

Optimizing Coverage of Hunting Areas for Maximum Removal of Burmese Pythons in South Florida

94-881 Managing Analytics Projects (Spring 2020)

Patrick Campbell, Alice Beattie, Taylor Burandt

March 4, 2020

Table of Contents

Executive Summary

1. Problem Framing
2. Data Exploration
 - a. Data Variable Selection
 - b. Data Source Selection
 - c. Data Source Analysis
 - d. Data Integration
3. Analysis Choices and Justification
 - a. Generalized Linear Model Regression
 - b. Optimization
4. Visualizing and Communicating the Analytics
 - a. Geospatial Analysis
5. Project Plan
 - a. Stages, Milestones, and Deliverables
 - b. Stakeholder Analysis
 - c. Hardware and Software Resources
 - d. Staffing Requirements
 - e. Risk Management
6. Success Measurements
7. Lessons Learned
8. Limitations and Future Research

Bibliography

Executive Summary

The wetland ecosystems of South Florida are threatened by invasive Burmese pythons. The pythons were introduced through the exotic pet trade and are quickly eradicating native bird and mammal species, causing long-term damage to these fragile ecosystems.¹ In March of 2017, the South Florida Water Management District (SFWMD) launched a new program that pays local hunters a bounty to capture and eliminate Burmese pythons within a designated hunting area.² Under the current system, 25 contract hunters choose their own hunting locations based on previous success and where they believe pythons will be. Hunters use a custom smartphone app to record data about their activity in the field, including the time, locations, and environmental conditions for each capture.

In a future stage of the program, managers want to transition to a dispatch model in which hunters are allocated to specific locations to provide more even coverage across the hunting area. By analyzing the data they've collected over the past three years, program managers can predict where pythons are most likely to be found and guide the allocation of hunters across the hunting area to maximize the number of pythons caught and expected payout for hunters. Using generalized linear models, optimization, and GIS mapping, the SFWMD can more effectively manage the spread of invasive Burmese pythons through South Florida and improve the health of the sensitive ecosystems they impact.

1. Problem Framing

The framing of the problem was done in three parts: decision to be improved, decisionmaker, and the value of the improved decision.

1. **Decision to be improved:** How should the project manager dispatch 25 hunters in time and space throughout the project coverage area to maximize the expected catch of Burmese pythons (in a measurement of snake-feet) on a given day?
2. **Decisionmaker:** Project manager for SFWMD Burmese python elimination program
3. **Value of Improved Decision:** The value of improving this decision includes the following outcomes:
 - a. Quantitative
 - i. Increase in the number of Burmese pythons removed on average each year
 - ii. Higher payouts for hunters
 - iii. More equal/fair distribution of risk and opportunity among hunters, resulting in more uniform per unit effort earnings across team
 - b. Qualitative

¹ Michael E. Dorcas, John D. Willson, Robert N. Reed, Ray W. Snow, Michael R. Rochford, Melissa A. Miller, Walter E. Meshaka, Paul T. Andreadis, Frank J. Mazzotti, Christina M. Romagosa, Kristen M. Hart, "Severe mammal declines coincide with proliferation of invasive Burmese pythons in Everglades National Park," *Proceedings of the National Academy of Sciences*, Feb 2012, 109 (7) 2418-2422.

² "Python Elimination Program," South Florida Water Management District, accessed Feb. 28, 2020, <https://www.sfwmd.gov/our-work/python-program>.

- i. Improved health for native wildlife and ecosystems
- ii. Increase in number of native mammal and bird observations in impacted areas
- iii. Fewer conflicts between hunters due to territorial disputes and bottlenecks

Hypothesis: By providing program managers with a model of how to allocate its 25 hunters across the program coverage area in order to optimize expected value³, we can help improve the program's effectiveness in terms of total number and biomass of pythons removed; more evenly distribute risks and opportunity among hunters, resulting in more uniform per unit effort earning across the team; decrease the number and frequency of conflicts between hunters due to territorial disputes and bottlenecks; and improve the health of native wildlife and ecosystems.

2. Data Exploration

2.1 Data Variable Selection

The next phase of project development is the exploration of data. Before any data was received, the following variables were identified as essential and/or important to the analysis and future implementation of the hypothesis:

1. Dependent variables
 - a. Hunter payout per session (in U.S. dollars)
 - b. Number/biomass/length of snakes captured in a session
 - c. Number and frequency of conflicts between hunters (if available)
 - d. Number and frequency of bottleneck events (as measured by the density of hunters in a single location, if available)
 - e. Number of native mammal and bird observations in impacted areas⁴
2. Independent variables
 - a. Capture locations of pythons (lat/long)
 - b. Date and time of capture
 - c. Size of python (length and weight)
 - d. Sex of python
 - e. Bounty price for pythons (for expected payoff calculations)
 - f. Temperature day python captured
 - g. Regional precipitation on day python captured
 - h. Cloud conditions on day python captured

³ Expected value is being treated as a placeholder until we can resolve what the desired value to optimize is, given that what is most important to the managers might not be what is most important to the hunters. Our final analysis may attempt to strike the optimal balance of multiple values/objectives, e.g., expected payout to hunters (\$) as well as number of snakes caught. As it happens, these values are not necessarily consistent with one another.

⁴ Possible data sources are cited in Dorcas et al. 2012, including "records of road-killed mammals from surveys conducted by National Park Service rangers within ENP from 1993–1999 and the authors' own data compiled from various sources.

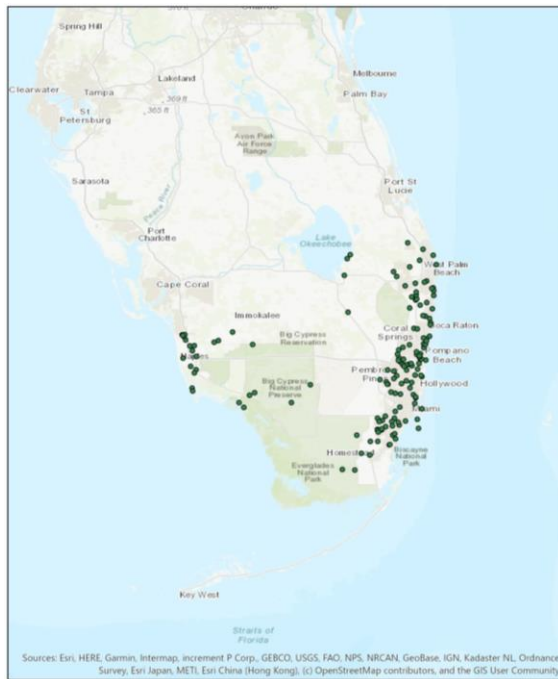
i. Water levels across hunting areas

2.2 Data Source Selection

Through the identification of the variables and discussions with stakeholders at SFWMD, the following datasets and sources were identified:

- A. SFWMD Python Elimination Program (PEP) hunter data
- B. SFWMD Survey123 data (csv)
- C. EDDMapS Burmese python distribution data (csv)⁵
- D. DBHYDRO water level data (csv)⁶
- E. NOAA temperature & precipitation data for weather stations throughout Florida over the reporting period (csv)⁷

Figure 1. Weather stations in South Florida⁸



2.3 Data Source Analysis

⁵ EDDMapS (2020), Early Detection & Distribution Mapping System, The University of Georgia - Center for Invasive Species and Ecosystem Health, accessed January 27, 2020.

