When is a good day to go fishing?

To answer this question, we analyzed fish count data from the Bonneville Dam, the most downstream dam on the Columbia River.

- Biologists count migrating fish, including salmon (coho, chinook), and steelhead
- Upper headlands with cold water and clean gravel are the best place for young fish to spend the first months of their life prior to migrating to the ocean
- After feeding in the ocean for 3-4 years, adult fish return to the same nursery/spawning grounds where they were raised

Jeremy, Mike, Lasitha and I analyzed weather data from the National Atmospheric and Oceanic Administration as well as data on fish in the Columbia River.

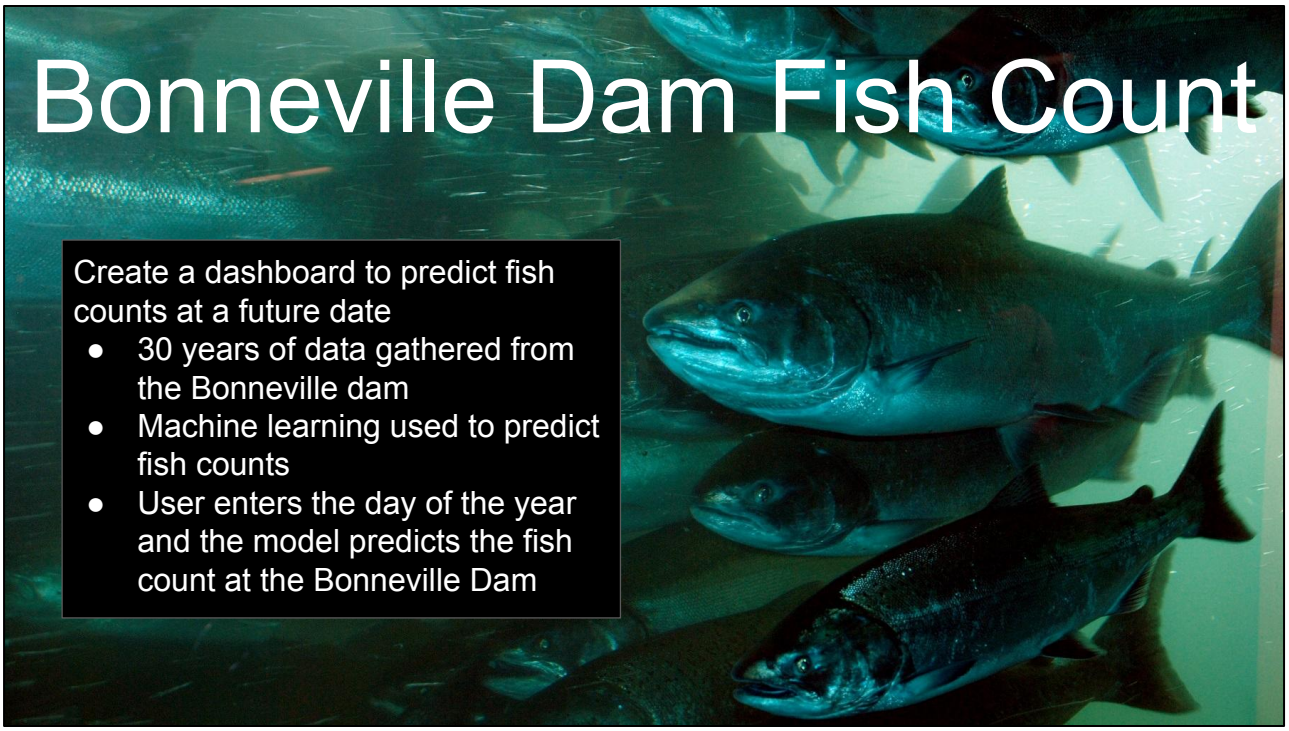
Specifically, we looked at:

- Steelhead, Coho and Chinook Salmon counts from the Bonneville Dam, Steelhead Catch Data from the Columbia River Basin

- The precipitation, minimum and maximum air temperature, and water temperature numbers were from Bonneville

