

# CUSTOMER ANALYTICS

CLV vs RFM



By:

Hossein Alibandehloo (40166331004)

Sara Houshmand (40166221014)

Ghazal Mamaghani (40166331010)

Fatemeh Ghasemi (40166331007)

Pegah Dehban (40166331002)

## Question 1

# CLV CALCULATION

	product_id	time_id	store_id	store_sales	store_cost	unit_sales	period_id	the_date
customer_id								
2586	820	935	7	2.10	0.903	3	19	1998-07-23
6685	703	836	9	11.88	5.346	4	16	1998-04-15
2656	1470	879	13	4.82	1.783	2	17	1998-05-28
1256	394	1072	7	12.52	3.756	4	24	1998-12-07
1346	1323	636	15	5.88	2.293	2	9	1997-09-27
9073	926	741	11	2.24	1.008	4	13	1998-01-10
5134	643	681	3	9.20	4.416	5	11	1997-11-11
8963	195	724	17	5.31	1.858	3	12	1997-12-24
2289	1518	1082	7	10.56	4.224	3	24	1998-12-17
3593	57	922	3	1.83	0.586	3	19	1998-07-10

Join “period\_id” from time information sheet to sales data

For the first month of 1997, period\_id is 1 and for the last month of 1998 is 24

## Question 1

# CLV CALCULATION

				R	C		i		
				store_sales	store_cost	unit_sales	period_id	the_date	CLV
customer_id	product_id	time_id	store_id						
573	1513	1095	11	3.80	1.444	2	24	1998-12-30	0.032465
815	812	1095	10	6.68	2.472	2	24	1998-12-30	0.057985
815	1055	1095	10	4.86	2.090	3	24	1998-12-30	0.038170
815	739	1095	10	7.14	2.499	3	24	1998-12-30	0.063952
815	130	1095	10	6.69	3.211	3	24	1998-12-30	0.047940
...	...	...	...	...	...	...	...	...	...
9068	131	367	6	9.60	3.552	4	1	1997-01-01	5.521043
9068	1179	367	6	7.89	3.472	3	1	1997-01-01	4.033064
9068	830	367	6	7.35	2.352	3	1	1997-01-01	4.562529
9068	574	367	6	10.68	4.486	4	1	1997-01-01	5.654323
9068	1217	367	6	2.54	1.219	2	1	1997-01-01	1.205902

$$CLV = \sum_{i=1}^n \frac{(R_i - C_i)}{(1 + d)^{i-0.5}}$$

d = 20% per month = 0.2

Calculate CLV for each transaction

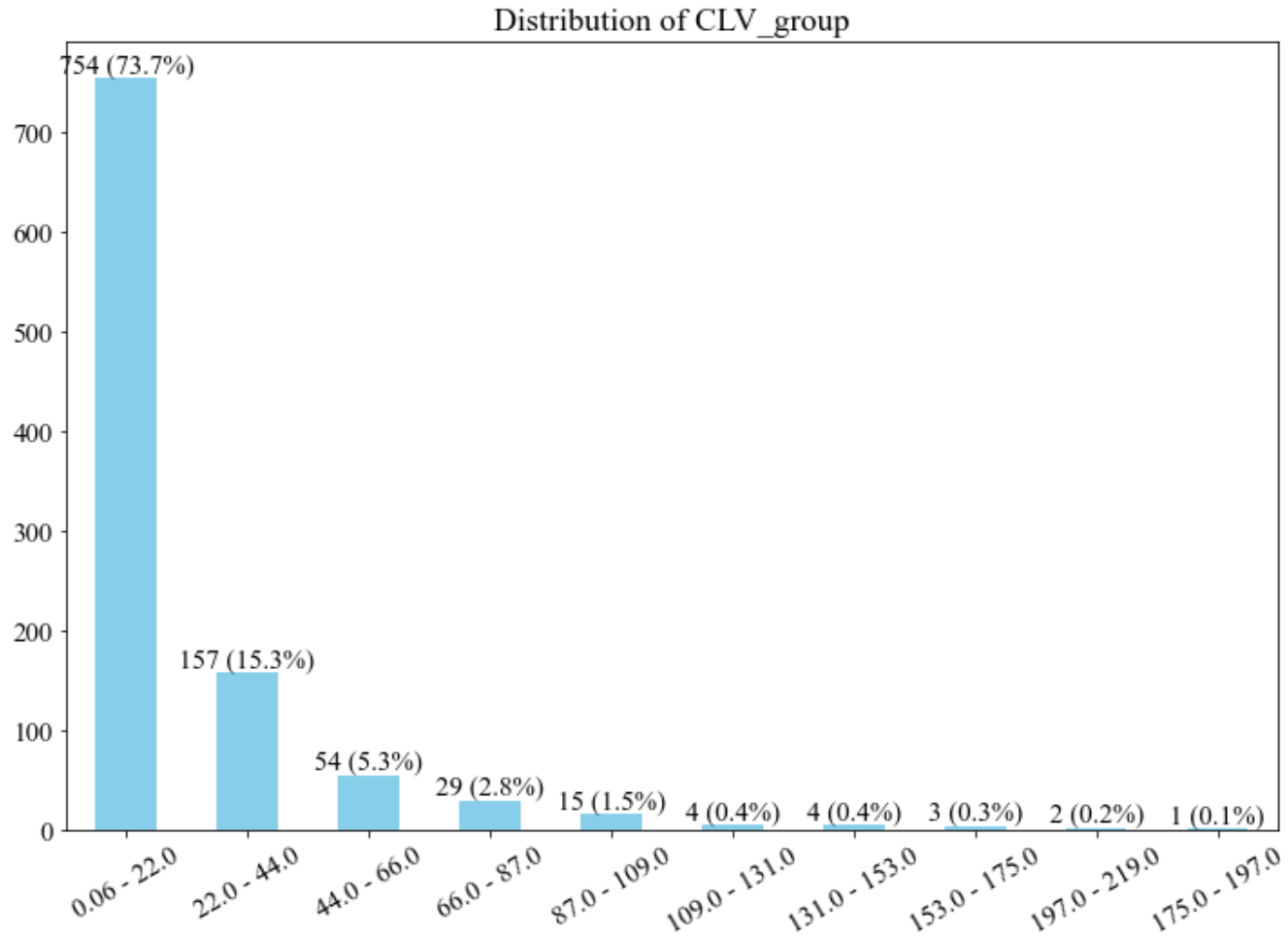
# CLV CALCULATION

	product_id	time_id	store_id	store_sales	store_cost	unit_sales	period_id	CLV
customer_id								
9	7490	12014	247	89.58	34.193	39	244	2.086822
20	25813	30584	864	250.72	101.560	112	593	23.256533
23	8724	6474	216	76.96	32.326	30	111	11.308382
25	2498	6306	114	35.55	13.861	20	138	0.358642
40	5468	3262	98	21.86	8.656	11	28	6.975433

Aggregation of CLV of each transaction based on customer\_id is the final CLV

## Question 1

# CLV DISTRIBUTION



- CLV spans from 0 to 219 for individual customers
- Around 73% of customers exhibit a low CLV.
- Customer Equity (CE) is \$18,612

# PARETO PRINCIPAL

								1	2	3
								CLV	CLV_percentage	cumulative_CLV_percentage
customer_id	product_id	time_id	store_id	store_sales	store_cost	unit_sales	period_id			
1297	182872	184931	4016	1826.31	730.667	891	3190	218.544999	1.174190	1.174190
4868	185706	162325	3552	1389.46	551.305	659	2762	205.264974	1.102840	2.277029
8511	183401	181301	3792	1509.47	597.658	698	3209	194.833901	1.046796	3.323825
2201	148021	136555	3040	1184.92	471.408	558	2299	169.309441	0.909659	4.233484
9631	163672	147811	3152	1371.25	557.852	678	2581	158.269715	0.850345	5.083830
...	...	...	...	...	...	...	...	...	...	...
5572	22366	17627	576	181.60	70.530	94	308	17.279220	0.092837	79.603774
4039	9115	5026	264	54.09	23.668	31	39	17.255268	0.092708	79.696483
2634	56313	67618	76	526.66	211.150	263	1344	17.246303	0.092660	79.789143
4304	24468	22725	435	228.62	86.115	87	414	17.173648	0.092270	79.881413
3918	2478	1784	28	41.33	14.431	15	12	17.052302	0.091618	79.973031

344 rows × 10 columns

**To identify 20% of customers leading 80% of customer equity**

1. Sort CLV in descending order
2. Calculate share of CLV for each customer (CLV\_percentage)
3. Calculate Cumulative percentages (Cumulative\_CLV\_percentage)

## Question 2

# CHECK PARETO PRINCIPAL

- 80% of CE is \$14,900
- 344 of 1023 customers (33%) lead 80% of Customer Equity



### Question 3

# RFM

	product_id	time_id	store_id	store_sales	store_cost	unit_sales	period_id	CLV	CLV_percentage	cumulative_CLV_percentage	CLV_group	Recency
customer_id												
1297	182872	184931	4016	1826.31	730.667	891	3190	218.544999	1.174190	1.174190	197.0 - 219.0	33.0
4868	185706	162325	3552	1389.46	551.305	659	2762	205.264974	1.102840	2.277029	197.0 - 219.0	4.0
8511	183401	181301	3792	1509.47	597.658	698	3209	194.833901	1.046796	3.323825	175.0 - 197.0	21.0
2201	148021	136555	3040	1184.92	471.408	558	2299	169.309441	0.909659	4.233484	153.0 - 175.0	4.0
9631	163672	147811	3152	1371.25	557.852	678	2581	158.269715	0.850345	5.083830	153.0 - 175.0	25.0
...	...	...	...	...	...	...	...	...	...	...	...	...
9196	2612	2799	42	10.55	4.665	4	72	0.081094	0.000436	99.998586	0.06 - 22.0	163.0
5916	2953	4176	8	7.44	3.329	5	92	0.067978	0.000365	99.998952	0.06 - 22.0	52.0
9994	357	1798	28	2.53	0.899	2	36	0.067109	0.000361	99.999312	0.06 - 22.0	197.0
9731	778	2102	28	7.27	3.324	3	46	0.065250	0.000351	99.999663	0.06 - 22.0	45.0
738	568	1055	20	6.66	2.864	2	23	0.062769	0.000337	100.000000	0.06 - 22.0	41.0

1023 rows × 12 columns

df\_sales\_data\_ungrouped.loc[1297]['the\_date'].max()

✓ 0.0s

Timestamp('1998-11-28 00:00:00')

## Recency:

Distance between the time of “1998-12-31” and the last purchase of customer

Check for customer id of “1297” that has recency of 33

## Question 3

# RFM

	product_id	time_id	store_id	store_sales	store_cost	unit_sales	period_id	CLV	CLV_percentage	cumulative_CLV_percentage	CLV_group	Recency
customer_id												
1297	182872	184931	4016	1826.31	730.667	891	3190	218.544999	1.174190	1.174190	197.0 - 219.0	0.030303
4868	185706	162325	3552	1389.46	551.305	659	2762	205.264974	1.102840	2.277029	197.0 - 219.0	0.250000
8511	183401	181301	3792	1509.47	597.658	698	3209	194.833901	1.046796	3.323825	175.0 - 197.0	0.047619
2201	148021	136555	3040	1184.92	471.408	558	2299	169.309441	0.909659	4.233484	153.0 - 175.0	0.250000
9631	163672	147811	3152	1371.25	557.852	678	2581	158.269715	0.850345	5.083830	153.0 - 175.0	0.040000
...	...	...	...	...	...	...	...	...	...	...	...	...
9196	2612	2799	42	10.55	4.665	4	72	0.081094	0.000436	99.998586	0.06 - 22.0	0.006135
5916	2953	4176	8	7.44	3.329	5	92	0.067978	0.000365	99.998952	0.06 - 22.0	0.019231
9994	357	1798	28	2.53	0.899	2	36	0.067109	0.000361	99.999312	0.06 - 22.0	0.005076
9731	778	2102	28	7.27	3.324	3	46	0.065250	0.000351	99.999663	0.06 - 22.0	0.022222
738	568	1055	20	6.66	2.864	2	23	0.062769	0.000337	100.000000	0.06 - 22.0	0.024390

1023 rows × 12 columns

## Recency:

The less recency, the better customer. Thus, we need to calculate  $1/R$

## Question 3

## RFM

	product_id	time_id	store_id	store_sales	store_cost	unit_sales	period_id	CLV	CLV_percentage	cumulative_CLV_percentage	CLV_group	Recency	Frequency	Monetary
customer_id														
1297	182872	184931	4016	1826.31	730.667	891	3190	218.544999	1.174190	1.174190	197.0 - 219.0	0.030303	251	1826.31
4868	185706	162325	3552	1389.46	551.305	659	2762	205.264974	1.102840	2.277029	197.0 - 219.0	0.250000	222	1389.46
8511	183401	181301	3792	1509.47	597.658	698	3209	194.833901	1.046796	3.323825	175.0 - 197.0	0.047619	237	1509.47
2201	148021	136555	3040	1184.92	471.408	558	2299	169.309441	0.909659	4.233484	153.0 - 175.0	0.250000	190	1184.92
9631	163672	147811	3152	1371.25	557.852	678	2581	158.269715	0.850345	5.083830	153.0 - 175.0	0.040000	197	1371.25
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
9196	2612	2799	42	10.55	4.665	4	72	0.081094	0.000436	99.998586	0.06 - 22.0	0.006135	3	10.55
5916	2953	4176	8	7.44	3.329	5	92	0.067978	0.000365	99.998952	0.06 - 22.0	0.019231	4	7.44
9994	357	1798	28	2.53	0.899	2	36	0.067109	0.000361	99.999312	0.06 - 22.0	0.005076	2	2.53
9731	778	2102	28	7.27	3.324	3	46	0.065250	0.000351	99.999663	0.06 - 22.0	0.022222	2	7.27
738	568	1055	20	6.66	2.864	2	23	0.062769	0.000337	100.000000	0.06 - 22.0	0.024390	1	6.66

1023 rows × 14 columns

**Monetary:**

Total amount of money paid by each customer

### Question 3

## RFM

Recency	Frequency	Monetary	Recency_status	Frequency_status	Monetary_status	cluster
0.050000	15	105.35	L	L	L	LLL
0.047619	88	745.97	L	H	H	LHH
0.071429	49	138.89	H	H	L	HHL
0.005263	11	39.05	L	L	L	LLL
0.002088	4	14.94	L	L	L	LLL
0.014493	8	43.31	L	L	L	LLL
0.100000	8	51.26	H	L	L	HLL
0.062500	2	12.21	H	L	L	HLL
0.002421	19	143.14	L	L	L	LLL
0.020000	28	194.09	L	L	L	LLL

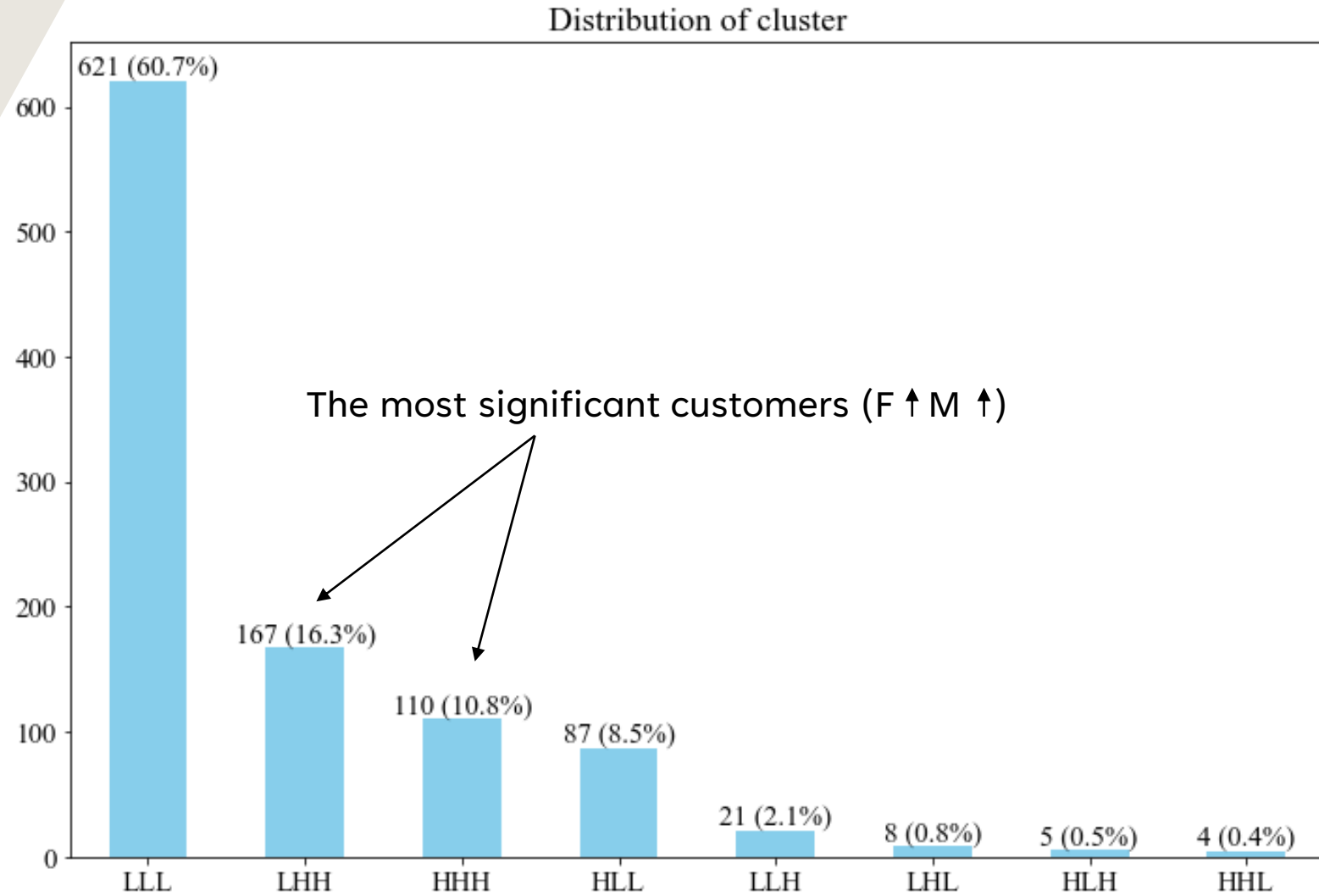
For each R, F and M if the value is less than its average, it will be labeled "L". Otherwise, the label will be "H"

Attribute	Mean
Recency	0.053
Frequency	32.2
Monetary	211.8

### Question 3

## RFM MODEL

Cluster	Count
LLL	621
LHH	167
HHH	110
HLL	87
LLH	21
LHL	8
HLH	5
HHL	4

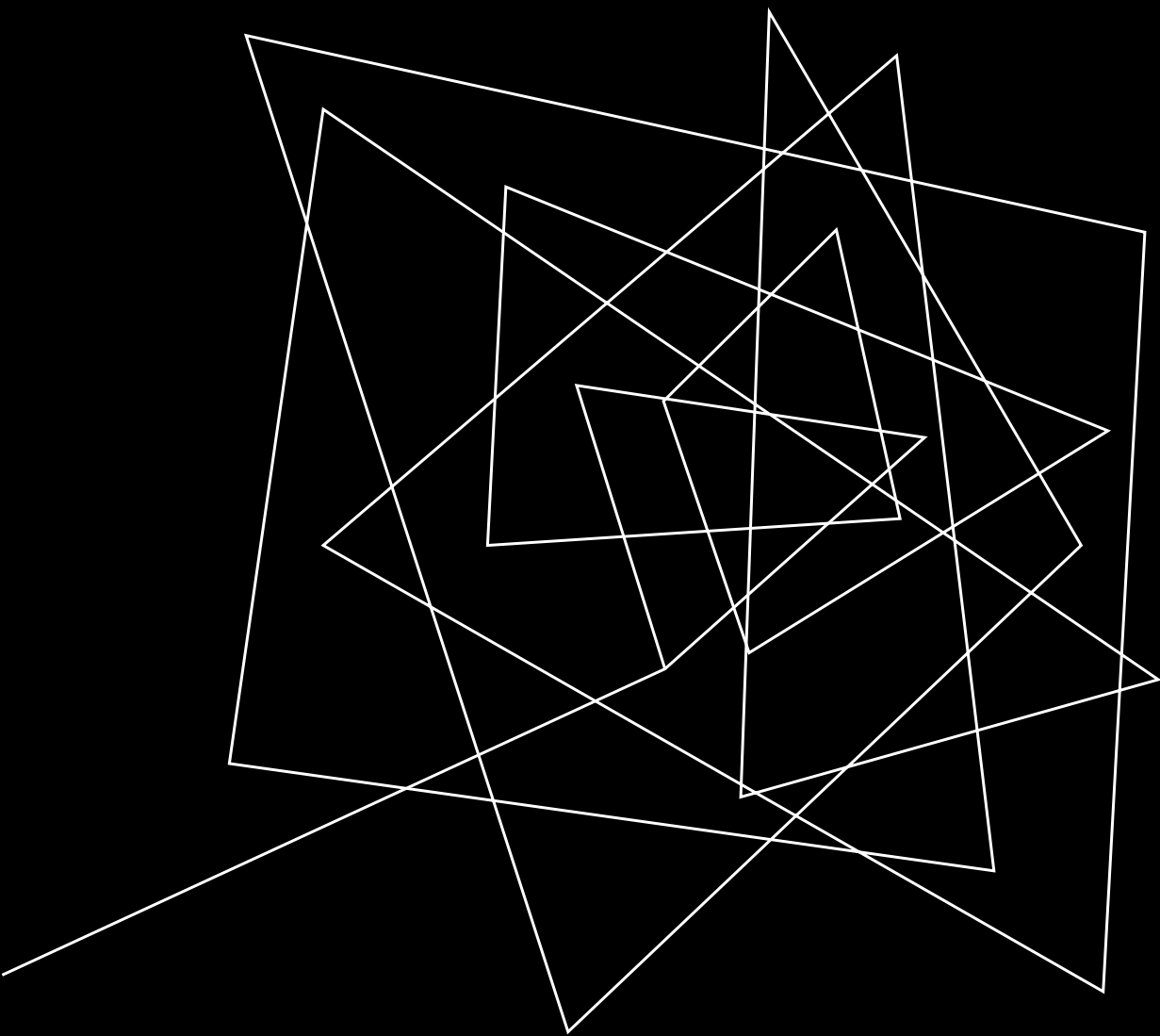


### Question 3

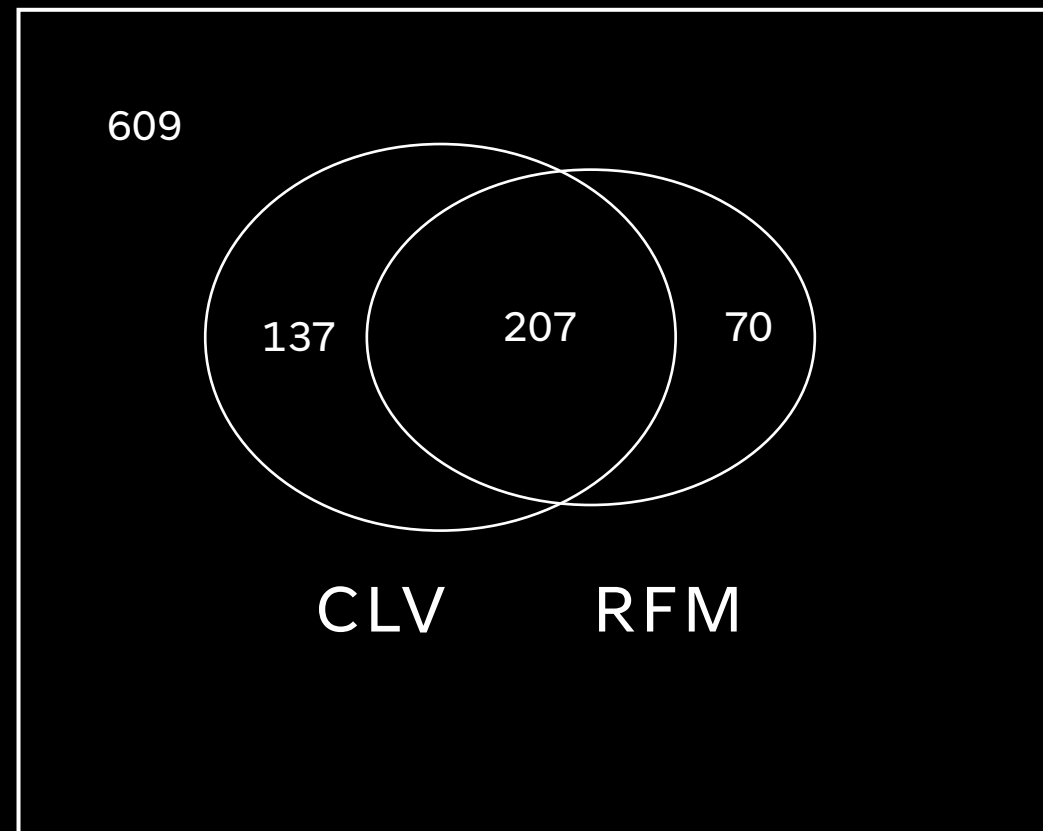
FOCUS ON

HHH & LHH

- 277 Customers (27% of Total)
- Total CLV: 11832 (63% of Total)
- 207 intersections with customers of top 80% of CLV
- with top 80% of CLV



