final-Copy1

September 26, 2020

1 Bird Collision vs Artificial Light

The below project is a part of Tiger Machine Learning Engineer position. Winger et al, 2019 examined nocturnal flight-calling behavior and vulnerability to artificial light in migratory birds.

> "Understanding interactions between biota and the built environment is increasingly important as human modification of the landscape expands in extent and intensity. For migratory birds, collisions with lighted structures are a major cause of mortality, but the mechanisms behind these collisions are poorly understood. Using 40 years of collision records of passerine birds, we investigated the importance of species' behavioural ecologies in predicting rates of building collisions during nocturnal migration through Chicago, IL and Cleveland, OH, USA. "

> "One of the few means to examine species-specific dynamics of social biology during nocturnal bird migration is through the study of short vocalizations made in flight by migrating birds. Many species of birds, especially passerines (order Passeriformes), produce such vocal signals during their nocturnal migrations. These calls (hereafter, 'flight calls') are hypothesized to function as important social cues for migrating birds that may aid in orientation, navigation and other decision-making behaviours.not all nocturnally migratory species make flight calls, raising the possibility that different lineages of migratory birds vary in the degree to which social cues and collective decisions are important for accomplishing migration. "

As per the researcher the bird collision with buildings in the metropolitan cities are mainly caused by artificial light. Birds react to this artificial light and initiate a flight call which attracts other birds for flight and results in collision with the building structure.

1.1 Citations

When using this data, please cite the original publication:

> Winger BM, Weeks BC, Farnsworth A, Jones AW, Hennen M, Willard DE (2019) Nocturnal flight-calling behaviour predicts vulnerability to artificial light in migratory birds. Proceedings of the Royal Society B 286(1900): 20190364. https://doi.org/10.1098/rspb.2019.0364

If using the data alone, please cite the Dryad data package:

Winger BM, Weeks BC, Farnsworth A, Jones AW, Hennen M, Willard DE (2019) Data from: Nocturnal flight-calling behaviour predicts vulnerability to artificial light in migratory birds. Dryad Digital Repository. https://doi.org/10.5061/dryad.8rr0498

1.2 Libraries

Below are the libraries used as a part of this project.

- pandas
- numpy
- matplotlib
- seaborn
- logging
- os
- datetime
- zipfile
- json
- subprocess
- sys

1.3 Data Cleaning and Preprocessing

Below are the following steps used to clean and preprocess the data.

1.3.1 1. UnZipping

The data files are extracted from zip, they are placed in *./Data/Input/* directory and the originial zip file is deleted.

1.3.2 2. JSON to Data Frame

The files received are in JSON format. These files have to be processed and converted into data frame for analysis purposes. I have created *my_func.json_df* that take path of json files from config file and returns a data frame

1.3.3 3. Initial Inspection

```
In [6]: flight_call.head()
```

Out[6]:

	Species	Family	Collisions	Flight Call	Habitat	Stratum
0	Zonotrichia albicollis	Passerellidae	10133	Yes	Forest	Lower
1	Junco hyemalis	Passerellidae	6303	Yes	Edge	Lower
2	Melospiza melodia	Passerellidae	5124	Yes	Edge	Lower
3	Melospiza georgiana	Passerellidae	4910	Yes	Open	Lower
4	Seiurus aurocapilla	Parulidae	4580	Yes	Forest	Lower