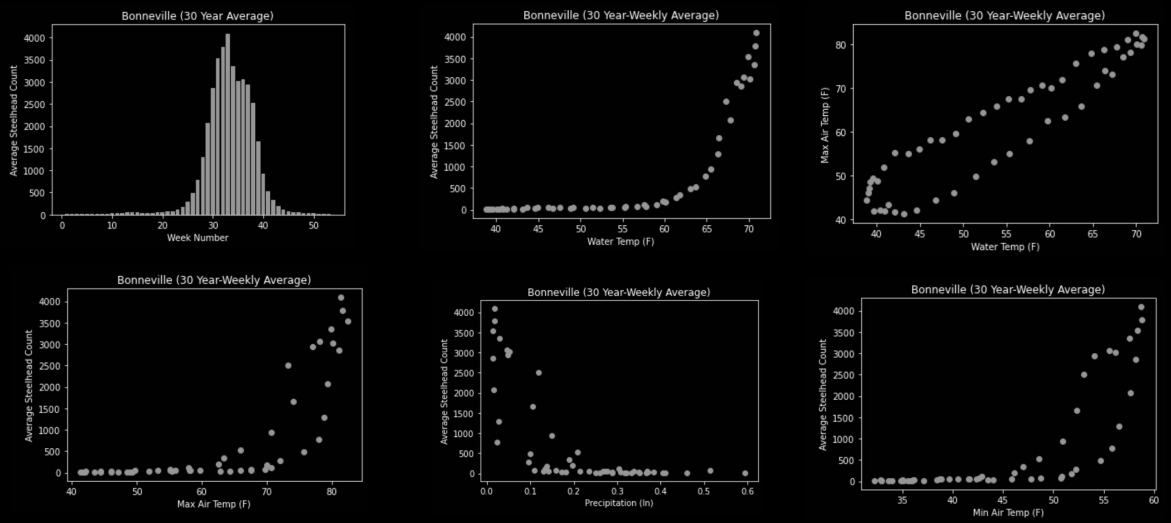
Columbia River

- To compare with fish count, we analyzed historical weather data for the Bonneville area, including air temp, precipitation and water temp
- Null Hypothesis: Weather conditions do not affect weekly fish counts.
- Fish count is our dependent variable (y); weather data is our independent variable (X)



Slide to discuss data and goal of the project.

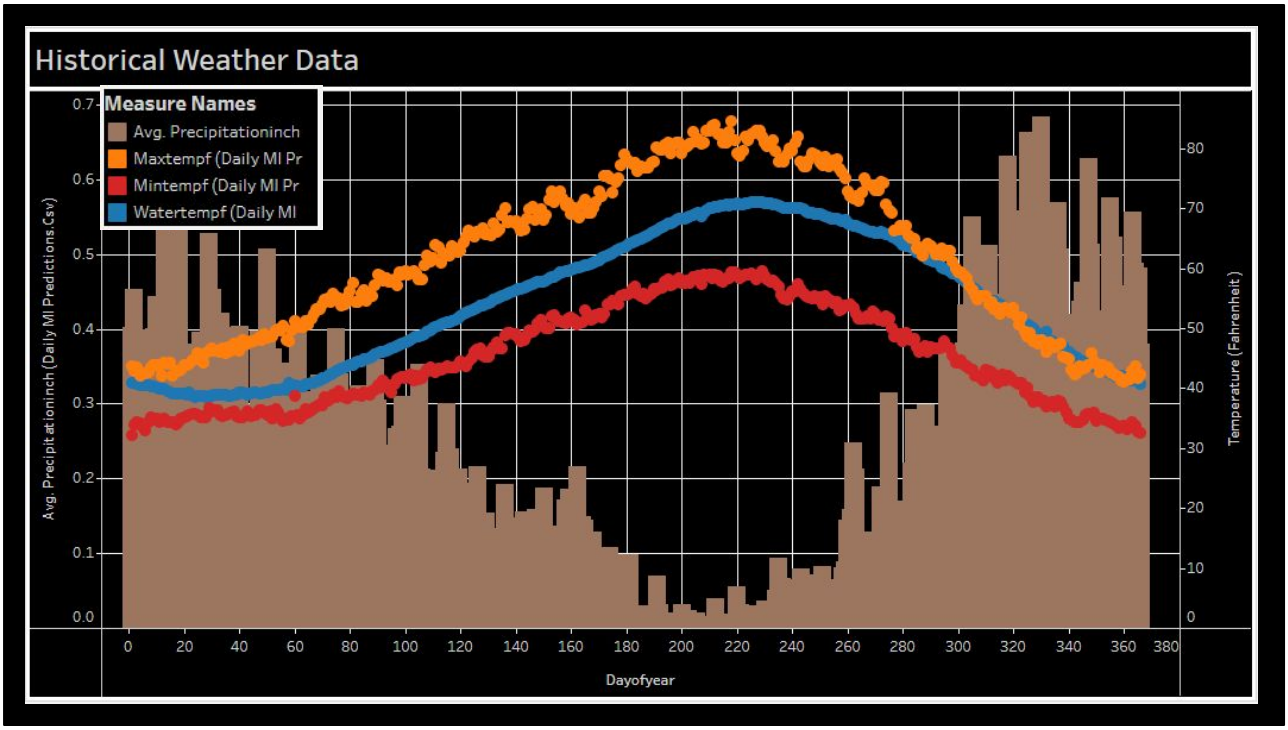
Visualizations



Initial analysis of the data based on Steelhead counts at Bonneville Dam:

