

Overview

Brought to you by [YData](#)

Overview

Alerts 10

Reproduction

Dataset statistics

Number of variables	11
Number of observations	999999
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	316.6 MiB
Average record size in memory	332.0 B

Variable types

Text	1
Categorical	1
DateTime	2
Numeric	6
Boolean	1

Variables

Select Columns



id

Text

Unique

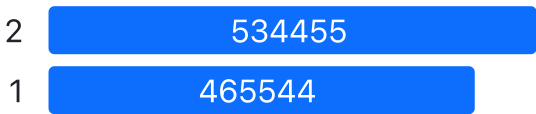
Distinct	999999
Distinct (%)	100.0%
Missing	0
Missing (%)	0.0%
Memory size	62.9 MiB



vendor_id

Categorical

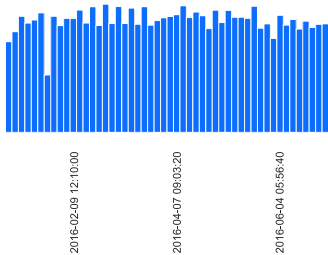
Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	55.3 MiB



pickup_datetime

Date

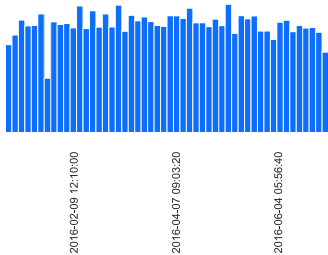
Distinct	962811	Minimum	2016-01-01 00:00:00
Distinct (%)	96.3%	Maximum	2016-06-30 23:59:59
Missing	0	Invalid dates	0
Missing (%)	0.0%	Invalid dates (%)	0.0%
Memory size	7.6 MiB		



dropoff_datetime

Date

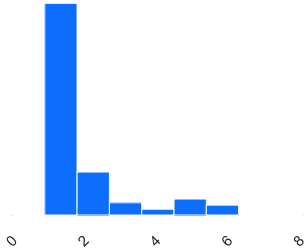
Distinct	962603	Minimum	2016-01-01 00:00:00
Distinct (%)	96.3%	Maximum	2016-07-01 23:59:59
Missing	0	Invalid dates	0
Missing (%)	0.0%	Invalid dates (%)	0.0%
Memory size	7.6 MiB		



passenger_count

Real number (\mathbb{R})

Distinct	10	Minimum	0
Distinct (%)	< 0.1%	Maximum	9
Missing	0	Zeros	37
Missing (%)	0.0%	Zeros (%)	< 0.1%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	1.6645377	Memory size	7.6 MiB

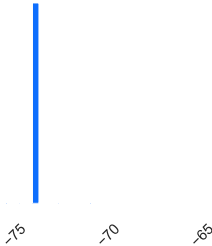


pickup_longitude

Real number (\mathbb{R})

High correlation Skewed

Distinct	21149	Minimum	-78.547401
Distinct (%)	2.1%	Maximum	-61.335529
Missing	0	Zeros	0
Missing (%)	0.0%	Zeros (%)	0.0%
Infinite	0	Negative	999999
Infinite (%)	0.0%	Negative (%)	100.0%
Mean	-73.973426	Memory size	7.6 MiB

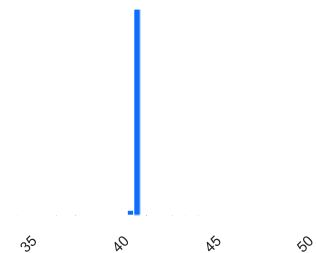


pickup_latitude

Real number (\mathbb{R})

High correlation

Distinct	42234	Minimum	34.359695
Distinct (%)	4.2%	Maximum	51.881084
Missing	0	Zeros	0
Missing (%)	0.0%	Zeros (%)	0.0%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	40.750937	Memory size	7.6 MiB



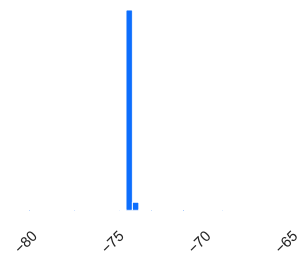
dropoff_longitude

Real number (\mathbb{R})

High correlation

Skewed

Distinct	30526	Minimum	-79.817978
Distinct (%)	3.1%	Maximum	-61.335529
Missing	0	Zeros	0
Missing (%)	0.0%	Zeros (%)	0.0%
Infinite	0	Negative	999999
Infinite (%)	0.0%	Negative (%)	100.0%
Mean	-73.973351	Memory size	7.6 MiB

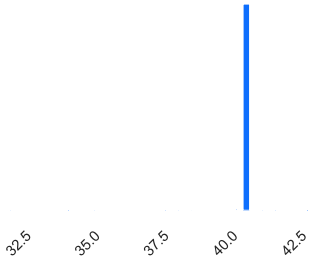


dropoff_latitude

Real number (ℝ)

