

Student Name: Akhil Montrose
Student ID: 81788
Course: Data Science (DASI3001)
Instructor: Dr. Ken Sooknanan
Document: Assignment 3 (Cyclist & Weather Dataset)

Table of Contents

Problem	3
Methodology	4
Flowchart.....	5
Results.....	6
Total Cyclist vs Average Cyclist.....	6
Total Cyclist vs Average Cyclist vs High Temperature Days	7
Observations.....	8
Conclusion	9
Appendix.....	10
Question 1	10

Problem

In this section, we will be discussing the problem that is to be solved.

The city of Cambridge is trying to promote tourism in their busy month of August by renting bicycles to tourist to cycle to the various attractions. Some of these tourist attractions include museums, art galleries, parks, hiking and cycling trails, that are located along very scenic routes. For tourist that are unable to cycle, such as senior citizens, the city plans to utilize hop-on/hop-off buses to traverse the various attractions. These buses will also be utilized by cyclist when the routes become too difficult for cycling. To schedule this bus service adequately for the cyclist and the non-cycling tourist, the city requires you examine the weather (weatherData) and cyclist (Eco-Totem_Broadway_Bicycle_Count) datasets to determine if temperature affects cycling in the city.

Plot and analyse per day for the month of August 2017 the: i) total cyclist, ii) Average cyclist for the month, and iii) days when there were high temperatures i.e. above 60 (F). Superimpose these plots (i.e. all plots on the same graph, using different colours, so that trends can be easily spotted). From the plot discuss if high temperatures affected cycling in August 2017.

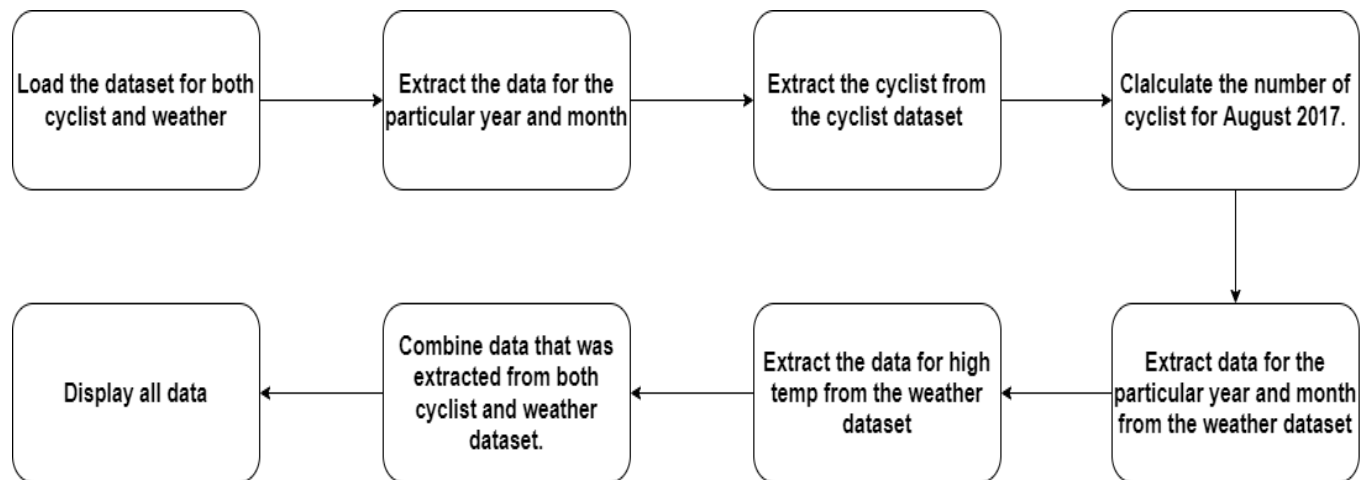
Methodology

For this problem, we need to do the following:

1. Load the datasets for both cyclist and weather.
2. Extract the data for the particular year and month.
3. Extract the cyclists from the cyclist dataset.
4. Calculate the number of cyclists for the specific month in question.
5. Extract the data for the particular year and month from the weather dataset.
6. Extract the data for high temp from the weather dataset.
7. Combine data that was extracted from both cyclist and weather dataset.
8. Display all data.

