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://bigquery.cloud.google.com/table/bigquery-public-data:san_francisco.bikeshare_trips

Business goal

This data was collected to answer the question:

How can we predict the target destination/station of the next rider?

Data description

This data set is a data frame of 11 variables over 983,648 rows.
Each row represent a bike rental from Bay Area Bike-Share Company in California.

We filtered this data set to the city of San Francisco only, where most of the rentals occurred, and got updated data frame of **891,200** rows.

There are no missing values in the data.

We combined this data with complimentary tables as shown below.

Variables description

San Francisco Bay Area bike-sharing trips

Variable Name	Type	Description
trip_id	Integer	Numeric ID of bike trip
duration_sec	Integer	Time of trip in seconds
start_date	Character	Start date of trip with date and time, in PST
start_station_name	Character	Station name of start station
start_station_id	Integer	Numeric reference for start station
end_date	Character	End date of trip with date and time, in PST
end_station_name	Character	Station name for end station
end_station_id	Integer	Numeric reference for end station
bike_number	Integer	ID of bike used
zip_code	Integer	Home zip code of subscriber (customers can choose to manually enter zip at kiosk however data is unreliable)
subscriber_type	Integer	Subscriber = annual or 30-day member; Customer = 24-hour or 3-day member

Complimentary Data Overview

Links to complimentary data:

- San Francisco Bay Area bike's stations
  https://github.com/YardenIsraeli/Business-Analytics-Bikeshare-Project/blob/master/dataSets/sanFrancisco/sanFran_stations.csv

- San Francisco Bart (Bay Area Rapid Transit) stations
  https://github.com/YardenIsraeli/Business-Analytics-Bikeshare-Project/blob/master/dataSets/raw_data/Stops%20bart.csv

- San Francisco Golden Gate Transit stations
  https://github.com/YardenIsraeli/Business-Analytics-Bikeshare-Project/blob/master/dataSets/raw_data/Stops%20goldengatetransit.csv

- San Francisco Muni (Municipal Railway) stations
  https://github.com/YardenIsraeli/Business-Analytics-Bikeshare-Project/blob/master/dataSets/raw_data/Stops%20muni.csv

- Bike-share stations Availability
  https://bigquery.cloud.google.com/table/bigquery-public-data:san_francisco.bikeshare_status

Variables description

San Francisco Bay Area bike's stations

Variable Name	Type	Description
station_id	Integer	Station ID number
Name	Character	Name of station
latitude	Float	Latitude of station
longitude	Float	Longitude of station
Dockcount	Integer	Number of total docks at station
landmark	Character	City (San Francisco)
Installation_data	Character	Original date that station was installed

San Francisco Bart (Bay Area Rapid Transit) stations

Variable Name	Type	Description
stop_id	Character	Station ID
stop_name	Character	Station name
stop_lat	Numeric	Latitude of stop station
stop_lon	Numeric	Longitude of stop station

San Francisco Golden Gate Transit stations

Variable Name	Type	Description
stop_id	Character	Station ID
stop_name	Character	Station name
stop_lat	Numeric	Latitude of stop station
stop_lon	Numeric	Longitude of stop station

San Francisco Muni (Municipal Railway) stations

Variable Name	Type	Description
stop_id	Character	Station ID
stop_name	Character	Station name
stop_lat	Numeric	Latitude of stop station
stop_lon	Numeric	Longitude of stop station

Bike-share stations Availability

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	Character	Date and time, PST