High correlation Skewed

Distinct	57733	Minimum	32.181141
Distinct (%)	5.8%	Maximum	43.911762
Missing	0	Zeros	0
Missing (%)	0.0%	Zeros (%)	0.0%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	40.751824	Memory size	7.6 MiB



store_and_fwd_flag

Boolean

Imbalance

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	976.7 KiB

False	994459
True	5540

trip_duration

Real number (ℝ)

Skewed

Distinct	6798	Minimum	1
Distinct (%)	0.7%	Maximum	3526282
Missing	0	Zeros	0
Missing (%)	0.0%	Zeros (%)	0.0%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	962.942	Memory size	7.6 MiB



Interactions

passenger_count

pickup_longitude

pickup_latitude

dropoff_longitude

dropoff_latitude

trip_duration

trip_duration

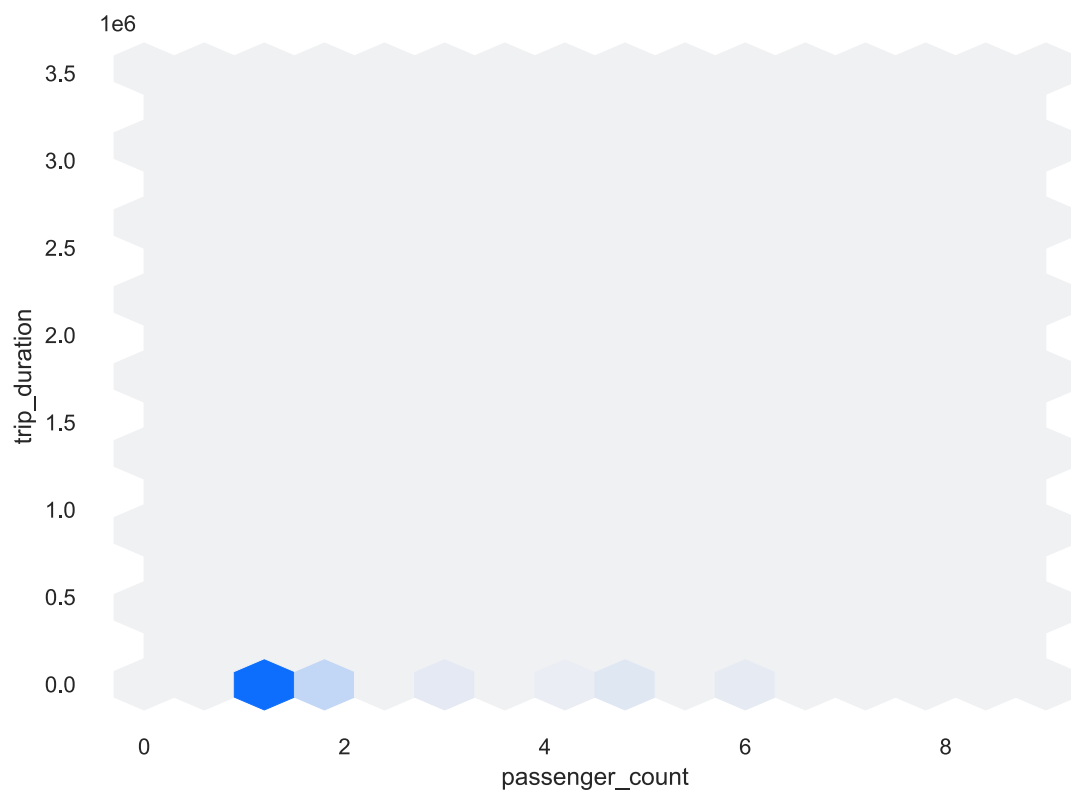
passenger_count

pickup_longitude

pickup_latitude

dropoff_longitude

dropoff_latitude

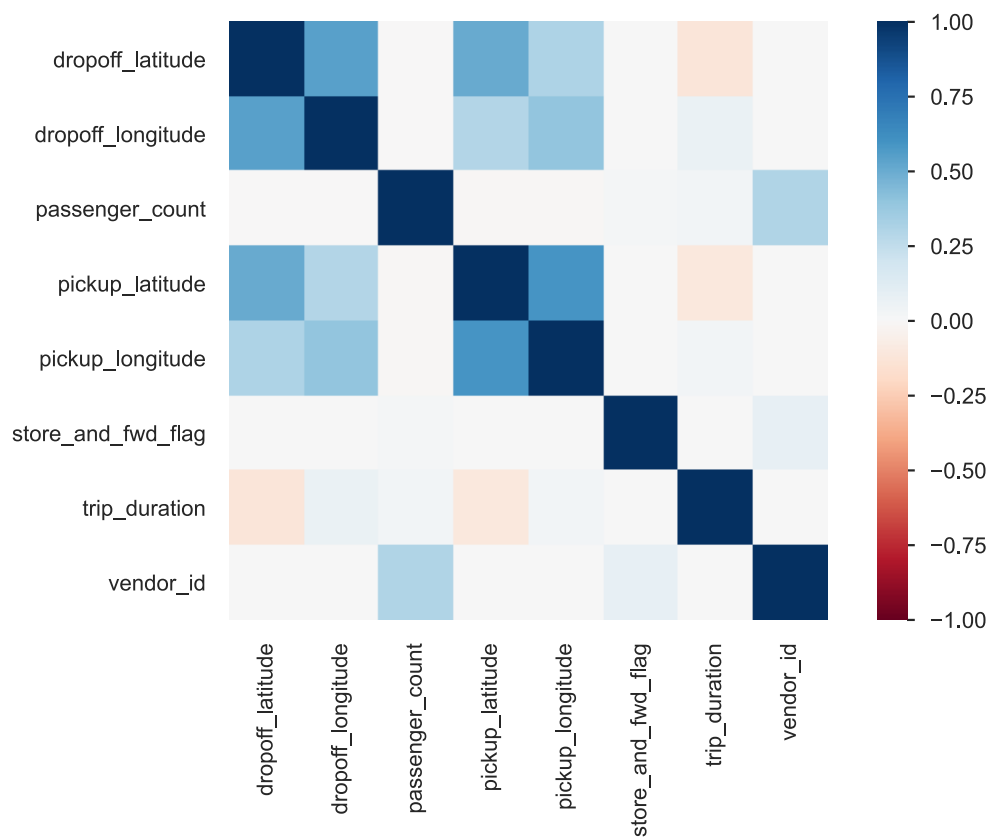