Cleaning and Mapping Columns We see that flight_call has its column order wrong. I have used a dictionary to map the columns correctly. I also stripped the white space in columns which helps in standardizing column names.

```
In [8]: flight_call.head()
```

```
Out[8]:          Genus      Species         Family  Collisions Flight Call Habitat  \
        0  Zonotrichia   albicollis  Passerellidae       10133         Yes  Forest
        1        Junco     hyemalis  Passerellidae        6303         Yes    Edge
        2    Melospiza      melodia  Passerellidae        5124         Yes    Edge
        3    Melospiza    georgiana  Passerellidae        4910         Yes    Open
        4      Seiurus  aurocapilla      Parulidae        4580         Yes  Forest

          Stratum
        0   Lower
        1   Lower
        2   Lower
        3   Lower
        4   Lower
```

```
In [10]: collision.head()
```

```
Out[10]:          Genus  Species              Date Locality
         0  Ammodramus  nelsoni  10/03/1982 00:00       MP
         1  Ammodramus  nelsoni  05/21/1984 00:00      CHI
         2  Ammodramus  nelsoni  05/25/1984 00:00       MP
         3  Ammodramus  nelsoni  10/08/1985 00:00       MP
         4  Ammodramus  nelsoni  09/10/1986 00:00       MP
```

```
In [14]: light_level.head()
```

```
Out[14]:          Date  Light Score
         0  2010-09-26          4.0
         1  2015-05-07          6.0
         2  2012-03-07          NaN
         3  2000-09-18         15.0
         4  2000-11-10         15.0
```

Capitalizing We find that there are some entries that are Capitalized and lowercase for same category. For example, there is zonotrichia and Zonotrichia. There we can Capitalize the first alphabet of every non-numeric to standardize the data and reduce data errors.

Dropping duplicates As a best practice, it is always recommended to drop duplicate records if, it has any unique or key values. In our case, flight_call and light_level have data column as unique value. The function data_process.drop_dup take in data frame and drop duplicate records.

Interpolating As a best practice, it is always recommended to interpolate missing values where ever deemed necessary, but with extreme caution. In our case, light_level has missing records. We can use a simple linear interpolate method because the light_source levels, on an average don't change drastically from the previous few days.

```
In [24]: flight_call['Flight Call'].unique()
```

```
Out[24]: array(['Yes', 'No', 'Rare'], dtype=object)
```

It is observed that flight_call column of flight_call data has an extra Rare factor. Since it can only have Yes/No, we can assume that Rare is Yes.

1.4 Merging the Data Frames

After doing the data preprocessing and clean, we obtain clean files that we can merge. my_func.file_merge takes in 3 data frames and output 1 final data frame on which we can do our analysis.

