Summary of Storm Events and Data Analysis

Daniel Clark

CTEC 298 Spring 2021

**Introduction**

Throughout this CTEC 298 class, the basics of python, additional python libraries, including matplotlib, pandas and numpy, as well as other data analysis applications such as Jupyter Notebook and Tableau have been covered. Each of these options provides a powerful method to manipulate and analyze data to provide strong conclusions from the data.

**Storm Events – CTEC 128 Report**

In my essay, “Storm Events – Data Science Report”, I tackled the rising threat of global warming and how it affects the insurance industry through increased severity of storm events.

With this concept map, I explained how the various factors of global warming affect and increase the severity of storm events. I also explained that the two main groups affected by these more severe storms are home and building owners and the government. I also explained the various factors for both parties that storm events would affect. The questions that guided my research for this topic were:

1. Is there an upward trend in the intensity of storm events?
2. Given selected states to study, which has the highest property damage in a specified time period?
3. Which of those five states has the highest number of deaths and injuries?
4. Which are the highest costing storm events within the five states?
5. Should insurance policies be changed based on the analysis?

The data used in the essay was retrieved from the National Oceanic and Atmospheric Administration (NOAA) database. The data retrieved contained information on every storm event in the United States from 1950 to the present year, 2018 at the time. However, the time period I focused on was 2005-2018, 2005 being the year Hurricane Katrina devastated the Louisiana region. The five states chosen for analysis were populous and large area states from the West, Midwest, Southwest, Northeast, and Southeast. The states chosen were California for the West, Michigan for the Midwest, Texas for the Southwest, New York for the Northeast, and Florida for the Southeast.

After filtering and analyzing the data, different visualizations were created to review the trends of storm events within the five selected states.

The first visualization shows the average magnitude for all applicable storm events for each year within the states. This shows that storm events were rising moderately before 2016 and jumped for both 2017 and 18.

The next visualization shows the amount of property damage in each state for 2018. For Florida, Hurricane Michael was removed from the dataset to reduce bias from the outlier. This shows that property costs are around the same for Michigan, New York, and Texas, while Florida and California invoke lower costs.

The third visualization depicts the number of direct and indirect deaths caused by storm events for the entire selected time period grouped by state. In this figure, Texas outclasses the other states, but California and Florida hold higher numbers of deaths compared to property costs.

The fourth visualization depicts the number of direct and indirect injuries caused by storm events for the entire selected time period grouped by state. Texas outclasses the other states again, but the numbers for the four other states are closer in value than before.

Lastly, the highest costing individual storm events for the states within 2018 were depicted. This shows what storm events are most likely to affect what area for insurance companies to adjust their rates and protections by region.

Based on the data, it was determined that Texas is a very high-risk state in terms of storm events. Florida was also determined to be a higher-than-average risk. Michigan and New York were found to have more damages relating to rain and windstorm events, while California was determined to have costs mainly relating to fires. This information was then used to recommend policy changes and coverage for insurance within these regions.

**CTEC 298 Material**

In this CTEC 298 class, I was responsible for completing the python tutorials within Dataquest, python tutorials on conditional statements, dictionaries, functions, tutorials on juptyer notebook, matplotlib, pandas, numpy, and Tableau.

The python tutorials within Dataquest, as well as the ones relating to conditional statements, dictionaries and functions were all geared towards learning the basics of python. These tutorials went over the basics of the language such as variables, loop statements, user defined functions, etc.

The jupyter notebook tutorial explained how to set up a jupyter notebook. The tutorial also detailed how to interact, store, and save data within the jupyter notebook.

The matplotlib tutorials explained how to install and use the matplotlib library within python. This library can be used to create graphs and charts from data for powerful analysis. The tutorials showed how to create basic plots, stack plots, pie charts, bar graphs, scatter plots, and histograms.

