

SHARKNADO

AGATA
SAMSUDEEN
SEBASTIAN
SHACHAM

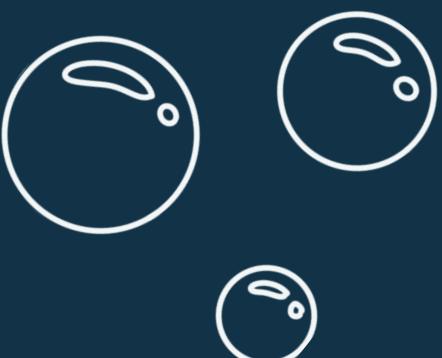
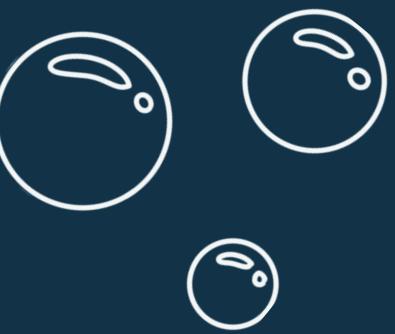


SUICIDE BY SHARK

— Where? —

what
are the
odds?

How?



But first... let's clean!

Problem:
too many nan!



Let's remove
some columns

```
# Looking at missing values before any operations
df.isna().sum()

Date           1
Year          3
Type         19
Country      51
State        482
Location     565
Activity     585
Name         220
Sex          580
Age         2987
Injury        36
Unnamed: 11   563
Time        3519
Species      3121
Source        20
pdf          108
href formula 112
href          111
Case Number   109
Case Number.1 110
original order 108
Unnamed: 21    6906
Unnamed: 22    6905
dtype: int64
```

```
] # Dropping columns of no use for us
df = df.drop(['Unnamed: 11', 'Source', 'pdf', 'href formula', 'href', 'Case Number', 'Case Number.1'],
             axis=1)
df.isna().sum()

Date           1
Year          3
Type         19
Country      51
State        482
Location     565
Activity     585
Name         220
Sex          580
Age         2987
Injury        36
Time        3519
Species      3121
dtype: int64
```

Much better!



Rows with no values

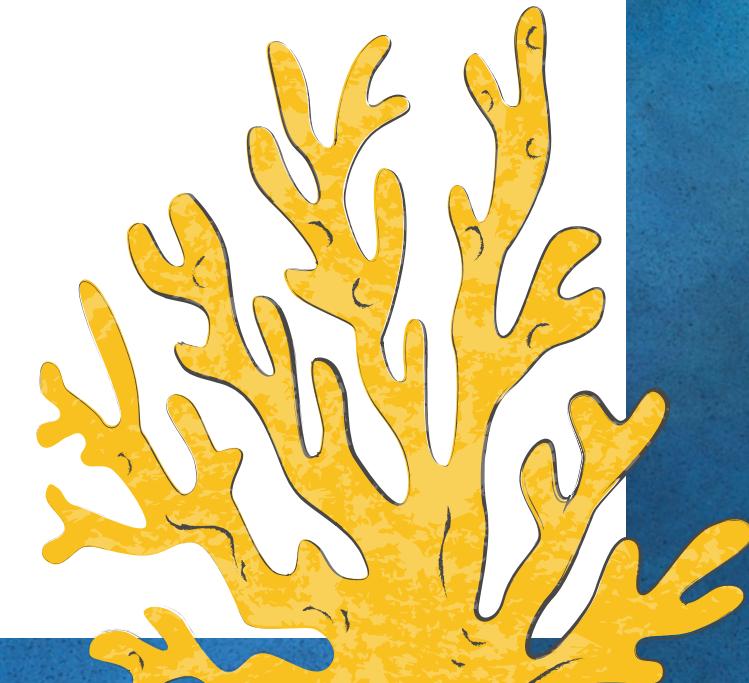


```
In [26]: df.style.applymap(lambda x: 'color: red' if pd.isna(x) else '')
```

formula	href	Case Number	Case Number.1	original order	Unnamed: 21	Unnamed: 22
nan	nan	nan	nan	nan	nan	nan
nan	nan	nan	nan	nan	nan	nan
nan	nan	nan	nan	nan	nan	nan
nan	nan	nan	nan	nan	nan	nan
nan	nan	nan	nan	nan	nan	nan
nan	nan	nan	nan	nan	nan	nan
...

```
# Dropping rows with 7 null values or more  
df=df.dropna(thresh=7)  
df.isna().sum()
```

Date	0
Year	2
Type	18
Country	44
State	460
Location	544
Activity	569
Name	203
Sex	561
Age	2962
Injury	30
Time	3494
Species	3097
	dtype: int64



Swim away!