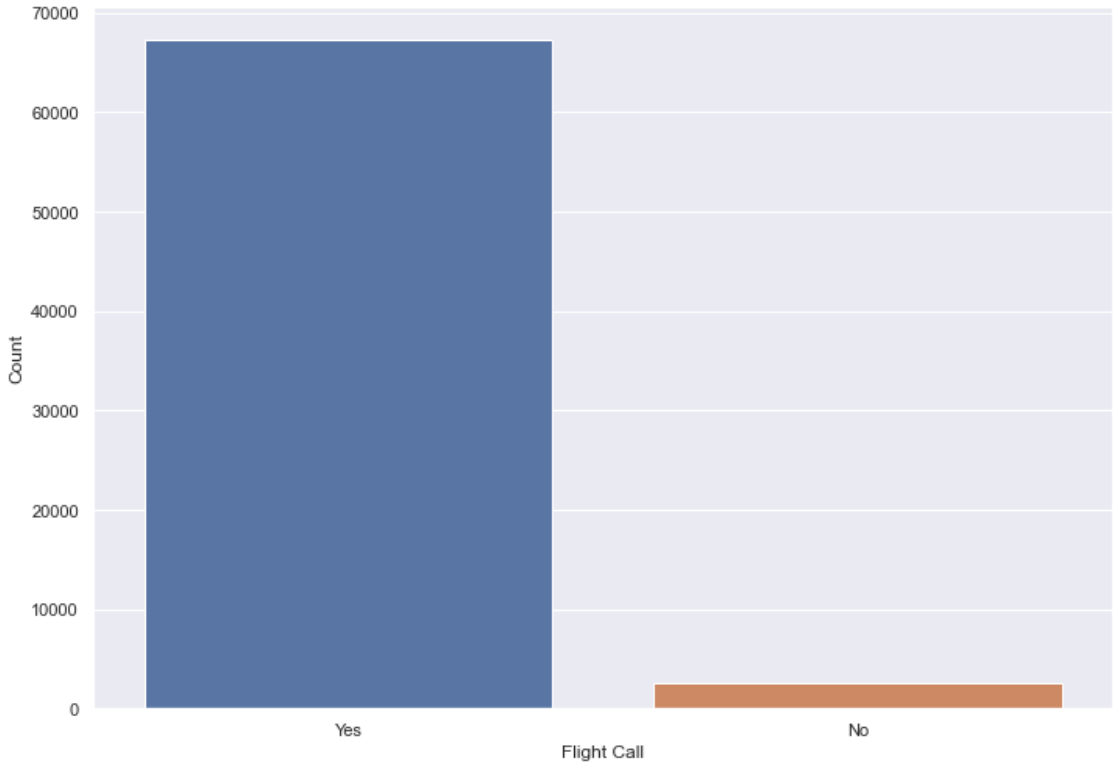
1.5 Generating Summary File and Plots

The final step is generate the result and plots. my_func.summary_stats.summarize generates the summary file as a csv because it is very easy to interpret and do custom analysis on csv. my_func.summary_stats.count_plot generate bar plot of different features like family, genus, Locality etc.

1.6 Insights

Collision by Flight Call of Birds We can see that birds that employ flight calls have significantly (almost 35000 times) more collision than the birds that don't employ flight

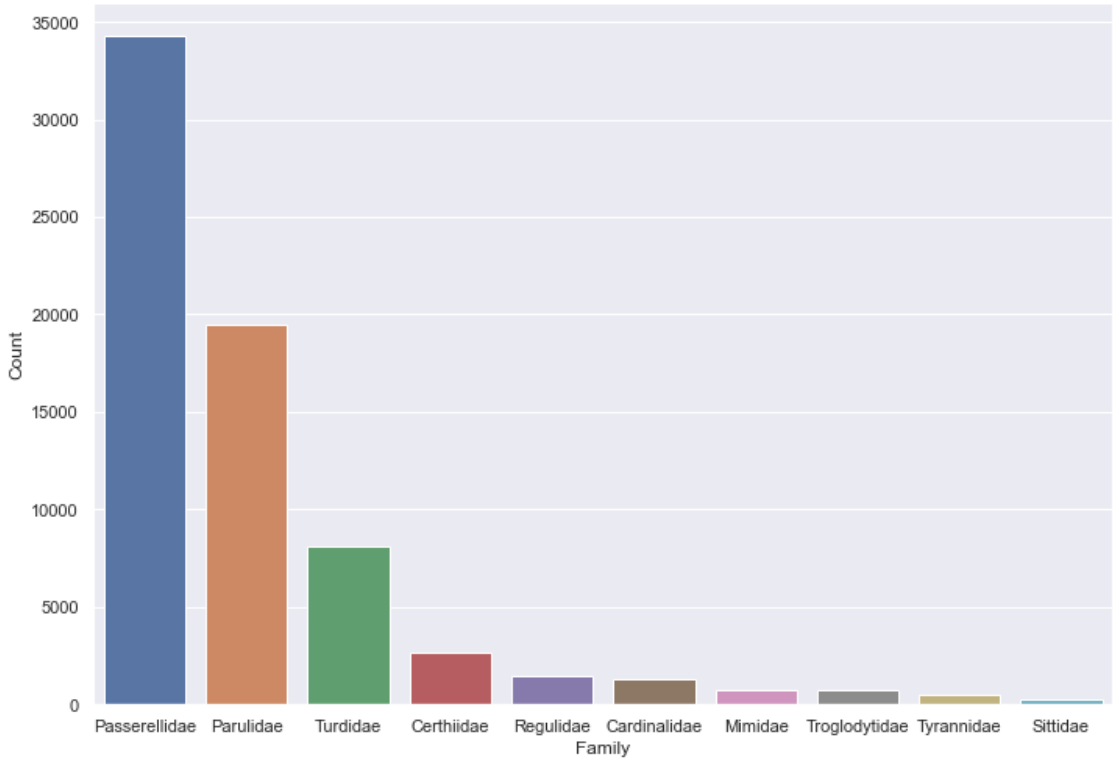
```
In [33]: br_plt('Flight Call')
```



We can see here that the flight call is a significant factor but we may need futher testing to be sure.

Collision by Family of Birds We can see that Passerellidae Family of birds have the highest collisions followed by Parulidae and Turdidae.

