Exploratory Data Analysis Report

Exploratory data analysis is an approach to quickly summarize the main characteristics of a dataset, comnbining visual and analytic methods.Through this report you will be able to check the main characteristics of your dataset, helping you to obtain further insights from your data

Dataset Description

Variable	Metrics
Total Observartions	119390
Total Columns	32
Total Nulls	129425

Continous Variables Description

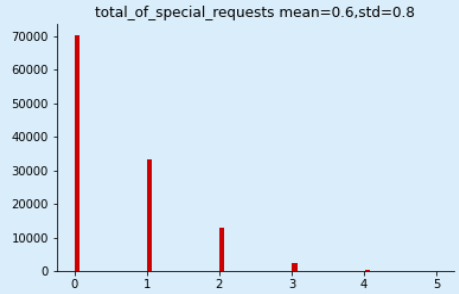
Variable	count	mean	std	min	25%	50%	75%	max
is_canceled	119390	0	0	0	0	0	1	1
lead_time	119390	104	106	0	18	69	160	737
arrival_date	119390	2016	0	2015	2016	2016	2017	2017
arrival_date	119390	27	13	1	16	28	38	53
arrival_date	119390	15	8	1	8	16	23	31
stays_in_wee	119390	0	0	0	0	1	2	19
stays_in_wee	119390	2	1	0	1	2	3	50
adults	119390	1	0	0	2	2	2	55
children	119386	0	0	0	0	0	0	10
babies	119390	0	0	0	0	0	0	10
is_repeated_	119390	0	0	0	0	0	0	1
previous_can	119390	0	0	0	0	0	0	26
previous_boo	119390	0	1	0	0	0	0	72
booking_chan	119390	0	0	0	0	0	0	21
agent	103050	86	110	1	9	14	229	535
company	6797	189	131	6	62	179	270	543
days_in_wait	119390	2	17	0	0	0	0	391
adr	119390	101	50	-6	69	94	126	5400
required_car	119390	0	0	0	0	0	0	8
total_of_spe	119390	0	0	0	0	0	1	5

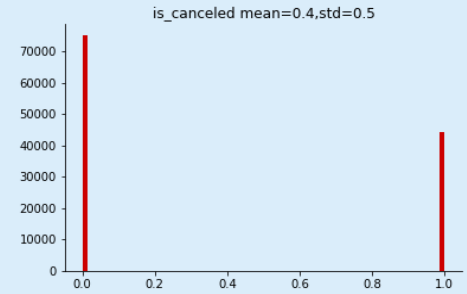
Categorical Variables Description

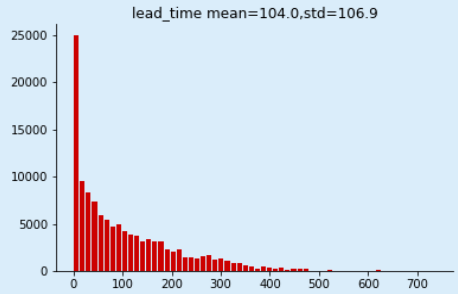
Variable	count	unique	top	freq
hotel	119390	2	City Hotel	79330
arrival_date	119390	12	August	13877
meal	119390	5	BB	92310
country	118902	177	PRT	48590
market_segme	119390	8	Online TA	56477
distribution	119390	5	TA/TO	97870
reserved_roo	119390	10	A	85994
assigned_roo	119390	12	A	74053
deposit_type	119390	3	No Deposit	104641
customer_typ	119390	4	Transient	89613
reservation_	119390	3	Check-Out	75166
reservation_	119390	926	2015-10-21	1461

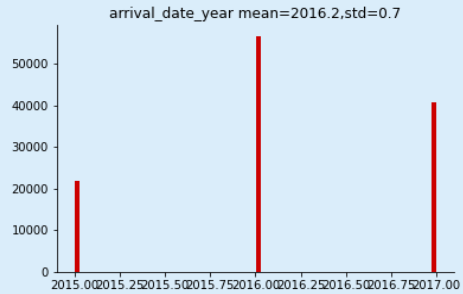
Histogram

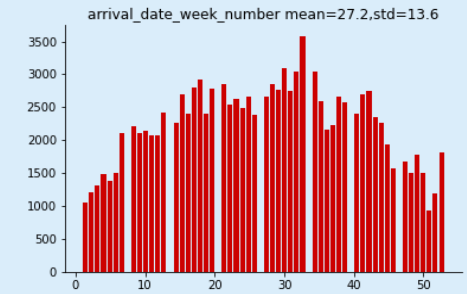
Histograms provide a visual interpretation of numerical data by indicating the number of data points that lie within a range of values. These ranges are called classes or bins. In this case data is represented with 60 bins

