# Module 5: PyBer with matplotlib

5.2.2 Project Overview

* Import your data into a Pandas DataFrame.
* Merge your DataFrames.
* Create a bubble chart that showcases the average fare versus the total number of rides with bubble size based on the total number of drivers for each city type, including urban, suburban, and rural.
* Determine the mean, median, and mode for the following:
  + The total number of rides for each city type.
  + The average fares for each city type.
  + The total number of drivers for each city type.
* Create box-and-whisker plots that visualize each of the following to determine if there are any outliers:
  + The number of rides for each city type.
  + The fares for each city type.
  + The number of drivers for each city type.
* Create a pie chart that visualizes each of the following data for each city type:
  + The percent of total fares.
  + The percent of total rides.
  + The percent of total drivers.

5.2.3 Load and Read the CSV files

* Don’t forget: to use os.path.join() to load the csv files the os module has to be imported with: import os
* Read in a csv file with: pd.read\_csv()

5.2.4 Inspect the Data

* Get all the rows that contain null values and check data types, and how many data points
* df.count() method for null values
  + another option is: df.isnull().sum() method and chaining
* dtypes for data types
* unique() method for data points
* when merging dataframes combine on the column that has the same data and label
* merging syntax: new\_df = pd.merge(leftdf, rightdf, on=["column\_leftdf", "column\_rightdf"])
  + changing the order will change the order of end df