DataScienceBikes
360'
upload
Predict the net rate of bike renting

# Assignment

BeachBoys BikeShare is a bike share service provider where users can take and return bikes at any of the 70 stations on their network. The company wants to leverage their data to better understand and, hopefully, optimize their operations. They are interested in exploring the potential of machine learning to predict the number of bikes taken and returned to each station.

If this project is successful they would like to use the output of the predictive model to help the logistics department schedule the redistribution of bikes between stations. This in turn will help ensure there are bikes and return docks available when and where users need them.

To arrange this schedule, they need an estimation of the net change in the stock of bikes at a station for the time window between two pick-up/drop-off visits so the station may face the upcoming demand. Naturally, the number of visits for each station and thus these time windows will depend on the intensity of use of that station, the use of other stations in the network, as well as resources available.

As a first step towards tackling this challenge, the company has asked you to develop a model capable of outputting the net rate of bike renting for a given station (net rate is defined as trips ended minus trips started at the station for a given hour). That is, at any time the logistics team of the company should be able to make a statement such as *'In the next hour, the net stock of bikes at station A will change by X'*.

Results should be accompanied by the root-mean-square error (RMSE) metric as a measure of performance of each presented model. You are free to use other metrics to discuss a model's merit.

# Deliverables

You are asked to upload the following deliverables in a .zip file:

a) A PDF report of max. 5 pages detailing:

- an exploration of the data

- the modelling approach

- the performance of the model

- a discussion on potential improvements

- your conclusions

b) The source code used to produce the results

c) Relevant additional files, including a clear description of each file in the appendix of the report

Reproducible documents are encouraged, including, but not limited to, [Rmarkdown](#), [Jupyter Notebook](#) or [knitpy](#). Markdown and Rmarkdown templates for your report are being provided with the data. Please note that a rendered version of the report should always be included in PDF format. You are free to use the language and/or framework of your preference.

# Data

The company has collected information on the stations, trips taken, and on weather conditions in each of the cities where the scheme operated from September 2014 to August 2015. Below, you will find detailed information on all the fields available in the dataset. The way you include this information in your model is up to you and should be clearly justified and documented in your report. You free to use any other data sources provided you specify a link to this information in your report.

All the provided data is also publicly available on bayareabikeshare.com/open-data.

# Station information

- Id: station ID number

- Name: name of station

- Lat: latitude

- Long: longitude

- Dock Count: number of total docks at station

- City: one of San Francisco, Redwood City, Palo Alto, Mountain View, or San Jose

Note: During the period covered by the dataset, several stations were moved. Stations 23, 25, 49, 69, and 72 became respectively stations 85, 86, 87, 88, 89 (which in turn became 90 after a second move).

# Trip data

- Trip ID: numeric ID of bike trip

- Duration: time of trip in seconds

- Start Date: start date of trip with date and time, in Pacific Standard Time

- Start Station: station id of start station

- Start Terminal: numeric reference for start station

- End Date: end date of trip with date and time, in Pacific Standard Time

- End Station: station id for end station

- Subscription Type: Subscriber (annual or 30-day member) or Customer (24-hour or 3-day member)

# Daily weather measurements

- Date: day for which the weather is being reported

- Temperature (day min, mean and max): in °F

- Dew point (day min, mean and max): Temperature in °F below which dew can form

- Humidity (day min, mean and max): in %

- Pressure (day min, mean and max): Atmospheric pressure at sea level in inches of mercury

- Visibility (day min, mean and max): distance in miles

- Wind Speed (day max and mean): in mph

- Max Gust Speed: in mph

- Precipitation: total amount of precipitations in inches

- Cloud Cover: scale of 0 (clear) to 8 (totally covered)

- Events: Special meteorological events

- Wind Direction: in degrees

- Zip: area code for San Francisco (94107), Redwood City (94063), Palo Alto (94301), Mountain View (94041), and San Jose (95113)