

7) Bike Sharing in the Bay Area

Vitor Kamada

December 2019

Reference

Tables, Graphics, and Figures from

Computational and Inferential Thinking: The Foundations of Data Science

Adhikari & DeNero (2019): Ch 8.5 Bike Sharing
in the Bay Area

<https://www.inferentialthinking.com/>

Bay Area Bike Share Dataset

```
pip install datascience
```

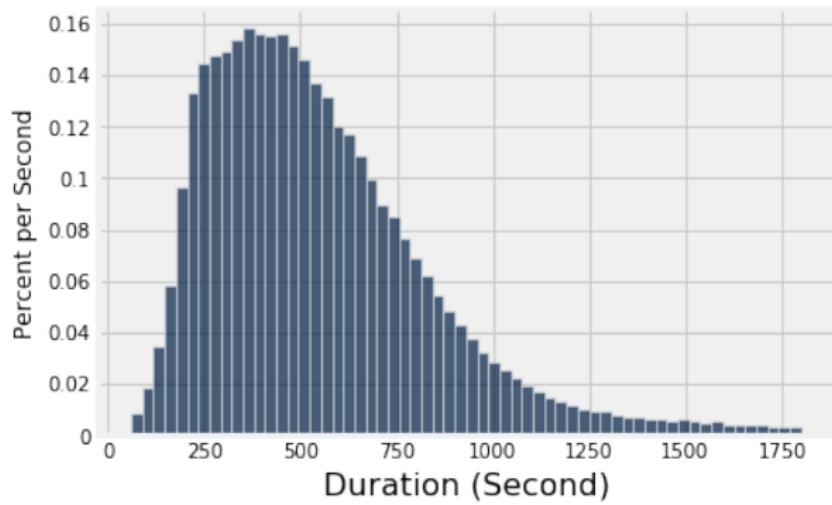
```
from datascience import *
path_data = 'https://github.com/data-8/textbook/raw/gh-pages/data/'
trips = Table.read_table(path_data + 'trip.csv')
```

Trip ID	Duration	Start Date	Start Station	Start Terminal	End Date	End Station
913460	765	8/31/2015 23:26	Harry Bridges Plaza (Ferry Building)	50	8/31/2015 23:39	San Francisco Caltrain (Townsend at 4th)
913459	1036	8/31/2015 23:11	San Antonio Shopping Center	31	8/31/2015 23:28	Mountain View City Hall

354,152 rentals between 9/2014 to 8/2015

30 min = 1800 secs = Free Trip

```
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
%matplotlib inline
commute = trips.where('Duration', are.below(1800))
commute.hist('Duration', bins=60, unit='Second')
```



```
commute.group(['Start Station', 'End Station'])
```

Start Station	End Station	count
2nd at Folsom	2nd at Folsom	54
2nd at Folsom	2nd at South Park	295
2nd at Folsom	2nd at Townsend	437

```
commute.pivot('Start Station', 'End Station')
```

	2nd		
End Station	2nd at Folsom	at South Park	2nd at Townsend

2nd at Folsom	54	190	554
---------------	----	-----	-----

2nd at South	295	164	
--------------	-----	-----	--

Shortest Time of the Rides

```
commute.pivot('Start Station', 'End Station', 'Duration', min)
```

End Station	2nd at Folsom	at South Park	2nd at Townsend
-------------	---------------	---------------	-----------------

2nd at Folsom	61	97	164
---------------	----	----	-----

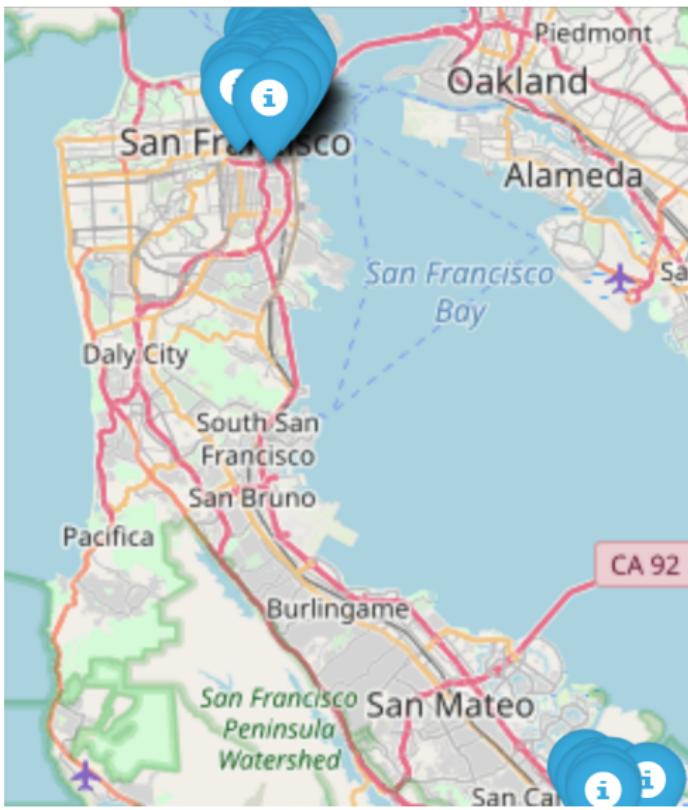
2nd at South Park	61	60	77
-------------------	----	----	----

