Syracuse University

Memo

|  |  |
| --- | --- |
| To: | Dr. Landowski |
| From: | Katie Haugh, Mark Roberts, Sandy Spicer |
| Date: | May 11, 2021 |
| Re: | Project Proposal |

**Topic: Analyze the Central Park Squirrel Census**

The Squirrel Census, an Atlanta-based organization with a mission to count and acknowledge the Eastern gray squirrel, set out to expand their mission to the Big Apple. In 2018, an 11-day squirrel census was conducted in Central Park in New York City. The organization engaged 323 volunteers who surveyed the 843-acre park for squirrels. This project presents the survey results.

**Data Description:**  
Two data sets are included for analysis.

*2018\_Central\_Park\_Squirrel\_Census\_-\_Squirrel\_Data.csv*

The first is a .csv file containing the location, age, color, and other characteristics of 3,023 squirrels found in Central Park. The variables include:

|  |  |  |  |
| --- | --- | --- | --- |
| Fields | Description | Type | Example |
| X | Longitude coordinate for squirrel sighting point | Number | -73.95613449 |
| Y | Latitude coordinate for squirrel sighting point | Number | 40.79408239 |
| Unique Squirrel ID | Identification tag for each squirrel sightings. The tag is comprised of "Hectare ID" + "Shift" + "Date" + "Hectare Squirrel Number." | Plain Text | 37F-PM-1014-03 |
| Hectare | ID tag, which is derived from the hectare grid used to divide and count the park area. One axis that runs predominantly north-to-south is numerical (1-42), and the axis that runs predominantly east-to-west is Roman characters (A-I). | Plain Text | 37F |
| Shift | Value is either "AM" or "PM," to communicate whether or not the sighting session occurred in the morning or late afternoon. | Plain Text | PM |
| Date | Concatenation of the sighting session day and month. | Plain Text | 10142018 |
| Hectare Squirrel Number | Number within the chronological sequence of squirrel sightings for a discrete sighting session. | Number | 3 |
| Age | Value is either "Adult" or "Juvenile." | Plain Text | Adult |
| Primary Fur Color | Value is either "Gray," "Cinnamon" or "Black." | Plain Text | Gray |
| Highlight Fur Color | Discrete value or string values comprised of "Gray," "Cinnamon" or "Black." | Plain Text | Black |
| Combination of Primary and Highlight Color | A combination of the previous two columns; this column gives the total permutations of primary and highlight colors observed. | Plain Text | + |
| Color notes | Sighters occasionally added commentary on the squirrel fur conditions. These notes are provided here. | Plain Text | Nothing selected as Primary. Gray selected as Highlights. Made executive adjustments. |
| Location | Value is either "Ground Plane" or "Above Ground." Sighters were instructed to indicate the location of where the squirrel was when first sighted. | Plain Text | Ground Plane |
| Above Ground Sighter Measurement | For squirrel sightings on the ground plane, fields were populated with a value of “FALSE.” | Plain Text | 10 |
| Specific Location | Sighters occasionally added commentary on the squirrel location. These notes are provided here. | Plain Text | Branch |
| Running | Squirrel was seen running. | Checkbox | TRUE |
| Chasing | Squirrel was seen chasing another squirrel. | Checkbox | TRUE |
| Climbing | Squirrel was seen climbing a tree or other environmental landmark. | Checkbox | FALSE |
| Eating | Squirrel was seen eating. | Checkbox | FALSE |
| Foraging | Squirrel was seen foraging for food. | Checkbox | FALSE |
| Other Activities |  | Plain Text | wrestling with mother |
| Kuks | Squirrel was heard kukking, a chirpy vocal communication used for a variety of reasons. | Checkbox | TRUE |
| Quaas | Squirrel was heard quaaing, an elongated vocal communication which can indicate the presence of a ground predator such as a dog. | Checkbox | FALSE |
| Moans | Squirrel was heard moaning, a high-pitched vocal communication which can indicate the presence of an air predator such as a hawk. | Checkbox | FALSE |
| Tail flags | Squirrel was seen flagging its tail. Flagging is a whipping motion used to exaggerate squirrel's size and confuse rivals or predators. Looks as if the squirrel is scribbling with tail into the air. | Checkbox | TRUE |
| Tail twitches | Squirrel was seen twitching its tail. Looks like a wave running through the tail, like a breakdancer doing the arm wave. Often used to communicate interest, curiosity. | Checkbox | FALSE |
| Approaches | Squirrel was seen approaching human, seeking food. | Checkbox | FALSE |
| Indifferent | Squirrel was indifferent to human presence. | Checkbox | TRUE |
| Runs from | Squirrel was seen running from humans, seeing them as a threat. | Checkbox | FALSE |
| Other Interactions | Sighter notes on other types of interactions between squirrels and humans. | Plain Text | ran from dog-walker |
| Lat/Long | Latitude and longitude | Point | POINT (-73.9561344937861 40.7940823884086) |

