

## Data Appendix

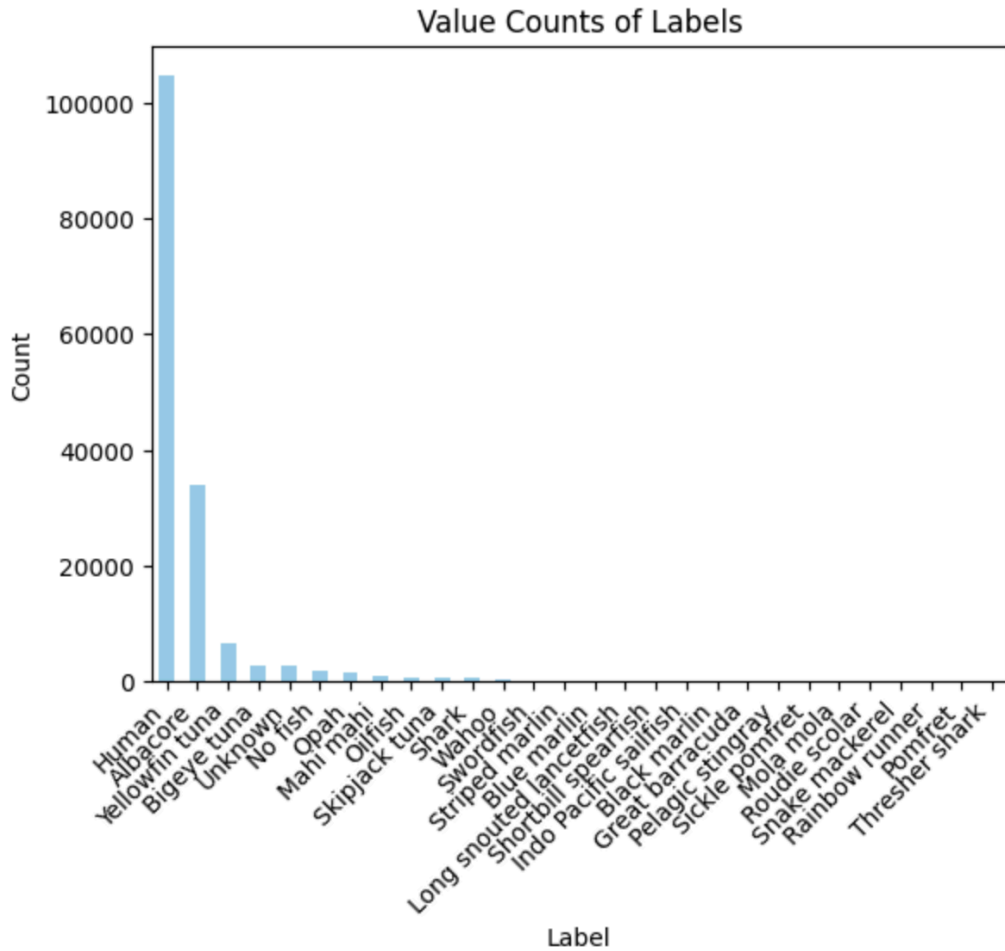
Our final dataset, `cleaned_foid.csv`, reflects the data cleaning from our original datasets, `foid_labels_bbox_v012.csv` and `foid_labels_bbox_v012_freq.csv`. The final dataset includes 12 columns with each row representing a single image and its corresponding data. The following variables from our final dataset were used in our analysis script.

### Variables Used in `cleaned_foid.csv`:

- **img\_id**
  - *Definition:* The unique identifier for each image (e.g., “94a69b66-23f0-11e9-a78e-2f2b7983ac0d”). Each image jpeg is named with this corresponding ID.
  - *Missing values:* 0
- **label**
  - *Definition:* This labeling class contains the full label for the object within the bounding box, including the 4 types of tuna as well as 25 additional species names. (e.g., “No fish”, “Yellowfin Tuna”, “Human”).
  - *Missing values:* 0
  - *Frequency table:*

	Value	Frequency
0	Human	104503
1	Albacore	33943
2	Yellowfin tuna	6693
3	Bigeye tuna	2772
4	Unknown	2702
5	No fish	1879
6	Opah	1606
7	Mahi mahi	1112
8	Oilfish	778
9	Skipjack tuna	693
10	Shark	623
11	Wahoo	468
12	Swordfish	247
13	Striped marlin	224
14	Blue marlin	177
15	Long snouted lancetfish	176
16	Shortbill spearfish	142
17	Indo Pacific sailfish	129
18	Black marlin	80
19	Great barracuda	76
20	Pelagic stingray	40
21	Sickle pomfret	39
22	Mola mola	6
23	Roudie scolar	4
24	Snake mackerel	2
25	Rainbow runner	2
26	Pomfret	2
27	Thresher shark	1