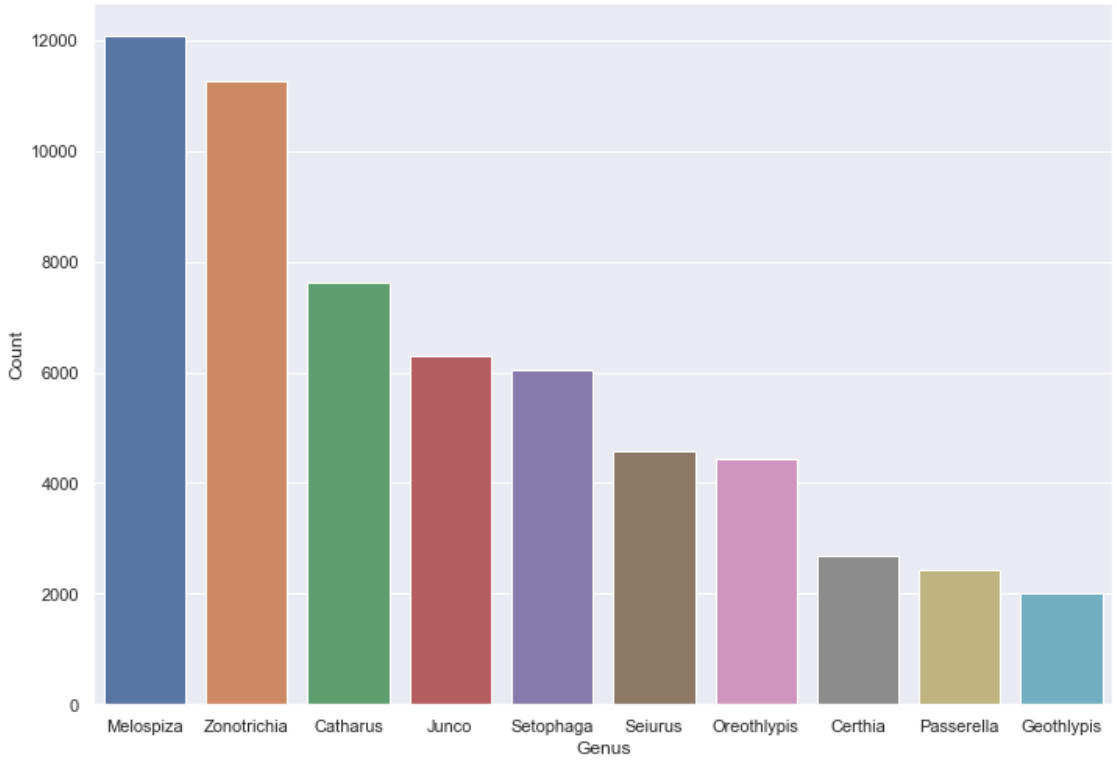
```
In [39]: br_plt('Family')
```



Here we observe that Passerellidae is most common family that have collisions. But its is highly possible that Passerellidae may be majority in the observed cities. If we an estimate of percentage distribution of birds by family we can take a weighted proportion and then check for most common family of birds.

Collision by Genus of Birds We can see that Melospiza Genus of birds have the highest collisions followed by Zonotrichia and Catharus.