<https://www.eddmaps.org/distribution/viewmap.cfm?sub=20461>

⁶ "Water Levels," South Florida Water Management District, accessed January, 27, 2020,

<https://www.sfwmd.gov/science-data/levels>.

⁷ "National Centers for Environmental Information," NOAA, accessed February 5, 2020,

<https://www.ncdc.noaa.gov/>.

⁸ *ibid.*

A. SFWMD hunter data and SFWMD Survey123 data

The two SFWMD hunter datasets were very granular with specific latitudes and longitudes reported for each capture in addition to very specific python information. The only issue with this data is the potential of human error and null values, which is to be expected with most data sets. However, weather station data will cover a small region of the state. Given that snakes are individuals, and a dead snake will not reappear where it's already been caught, we want to advise hunters on what regions to aim for, not specific location points.

Additionally, weather data will be specific and granular when it comes to daily temperature and precipitation. We will need to create categories of temperature and precipitation levels in order to allow a hunter to make a meaningful decision (i.e., if it's between 50 and 60 degrees fahrenheit, and there have been 1- 2 inches of rain in the past day, the optimal location is X region). This will require additional research on python biology and behavior (i.e., is there a temperature below which all snakes will be more docile? Or, below which, python eggs won't hatch?). This may be outside the scope for this project.

This data was relatively clean and organized when received. This data allowed some initial conclusions to be drawn. For example, the average python caught is of adult age (based on its length), thus a higher bounty and informs the current monetary incentive python hunters have in capturing the snakes.⁹ Another example is that the data also benchmarks the number of pythons caught annually. The data ranges from March 2017 - January 2020, with 2,779 caught in that timeframe. This data point is important when evaluating the potential success of maximizing the number of pythons caught.

Data information overview: 2,779 records of python captures described by 18 total attributes, including but not limited to weight, length, sex, capture date, time and location. Using excel, multiple summary statistics were calculated as seen in Tables 1 and 2. (The data has many other attributes that do not appear to be relevant to the analysis at this time.)

Table 1. SFWMD Summary statistics

Table 2. Breakdown of captures by sex¹⁰

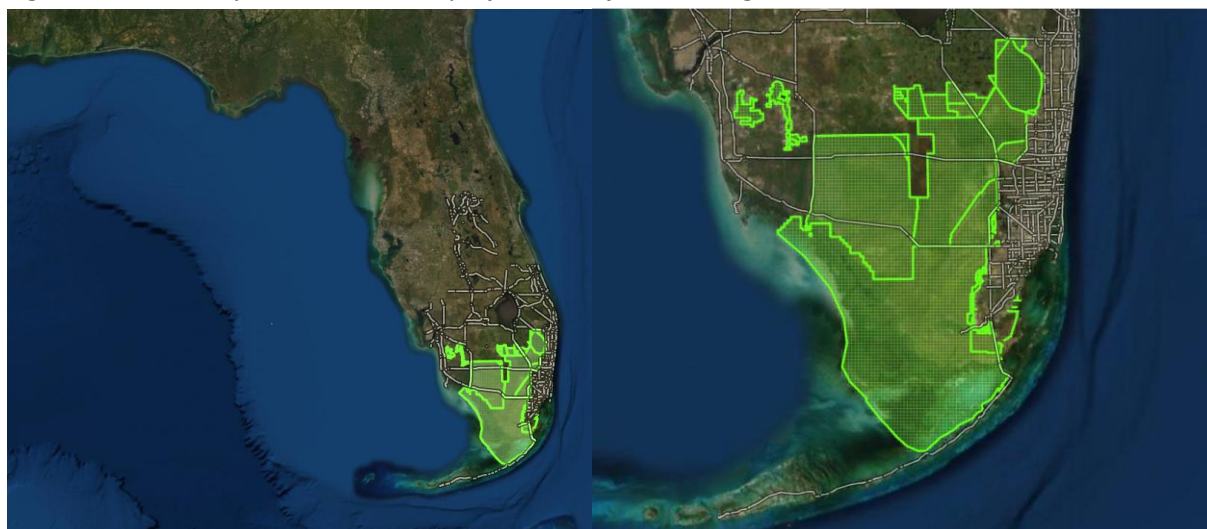
Count Female	Count Male
1,381	1,320

⁹ "Python Elimination Program," South Florida Water Management District, accessed February 4, 2020, <https://www.sfwmd.gov/our-work/python-program>.

¹⁰ The total count discrepancy in total number of pythons caught and the total of both males and females is due to blank values in the sex column in the data.

	Total Length (ft)	Weight (lbs)
Average	6.28	11.78
Median	6.58	7.00
Max	17.42	144.00
Min	1.33	0.18
Range	16.09	143.82

Figure 2. SFWMD Python Elimination projected Project Coverage Area, 2020



B. NOAA Historical Weather Data

Burmese pythons' behavior and activity levels are strongly influenced by temperature and weather conditions. Of many South Florida weather stations reporting to NOAA, nine weather stations consistently reported minimum, average and maximum temperatures. Because pythons are ectotherms, and therefore reliant on external sources to control their body temperature and metabolism, it is of interest to understand the average minimum temperatures reached at each station and the standard deviation.

Table 3. Average and Standard Deviations of Minimum Temperatures

NAME	ave_minimum_tempsd_min_temp	
CACHE FLORIDA, FL US	67.87	8.36
CHEKIKA FLORIDA, FL US	69.42	8.91
FORT LAUDERDALE INTERNATIONAL AIRPORT, FL US	71.71	7.98
MIAMI INTERNATIONAL AIRPORT, FL US	71.85	7.54
MILES CITY FLORIDA, FL US	65.80	9.43
OASIS FLORIDA, FL US	66.39	9.86
OCHOPEE FLORIDA, FL US	67.48	9.60
RACCOON POINT FLORIDA, FL US	67.34	9.81
WEST PALM BEACH INTERNATIONAL AIRPORT, FL US	69.69	8.84

Similarly, minimum temperature and overall variation day-to-day of temperature are different throughout the year. In August, for example, the standard deviation of minimum temperature is less than 3 degrees Fahrenheit. This suggests that temperature will be more useful in dispatching hunters in the fall and winter months when temperatures are both lower and have greater variance.

Data information overview: 9,854 daily weather summaries for nine weather stations in Florida, recorded between January 1, 2017 and December 31, 2019.

