



Bike-sharing **target location prediction**

Data Analytics for Business



The bike-sharing BOOM

A group of people, including adults and a child, are riding bicycles on a city street. The background shows European-style buildings and traffic lights. The scene is captured in a cinematic style with soft lighting.

The bike-sharing industry is exploding

Offering short-term bicycle rental at urban locations for minimal price

A high-angle photograph showing a vast number of green bicycles parked in neat, long rows on a paved surface. The bicycles are arranged in a grid-like pattern, filling most of the frame. In the lower right corner, a man in a light blue uniform is riding one of the bicycles. The text "50 million people are regular users !" is overlaid in the center of the image.

50 million
people are
regular
users !

The bike-sharing BOOM

Millions of users
BUT

**No company has yet been able to make profit,
using only revenues generated from membership
subscription or user fees**

OR
A bubble is about to explode ??

The bike-sharing BOOM

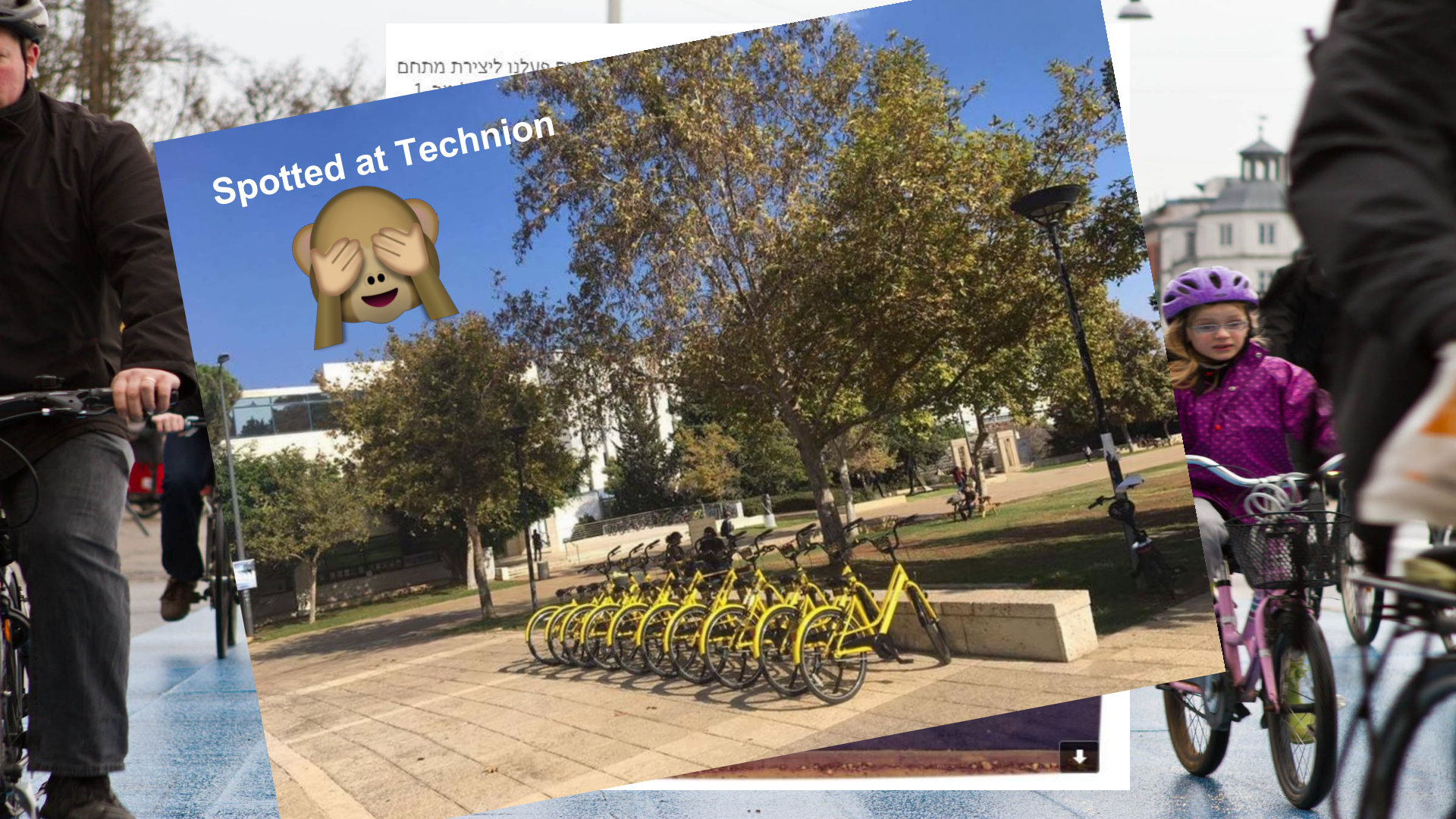
A GOLDEN OPPORTUNITY

We can feel it ! Can you too ?

Spotted at Technion



פוטולנו ליצירת מתחם
1.5





To cut long story short ...

A group of people, including adults and a child, are riding bicycles on a city street. The street has a blue-painted crosswalk. In the background, there are European-style buildings and traffic lights. A large, stylized graphic is overlaid in the center of the image. The graphic consists of a purple triangle with a gold glittery interior. Inside the triangle, the word "KNOWLEDGE" is written in blue capital letters, "is" is written in a gold cursive font, and "POWER" is written in pink capital letters. The entire graphic has a white dotted border.

KNOWLEDGE
is
