**Applied Data Science Capstone: *Ontario Hunting Data Visualization***

**Introduction:** In my home province of Ontario, Canada the traditional pastime of hunting has seen a significant decrease in participation over the past few decades. Apart from providing affordable sustenance for participants, hunting plays a highly impactful role in the conservation and management of the various animals that inhabit Ontario. Through statistical analysis and careful research, the Ontario Ministry of Natural Resources publishes specifically crafted figures for the number of different game animals that should be harvested each season towards ensuring the sustainability of these populations and thus conservation of the fragile ecosystems which they inhabit.

If one wildlife species was to become too highly populated in comparison to those species who exist both above and below it on the food chain, there can be significant side effects which are felt in several different ways. As the population of one species swells beyond manageable numbers, it will often consume its food source at an unsustainable rate frequently resulting in animals wandering into unnatural regions where they inadvertently cause property damage and pose risk to human life. To maintain a population level that most effectively balances the needs of an ecosystems at many different levels of the food chain, environmental conservation organizations such as the Ontario Ministry of Natural Resources rely on activities such as hunting and fishing as a powerful conservation tool to guide the natural world towards the most equitable outcomes.

With the gradual decline in popularity of activities such as hunting and fishing, this important tool for the conservation of the natural world becomes far less effective. Amongst several reasons suggested as being behind this decline is the vast complexity of regulations that are imposed on these types of activities. The same careful research and statistical inference that have resulted in hunting being an incredibly effective force towards conservation, has also produced a considerable barrier to entry as Ontario has been segmented into ~150 different “wildlife management units” (WMU) , each with different regulations and season dates for different game animals, once again differing in rules and regulations according to province of origin of a participant.

To complete my capstone project, I will perform data mining to determine the exact geographic boundaries for each of these WMU’s, aggregate the many regulations and season dates through the creation of automation software, in order to perform data integration bringing these disparate sources of information into one unified dataset. Once this dataset is completion, I will create a unique data visualization map to empower a prospective hunter to quickly and effectively obtain all the information that they would need to adopt the hobby of hunting in my home province.

**Data Sources:**

**1. Geographic Regions for WMU’s:** The Ministry of Natural resources posts maps of the various boundaries that form the wildlife management units, but these are not available in a raw format for this application. Through use of the “ArcGIS” geographic dataset tool I was able to extract the layers that represent these regions in a raw format as a ‘.SHP’ shapefile.

|  |
| --- |
|  |
| **WILDLIFE\_MGMT\_UNIT.shp**  (available: https://www.arcgis.com/home/item.html?id=b5e6b1be4e22418fb52c506138dac7e8) |

**2. Dates & Regulations for each WMU:** the open season dates for each WMU are typically published in a physical catalog however there is also a digital distribution of the document in ‘.pdf’ format. Rendering this online document through a web browser will enable us to perform data mining using python libraries such as Selenium and BeautifulSoup much the same way as it was performed in the ‘Toronto Neighbourhood Clustering’ assignment.

|  |
| --- |
|  |
| **mnrf-2019-hunting-regulations-en-04-0-2019.pdf**  (available: https://files.ontario.ca/mnrf-2019-hunting-regulations-en-04-0-2019.pdf) |

**3. POI’s via FourSquare API:** Using the FourSquare API through python we can obtain data from a massive volume of venues that are available through the service. For the purposes of this assignment the venues search call with the ‘browse’ parameter will be implemented to determine many results that center around the areas passed to the API.

|  |
| --- |
|  |
| **FourSquare API call & result set**  (available: https://developer.foursquare.com/) |

**Methodology:**

**1. Working with Shapefile** **& Converting to GeoJSON:** While I was able to obtain some geographical representation of the WMU locations that segment the province they were in a format that is not compatible with Folium, the geographical visualization library that I had intended to work with. The WMU layer was in the form of a ‘shapefiles’ (.shp) which is a geospatial vector data format for GIS software. While being a commonly used file format in many mapping applications in its current form it would not be suitable for this project.