- *Frequency distribution:*

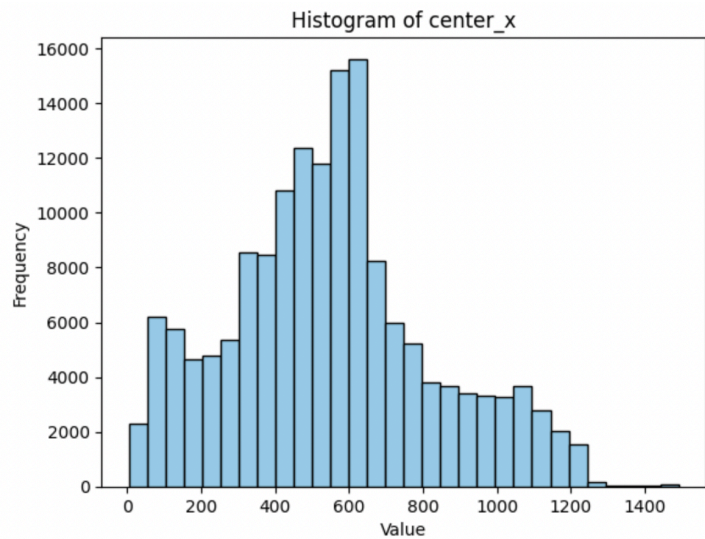


- **center\_x**

- *Definition:* The center point of the x coordinates of the bounding box.
- *Missing values:* 0
- *Summary statistics:*

count	159119.000000
mean	550.357877
std	276.421260
min	4.500000
25%	366.000000
50%	544.000000
75%	693.500000
max	1491.500000
Name: center_x, dtype: float64	

- *Histogram:*

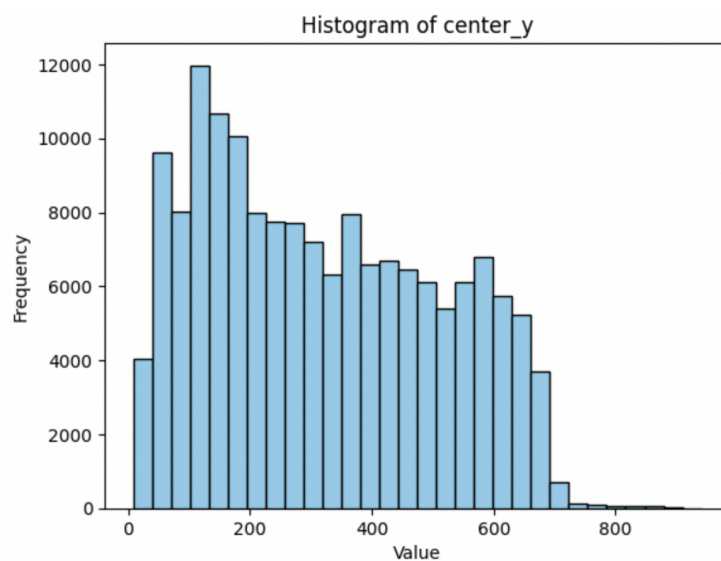


- **center\_y**

- *Definition:* The center point of the y coordinates of the bounding box.
- *Missing values:* 0
- *Summary statistics:*

```
count    159119.000000
mean      318.741439
std       190.772199
min         8.000000
25%      150.500000
50%      294.500000
75%      476.500000
max      941.000000
Name: center_y, dtype: float64
```

- *Histogram:*

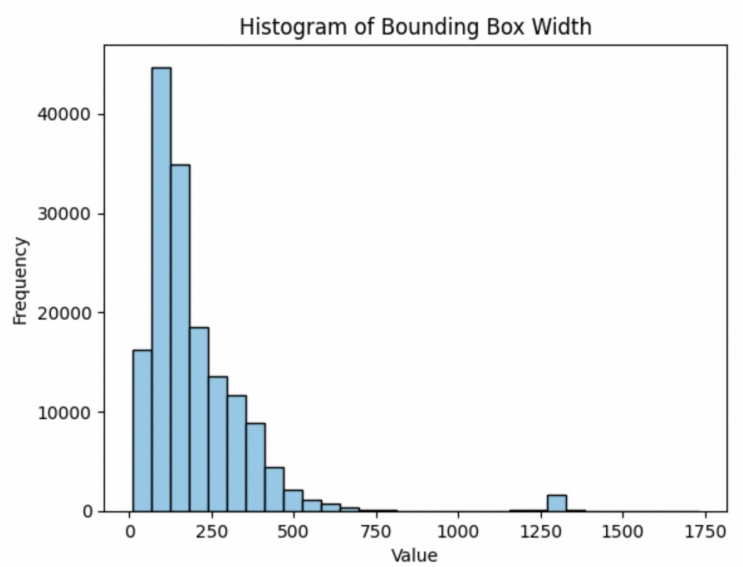


- **width**

- *Definition:* The width of the bounding box.
- *Missing values:* 0
- *Summary statistics:*

count	159119.000000
mean	199.177169
std	168.010889
min	9.000000
25%	98.000000
50%	150.000000
75%	259.000000
max	1731.000000
Name: width, dtype: float64	

- *Histogram:*



- **height**

- *Definition:* The height of the bounding box.
- *Missing values:* 0
- *Summary statistics:*

count	159119.000000
mean	184.059679
std	119.887359
min	6.000000
25%	109.000000
50%	151.000000
75%	220.000000
max	973.000000
Name: height, dtype: float64	

○ *Histogram:*

