Course Project

2022-11-25

In 2020, for the SAS EMEA Hackathon, Amesto NextBridge, a business solutions firm, and Beefutures, a beekeeping-focused biotechnology company, both based in Norway, partnered to model bee waggle movements and optimize locations for bee hives to support bee population preservation. They found two types of waggle movements could indicate bees communicating flower locations, which are food sources, to their hive-mates. However, due to many factors, notably monoculture farming, which is where farmers farm one type of plant and reduce much-needed biodiversity in the area, bee populations seemed to be be in decline, which could be detrimental to human food sources and wellbeing.

In order to attempt to intervene in the bee population decline by optimizing bee hive locations, they had to employ a number of different analytic modeling techniques to reach their goal or optimizing bee hive locations according to bee waggle movements and pollinating plant type locations. A potential pathway to this analytics solution would be using time series data to model trends in bee populations over time, a regression model for how different environmental and societal factors affect the probability of bee population growth/decline, another regression model to demonstrate the relationship between different flower locations and bee waggle movements, and lastly an optimization model to place bee hives in the best locations according to their waggle movements and the available plant life for pollination.

Bee populations have been noticeably, but slowly, declining since the early 2000s, which is what inspired this collaboration. To establish this trend before moving forward with analytic solutions to the problem, the team could use a seasonal ARIMA model to model the bee populations over time. They would use a seasonal ARIMA model to take into consideration the different seasons, which could also affect bee populations given that different seasons may bring about different plant life, as well as weather patterns and other factors such as potential tourism that could affect human population at certain points of the year which could in turn impact bees and how they function and thrive on a day-to-day basis. The data needed for this model would be the bee count across each year from 1980 to 2020, to account for the baseline number of bees before the noticeable decline began. Bee population data is collected by beekeepers by way of estimation, using the average count of bees leaving the hive, average number of flights per bee per day, and the amount of time the bees spend foraging, as this can help calculate the number of bees in a colony, since the number of foragers is typically a third of the total population. This estimate of hive populations can be used in the model. This model could be updated annually to track the bee population increases and decreases over time, as well as to see how different potential interventions may be affecting the trends moving forward.

Table 1. Establishing Trend in Bee Population Over Time

Given	Bee population data from 1980-2020
Use	Seasonal ARIMA
To	Model bee population increases and decreases over time with consideration for seasonalities

The next model would be a logistic regression model. Some data that would be useful for this model would be different environmental and societal factors, such as tourism counts, percentage of land farmed and the plants that are farmed, is the farming monoculture (meaning they only plant one type of plant) or not, distance from hives to flowers, and bee population counts. After going through the data and performing scaling and imputation as necessary, they could fit the data to a stepwise regression model in order to find the significant variables that affect the bee populations, and from there, they could fit the selected variables to the logistic regression model and be able to investigate the probabilities of the bee hives continuing to thrive given the predictors contributing to the model. This model would give information to the team about which hives are most at risk, as well as which factors are most affecting the bee populations. This model could continue to be updated periodically by watching the previously-made time series model, so if and when a new trend is noticed in the time series model, this model could be updated to check how the factors are correlating to the population, as well as the probabilities of bee populations thriving. Since hive location (distance from hives to food sources) would be a factor in this model, this could also be updated as hives are relocated with the intervention to optimize hive locations in order to check how the relocations are affecting survival probabilities.

Table 2. Factors Affecting Bee Populations

Given	Tourism counts, percentage of land farmed and the plants that are farmed, is the farming monoculture (meaning they only plant one type of plant) or not, distance from hives to flowers, and bee population count (along with other possible contributing local variables)
Use	Stepwise Regression, Logistic Regression
To	Model probability of bee populations increasing/decreasing as a result of environmental and societal factors

Before they could begin to optimize the hive locations, they would need to first look at the relationship between the two different waggle movements and the flower locations that those bees seem to prefer/need for survival. To do this, they could perform a principal component analysis to look at how each of the variables correlate with one another. Variables could include waggle movement type, bee death rate, bee birth rate, hive population, hive location, flower type, flower location, and distance between hive and flowers. Performing the PCA would allow the investigators to see which variables in the principal components are most closely correlated in order to help in selecting variables for two different linear regression models. Each waggle movement would have its own linear regression model and add in the variables that are most closely correlated with that specific waggle movement in the PCA in order to look at the relationship. From here, they would be able to know which bees do which waggle movement most frequently and thrive the most with which type of flower and at which distance as well.

Table 3. Flower Location and Bee Waggle Movement Correlation

Given	Waggle movement type, bee death rate, bee birth rate, hive population, hive location, flower type, flower location, and distance between hive and flowers
Use	Principal Component Analysis, Linear Regression
To	Model correlation between flower locations and bee waggle movements

Finally, they could use the correlations yielded from the linear regression model and plug those into a stochastic optimization model to model the optimal bee hive locations with regards to preferred flower type based on waggle movements. Some data that would be needed to perform the optimization would be bee hive location, flower location, bee waggle movement type, flower type, bee movement and flower type correlation, and likelihood for bees to survive. Some possible constraints to add into the model could include requiring a minimum distance between hives or requiring a minimum/maximum distance between the hive and the flowers. All of these factors plugged into the model, alongside the stochastic element accounting for any randomness and unpredictability since nature can be very unpredictable, could be used to help the investigators figure which bee hives should be how close to which flowers in order to encourage bee population growth.

Table 4. Hive Location Optimization

Given	Bee hive location, flower location, bee waggle movement type, flower type, bee movement and flower type correlation, and likelihood for bees to survive
Use	Stochastic Optimization Model
To	Model optimal bee hive locations with regards to preferred flower type based on waggle movements

From there, the investigators could update the models and rerun annually in order to check bee population changes and monitor the effect of the intervention in place. Encouragingly, most of the data is already available or easy to gather, as farming land and weather patterns are public data. The investigators also already keep up with bees and their waggle movements and populations by way of in-hive cameras and historical data, so this would all be convenient for building these models in order to help increase sustainability of bee populations.

References:

- https://www.sas.com/content/sascom/en_us/customers/beefutures.html (https://www.sas.com/content/sascom/en_us/customers/beefutures.html)
- https://www.buzzaboutbees.net/how-many-bees-in-a-hive.html (https://www.buzzaboutbees.net/how-many-bees-in-a-hive.html)