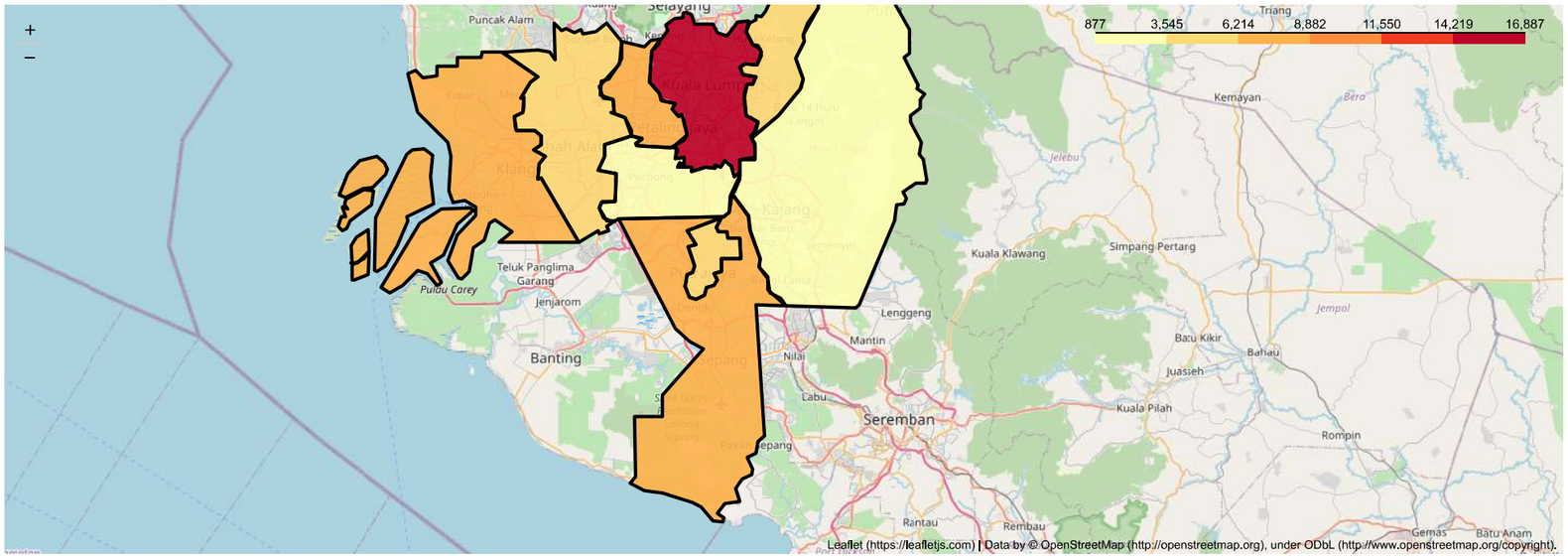
Data Analysis

Dataset

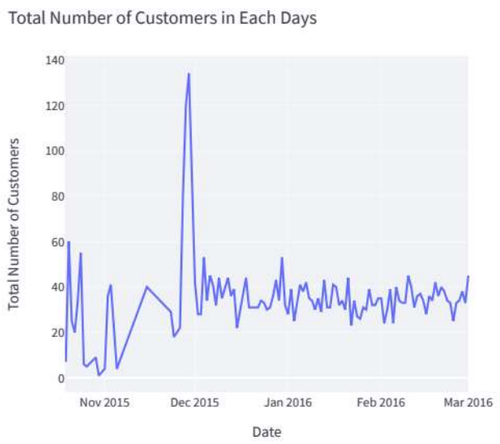
	date	time	race	gender	body_size	age_range	with_kids	kids_category	basket_size	basket_	attire	shirt_colour	shirt_type	pants_colour	pants_type	wash_item	wash	drye	s
0	2015-10-19	20:17:50	malay	male	moderate	28.0000	yes	young	big	red	casual	blue	short_sleeve	black	short	clothes	3	10	n
1	2015-10-19	20:28:42	<NA>	male	thin	32.0000	no	no_kids	big	<NA>	casual	white	short_sleeve	blue_jeans	long	<NA>	6	9	n
2	2015-10-19	20:58:31	malay	female	moderate	30.0000	no	no_kids	big	blue	casual	red	short_sleeve	black	long	<NA>	4	10	n
3	2015-10-19	21:31:28	indian	male	thin	51.0000	no	no_kids	<NA>	black	casual	black	short_sleeve	yellow	short	clothes	5	9	n
4	2015-10-19	21:40:28	indian	male	<NA>	34.0000	no	no_kids	big	blue	casual	blue	<NA>	white	long	clothes	3	9	<
5	2015-10-19	23:45:38	<NA>	male	fat	50.0000	no	no_kids	<NA>	<NA>	casual	brown	<NA>	blue_jeans	long	clothes	4	8	n
6	2015-10-19	23:53:38	<NA>	female	fat	45.0000	no	no_kids	big	pink	traditional	blue	<NA>	black	long	clothes	5	10	n
7	2015-10-20	0:24:01	indian	male	thin	49.0000	<NA>	<NA>	big	blue	casual	white	short_sleeve	brown	long	clothes	3	9	<
8	2015-10-20	0:45:19	indian	male	thin	30.0000	no	no_kids	small	purple	casual	<NA>	short_sleeve	black	short	<NA>	6	10	n
9	2015-10-20	1:17:50	chinese	<NA>	moderate	33.0000	no	no_kids	small	pink	casual	white	short_sleeve	blue_jeans	short	<NA>	5	7	n