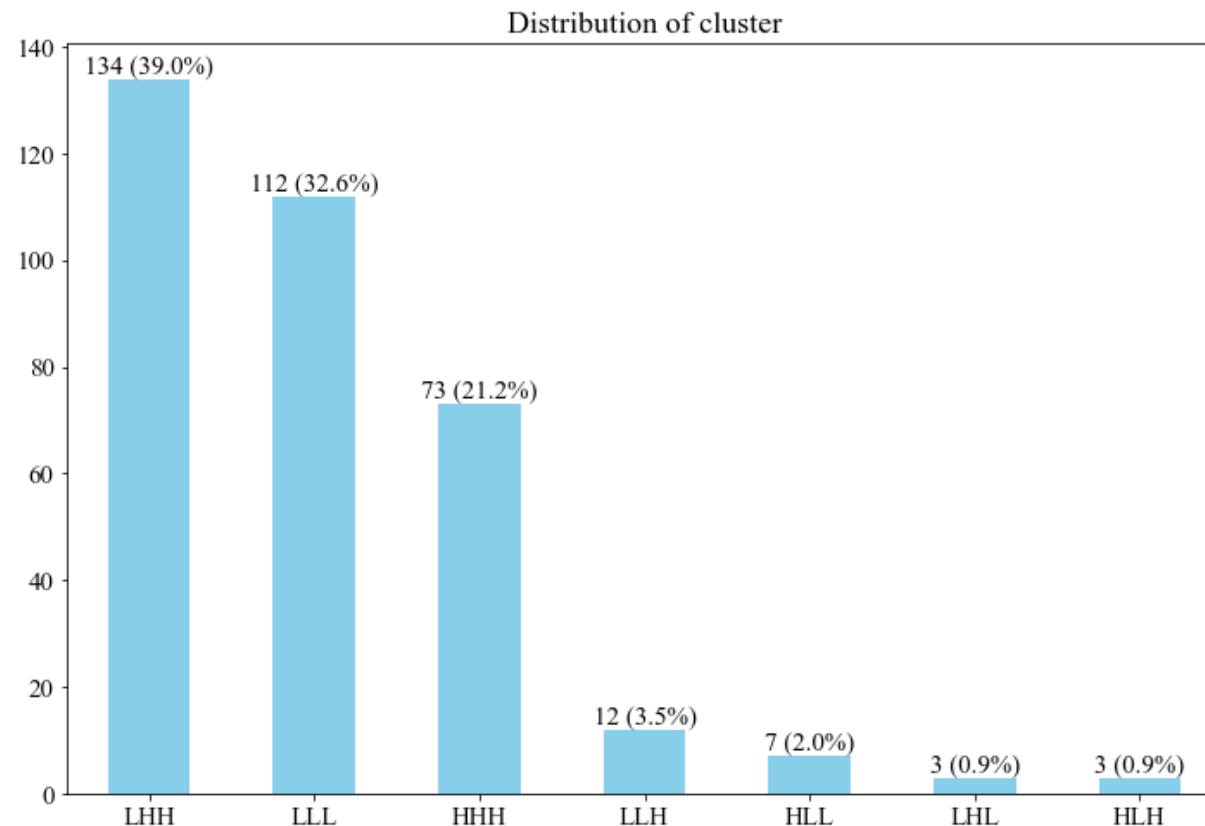
## ALL CUSTOMERS



### Question 3

## TOP 80% OF CUSTOMERS BASED ON CLV

Attribute	Average
R	0.015
F	20
M	141



- There are 112 customers of “LLL” which reside in top 80% of CLV!
- 60% of top 80% of customers are of “HHH” and “LHH”
- Maybe it’s better to label clusters based on median



# CALCULATE CLV BASE ON RFM

Recency_status	Frequency_status	Monetary_status	cluster	Recency_scaled	Frequency_scaled	Monetary_scaled	CLV_RFM
H	L	L	HLL	0.070094	0.070094	0.070094	97.120714
L	L	H	LLH	0.007505	0.007505	0.007505	117.231339
L	L	L	LLL	0.016771	0.016771	0.016771	62.942727
L	L	L	LLL	0.002771	0.002771	0.002771	66.830630
L	L	L	LLL	0.002634	0.002634	0.002634	15.200610
L	L	L	LLL	0.000562	0.000562	0.000562	25.055299
H	H	H	HHH	0.248922	0.248922	0.248922	145.387500

1. Scale R, F and M to 0 to 1
2. Calculate CLV\_RFM based on the formula below:
3.  $CLV\_RFM = 0.15 \times R + 0.35 \times F + 0.5 \times M$

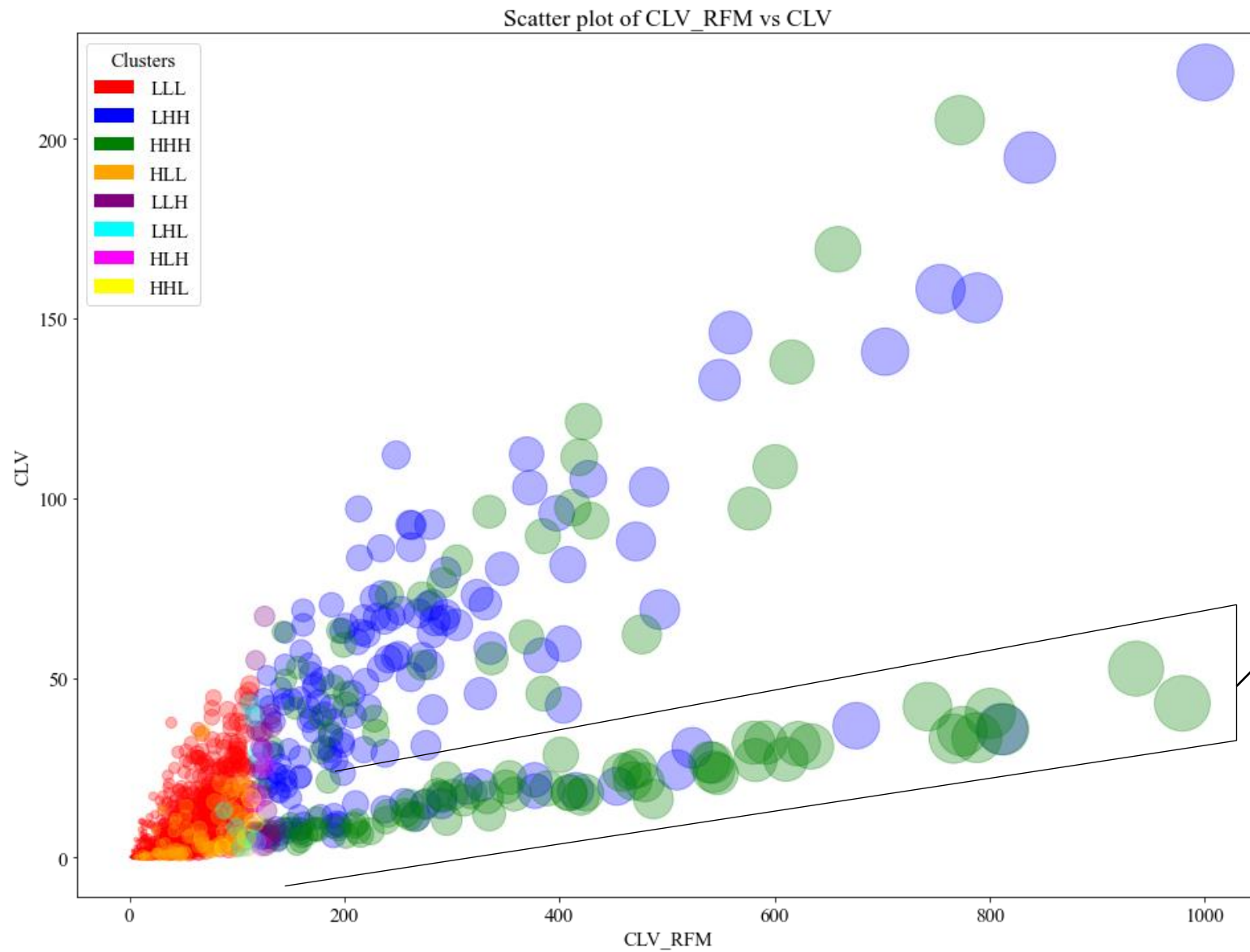


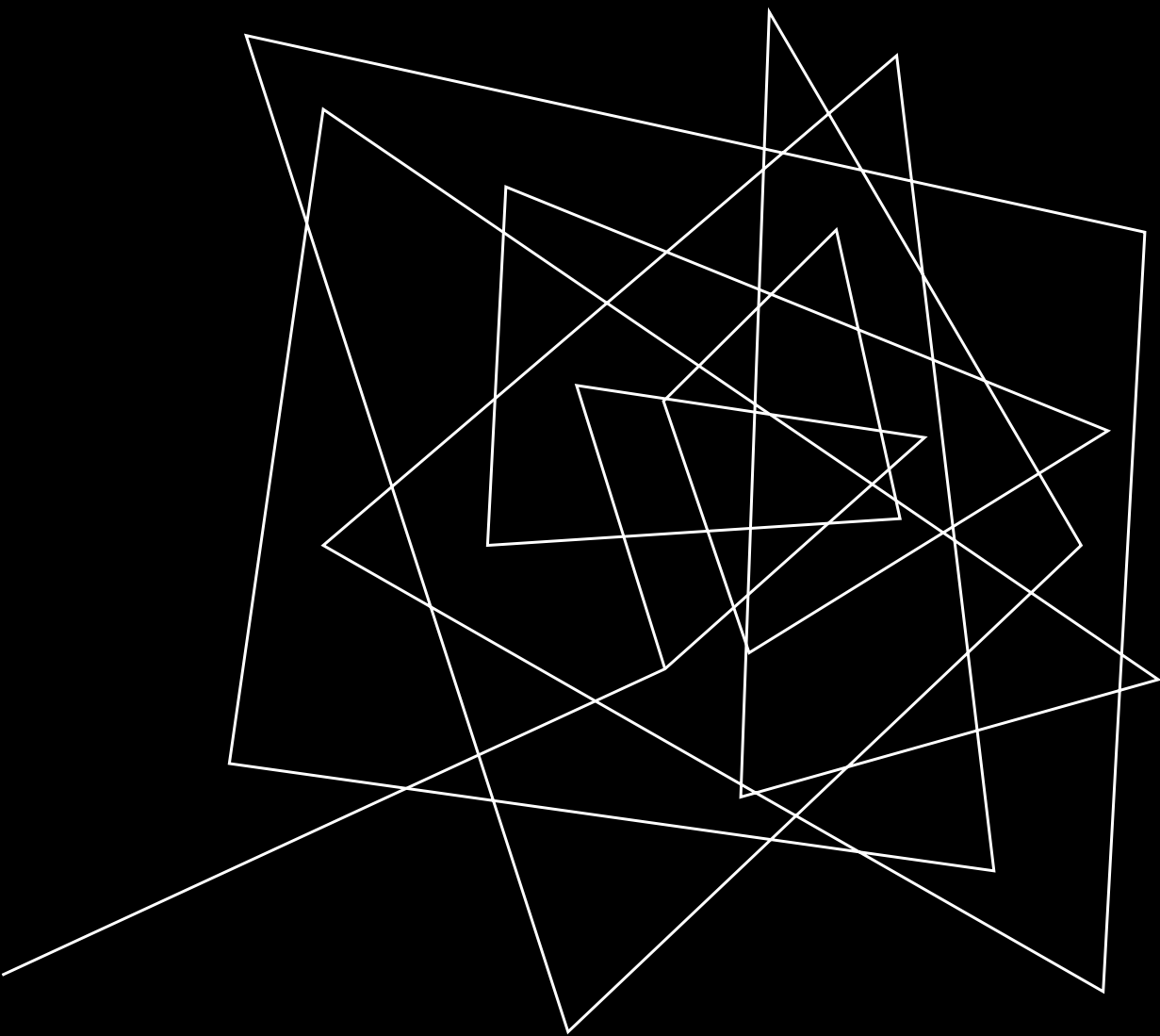
Question 4

**CLV VS RFM**

Correlation = 69%

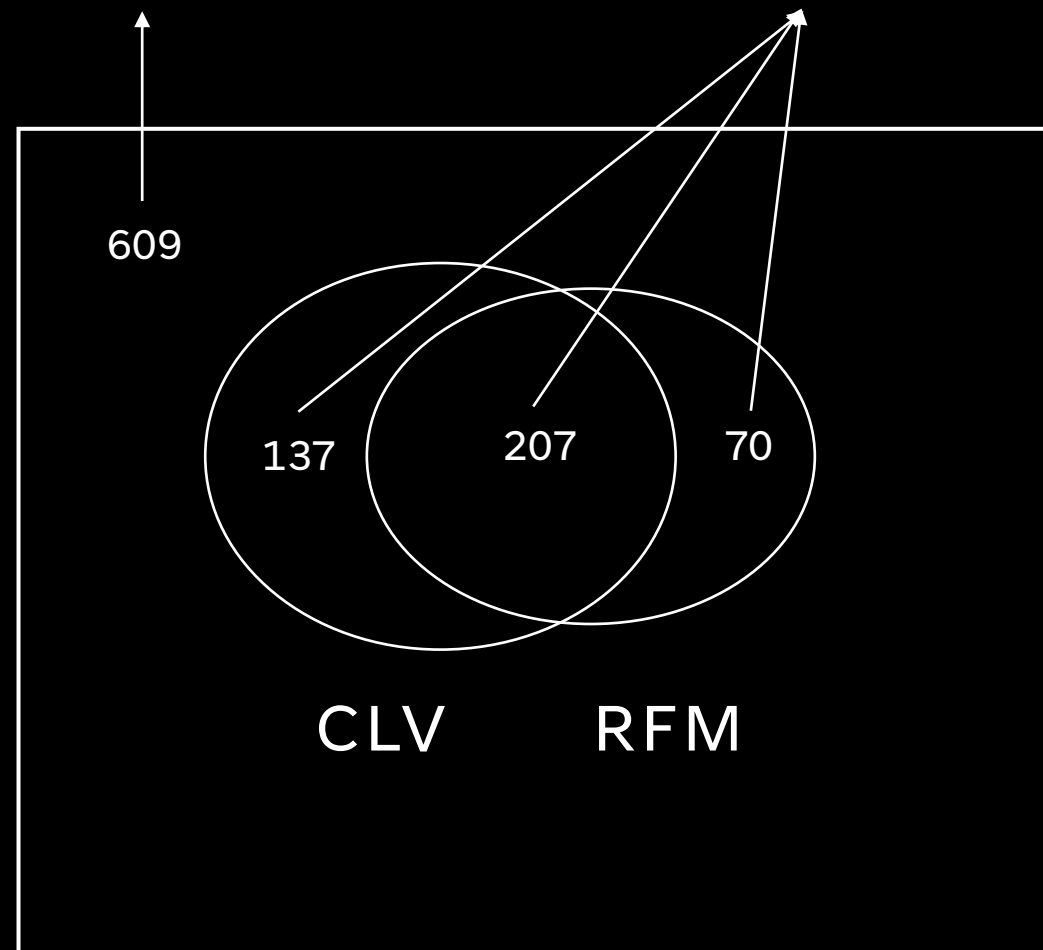
# RFM VS CLV





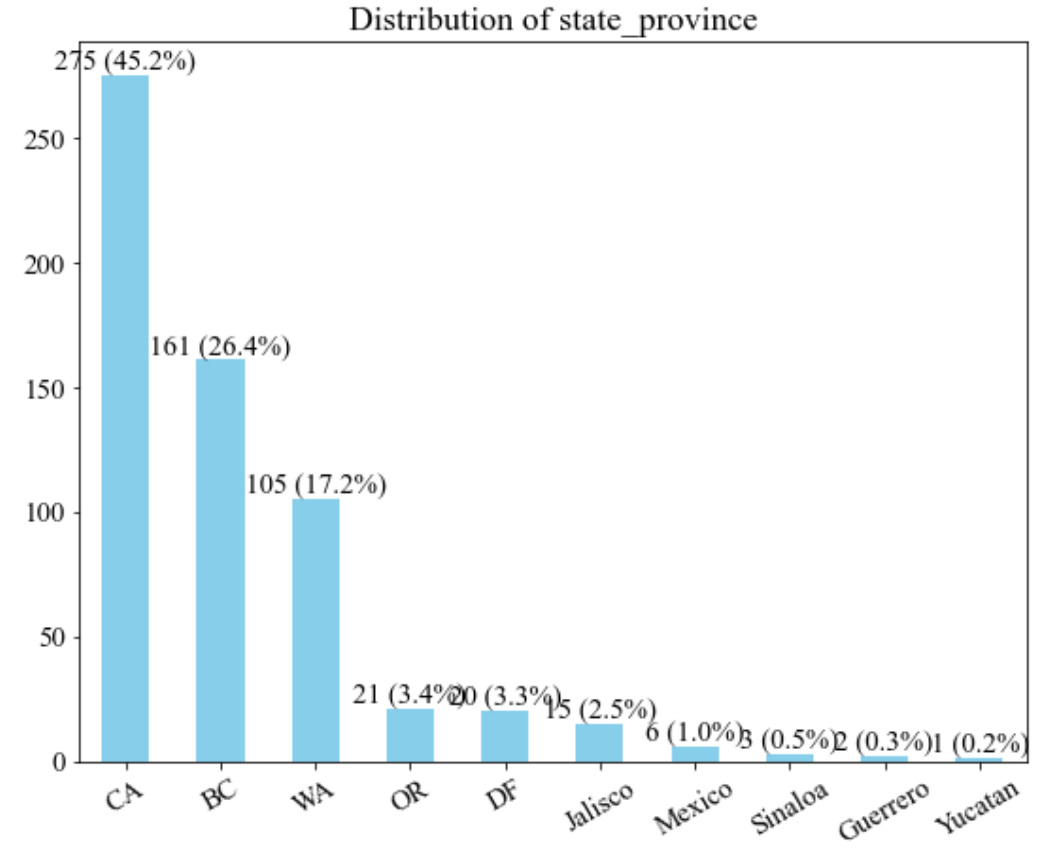
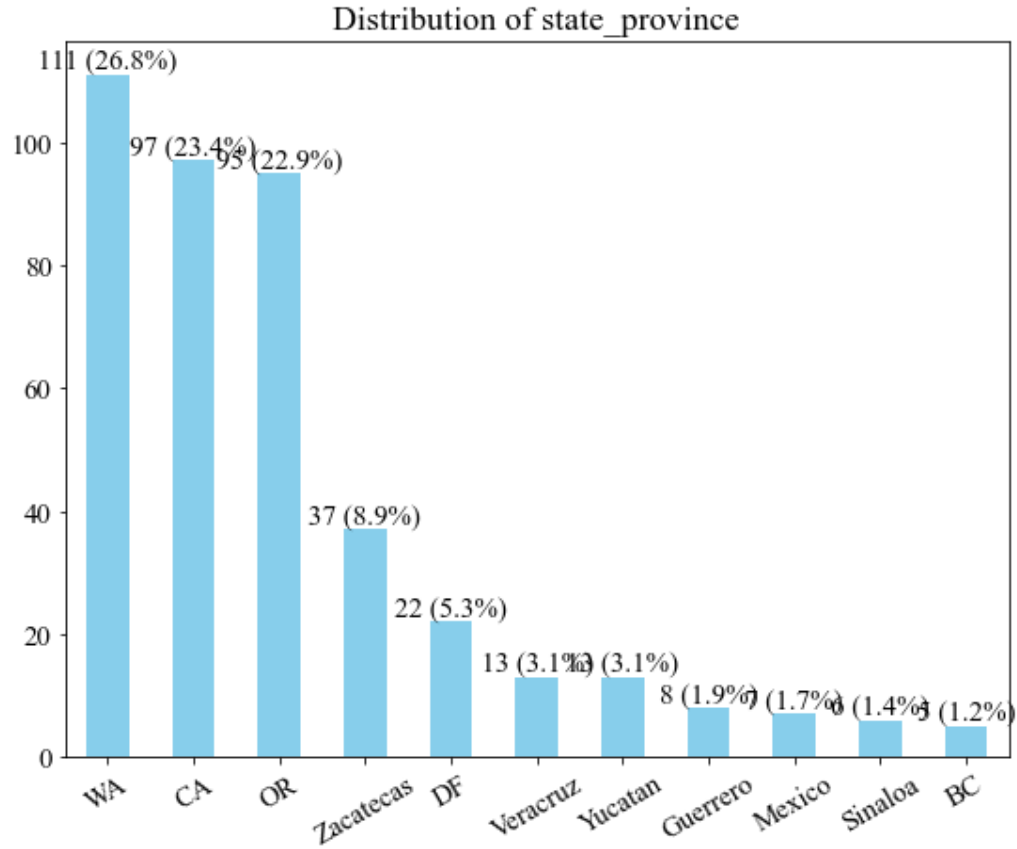
Low value customers

High value customers



## HIGH VALUE CUSTOMER

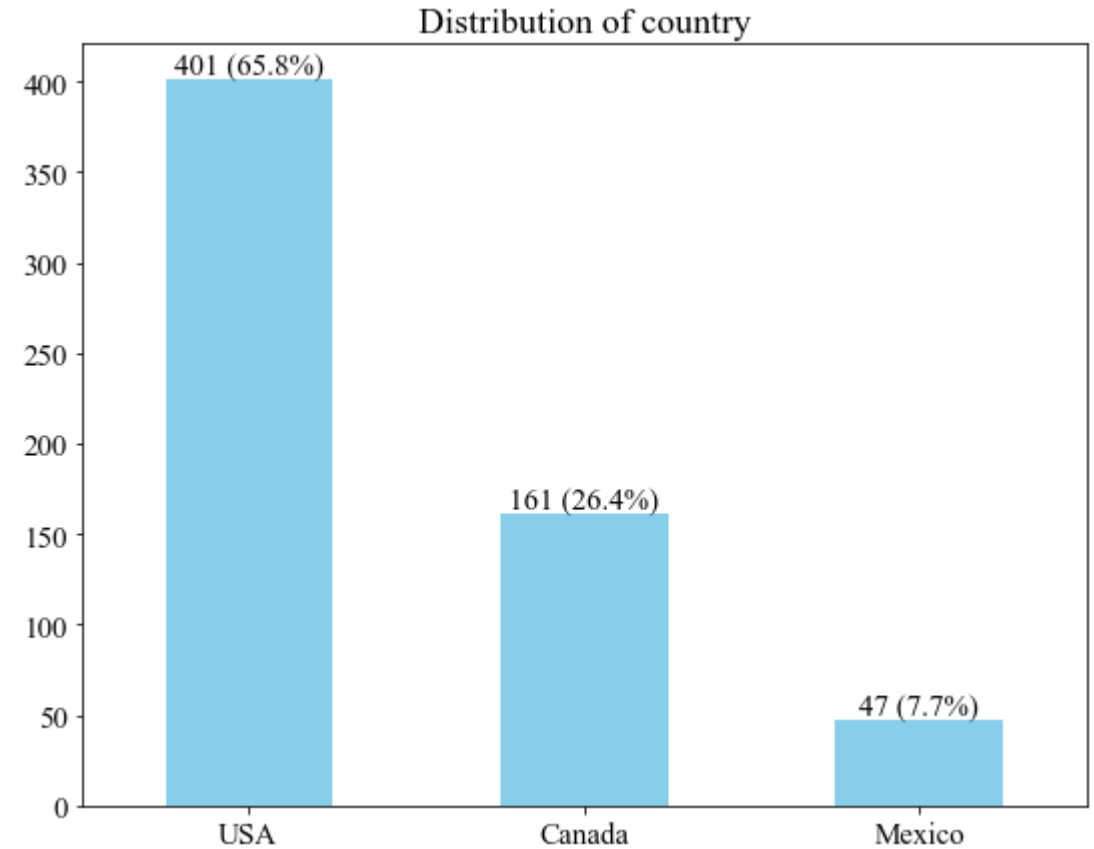
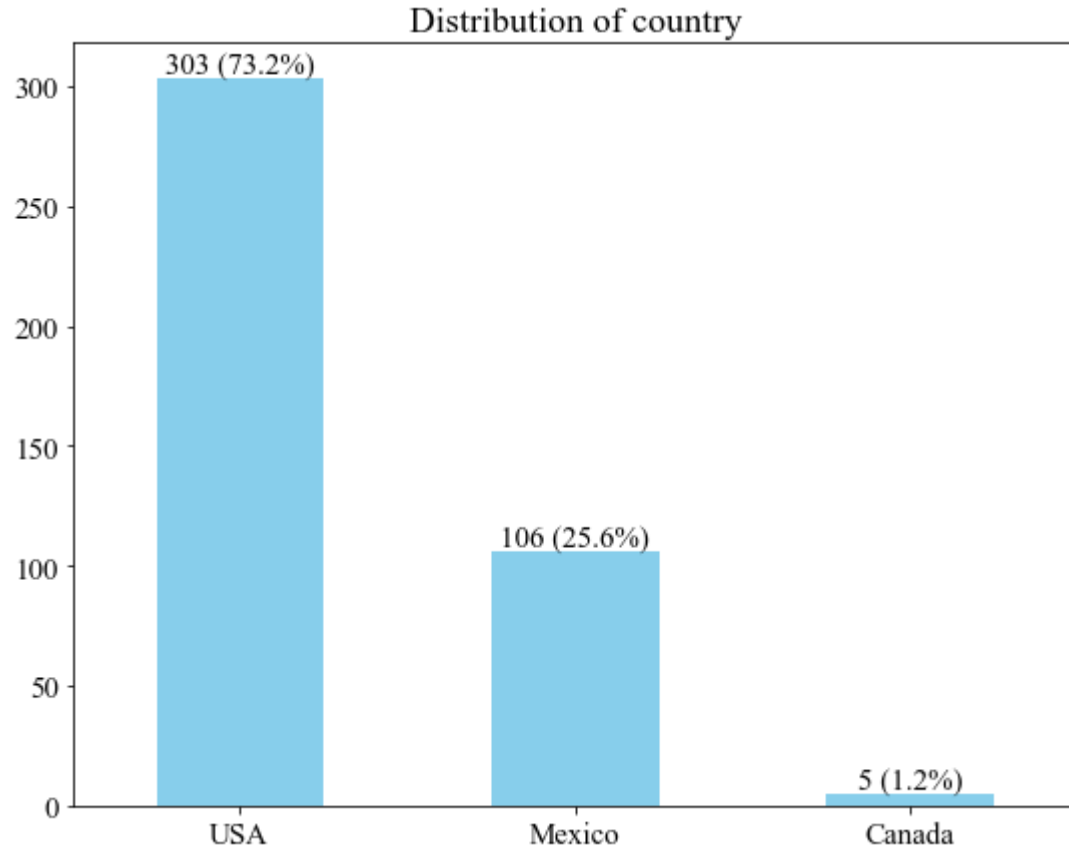
## LOW VALUE CUSTOMER



