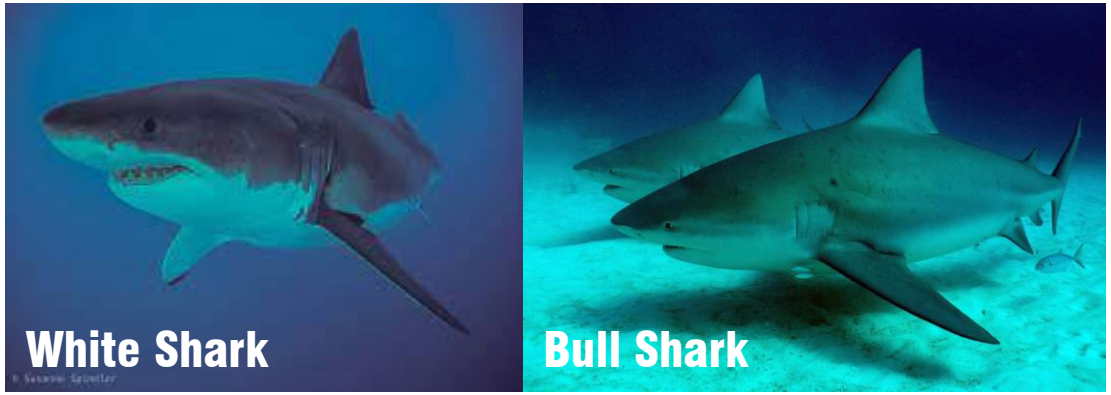
Out[874]:
```
shark involvement prior to death unconfirmed          10
white shark                                            9
bull shark                                             7
bronze whaler shark                                    6
wobbegong shark                                        6
grey nurse shark                                       5
invalid                                                5
12' shark                                              5
tiger shark                                            4
shark involvement prior to death was not confirmed     3
Name: species, dtype: int64
```

In [874]:
```
sydney_newcastle['species'].value_counts().head(10)
```

Out[874]:
```
shark involvement prior to death unconfirmed          10
white shark                                            9
bull shark                                             7
bronze whaler shark                                    6
wobbegong shark                                        6
grey nurse shark                                       5
invalid                                                5
12' shark                                              5
tiger shark                                            4
shark involvement prior to death was not confirmed     3
Name: species, dtype: int64
```

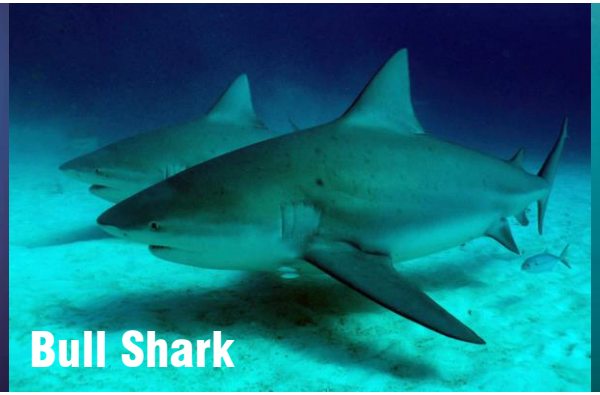
```
In [874]: sydney_newcastle['species'].value_counts().head(10)
```

```
Out[874]: shark involvement prior to death unconfirmed          10
          white shark                                            9
          bull shark                                             7
          bronze whaler shark                                    6
          wobbegong shark                                        6
          grey nurse shark                                       5
          invalid                                                5
          12' shark                                              5
          tiger shark                                            4
          shark involvement prior to death was not confirmed     3
          Name: species, dtype: int64
```