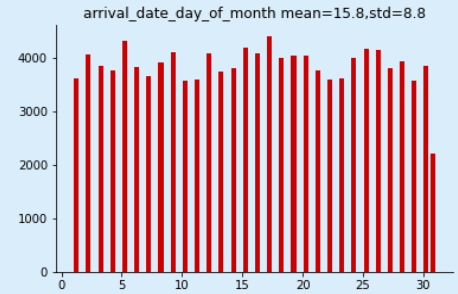


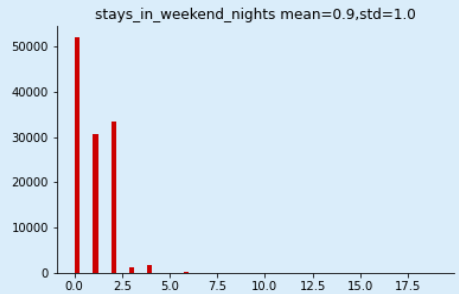


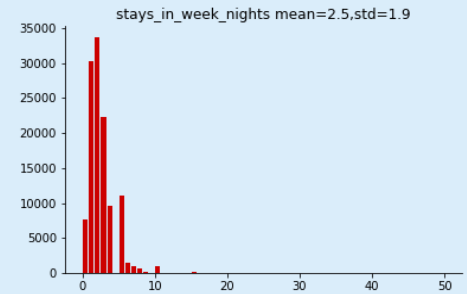


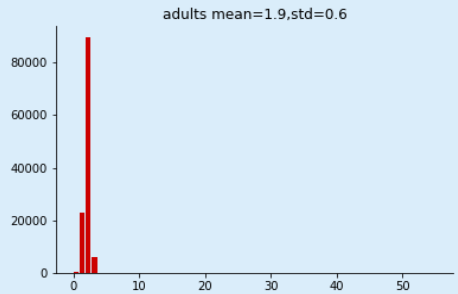


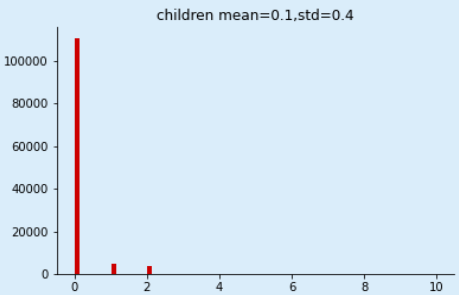


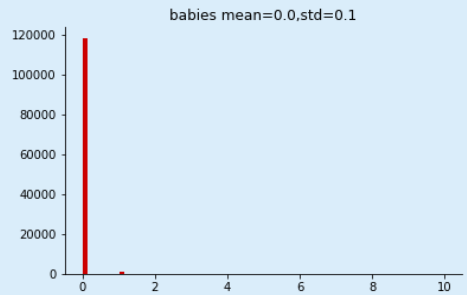


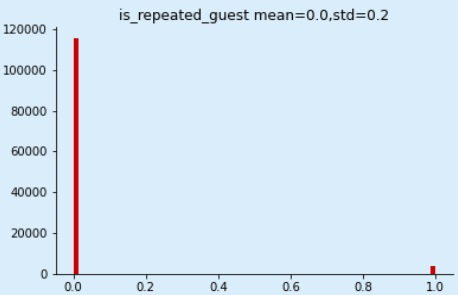


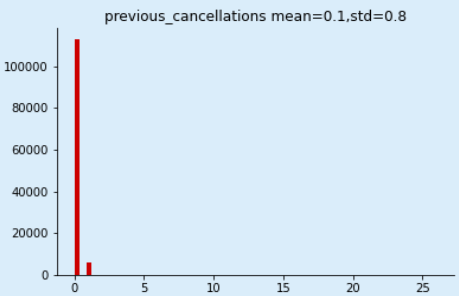


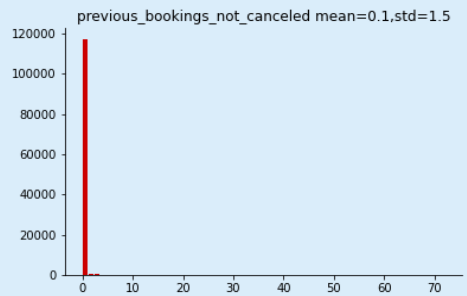


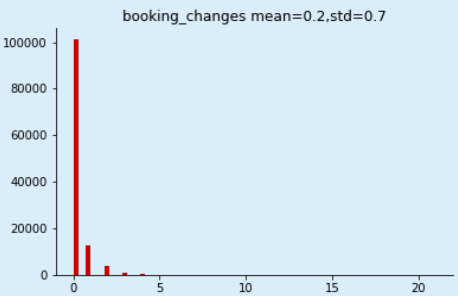


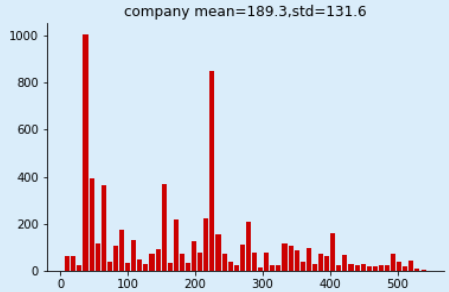


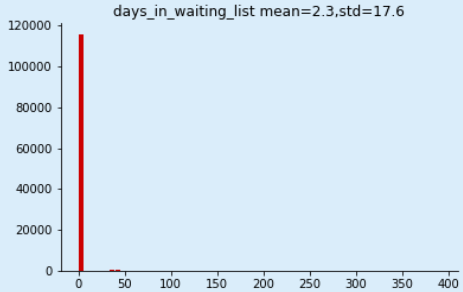


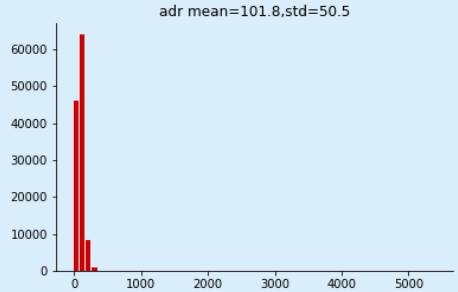