Correlations

Auto

Heatmap

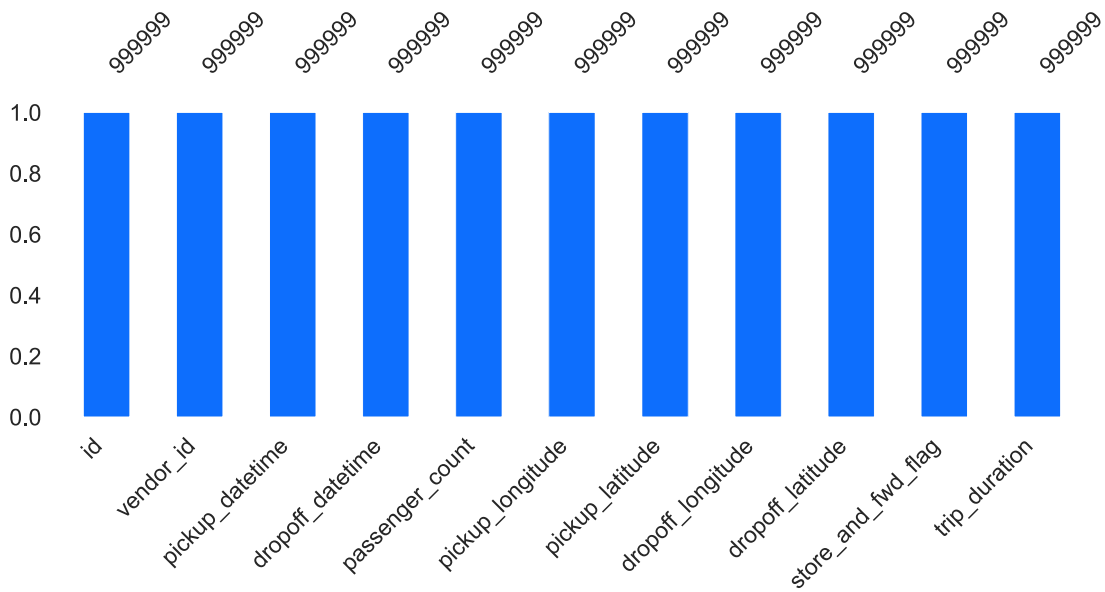
[Table](#)



Missing values

Count

[Matrix](#)



A simple visualization of nullity by column.

Sample

First rows

[Last rows](#)

	id	vendor_id	pickup_datetime	dropoff_datetime	passenger
0	id2875421	2	2016-03-14 17:24:55	2016-03-14 17:32:30	1
1	id2377394	1	2016-06-12 00:43:35	2016-06-12 00:54:38	1
2	id3858529	2	2016-01-19 11:35:24	2016-01-19 12:10:48	1
3	id3504673	2	2016-04-06 19:32:31	2016-04-06 19:39:40	1
4	id2181028	2	2016-03-26 13:30:55	2016-03-26 13:38:10	1
5	id0801584	2	2016-01-30 22:01:40	2016-01-30 22:09:03	6
6	id1813257	1	2016-06-17 22:34:59	2016-06-17 22:40:40	4

	id	vendor_id	pickup_datetime	dropoff_datetime	passenger
7	id1324603	2	2016-05-21 07:54:58	2016-05-21 08:20:49	1
8	id1301050	1	2016-05-27 23:12:23	2016-05-27 23:16:38	1
9	id0012891	2	2016-03-10 21:45:01	2016-03-10 22:05:26	1

Report generated by [YData](#).