POWER

To cut long story short ...



Who We Are

#datascience

#businessthinker

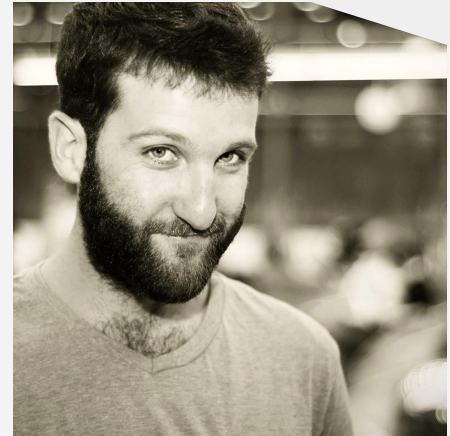
#materna#cheaf#datascience



Omer Klein



Yarden Israeli



Amitai Serphos

What is our SuperPower ?

A group of people, including men, women, and a child, are riding bicycles on a city street. They are wearing winter clothing like jackets and scarves. In the background, there are European-style buildings and traffic lights. The scene is captured in a cinematic style with soft lighting.

**Predict DESTINATION
of the next rider to use the bikeshare**

Given rental information

Our DATA

```
## 'data.frame':    172966 obs. of  18 variables:
## $ trip_id      : int  1059051 1311233 1072279 1105120 1258908 1122795 1236506 1143485 1294058 1218986 ...
## $ start_station_id: int  60 74 59 70 69 69 57 45 48 54 ...
## $ start_lat     : num  37.8 37.8 37.8 37.8 37.8 ...
## $ start_lon     : num  -122 -122 -122 -122 -122 ...
## $ duration_minutes: num  8.65 27.7 13.42 9.32 6.63 ...
## $ start_dockcount : int  15 23 23 19 23 23 15 15 15 15 ...
## $ month         : int  1 8 1 2 6 3 6 3 7 5 ...
## $ day_status    : int  1 1 1 1 1 1 0 1 1 0 ...
## $ hour          : int  18 7 18 7 8 7 15 17 9 14 ...
## $ minDist_bart  : num  1429 267 433 1419 1405 ...
## $ minDist_caltrain: num  3196.1 1929.2 2157.6 44 62.2 ...
## $ minDist_ferry : num  1352 171 2717 696 715 ...
## $ minDist_muni  : num  1891 503 2427 1466 1470 ...
## $ zipcode       : int  4555 94538 94610 94401 94303 4561 94105 4578 94105 94105 ...
## $ bike_id       : int  677 675 378 332 274 659 420 451 488 455 ...
## $ end_station_id : int  77 65 50 47 57 50 62 74 54 68 ...
## $ X_end         : num  552757 552603 553335 552525 552383 ...
## $ Y_end         : num  4182643 4180582 4183287 4182569 4181767 ...
```


