HUB DATA VISUALIZATION CHALLENGE[1]

Every time a Hubway[2] user checks a bike out from a station, the system records basic information about the trip. Those anonymous data points have been exported into the spreadsheet. Please note, all private data including member names have been removed from these files.

What can the data tell us?

The CSV file contains data for every Hubway (Blue Bikes, originally Hubway, is a bicycle sharing system in the Boston, Massachusetts metropolitan area) trip from the system launch on July the 28th of 2011, through the end of September 2012. The file contains the data points listed below for each trip. We've also posed some of the questions you must answer with this dataset - we're sure you'll have lots more of your own.

- **Duration** - Duration of trip.
  - *What's the average trip duration for annual members vs. casual users?*
- **Start date** - Includes start date and time.
  - *What are the peak Hubway hours?*
- **End date** - Includes end date and time.
  - *Which days of the week get the most Hubway traffic?*
- **Start station** - Includes starting station name and number.
  - *Which stations are most popular?*
  - *Which stations make up the most popular origin/destination pairs?*
- **End station** - Includes ending station name and number.
  - *Which stations are the most asymmetric - more trips start there than end there, or vice versa?*
  - *Are they all at the top of hills?*
- **Bike Nr** - Includes ID number of bike used for the trip.
  - *What does a year in the life of one Hubway bike look like?*
- **Member Type** - Lists whether user was an Annual or Casual (1 or 3 day) member.
  - *Which stations get the most tourist traffic, and*
  - *which get the most commuters?*
- **Zip code** - Lists the zip code for annual members only.
  - *How far does Hubway really reach?*
  - *Which community should be the next to get Hubway stations?*
- **Birthdate** - Lists the year in which annual members were born.
  - *Are all of the Hubway rentals at 2:00am by people under 25?*
- **Gender** - Lists gender for annual members only.
  - *Are there different top stations for male vs. female Hubway members?*

[1] Adapted from http://hubwaydatachallenge.org/trip-history-data/
[2] https://en.wikipedia.org/wiki/Blue_Bikes

The following is an example of the data file “hubway_trips.csv”

seq_id	hubway_id	status	duration	start_date	strt_statn	end_date	end_statn	bike_nr	subsc_type	zip_code	birth_date	gender
1	8	Closed	9	7/28/2011 10:12	23	7/28/2011 10:12	23	B00468	Registered	'97217	1976	Male
2	9	Closed	220	7/28/2011 10:21	23	7/28/2011 10:25	23	B00554	Registered	'02215	1966	Male
3	10	Closed	56	7/28/2011 10:33	23	7/28/2011 10:34	23	B00456	Registered	'02108	1943	Male
4	11	Closed	64	7/28/2011 10:35	23	7/28/2011 10:36	23	B00554	Registered	'02116	1981	Female
5	12	Closed	12	7/28/2011 10:37	23	7/28/2011 10:37	23	B00554	Registered	'97214	1983	Female
6	13	Closed	19	7/28/2011 10:39	23	7/28/2011 10:39	23	B00456	Registered	'02021	1951	Male
7	14	Closed	24	7/28/2011 10:47	23	7/28/2011 10:47	23	B00554	Registered	'02140	1971	Female
8	15	Closed	7	7/28/2011 10:48	23	7/28/2011 10:48	23	B00554	Registered	'02140	1971	Female
9	16	Closed	8	7/28/2011 11:01	23	7/28/2011 11:01	23	B00554	Registered	'97214	1983	Female
10	17	Closed	1108	7/28/2011 11:55	47	7/28/2011 12:13	40	B00550	Registered	'01867	1994	Male
11	18	Closed	1055	7/28/2011 11:55	47	7/28/2011 12:13	40	B00580	Registered	'01867	1956	Male
12	19	Closed	1042	7/28/2011 11:55	47	7/28/2011 12:12	40	B00539	Registered	'01867	1959	Female
13	23	Closed	994	7/28/2011 12:00	40	7/28/2011 12:16	47	B00368	Casual			
14	25	Closed	15	7/28/2011 12:00	22	7/28/2011 12:00	22	B00442	Registered	'02446	1982	Male
15	27	Closed	952	7/28/2011 12:00	40	7/28/2011 12:16	23	B00556	Registered	'02128	1944	Male
16	28	Closed	32	7/28/2011 12:00	22	7/28/2011 12:01	22	B00316	Registered	'02115	1960	Female
17	29	Closed	1261	7/28/2011 12:00	22	7/28/2011 12:21	45	B00454	Registered	'02492	1975	Male
18	30	Closed	7	7/28/2011 12:00	22	7/28/2011 12:00	22	B00404	Registered	'02446	1956	Male
19	31	Closed	1020	7/28/2011 12:01	38	7/28/2011 12:18	36	B00147	Registered	'02118	1987	Female
20	33	Closed	1264	7/28/2011 12:01	38	7/28/2011 12:22	44	B00165	Registered	'02139	1985	Female
21	34	Closed	1043	7/28/2011 12:01	38	7/28/2011 12:18	44	B00527	Registered	'02351	1978	Male