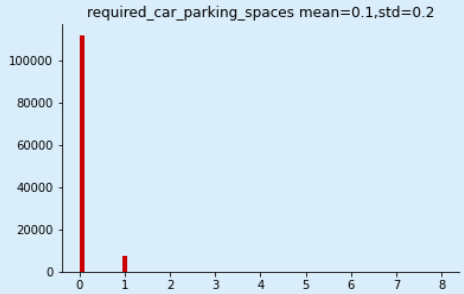





company mean=189.3,std=131.6
1000
800
600
400
200
0
0
100
200
300
400
500

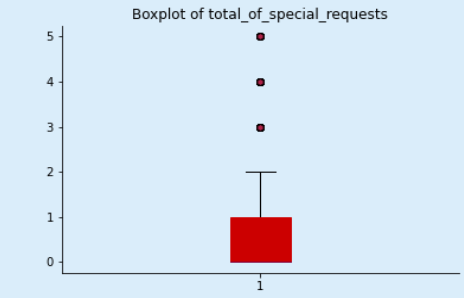

days_in_waiting_list mean=2.3,std=17.6
120000
100000
80000
60000
40000
20000
0
0
50
100
150
200
250
300
350
400

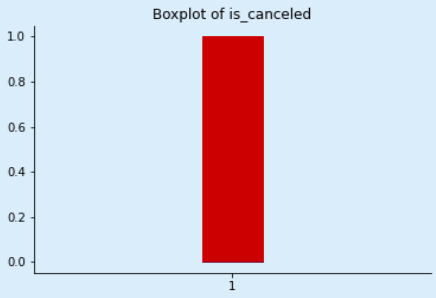

adr mean=101.8,std=50.5
60000
50000
40000
30000
20000
10000
0
0
1000
2000
3000
4000
5000

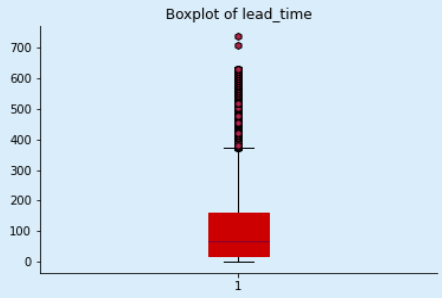

required_car_parking_spaces mean=0.1,std=0.2
100000
80000
60000
40000
20000
0
0
1
2
3
4
5
6
7
8

