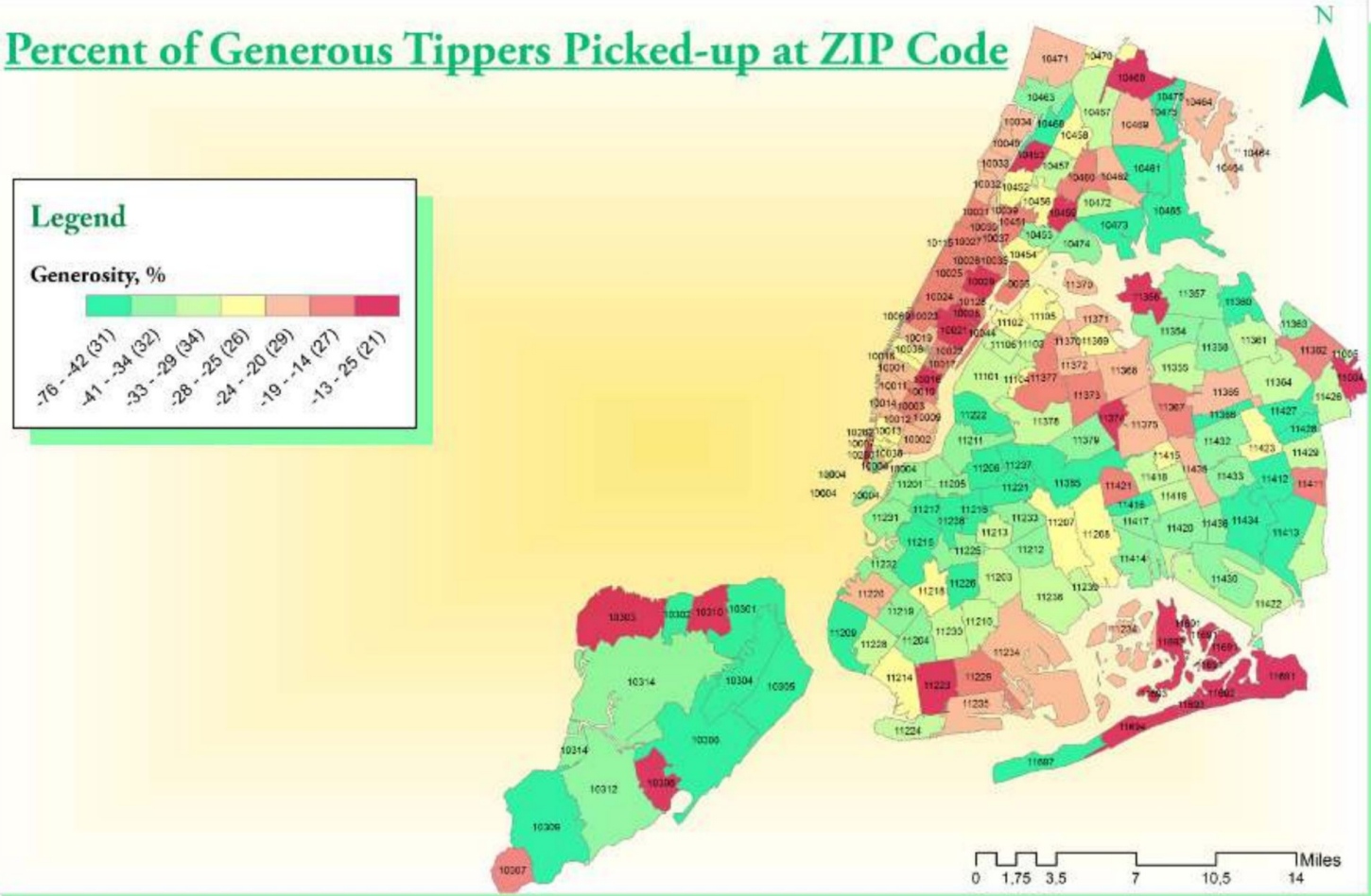
Vendor ID	passenger _count	trip_ distance	pickup_ longitude	pickup_ latitude	dropoff_ longitude	dropoff_ latitude	payment _type	tip_ amount	total_ amount
2	1	7.22	-73.9998	40.74334	-73.9428	40.80662	2	0	30.8
1	1	2.3	-73.977	40.7749	-73.9783	40.74986	1	2.93	16.23
1	1	1.5	-73.9591	40.77513	-73.9804	40.78231	1	1.65	9.95
1	1	0.9	-73.9766	40.78075	-73.9706	40.78885	1	1.45	8.75
2	1	2.44	-73.9786	40.78592	-73.9974	40.7563	1	2	16.3
2	1	3.36	-73.9764	40.78589	-73.9424	40.82209	1	3.58	17.88
2	2	2.34	-73.9862	40.76087	-73.9569	40.77156	1	1	13.8
2	1	10.19	-73.79	40.64406	-73.9312	40.67588	2	0	32.8
1	2	3.3	-73.9937	40.72738	-73.9982	40.7641	1	2	21.3
1	1	1.8	-73.9949	40.74006	-73.9767	40.74934	1	1.85	11.15

Ukážka z data setu “NY city taxi cab”
Skiena: Data Science Design Manual

variable	LR coefficient
(intercept)	0.08370835
duration	0.00000035
distance	0.00000004
fare	0.17503086
tolls	0.06267343
surcharge	0.01924337
weekends	−0.02823731
business day	0.06977724
rush hour	0.01281997
late night	0.04967453
# of passengers	−0.00657358

Percent of Generous Tippers Picked-up at ZIP Code



variable	LR coefficient	LASSO
(intercept)	0.08370835	0.079601141
duration	0.00000035	0.00000035
distance	0.00000004	0.00000004
fare	0.17503086	0.17804921
tolls	0.06267343	0.00000000
surcharge	0.01924337	0.00000000
weekends	−0.02823731	0.00000000
business day	0.06977724	0.00000000
rush hour	0.01281997	0.00000000
late night	0.04967453	0.00000000
# of passengers	−0.00657358	0.00000000