Twelve attribute fields and summary statistics include the following:

Table 4. NOAA Data Attributes and Summary Statistics

STATION	NAME	LATITUDE	LONGITUDE	ELEVATION	MONTH
Length:9854	Length:9854	Min. :25.39	Min. :-81.30	Min. :1.500	Min. : 1.000
Class :character	Class :character	1st Qu.:25.79	1st Qu.: -81.03	1st Qu.:2.100	1st Qu.: 4.000
Mode :character	Mode :character	Median :25.92	Median :-80.68	Median :2.400	Median : 7.000
		Mean :25.95	Mean :-80.71	Mean :3.578	Mean : 6.526
		3rd Qu.:26.08	3rd Qu.: -80.32	3rd Qu.:4.600	3rd Qu.:10.000
		Max. :26.68	Max. :-80.10	Max. :8.800	Max. :12.000
DAY	YEAR	PRCP	AVERAGE_TEMP	MAX_TEMP	MIN_TEMP
Min. : 1.00	Min. :2017	Min. :0.000	Min. :47.00	Min. : 52.00	Min. :13.00
1st Qu.: 8.00	1st Qu.:2017	1st Qu.:0.000	1st Qu.:72.00	1st Qu.: 82.00	1st Qu.:63.00
Median :16.00	Median :2018	Median :0.000	Median :78.00	Median : 87.00	Median :71.00
Mean :15.72	Mean :2018	Mean :0.172	Mean :76.33	Mean : 85.64	Mean :68.62
3rd Qu.:23.00	3rd Qu.:2019	3rd Qu.:0.080	3rd Qu.:82.00	3rd Qu.: 90.00	3rd Qu.:76.00
Max. :31.00	Max. :2019	Max. :7.130	Max. :90.00	Max. :100.00	Max. :84.00
NA's :6572					

While the data was mostly clean, well-organized and well-documented, precipitation data is missing for 6,572 observation-days so it will be difficult to include in further analysis. However, python behavior and hunter preference align in that neither like rainy days. Pythons are hard to find and hunters are not likely to be in the field on rainy days so this analysis will probably not need to depend on granular precipitation data.

In addition, NOAA data includes fields for snow, snow accumulation, and wind. Snow and snow accumulation were 0 for all observation-days in this set, so that attribute was removed. Additionally, wind was deemed not relevant to our analysis so it was also removed. NOAA data is free and publicly available at <https://www.ncdc.noaa.gov/>.

C. EDDMapS Burmese Python Distribution Data

This data contains the shapefiles of South Florida. These shapefiles being pre-made saves a lot of time in that this project team does not have to geolocate the areas of interest. The map data does not provide additional conclusions at this time, but is integral in preparing the visualizations of the data findings.

D. DBHYDRO Data

The DBHYDRO database contains valuable environmental data collected routinely by the SFWMD. For our purposes, we were interested primarily in the rain gauge and water level (stage) data for all monitoring stations within the project coverage area, as these have the most direct impact on the snakes' and hunters' behavior. Unfortunately, these data are organized by station names and codes that we are unfamiliar with, making it difficult to query. As of the preparation of this report, we were still working out the best way to identify the data that corresponded to our desired area. The strategy we are currently exploring is using the monitoring station map and index provided [here](#) along with the maps

of the project area to filter the relevant station names, then using these names in our query to return the desired rows. The results of this process, if successful, will be incorporated into all future products.

2.4 Data Integration

The data sources used for this analysis includes data in both CSV and SHP (shapefile) formats. All of the CSV data sets will be joined in R using a combination of primary keys (when available) and unique attribute sets (e.g., date, time, and location of capture). After processing, all CSV data will be imported into ArcGIS Pro and mapped onto the program coverage area using the data's latitude and longitude fields.

In order to prepare the integrated data for analytic modeling, we will join the latitude/longitude data from the SFWMD's Survey123 dataset with their python capture dataset. Presently, it appears the Survey123 dataset is not set up to do this join since the two data sets do not share any single field in common, and multiple fields must be used. Our prospective solutions to this problem are described below, but will likely require some additional troubleshooting.

Only once we have these two datasets joined will we be able to move the data onto our preferred analytical platform, ArcGIS Pro. Seeing as latitude/longitude is provided in the joined python data and the weather data, additional data cleaning is likely to be minimal in GIS.

In order to assess which weather station's data should be assigned to each capture, spatial analysis will be used to determine the nearest weather station to each capture location. Additionally, ArcGIS density analysis will be used to divide South Florida into regions that are meaningful and actionable for dispatchers based on historical python capture density (i.e. a geographic area with high capture density should probably be split into multiple dispatch regions so multiple hunters can maximize captures).

This strategy did not work in every case, however. For example, there was no single field that would allow us to join the two main SFWMD hunter datasets. Although we have not settled on a single strategy for dealing with this problem, we have considered the following two solutions:

1. Substituting unique identifiers (e.g., primary keys) with unique sets of attributes (e.g., date, time, and location of capture).
2. Running separate analyses on the two data sets and then synthesizing our findings to tell a more complete story about what factors go into making a successful hunting strategy.

The following steps were taken to clean the data:

1. Eliminating extraneous attributes and records
 - a. The Survey123 data contains data fields that are irrelevant to our analysis (e.g., "take another picture," "creation date," "creator," "drop-off site," etc.). Removing these simplifies the data exploration process and analysis.
 - b. Likewise, the data included several records for other species of exotics, including Boa constrictors, ball pythons, and others. These records were all excluded.

2. Renaming column headers for easy reference in R (e.g., “Date Caught” changed to “Date_Caught”, etc.)
3. Standardizing NA entries
 - a. Different fields used different conventions for identifying null or missing values. In the Survey123 data, for example, “No submitter” was used for the SubmitterName field to indicate that no submitter was specified for the record. In order to more easily recognize null and missing data, we replaced all such entries with the text “NA.”
4. Removing and/or recoding NAs to preserve integrity of summary statistics and other analyses
 - a. While removing NAs isn’t always the best strategy for dealing with missing data, it turned out to be adequate for the purposes of our analysis. Because of how the data was collected (namely, through manual inputs into a mobile app, which pre-specified each field as either required or not required), it didn’t seem appropriate to read any further into missing fields or attempt to recode them in any way.
5. Combine separated fields
 - a. E.g., the Survey123 data contained fields that separated the foot and inches components of the snake’s length measurement, as well as a third field for the total measurement measured in feet. We consolidated these three fields into a single field for total length measured in inches.
6. Separate combined fields
 - a. E.g., the Survey123 data contained fields that combined the date and time of each capture. For the purposes of our analysis, we found it more useful to separate these into distinct time and date fields.
7. Change/standardize field formats
 - a. E.g., changing time formats to a standard 24:00 format to eliminate ambiguities between AM and PM entries
8. Convert numeric fields from text to numeric format
 - a. E.g., the Survey123 data contained a field for “temperature at time of capture,” which was recorded in text format (e.g., “seventies” instead of “70-79”). In cases like this, we found it useful to convert these fields to a numeric format.
9. Resolve inconsistencies and errors in data
 - a. E.g., the Survey123 data contains records in which there is a “capture date” recorded but no python captured.
 - b. Same identifier (Object ID) used for multiple entries

3. Analysis Choices and Justification

To address the problem, it was divided into three sub-problems that could be approached by different analytical methods. The sub-problems are as follows:

- A. **Problem 1:** *In each defined region (grid cells), how many snake(s) (measured in snake-feet; total feet of snake) are likely to be found on a specific day given certain environmental factors such as temperature, cloud cover, time of year, weather, water levels?*

- B. **Problem 2:** The second problem is: *Given snake-feet predictions for each grid cell, to which grid cells should certain numbers of hunters be dispatched in order to maximize the snake-feet catch for each day?*
- C. **Problem 3:** *How to communicate results of model outputs to be effective and valuable to decision-makers who may not be familiar with analytic tools or models.*

Because the analysis is based on a specific target variable of interest (snake-feet) and historical data, supervised learning methods are most appropriate for this project to improve hunter dispatching decisions. This project comprises two main analytic problem types: prediction and optimization, both of which are built on initial geospatial analysis of existing data and will be visualized using geospatial analysis. Below is more detail on the specific models.