Focus on Oregon and Zacatecas rather than British Colombia

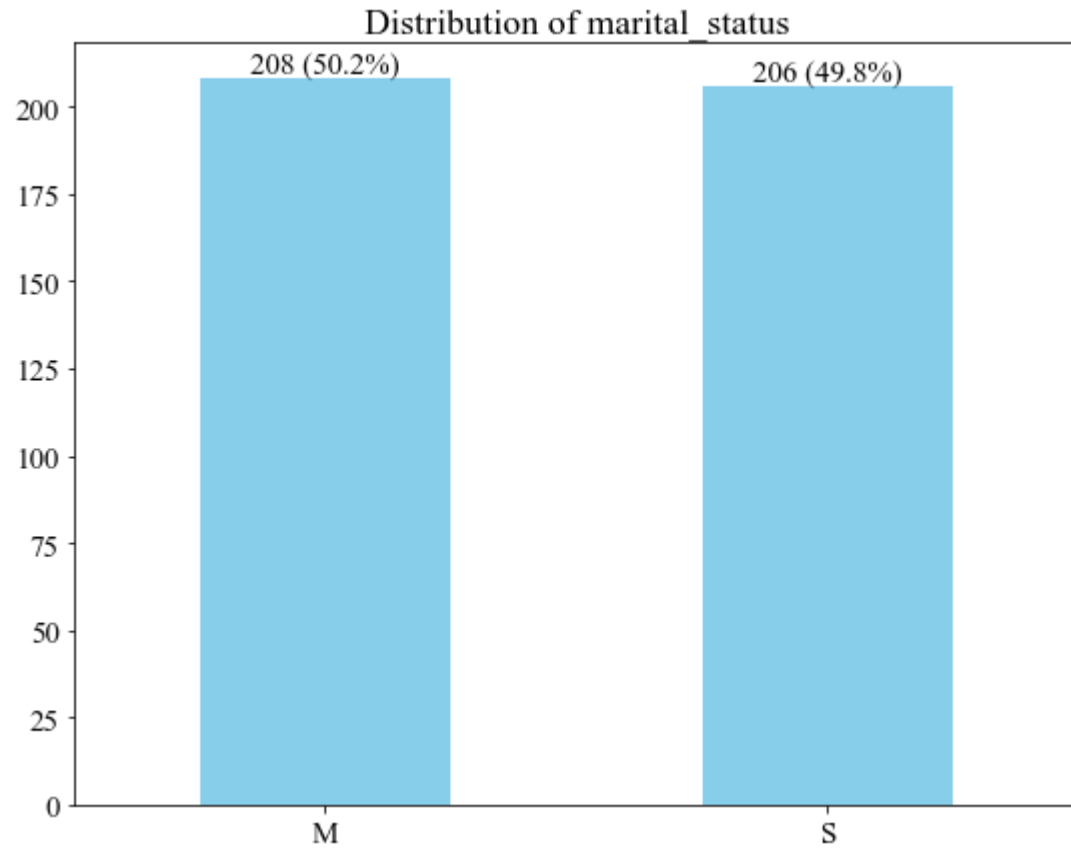
## HIGH VALUE CUSTOMER

## LOW VALUE CUSTOMER

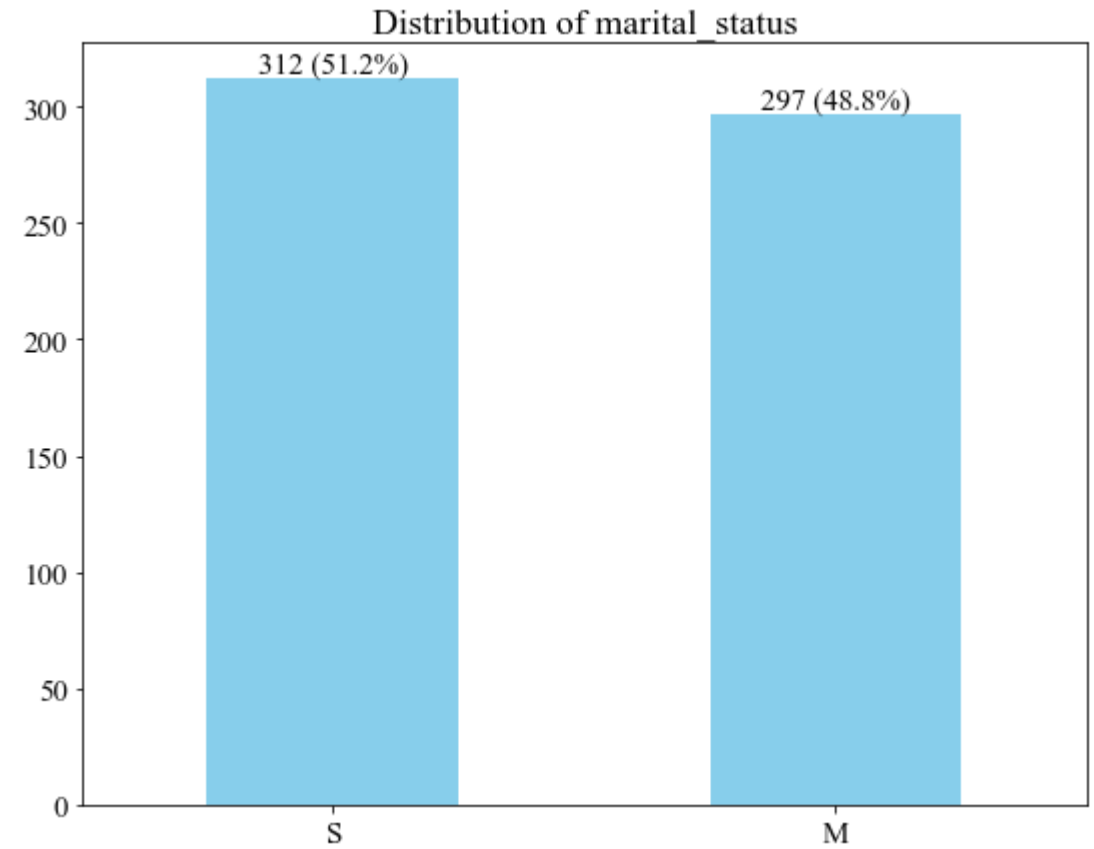


Focus on Mexico rather than Canada

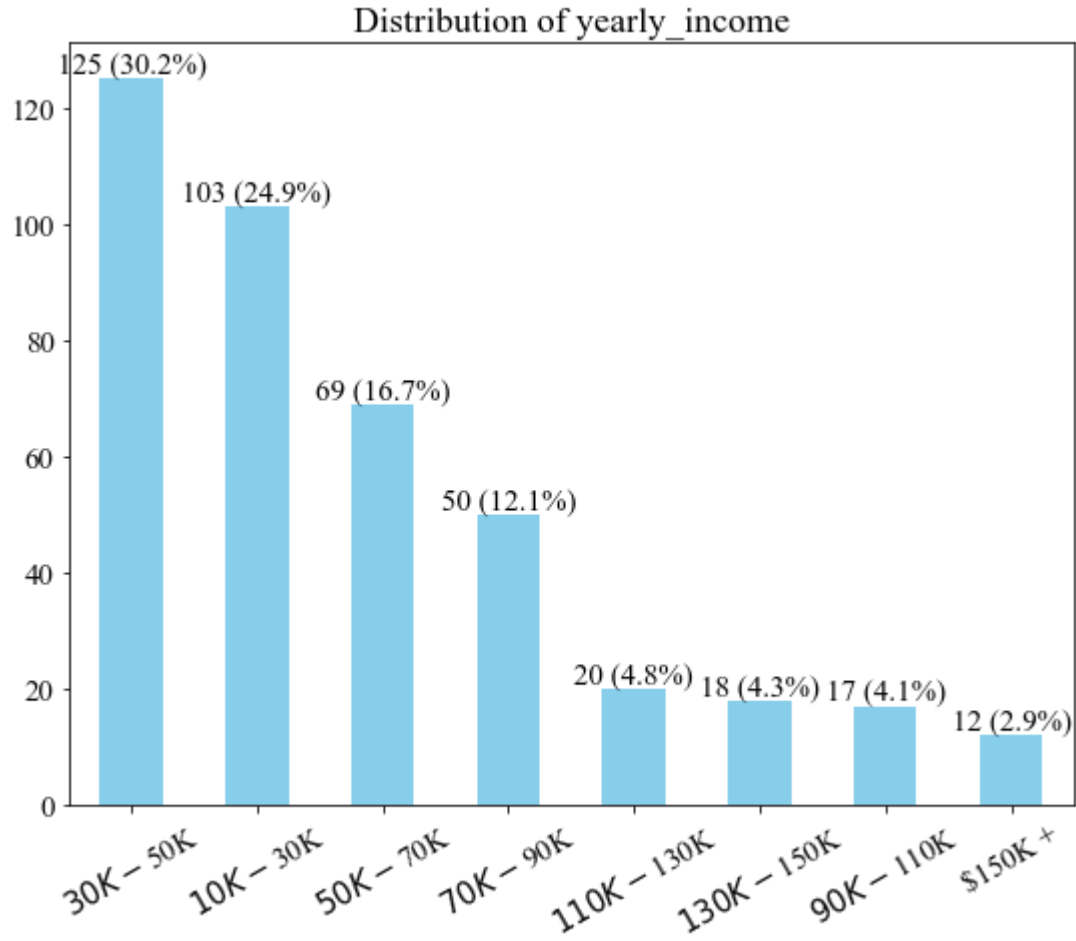
## HIGH VALUE CUSTOMER



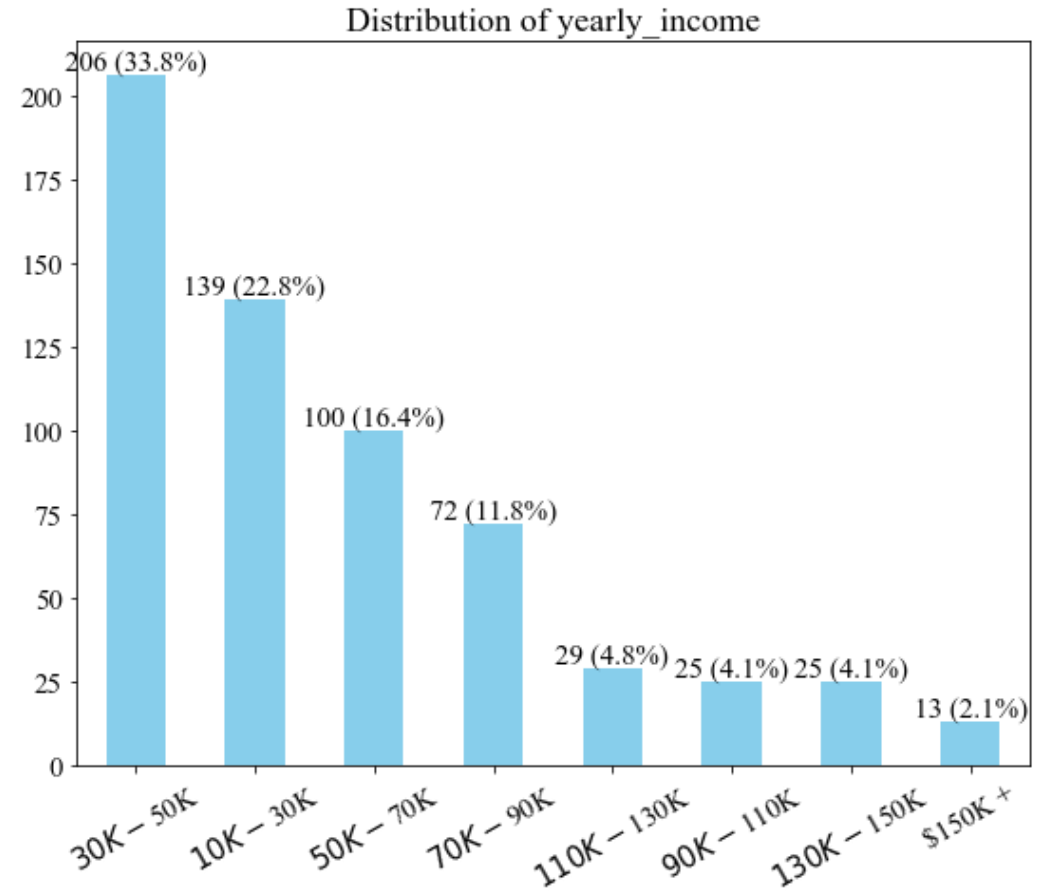
## LOW VALUE CUSTOMER



## HIGH VALUE CUSTOMER

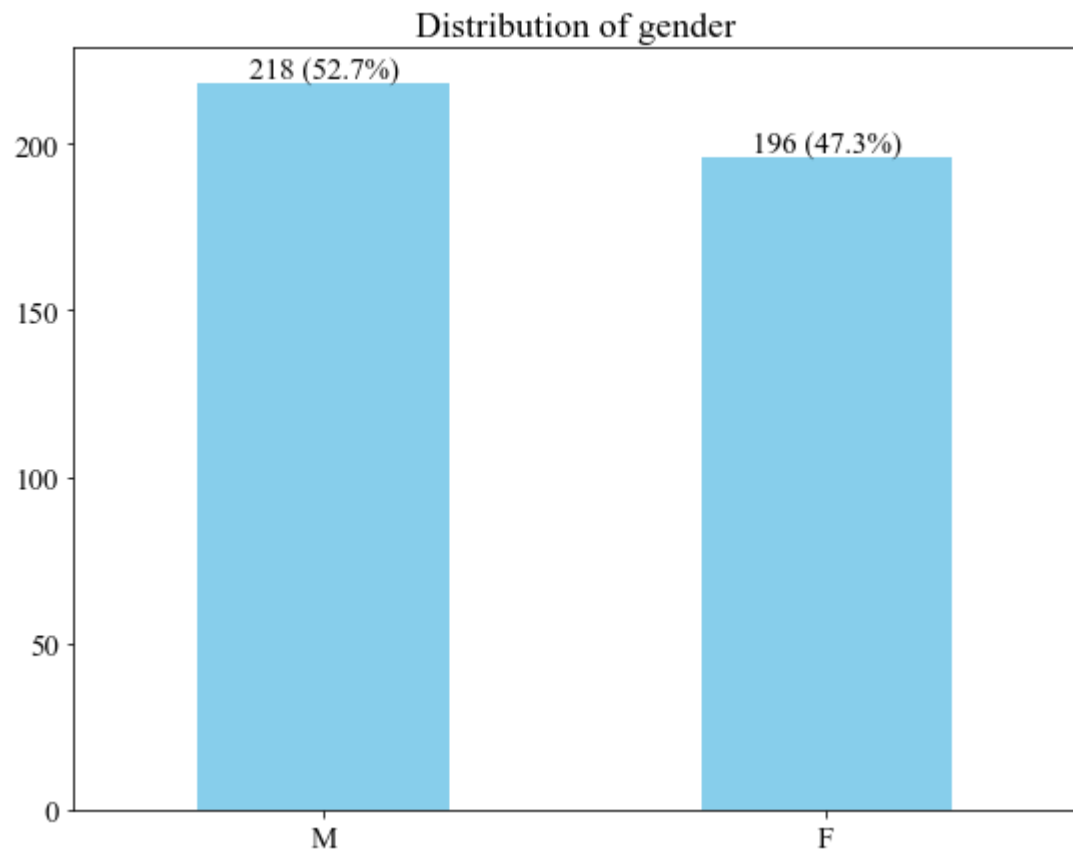


## LOW VALUE CUSTOMER

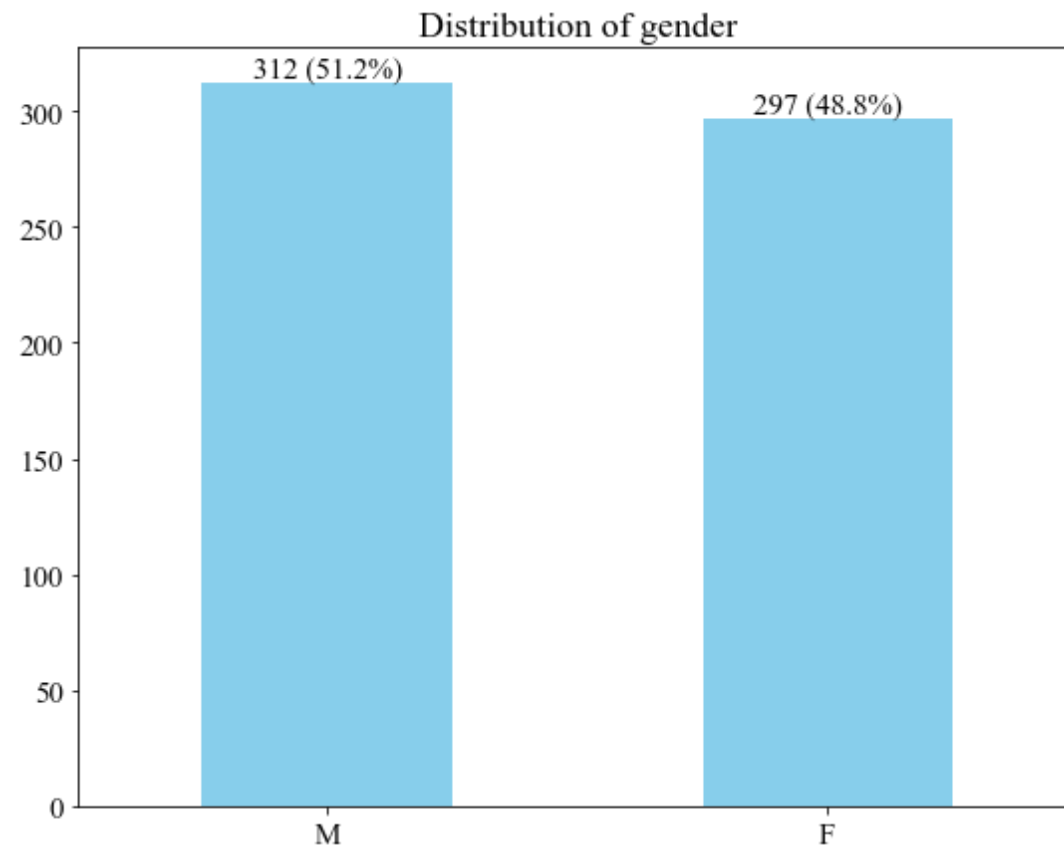




## HIGH VALUE CUSTOMER

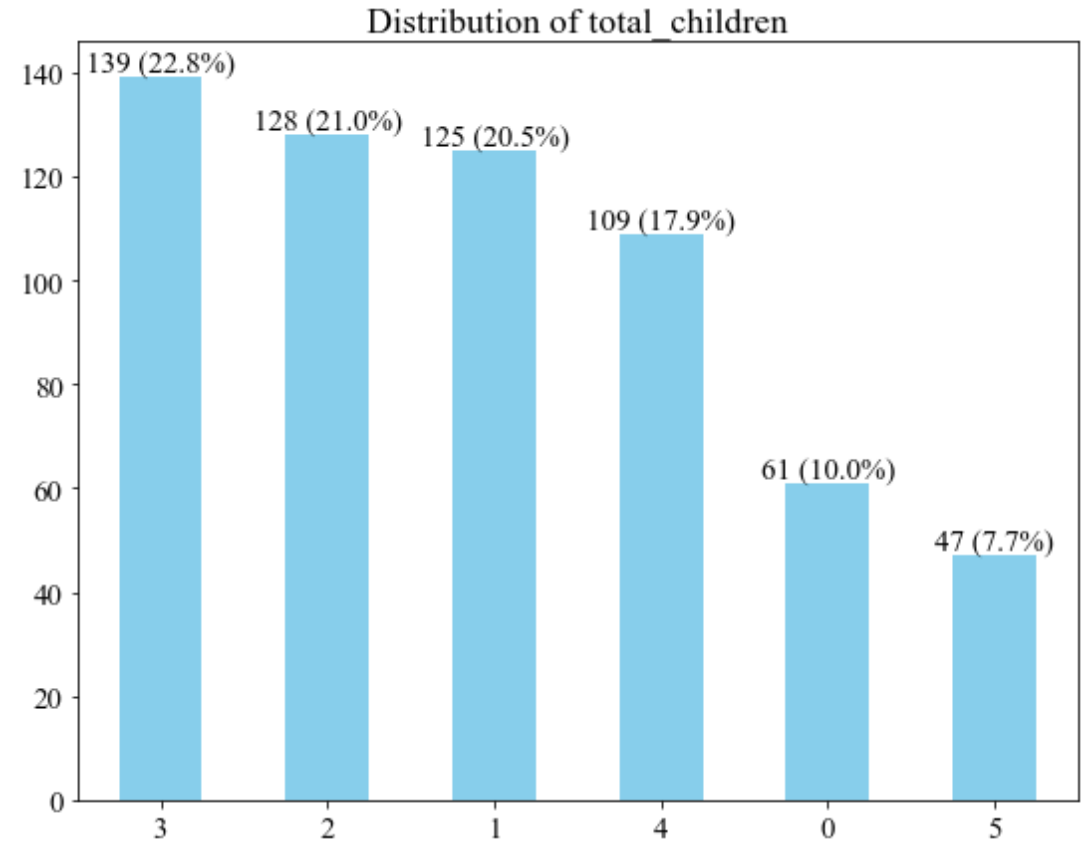
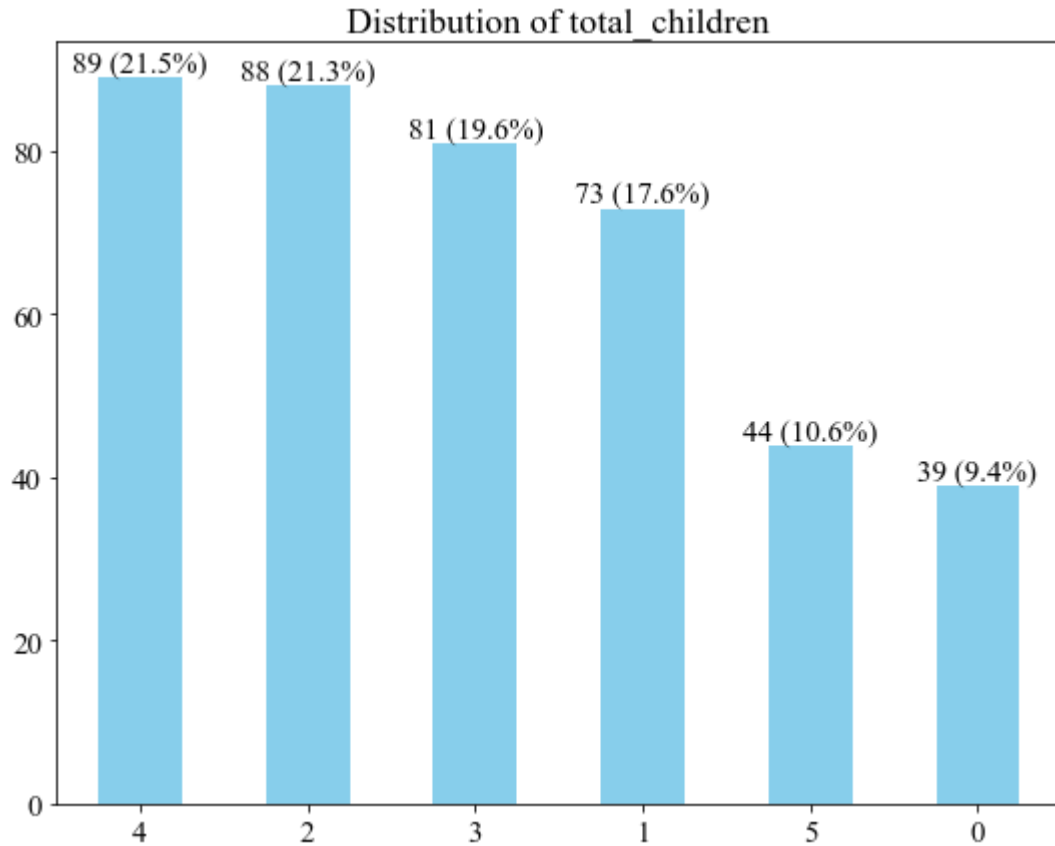


