What is the problem you want to solve?

Hubway is a bicycle sharing system in the Boston, Massachusetts metro area. The system is owned by the cities of Boston, Cambridge, Somerville and Town of Brookline, and operated by Motivate and uses technology provided by 8D Technologies, as well as PBSC Urban Solutions bikes and docking stations. The bike share program officially launched on July 28, 2011 with 61 stations and 600 bicycles. As of May 2017, the system has deployed 180 stations with a fleet of over 1,600 bikes. Bike share systems have been growing in popularity across the United States. The concept is simple. There are racks of bikes set up around the city, and people can rent a bike for a short period of time, even if only to get from point A to point B. Obviously, there are many benefits to bike shares. However, that's not to say they don't come with its disadvantages. Some of the advantages are improved air quality, Convenience, Better bike Laws, Healthier people whereas some disadvantages are traffic accidents due to first time bikers with no helmet and slow traffic. There are also several reviews found online which is about the poor customer service, over charged rental fees, lack of docking station.

One of the problems with commercial bike sharing programs is unequal riding patterns that result in unequal bicycle distribution at the end of the day. This means that unless the bikes are redistributed at night, there will be insufficient bikes at certain locations and too many bikes at other locations for the number of riders who wish to use them.

For this project, my focuses is on Hubway rideshare issues in the City of Boston and analyze the root cause of different issues faced by riders every day and provided a detailed report on the following:

1. A summarized visualization of daily trend to know how the demand for bike is on weekdays as compared to weekend or holiday
2. A comparison of 3-5 Hubway station with most demand against 3-5 least Hubway station with least demand
3. Identifying the most prominent time of the day having highest and lowest demand for bike
4. Identifying the district with the most change in demand and potential reasons for the change
5. Building a predictive model of Hubway bike demand in each station using machine learning
6. Finally, after addressing #5, I wanted to identify the most salient features/variables used by the model for predicting Hubway bike demand, within the limitations of my dataset

Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?

There are two different types of clients that could be interested in the findings from this project. The first type of clients would be the US online and print media that cover socioeconomic and urban issues. These clients are magazines that take an active interest in stories driven by socially relevant issues and are backed by data analytics, for creating awareness within the public while simultaneously enhancing the quality of their readership. For example, US online media such as US News and Analytics Vidhya would fall under this category. I also anticipate interest from Transportation department and private organization in Rideshare business who wants to invest in such program in other cities.

What data are you going to use for this? How will you acquire this data?

In 2012, Hubway and MAPC challenged the public to visualize half a million Hubway rides. The Challenge is now closed, but the data is still available for public use. Recently the organizers has posted new comprehensive trip history data. Data from Hubway's launch in 2011 through the end of the 2013 regular season are now available.

- The Hubway trip history data includes every trip taken through Nov 2013 – with date, time, origin and destination stations, plus the bike number and more. Related data (Census, neighborhoods, bike facilities, elevation, etc.) and Station status data, with information about available bikes and empty docks per station are also available:
- Since the demand for riding a bike is very much dependent on the weather, the daily temperature, humidity, wind speed details are collected by searching the historical weather details for Boston from Weather Underground website

In brief, outline your approach to solving this problem (knowing that this might change later).

1. I will study the Hubway trip data from 2011 to 2013 and identify top 3-5 station having demand and bottom 3-5 station having lowest demand.
2. Analyze the bike ride and investigate which among the following variables which are the having most significant impact:
    a. Route: Start station and end station
    b. Age: Age of the traveler
    c. Gender: Gender of the traveler
    d. Date time: Date and hour in "mm/dd/yyyy hh:mm" format
    e. Season: Four categories-> 1 = spring, 2 = summer, 3 = fall, 4 = winter
    f. Holiday: Whether the day is a holiday or not (1/0)
    g. Working day: Whether the day is neither a weekend nor holiday (1/0)
    h. Weather: Four Categories of weather
        1. Clear, Few clouds, Partly cloudy, Partly cloudy
        2. Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
        3. Light Snow and Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
        4. Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog
    i. Temperature: Average temperature in Fahrenheit
    j. Apparent temperature: Average feels like temperature in Fahrenheit
    k. Humidity: Average relative humidity
    l. Wind speed: Average wind speed
3. Initial Data Exploration and checking the missing values
4. Normalizing the Hubway trip data if required
5. Exploring further into the factors contributing high demand
6. Comparison of station having most and least the demand
7. Find stations with maximum change in demand over time
8. Identify prominent age groups and economic factors behind using rideshare
9. Prediction using machine learning
10. Model Selection and evaluation

What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

1. The Python source code
2. Final paper with detailed analysis with limitations and future extensions
3. A presentation of key findings and recommendations