COMP 495 PROJECT PROPOSAL

Research topic- Weather/Climate Data Analysis

Background:

Whenever we leave our home, we tend to check how the external conditions are. By external conditions, I mean the weather conditions. Specially living in a province like Alberta, where a lot of times we must face extreme weather conditions. On days when it snows, people tend to avoid travel and stay indoors. This proves that how important it is to be on top of the weather data and learn from the trends that have occurred in the past.

The purpose of this project is also the same- I will be analyzing various weather conditions which govern how people should be prepared, and what to expect. The major weather factors that I will be working with in this project are wind speed, rain, humidity, visibility, snow, and sky conditions.

In this project, I will be first looking to explore a suitable weather data source which has been recorded over a period. After I get the data, I will look for null values and clean the data for analysis. I would then play around and try to find various combinations of wind speed, humidity with other factors and find mean, and standard deviation of some of the weather conditions.

Technology used to make the project: Python and Python Libraries

Through COMP 495, I intend to demonstrate the core programming language that is used widely in the Data Science field. In COMP 495, I will be using Python as a programming language to perform the operations and tasks within the project. In Python, I will be importing the Python libraries such as Numpy, Pandas, Matplotlib, and Seaborn. Numpy and Pandas will be used for data storage and data manipulation. I will be looking for the dataset and will import that using the CSV or Excel file into the Jupyter Notebook.
After I perform my data analysis operations on the dataset, I will be exploring different avenues related to the data and then visualize them using Python libraries such as Matplotlib and Seaborn.

Project Concept:

Data Analysis plays a huge part in all cities and countries around the globe. Some governments use the weather data for the safety of its citizens and to make them more informed to reduce the number of accidents. Some businesses, for instance, agricultural companies would want to cultivate a crop in a specific region based on the historical data in the past and the yield value there. Therefore, weather analysis is very important and has huge real-life value.

Focus:

The focus of this project would be to use Python programming language as well its library and demonstrate those in getting the desired analysis about the weather conditions.

With Pandas and Numpy, I will be storing data into arrays and other data types so that analysis can be performed on those. After that has been completed, the focus would be to demonstrate the results in the form of visualizations. Data visualization is the most crucial part as this is the part where we try to showcase the inference of our findings to the audience.

Content:

The content of the project would be all the data related to the weather conditions of a gjven Canadian city. There are a lot of sources of weather data on the web, but I will probably try to scrape it out of web or pick the data from Kaggle. After I import the data, I will use Python libraries to modify, clean, transform, analyze and visualize the weather trends.

Preliminary Design:

The design would follow a basic structure of a Jupyter notebook in Python. Cells would be used to write the code and then displaying the findings in the form of a table showing the top 5 entries and so on. After displaying the table, I will also be displaying the visualizations in the cells below by typing the code in python libraries- Matplotlib, and Seaborn.

Planned Look and Feel:

The final report will have a Weather poster in the beginning, which will give the feel of a weather page. Below that, the design structure that was discussed in the Preliminary Design section will be followed by showing the findings in the form of tabular structure, followed by visualizations.

Milestones and Timelines:

1-4 weeks: Project planning, documentation, and research about the data source

2-8 weeks: Learning about Numpy, Pandas, Matplotlib and Seaborn. Implement Numpy and Pandas to demonstrate Data Analysis on the data collected and show them in the tabular form in the Jupyter Notebook.

9-12 weeks: Perform Data Visualization by using Matplotlib and Seaborn. Along with this, working on the documentation of the project and the visuals.

12-15 week: Preparing to present the project. Prepare a report which would present the final findings of the research and present it to the supervisor.

Suggested Evaluation Criteria:

Correct implementation of the Python language and its library. Valid representation of different weather factors such as Rain, Humidity, Snow, etc. and their stats in the form of table and visualizations.
Proper information being provided at each step in the form of Markdown cells which would present the fact about the finding and conclude the result, which would make the decision making easier.

Importance of this research:

This type of research in the field of weather analysis is very crucial. While there is data about a lot of popular places on the web, not many articles or posts have analyzed data for all the cities. Through this research project on weather for a Canadian city, I would demonstrate my learning of Python language and also demonstrate coding practices that I have learnt during my degree. With my weather analysis, I want to make my friends/family more aware about the weather patterns during different months which would be an interesting face to know about for everyone.

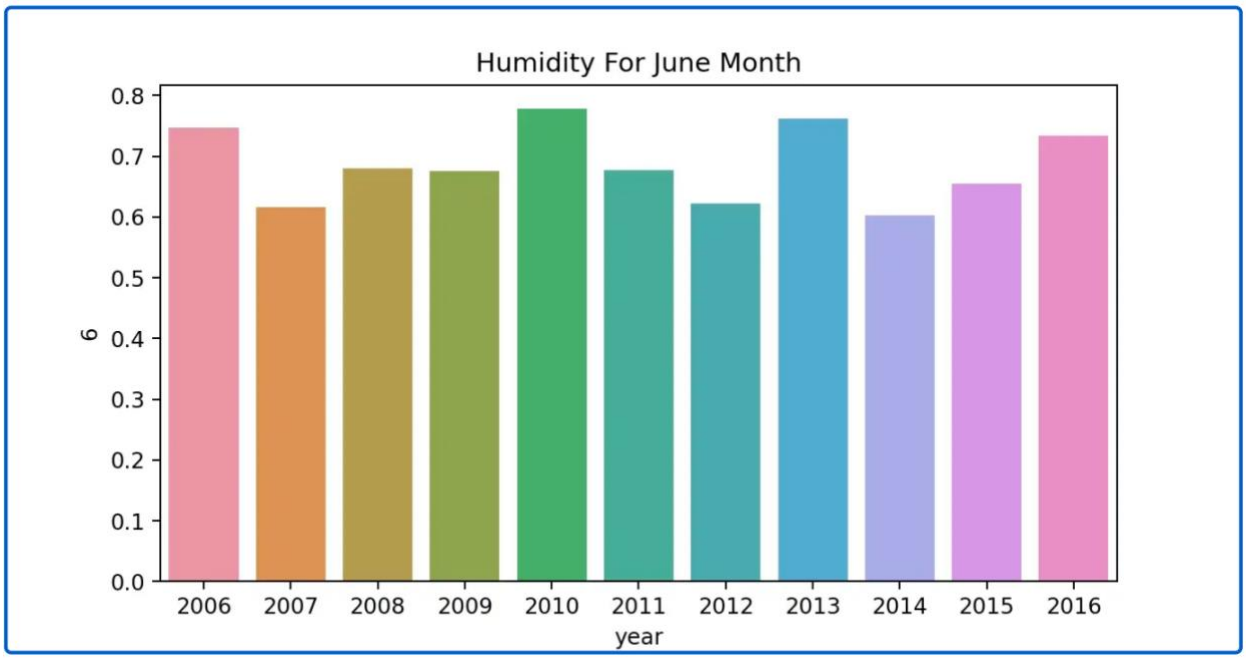
Data source:

The data for this project would be taken from Kaggle, which contains 80 years of Canadian Climate Data.
Source: Kaggle- 80 Years of Canadian Climate Data

Visualization:

My goal is to show my friends and family the trends and patterns of weather in a given Canadian city and the best way to show it would be in the form of visualizations. I will be doing the analysis and then presenting the visualizations in the form of bar graphs, line charts, etc.

A sample of a weather visualization is shown below. My work would have similar sort of bar graphs and other visualizations.



The above visualization has been performed using Tableau, whereas the visualizations that I will be working with in my project will be made solely using Python and its libraries.