The following is an example of the data file “hubway_stations.csv”

id	terminal	station	municipal	lat	lng	status
3	B32006	Colleges of the Fenway	Boston	42.340021	-71.100812	Existing
4	C32000	Tremont St. at Berkeley St.	Boston	42.345392	-71.069616	Existing
5	B32012	Northeastern U / North Parking Lot	Boston	42.341814	-71.090179	Existing
6	D32000	Cambridge St. at Joy St.	Boston	42.361285	-71.06514	Existing
7	A32000	Fan Pier	Boston	42.353412	-71.044624	Existing
8	A32001	Union Square - Brighton Ave. at Cambridge St.	Boston	42.353334	-71.137313	Existing
9	A32002	Agganis Arena - 925 Comm Ave.	Boston	42.351313	-71.116174	Existing
10	A32003	B.U. Central - 725 Comm. Ave.	Boston	42.350075	-71.105884	Existing
11	A32004	Longwood Ave / Binney St	Boston	42.338629	-71.1065	Existing
12	B32002	Ruggles Station / Columbus Ave.	Boston	42.335911	-71.088496	Existing
13	C32002	Boston Medical Center - 721 Mass. Ave.	Boston	42.334057	-71.07403	Removed
14	B32003	HMS / HSPH - Ave. Louis Pasteur at Longwood Ave.	Boston	42.337171	-71.102797	Existing
15	A32005	Harvard Real Estate - Brighton Mills - 370 Western Ave	Boston	42.361667	-71.13802	Existing
16	C32003	Back Bay / South End Station	Boston	42.347433	-71.076163	Existing
17	A32006	Harvard University Housing - 111 Western Ave. at Soldiers Field Park	Boston	42.365074	-71.119581	Existing
18	A32007	Harvard Real Estate - 219 Western Ave. at North Harvard St.	Boston	42.36337	-71.129791	Existing
19	A32008	Buswell Park	Boston	42.347527	-71.105828	Existing

.

PROJECT REQUIREMENTS

1. Read about the Hubway project and understand the information generated by the project.
2. Using Jupyter Notebook or Google Colaboratory (https://colab.research.google.com/ [3]) answer each question using the knowledge acquired in class.
3. Each answer must include:
    a. Analysis of data
    b. Approach used to generate the answer.
    c. Python source code
    d. Data visualization
    e. Answer/conclusion
4. Make sure to use efficiently the available Python toolboxes. Use functions and appropriate data structures.
5. Make sure to generate the appropriate plots and use appropriate scales and labels.
6. Remember to use Python 3
7. Bonus point may be given to new questions formulated and answered by each team.

SUBMISSION

Students mus send a link to the google Colaboratory notebook. I you decide to work with Jupyter Notebook, make sure to import the Jupyter notebook with your test to the colaboratory notebook.

Deliverable

- Jupyter Notebook imported to google colaboratory with the answers according to the project requirement.

Deadline:
Friday, February 22 – 2019, 11:30PM

[3] Colaboratory is a research tool for machine learning education and research. It is a Jupyter notebook environment that requires no setup to use.