THE WORLD UNDER THE OCEAN

CORAL SPAWNING

Image: https://www.whitsundayescape.com/news/coral-spawning/

Today's Discussion

Introduction

- Background Story
- Problem / Goal

Process

Result / Conclusion

What is coral spawning?

Image: https://www.diversden.com.au/travel-guide/great-barrier-reef-coral-spawning-guide/

Timing!

Image: https://climatekids.nasa.gov/career-ocean-scientist/

Dataset

Column	Non-Null Count	Dtype
Obs_ID	6178	int64
Ecoregion	6178	object
Country	6178	object
Site	6178	object
Subsite	6178	object
Latitude	6178	float64
Longitude	6178	float64
Subsite_ID	6178	int64
Genus	6178	object
O_n	261	object
Species	6178	object
Taxon	6178	object
Coral_ID	6178	int64
N	6177	float64
N_min	6177	float64
Date	6178	datetime64[ns]
Start_decimal	4484	float64
STrSS	4396	float64
STrSR	88	float64
No_start	6178	bool
Quality_start	4484	object
End_decimal	2842	float64
ETrSS	2771	float64
ETrSR	70	float64
No_end	6178	bool
Quality_end	2839	object
Gamete_release	6178	object
Situation	6178	object
DoNFM	6178	datetime64[ns]
DoSRtNFM	6178	int64
ToSS_decimal	6178	float64
ToSR_decimal	6178	float64
Timezone	6178	float64
Reference	6178	object
Ref_ID	6178	int64
Comments	479	object

Goal

>> Predict coral spawning time

Process

Where we go from here

PROJECT PLANNING

Tool - Trello

DATA SCIENCE CYCLE

Understand problem
Goal setting (>> then plan the project)
Data collection/Literature review
Data cleaning/EDA/processing
Build models/improve model
Data collection II
Check the outcome/go back/check the whole cycle

REPORT

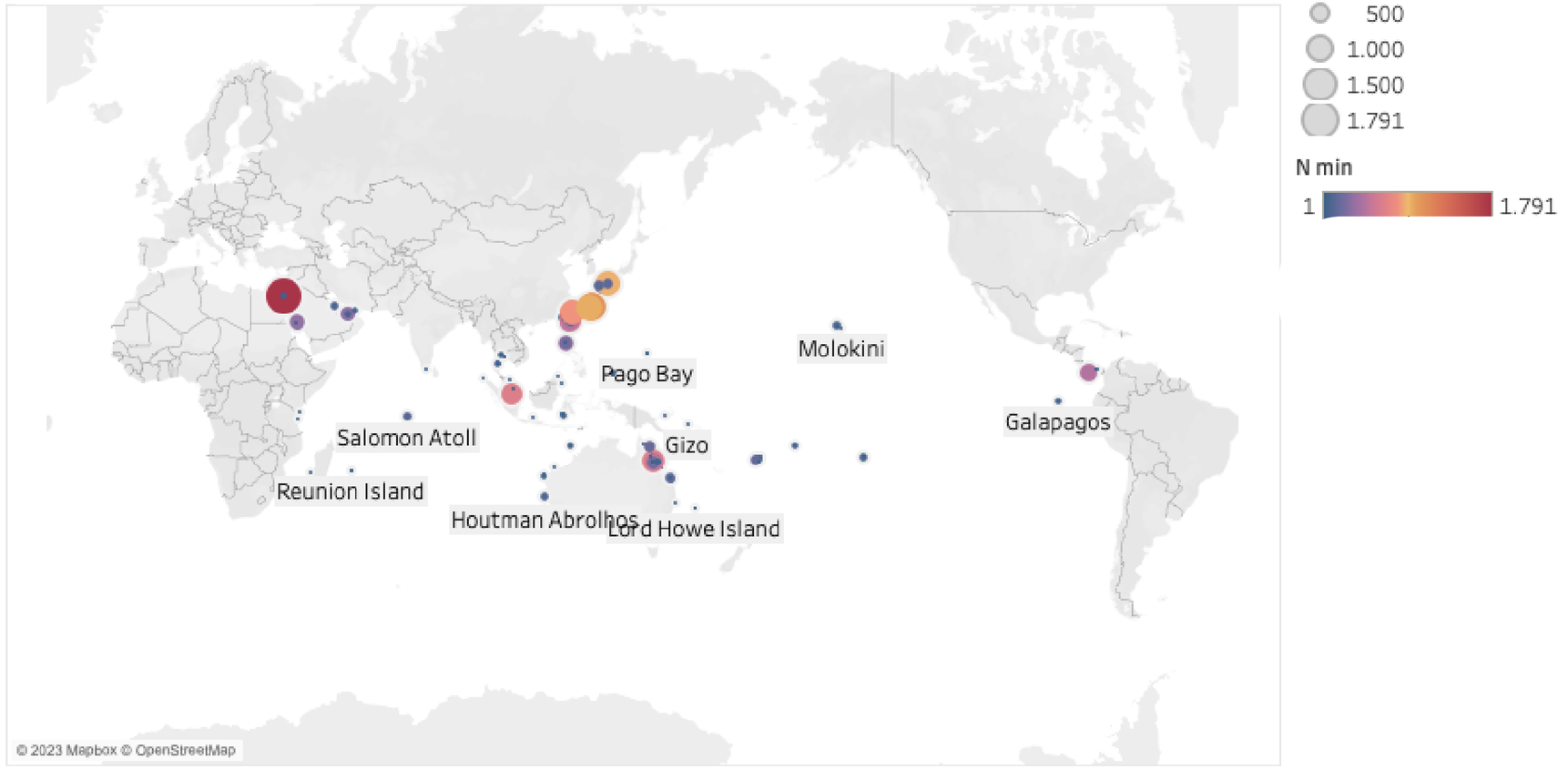
Writing report (Jupyter Notebook)
Presentation