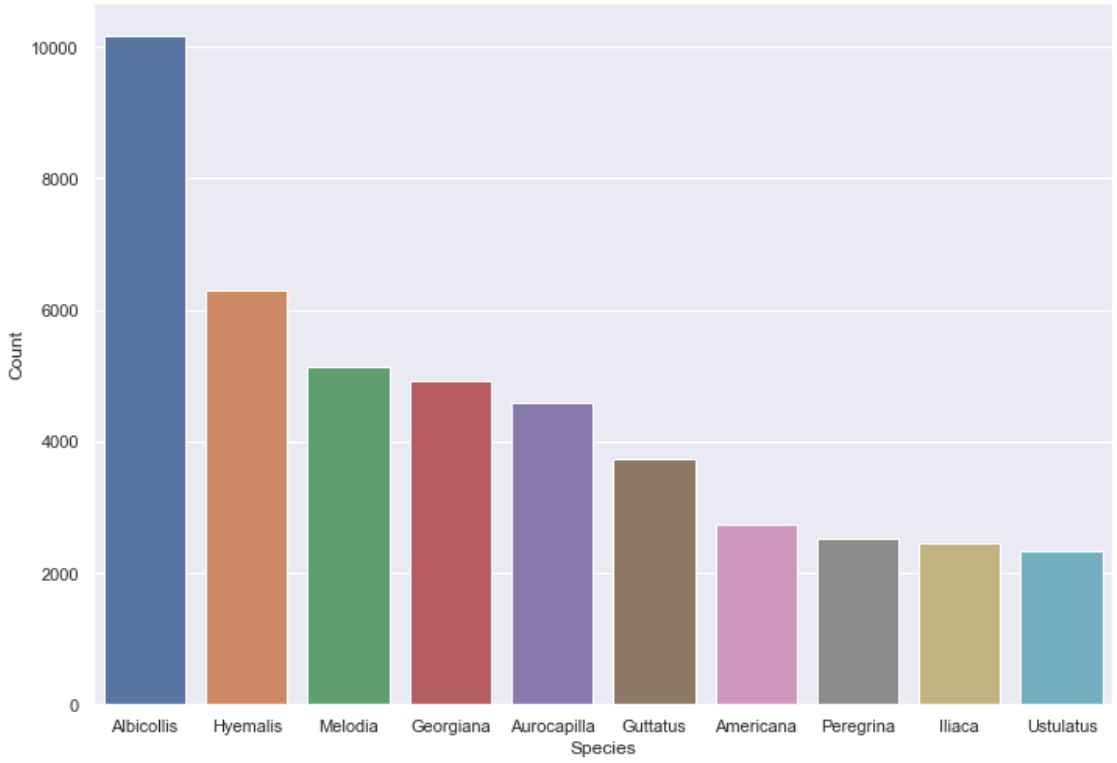
```
In [40]: br_plt('Genus')
```



Here we observe that Melospiza is most common genus that have collisions. But its is highly possible that Melospiza may be majority in the observed cities. If we an estimate of percentage distribution of genus by family we can take a weighted proportion and then check for most common genus of birds.

Collision by Species of Birds We can see that Albicollis species of birds have the highest collisions followed by Hyemalis and Melodia.