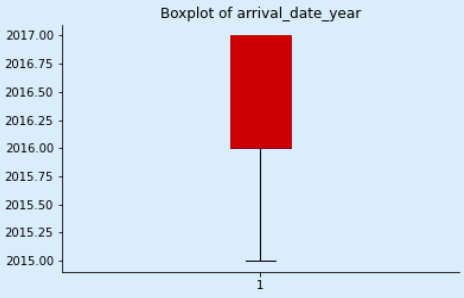
Boxplot

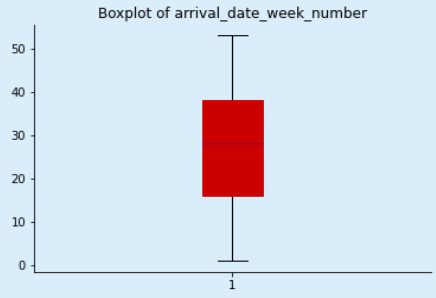
Boxplot is a method for graphically depicting groups of numerical values through their quantiles. Line in orange represents the median, while highlighted values (red) correspo
to outliers

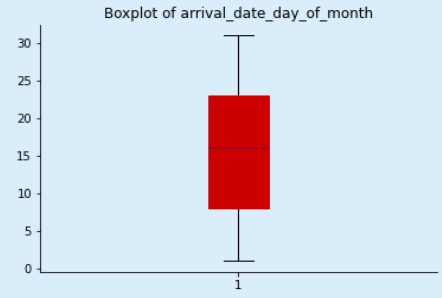


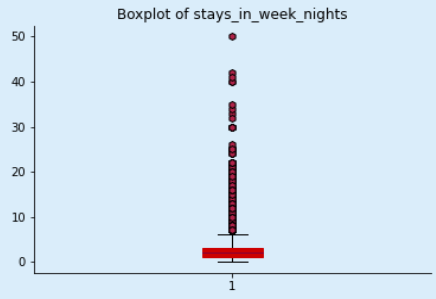


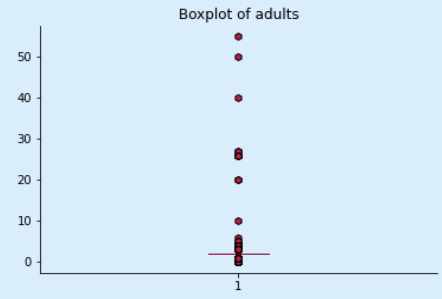


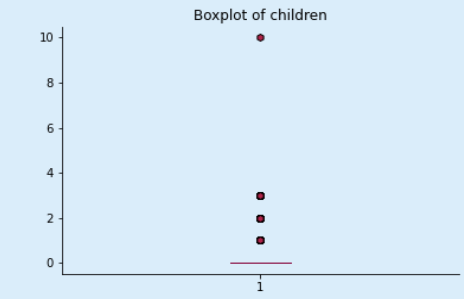


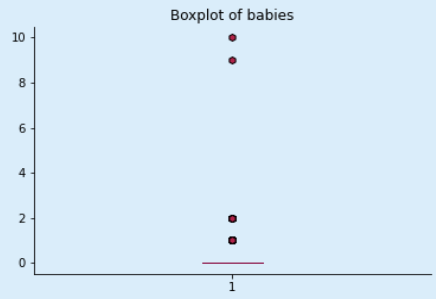


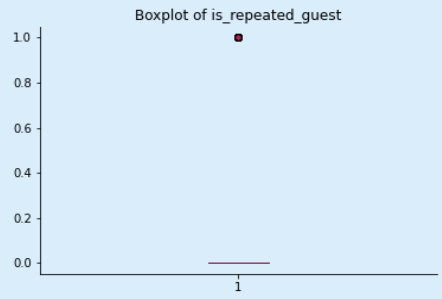


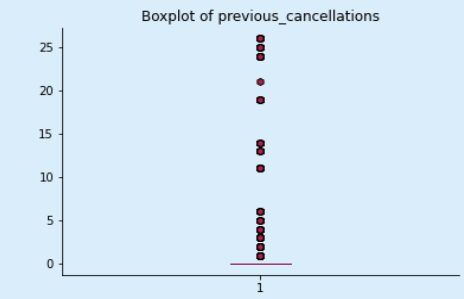


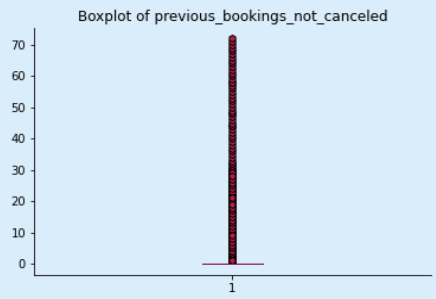