SUMMARY & CLOSE THE PROJECT

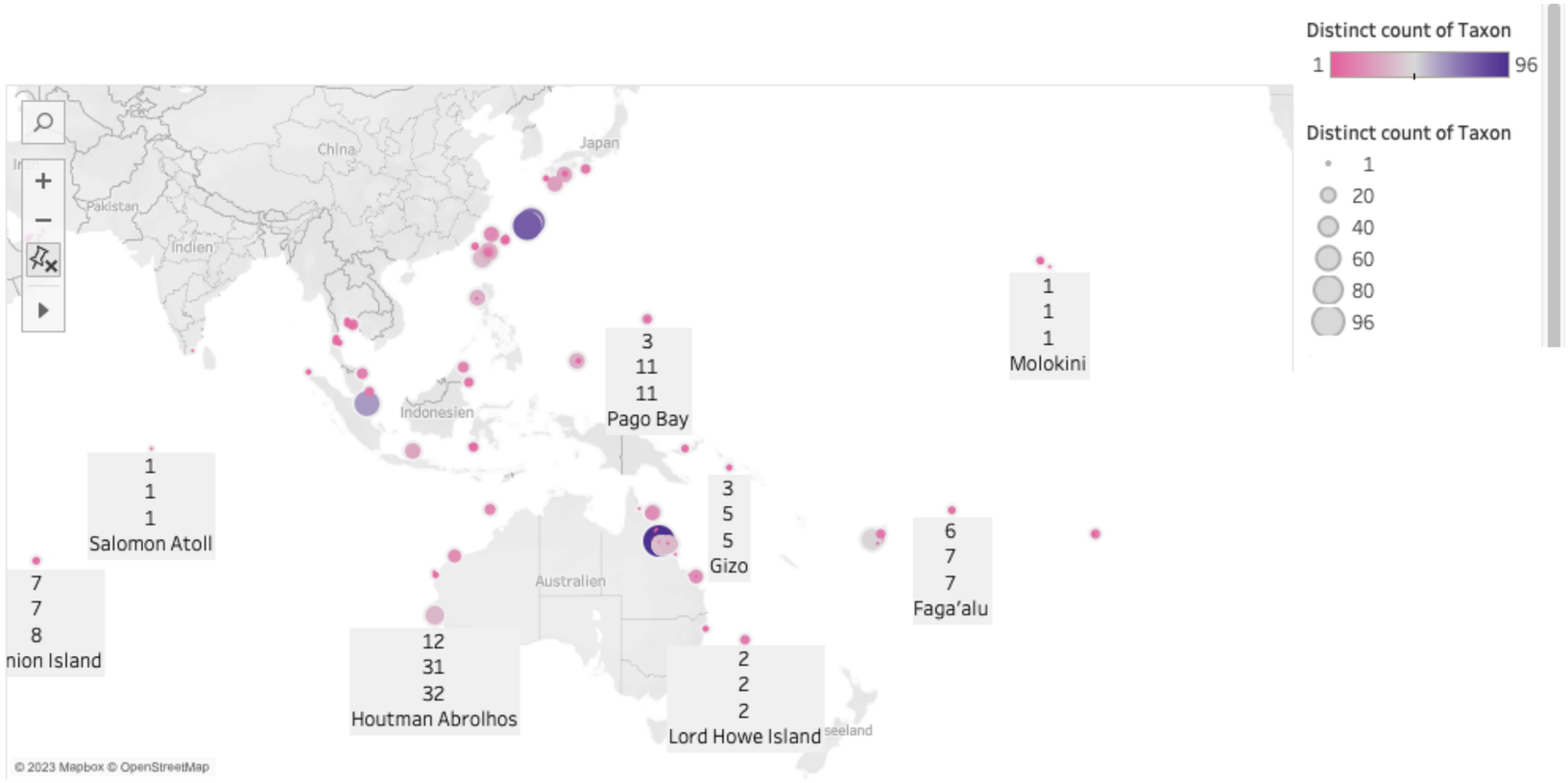
Delivery:
Presentation
Slides
Report

