**Codebook for Bay Area bike-sharing Dataset**

# Data Overview

## Credentials

This data set can be downloaded from BigQuery – Google’s data warehouse.

* The direct link to exploring this data in the BigQuery console is: <https://cloud.google.com/bigquery/public-data/bay-bike-share>
* The direct link to exported table is (using SQL & R skills):  
  https://github.com/AmitaiSe/Business-Data-Analytics-Group-5/blob/master/bikeshare\_trips.csv

## Business goal

This data was collected to answer the question:

## **How to increase usage of bike-sharing in San Francisco?**

In order to answer this question we will go through some leading questions such as:

* What are the busiest stations per time of day?
* What is the busiest time per day?
* How can we predict stations ought to run out of bikes?
* How can we predict stations where no rental will occur?
* What is the average riding time?
* How rental activity varies between customer and subscriber?

## Data description

This data set is a data frame of 11 variables over 100,000 rows.  
Each row represent a rider’s trip on rented bike from Bay Area Bike Share Company in San Francisco.  
There are no missing values in the data.

We will combine this data with a complimentary table showcasing availability of bikes / docks.

# Variables description

# Table 1:

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Type** | **Description** |
| **trip\_id** | INTEGER | Numeric ID of bike trip |
| **duration\_sec** | INTEGER | Time of trip in seconds |
| **start\_date** | TIMESTAMP | Start date of trip with date and time, in PST |
| **start\_station\_name** | STRING | Station name of start station |
| **start\_station\_id** | INTEGER | Numeric reference for start station |
| **end\_date** | TIMESTAMP | End date of trip with date and time, in PST |
| **end\_station\_name** | STRING | Station name for end station |
| **end\_station\_id** | INTEGER | Numeric reference for end station |
| **bike\_number** | INTEGER | ID of bike used |
| **zip\_code** | STRING | Home zip code of subscriber (customers can choose to manually enter zip at kiosk however data is unreliable) |
| **subscriber\_type** | STRING | Subscriber = annual or 30-day member; Customer = 24-hour or 3-day member |

# Table 2:

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Type** | **Description** |
| **station\_id** | INTEGER | Station ID number |
| **bikes\_available** | INTEGER | Number of available bikes |
| **docks\_available** | INTEGER | Number of available docks |
| **time** | TIMESTAMP | Date and time, PST |