CSIT 356: Introduction to Data Science

Project Report

Project Title: Storms in Florida

Team Members: *Karen Serrano, Winnie Delinois, and Jessica Ariza*

---

1. Description

1.1 Basic Information

This project looks into the storm data, provided by NOAA, particularly taking an interest in Florida.

1.2 Project Objectives

The primary questions we were trying to answer on our project is to focus the data on Florida out of all the states. Once we narrowed down our data we worked on figuring out how long each weather event occurred for and sort it by longest durations to shortest duration. We also questioned which month had the most activity. Another question we had was which weather event had the highest count.

1.3 Description of the Data Set

In the dataset, it shows every storm event across all states, however, Florida is the main focus. The dataset also includes the month, day, and year of these storm events.

2. Exploration of Data Analysis

2.1 Data Preparation

We scaled the data down to Florida and took out data we did not understand or felt was unnecessary to keep. We wanted to focus on Florida out of all the states and we wanted to see the event type, the start and end of the event, the damages, and magnitude.

To select only Florida, we made a new dataframe and called it df2. We equaled df2 to our first dataframe, df1, and selected the column “STATE”. Afterwards we equaled it to “FLORIDA”, so the data would only be retrieved from the state of Florida. We called it df2, so it could print the result.

In addition to prepping the dataset, we dropped many columns that were not relevant to what we were looking for and didn't have much information to use. These columns were dropped using the drop function.

2.2 Data Analysis using Descriptive Statistics

In order to figure out the duration of the weather event we had to convert the BEGIN_DATE_TIME and the END_DATE_TIME columns in our data to datetime using pd.to_datetime(). Once the duration was calculated we used the following code to sort the duration from highest to lowest, f7=df2.sort_values(by= 'DURATION', ascending=False). We then worked on figuring out the count of the month which we did by applying this code, df3 = df2['MONTH_NAME'].value_counts().rename_axis('MONTH').reset_index(name='COUNT')

df3.set_index("MONTH",inplace=True). Once we had a dataframe with the months and count, we created a horizontal bar graph to represent the data. We did this by making df4 = df3.sort_values(by='COUNT', ascending=False)

df4 = df4[:10]

df4.plot.barh(title="Storm count per month",rot=0). We worked on getting the count for the type of event. We used, df5 = df2['EVENT_TYPE'].value_counts().rename_axis('EVENT').reset_index(name='COUNT')

df5.set_index("EVENT",inplace=True), this code to get the count of each event and create a dataframe for it. We then took this new dataframe and used to make a bar graph for the types of events and their count using the following code, df6 = df5.sort_values(by='COUNT', ascending=False)

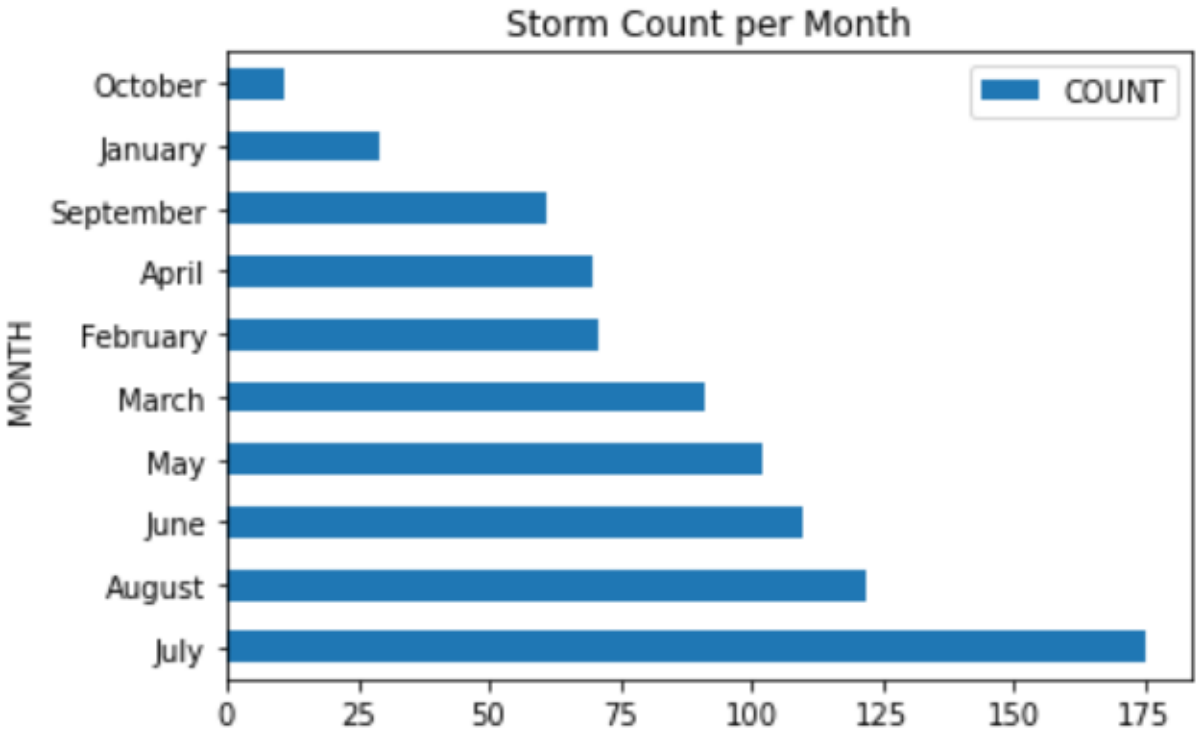
df6 = df6[:20]

df6.plot.bar(title="Event count",rot=90)