Data Exploration & Analysis

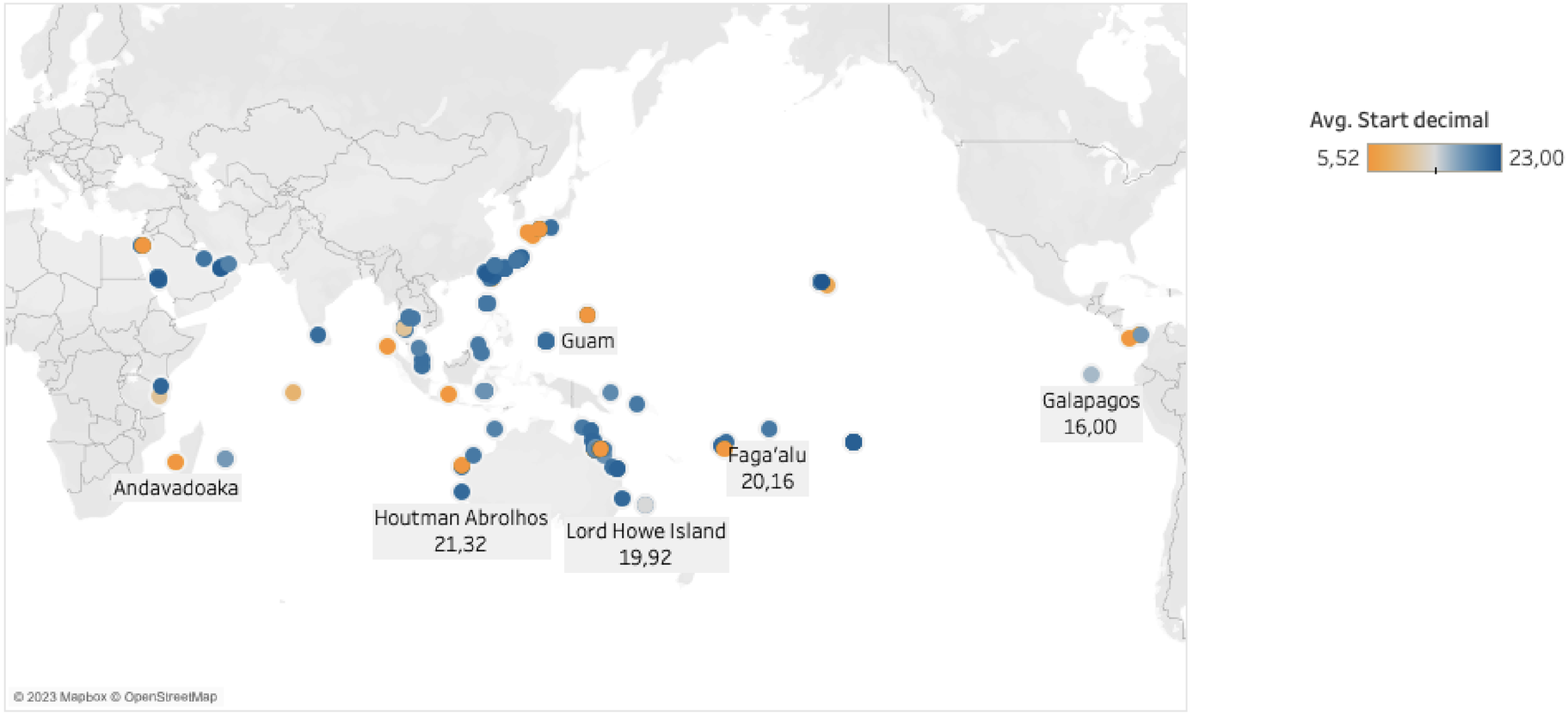
Where will you find coral colony?



