Python Project Plan

import pandas as pd
import numpy as np
import myfunctions as func
 this file input / output

load chicago csv
load new york city csv
load washington csv

ask for first input: What city would you like to get statistics from ?
Type C for Chicago, W for Washington or N for New York City.
You can combine them, eg CW will give you statistic for both Chicago and Washington

ask for second input: What do you want to know?
1 - Popular times of travel?
2 - Popular stations and trip?
3 - Trip duration
4 - User info
5 - All
Here you cannot combine them, it is either one or all.

Output (if all, type one batch for each city)

CHICAGO
1 - Popular times of travel?
Most common...
month: July
day of week: Tuesday
hour of day: 08:00

2 - Popular stations and trip?
Most common...
start station: central
end station: industri
trip from start to end : central - brunnsparken

3 - Trip duration
Total travel time: 342 hours
Average travel time: 45 min

4 - User info
Counts of each user type: Subscribers - 675, Customer - 395

Counts of each gender: Female - 40, Male - 100

5 - All ^

New query ? Yes or No
Yes - start over
No - Thank you, have a good day!

#1 Popular times of travel

input_city = input(string, string, string)
start_dates = Get start from city.csv, city2.csv

Month_most-common_func
count month in start_dates(city from input)
do not count NaN or null
group by month
sort with largest first
Print the first

Weekday_most-common_func
get start time check what day of week with day method
count start time per days
group by days
sort with largest first
Print the first

Hour_ most-common_func
from city.csv
get hour from start time
count hours in start time
group by hours
sort largest count
Print the first

#2 Popular Stations and trips

stations = get start, end from input-city.csv

Start-station_mostcommon_func
from city.csv
get start station
count
group by start
sort largest count

Print the first

End-station_mostcommon_func
from city.csv
get end station
count
group by end
sort largest count
Print the first

StartToEnd-stations_mostcommon_func
from city.csv
get start station and end station
count
group by start and end
sort largest count
Print the first

#3 Trip Duration

end_dates = Get end from city.csv, city2.csv
duration[] = calculate time between start_dates and end_dates
put in a list for each city

Trip-duration_sum_func
Print duration sum(duration[city])

Trip-duration_avg_func
Print avg time(duration[city])

#4 User Info

usertypes = store user types in a dictionary. ignore nan, null
genders = store gender in a dictionary if receive that does not exist - ignore
birthyear = store gender in a dictionary if receive that does not exist - ignore

usertype_count_func
count usertypes (city)
Print count per city

user_gender_func
count gender
Print count per city

by_early_func

Print ascending, show 1

by_recent
Print descending, show 1

by_common
count year
Print 1 per city descending