- Count vs Week
- Count vs Water Temp
- Max air temp vs Water Temp
- Count vs Max air temp
- Count vs Precipitation
- Count vs Min air temp

These plots help show the count is related to each of these independent variables.



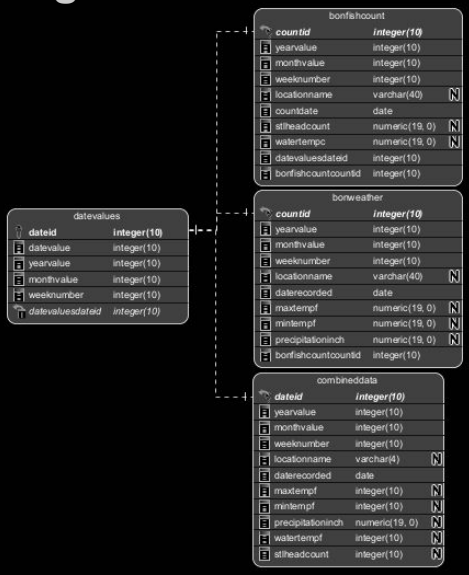
The averages of the input data precipitation, max temp, min temp and water temp. We see in the later slides how this distribution affects the modeling

Database Connections

1. Two Data Sources:
   a. Fish Count Data => DART Adult Passage Daily Counts for All Species | Columbia Basin Research (washington.edu)
   b. Weather Data => Past Weather | National Centers for Environmental Information (NCEI) (noaa.gov)
2. PostgreSQL and pgAdmin as the database and the data management tools for the Heroku database.
3. Used SQL to create new table with combined fish count and weather data.
4. Jupyter Notebook to extract the data from database for analysis.

Lasitha Starts Here

Entity Relationship Diagram



Five tables in total were need to house the data. The Fish data and Weather data were loaded into bonfishcount and bonweather tables. A new table was created to house the date information broken into a dateid, datevalue, yearvalue, monthvalue and weeknumber

Extract Transform & Load (ETL)

```
# Define data file directory
file_dir1 = '../Resources/refactored_data'
file_dir2 = '../Resources/weather'
```

```
# Create total fish count DataFrame
fish_df = pd.read_csv(f'{file_dir1}/total_data.csv', low_memory=False)
fish_df.head()
```

```
# Convert TempC, Celcius values into TempF, Fahrenheit values
fish_df['TempF'] = fish_df['TempC'] * (9/5) + 32
fish_df.head()
```

```
# Drop columns that are not used
fish_df.drop(['Chinook Run','JChin','WStlhd','JCoho','Lmpry','BTrout','Chum','Pink','TempC'], axis=1, inplace=True)
fish_df.head()
```

```
# Create weather data DataFrame
weather_df = pd.read_csv(f'{file_dir2}/bonWeather.csv', low_memory=False)
weather_df.head()
```

Weather and Fish data upload to Jupyter Notebook. Convert temperature from Celsius to Fahrenheit, drop extra columns.