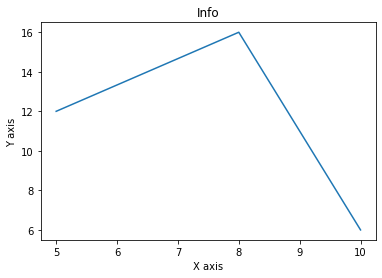
The pandas tutorial showed how to install and use the pandas library within python which allows for powerful and easy data manipulation within the language. This library can be used to create or import dataframes. This can also be paired with other libraries for even more powerful analysis.

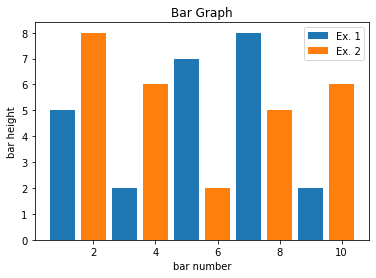
The numpy tutorial explained how to install and use the numpy library within python which provides support for larger and multi-dimensional arrays and matrices. Numpy also allows for powerful mathematical calculations and operational analysis on various numbers and figures within python. Once again, this library can be combined with others to provide even better analysis.

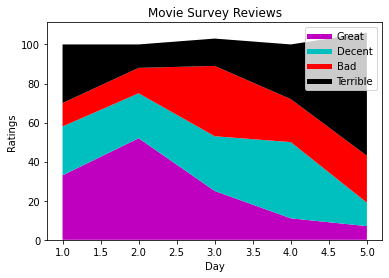
Lastly, the Tableau tutorial showed how to install and use Tableau to create visualizations that can be useful for data analysis. Tableau provides a user interface approach to create visualizations, while matplotlib focuses on the coding interface. Multiple file formats can be imported to Tableau, providing a very powerful and compatible analysis tool.

**Matplotlib Plots**

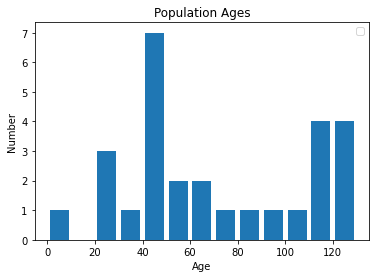
A simple plot is used to chart data on a table and can compare values to each other, or potentially show trends within data.



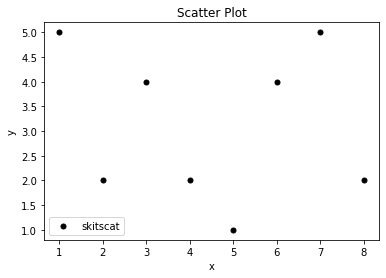
A bar graph represents the values of data as rectangles with different heights. These graphs can be used to depict which category or piece of data has more or less value. 

Stack graphs show values of data on top of each other or next to each other in the same column. These graphs can be used to portray the importance of individual pieces as well as the whole value. 

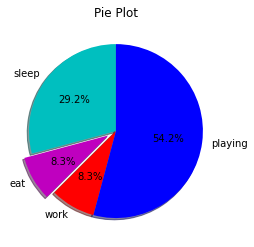
Histograms are graphs that show a representation of grouped data based on certain ranges. Similar to a bar graph, it depicts the data as rectangles, representative of the various ranges. However, unlike a bar graph, histograms show the value of a range of data, rather than an individual value.



Scatter plots are graphs that place points of data based on two important factors. These graphs can show correlations between the two selected factors based on the closeness and pattern of points on the graph.

