User Guide

- DS-GA-1007 Final Project: Citibike analysis
- Team Member:
    - Muhe Xie(mx419) mx419@nyu.edu
    - Sida Ye(sy1743) sy1743@nyu.edu
    - Junchao Zheng(jz2327) jz2327@nyu.edu

Part 1: Environment Setup

1. run the following commands to download our projects

```
$ git clone [git-repo-url]
$ cd sy1743
```

2. Set up memory of VirtualBox as 6GB

3. download data from google drive, click here to download.

4. move data file (**Citibike_final.csv** and **station_dictionary.p**) into our repo sy1743

5. Install basemap package, see instruction basemap_installation_guide.md here

Part 2 : Usage of the Application

1. This data visualization will generate four graphs with two pie plots and two bar plots: gender distribution, user type distribution, daily usage and daily miles.

2. The station frequency visualization part (option 2) will print the name of top 5 high frequency stations and generate 3 plots automatically (close one to see the next plot).

    - Plot 1: The points of citi bike stations on the map
    - Plot 2: The top 5 high frequency stations on the map
    - Plot 3: The heat map of the station frequency

3. Recommendation and predication
    - Get information of usage of the station on that particular date on historical date and get recommendation on the station.
    - Get two alternative stations nearby which meet with the criterion: I. within 15-minute walk, II. predicted to be recommended.

Part 3 : Configuration

1. Pandas to access data.
2. Numpy to perform statistical analysis.
3. Matplotlib for graphics including pie plots and bar plots.
4. Basemap for graphics including geometric maps.

Part 4: How to run the program

In terminal, enter the command to run the program:

```
$ cd sy1743
$ python main.py
```

Main Menu:

- Enter 1 to go to a sub menu of monthly data visualization.
- Enter 2 to go to a sub menu of station frequency visualization.
- Enter 3 to go to a sub menu of prediction and recommendation.

Option 1: Monthly data visualization

- Input year and month between 2013/7 and 2015/10:
    - Enter a year between 2013, 2014, 2015.
    - Enter an integer from 1 to 12 as month.
- Enter back: go back to main menu.
- Enter quit: exit the program.

Option 2: Station frequency visualization

- Input year and month between 2013/7 and 2015/10:
    - Enter a year between 2013, 2014, 2015.
    - Enter an integer from 1 to 12 as month.
- Enter back: go back to main menu.
- Enter quit: exit the program.

Option 3: Prediction and recommendation

- **Note**: When customers go to a bike station, they can see the station id. So, we just use the station id as our input, which is more convenient for customers. ***You can find the station id information, station_information.pdf here***
- Enter 1 to run the prediction function, enter station ID, day, month and each end with return.
- Enter 2 to run the recommendation function, enter station ID, day, month and each end with return.
- Enter back: go back to main menu.
- Enter quit: exit the program.

Part 5: Q&A

1. Where can I find data?

- You can find it on google drive or follow the link from Part 1.

2. How to install the basemap?

- Please follow the instruction here

3. What is the result for the program?

- You can find some sample plots under the sample_figures file.

4. How to test the program?

- You can run the test by enter

```
$ python test.py
```

Acknowledgement

- Data resource from citibike website