ETL Continued…

```sql
CREATE TABLE DateValues (
    DateId INT NOT NULL,
    DateValue DATE NOT NULL,
    YearValue VARCHAR(4) NOT NULL,
    MonthValue VARCHAR(2) NOT NULL,
    WeekNumber NUMERIC NOT NULL,
    PRIMARY KEY (DateId)
);

CREATE TABLE BonFishCounts (
    CountId INT NOT NULL,
    YearWkValue VARCHAR(7) NOT NULL,
    YearValue VARCHAR(4) NOT NULL,
    MonthValue VARCHAR(2) NOT NULL,
    WeekNumber NUMERIC NOT NULL,
    LocationName VARCHAR(40),
    CountDate DATE NOT NULL,
    ChinookCount NUMERIC,
    StlheadCount NUMERIC,
    SockeyeCount NUMERIC,
    CohoCount NUMERIC,
    ChadCount NUMERIC,
    WaterTempF NUMERIC,
    PRIMARY KEY (CountId)
);
```

```sql
CREATE TABLE BonWeather (
    CountId INT NOT NULL,
    YearValue VARCHAR(4) NOT NULL,
    MonthValue VARCHAR(2) NOT NULL,
    WeekNumber NUMERIC NOT NULL,
    LocationName VARCHAR(40),
    DateRecorded DATE NOT NULL,
    MaxTempF NUMERIC,
    MinTempF NUMERIC,
    PrecipitationInch NUMERIC,
    PRIMARY KEY (CountId)
);

CREATE TABLE CombinedData2 (
    CountId INT NOT NULL,
    YearValue VARCHAR(4) NOT NULL,
    MonthValue VARCHAR(2) NOT NULL,
    WeekNumber NUMERIC NOT NULL,
    LocationName VARCHAR(40),
    DateRecorded DATE NOT NULL,
    MaxTempF NUMERIC,
    MinTempF NUMERIC,
    PrecipitationInch NUMERIC,
    WaterTempF NUMERIC,
    StlheadCount NUMERIC,
    ChinookCount NUMERIC,
    SockeyeCount NUMERIC,
    CohoCount NUMERIC,
    ChadCount NUMERIC,
    PRIMARY KEY (CountId)
);
```

```sql
CREATE TABLE daily_ml_predictions (
    DayofYear NUMERIC NOT NULL,
    WeekNumber NUMERIC NOT NULL,
    DateRecorded DATE NOT NULL,
    MaxTempF NUMERIC,
    MinTempF NUMERIC,
    PrecipitationInch NUMERIC,
    WaterTempF NUMERIC,
    StlheadPredict NUMERIC,
    ChinookPredict NUMERIC,
    CohoPredict NUMERIC,
    ChadPredict NUMERIC,
    SockeyePredict NUMERIC,
    PRIMARY KEY (DayofYear)
);
```

ETL Continued…

```
INSERT INTO combineddata2
(
select
    dv.dateid,
    dv.yearvalue,
    dv.monthvalue,
    dv.weeknumber,
    'Bonneville' as locationname,
    dv.datevalue,
    bw.maxtempf,
    bw.mintempf,
    bw.precipitationinch,
    bf.watertempf,
    bf.stlheadcount,
    bf.chinookcount,
    bf.sockeyecount,
    bf.cohocount,
    bf.chadcount
from datevalues dv
    left join
        bonfishcounts bf
        on dv.dateid = bf.countid
    left join
        bonweather bw
        on dv.dateid = bw.countid
);
```

```
# Define a file path to save the combined data file
filepath1 = Path(f'{file_dir1}/combineddata2.csv')
```

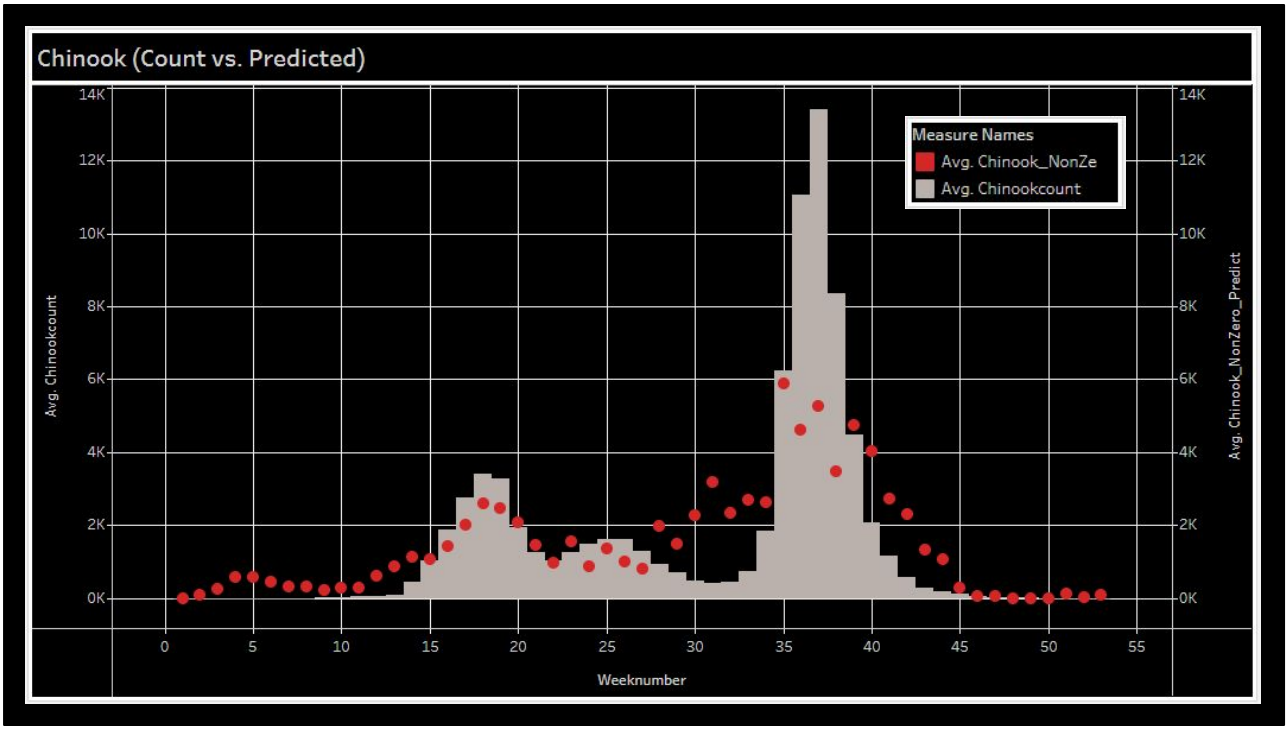
```
# Save the combined data as a CSV file
combined_df.to_csv(filepath1,index=False)
```