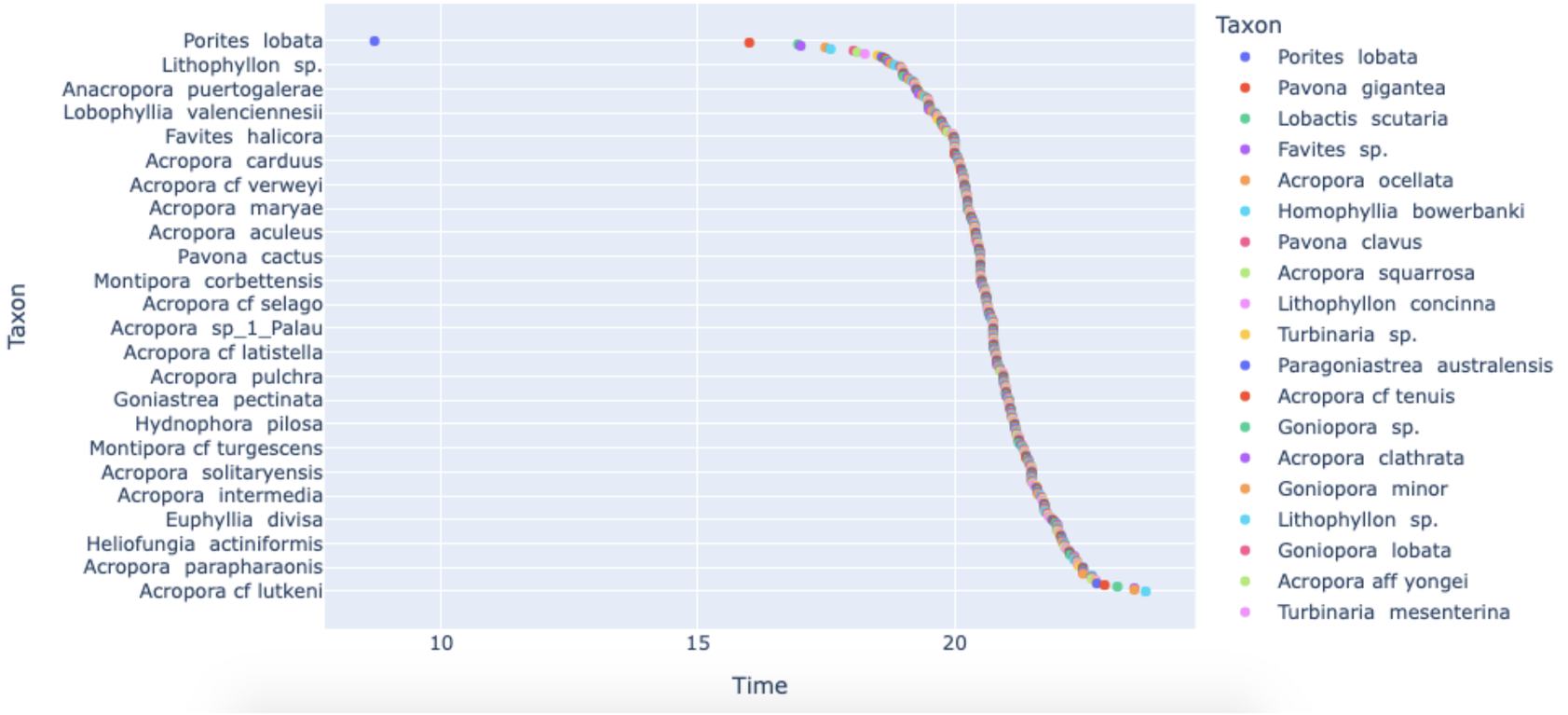
Which site has the most coral diversity?



Do corals in all sites spawn at the same time?