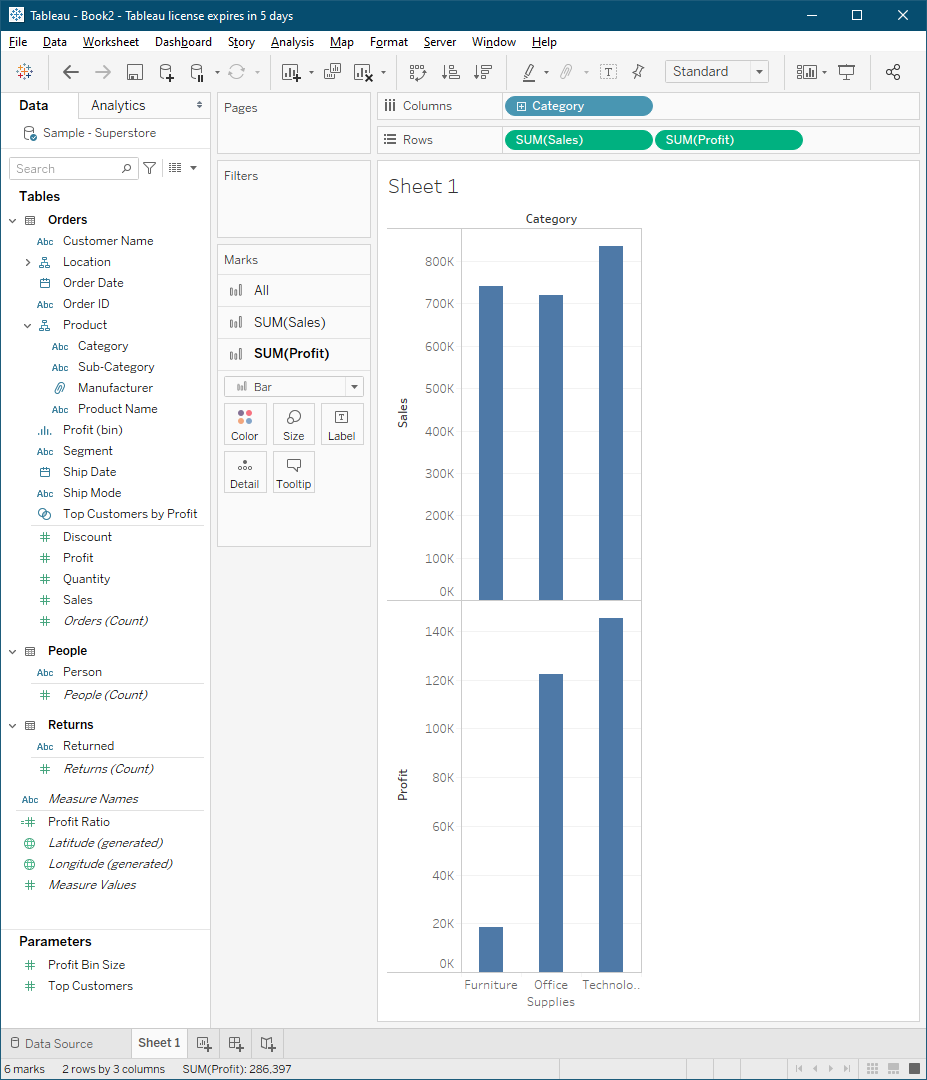


Lastly, pie charts are graphs that show data as pieces of a whole shape. These graphs can be used to show the importance of individual pieces compared to each other and the whole.