## LOW VALUE CUSTOMER



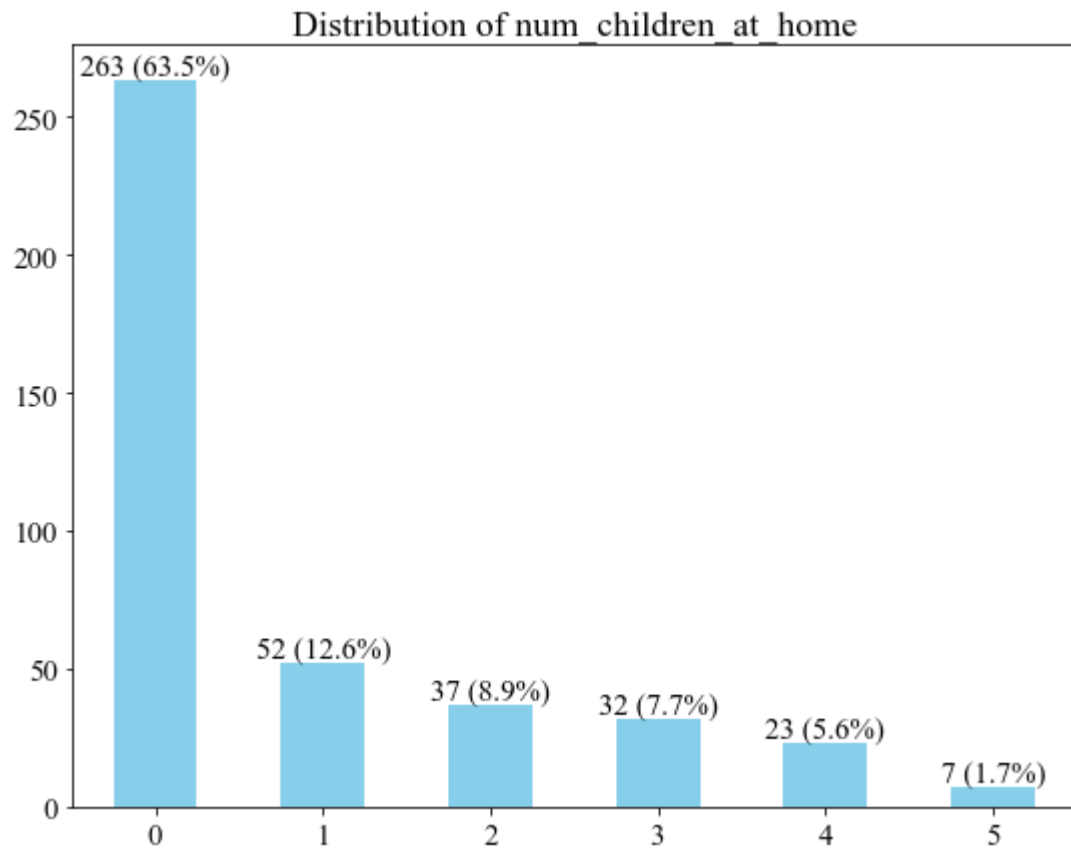
## HIGH VALUE CUSTOMER

## LOW VALUE CUSTOMER

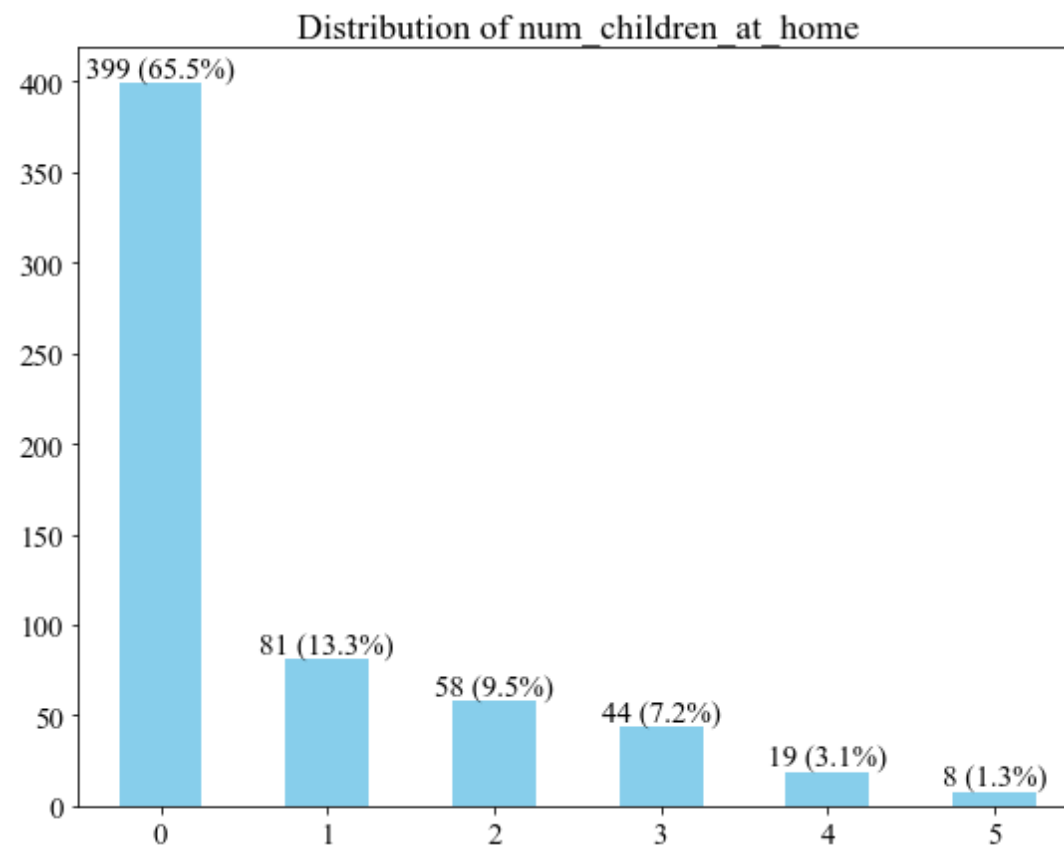


High value customers have more children

## HIGH VALUE CUSTOMER



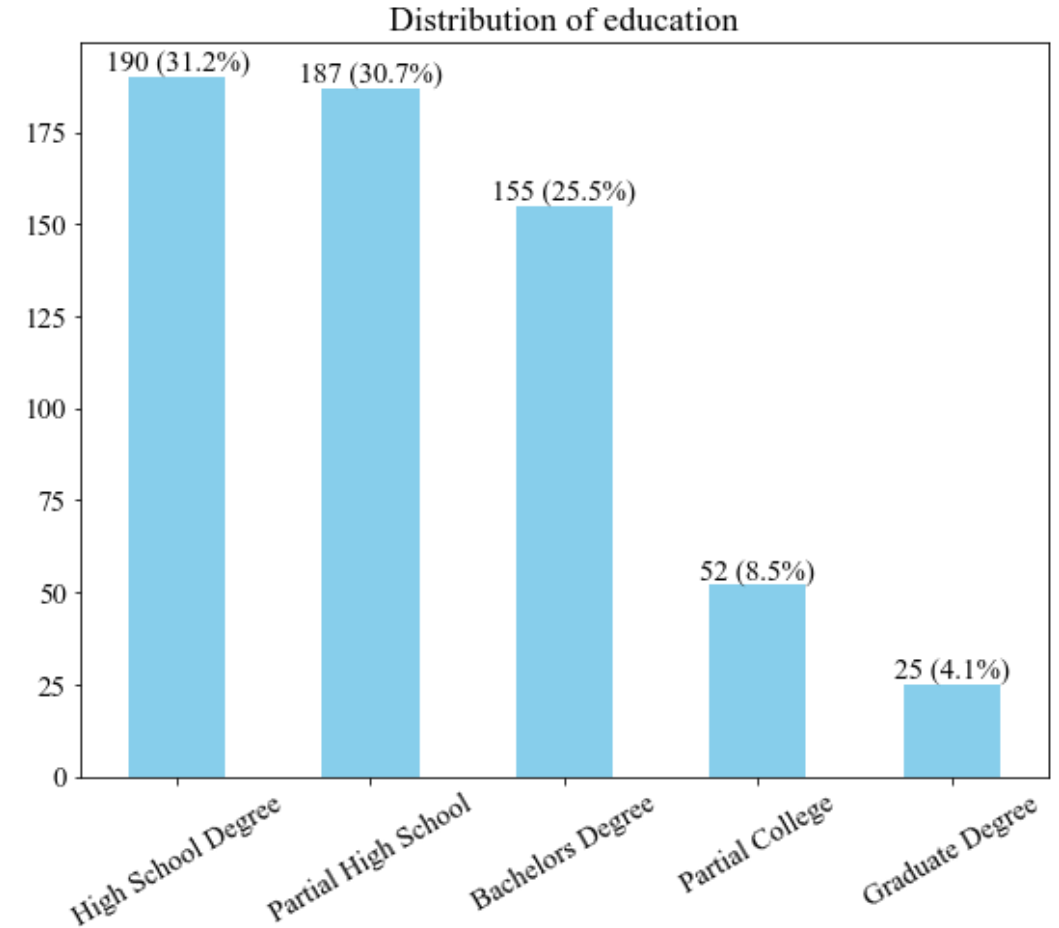
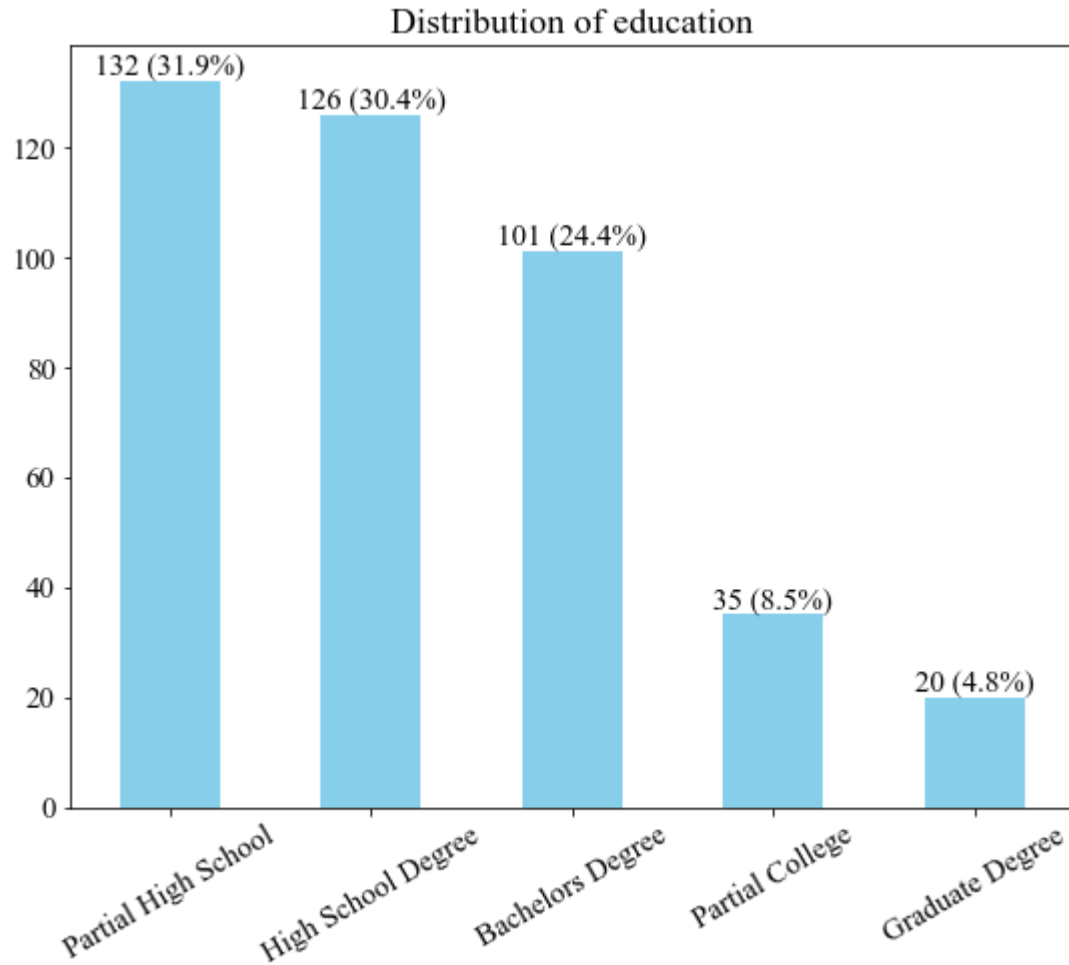
## LOW VALUE CUSTOMER



## HIGH VALUE CUSTOMER



## LOW VALUE CUSTOMER

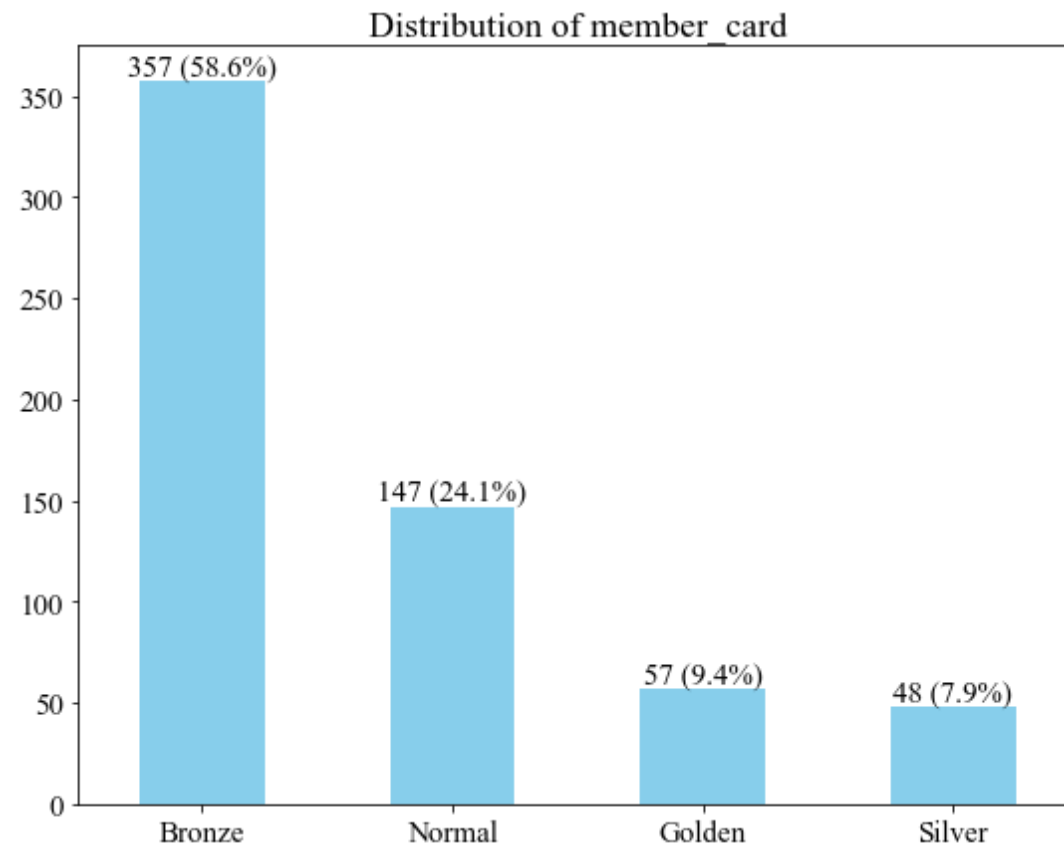
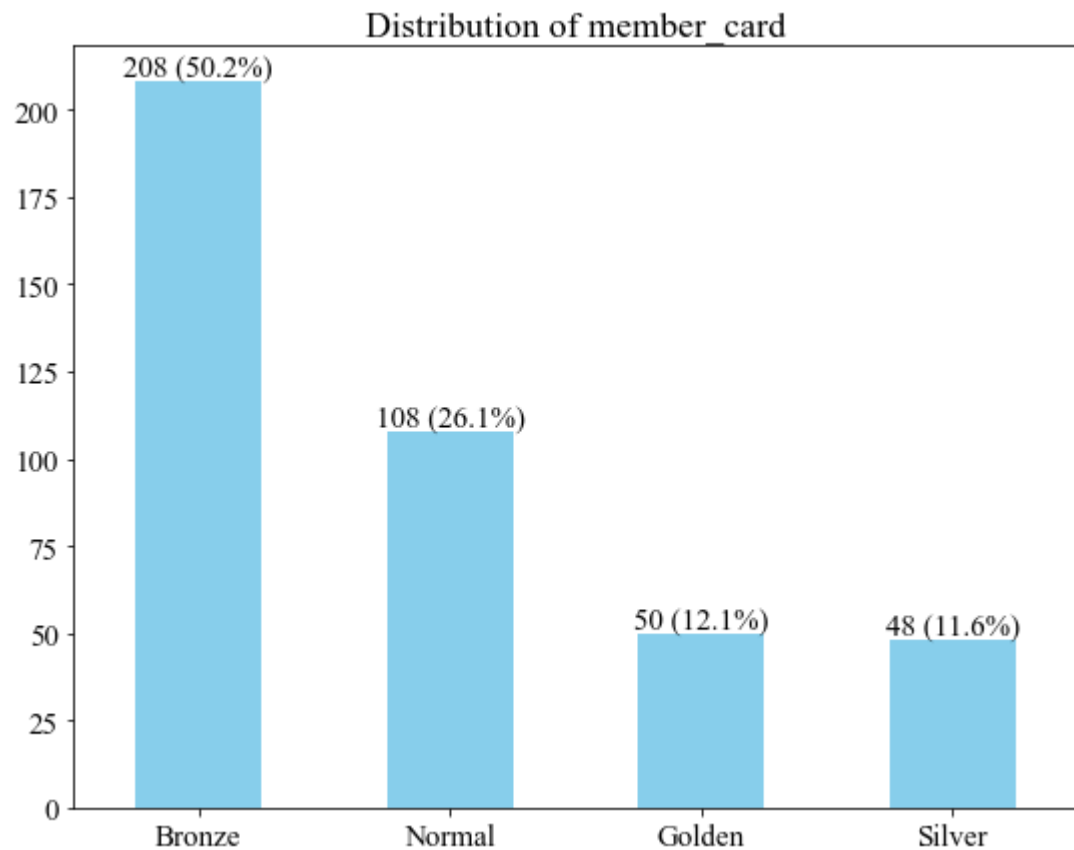