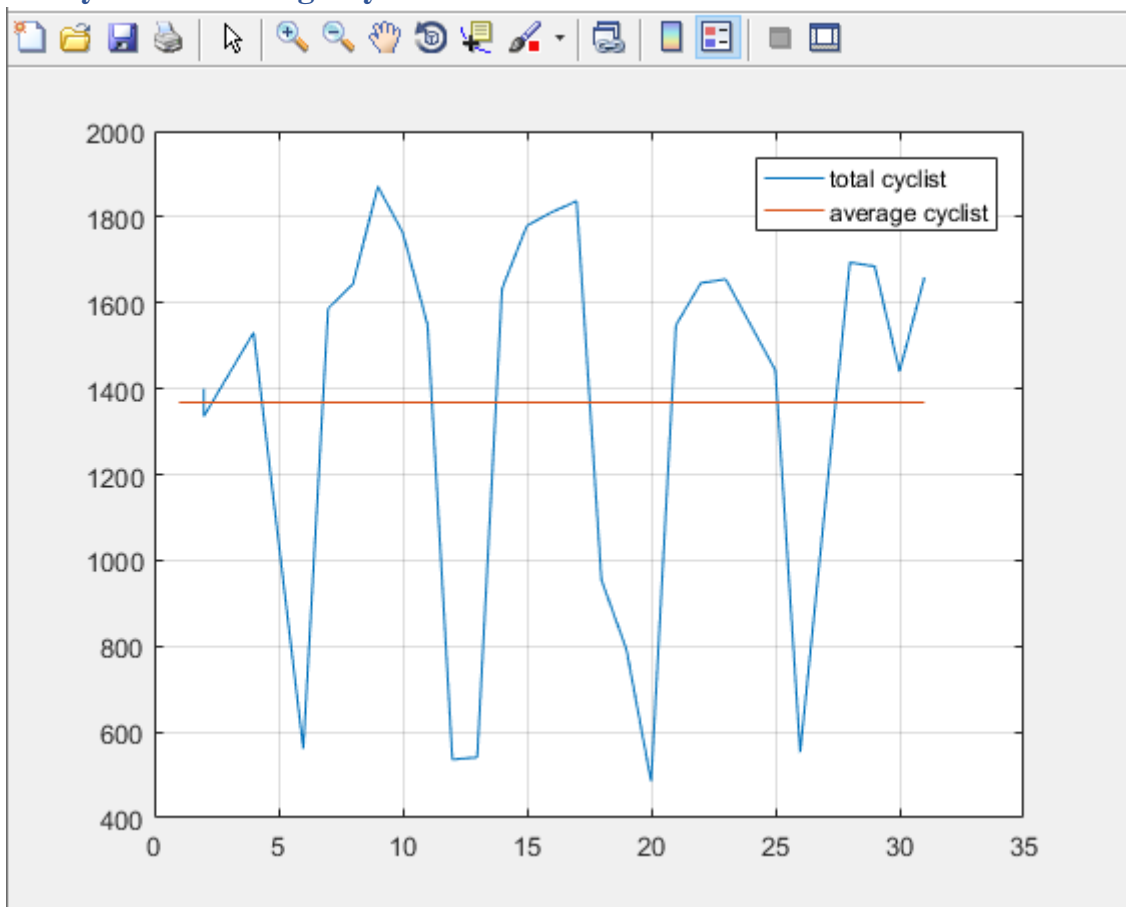
Flowchart

Below is a flowchart showing the steps outlined in the methodology.