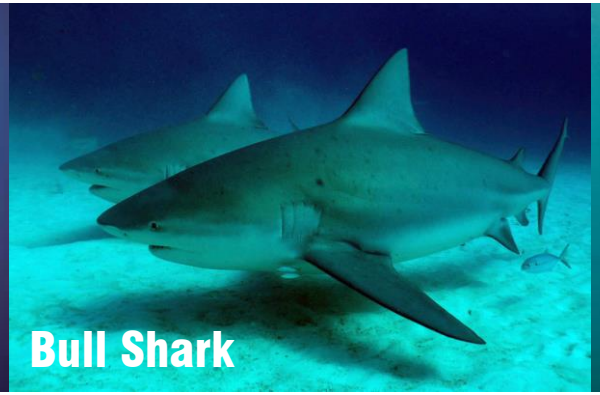
```
In [874]: sydney_newcastle['species'].value_counts().head(10)
```

```
Out[874]: shark involvement prior to death unconfirmed          10
          white shark                                            9
          bull shark                                             7
          bronze whaler shark                                    6
          wobbegong shark                                        6
          grey nurse shark                                       5
          invalid                                                5
          12' shark                                              5
          tiger shark                                            4
          shark involvement prior to death was not confirmed     3
          Name: species, dtype: int64
```



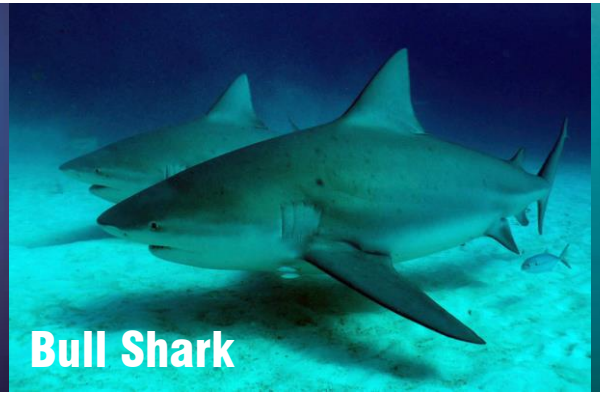
```
In [874]: sydney_newcastle['species'].value_counts().head(10)
```

```
Out[874]: shark involvement prior to death unconfirmed          10
          white shark                                            9
          bull shark                                             7
          bronze whaler shark                                    6
          wobbegong shark                                        6
          grey nurse shark                                       5
          invalid                                                5
          12' shark                                              5
          tiger shark                                            4
          shark involvement prior to death was not confirmed     3
          Name: species, dtype: int64
```



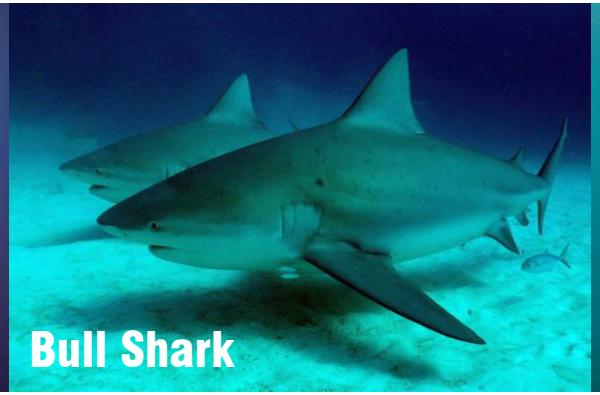
```
In [874]: sydney_newcastle['species'].value_counts().head(10)
```

```
Out[874]: shark involvement prior to death unconfirmed          10
          white shark                                            9
          bull shark                                             7
          bronze whaler shark                                    6
          wobbegong shark                                        6
          grey nurse shark                                       5
          invalid                                                5
          12' shark                                              5
          tiger shark                                            4
          shark involvement prior to death was not confirmed     3
          Name: species, dtype: int64
```



In [874]:
```
sydney_newcastle['species'].value_counts().head(10)
```

Out[874]:
```
shark involvement prior to death unconfirmed          10
white shark                                            9
bull shark                                             7
bronze whaler shark                                    6
wobbegong shark                                        6
grey nurse shark                                       5
invalid                                                5
12' shark                                              5
tiger shark                                            4
shark involvement prior to death was not confirmed     3
Name: species, dtype: int64
```



Obrigado ;)