3. Data Visualization

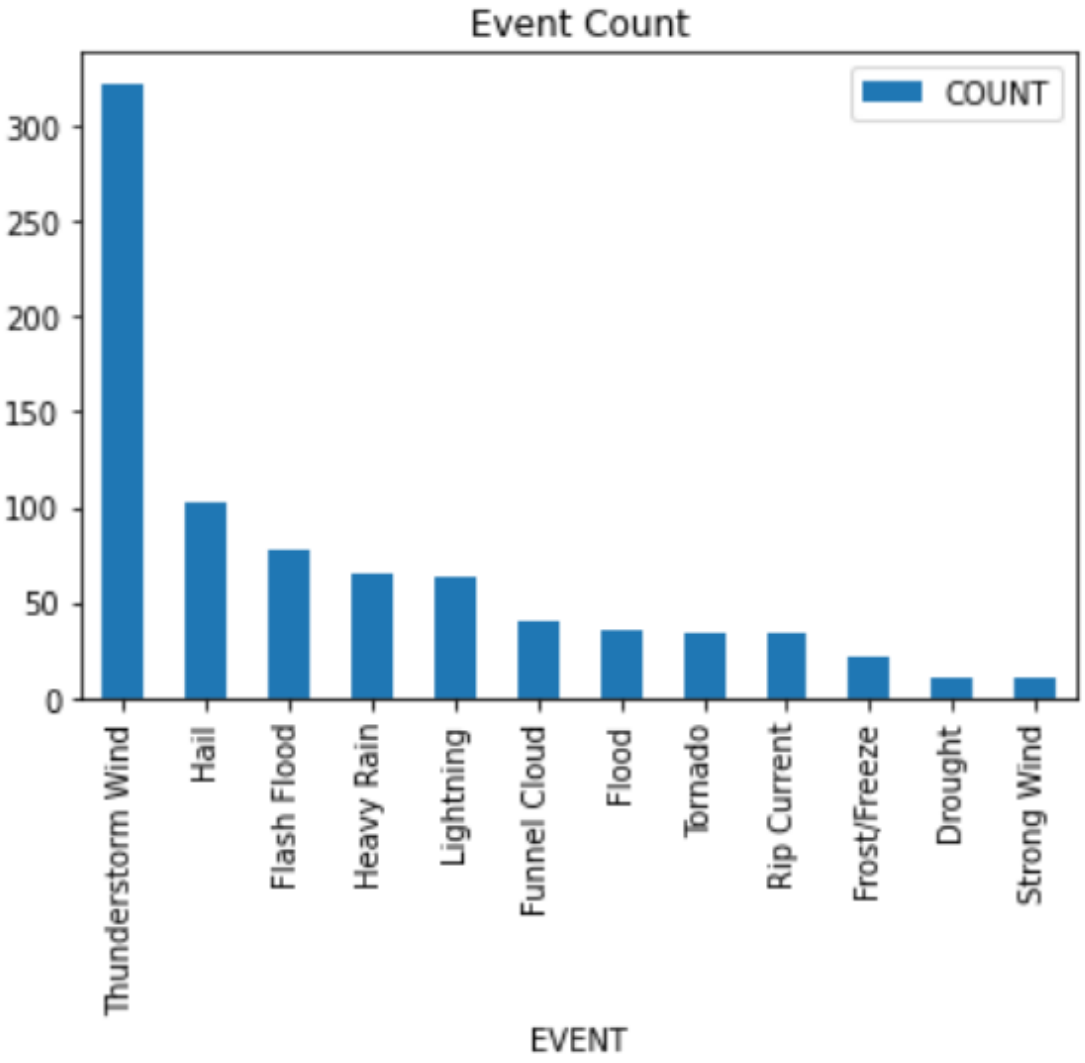
```
In [78]: florida6 = florida5.sort_values(by='COUNT', ascending=False)
         florida6 = florida6[:10]
         florida6.plot.barh(title="Storm Count per Month",rot=0)
```

```
Out[78]: <AxesSubplot:title={'center':'Storm Count per Month'}, ylabel='MONTH'>
```




```
In [77]: florida8 = florida7.sort_values(by='COUNT', ascending=False)
         florida8 = florida8[:12]
         florida8.plot.bar(title="Event Count",rot=90)
```

```
Out[77]: <AxesSubplot:title={'center':'Event Count'}, xlabel='EVENT'>
```



4. Conclusion

Based on the graphs, we can conclude that most of the storm events occurred in the month of July and the most common event that occurred was thunderstorm winds, with hail being a close second that occurred in 2013. In the beginning of 2013, the longest duration of a drought was for 25 days. The property damage after a lightning storm was the most costly, being $95,000. Magnitude was used to measure wind speeds and hail size and in 2013, the average magnitude was 17.144878 and the maximum was 105. Overall, these kinds of events are seen more frequently due to climate change.