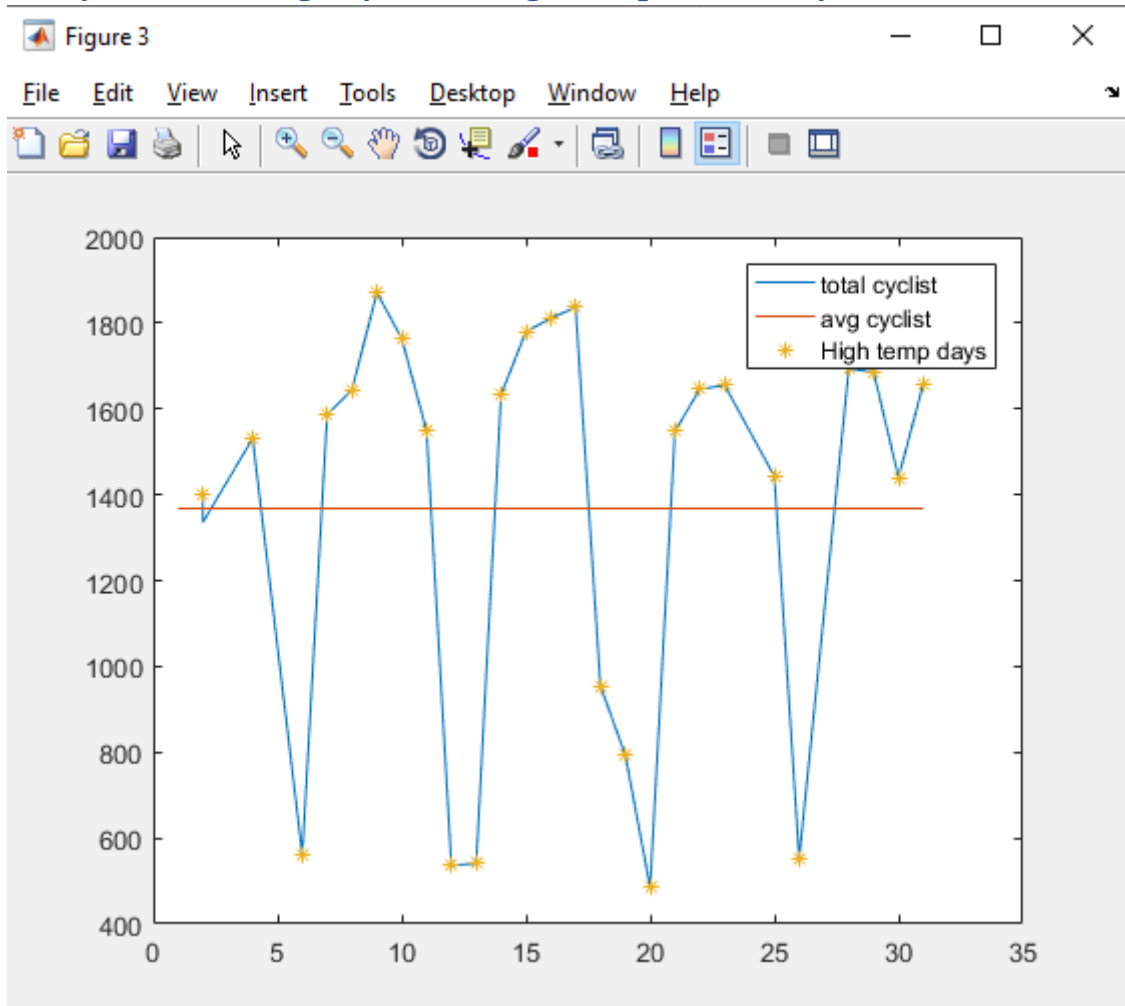


Results

Total Cyclist vs Average Cyclist



Total Cyclist vs Average Cyclist vs High Temperature Days



Observations

In the above diagram, it was observed that in month August 2017, the average cyclist was 1368 while the highest would range around 1900 cyclist a day and the lowest would be around 450 per day.

There were 26 high temperature days. Out of all of those days, 19 of them, the cyclist were above the average while 7 of them were below the average.

Conclusion

We can now conclude that the high temperature days didn't have an effect on cyclist. Matter of fact, there were more cyclist on the road when the temperature was above 60 degrees Fahrenheit.