Most of our customers have low level of educations

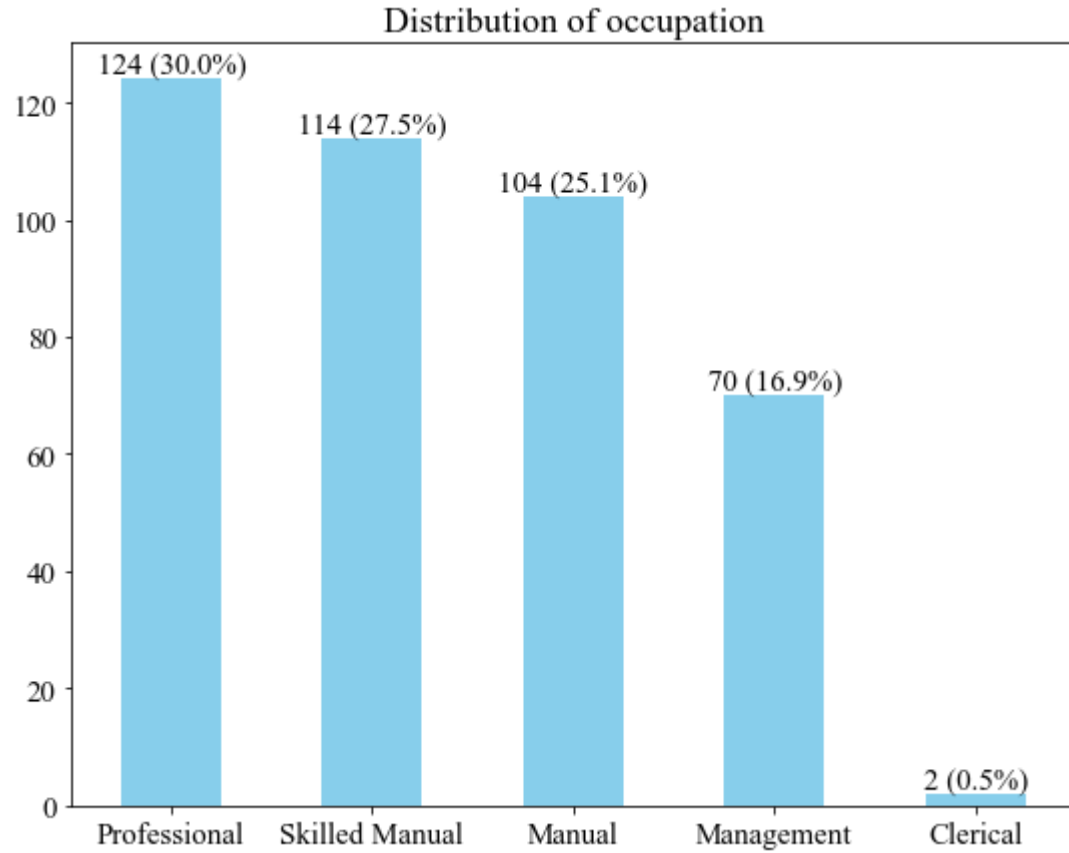
## HIGH VALUE CUSTOMER



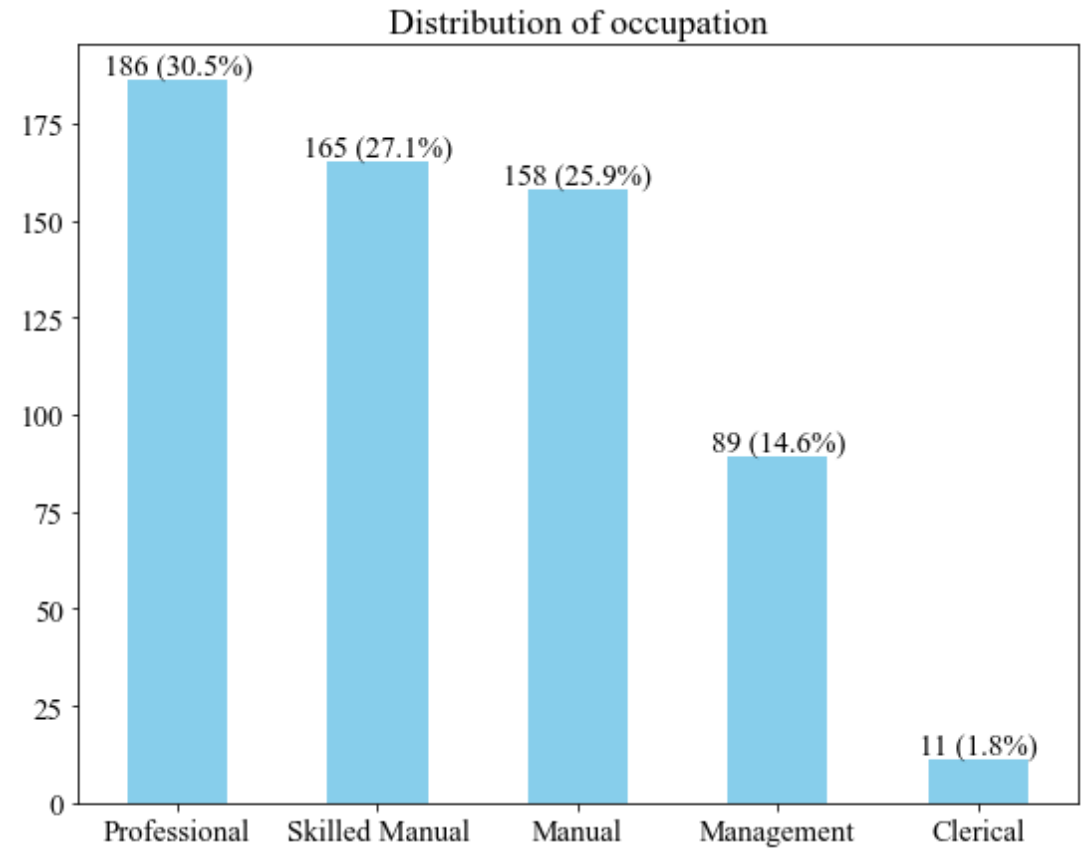
## LOW VALUE CUSTOMER



## HIGH VALUE CUSTOMER

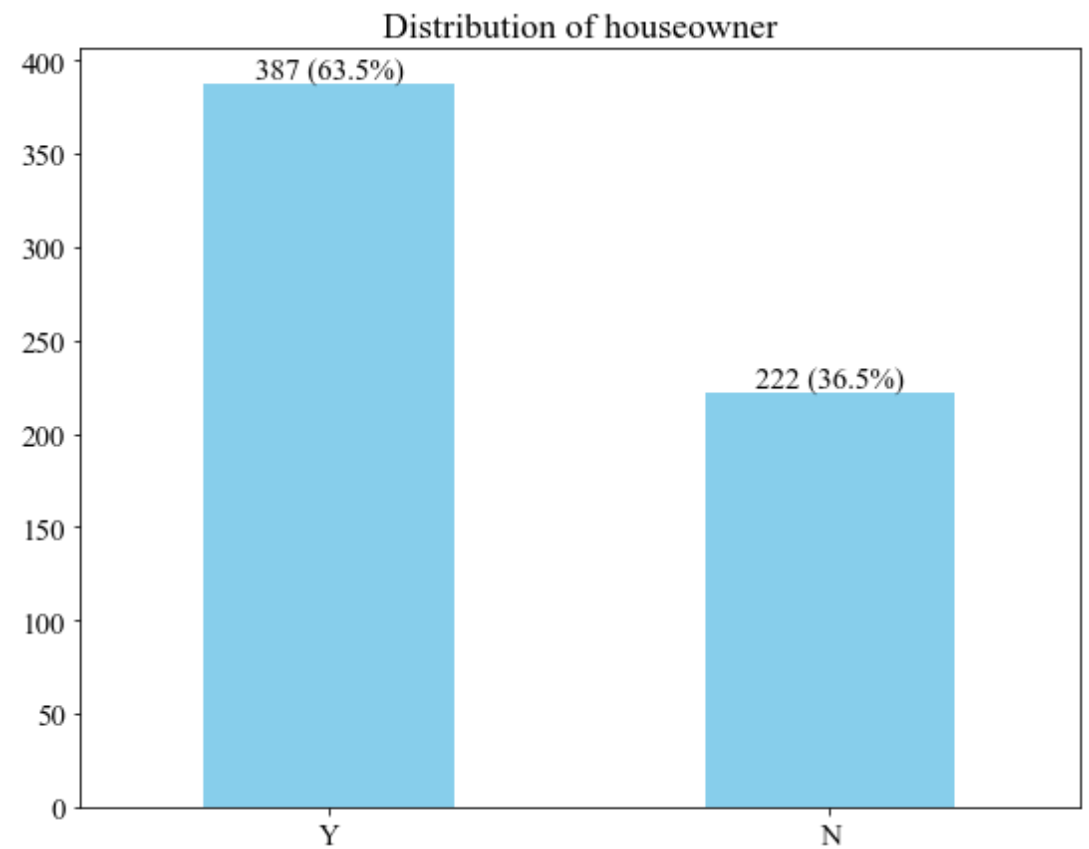
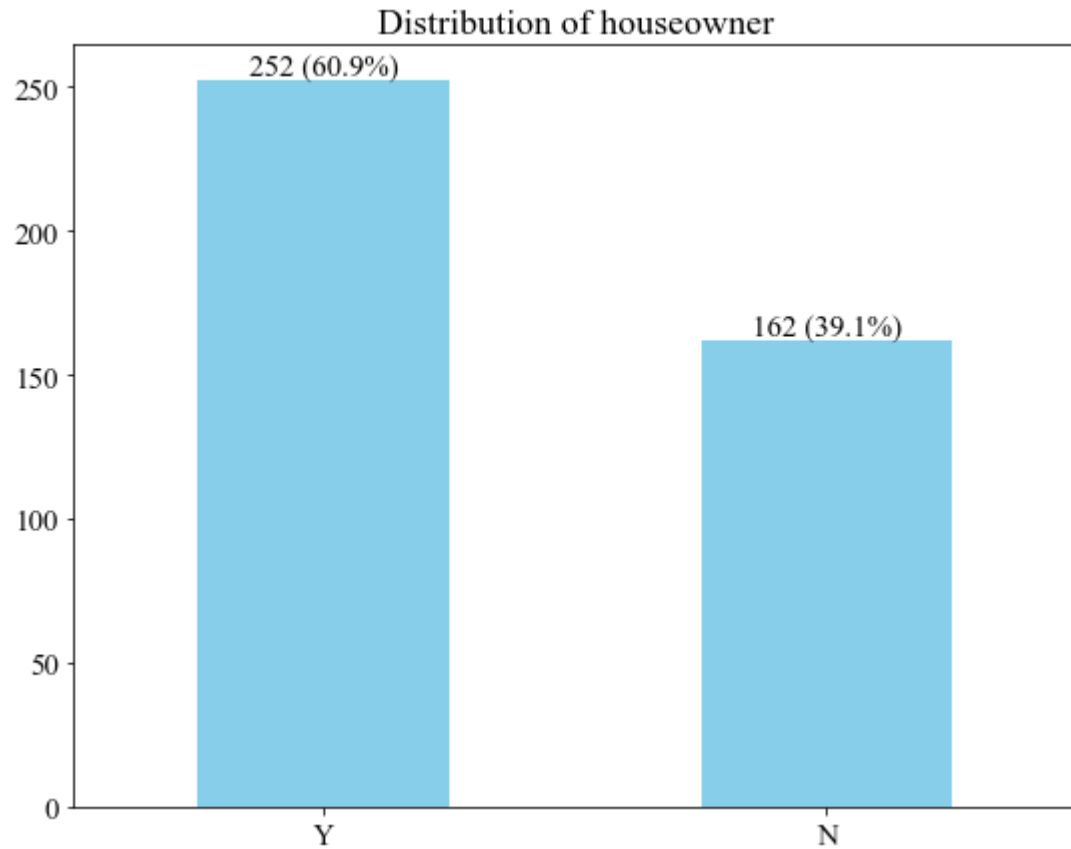


## LOW VALUE CUSTOMER



## HIGH VALUE CUSTOMER

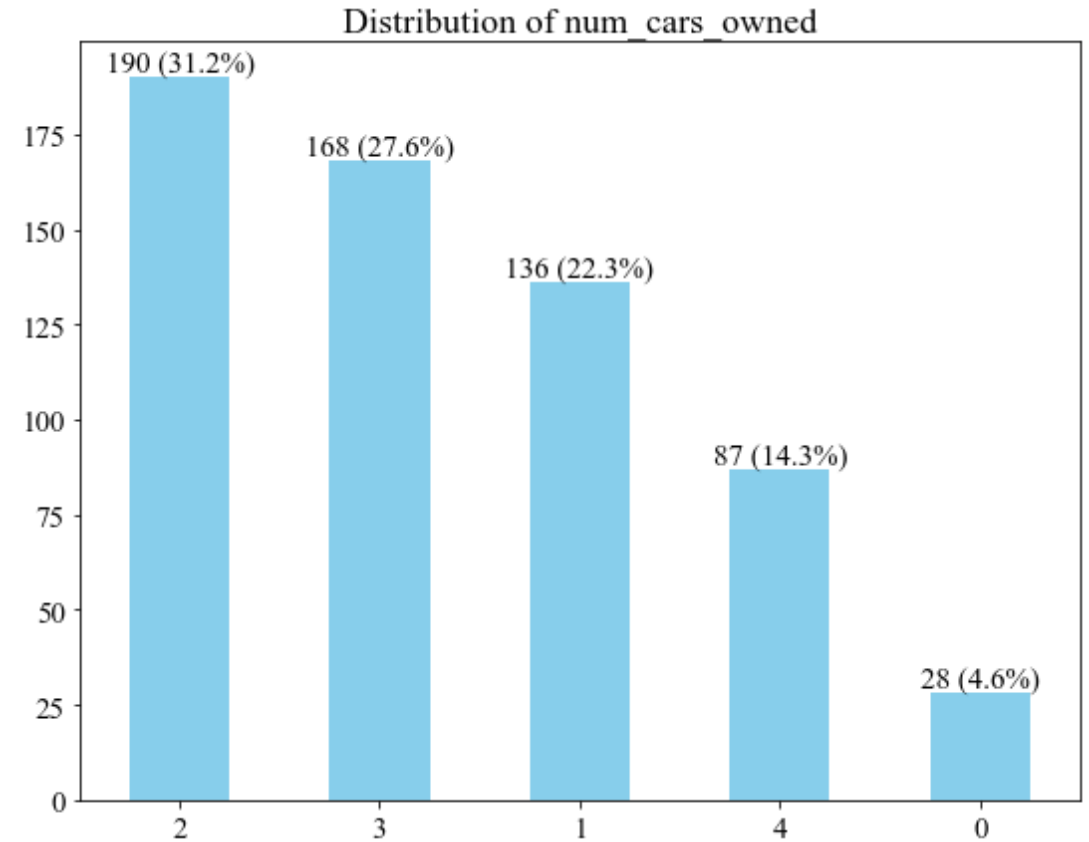
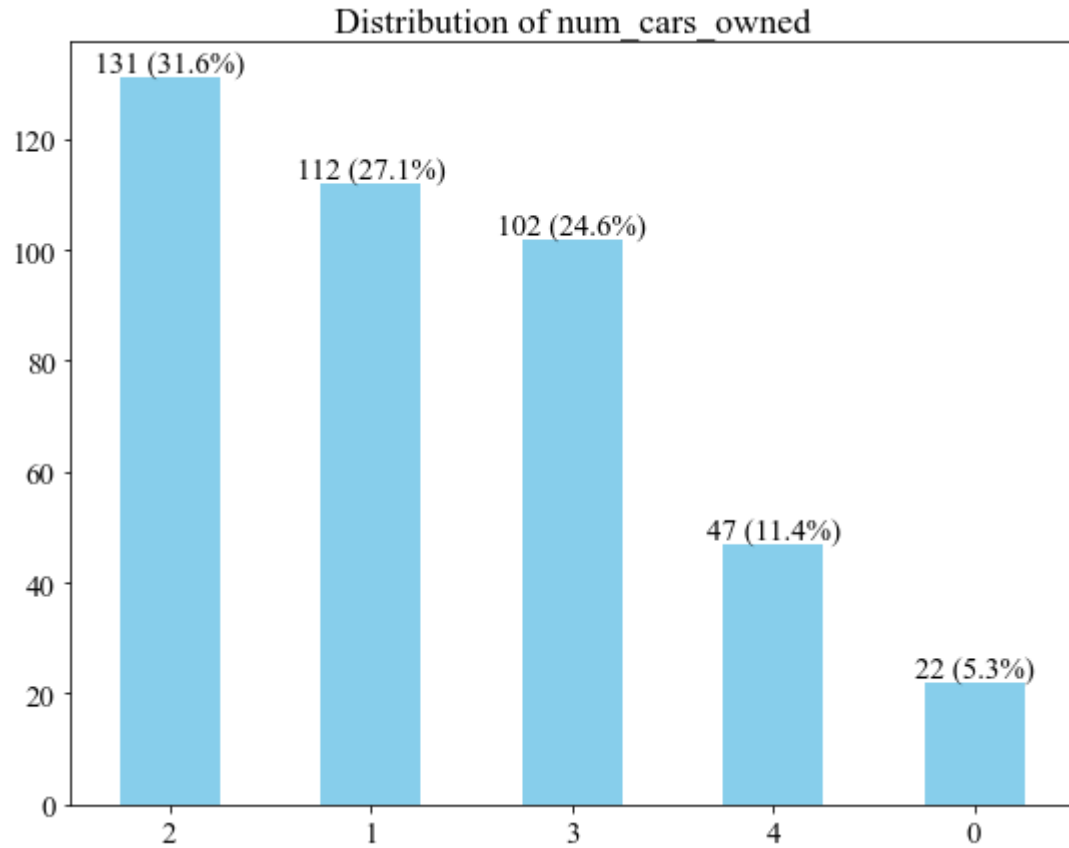
## LOW VALUE CUSTOMER



Most of our customers are houseowner

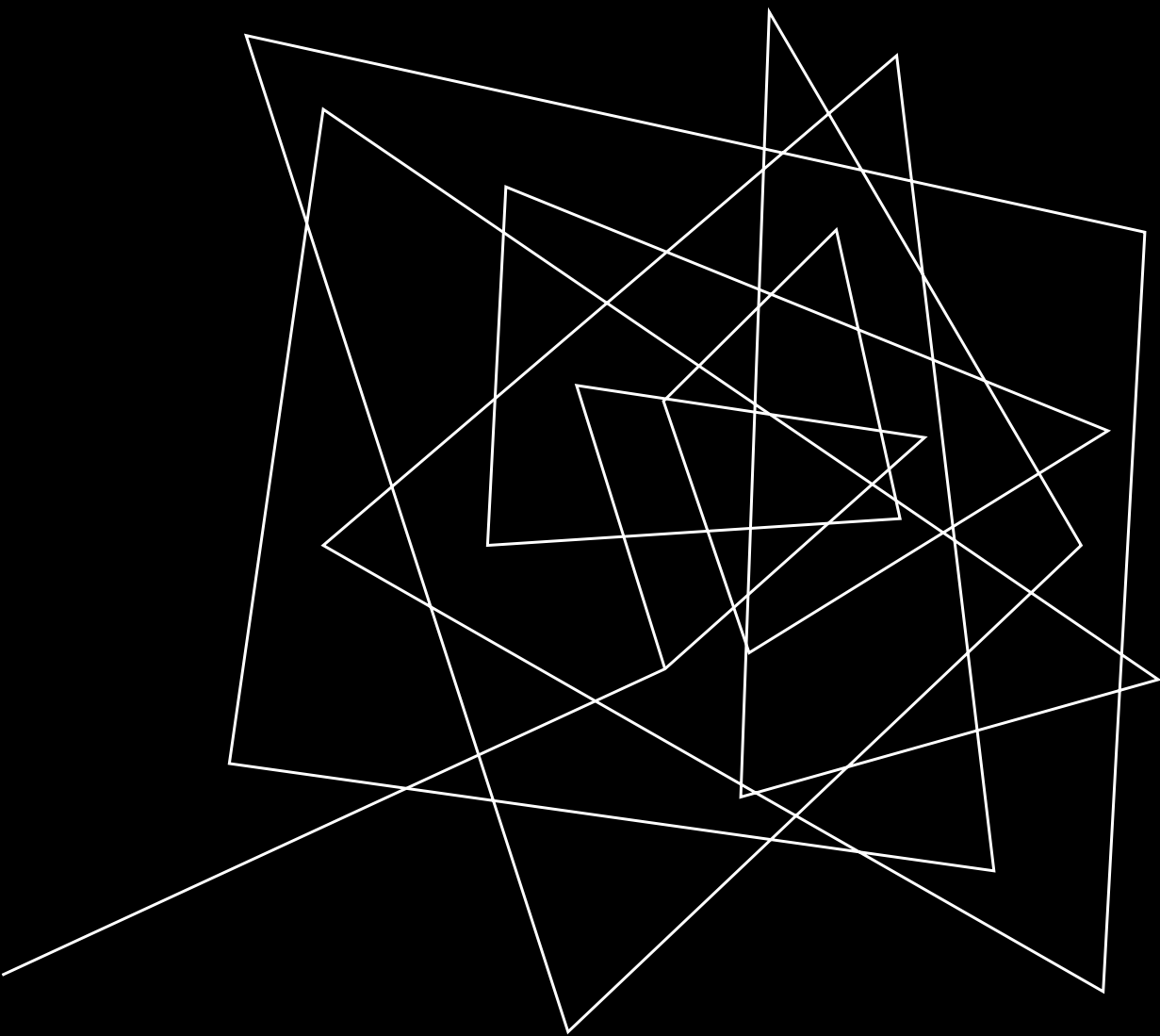
## HIGH VALUE CUSTOMER

## LOW VALUE CUSTOMER



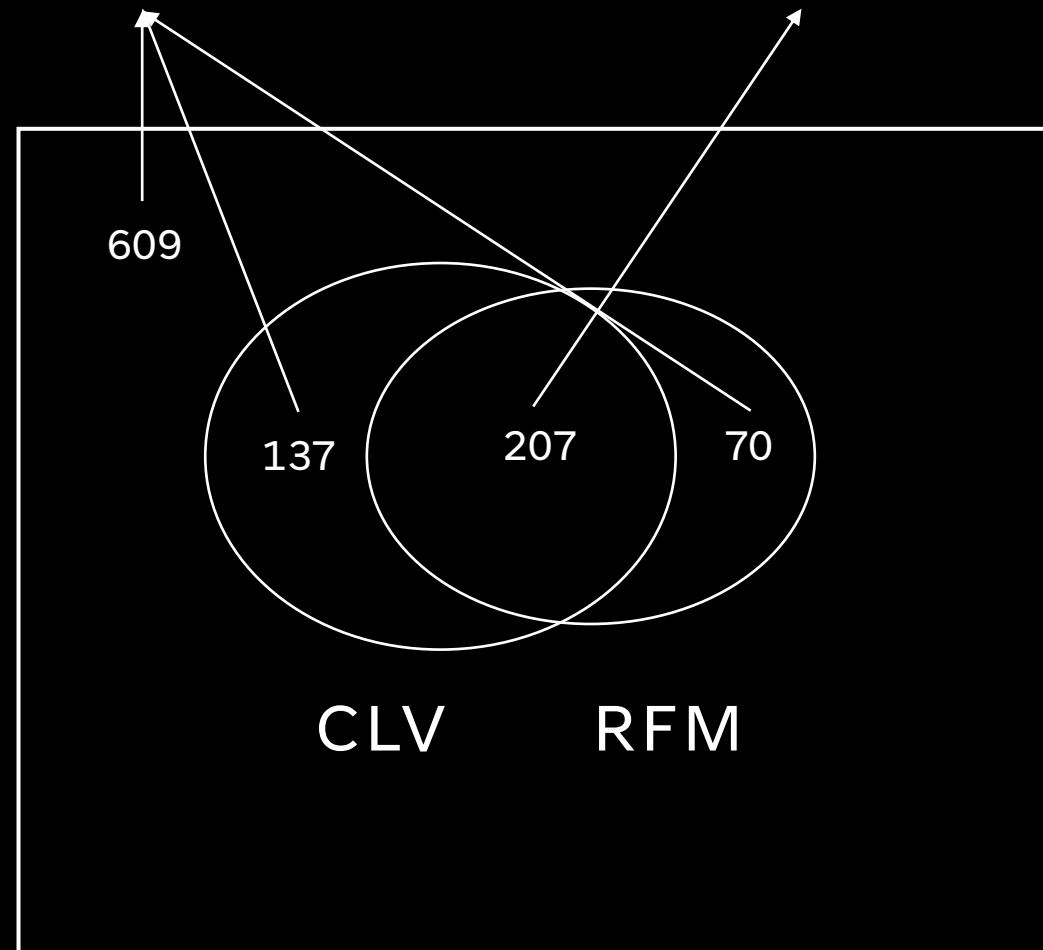
Most of our customers have at least 1 car