Source: <https://data.cityofnewyork.us/Environment/2018-Central-Park-Squirrel-Census-Squirrel-Data/vfnx-vebw>

*2018\_Central\_Park\_Squirrel\_Census\_-\_Hectare\_Data.csv*

The second data set includes location, weather, and environmental information on the 350 hectares of Central Park during the census:

|  |  |  |  |
| --- | --- | --- | --- |
| Fields | Description | Type | Example |
| Hectare | ID tag, which is derived from the hectare grid used to divide and count the park area. One axis that runs predominantly north-to-south is numerical (1-42), and the axis that runs predominantly east-to-west | Plain Text | 01A |
| Shift | Value is either "AM" or "PM," to communicate whether the sighting session occurred in the morning or late afternoon. | Plain Text | AM |
| Date | Concatenation of the sighting session month, day, and year (MMDDYYYY). | Plain Text | 10072018 |
| Anonymized Sighter | Sighter names were alphabetized then assigned a number (1-267) so that the individual abilities and characteristics of each sighter could be taken into account. | Number | 110 |
| Sighter Observed Weather Data | Each sighter was asked to describe the weather conditions before starting each sighting session. Weather values varied widely, and are in some cases inaccurate. | Plain Text | 70¬∫ F, Foggy |
| Litter | Value can be "None", "Some" or "Abundant." | Plain Text | Some |
| Litter Notes | Sighters occasionally added commentary on the amount or quality of litter. These notes are provided here. | Plain Text |  |
| Other Animal Sightings | Comma separated values indicating other animals sighted in the hectare. A list of other animals common to the park were provided; however, other animals were listed as well. | Plain Text | Humans, Pigeons |
| Hectare Conditions | Value is one of the following: “Calm”, “Moderate”, “Calm, busy” or “Busy.” The difference between “moderate” and “calm, busy” is as follows: “moderate” was either entered by the sighter, or translated from an entry that described a condition somewhere in-between busy and calm. “Calm, busy” was either entered by the sighter or translated from an entry that described discrete hectare sections as busy, but other hectare areas as separately calm. | Plain Text | Busy |
| Hectare Conditions Notes | Sighters occasionally added commentary on the hectare conditions. These notes are provided here. | Plain Text | Construction |
| Number of sighters | The number of sighters that observed the hectare for the sighting session. Values range from "1" to "3." | Number | 1 |
| Number of Squirrels | The number of squirrels sighted in the hectare during the sighting session. | Number | 4 |
| Total Time of Sighting | The sighting session duration, in minutes. | Number | 22 |

Source: <https://data.cityofnewyork.us/Environment/2018-Central-Park-Squirrel-Census-Hectare-Data/ej9h-v6g2>

**Research Questions:**

* What are the primary visual characteristics of the squirrels? What color combinations were most popular? Were squirrels of similar markings found in a particular area of the park?
* Is there a relationship between the number of squirrels seen and the time of day? Does the weather at the time have any influence?
* Is there a section of the park that tends to have more squirrels? What are the characteristics of those hectares? How busy were those areas? Were other animals present?
* What verbal noises were most popular? Were the noises correlated with the presence of any other animals in the area? How does this correlate with their tail movements?
* What were the most common interactions with humans?

**Data Preparation Plan**

1. Load the data into Python.
2. Clean the data:
   1. Remove columns that are unnecessary for our analysis.
   2. Replace missing values that can be replaced.
   3. Normalize columns (several columns contain more than one entry per observation, so normalizing those variables is important).
   4. Transform fur color columns to account for those squirrels that were only one color.
   5. Separate out temperature from the weather conditions (sunny/cloudy) in the ‘Sighter Observed Weather Data’ column.
3. Observe the descriptive statistics of the data set to check for outliers and distribution.