**Code for recreational fishing trip identification as implemented in “Fishing for Anglers in a Sea of Data: Using Mobility Data to Identify and Track Recreational Fishing in the Gulf of Mexico”**

The Jupyter Notebooks in this folder are used to identify recreational fishing trips following the protocol described in Ahmadiani and Woodward “Fishing for Anglers in a Sea of Data: Using Mobility Data to Identify and Track Recreational Fishing in the Gulf of Mexico.”

The primary data used were obtained under a contract from Cuebiq.com and all analysis with their data was carried out on their server in Jupyter Lab. During the process of developing the data, protocols and code changed and we have not attempted to bring all code up to date. In particular, during the period over which the code was written, Cuebiq/Spectus changed their SQL platform from Trino to Snowflake. Python packages change over time and, therefore, code may not be correct if run at a later date. In particular, the H3 code underwent significant changes over time.

**Summary of the notebooks provided**

0A AIS data download, clean & process.ipynb

* *Downloads, cleans and develops features for AIS data*
* Input: zipped files of pings from https://marinecadastre.gov/accessais
* Output: Indicators2019\_All.C.csv.csv
* *In the middle of this notebook, reference is made to Linux commands were used to efficiently concatenate all files.*

0B AIS Machine Learning

* *Carries out ML Classification on the AIS Data*
* Input: Indicators2019\_All.C.csv.csv
* Output: rf\_model\_AIS\_2019.pkl

1 Random Sample from OurTables.04.ipynb

* *Gathers pings from 300,000 randomly sampled devices*
* Input: Raw data pulled from Spectus’s V2 data within the Gulf
* Output: Pings\_OurTable\_Gulf\_ALL.csv

2 CreatingIndicators\_OurTables.04

* *Creates indicators for the trips that have enough pings*
* Input: Pings\_OurTable\_Gulf\_ALL.csv
* Output: Indicators\_OurTable.csv

3 MachineLearningClassification.02

* *Uses the ML classifier created with the AIS data to predict the vessel class for the mobility data trips*
* Input: Indicators\_OurTable.csv
* Input: rf\_model\_AIS\_2019.pkl
* Output: Indicators\_OurTable.Predictions.csv

4 RecTrip\_Identification&Selection.04

* *Develops and implements the non-ML criteria for recreational trip identification. Creates summary stats tables for identified trips*
* Input: Indicators\_OurTable.Predictions.csv
* Input: Cuebiq's Device table – raw data for pings outside the Gulf polygon
* Output: rec\_indicators\_with\_V3
* Output: DisappearanceIndicators.csv
* Output: Stop\_Trawls\_Indicators.csv

5 Identify\_Stops\_Near\_Stations.08

* *File carries out the process for distinguishing station-* *and nonstation-trips.*
* Input: rec\_indicators\_with\_V3
* Input: DisappearanceIndicators.csv
* Output: Station\_NonStationAnalysis\_full.csv

6 Validation.06

* *This carries out the validation analysis by comparing with MRIP and TPWD data*
* Input: rec\_indicators\_with\_V3
* Input: DisappearanceIndicators.csv
* Input: Station\_NonStationAnalysis\_full.csv