3.1 Generalized Linear Model

Predict EV of target variable “snake-feet” based on environmental factors such as location, temperature, time of day, and cloud cover.

A Generalized Linear Model (GLM) was selected to analyze the data because the goal is to understand the relationship between snake size and the multiple factors that needed to be incorporated into the analysis in order to make an accurate prediction (see above). A GLM is more appropriate for this data because the features and outcome variable can be connected using a non-linear function, which is more helpful than simpler models which assume linearity and a normal distribution of error terms.

The GLM method is relatively fast to implement so long as the data is clean. Standard programming platforms (e.g., R) are able to run these quickly. These platforms are widely used by organizations and/or free, making this an ideal analytics approach.

The largest area of concern for this project is the data being used and its impact on accuracy. All of the data is manually entered by a variety of people and there are some consistency issues observed. All of the models are trained on this data, so if there are consistent errors that are unknown to the team, then the entire model may be inaccurate. However, due to a teammate’s personal experiences, the data appears to be aligned with what is occurring; this also helped avoid processing errors.

By frequently validating model inputs, design and results with python biologists and hunting program stakeholders, most potential confounding will be identified and added to the model as appropriate.

Two other areas of concern when using regression in this case is overfitting and collinearity. Overfitting may be an issue because all of the training data is based upon catching pythons and there is no training data on unsuccessful sites. So, depending on if the research question ever changed, this model may not produce accurate results. In this specific question, the results will be as accurate as the inputted data.

In regard to collinearity, variables such as cloud cover and temperature may not be independent of one another, causing the model to be inaccurate. This will have to be analyzed during testing and

coefficients may be eliminated depending upon initial findings. Lastly, a final form of inaccuracy that may come from the model is an inference error if this model does make incorrect predictions.

3.2 Optimization

Deterministic (linear programming [LP]) optimization (based on the expected value of target variable, deciding if and how many hunters to send.)

In addition to doing a regression, an LP optimization will take the findings and advances them to a point in which decisions can be made. While the GLM predicts snake-feet considering the variables, the optimization will be able to allocate the appropriate number of hunters based on the prediction. This is possible because there is a set number of hunters. This allows management to assign hunters and maximize the predicted snake-length of each 1 square mile grid unit, or cell. The results of the optimization may show the issue of coverage and/or sampling error in that there may have been some hunters who were not captured in the data; therefore, the locations and number of hunters suggested may not be reflective had more data been contributed to the model. Similarly to the geospatial analysis, the optimization is only as accurate as the GLM outputs and the data that went into the GLM equation.

Although this can be done manually with the current data set with a small amount of a time commitment, that is not the intent long-term. Developing an automated process between this and the other functions would be important, but would require a large amount of technical, coding skill. Ideally, once running successfully it would not need consistent maintenance.

4. Visualizing and Communicating the Analytics

4.1 Geospatial Analysis

ArcGIS Geospatial analysis (Spatial join to classify existing python capture data into grid cells.)

ArcGIS Pro was selected to perform the geospatial analysis because it visualizes the output of the regression and optimization models as well as allows the team to easily develop a “grid” in which to categorize the different areas within South Florida. The visualization is very important because it relays the complex information of the models to the stakeholders and/or decision-makers. This is vital because understanding of the results is essential to making informed decisions, the ultimate goal of any data analytics projects.

ArcGIS is very accurate when provided with latitude and longitude data, as is available in the project. However, the mapping is only as accurate as the output of the regression and optimization models. More on the accuracy of the models is described below. ArcGIS can be very timely in that an internal, automated model within the program can be developed that gathers data (from the database that is holding the python data) and performs the prescribed mapping techniques automatically. Although convenient, the set-up of this model can take some time. Additionally, another constraint of using ArcGIS Pro is the high cost of the program. Small organizations generally do not have many (if any)

licenses, so this analysis may be limited to one person. This limitation may slow a continuous analytics project if there are any issues in this phase of the analysis if it is dependent on one person.

Below we have provided several examples of how we might visualize our data along with brief descriptions of how each one might be used by the decision-maker and stakeholders.

Figure 3 displays the data as a simple density heatmap. While pleasing to look at, this effect communicates some misleading information, e.g., that snakes were captured within a more diffuse zone than they actually were. It is also not particularly helpful for informing project managers about how to best allocate hunters in order to optimize expected catch. For this, we would need to divide the coverage area into smaller units that could be used as a quick reference point for dispatchers.