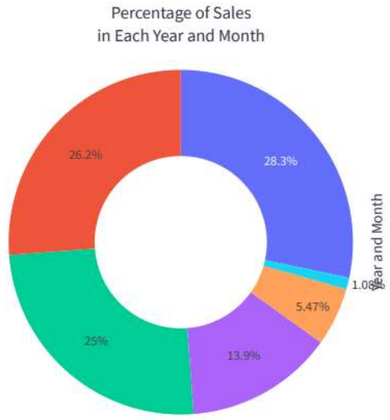
Total Sales in Each Area

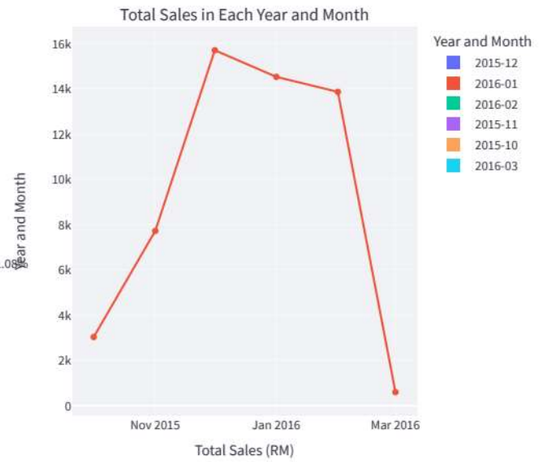


Total Number of Customers in each Days

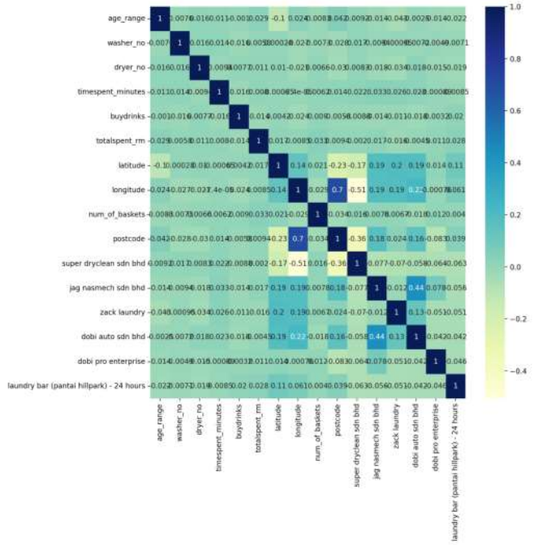
Percentage of Sales in Each Month and Year







Relationships between Variables

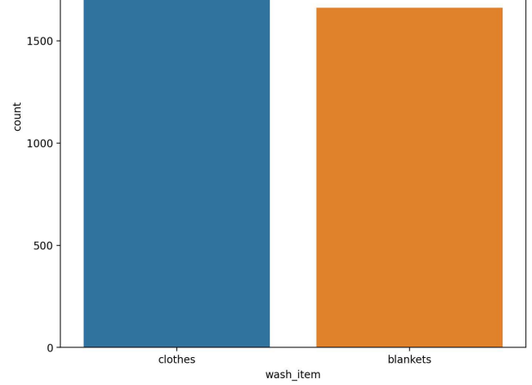


What types of customers will likely to choose Washer No. 3 and Dryer No. 10?