The first data scientist (EDA)

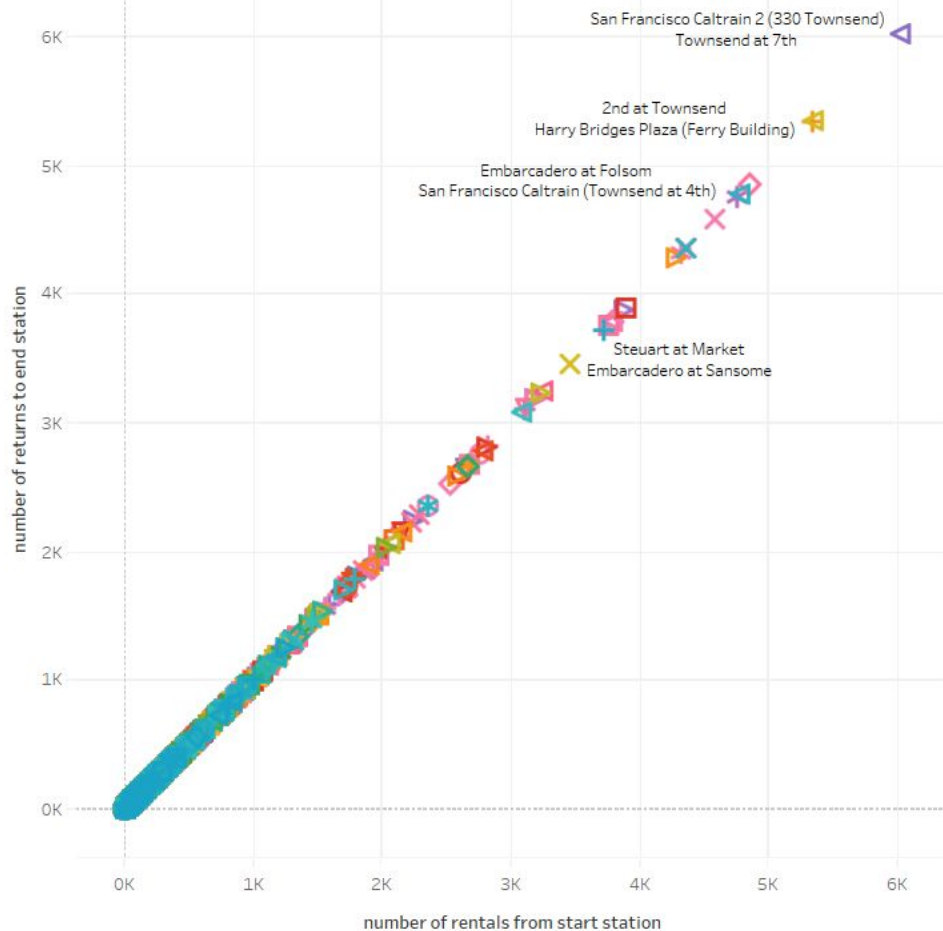
⁶ E. W. Gilbert, "Pioneer Maps of Health and Disease in England," *Geographical Journal*, 124 (1958), 172–183. Shown here is a redrawing of John Snow's map. For a reproduction and detailed analysis of the original map, see Edward Tufte, *Visual Explanations: Images and Quantities, Evidence and Narrative* (Cheshire, Connecticut, 1997), Chapter 2. Ideally, see John Snow, *On the Mode of Communication of Cholera* (London, 1855).



Using static maps?!