Lasitha Ends Here

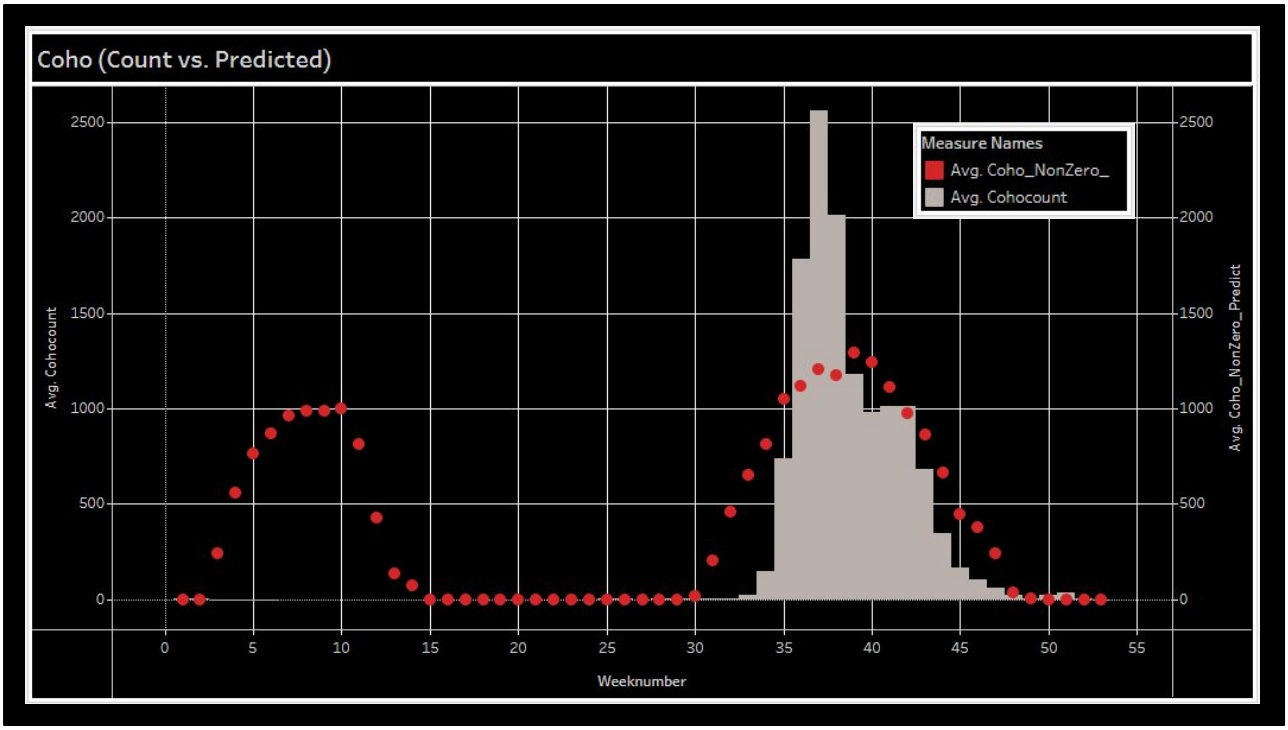
- Step One:
  - Dependent Variable - Steelhead count
  - Independent variables - Max temp, min temp, precipitation, water temp
  - Train, Test, Split - random state = 42, default size
  - Polynomial regression - exponential growth and multivariate data
    - A form of linear regression(actually creates features/flattens out the curve)
    - Limitations - same as linear regression:
      - the presence of a few outliers will have large effects on results
      - There are fewer validation tools for detecting outliers than other linear models
  - other machine learning models we tried: Non-Negative Least Squares/Ridge Regression and Classification/Lasso/Naive Random Oversampling/SMOTE Oversampling/Undersampling/SMOTEEN/Balanced Random Forest Classifier/Easy Ensemble AdaBoost Classifier
  - other scalers we tried: MinMaxScaler/MaxABSScaler/Robust Scaler/Power Transformer with the yeo-johnson method/Power Transformer with the box-cox method/Quantile Transformer with uniform distribution/Quantile Transformer with normal distribution/Normalizer