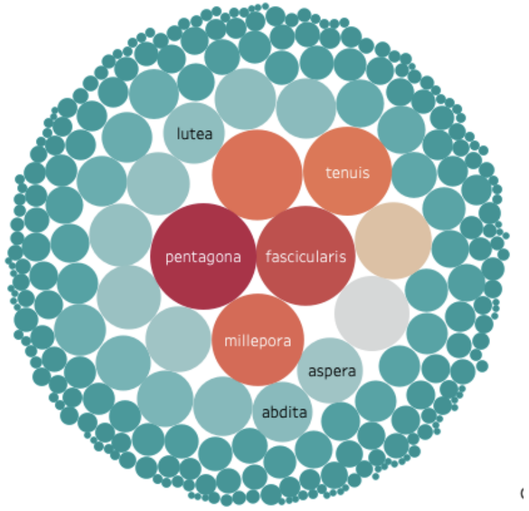


Do the different species spawn at different times?

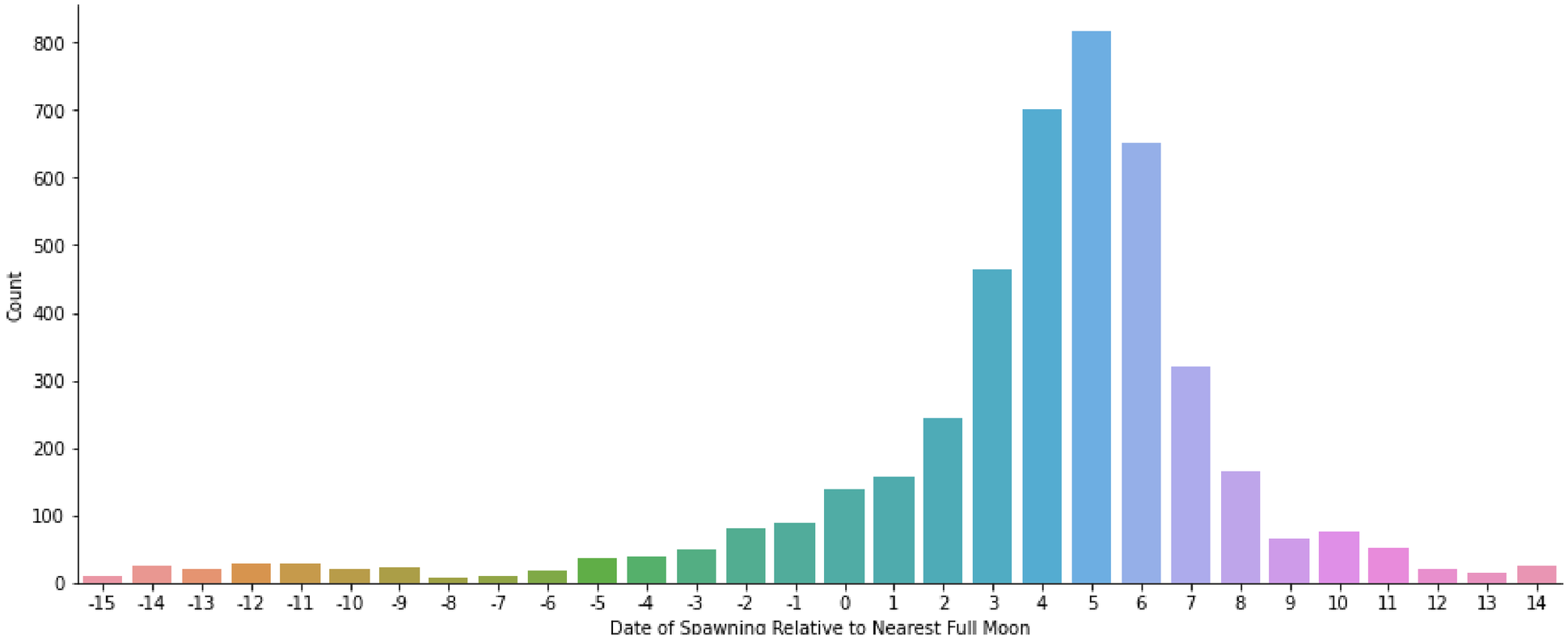


GenusSpeciesCount

SpeciesCount

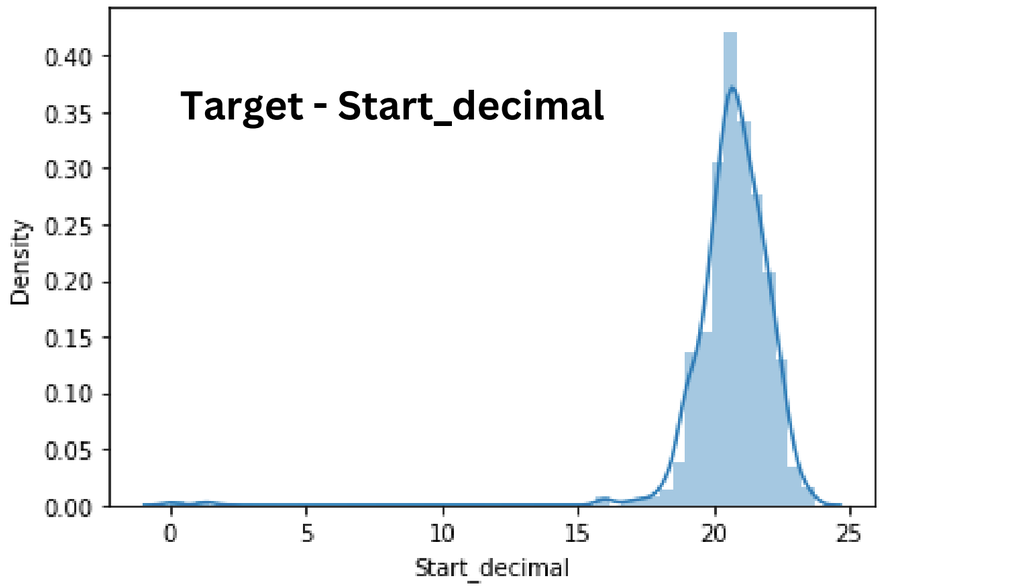