Cycling Routes



Start Station Name

- 2nd at Folsom
- 2nd at South Park
- + 2nd at Townsend
- × 5th at Howard
- * 5th St at Folsom ..
- ◇ Beale at Market
- △ Broadway St at B..
- ▽ Civic Center BAR..
- ◁ Clay at Battery
- ▷ Commercial at M..
- Cyril Magnin St a..
- Davis at Jackson
- .

End Station Name

- 2nd at Folsom
- 2nd at South Park
- 2nd at Townsend
- 5th at Howard
- 5th St at Folsom St
- Beale at Market
- Broadway St at B..
- Civic Center BART..
- Clay at Battery
- Commercial at M..
- Cyril Magnin St at..
- Davis at Jackson
-

Classical problem of supervised learning

##	tree_test_results										
##		47	50	61	66	68	73	74	76	77	90
##	47	0	78	24	0	0	0	5	0	71	0
##	50	0	687	43	0	0	0	96	0	47	0
##	61	0	69	64	0	0	0	40	0	68	0
##	66	0	83	72	0	0	0	0	0	28	0
##	68	0	102	9	0	0	0	5	0	60	0
##	73	0	12	60	0	0	0	7	0	13	0
##	74	0	433	37	0	0	0	108	0	77	0
##	76	0	87	20	0	0	0	19	0	129	0
##	77	0	405	39	0	0	0	11	0	156	0
##	90	0	5	3	0	0	0	1	0	3	0

##	stations_id		TP	FN	FP
##	[1,]	"47"	"0"	"178"	"0"
##	[2,]	"50"	"687"	"186"	"1274"
##	[3,]	"61"	"64"	"177"	"307"
##	[4,]	"66"	"0"	"183"	"0"
##	[5,]	"68"	"0"	"176"	"0"
##	[6,]	"73"	"0"	"92"	"0"
##	[7,]	"74"	"108"	"547"	"184"
##	[8,]	"76"	"0"	"255"	"0"
##	[9,]	"77"	"156"	"455"	"496"
##	[10,]	"90"	"0"	"12"	"0"

```
recall = mean(recall4each)
recall
```

```
## [1] 0.1472706
```


RF_test_results															
	47	50	61	66	68	73	74	76	77	90		stations_id	TP	FN	FP
47	16	82	9	13	0	2	4	13	39	0	[1,]	"47"	"16"	"162"	"18"
50	0	769	16	16	0	1	47	4	20	0	[2,]	"50"	"769"	"104"	"1315"
61	0	94	51	28	0	1	19	15	33	0	[3,]	"61"	"51"	"190"	"63"
66	0	84	1	71	1	2	1	4	19	0	[4,]	"66"	"71"	"112"	"131"
68	7	105	8	2	9	1	3	17	24	0	[5,]	"68"	"9"	"167"	"1"
73	0	15	7	14	0	40	5	5	6	0	[6,]	"73"	"40"	"52"	"13"
74	2	427	0	20	0	3	176	6	21	0	[7,]	"74"	"176"	"479"	"115"
76	2	94	11	10	0	3	24	68	43	0	[8,]	"76"	"68"	"187"	"81"
77	7	408	11	25	0	0	12	16	132	0	[9,]	"77"	"132"	"479"	"207"
90	0	6	0	3	0	0	0	1	2	0	[10,]	"90"	"0"	"12"	"0"

```
recall_rf = mean(recall4each_rf)
recall_rf
```

```
## [1] 0.2807682
```

From

classification

to

regression

Step 1 regression

Model	Error X	Error Y	Error XY
Comlex Tree	572.36 m	806.75 m	989.16 m
RF	505.7 m	713.6 m	874.6 m