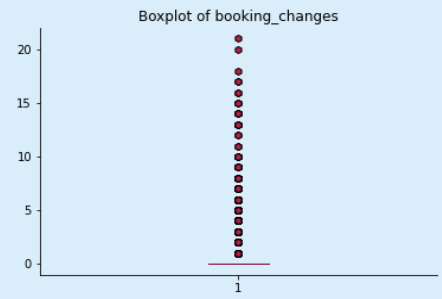


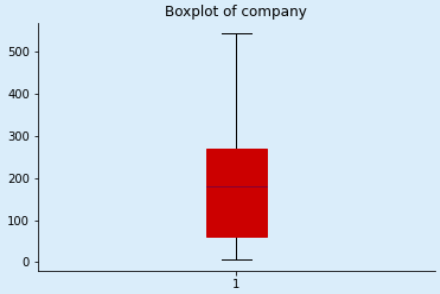


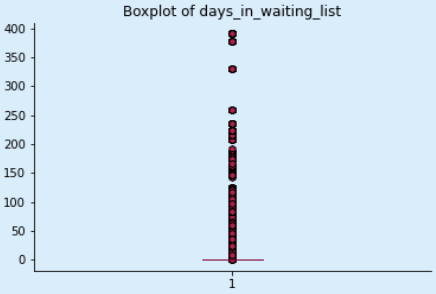


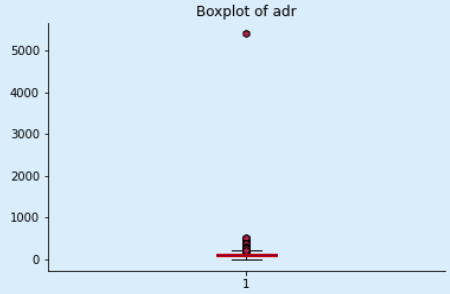


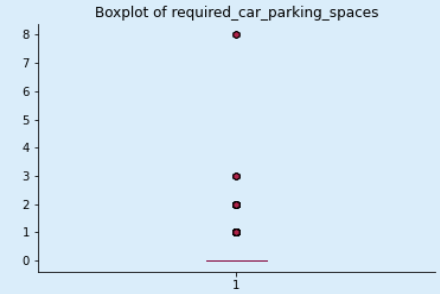


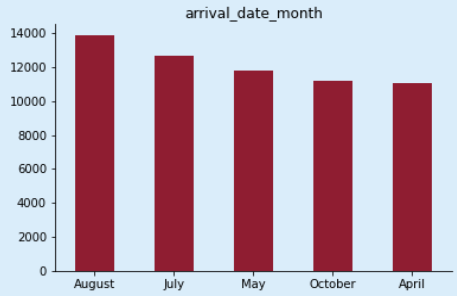
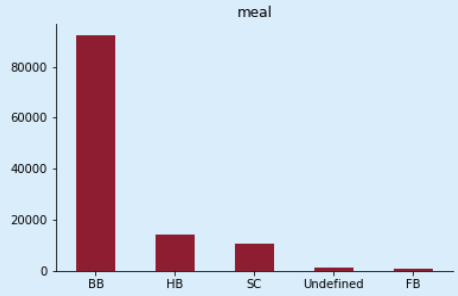
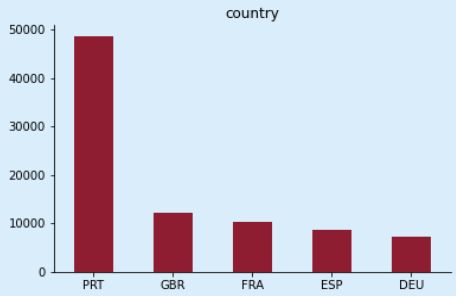
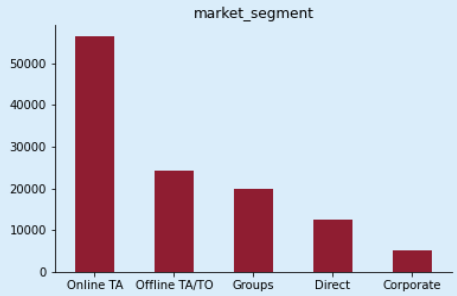
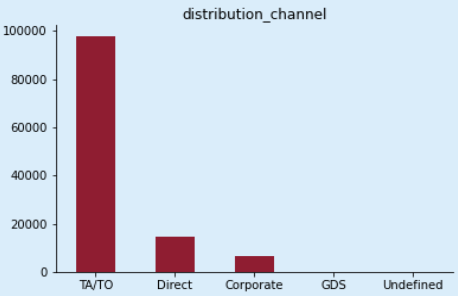
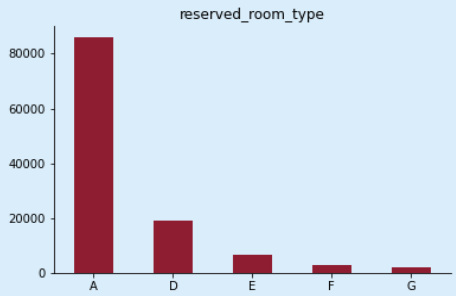
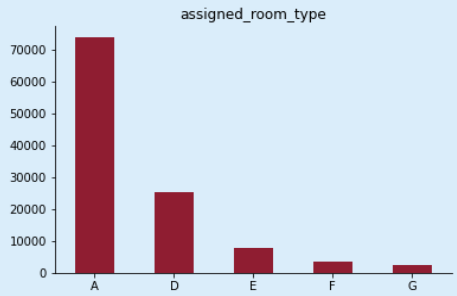
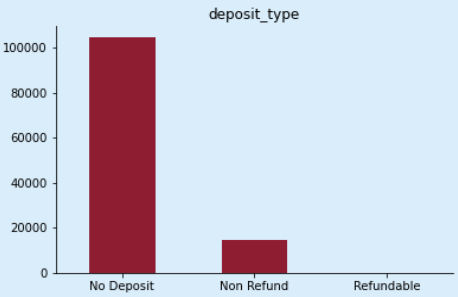
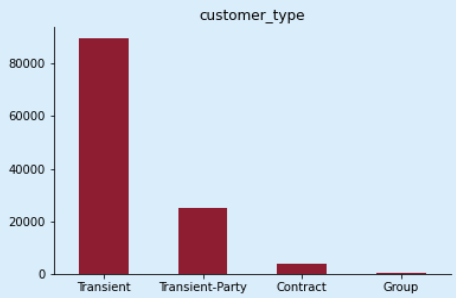