Figure 3. Density heatmap

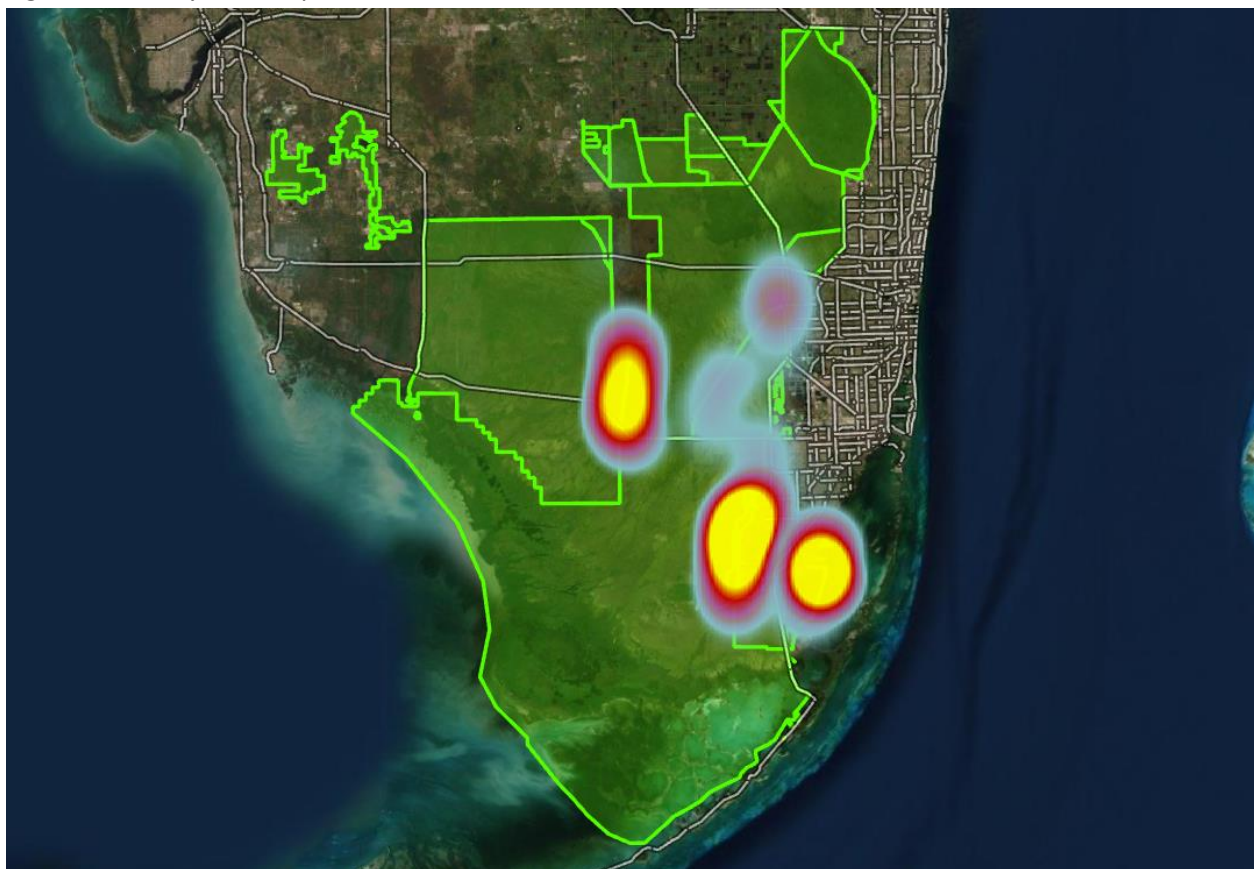
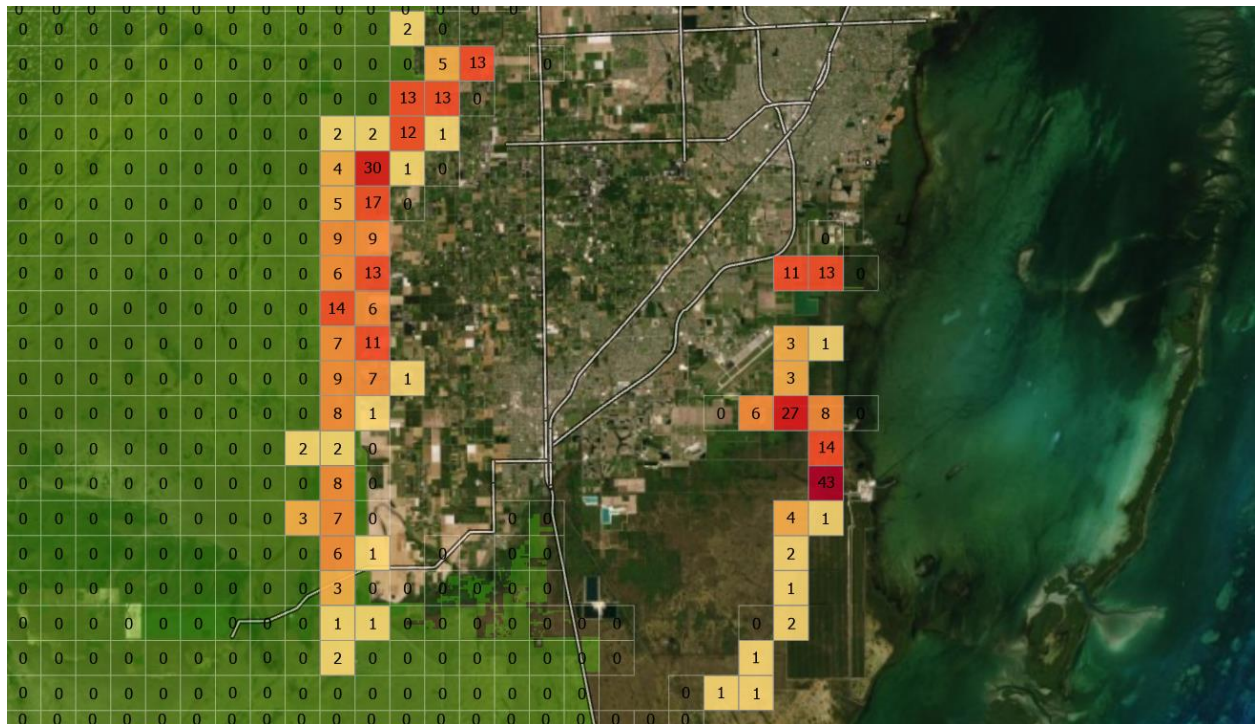


Figure 4 illustrates how the number of python captures varies across the hunting area using a 1-square mile cell-level break down. These values were calculated in ArcGIS Pro by, first, creating a grid overlay feature and mapping it on to the program coverage area (shown in light green), then using the summarize within function to count the number of pythons captured in each cell of the grid (total count is displayed in the center of each cell).

Figure 4. Cell-level scoring by total number of captures using 1 square mile grid system



Another benefit to organizing the data in this way is that it concentrates the heatmap effect to just those areas within each site where snakes were actually captured, down to a 1-square mile of granularity. It also shows more clearly than the previous heatmap how the density of snakes captured varies within areas of high activity. While the entire Rocky Glades area (the vertical area on the left side of the image) is active, the vast majority of snakes captured within this area is restricted to just a few dense pockets, likely representing sites where nests were present (large numbers of captures within a very small area often represent hatchlings that were just starting to emerge from their nest).

The same technique could be used to summarize other attributes of interest, including the total length of pythons captured in each cell (Figure 5-a), total weight, average payout to the hunter (Figure 5-b); or, alternatively, some informative calculated rate, such as catch per unit effort.

