**DSCI 230 Final Project**

You have been contracted by a mountain biking company in Bentonville, AR where there are world-class mountain biking trails. The company has customers who register to go biking before arriving, and customers who walk in without a reservation. For this project, there are **registered** and **causal** users, with registered being those who made reservations and casual users who did not.

The business owner would like to know if the weather affects one group of customers more than the other. They would also like to identify the thresholds in which the groups of users will decide not to come, whether it be snow, rain, or excessive temperatures. The analysis part is up to you and how you interpret the data.

A person riding a bike on a bridge in the woods

Description automatically generated

For your final project, please construct a data analysis example using linear regression to explore a relationship between one of the weather attributes, and the registered and/or casual users. The due date for this project is **April 29th by 8pm EST.**

**Deliverables:**

Your project should include a Jupyter notebook and a Word doc write-up of your methodology and analysis outcome.

Please include, in not particular order, examples of the following in your code using the dataset provided:

* Sorting
* Slicing and Indexing
* Create a Function
* Create a DataFrame
* Index method to compute a Boolean array indicating whether each value is contained in the passed collection
* Indexing options with DataFrame for selecting a single column or subset of columns by label
* Filtering out missing data
* Filling missing data
* Detect and filter outliers
* Demonstrate two different types of joins with the *how* argument
* Create a line plot with point markers
* Create a line plot with the tick labels adjusted, color added, and a legend displayed
* Create a vertical and horizontal bar plot
* Create a scatter plot with a linear regression line fitted
* Create a boxplot
* Choose two *groupby* methods and demonstrate using dataset
* Perform slicing using *datetime*
* Show an example of using periods and period arithmetic
* Create a period index
* Form a linear model
* Show summary statistics for the model
* Score the model