- Step Two:

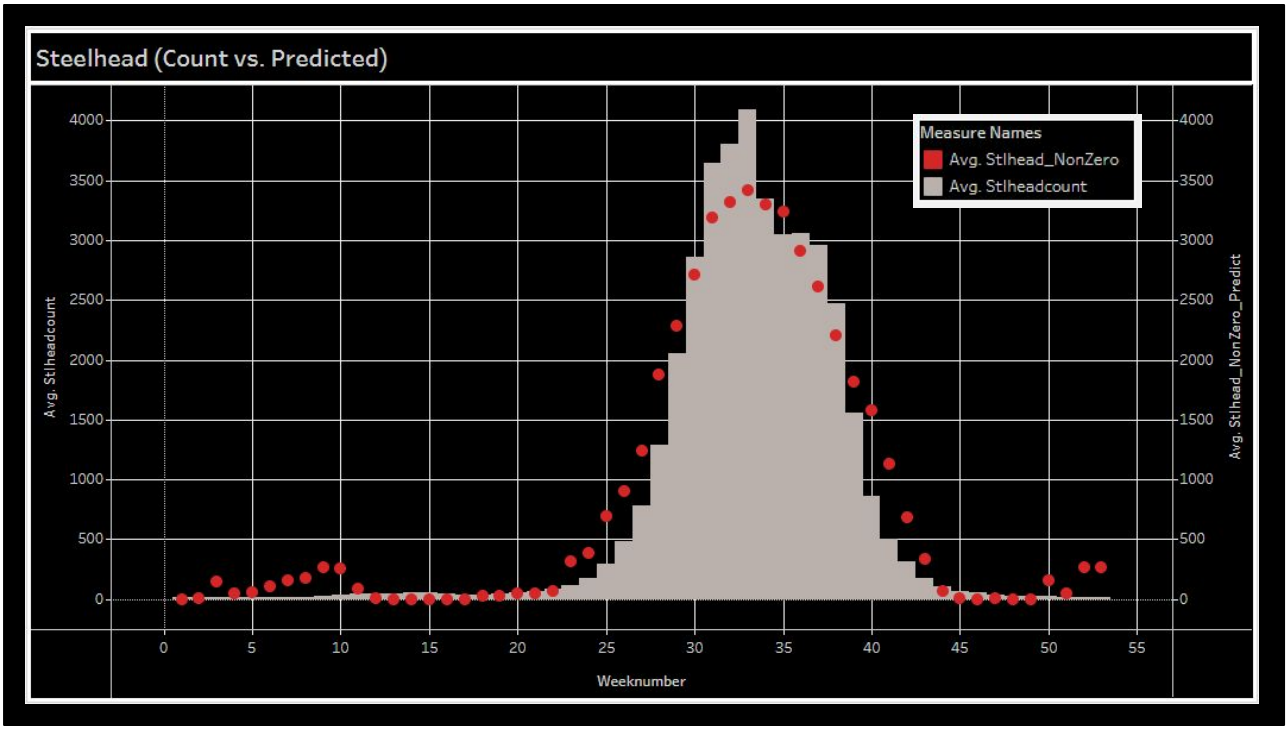
- Dependent Variable - Steelhead count, coho count, chinook count, shad count, and sockeye count
- Independent variables - Max temp, min temp, precipitation, water temp
- Train, Test, Split - random state = looped through 1 to 5000 to find best accuracy for each
- Result - Decided to keep steelhead, chinook, and coho



The graph shows how closely the actual count of Chinoon tracks with the machine predicted outcomes for a given day of the year. The prediction done through the machine learning fits the data well into the bimodal distribution of the actual data.