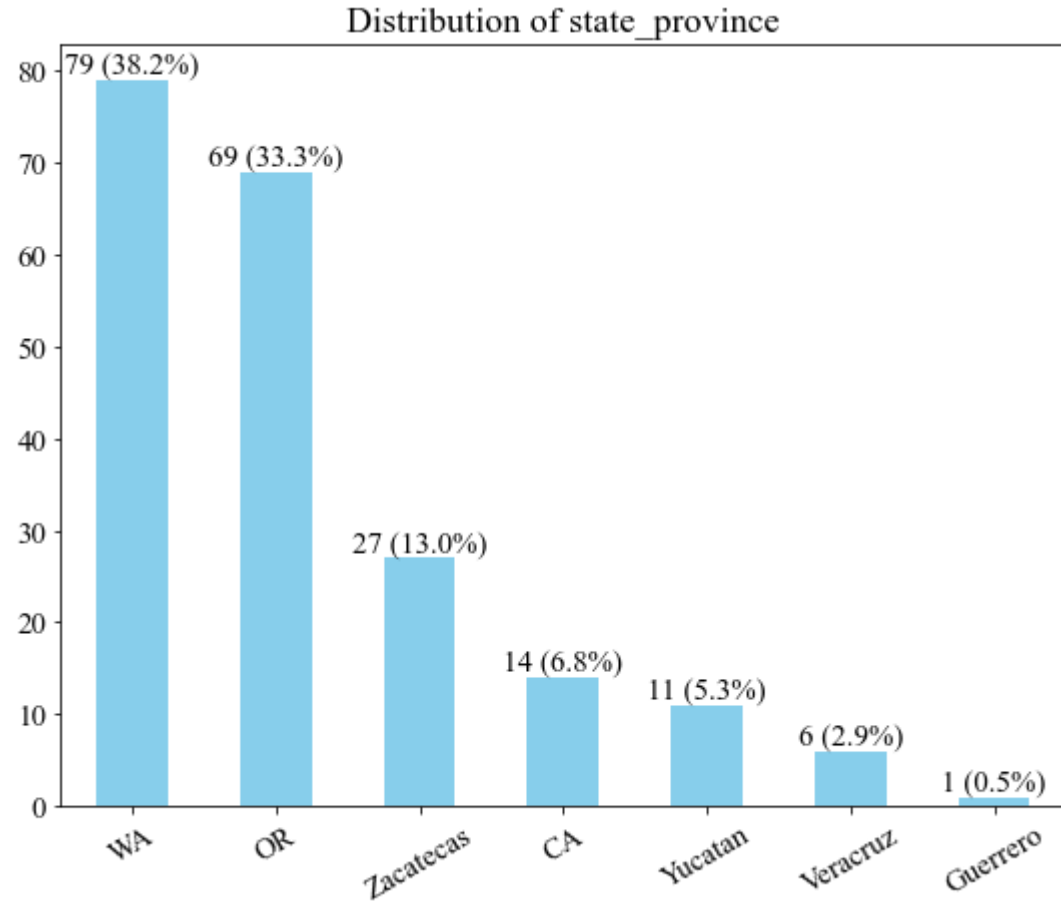


Low value customers

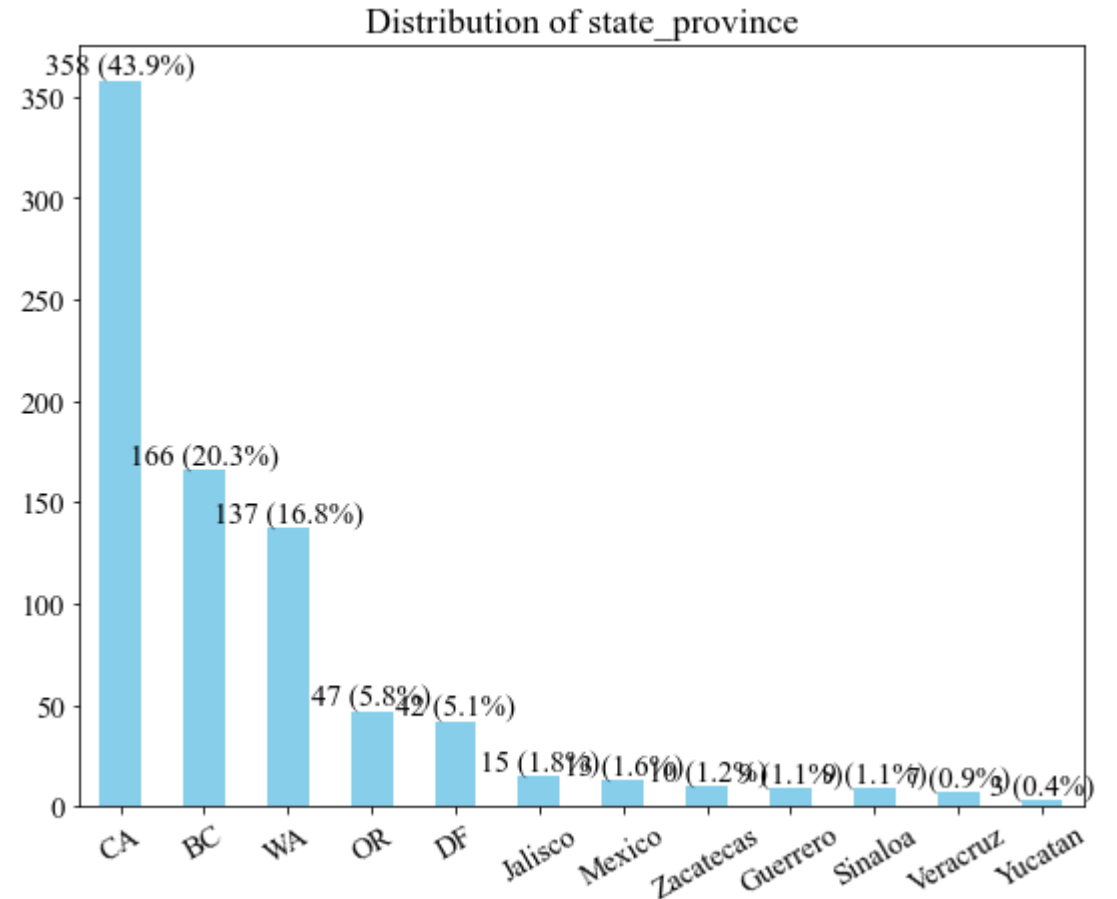
High value customers



## HIGH VALUE CUSTOMER

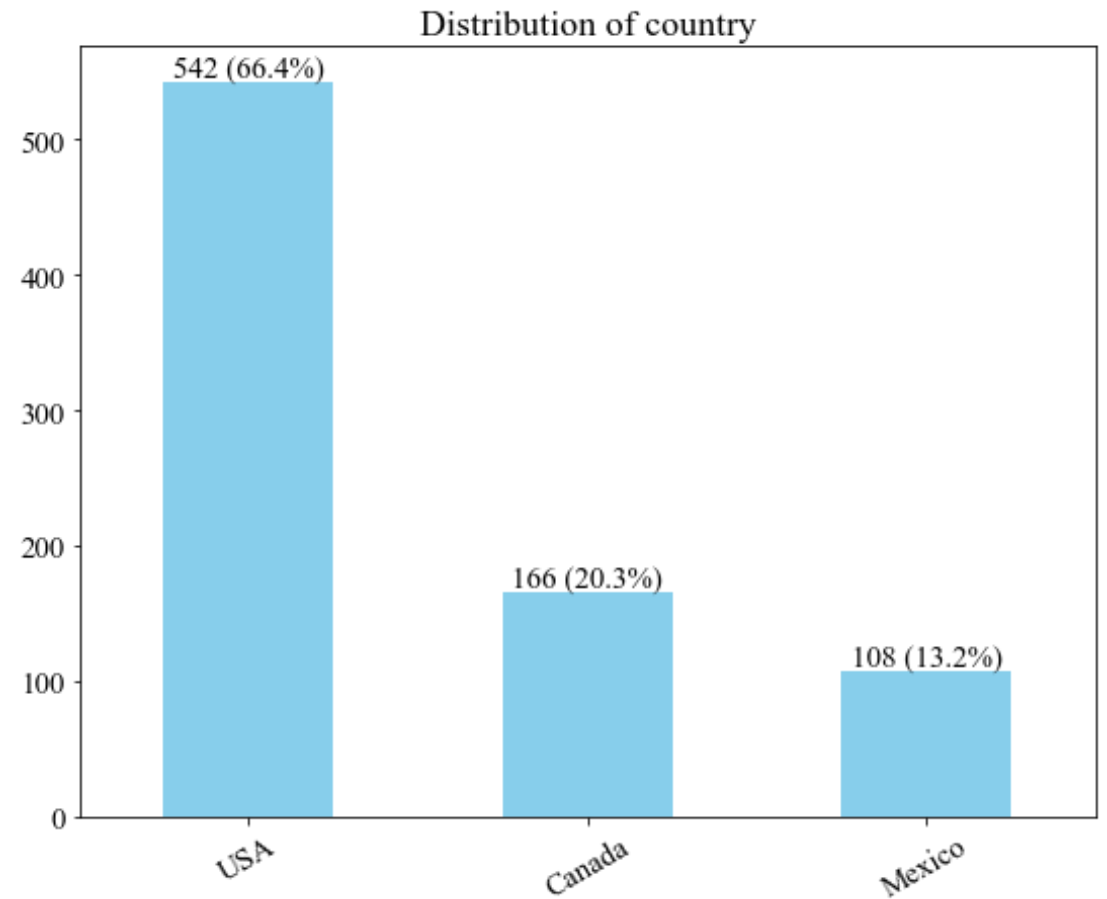
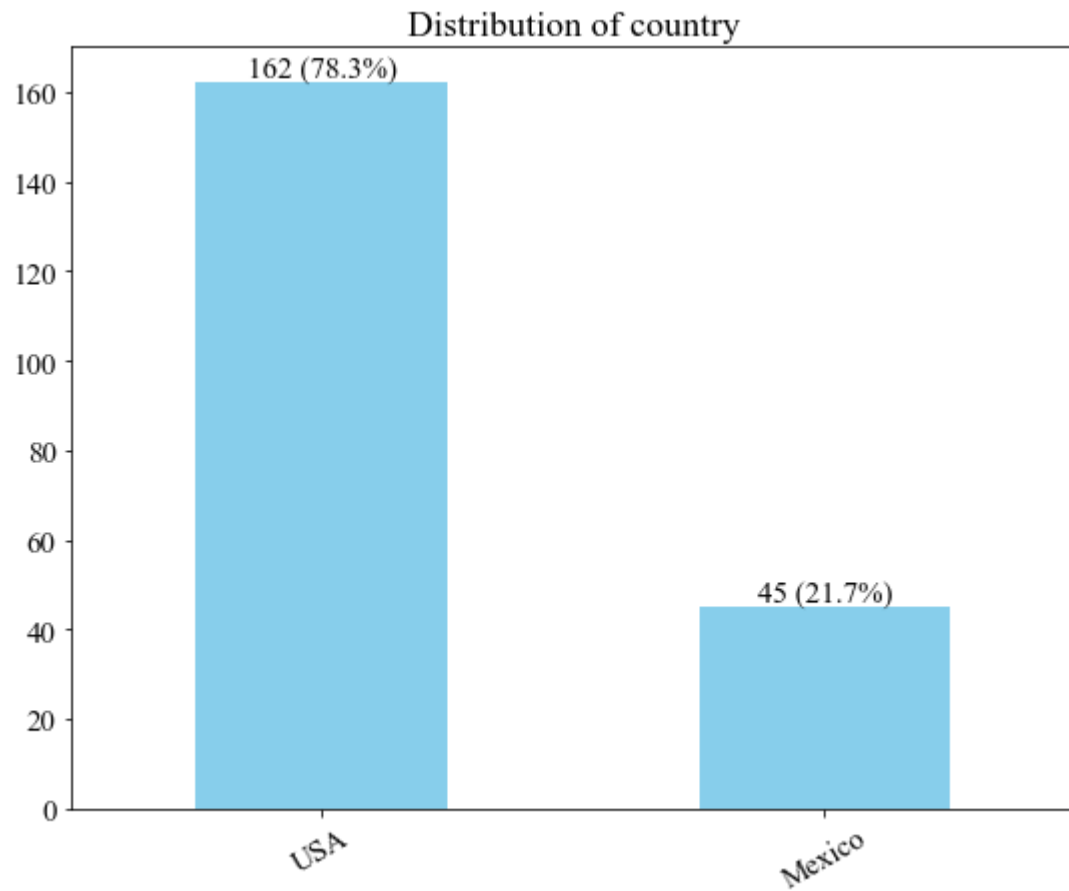


## LOW VALUE CUSTOMER



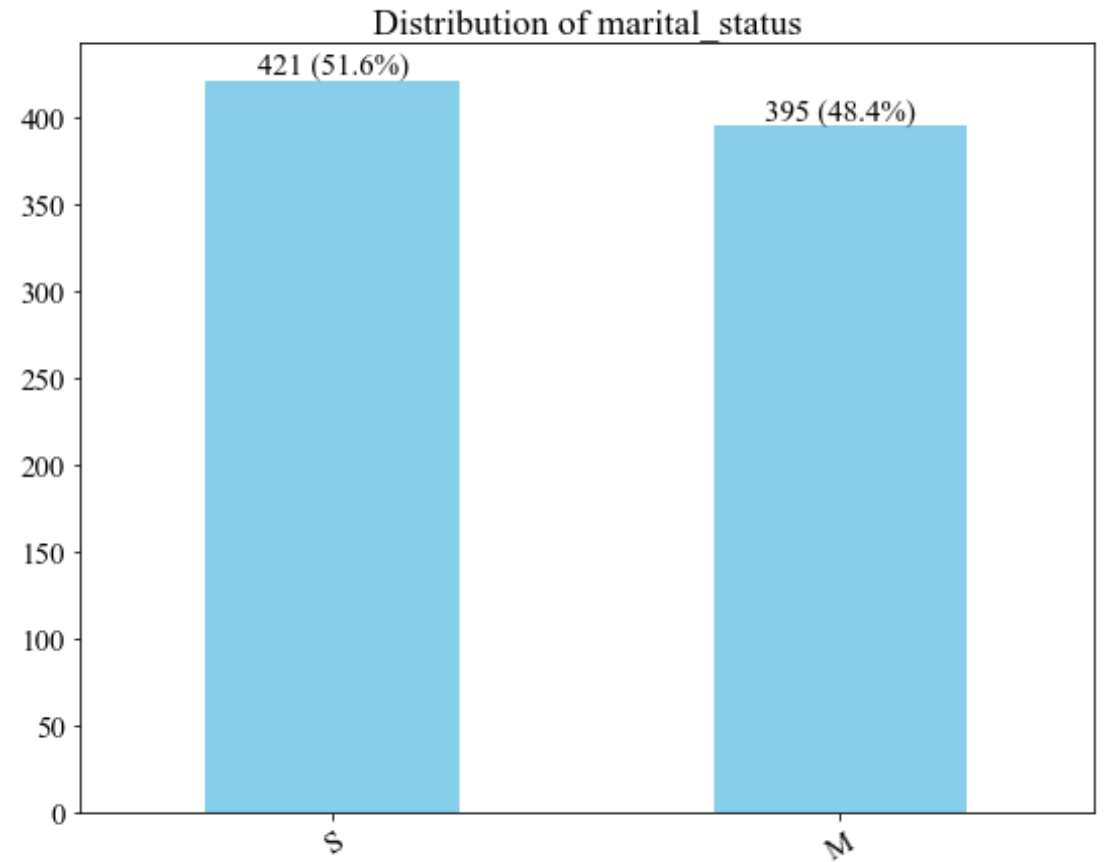
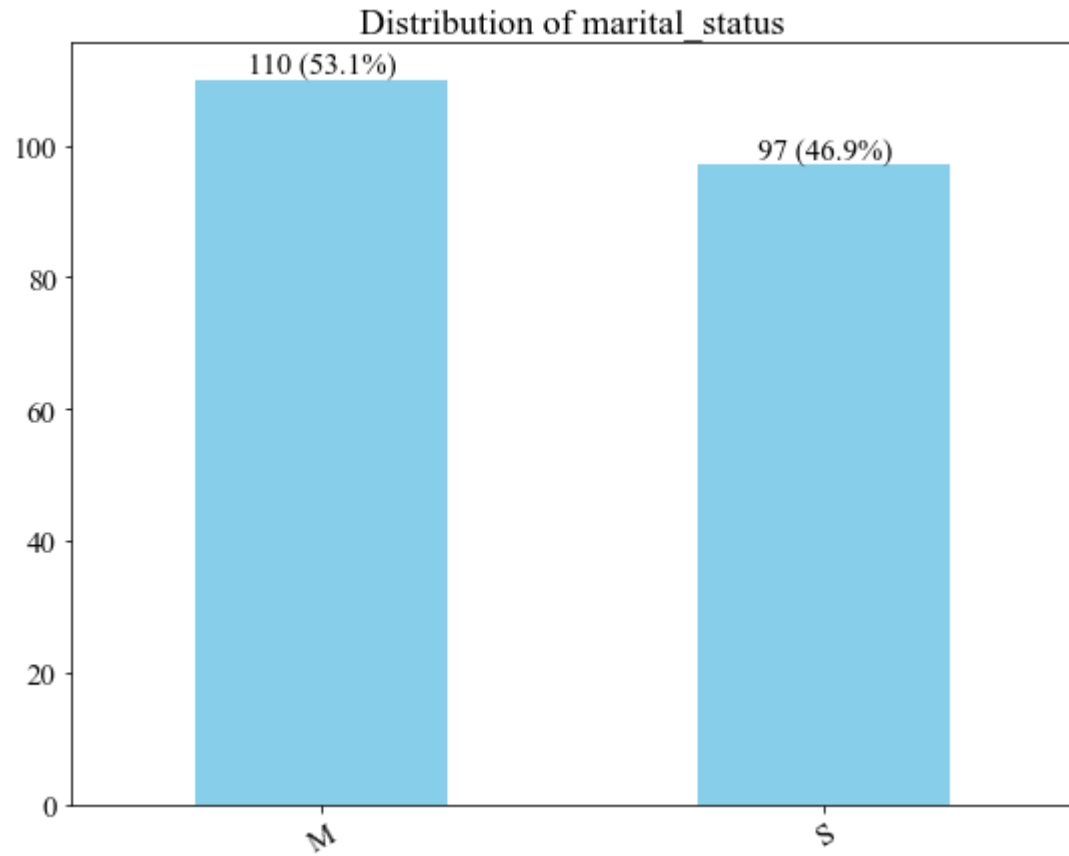
## HIGH VALUE CUSTOMER

## LOW VALUE CUSTOMER



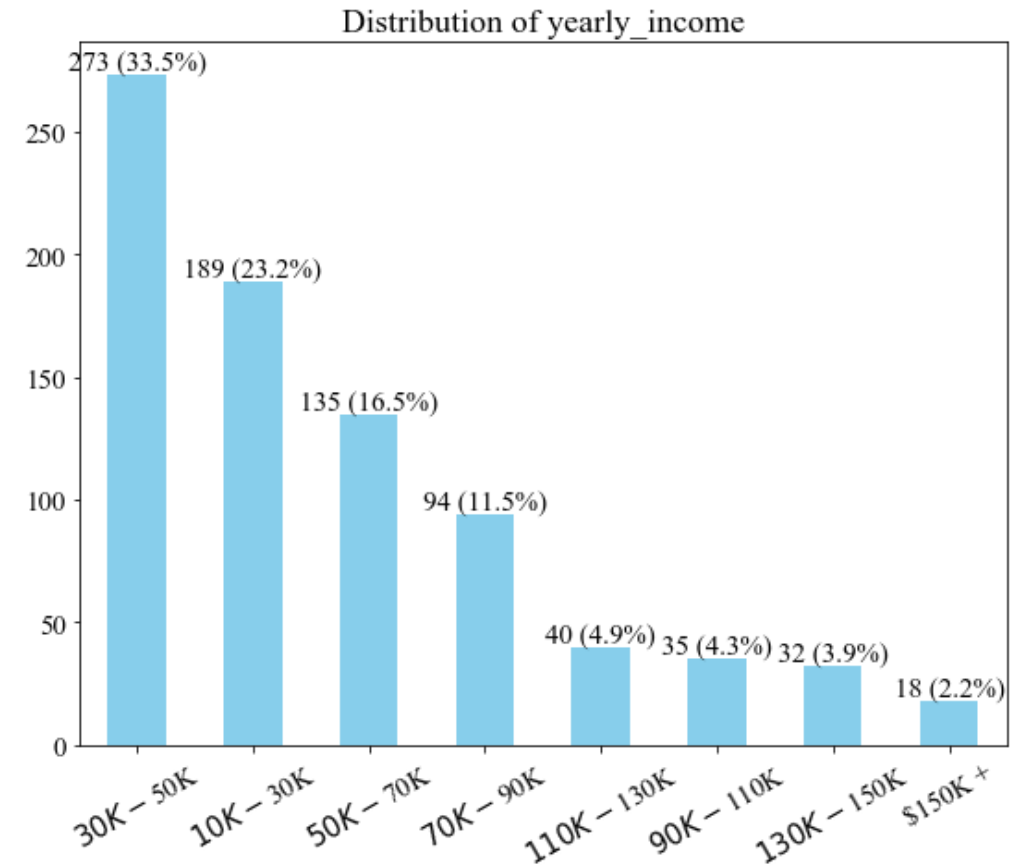
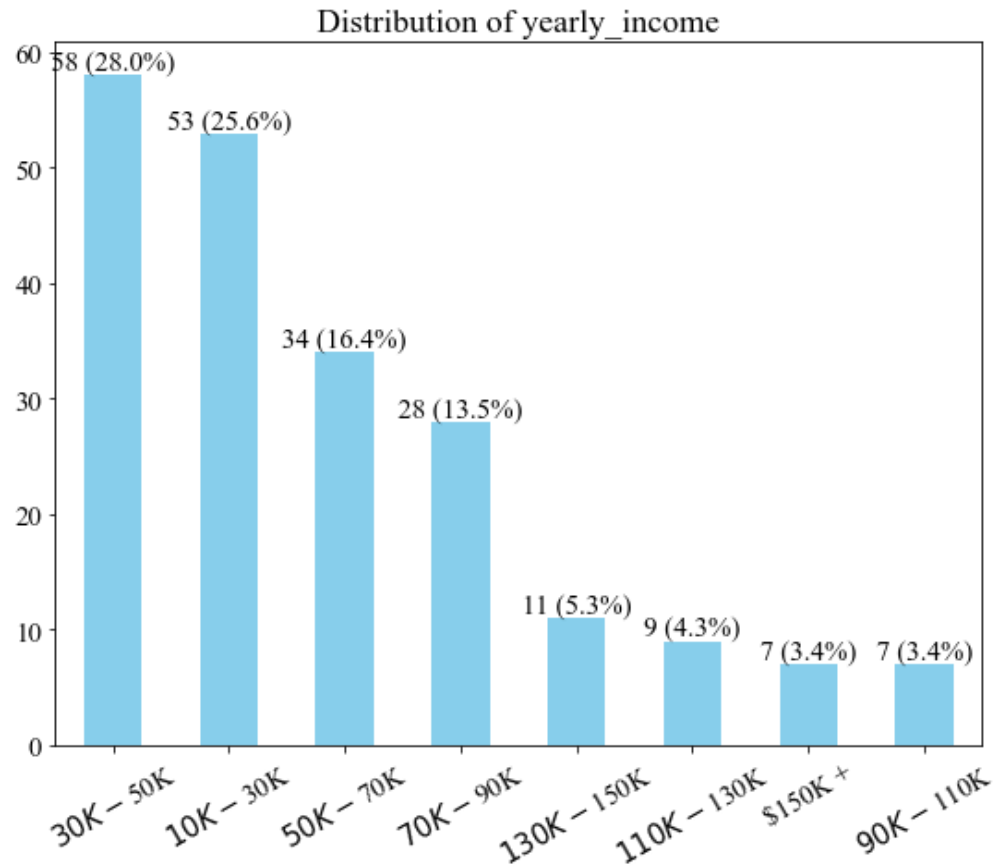
## HIGH VALUE CUSTOMER

## LOW VALUE CUSTOMER



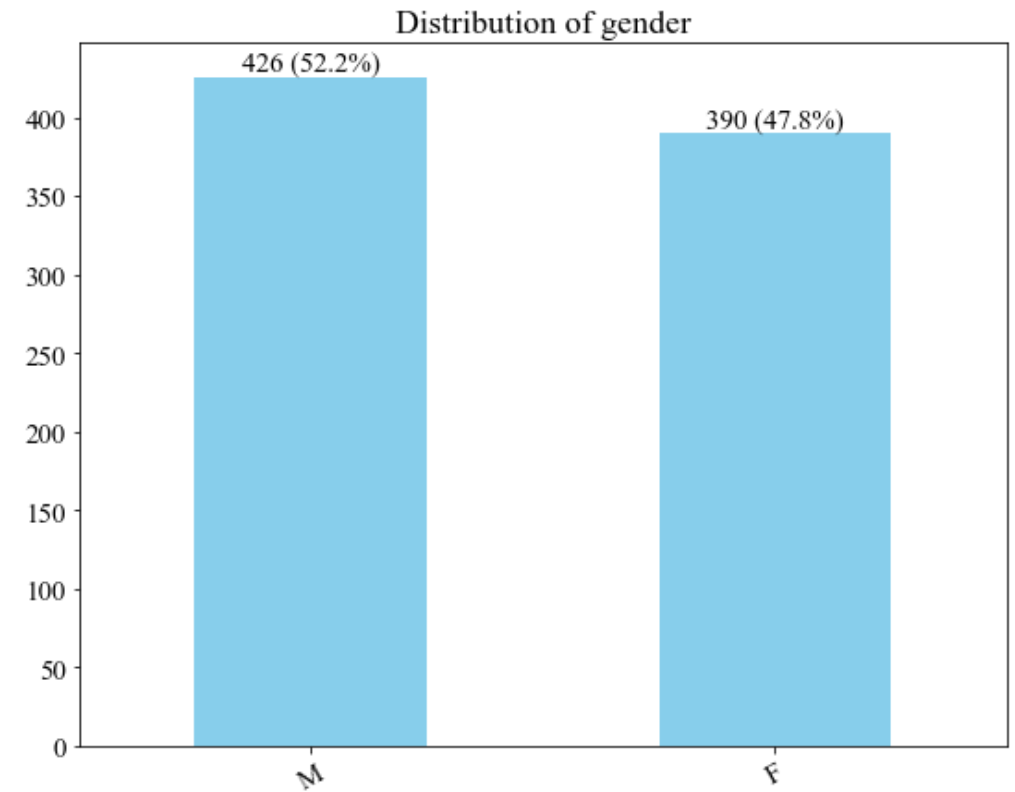
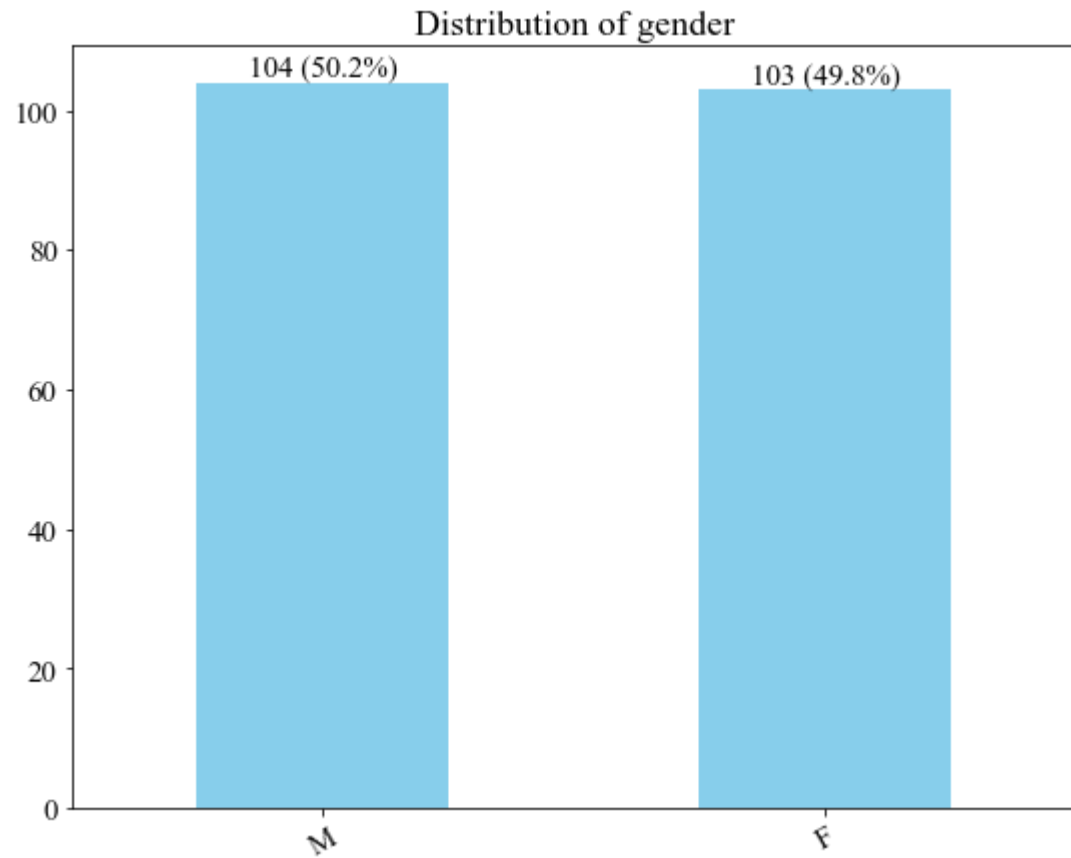
## HIGH VALUE CUSTOMER

## LOW VALUE CUSTOMER



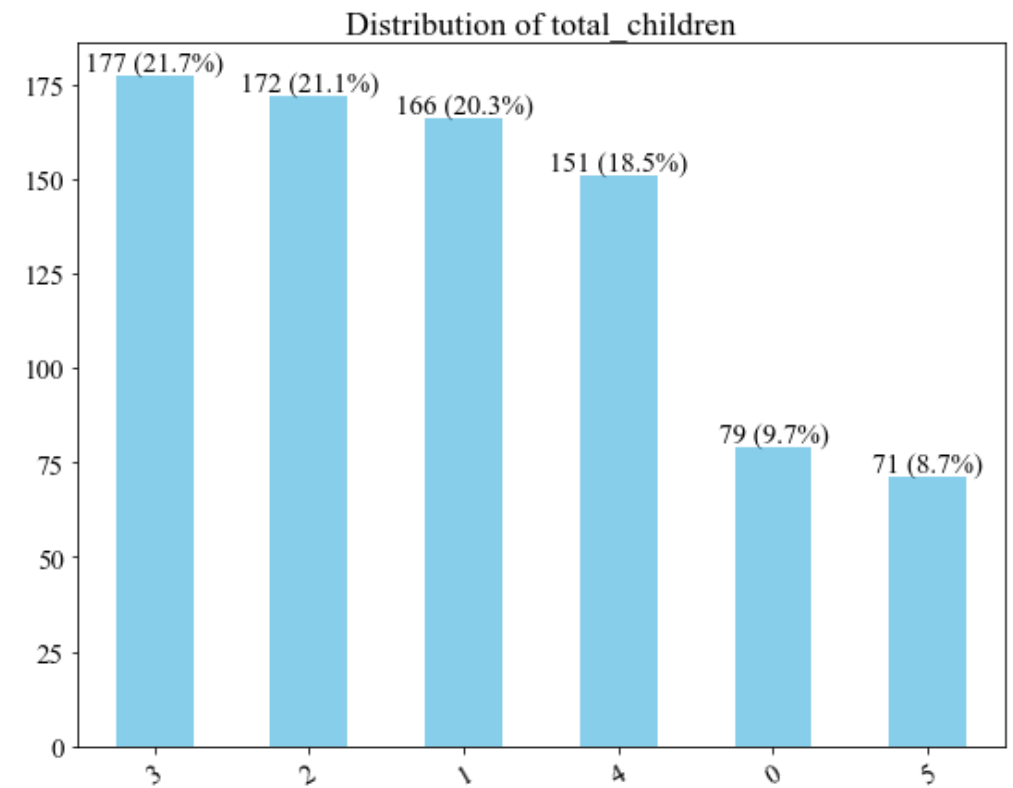
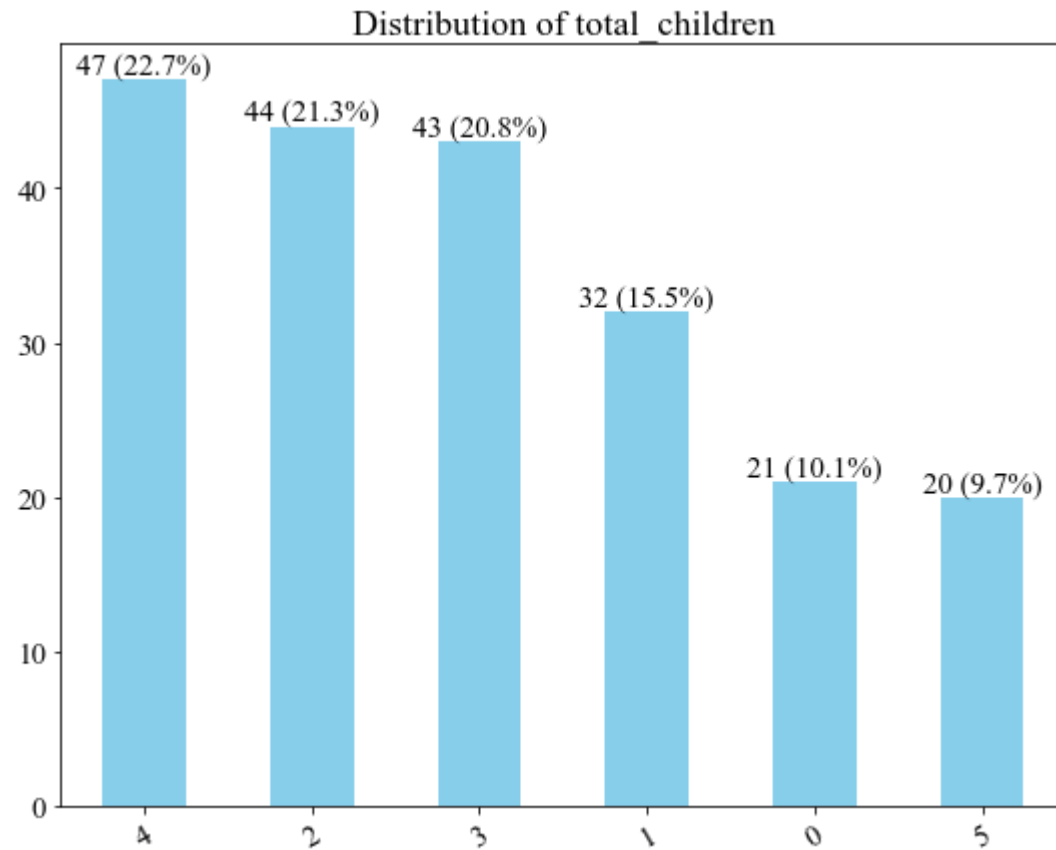
## HIGH VALUE CUSTOMER