No Doppelgängers!

```
# Removing duplicates  
df.drop_duplicates(inplace=True)
```

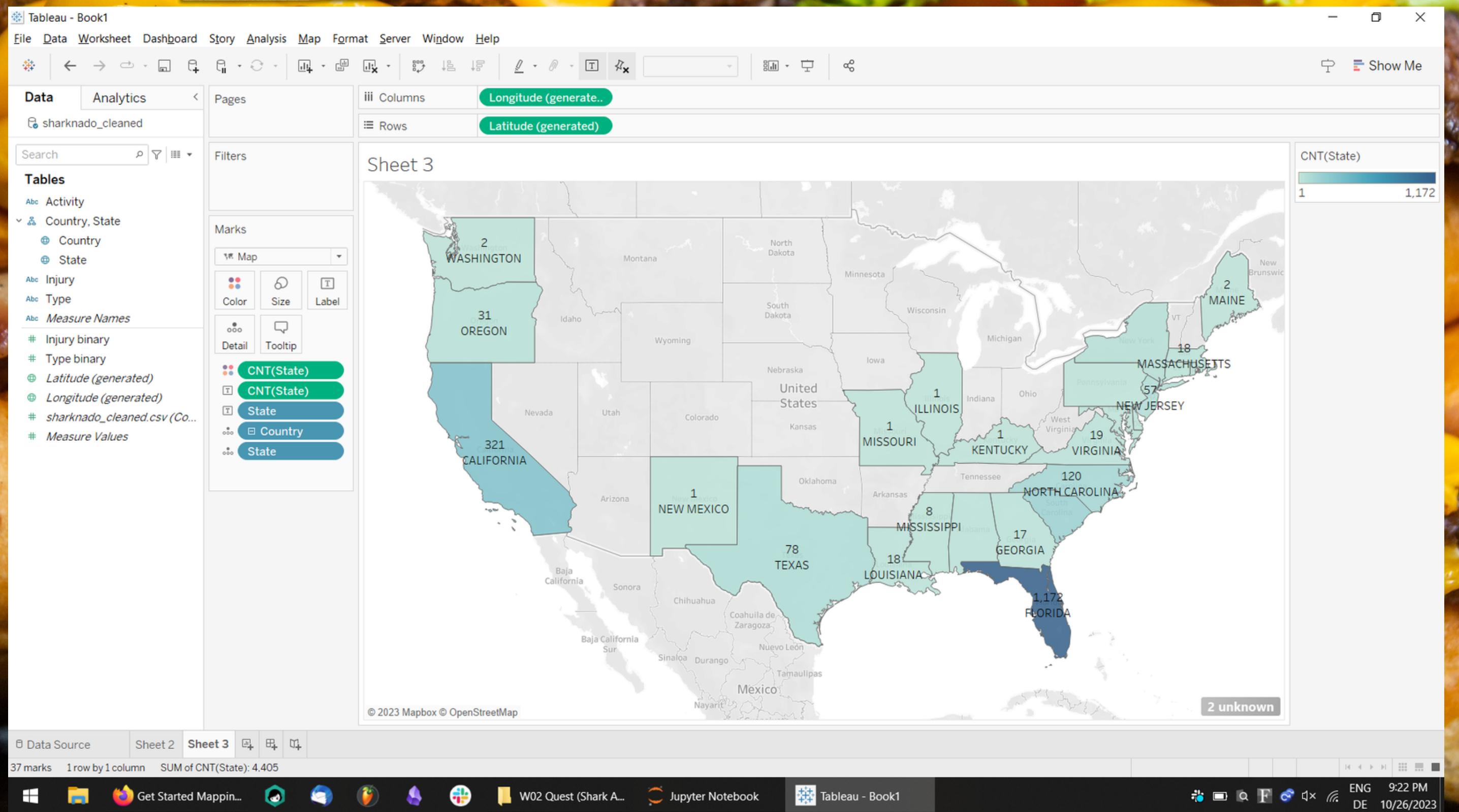
Stay in shape!

```
# Unifying capping (all uppercase) to further reduce unique values  
df = df.applymap(lambda x: x.upper() if isinstance(x, str) else x)  
df.head()
```