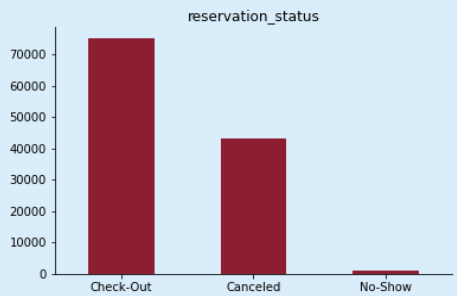
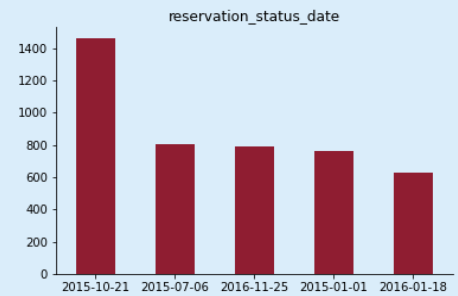



Boxplot of company
500
400
300
200
100
0
1


Boxplot of days_in_waiting_list
400
350
300
250
200
150
100
50
0
1


Boxplot of adr
5000
4000
3000
2000
1000
0
1


Boxplot of required_car_parking_spaces
8
7
6
5
4
3
2
1
0
1

Barchart

Barchar represent categorical data with rectangular bars with heights and lenghts proportional to the values that they represent . Here we only plot the top 5 categories in terms of ocurrencies, to check more details about the number of levels, refer to the categorical variables table