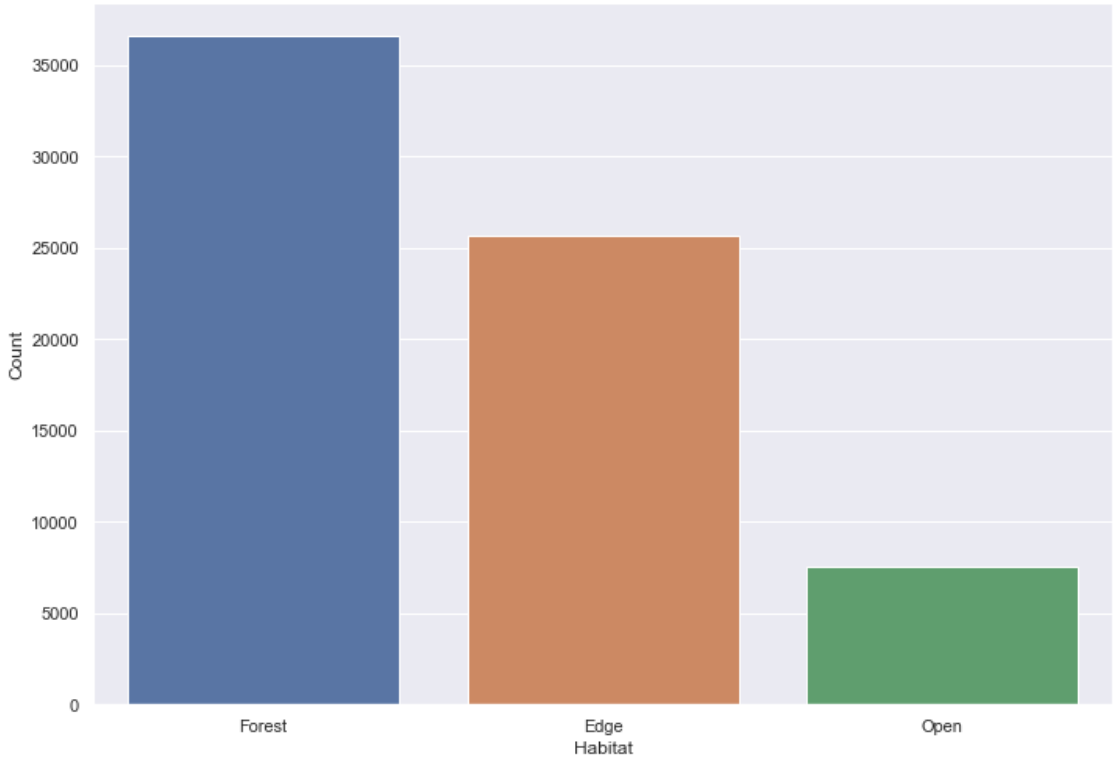
```
In [41]: br_plt('Species')
```



Here we observe that Albicollis is most common species that have collisions. But its is highly possible that Albicollis may be majority in the observed cities. If we an estimate of percentage distribution of species by family we can take a weighted proportion and then check for most common species of birds.

Collision by Habitat of Birds We can see that birds who usually dwell in Forest are twice as much as birds who live on edge and almost 7 times as much as birds who live in the open.

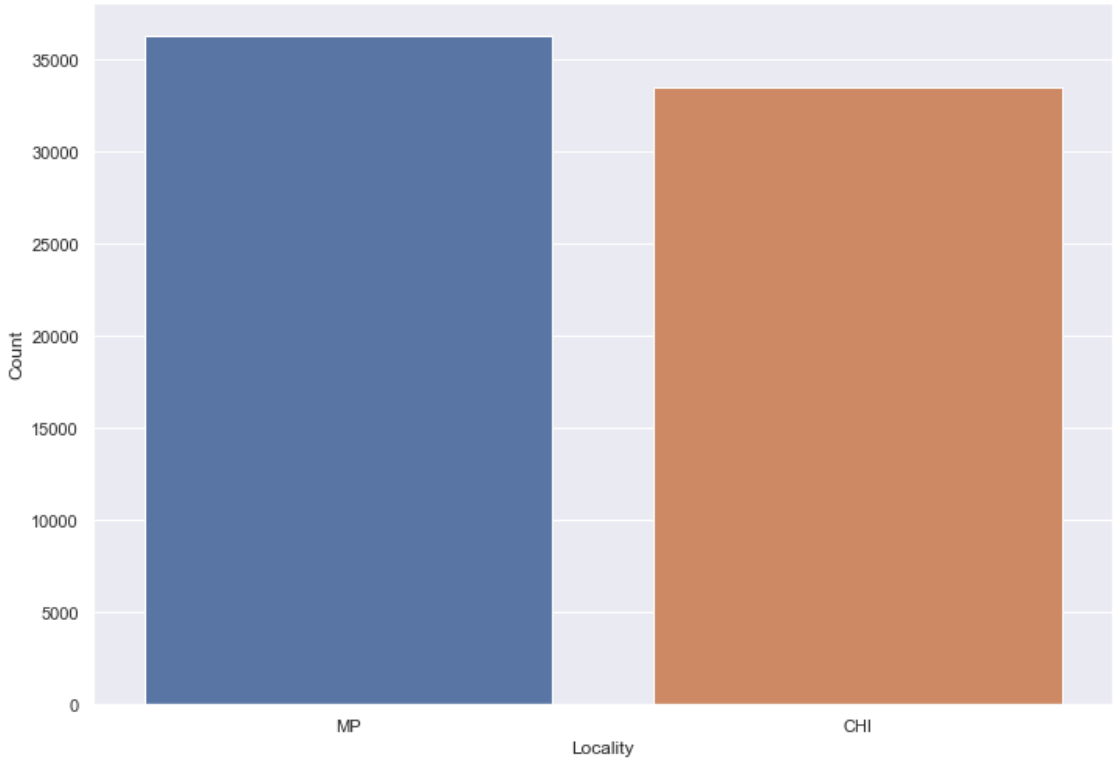
```
In [43]: br_plt('Habitat')
```



Here we can see that birds who are not used to artificial lighting are more susuptible for collisions. This could be a compelling feature.