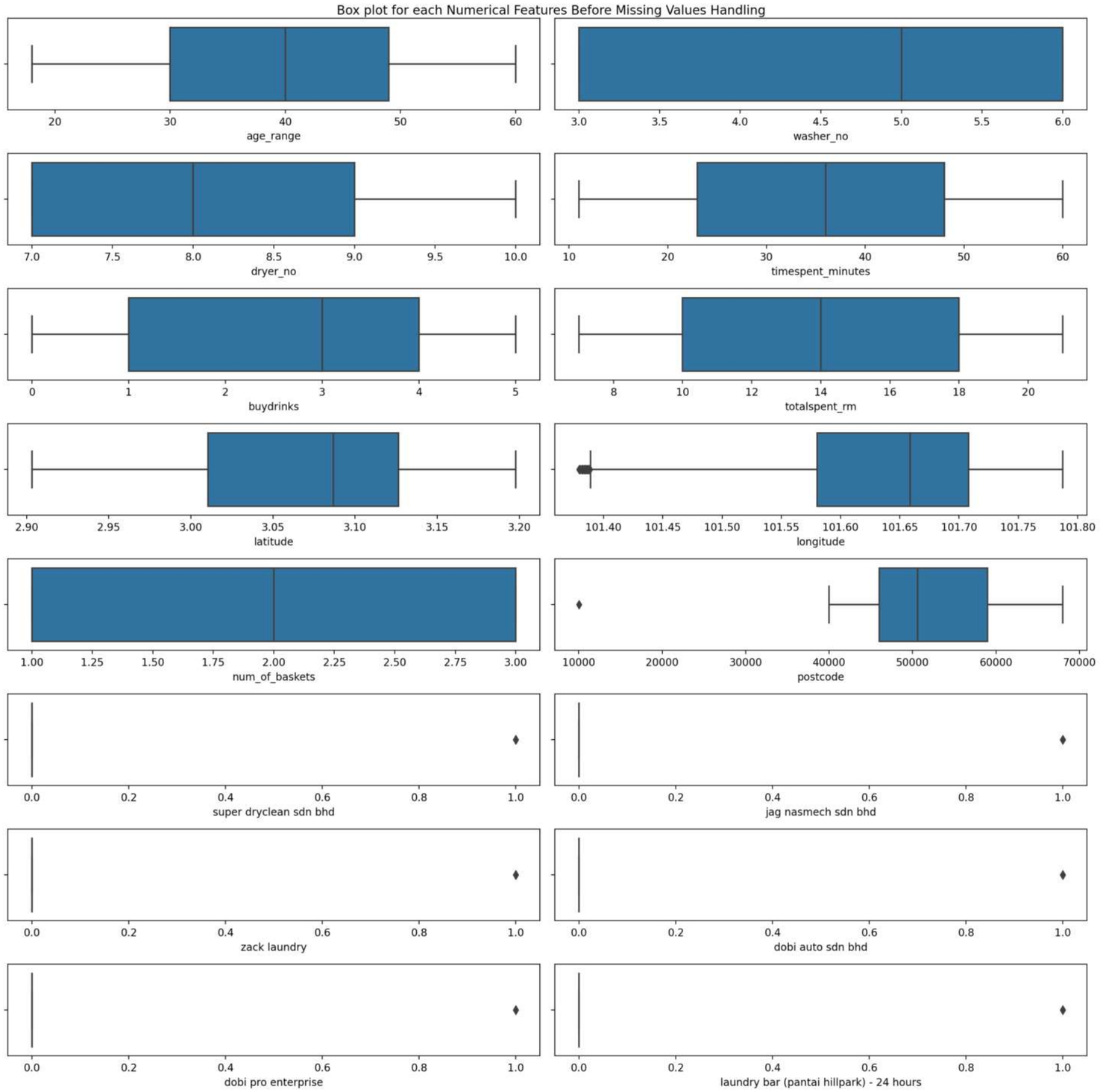
	long	short
baby	23	17
no_kids	37	31
toddler	38	36
young	26	16

Most of customers who wear long pants will likely to choose Washer No.3 and Dryer No.10. From those customers, most of customers who are having toddler will likely to choose Washer No.3 and Dryer No.10.

