# TABLE OF CONTENTS

[**TABLE OF CONTENTS**](#_ft094pd0yk7r) **1**

[**Datasets**](#_jne7rc4wc62l) **1**

[Trees](#_n1o9j6nhe7i1) 1

[eBirds](#_5g1a8vrrjgt9) 2

[**Visuals**](#_ttx4p4kfs5hk) **4**

[Insights](#_bn05qk9k4uab) 7

# Datasets

Places Boundaries

<https://data.ca.gov/dataset/ca-geographic-boundaries> - shape file

Zip Codes Boundaries

<https://data.ca.gov/dataset/county-and-zip-code-references> - Just a complete list

'<https://raw.githubusercontent.com/OpenDataDE/State-zip-code-GeoJSON/master/ca_california_zip_codes_geo.min.json>' - Actual geometric shape

## Trees

Possible Keys to Aggregate

1139 Distinct **Zip Codes** in dataset**, 1769 in CA**

47 Distinct **Counties** in dataset**, 69 in CA**

702 Distinct **Places** in dataset**, 1522 in CA**

All Zip Codes in the Dataset show up in the Zip Code dataset that we found, except a Zip Code labeled as None (or null). There are 1939 null zips.

There are around 20 Places in the Dataset that don’t match up with the Places dataset that we found, but only because of formatting. We can easily get them to match

Zip Codes seem like the best key to aggregate to for trees. Least number of nulls and most coverage over California (reference visuals).

287 Observations were created in 2070. But all of them were then updated in 2018. They were all observed in San Jose.

13 Observations were created on December 29, 1899. All of them are in San Jose spread across two Zip Codes. Only one of the observations has a good amount of data. All other 12 have

246 Observations were created on January 1st, 1900. All of them are in San Diego, and seem to contain valid information

## eBirds

eBirds Dataset Download

[*https://www.gbif.org/occurrence/download?country=US&dataset\_key=4fa7b334-ce0d-4e88-aaae-2e0c138d049e&year=1000,2022&gadm\_gid=USA.5\_1*](https://www.gbif.org/occurrence/download?country=US&dataset_key=4fa7b334-ce0d-4e88-aaae-2e0c138d049e&year=1000,2022&gadm_gid=USA.5_1)

Links to Upload and Grab the Dataset

<https://cloud.google.com/storage/docs/uploading-objects#storage-upload-object-python>

<https://stackoverflow.com/questions/49541026/how-do-i-unzip-a-zip-file-in-google-cloud-storage>

73,611,972 Rows for Ebird in California

66,492,438 Rows for Chordata-Aves (Birds Only)

gbifID is the unique identifier for each row

species has more nulls than verbatimScientific name, they give the same information

Data from 1800 - 2021

All occurrenceStatus are labeled as PRESENT

Only 1,252,073 distinct pairs of Latitude, Longitude

80,296 Pairs of Latitude, Longitude did not match up with the ZipCodes in our ZipCodes dataset.

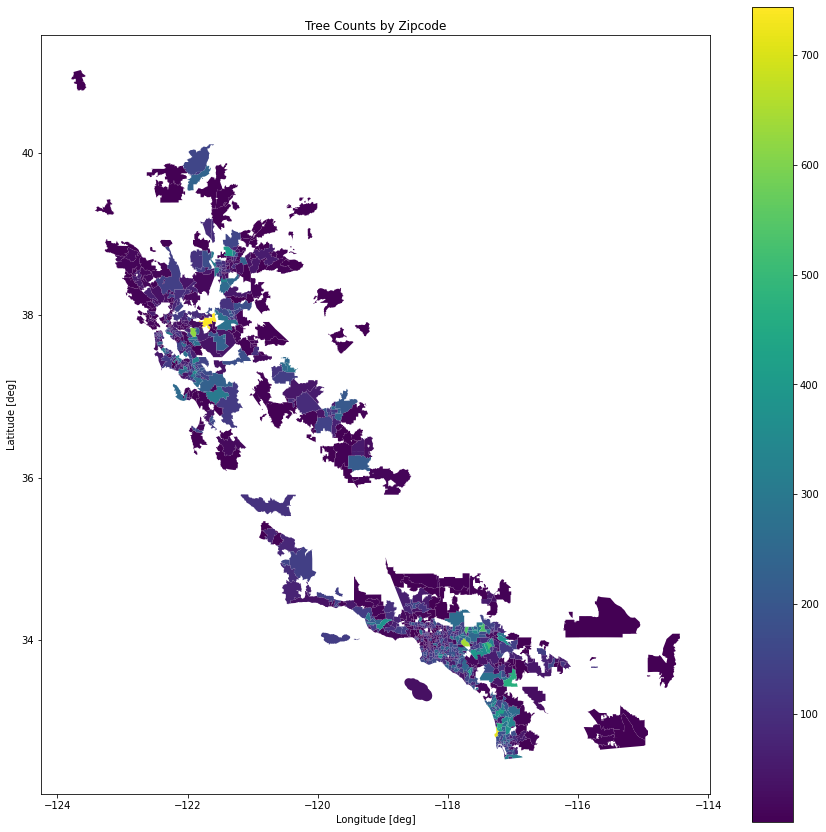
1765 Unique Zip Codes in the Dataset

1136 Zip Codes Overlap with the trees dataset

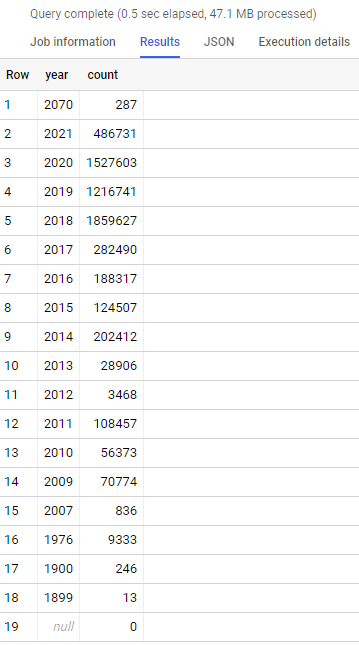
Using Zip Codes we find that 922,810 of the 1,252,073 distinct Lat/Long coordinates are apart of the overlap

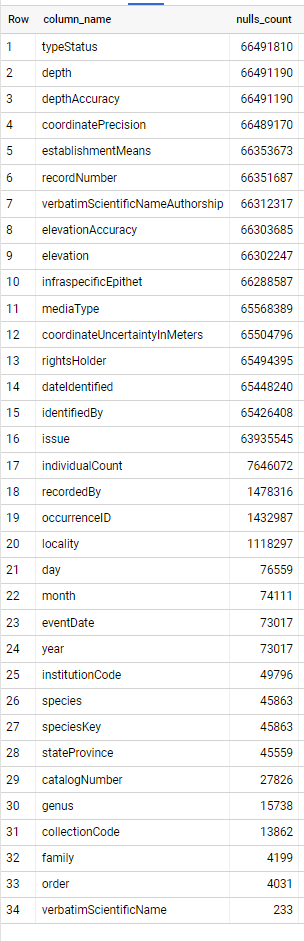
# Visuals

## 



## Insights



**