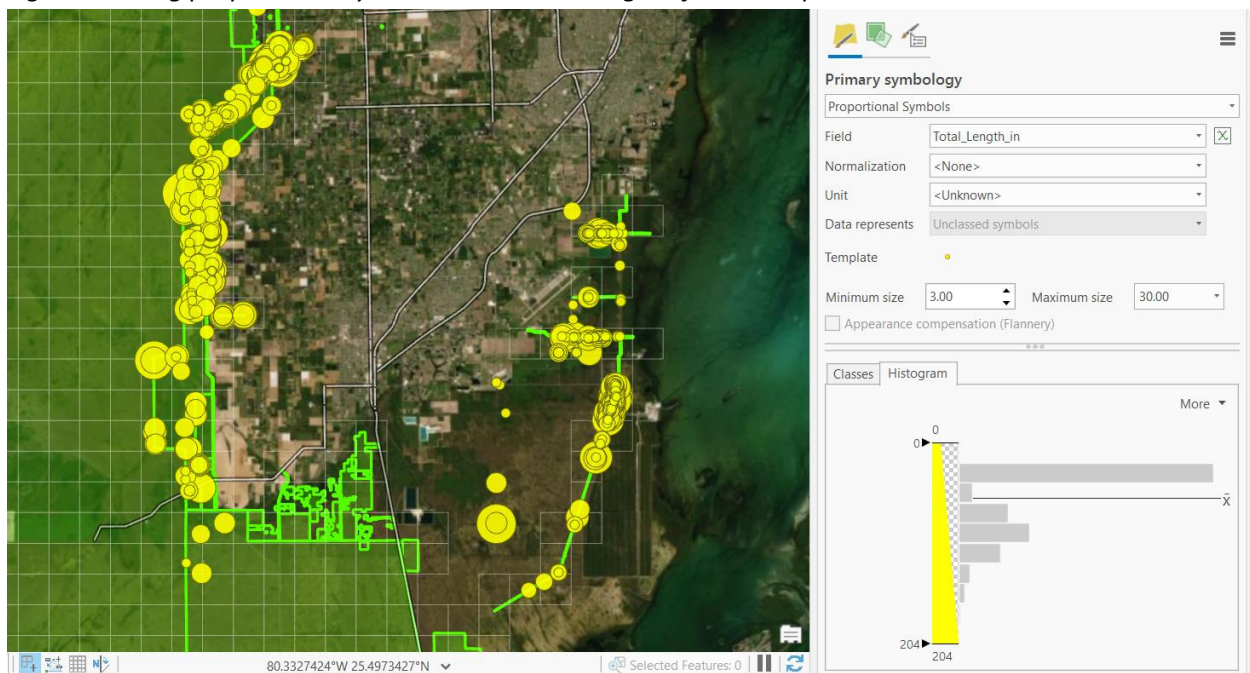
Figures 5-a and 5-b. Cell-level scoring by total length of pythons captured (a) and average payout (b)



Similarly, we could filter the data by date and time of capture in order to show how these values fluctuate over different relevant time spans, e.g., over the course of a day, month, or year. We could even use this approach to show how the size and distribution of the python population has changed since the program began in order to provide a rough measure of the program's impact and effectiveness.

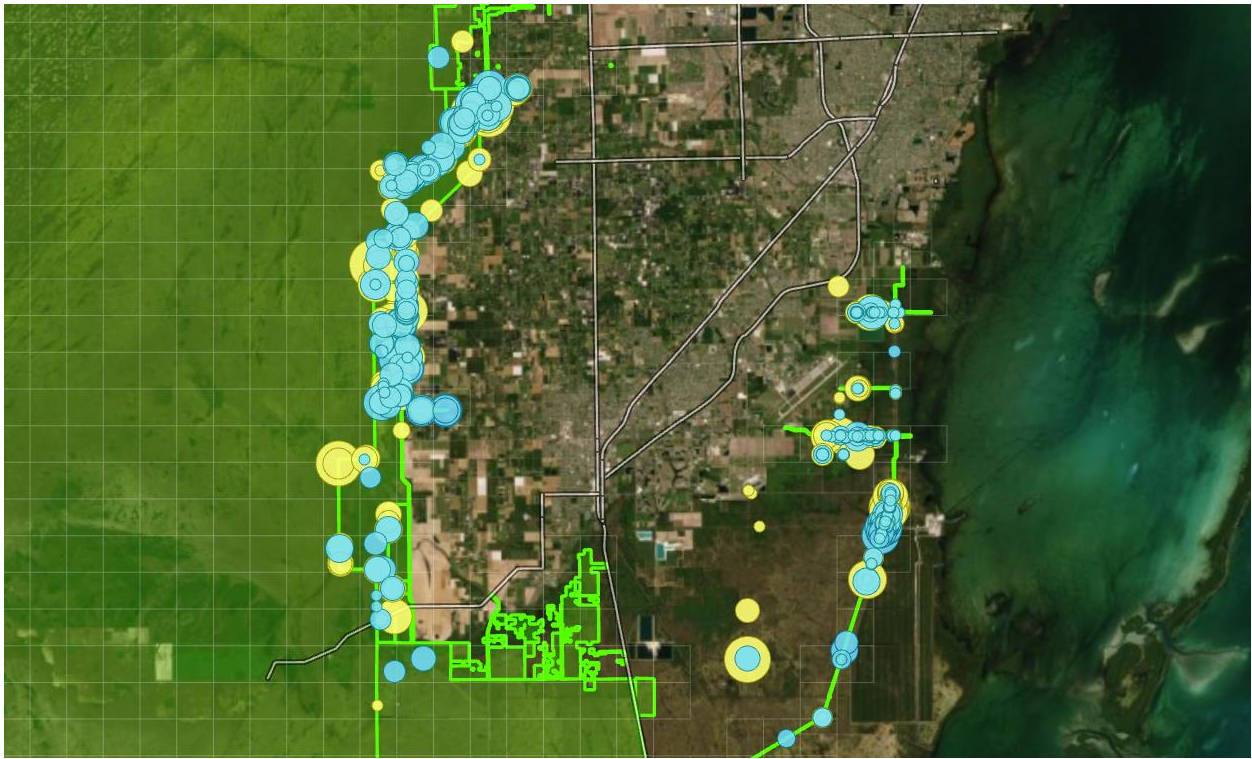
Figure 6 illustrates a further alternative in which proportional symbols are used to visualize the length of the snake captured. In the menu to the right, summary statistics are provided showing the distribution of sizes across the larger population of snakes captured.

Figure 6. Using proportional symbols to visualize length of snake captured



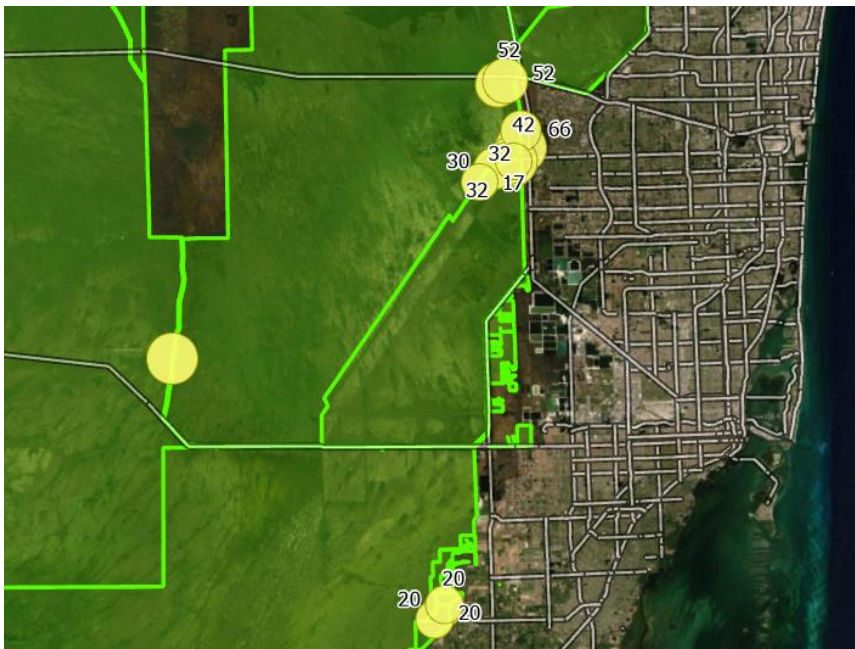
This approach can be combined with additional filters to provide a still finer-detailed representation of the data. Figure 7, for example, shows the same data with additional color coding to distinguish male and female snakes. Using this sort of multi-attribute symbology can be used to reveal interesting correlations between these attributes that may not otherwise be obvious, e.g., that the longest snakes tend to be female and there is less variation in size among male snakes in areas where large female snakes are found, particularly during the breeding season (although visualizing this in our map would require further filtering by date). Discovering patterns like this can help project managers identify the most active breeding grounds, which can be used to target sexually mature snakes and, during the nesting season, gravid females.