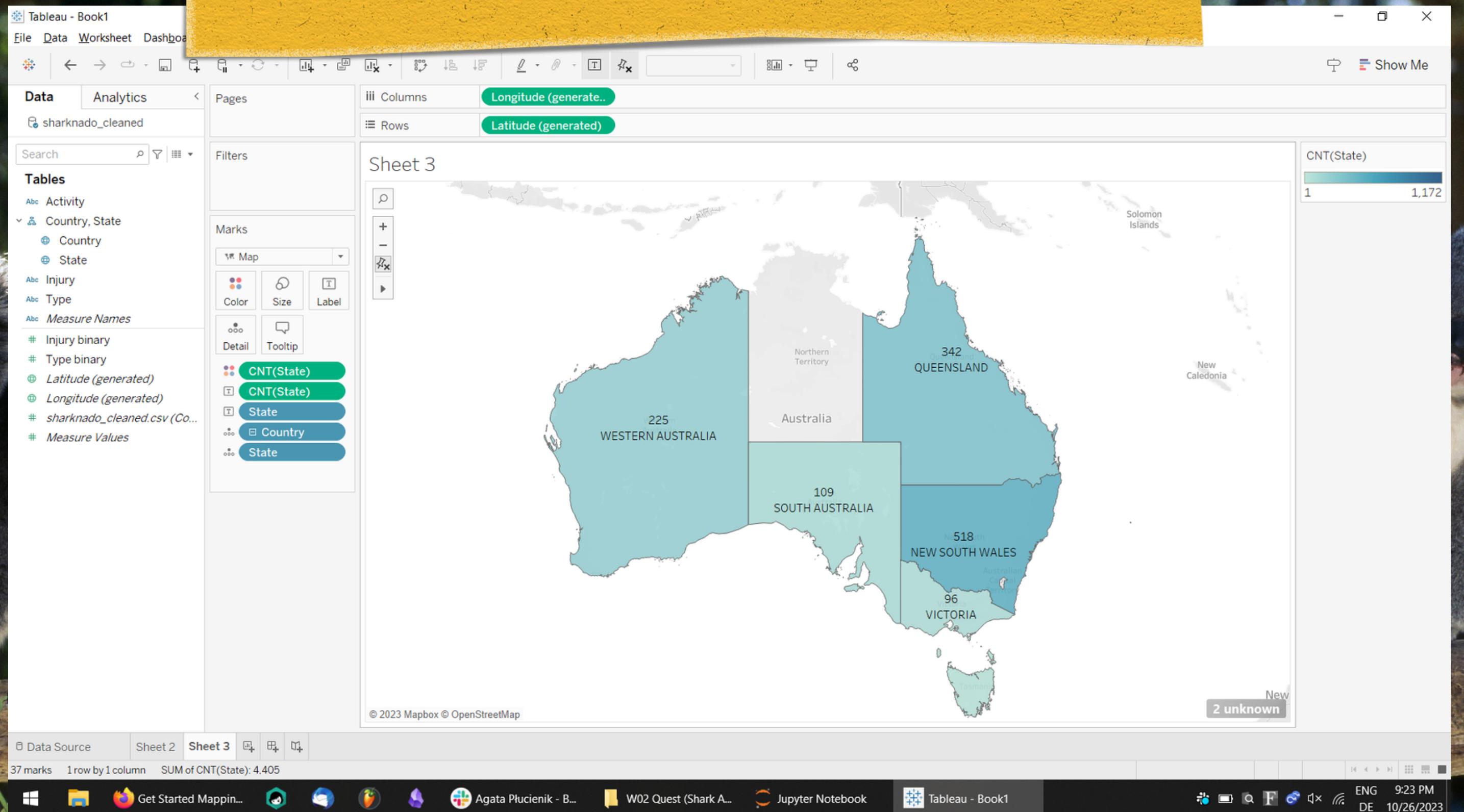


LOCATION

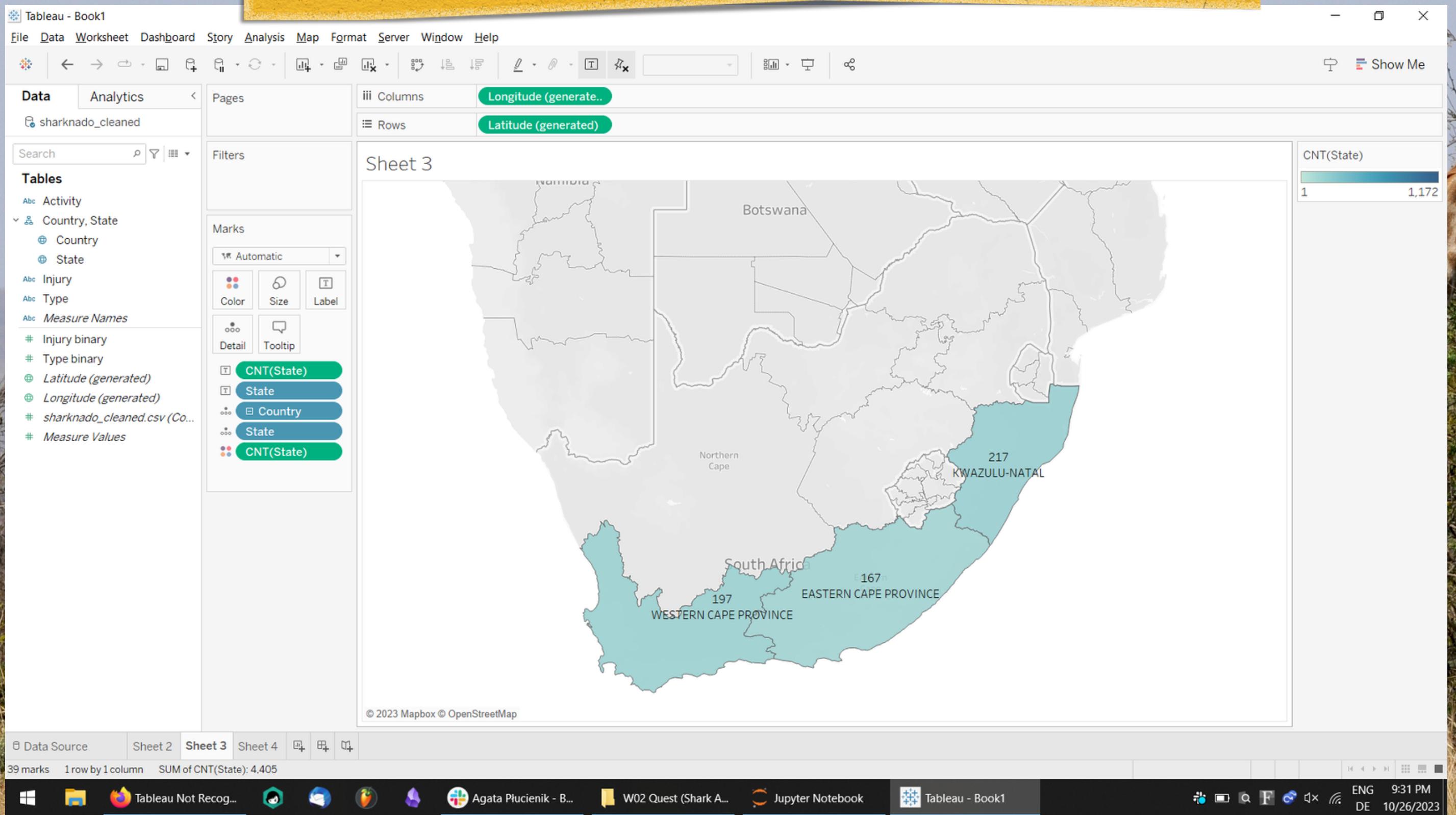




AUSTRALIA



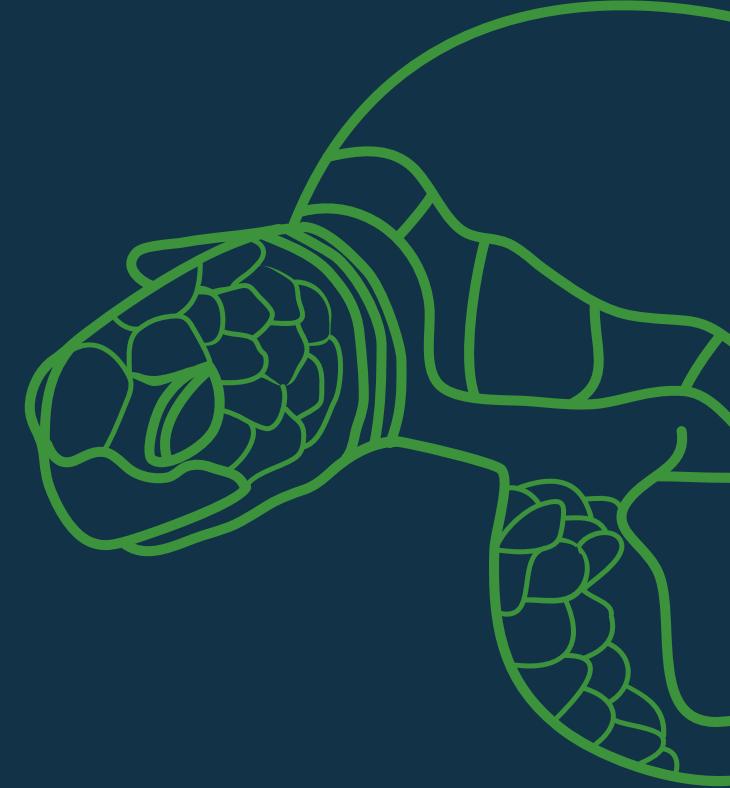
