Thomas Truong
Leaf Darnelle
Hayden Beets
4/8/22
Section 008

Effort Sheets Week 1

We spent this week deciding which roles each member would be assigned. Since there are three main deliverables to complete, we have split the work in three ways. Hayden will do the Literature Research about the streamflow forecasting, Leaf will do the Exploratory Data Analysis (EDA) summarizing the essential information of the data set,and Thomas will start the Data modeling.

Hayden has researched the importance and challenges to streamflow forecasting. Forecasting is important because it uses past and current data to create models that can predict future streamflow data. These predictions help create more advanced flood and mitigation systems, provide necessary data to determine potential areas for hydroelectric plants to be built, track sediment transportation, help manage irrigation decisions, and direct the planning and operation of reservoirs. One challenge with the conventional methods used in modeling streamflow forecasting is the use of linear models. Many linear models are used to model streamflow data, however the data collected is not linear.

The progress report on what Leaf has discovered on the Exploratory Data Analysis (EDA) is that the ColoradoRiverData CSV file contains 373 rows of data detailing the monthly data of streamflow. The dataset contains 10 different columns of information regarding the date, the total recorded precipitation in inches for each month, the minimum recorded temperature in degrees Fahrenheit for each month, the average recorded temperature in degrees Fahrenheit for each month, the maximum recorded temperature in degrees Fahrenheit for each month, the difference between the recorded highest and lowest temperatures, the Soil moisture index, the evapotranspiration, the average recorded streamflow in cubic feet per second for each month,

and the periods of flow rate which can either be 1 or 0. Using the .describe function on the dataset then reveals the mean, standard deviation, minimum, maximum, and quartile percentiles. The most notable information reveals that the average precipitation is 42.48, the average temperature is 17.77 degrees Fahrenheit, and the average recorded streamflow is 13.15 cubic feet per second. However, the dataset spans three decades including the 1990s, 2000s, and 2010s which allows for more analysis throughout time to figure out if instances of drought are increasing or the streamflow is slowly decreasing. It is shown that the 1990s has 120 rows of data, the 2000s has 120 rows, and the 2010s have 111 rows. Another crucial data value to look at is the "Flow?" column which tells whether or not a flowrate existed, after analysis, there exists 120 rows of non-existent flow rates and 253 rows where it does exist.

Thomas has reviewed the data models taught in class and has determined that in order to predict the flow rates of the Colorado River, the usage of hypothesis test models should be applied. The normality of the data must be measured, then based on the accuracy of the Gaussian distribution, parametric or non-parametric tests would be applied. As of now, no data models have been applied on the data, however once the model quality has been assessed, the estimated flow rate will be predicted based on the value of the information in the "ppt" column.

Signatures: Leaf Darnelle

Hayden Beets

Thomas Thruong