Appendix

Question 1

```
% obtain the data on August 2017
dateC = EcoTotemBroadwayBicycleCount(:,1);
% convert to numeric array to search through
dateC = datevec(dateC);
%%% obtain a logical vector for the year 2017
c_2017 = (dateC(:,1)==2017);
%%% obtain a logical vector for August
c_Aug = (dateC(:,2)==8);
%%% obtain a logical vector for August 2017
cAug_2017 = c_2017.*c_Aug;
cAug_2017 = cAug_2017>0;
%%% pull out the dates(days) for August 2017
daysData = dateC(:,3);
daysAug2017 = daysData(cAug_2017);
%%% obtain the total cyclist data for August 2017
tC = EcoTotemBroadwayBicycleCount(:,5);
% convert to numeric array to search through
tC = cell2mat(tC);
tC_Aug_2017 = tC(cAug_2017);
% to get the total cyclist daily: first group cyclist data per day
divPerDay = 96;% 24x4
NoDays = round(length(tC_Aug_2017)/divPerDay);% 27 days
cyclistT_Each_Day_Aug_2017 = zeros(NoDays,2);
% do a for loop to sum the number of enteries on each respective Wed.
for i=1:1:NoDays
    start_P = i + (i - 1)*(divPerDay-1);% Arithmetic Progressioon
    end_P = start_P + (divPerDay-1);
    cyclistT_Each_Day_Aug_2017(i,1) = sum(tC_Aug_2017(start_P:end_P,1));
    cyclistT_Each_Day_Aug_2017(i,2) = daysAug2017(start_P,1);
end
%%% now to obtain the average total cyclist per day in August 2017
avgCylistT_Aug2017 = mean(cyclistT_Each_Day_Aug_2017(:,1));%1368
noDays = 31;
avgC = zeros(1,noDays);
avgC(:) = avgCylistT_Aug2017;
figure(2);
plot(cyclistT_Each_Day_Aug_2017(:,2),cyclistT_Each_Day_Aug_2017(:,1)); hold
on;
plot(1:31, avgC); legend('total cyclist','average cyclist'); grid on;
%%% obtain Temp information for August 2017
% 1-a) obtain the year 2017 data
yearW = weatherData(:,1);
yearW = cell2mat(yearW);
w2017 = yearW==2017;
% 1-b) obtain August month data
monthW = weatherData(:,2);
monthW = cell2mat(monthW);
wAug = monthW==8;
%%% 1-c) obtain a logical vector for the August 2017
wAug_2017 = w2017.*wAug;
wAug_2017 = wAug_2017>0;
%%% 1-d) obtain the Temp data for May 2017
wTemp = weatherData(:,4);
wTemp = cell2mat(wTemp);
wTempAug2017 = wTemp(wAug_2017);
%%% 1-e) obtain days with high temp i.e. greater than threshold
wTempHval = 60;
%%% logical vector for the days that the Temp is above threshold
```

```

wTempAug2017H = wTempAug2017 > wTempHval;
sum(wTempAug2017H);
wDays = weatherData(:,3); % days
wDays = cell2mat(wDays); % convert to numerical
wDaysAug2017 = wDays(wAug_2017); % get all days in Aug 2017
wDaysAug2017H = wDaysAug2017(wTempAug2017H); %get days with high temp. in
Aug 2017
%%% now to get the total cyclst data on the days with prec
noTempDays = length(wDaysAug2017H);
TcTempDaysAug2017 = zeros(noTempDays,2);
for i=1:1:noTempDays
    %%% identify the days with high temperature
    TempDay = cyclistT_Each_Day_Aug_2017(:,2)==wDaysAug2017H(i,1);
    %%% extract the total cyclist information
    TcTempDaysAug2017(i,2) = wDaysAug2017H(i,1); % store the day
information
    TcTempDays = cyclistT_Each_Day_Aug_2017(TempDay,1);
    if (length(TcTempDays)>0)
        TcTempDaysAug2017(i,1) = TcTempDays(1);
    end
end
%%% eliminate any missing information
validInfo = TcTempDaysAug2017(:,1)>0;
TcTempDaysAug2017 = TcTempDaysAug2017(validInfo,:);
%%% now to plot all the info.
figure(3);
plot(cyclistT_Each_Day_Aug_2017(:,2),cyclistT_Each_Day_Aug_2017(:,1)); hold
on;
plot(1:31, avgC);hold on;
plot(TcTempDaysAug2017(:,2),TcTempDaysAug2017(:,1),'*'); legend('total
cyclist','avg cyclist','High temp days');

```