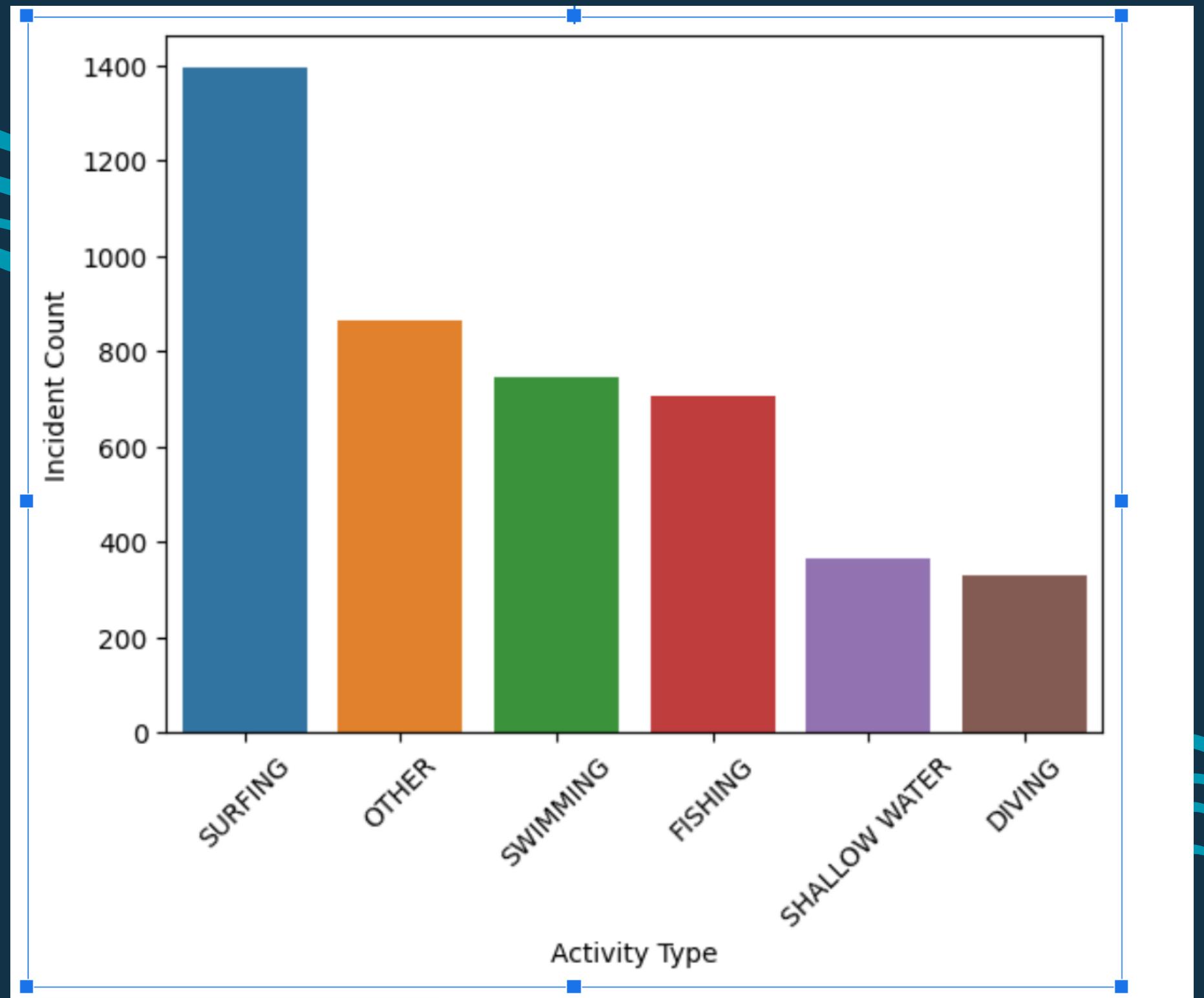
SOUTH AFRICA



ACTIVITY



.groupby()



COMMON INJURIES

Foot

Leg



Arm

Other:
fingers
toes
head



.NUNIQUE()

SUCCESS RATE

```
|df['Injury_binary'].value_counts(normalize=True) * 100
```

0	85.856981	'JUST' INJURY
1	14.143019	FATAL

Name: Injury_binary, dtype: float64



Best Wishes
& GOODLUCK