Figure 7. Combined proportional symbols and color coding by sex



Filtering the results by whether the snake captured was a gravid female, as we did in Figure 7, reveals a surprising amount of consolidation across the coverage area. The labels represent the number of eggs present. While this data is not exhaustive (females submitted were not consistently checked for the presence of eggs), it nonetheless suggests a pattern that would be of high interest to project managers.

Figure 8. Filtering by gravid females



5. Project Plan

5.1 Stages, Milestones, and Deliverables

The ten week long project will consist of five phases as explained in the previous pages: (1) question and scope formulation; (2) data exploration; (3) modeling; (4) interpretation and findings; and (5) communication of results and recommendations. The following describes the milestones and deliverables associated with each stage:

	Stage 1: Question Formulation
Milestones	<ul style="list-style-type: none">- Stakeholders agreed on a question that clearly (a) states the problem to be solved as a decision (choice among alternative actions) to be improved; (b) identifies the decision-maker; and (c) describes in measurable terms the value of improving the decision.- SFWMD receives and approves project proposal
Deliverables	<ul style="list-style-type: none">- Project proposal for approval by the SFWMD

	Stage 2: Data Exploration
Milestones	<p>Obtained and secured necessary permissions for the input datasets, including SFWMD hunter data (Survey123), NOAA weather and temperature data, DBHYDRO water level data, and SFWMD GIS shapefiles (project coverage area, waterways, roads/levees)</p> <p>Cleaned and merged all data sets (ran summary statistics, merged early-, middle, and late-stage Survey123 data into a single integrated dataset, removed erroneous records, flagged and verified outliers, etc.).</p>
Deliverables	<p>Clean, complete source data sets (QA/QCd)</p> <p>Data documentation (dataset descriptions, metadata, source location, etc.)</p> <p>Data brief with summary statistics</p>

	Stage 3: Data Modeling and Analytics
--	--------------------------------------

Milestones	Develop initial GLM mode Validate Model with ecology experts and field operators Refine model and iterate Develop initial Optimization model based on refined GLM outputs Validate optimization with field operators Refine model and iterate
Deliverables	Final analytical models (GLM and optimization), fully integrated with SFWMD's hardware/software systems Performance evaluation (report)

	Stage 4: Interpretation
Milestones	Model's outputs are translated into insights / recommendations that are understandable and actionable for program manager and other relevant stakeholders (higher management, hunters, field staff, and general public)
Deliverables	Users Guide (including standard interpretations of outputs, coefficients, etc.)

	Stage 5: Communication and Next Steps
Milestones	Deliver the model to the SFWMD / program manager Deploy the model in the field Validated model's superior performance in the field relative to previous methodology, both in terms of predictive power as well as soundness of recommendations and interpretability by managers and field staff/hunters (iterative)
Deliverables	Final report Final models

5.2 Stakeholder Analysis

Below are all of the stakeholders associated with the project and their expected impact and communication frequency:

Stakeholder	Role on the Project or Within the Organization	Contribution to the Project	Project Influence	Communication Plan (Frequency and Method)	Person Responsible
Python hunters	Recipient of	Implements the	High	Bi-weekly	ArcGIS

	project recommendations for implementation	project recommendations and provides valuable insights to logical application of the findings		meetings with five or more hunters (voluntary)	expert, Python developers
South Florida Water Management District Staff	Project sponsor	Guides the direction of the project and is a key part in the iteration process to ensure the output is understandable and usable to the organization	High	Weekly meetings	Project manager
Locator app developers (Esri)	Supplier of existing infrastructure information	Provides information on how data is currently gathered and ensures that this project and the existing infrastructure cooperate harmoniously	Medium	Weekly phone call	DevOps engineer, python developer
Conservationists	Knowledge source to ensure no unintended impacts occur	Provides best practices and/or insights on pythons that may not be captured with data	Low	Bi-weekly phone call	Project manager

5.3 Hardware and Software Resources

- A. Hardware and costs
 - a. Computers for staff: no additional cost (already have)

- b. Hand-held devices to log location and other information: no additional cost (already have - use personal smartphones)
- c. Amazon Web Services (AWS) (assume servers are off-site hosted by third party and not internally maintained): \$500
- B. Software and costs
 - a. SQL database, postgres (through AWS - no big data infrastructure is required): included in above AWS price
 - b. ArcGIS Pro Desktop: \$10,000 for enterprise license (estimation)
 - c. Open source python editor (e.g., PyCharm): free
 - d. ESRI Tracker application: no additional cost (existing)

5.4 Staffing Requirements

The following are the anticipated staffing requirements for the duration of the ten week project:

	Skills Necessary	Years of Experience Required	Time Needed
ArcGIS Integration, Analysis, and Visualization expert	<ul style="list-style-type: none"> - mobile applications - conservation data experience - regression & optimization models - spatial analysis in ArcGIS Pro 	6-8	10 weeks; 20 hours per week
Conservation biologist/python behavior expert	<ul style="list-style-type: none"> - PhD in field related to invasive python mitigation and/or behavior 	1+	7 weeks; 10 hours per week
Database engineer	<ul style="list-style-type: none"> - configuring and maintaining databases 	3-5	10 weeks; 15 hours per week
ESRI Tracker expert	<ul style="list-style-type: none"> - integrating ESRI Tracker app data into analytics models 	2+	10 weeks; 10 hours per week
Data analyst	<ul style="list-style-type: none"> - cleaning and validating data - knowledge of South Florida geography helpful 	1-3	10 weeks; 7 hours per week
Python developer	<ul style="list-style-type: none"> - developing and implementing predictive regression - optimization models in Python 	5+	10 weeks; 25 hours per week

	- experience coding in Python for ArcGIS		
DevOps Engineer	- managing development of multi platform analytics - working with existing infrastructure and systems (i.e. government)	3+	10 weeks; 11 hours per week
Project Manager	- managing development and analytics projects in a government setting - familiarity with python capture program and staff	3+	10 weeks; 20 hours per week

5.6 Risk Management

Below are the anticipated risks associated with the project. The scale is as follows: Likelihood Level and Severity: 1 = High, 2 = Medium, 3 = Low

Likelihood	Description of Risk	Severity	Response	Responsibility
1	Data quality: python capture data is reported from the field by hunters via an application which has been modified several times in the past few years. Data not related to snake bounty payments (like weather conditions and location) may be incorrect or inconsistent year-to-year which would impact the validity of any model based on that day, and may be difficult to detect.	1	Weather conditions reported via the app can be cross referenced against NOAA weather station data to determine likelihood of accuracy. If necessary, user-reported weather data could be replaced with NOAA validated weather data	Data analyst
1	Project rejection: Python hunters may feel resentment about being directed to different areas based on computer models instead of their own preferences, instincts or field knowledge.	1	Extensive stakeholder engagement must happen throughout all stages of the process to: - Gain hunter trust - Leverage hunters' input into the model	Project manager