How many days away from the full moon date does coral spawn?



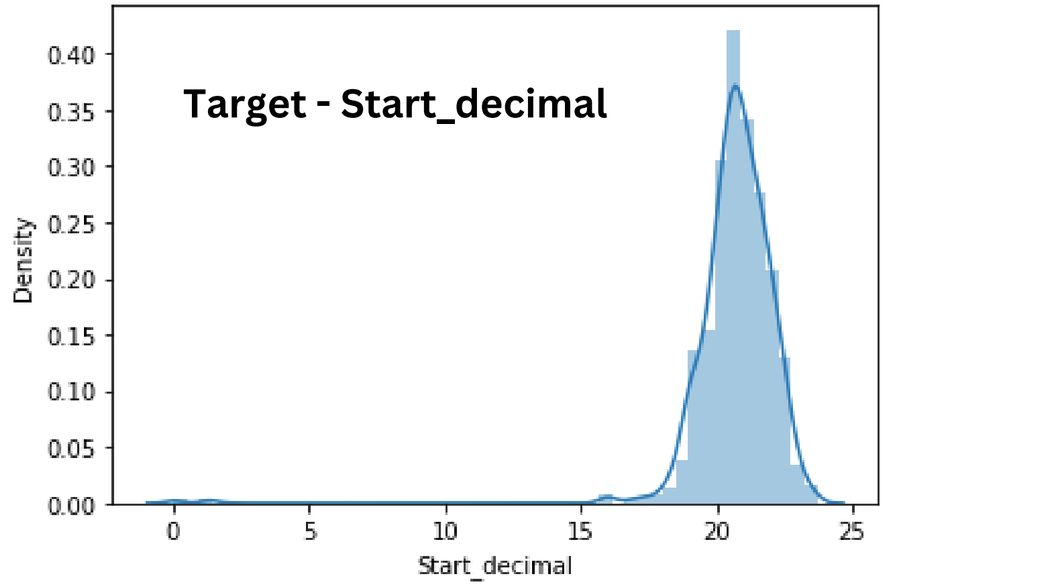
Building a baseline model

```
<AxesSubplot:xlabel='Start_decimal', ylabel='Density'>
```