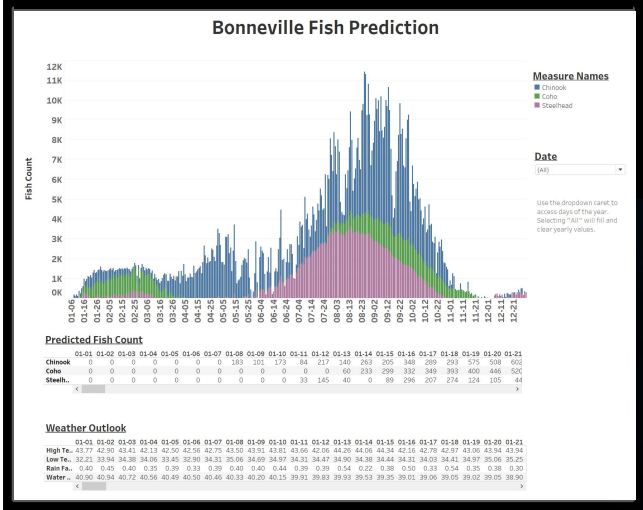


The predicted values of Coho are showing a signal detected in the 2 to 15 week date range even if there are no fish counted in that timeframe. Further investigation is required to understand this phenomenon. One possibility is that since the Coho run is in the fall where the temperatures are mild, the model predicted fish counts for the spring time weather where the temperature conditions were similar.

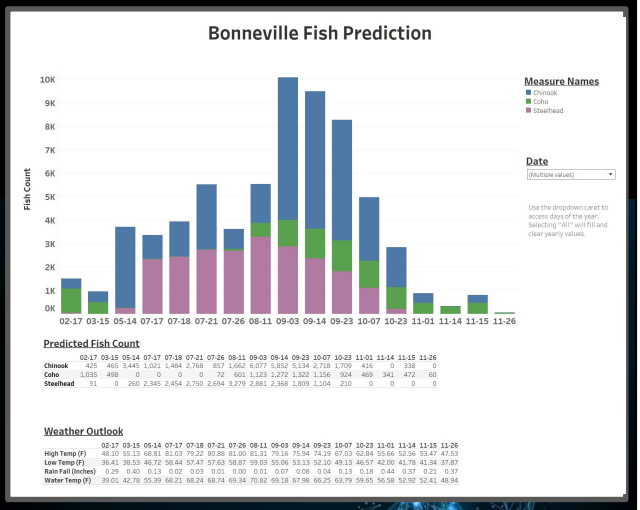


The steelhead distribution having a peak more towards the middle of summer does not show the bimodal predictions that was seen in the Coho count data set.

Count Predictions For Entire Year



Count Predictions For Selected Days



An entire year of ML data should load when the dashboard is opened. Multiple days can be selected through the “Date” filter.
(https://public.tableau.com/app/profile/jeremy3008/viz/BonnevilleFish1/BonnevilleFishPrediction?publish=yes)

Jeremy Starts at this slide:

User enters Month and Day to get the predicted count value.

Predicted fish counts are shown in the bar chart and in a table below the plot.

Long-range weather outlook will be shown in “Weather Outlook” table.

All dataframes filter on Date recorded.

Results

Link to dashboard for demonstration

Tableau Public:
https://public.tableau.com/app/profile/mike.thalken/viz/MT-Fish/Story1?publish=yes

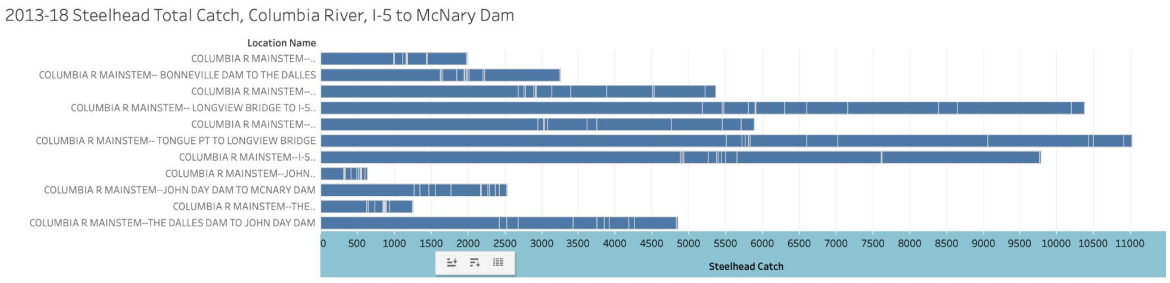
Tableau Online:
Tableau Story: Story 1 - Tableau Online