			- Explain the value of the improved decision to the goals of the program and hunters overall (i.e. more snakes caught and more bounties paid)	
3	Staff turnover: staff members may leave the organization at any time and their tasks may be left unassigned and incomplete.	3	In the event any staff member from the team leaves the organization, the project will be set up with transparency and crossover to ensure transfer knowledge is smooth and fast. Additionally, having a clause in the project contract regarding extensions due to employees leaving the company would allow for additional working days to complete the project.	Project manager
2	Modeling errors: the generalized linear model has a potential for error due to collinearity of the data. Although there are methods in place to lessen the degree of this error, it may still occur.	1	Reiterate and test the model design multiple times in order to detect and significant errors and debug.	DevOps engineer, python developers

6. Success Measurements and Project Impact

Our model is designed to improve the SFWMD's invasive species management efforts in and around Everglades and Big Cypress National Parks. The impact of our solution will therefore be measured largely in terms of how well it helps to control the size and spread of the Burmese python population within the project coverage area. Relevant metrics for measuring this impact include:

1. Greater number and total length of Burmese pythons removed on average over a sufficient time horizon
2. Increase in number of native mammal and bird observations in impacted areas, as reported in routine biological surveys

Because the SFWMD is a government entity reliant on public funding, it is also essential that the use of the model is not perceived to conflict in any way with the public's interests, or otherwise threaten the support the program currently enjoys from the political establishment, especially the current governor, Ron DeSantis. These impacts can be measured by such metrics as the amount of State funding invested in the program, the frequency of media coverage of the program, and the quality of the public's response to this coverage.

Finally, the District's use of the model has to win the approval of the hunters themselves. As contract workers with relatively low pay and no benefits, much of the success of the python elimination program to date owes to the enthusiasm and strong morale of the hunters. In order to measure the impact of our solution on the hunters, we should also measure the following metrics:

1. Average payout for hunters
2. Distribution of risk and opportunity across hunters, resulting in more uniform per unit effort earnings across team
3. Number of conflicts between hunters due to territorial disputes and bottlenecks
4. Hunters' rate of adherence to dispatcher's assignments
5. Hunter attrition and application rate for open positions

The criteria for determining when the project is complete include:

- A. Model performance (predictive accuracy, quality of recommendations): significantly better than previous solution over a reasonable time horizon, as measured by an increase in total snake feet removed (greater than 8,340 as caught in the previous year).¹¹ Detection rates using traditional methods (including baited traps, dogs, and visual survey) are estimated to be less than 1%,¹² and this tool doesn't do anything specifically to improve that rate. However, by directing hunters to the most population dense and active hunting locations within the coverage area, the expected catch should still be significantly higher than that achieved by the current method.
- B. Client satisfaction: measured by decision-makers' degree of adherence (at least 80% adherence) to model's recommendations and renewal of contract (if required)

7. Lessons Learned

During the course of this investigation, we learned the following lessons:

- A staff with a diverse set of skills is essential to this project, and likely many others. Without the varying skills this project would not be successful. (As demonstrated by this team's lack of analytics and/or programming skills.)

¹¹ "Python Elimination Program," South Florida Water Management District, accessed February 4, 2020, <https://www.sfwmd.gov/our-work/python-program>.

¹² Sophia C. M. Orzechowski, Peter C. Frederick, Robert M. Dorazio, Margaret E. Hunter, "Environmental DNA sampling reveals high occupancy rates of invasive Burmese pythons at wading bird breeding aggregations in the central Everglades," *PLOS One*, April 10, 2019.

- Stakeholders have the capability to make the best analytics projects unsuccessful due to unwillingness to implement. Especially in this project, the python hunters held the most influence in that they would choose not to use the recommendations. This stressed the importance of including stakeholders consistently throughout all stages of a project.

8. Limitations and Future Research

- The District probably needs to clean up its data management protocol. Three years in, the data is still spread out across 4 separate Excel files all of which used different attributes and data collection protocols (due to updates in the Survey123 e-form). No metadata is provided to explain these features of the datasets, and a number of confusing conventions have been applied for various ad hoc purposes (e.g., color coding, filtering, etc.). There is also no paper trail to refer to if the data files become corrupted, which doesn't seem particularly improbable.
- A lot of potentially useful data was lost by poor choices in software for the early stages of the program, particularly related to compatibility between software systems. One specific example is the incompatibility of the LaborSync app, which was used to track hunters' movements through the coverage area, and ArcGIS software, which was used to map captures. As the District considers integration of new software (e.g., the Esri Tracker app recommended by this proposal), they should give more careful consideration to other possible compatibility issues with legacy systems.

Bibliography

- EDDMapS. 2020. Early Detection & Distribution Mapping System. The University of Georgia - Center for Invasive Species and Ecosystem Health. Accessed January 27, 2020. <http://www.eddmaps.org/>.
- Kaley Fournier, Edward Hines, and Nicholas Stevenson. "Invasive Burmese Pythons Eat Their Way through Southern Florida: The Unexpected Effect on Our Health. – Debating Science." 2018.
- Mariya Yao, Marlene Jia, and Adelyn Zhou. *Applied Artificial Intelligence: A Handbook for Business Leaders*. Edited by Natalia Zhang. TOPBOTS Inc., 2018.
- Michael E. Dorcas, John D. Willson, Robert N. Reed, Ray W. Snow, Michael R. Rochford, Melissa A. Miller, Walter E. Meshaka, Paul T. Andreadis, Frank J. Mazzotti, Christina M. Romagosa, Kristen M. Hart. "Severe mammal declines coincide with proliferation of invasive Burmese pythons in Everglades National Park." *Proceedings of the National Academy of Sciences*. Feb 2012, 109 (7) 2418-2422.
- Miller, Melissa, Jenny Ketterlin, and Mike Kirkland. "Python Removal Presentations." 2018.
- "National Centers for Environmental Information." National Oceanic and Atmospheric Administration (NOAA). Accessed February 5, 2020. <https://www.ncdc.noaa.gov/>.
- Nishihara, Robert & Moritz, Philipp & Wang, Stephanie & Tumanov, Alexey & Paul, William & Schleier-Smith, Johann & Liaw, Richard & Jordan, Michael & Stoica, Ion. *Real-Time Machine Learning: The Missing Pieces*. 2017.
- "Python Elimination Program." South Florida Water Management District. Accessed February 4, 2020. <https://www.sfwmd.gov/our-work/python-program>.
- Sophia C. M. Orzechowski, Peter C. Frederick, Robert M. Dorazio, Margaret E. Hunter. "Environmental DNA sampling reveals high occupancy rates of invasive Burmese pythons at wading bird breeding aggregations in the central Everglades." *PLOS One*. April 10, 2019.
- "Water Levels." South Florida Water Management District. Accessed January, 27, 2020. <https://www.sfwmd.gov/science-data/levels>.