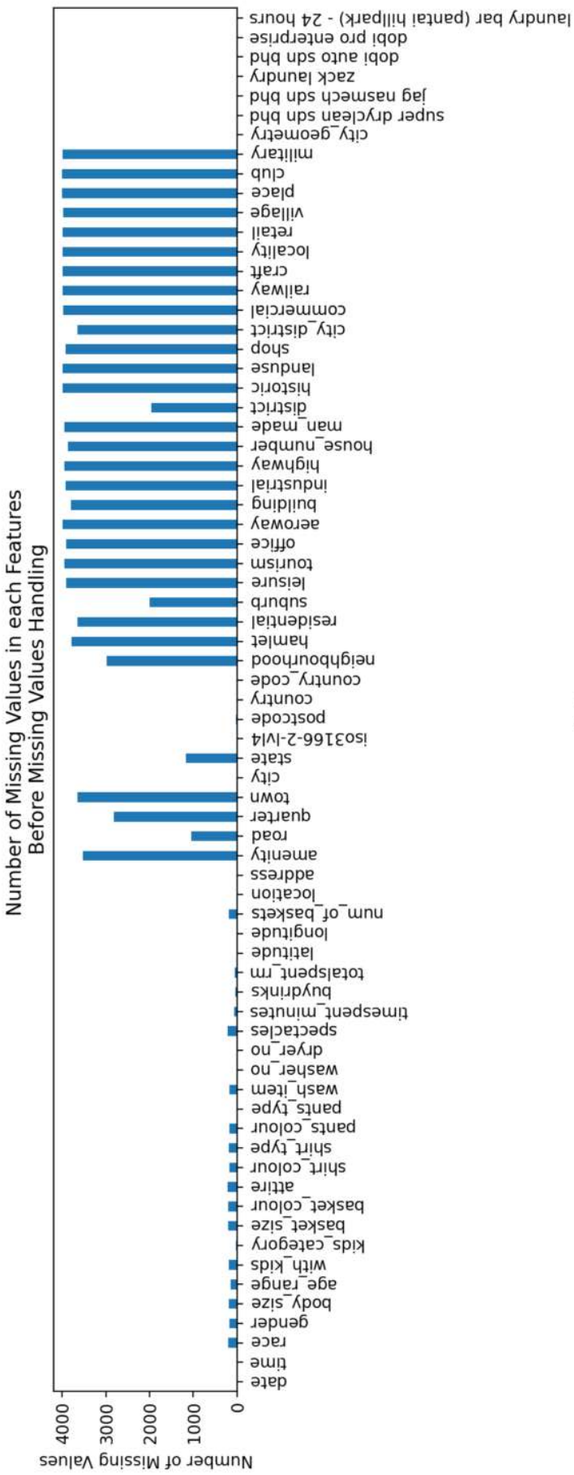
Do we need to perform data imbalance treatment?



