Citi Bike Docking Station Hourly Scoring

Chunqing Xu
Graduate Student
Center for Urban Science + Progress
NYU

Data

Citi Bike Trip Histories from Citi Bike System Data Web Page

Start Time, Stop Time,

Start Station ID, End Station ID

Scoring Function

1. Data Processing

Each station, Each hour

One trip arriving **+1**, One trip leaving **-1**

Net Value = **1** x (Number of Trips Arriving) **-1** x (Number of Trips Leaving)

Scoring Function

2.Formula

- For each station, at time N

  Prescore at N = Net Value at N-1 + Net Value at N-2 + Net Value at N-3

- For all stations, at time N

  Prescores at N would be a vector

- Sorted Prescores in the vector

Scoring Function

3.Scoring

- Set the scoring standard:

Score	-3	-2	-1	0	1	2	3
Percentage	%5	%10	%10	50%	10%	10%	5%

- Match the sorted prescores vector with the scoring standard

Enough of theories! How does it operate?

Demo

1.Data Processing

	tripduration	starttime	stoptime	startstationid	endstationid	bikeid	usertype	birthyear	gender
499999	754	2017-03-08 16:58:07	2017-03-08 17:10:42	3090	3117	26443	Subscriber	1964.0	2
499998	199	2017-03-08 16:58:09	2017-03-08 17:01:29	379	486	17720	Subscriber	1959.0	1
499997	1106	2017-03-08 16:58:09	2017-03-08 17:16:36	398	398	24918	Subscriber	1985.0	2
499996	364	2017-03-08 16:58:10	2017-03-08 17:04:14	434	368	27163	Subscriber	1983.0	1
499995	1386	2017-03-08 16:58:10	2017-03-08 17:21:17	457	3320	18154	Subscriber	1977.0	1

	2017030816	2017030817	2017030818	2017030819	2017030820	2017030821	2017030822	2017030823	2017030900	2017030901	...	20170401
72	1.0	2.0	-3.0	2.0	-2.0	0.0	0.0	0.0	0.0	1.0	...	0.0
79	1.0	0.0	0.0	2.0	-3.0	1.0	1.0	-2.0	0.0	2.0	...	0.0
82	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
83	2.0	-2.0	0.0	0.0	1.0	0.0	0.0	2.0	-3.0	0.0	...	0.0
116	0.0	1.0	4.0	-5.0	2.0	0.0	-2.0	0.0	0.0	1.0	...	0.0

Demo
2.Scoring Function

score(dh, df, w)

dh - **Date Hour** to score “2017030821”

df - **Data Frame of Prescores** (Station ID x Date Hour)

w - **Scoring Standard** w = [0.05, 0.1, 0.1, 0.5, 0.1, 0.1, 0.05]

Demo

3.Scoring example

score(dh, df, w)

dh - '2017030821' - "2017030905"

df - Prescores of 616 Citi Bike Stations at N-1, N-2 and N-2

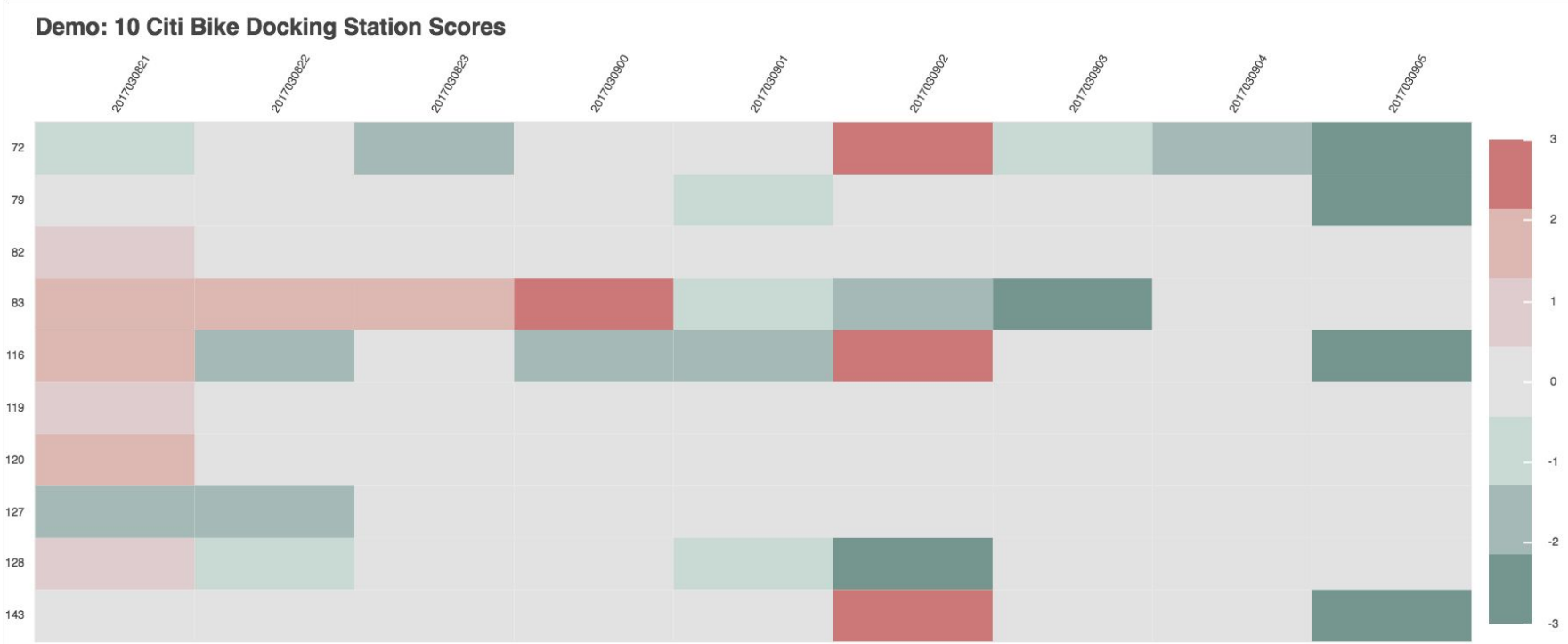
w - w = [0.05, 0.1, 0.1, 0.5, 0.1, 0.1, 0.05]

Demo
4.Scoring Table

	StationID	2017030821	2017030822	2017030823	2017030900	2017030901	2017030902	2017030903	2017030904	2017030905
0	72	-1	0	-2	0	0	3	-1	-2	-3
1	79	0	0	0	0	-1	0	0	0	-3
2	82	1	0	0	0	0	0	0	0	0
3	83	2	2	2	3	-1	-2	-3	0	0
4	116	2	-2	0	-2	-2	3	0	0	-3

Demo

5.Visualization (interactive)



Conclusion

Score of a certain hour of Citi Bike Stations

=

Scoring standard

+

Prescores of the past three hours *

Improvement on Scoring Function

Normalized Prescores Table

	id	totalDocks
0	72	39
1	79	33
2	82	27
3	83	62
4	116	39

Future Works

Scoring Function

Panel Regression

ARMA