Stations Data

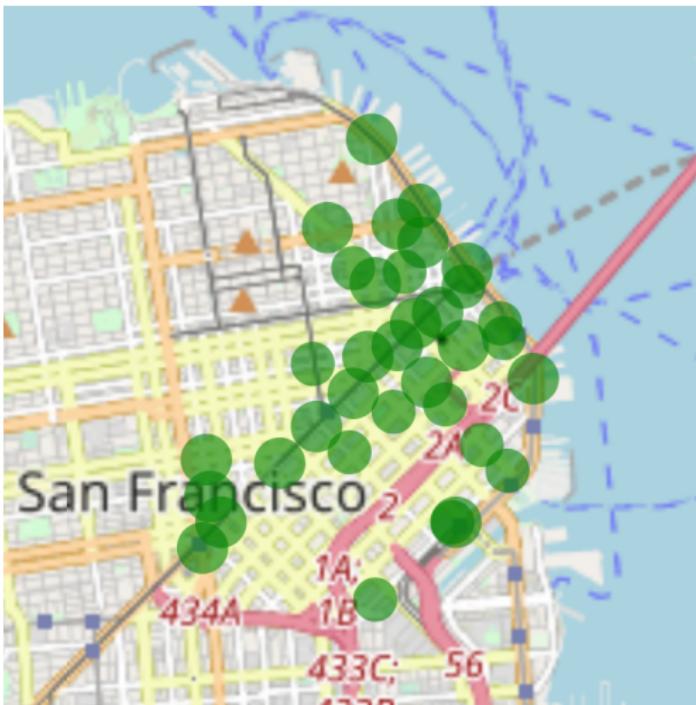
```
stations = Table.read_table(path_data + 'station.csv')
```

station_id	name	lat	long	dockcount	landmark	installation
2	San Jose Diridon Caltrain Station	37.3297	-121.902	27	San Jose	8/6/2013
3	San Jose Civic Center	37.3307	-121.889	15	San Jose	8/5/2013

```
Marker.map_table(stations.select('lat', 'long', 'name'))
```



```
sf = stations.where('landmark', are.equal_to('San Francisco'))  
sf_map_data = sf.select('lat', 'long', 'name')  
Circle.map_table(sf_map_data, color='green', radius=200)
```

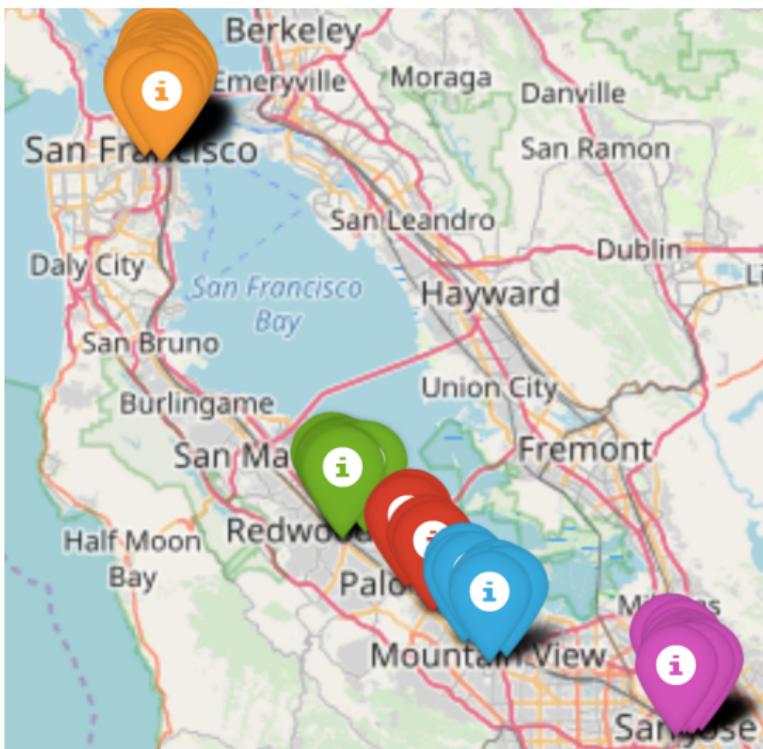


```
cities = stations.groupby('landmark').relabeled('landmark', 'city')
```

```
colors = cities.with_column('color',  
    make_array('blue', 'red', 'green', 'orange', 'purple'))
```

city	count	color
Mountain View	7	blue
Palo Alto	5	red
Redwood City	7	green
San Francisco	35	orange
San Jose	16	purple

```
joined = stations.join('landmark', colors, 'city')
colored = joined.select('lat', 'long', 'name', 'color')
Marker.map_table(colored)
```



```
starts = commute.groupby('Start Station').sort('count', descending=True)
```

	Start Station	count
San Francisco Caltrain (Townsend at 4th)	25858	
San Francisco Caltrain 2 (330 Townsend)	21523	
Harry Bridges Plaza (Ferry Building)	15543	

```
station_starts = stations.join('name', starts, 'Start Station')
```

name	station_id	lat	long	dockcount	landmark	installation
2nd at Folsom	62	37.7853	-122.396	19	San Francisco	8/22/2013
2nd at South Park	64	37.7823	-122.393	15	San Francisco	8/22/2013

Eastern Section: Capital of Bike Rentals

```
starts_map_data = station_starts.select('lat',
    'long', 'name').with_columns('color', 'blue',
    'area', station_starts.column('count') * 1000
)
starts_map_data.show(3)
Circle.map_table(starts_map_data)
```

lat	long	name	color	area
37.7853	-122.396	2nd at Folsom	blue	7841000
37.7823	-122.393	2nd at South Park	blue	9274000