## LOW VALUE CUSTOMER



## HIGH VALUE CUSTOMER

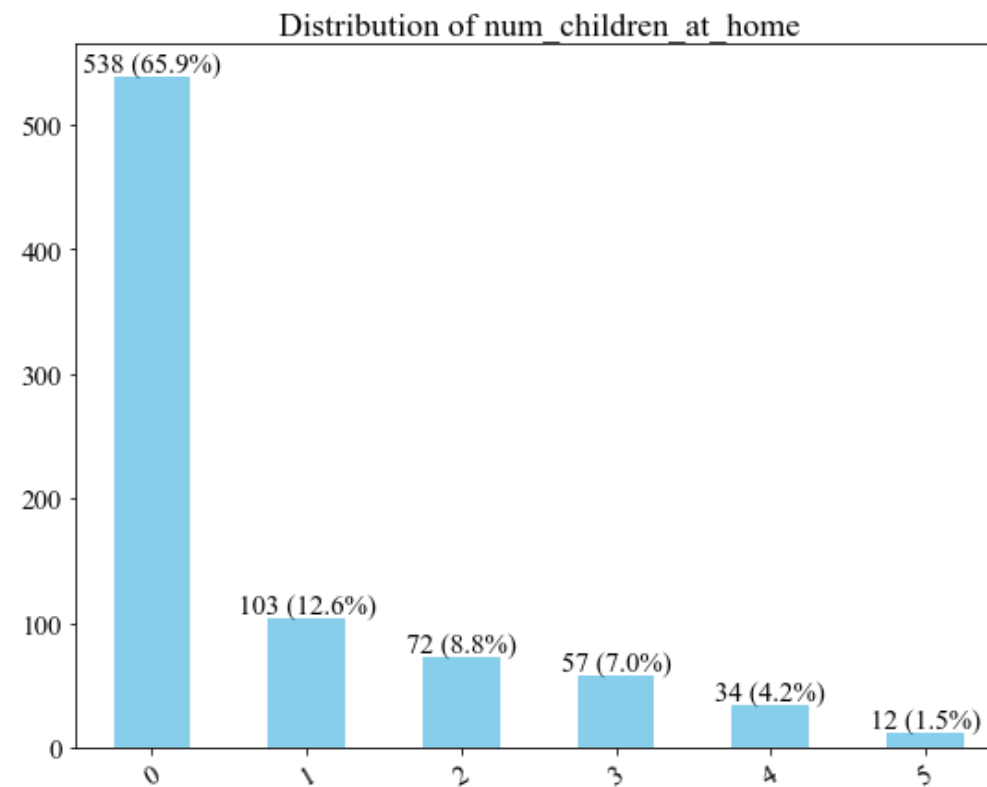
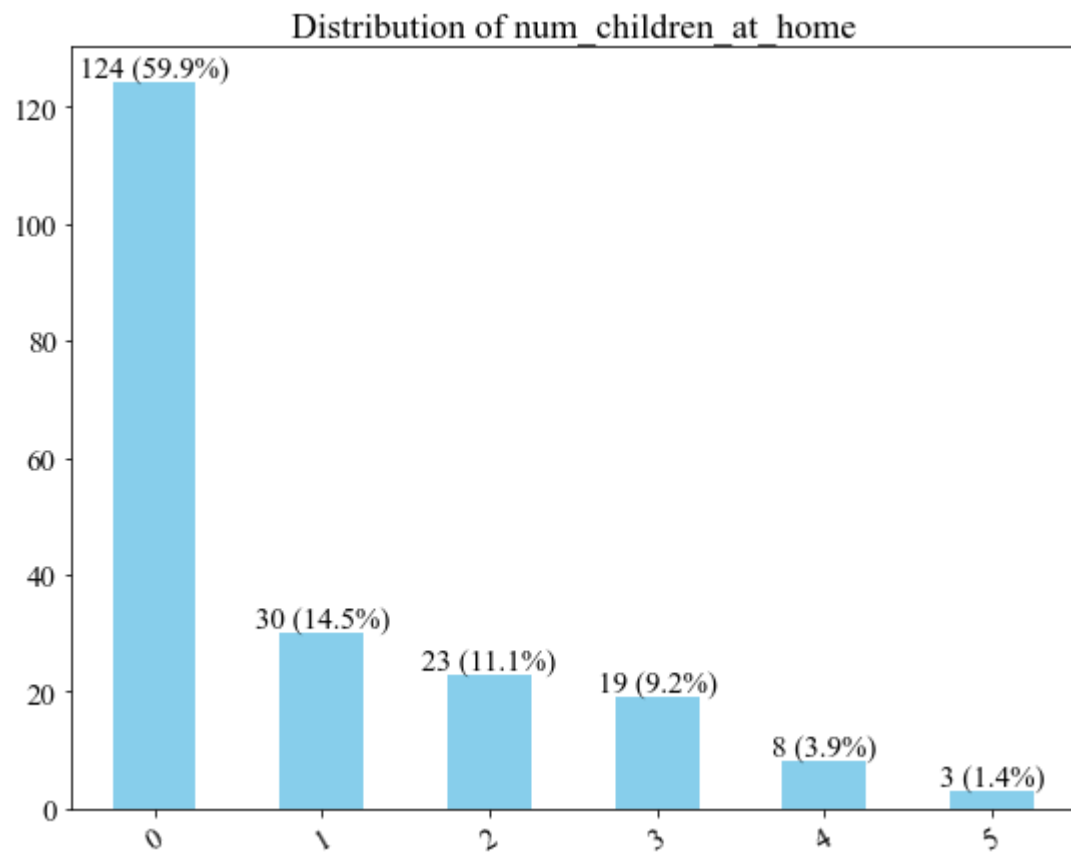
## LOW VALUE CUSTOMER



## HIGH VALUE CUSTOMER

|

## LOW VALUE CUSTOMER

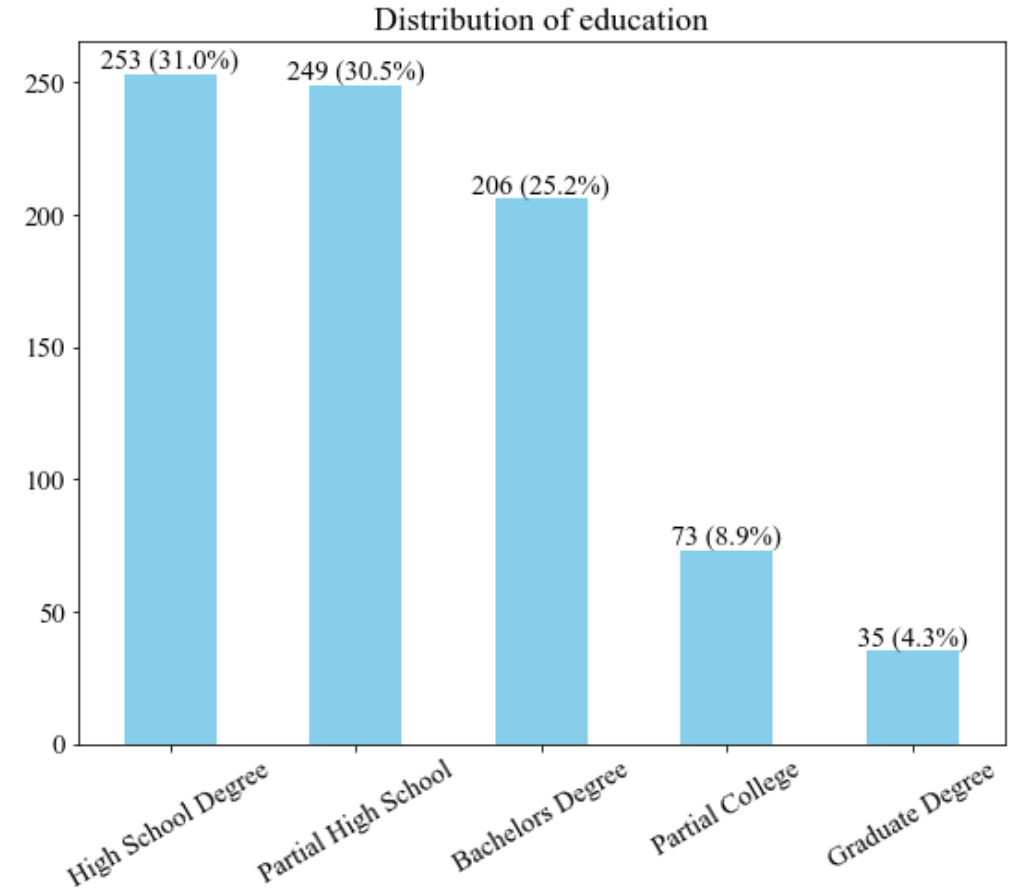
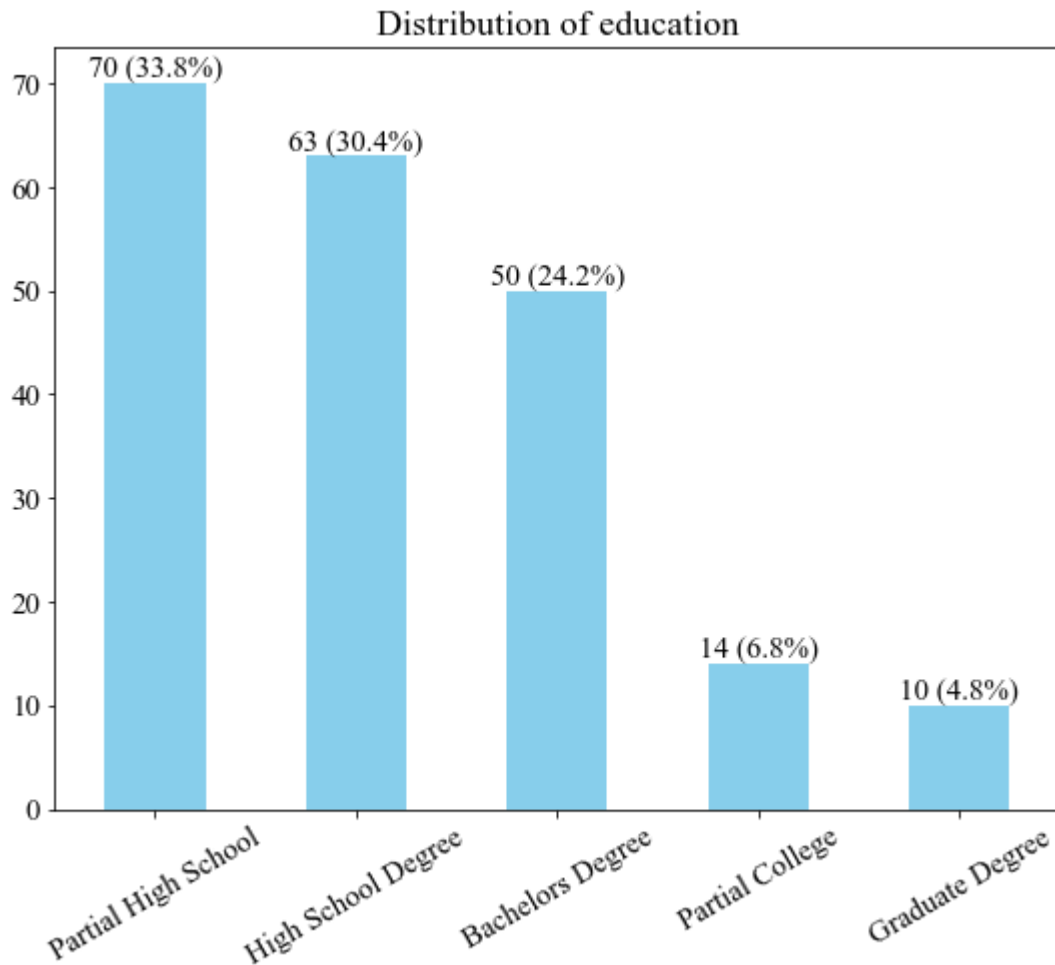




## HIGH VALUE CUSTOMER

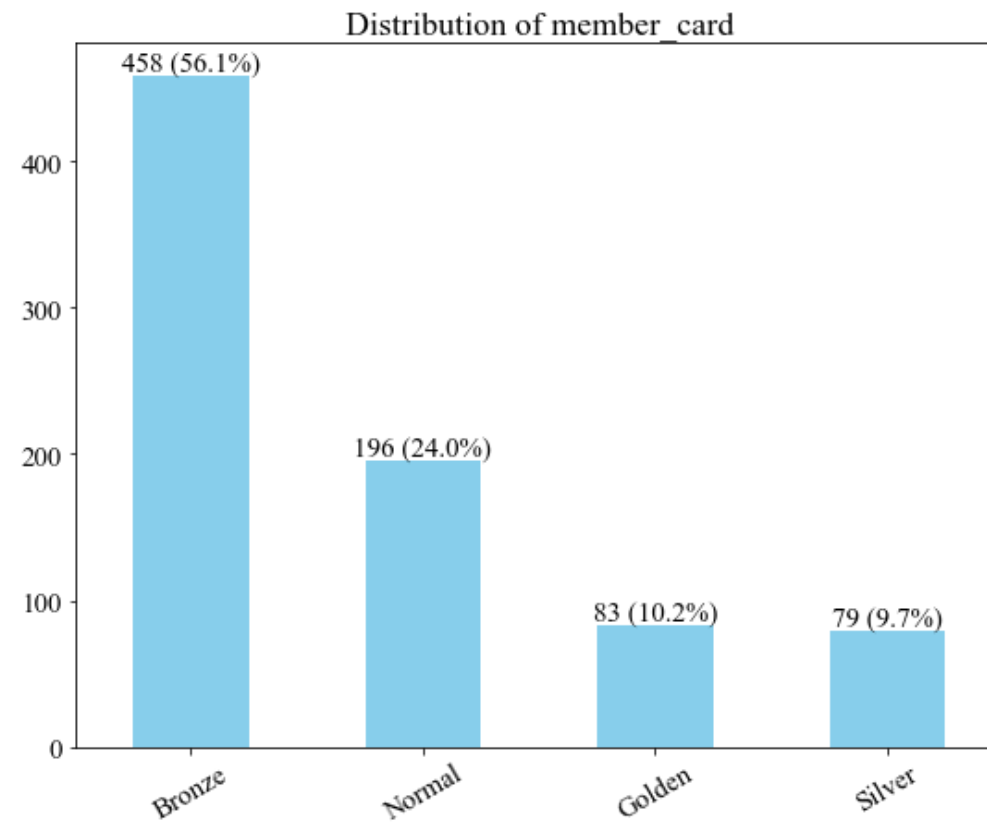
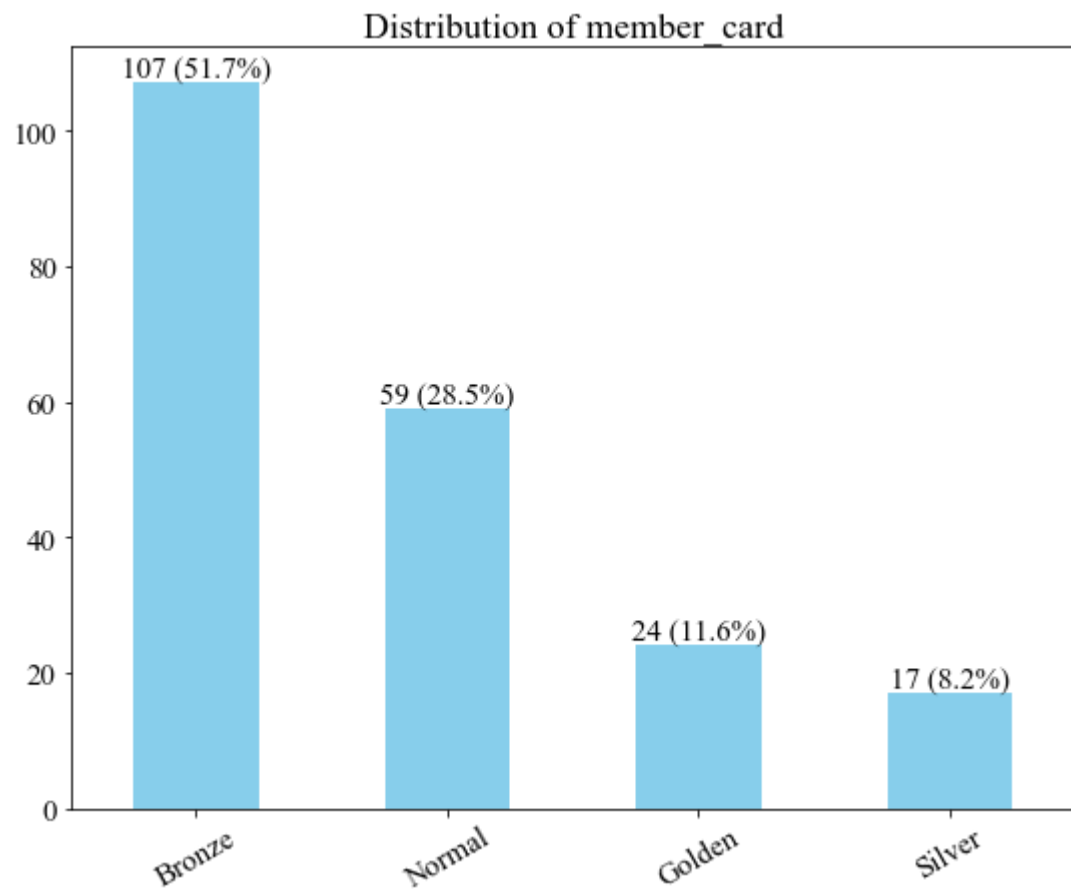


## LOW VALUE CUSTOMER



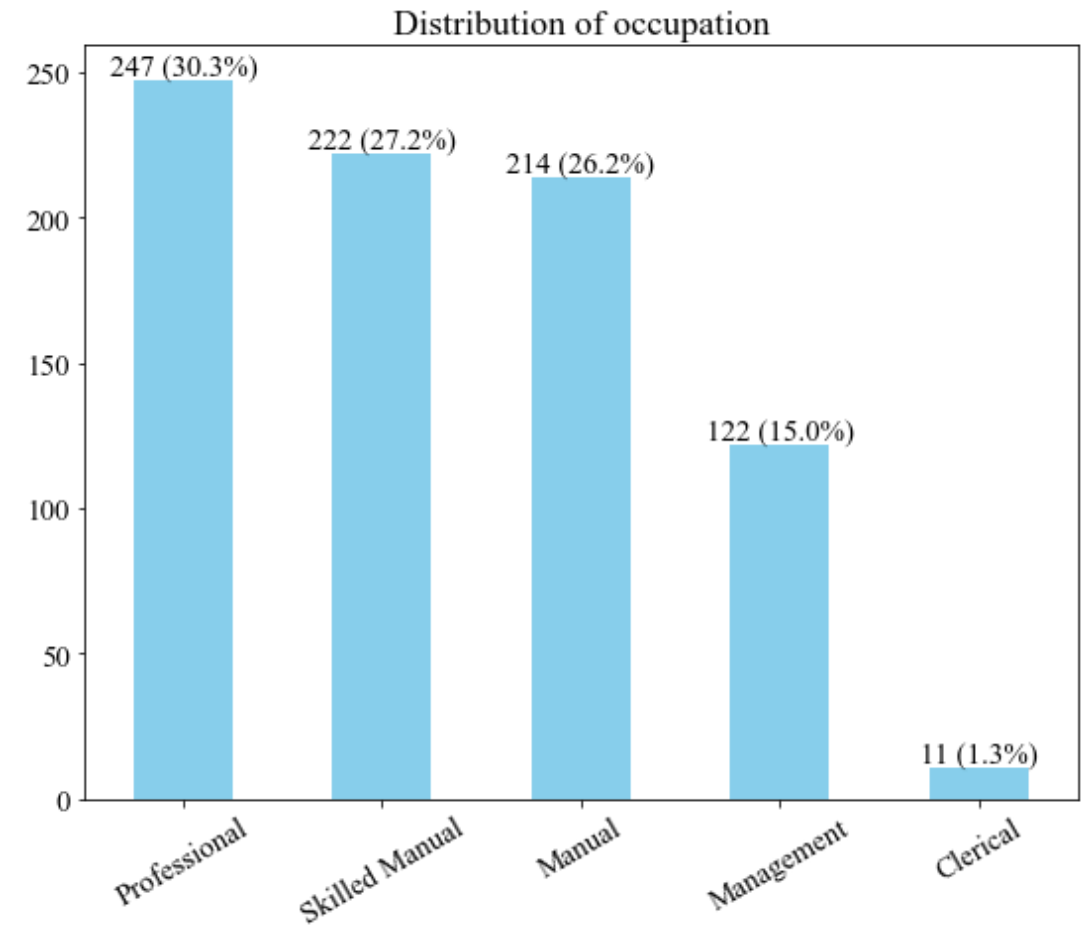
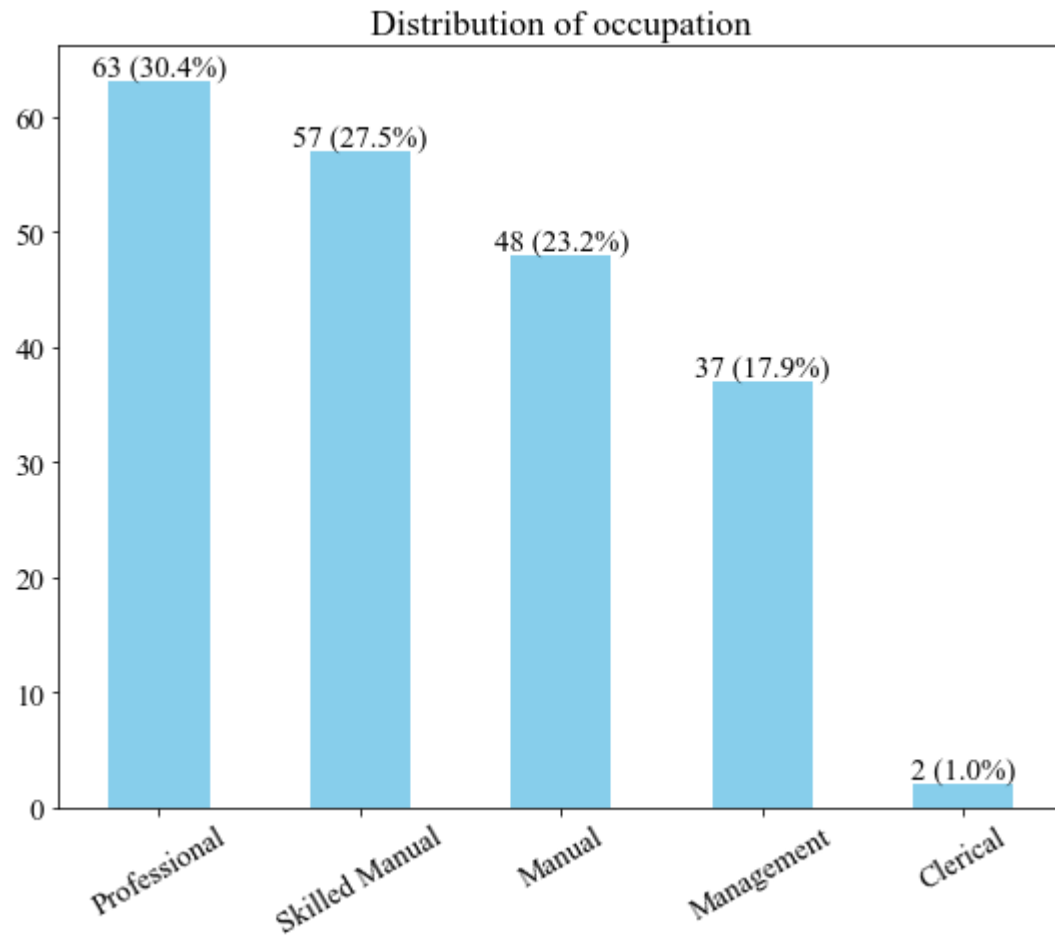
## HIGH VALUE CUSTOMER

