Final Exam – Go Forth and Do Data Science – Proposal Document

12/10/17 Austin Derbique A01967241

Introduction

A drone is an unmanned aerial vehicle (UAV) that have multiple rotors on them. They have become increasingly popular in the last 10 years with the improvement of Lithium batteries. This has caused a surge in ownership It is expected that the Drone market will be worth nearly $127,000,000,000 in 2020. With this much money in the playing field, it is important to use data science to make educated guesses on drone related topics to successfully compete in the market.

DJI Mavic Pro Drone

Purpose

For this final exam, I have decided to study the topic of drones. Specifically, I will be looking at the drone registry supplied by the FAA to perform analysis on ownership characteristics. The purpose of this is to analyze characteristics ranging from what type of people own drones, where are the most registrations, what is the most popular model, etc.

Source of Information. http://registry.faa.gov/aircraftinquiry/Aircraft_Inquiry.aspx

The FAA provides valuable knowledge of drone ownership information and makes it publicly available online. This includes the the name of the owner, where they live, serial and model number of the drone, and other details relevant to the matter. These files can be viewed online from the browser or downloaded as a CSV file.

Plan of Attack/ Methodology

Data Gathering. The website supports collecting data by a multitude of different options. The easiest option is to collect data for the past 30 days. This downloads an entire CSV file. The file for November and December contains about 3200 entries. Another option would be to iterate through all states and counties and download information that way. This would generate a CSV file for every county in the United States. This would be obtainable through a python script and some libraries out there that contain that necessary information.

Data Cleaning. Currently, the fields are as follows:

A. N Number

B. Manufacturer Name
C. Model Name

D. Certification Issue Date

E. Name of Owner

F. Street Address

G. City

H. State

I. Country

J. Zip Code

The data is already in a CSV file so that makes it easy. The cleaning would come into play for picking the data that is relevant to the particular data science you are trying to accomplish.

Data Processing. Using the different approaches learned in class, there is a variety of ways to process this data. More data processing could be done if outside information is pulled in such as populations of the counties. From here, you could see the highest ownership per capita, etc.

Results and Visualizations. Most importantly, it would be cool to see all registries placed on a map using the address given for each entry. This would show where most drones are registered. Additionally, you could also show the following:

the most popular drone in each state

average registries per user

percentage of Individual, commercial, government

Drone accidents compared to registries

Conclusion

There are many ways going about organizing the data available. This will surely be beneficial for better understanding drone ownership and give better predictions for the future on purchases. Attached in the folder are several sample CSV files to show that this can definitely be done.