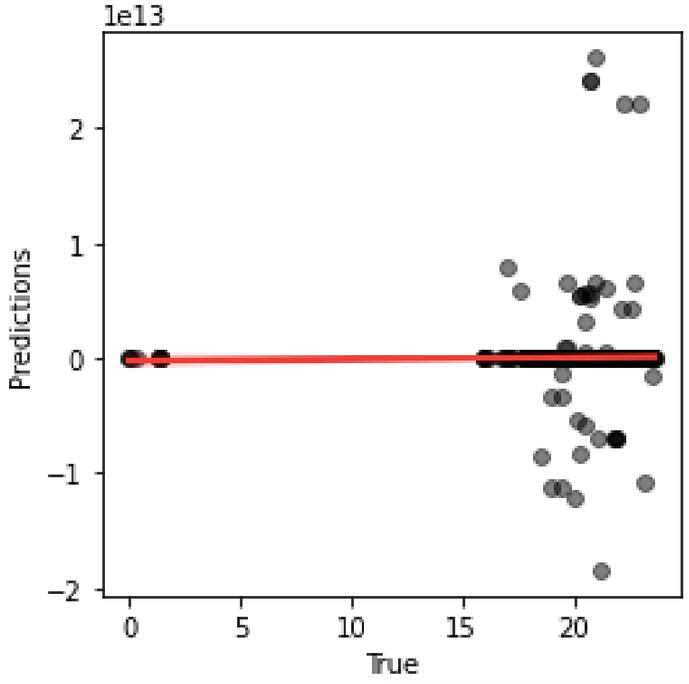
Building a baseline model

```
<AxesSubplot:xlabel='Start_decimal', ylabel='Density'>
```

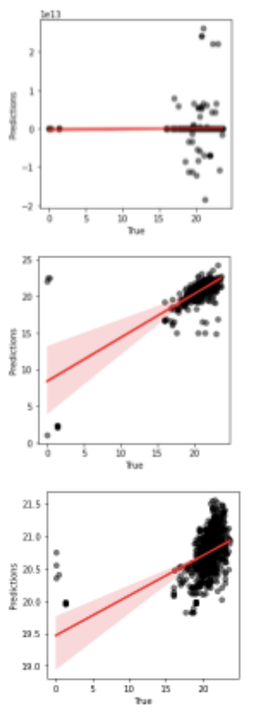



```
Baseline model (Linear Regression):

R^2:  -8.521397933708999e+23
MSE:  4.2196627240173417e+24
RMSE:  2054181765087.34
MAE:  306825665752.64
```



Different models



Linear Regression (baseline model):

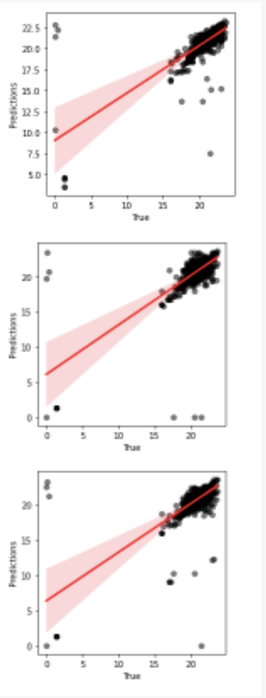
R^2: -8.521397933708999e+23
MSE: 4.2196627240173417e+24
RMSE: 2054181765087.34
MAE: 306825665752.64

SVR:

R^2: 0.56
MSE: 2.17
RMSE: 1.47
MAE: 0.72

Poisson Regressor:

R^2: 0.1
MSE: 4.46
RMSE: 2.11
MAE: 0.94



RandomForest Regressor:

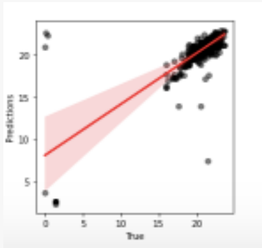
R^2: 0.56
MSE: 2.18
RMSE: 1.48
MAE: 0.53

DecisionTree Regressor:

R^2: 0.42
MSE: 2.88
RMSE: 1.7
MAE: 0.54

KNeighborsRegressor:

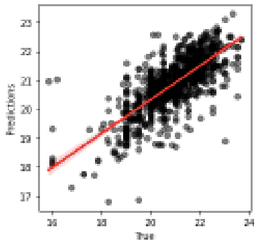
R^2: 0.44
MSE: 2.75
RMSE: 1.66
MAE: 0.53



XGBRegressor:

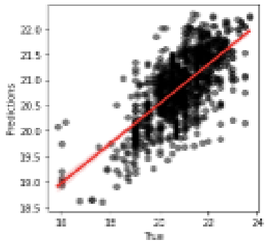
R^2: 0.6
MSE: 2.0
RMSE: 1.41
MAE: 0.56

Improved models



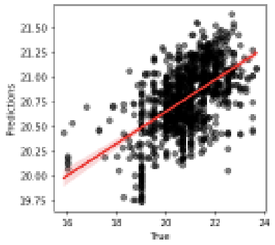
Linear Regression (no outliers):