## LOW VALUE CUSTOMER

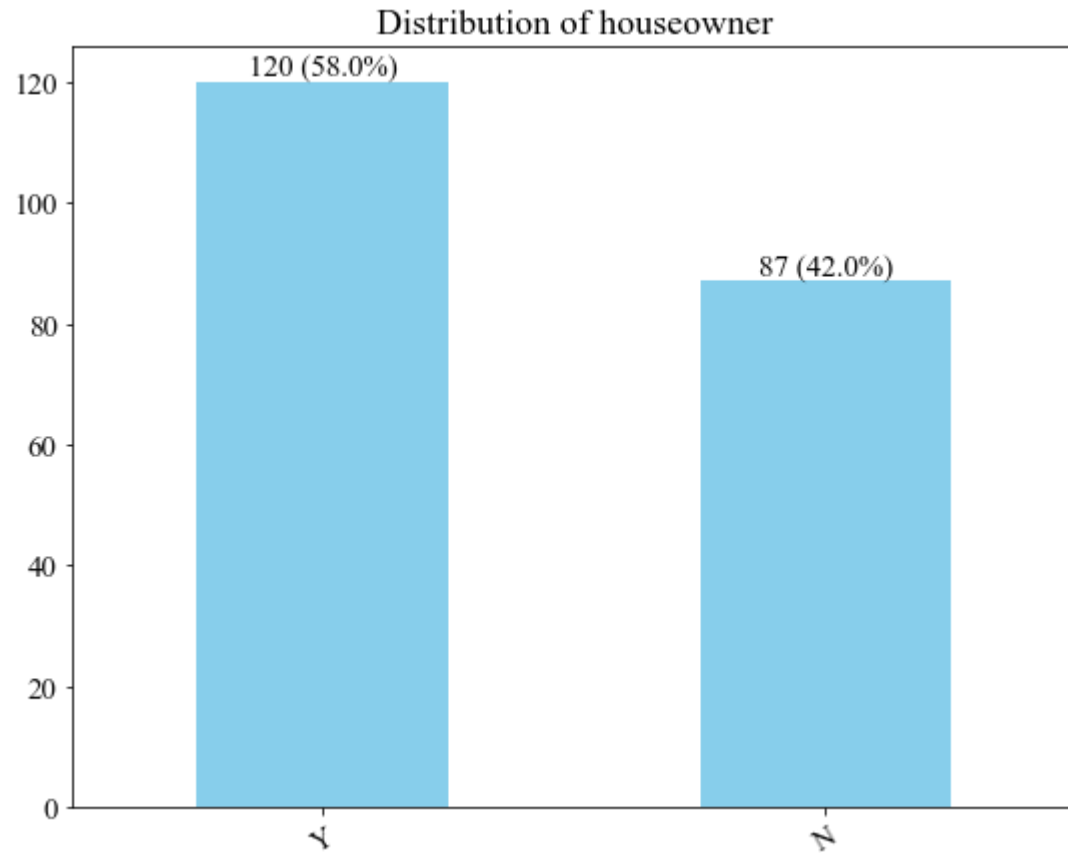


## HIGH VALUE CUSTOMER

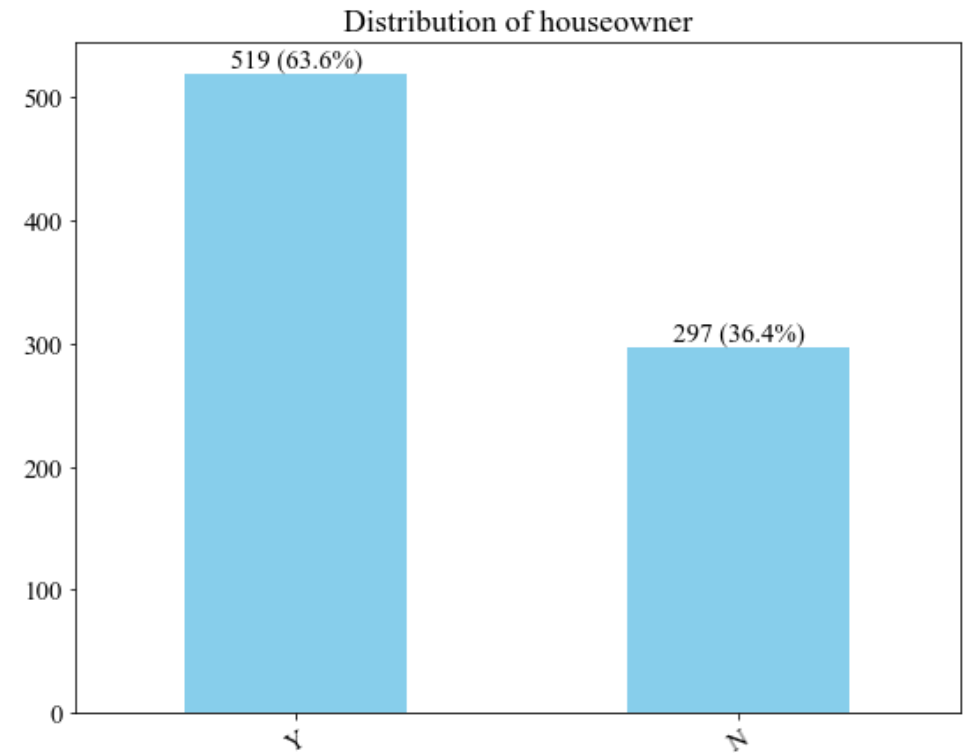
## LOW VALUE CUSTOMER



## HIGH VALUE CUSTOMER

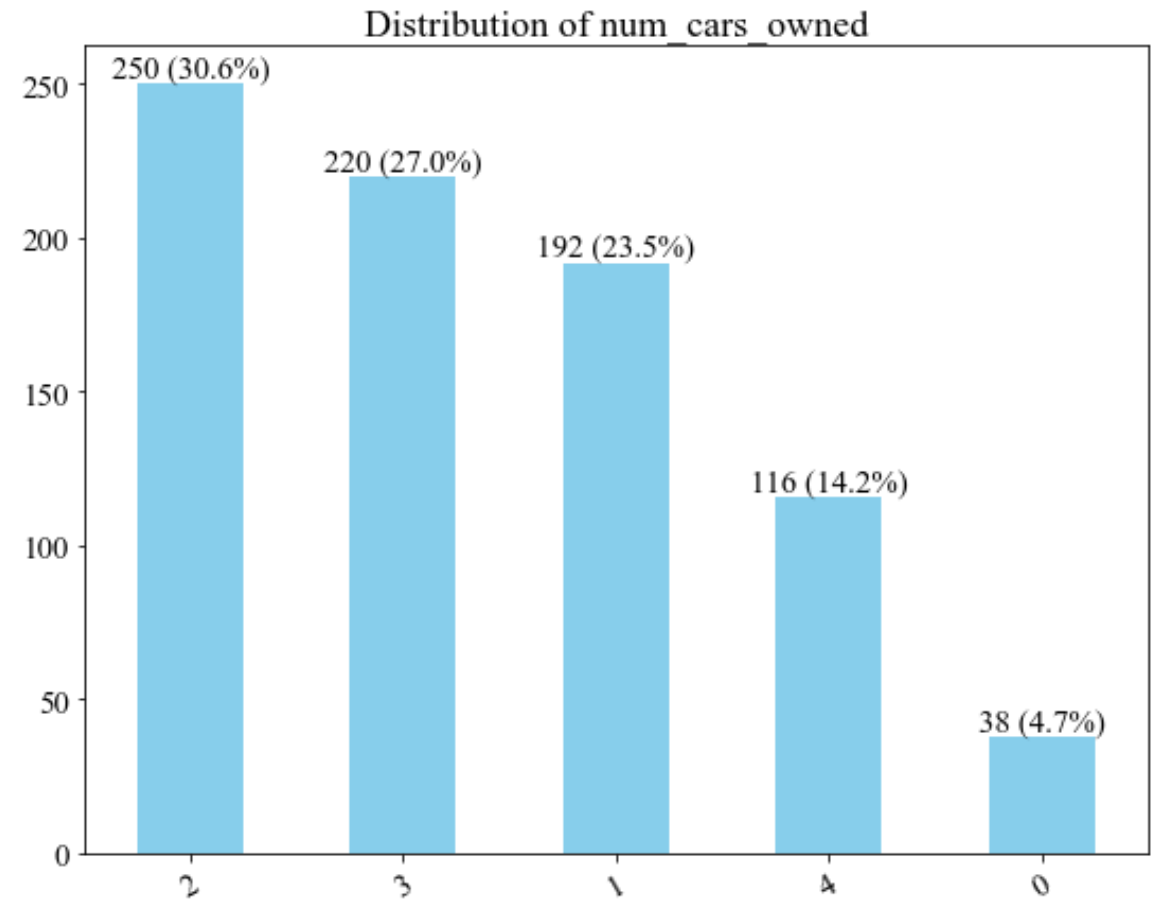
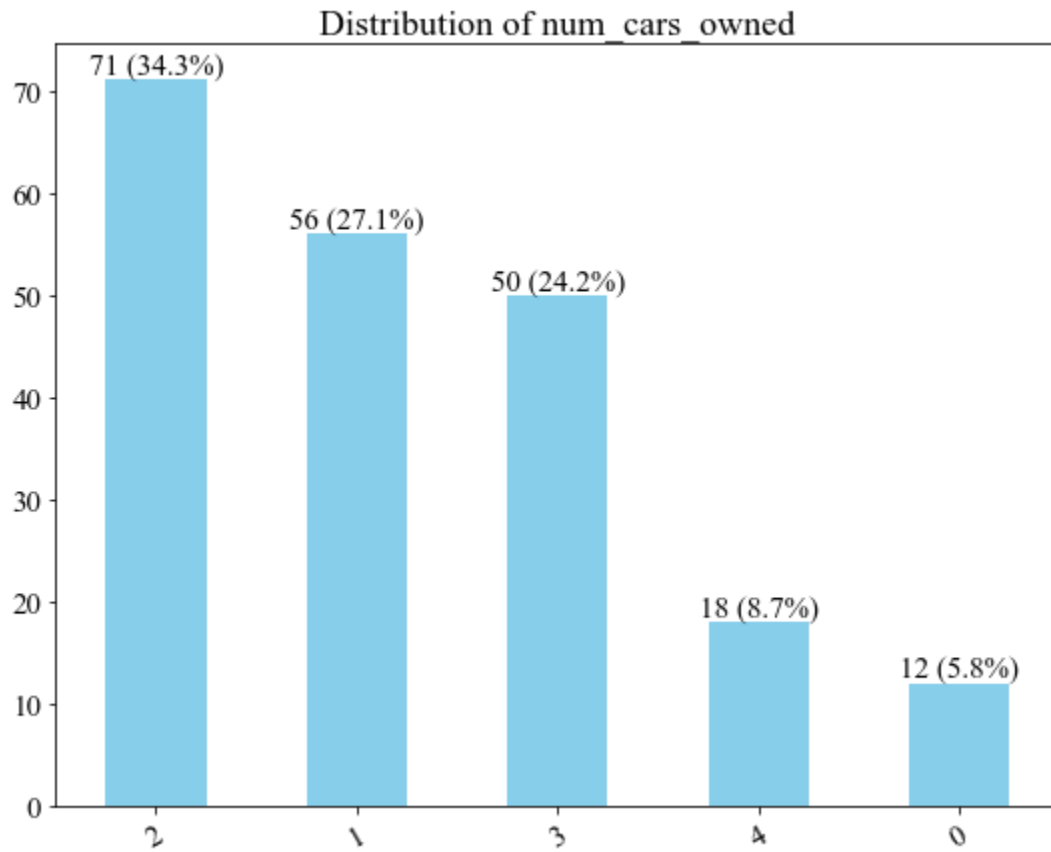


## LOW VALUE CUSTOMER



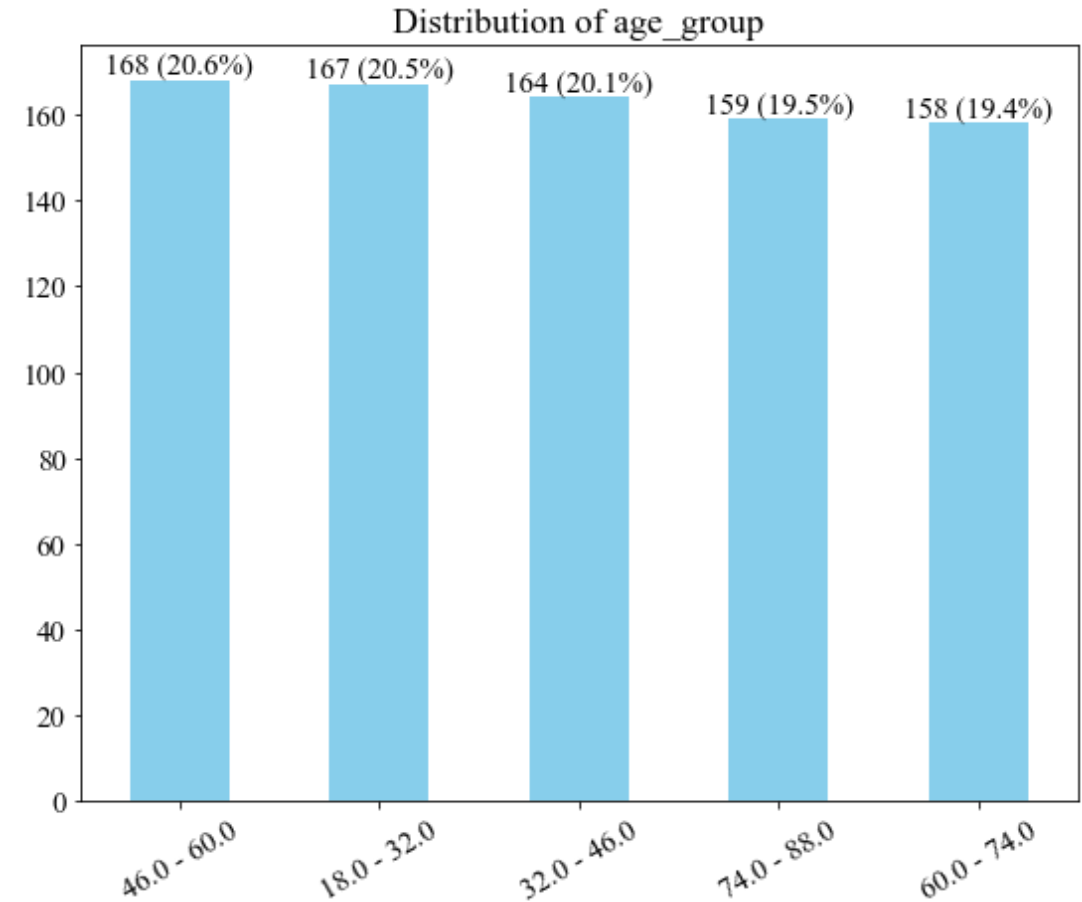
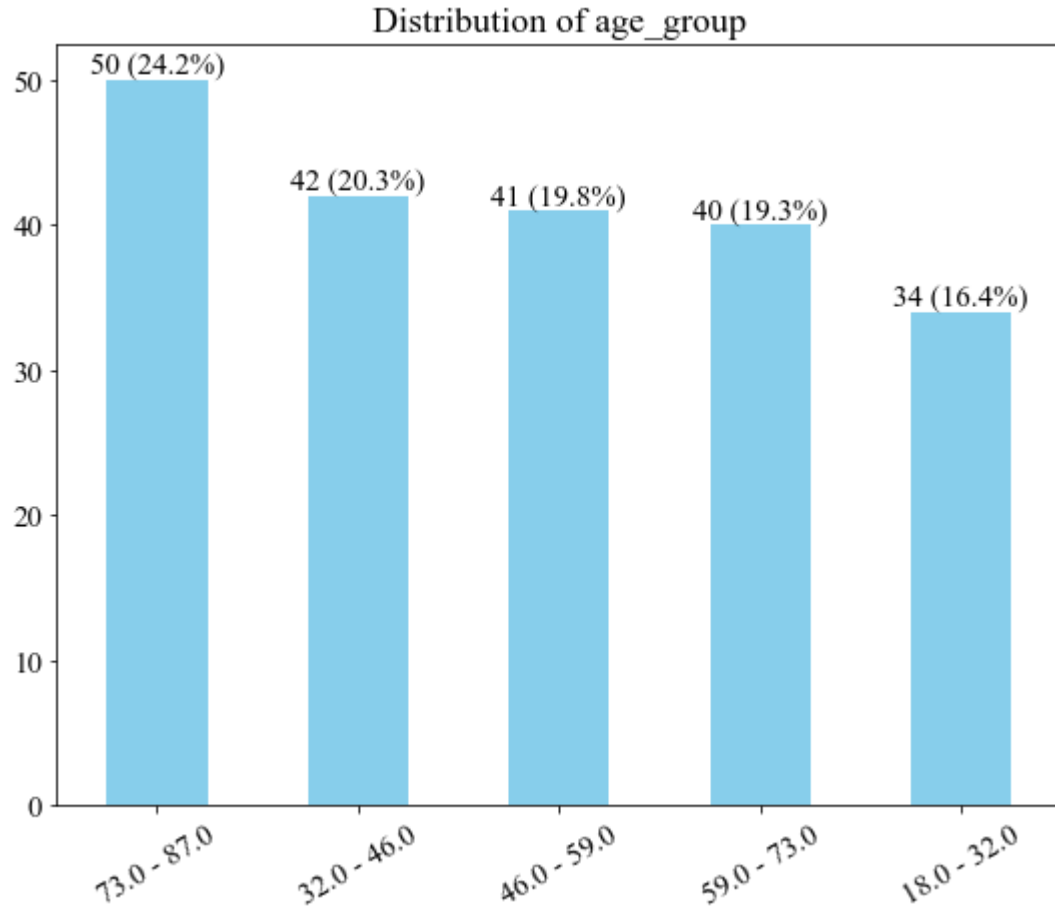
## HIGH VALUE CUSTOMER

## LOW VALUE CUSTOMER



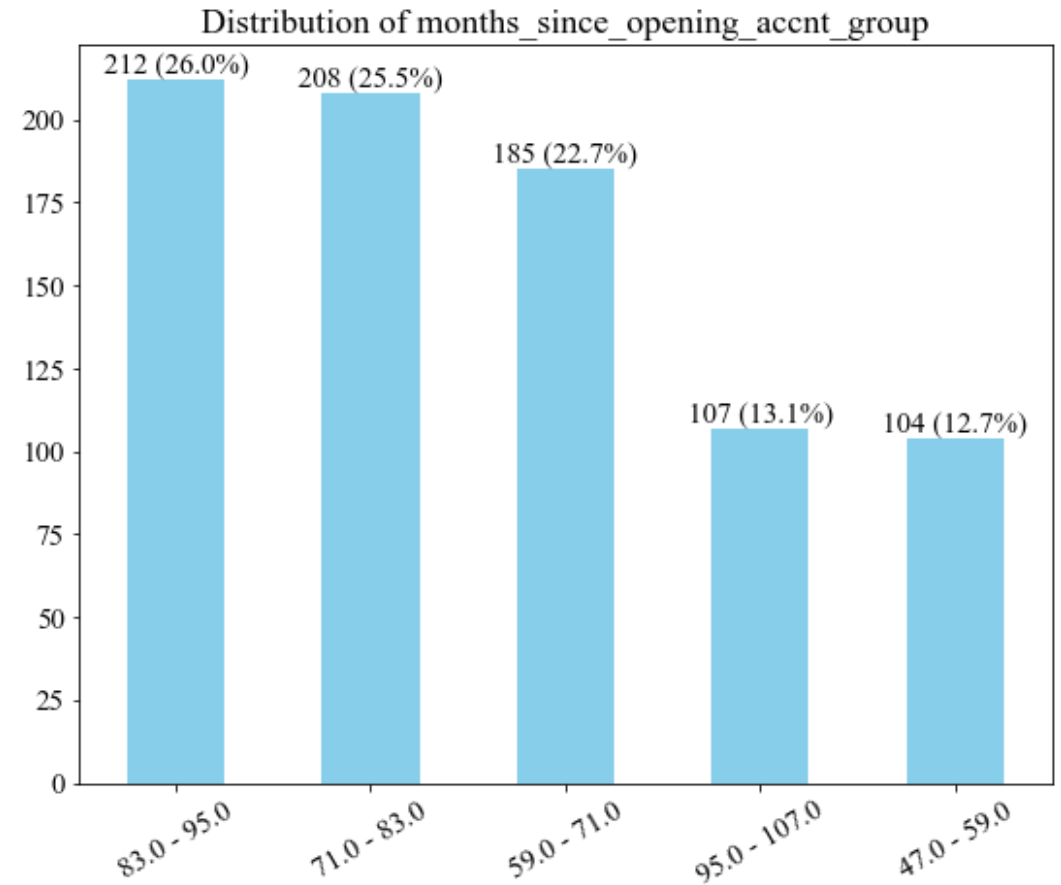
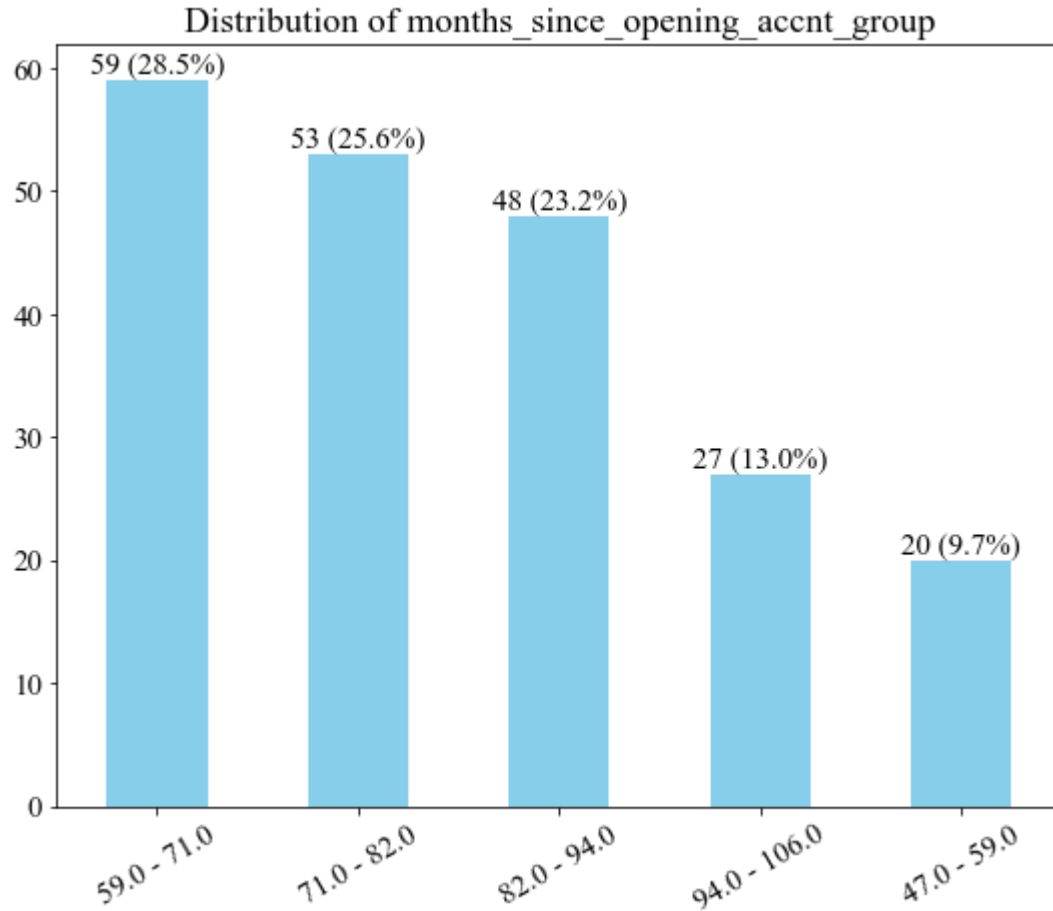
## HIGH VALUE CUSTOMER

## LOW VALUE CUSTOMER



## HIGH VALUE CUSTOMER

## LOW VALUE CUSTOMER



A series of white, thin, overlapping geometric lines on a black background, forming various polygons and intersecting points, primarily located on the left side of the slide.

THANK YOU