Using the open source program ‘ogr2ogr’ provided under the Geospatial Data Abstraction Library project (GDAL) I was able to wrangle this proprietary geographic information file into a ‘GeoJSON’ format which is a JSON object adhering to specific formatting restrictions containing a series of longitude / latitude coordinates that would be easier to manipulate with folium.

|  |  |
| --- | --- |
|  |  |
| **Visual Representation of shapefile (.shp)** | **GeoJSON Object** |

**2. Data Mining for Open Season Dates & Other Regulations:** Once I had the base layer for my visualization the next step was to gather the data for each of the map segments. The primary source of data was the ‘2019 Hunting Regulations Summary’ document available in several forms from the Ontario Ministry of the Environment website. From these online sources I was able to extract the required data through software automation and populated a ‘.csv’ file with the raw information. After reviewing the results and performing minor *data cleansing* (largely in the form of removing whitespaces before and after certain cell values) I was able to review the tabular record set to ensure that each region of the province was accounted for.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Raw tables of data available online:** | | **Organized tabular data in ‘.csv’ format.** |

**3. Additional Points of Interest using FourSquare API:** The final component of the visualization required some way to determine both the street address and folium-friendly latitude / longitude coordinates for many important POI’s to hunters. The FourSquare service providing this functionality through calls to the API with the ‘explore’ parameter set which allows the service to return venues close to a certain point based upon some supplied keyword. I decided that the most prudent point of interest for this map would be the locations of ‘ServiceOntario’ branches which are the locations where things such as permits, and hunting tags are acquired. Through a series of API calls supplied with coordinates all over the province I was able to efficiently populate a dataframe of useful locations that could be plotted on the map through the folium ‘Marker’ object.

|  |
| --- |
|  |
| **Dataframe (pandas) containing the results of FourSquare API calls:** |

**Results:** Now that I had obtained all required data and had organized it into formats that would be compatible with the folium library, I could begin creating the data visualization. Starting with the ‘OpenStreetMap’ base map I was able to plot the WMU’s using the folium ‘GeoJson’ module. Once this layer had been successfully applied, I was able to read in the ‘.csv’ data containing the season dates and regulations and created a mechanism to populate custom HTML segments for each individual WMU. Once this functionality had been tested, I bound the data to the main map through use of the folium ‘Popup’ module.

Next, I created an additional layer which would contain the points of interest provided through the FourSquare API. Using a list of coordinates from different regions in the province I was able to make many API calls to populate a dataframe with relevant points of interest from all stretches of Ontario. Using this dataframe I was able to then obtain the street addresses through further API calls and was able to add these records to the map as markers. Once the data was all added to the map, I tweaked the design slightly for ease of usage and added an external module to enable the ‘find my location’ button which allows users to zoom to their location when viewing the map.

|  |  |
| --- | --- |
|  |  |
| **Open Season Dates & Regulations:** | **Points of Interest from the User’s Location:** |

**Discussion:** The completion of this assignment primarily highlighted just how valuable it is to have access to data in raw and versatile forms such as JSON objects. The most difficult part in creating this visualization was to *wrangle* the data from proprietary formats to more ‘raw’ forms that were easily digestible by programming languages. The ease of creating this map once the data had successfully been converted into a more useful form leads me to the suggestion that regulatory agencies and conservation groups that outline the rules for each season should make this data accessible in a raw format as the world continues to become a more digitally connected place. While it may have been tradition to mark important dates like these into a physical calendar in the past, most modern-day hunters likely maintain some form of digital calendar such as Google Calendar or the iCloud calendar provided by Apple which would lend to raw forms of the information being far more useful, (and environmentally friendly).

**Conclusion:** This assignment provided me with a strong introduction to the geospatial information systems side of data science and allowed me to further explore the many modules of the folium python library. I found that I was able to create an attractive and informative representation of these key provincial regulations which would lend itself to more effectively informing a prospective hunter of the pieces of information that they would require to adopt the hobby.

Moving forward features that I intend to add would be additional tables under each WMU popup to summarize any additional regulations and important points of contact that may apply to these specific regions. Through this introduction to the GitHub version control software and its integration with the IBM Watson Studio Jupyter Notebook I can retain this project in perpetuity so if I wish to update it for successive years or with additional features it remains readily available.

**References & External Resources Used:**

**[1] ArcGIS Online** – available: <https://www.esri.com/en-us/arcgis/products/arcgis-online/overview>

**[2] ogr2ogr (GDAL module)** – available: <https://gdal.org/programs/ogr2ogr.html>

**[3] geopy (Python 3 library)** – available: <https://geopy.readthedocs.io/en/stable/>

**[4] BeautifulSoup (Python 3 library)** – available: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

**[5] FourSquare API client (Python 3 library)** – available: <https://github.com/mLewisLogic/foursquare>