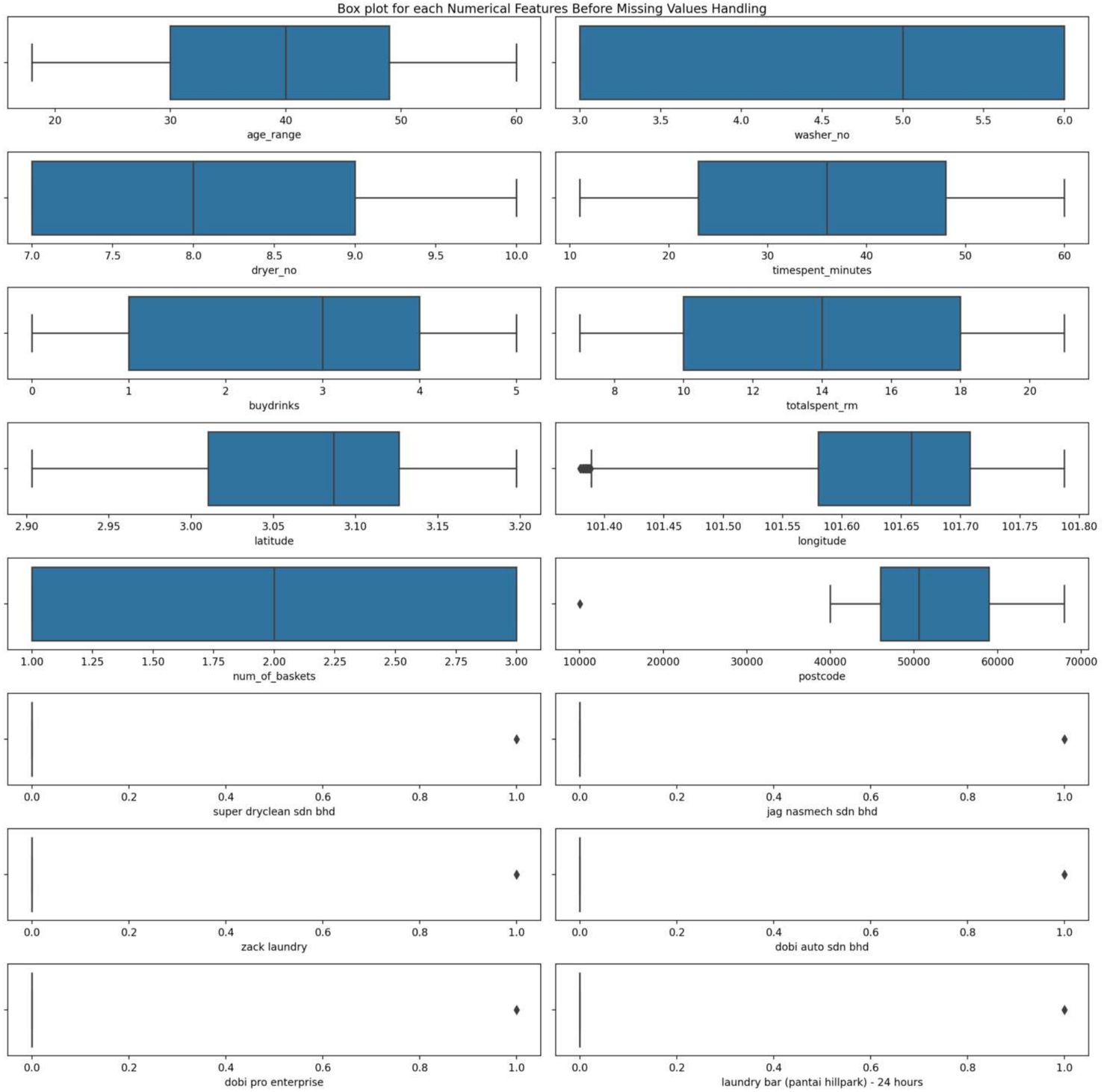
Outliers



Missing Values Handling


Number of Missing Values in each Features
Before Missing Values Handling
Number of Missing Values
Feature
date
time
race
gender
body_size
age_range
with_kids
kids_category
basket_size
basket_colour
attire
shirt_colour
shirt_type
pants_colour
pants_type
wash_item
washer_no
dryer_no
spectacles
timespent_minutes
buydrinks
totalspent_rm
latitude
longitude
num_of_baskets
location
address
amenity
road
quarter
town
city
state
iso3166-2-lvl4
postcode
country
country_code
neighbourhood
hamlet
residential
suburb
leisure
tourism
office
aeroway
building
industrial
highway
house_number
man_made
district
historic
landuse
shop
city_district
commercial
railway
craft
locality
retail
village
place
club
military
city_geometry
super dryclean sdn bhd
jag nasmech sdn bhd
zack laundry
dobi auto sdn bhd
dobi pro enterprise
laundry bar (pantai hillpark) - 24 hours
Number of Missing Values in each Features
After Missing Values Handling
Number of Missing Values
Feature
Number of Missing Values in each Features
Before Missing Values Handling
Number of Missing Values
Feature

Box plot for each Numerical Features Before Missing Values Handling



Box plot for each Numerical Features After Missing Values Handling