

## 5.2.2 Overview of the Project

**You** may only be in your second week at PyBer, but there's one thing you've learned along the way that you know applies here too: if you fail to plan, you plan to fail.

So before you dive into these massive datasets, you sit down and scope out the analytical project to make sure you and Omar are on the same page about the final deliverables.

It would be foolish to start messing around with such a large dataset without knowing exactly what you're trying to get out of it. So you and Omar have a head-to-head to clarify the process and outcomes for this project. You come up with the following list of steps and deliverables:

- 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.

Omar has approved the project scope. It's time to get to work!