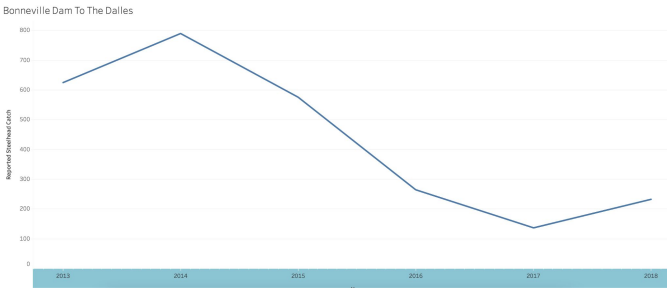
End of Presentation

Questions?

Next Steps

- Update ML model to evaluate days of the year with outlier weather conditions based historical dataset, i.e. extreme max/min temperatures or high rainfall.



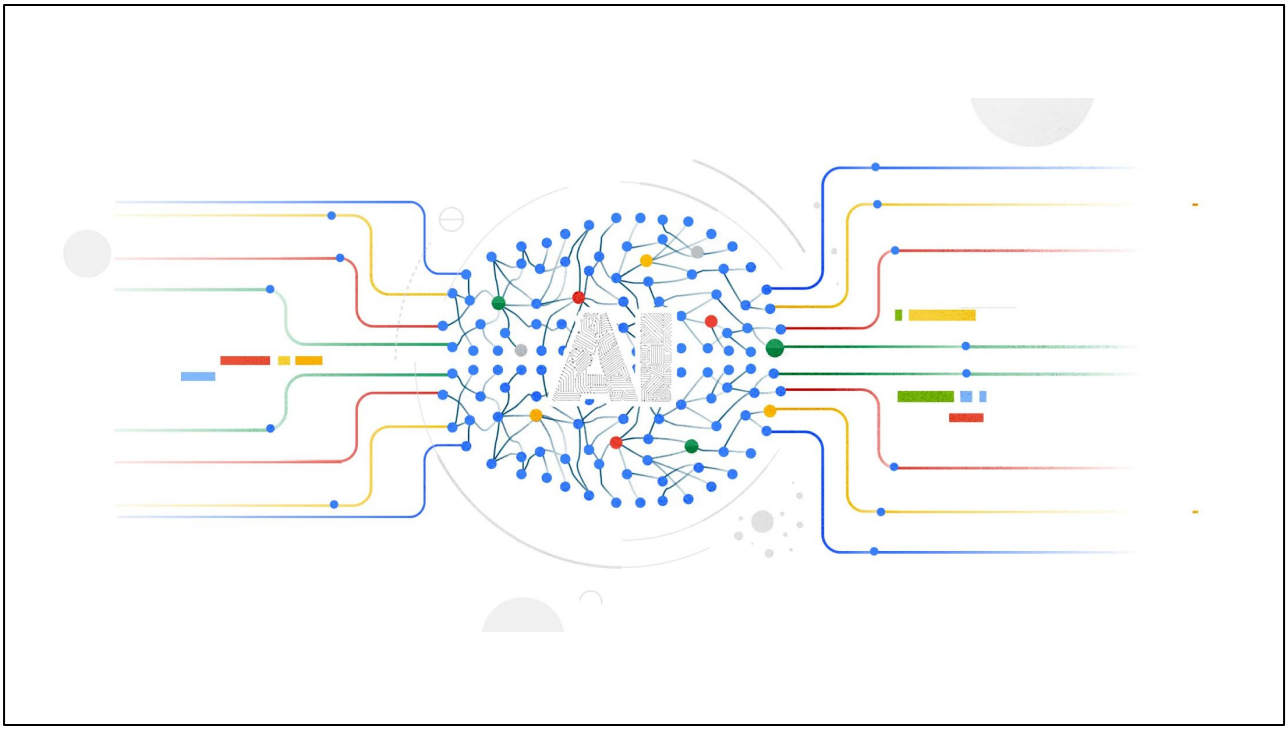


Steelhead, Reported Catch Data

Source: https://www.dfw.state.or.us/resources/fishing/sportcatch_archives.asp

Recommendation for future analysis

- Varying machine learned depending on fish type
- Add map and location information to dashboard
- Add catch counts for all fish types


AI