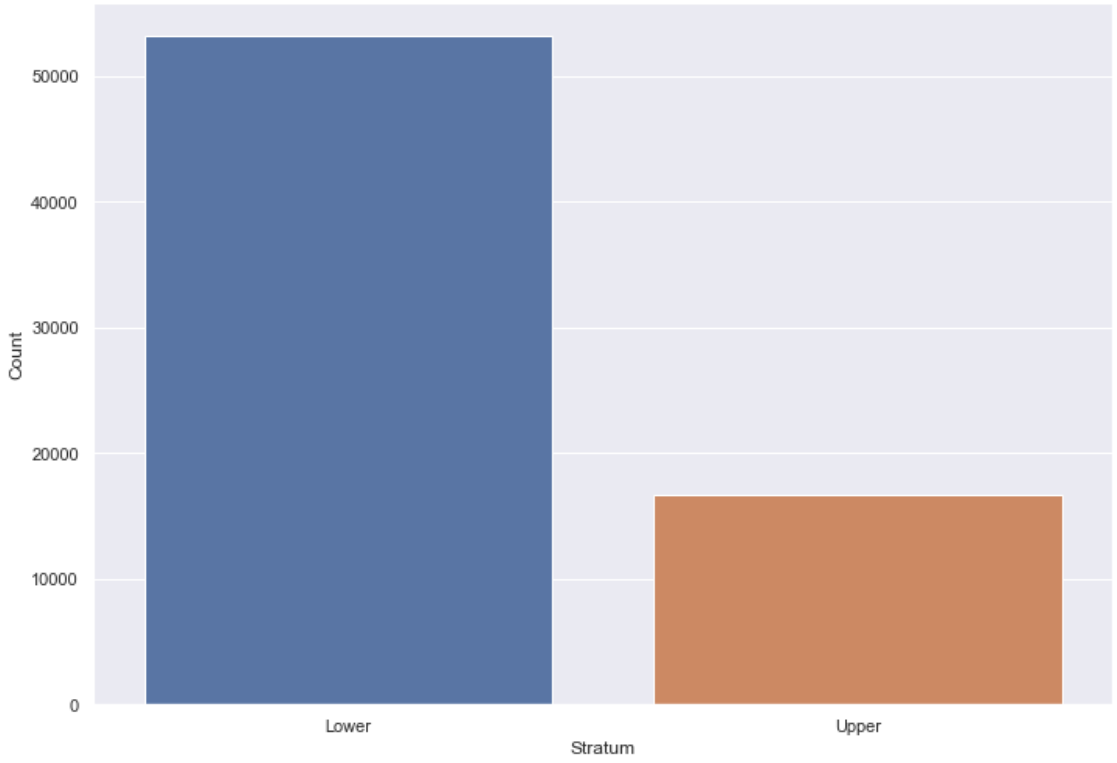
Collision by Locality of Birds We can see that there is an equal distribution of birds in both localities.

```
In [45]: br_plt('Locality')
```



Collision by Stratum of Birds We can see that the lower stratum birds have twice as much as collision than upper stratum birds.

```
In [47]: br_plt('Stratum')
```



Here we can see that birds who live on lower stratum are more susuptible for collisions. This could be a compelling feature.

2 Summary & Findings

1. Passerellidae Family of birds have the highest collisions followed by Parulidae and Turdidae

2. Melospiza Genus of birds have the highest collisions followed by Zonotrichia and Catharus

3. Albicollis species of birds have the highest collisions followed by Hyemalis and Melodia

4. Birds that usually dwell in Forest are twice as much as birds that live on edge and almost 7 times as much as birds who live in the open

5. The lower stratum birds have twice as much as collision than upper stratum birds

6. Birds who employ flight calls have twice as much as collision than the birds who don't employ flight calls