~11%



Step 2 regression

Model	Error XY
Tree	1130.5 m
MV Tree	785.9 m

~30%







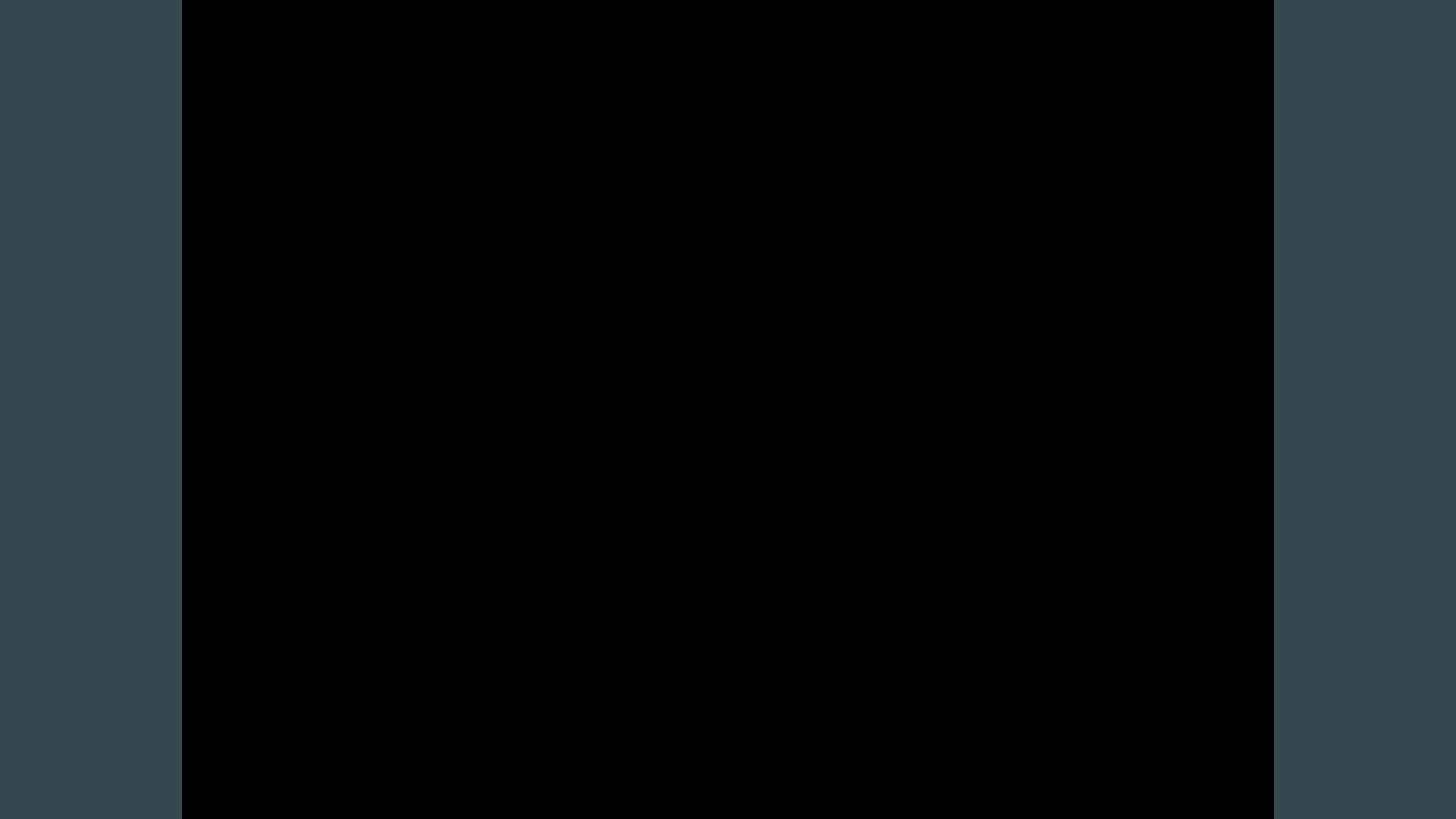
Destination Based Ad Network

Pre order your coffee



WAZE
who?!







Thank you