R^2: 0.56
MSE: 0.6
RMSE: 0.78
MAE: 0.55



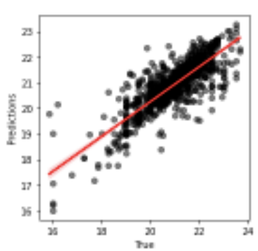
SVR (no outliers):

R^2: 0.46
MSE: 0.73
RMSE: 0.86
MAE: 0.67



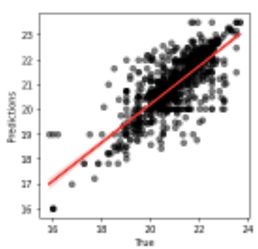
Poisson Regressor (no outliers):

R^2: 0.24
MSE: 1.04
RMSE: 1.02
MAE: 0.78



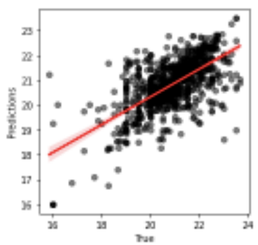
RandomForest Regressor (no outliers):

R^2: 0.7
MSE: 0.41
RMSE: 0.64
MAE: 0.41



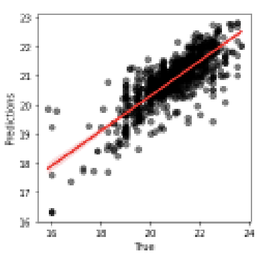
DecisionTree Regressor (no outliers):

R^2: 0.6
MSE: 0.55
RMSE: 0.74
MAE: 0.44



KNeighborsRegressor (no outliers):

R^2: 0.41
MSE: 0.81
RMSE: 0.9
MAE: 0.58



XGBRegressor (no outliers):

R^2: 0.66
MSE: 0.46
RMSE: 0.68
MAE: 0.47

Predict on Test Dataset

Taxon	N_min	Date	Start_decimal	STrSS	Gamete_release	DoSRtNFM	ToSS_decimal	ToSR_decimal	daylen_decimal	sunset	sunrise
Merulina ampliata	2.0	1995-11-12	NaN	NaN	Bundles	6	18.483333	5.733333	13.45	18.29	5.44
Acropora gemmifera	1.0	1982-12-06	NaN	NaN	Bundles	5	18.600000	5.383333	13.13	18.36	5.23
Acropora humilis	1.0	1982-12-06	NaN	NaN	Bundles	5	18.600000	5.383333	13.13	18.36	5.23

Final Outcome

	Ecoregion	Country	Site	Latitude	Longitude	Genus	Species	Taxon	N_min	Date	Start_decimal	STrSS	Gamete_release	Do
0	Tuvalu, Samoa and Tonga	American Samoa	Faga'alu	-14.279444	-170.700833	Merulina	ampliata	Merulina ampliata	2.0	1995-11-12	21.353000	NaN	Bundles	
1	Central and northern Great Barrier Reef	Australia	Big Broadhurst Reef	-18.929444	147.757222	Acropora	gemmifera	Acropora gemmifera	1.0	1982-12-06	18.510528	NaN	Bundles	
2	Central and northern Great Barrier Reef	Australia	Big Broadhurst Reef	-18.929444	147.757222	Acropora	humilis	Acropora humilis	1.0	1982-12-06	20.127976	NaN	Bundles	
3	Central and northern Great Barrier Reef	Australia	Big Broadhurst Reef	-18.929444	147.757222	Acropora	loripes	Acropora loripes	1.0	1982-12-06	21.440833	NaN	Bundles	
4	Central and northern Great Barrier Reef	Australia	Big Broadhurst Reef	-18.929444	147.757222	Acropora	nasuta	Acropora nasuta	1.0	1982-12-06	20.893500	NaN	Bundles	


HORNIMAN
MUSEUM
& GARDENS

Conclusion

- What I have learned
- Challenges
- Personal story

"**12% of the land** around the world is now protected but only **1% of the ocean** is fully protected"
- ***Sylvia Earle,*** *TED Prize 2009*

Portrait of Sylvia Earle, oceanographer and National Geographic Explorer at Large, in 2017.
MARK THIESSEN/NATIONAL GEOGRAPHIC

Thank you!

Image: https://hawaiianpaddlesports.com/maui/turtle-town/