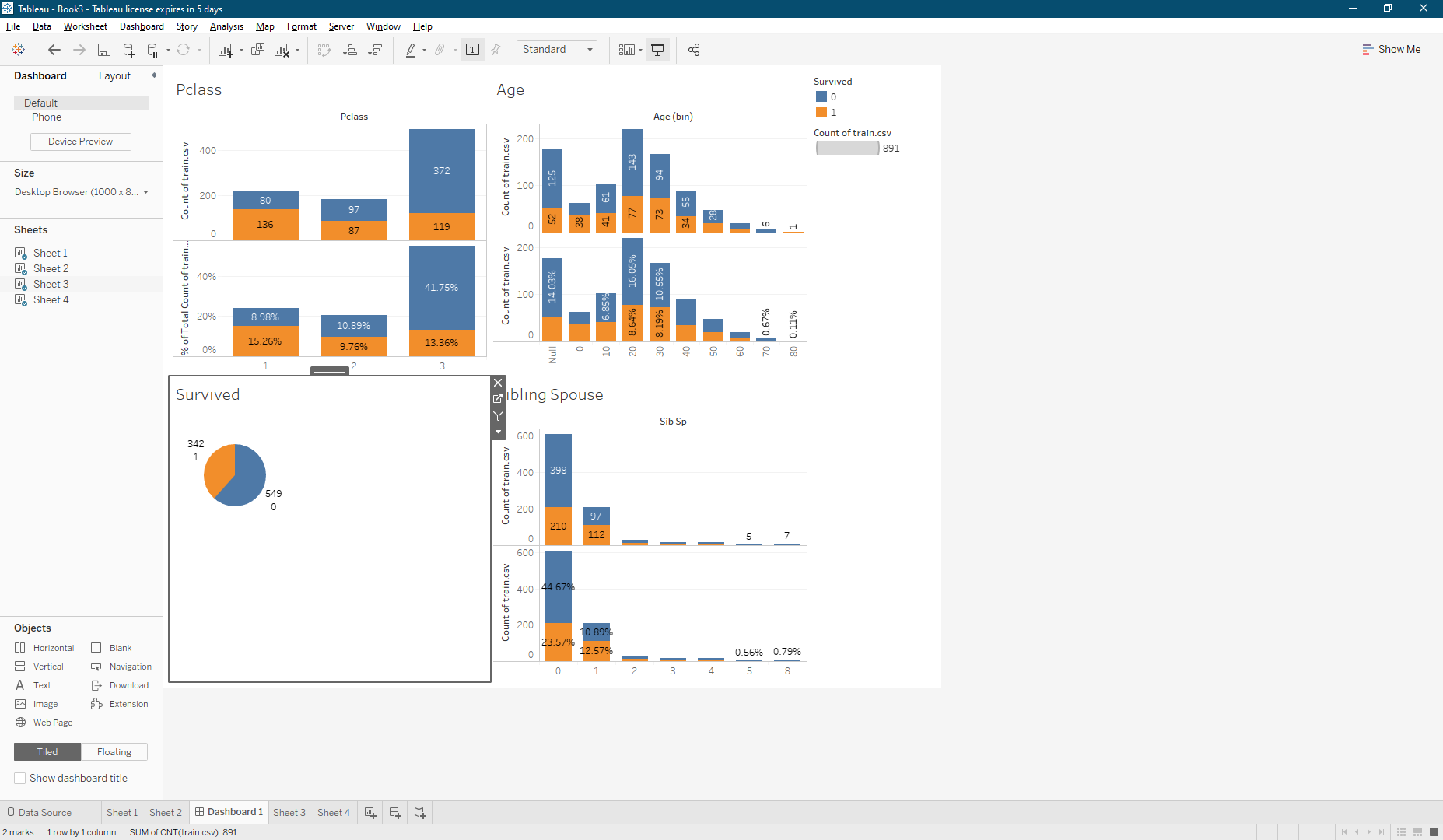


**Tableau Plots**

Within Tableau, two plots were made. The first being a bar graph from the Sample – Superstore dataset. This plot shows the amount of sales and profits by category for a store.



The second plot is a multiplot containing three stack plots and a pie chart created from a dataset for the titanic. The charts show the amount of survivors and deaths between different classes of passengers, ranges of age, sibling and spouse relationships, and the total amount of survivors compared to fatalities.



**Conclusion**

Overall, this class was used to cover data analysis through various means. Python is a programming language that can be added on through libraries and used as a powerful data manipulation and analysis tool. Python can be used for adding, deleting and altering data as well as calculations and creating analytic visualizations. The libraries that python utilizes for these tasks are matplotlib, numpy, and pandas. Jupyter notebook is a server based solution that is based in python. Therefore, Jupyter notebook can also utilize the same libraries as python. Lastly, Tableau provides a user interface approach to data analysis similar to matplotlib in python. Tableau can accept various file formats to create a powerful and compatibly data analysis tool.

References

* Edureka! (Director). (2017, April 19). *Python Matplotlib Tutorial | Matplotlib Tutorial | Python Tutorial | Python Training | Edureka* [Video file]. Retrieved February 24, 2021, from <https://www.youtube.com/watch?v=yZTBMMdPOww>
* NCEI. (n.d.). Storm Events Database. Retrieved from <https://www.ncdc.noaa.gov/stormevents/>