EXTRACT

In our database, we wanted to capture a full look at the city of Chicago and give users the ability to not only view the specifically discussed data, but find meaningful insights related to bird populations and flight datasets.

To compile the data necessary for this project, we used data.world for building-related bird strike data, the FAA for aviation-related bird strike data, the bureau of transportation for general flight data, and and the American Ornithology Society for bird populations.

CSV: https://data.world/animals/bird-building-collisions
CSV: https://wildlife.faa.gov/databaseSearch.aspx
CSV: https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp?pn=1
CSV: http://www.americanornithology.org/content/population-data-resources

TRANSFORM

White our downloaded data included bird species, variations in names were used for this field across the datasets. For each of the tables to be relatable at the species level, we first had to create a bird species key.

Next, after reading the bird strike CSV file into a Pandas dataframe, we filtered the Illinois airport data down to Chicago airports for the years 2004-2016 and created a new table that included only the columns that were relevant. Next we applied the same methodology to the building strike data.

To clean the flight data, we loaded the data into a pandas dataframe, we used the append function to combine Chicago O'Hare (ORD) data with Midway (MDW) and filtered for the same 2004-2016 timeframe from before. To complete this table we kept only relevant columns and renamed a column to make sense to anyone viewing it.

A complication we faced was receiving a key error due to spacing in the field name.

LOAD

Our final compilation consisted of four tables that exist within a SQL database. The current tables allow for inquires based on month, year, bird species, and airport. SQL was selected as our database for its ability to host a relational database.