#### Dataset loaded successfully!

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999

Columns: 117 entries, CUSTOMERID to MDAYSHOPPER\_L4Y dtypes: float64(49), int64(65), object(3) memory

usage: 8.9+ MB

#### List Down all the columns with missing values bold text

#### Columns with missing values:

STATE	59	
TENURE IN MONTHS	66	
CLOSESTSTOREDISTANCE	1545	
AGE	5945	
INCOME	3259	
LENGTH OF RESIDENCE	2872	
NUMBER OF PERSONS IN LIVING UNIT	2872	
NUMBER OF ADULTS IN LIVING UNIT	2872	
MOSAIC	1868	
CAPE: AGE: POP: MEDIAN AGE	2872	
CAPE: AGE: POP: % 0-17	2872	
CAPE: AGE: POP: % 18-99+	2872	
CAPE: AGE: POP: % 65-99+	2872	
CAPE: ETHNIC: POP: % WHITE ONLY	2872	
CAPE: ETHNIC: POP: % BLACK ONLY	2872	
CAPE: ETHNIC: POP: % ASIAN ONLY	2872	
CAPE: ETHNIC: POP: % HISPANIC	2872	
CAPE: DENSITY: PERSONS PER HH FOR POP IN HH	2872	
CAPE: HHSIZE: HH: AVERAGE HOUSEHOLD SIZE	2872	
CAPE: TYP: HH: % MARRIED COUPLE FAMILY	2872	
CAPE: CHILD: HH: % WITH PERSONS LT18	2872	
CAPE: CHILD: HH: % MARR COUPLE FAMW- PERSONS LT18	2872	
CAPE: CHILD: HH: % MARR COUPLE FAMW-O PERSONS LT18	2872	
CAPE: LANG: HH: % SPANISH SPEAKING	2872 CAPE: EDUC:	
POP25+: MEDIAN EDUCATION ATTAINED 2872		
CAPE: HOMVAL: OOHU: MEDIAN HOME VALUE	2872	
CAPE: BUILT: HU: MEDIAN HOUSING UNIT AGE	2872	
CAPE: TENANCY: OCCHU: % OWNER OCCUPIED	2872	
CAPE: TENANCY: OCCHU: % RENTER OCCUPIED	2872	
CAPE: EDUC: ISPSA	2872	
CAPE: EDUC: ISPSA DECILE	2872	
CAPE: INC: FAMILY INC STATE DECILE	2872	
CAPE: INC: HH: MEDIAN FAMILY HOUSEHOLD INCOME	2872	
dtype: int64		

#### Categorize columns based on their data types

Numeric Columns: ['CUSTOMERID', 'LCPCOUNT', 'TENURE IN MONTHS', 'CLOSESTSTOREDISTANCE', 'FEMALE', 'AGE', 'HS DIPLOMA', 'SOME COLLEGE', 'BACH GRAD DEG', 'LT HS DIPLOMA', 'MARRIED', 'MNGMNT OFFICEADMIN', 'TECH PROF', 'SALES JOB', 'BLUE COLLAR', 'FARMER', 'RETIRED', 'SFDU', 'MFDU', 'HOMEOWNER', 'INCOME', 'MAIL RESP MULTI', 'MAIL RESP SINGLE', 'LENGTH OF RESIDENCE', 'NUMBER OF PERSONS IN LIVING UNIT', 'NUMBER OF ADULTS IN LIVING UNIT', 'METRO', 'URBAN', 'MOR BANK: UPSCALE MERCHANDISE BUYER', 'MOR BANK: MALE MERCHANDISE BUYER', 'MOR BANK: FEMALE MERCHANDISE BUYER', 'MOR BANK: CRAFTS-HOBBY MERCHANDISE BUYER', 'MOR BANK: GARDENING-FARMING BUYER', 'MOR BANK: BOOK BUYER', 'MOR BANK: COLLECT-SPECIAL FOODS BUYER', 'MOR BANK: GIFTS AND GADGETS BUYER', 'MOR BANK: GENERAL MERCHANDISE BUYER', 'MOR BANK: FAMILY AND GENERAL MAGAZINE', 'MOR BANK: FEMALE ORIENTED MAGAZINE', 'MOR BANK: MALE SPORTS MAGAZINE', 'MOR BANK: RELIGIOUS MAGAZINE', 'MOR BANK: GARDENING-FARMING MAGAZINE', 'MOR BANK: CULINARY INTERESTS MAGAZINE', 'MOR BANK: HEALTH AND FITNESS MAGAZINE', 'MOR BANK: DO-IT-YOURSELFERS', 'MOR BANK: NEWS AND FINANCIAL', 'MOR BANK: PHOTOGRAPHY', 'MOR BANK: OPPORTUNITY SEEKERS AND CE', 'MOR BANK: RELIGIOUS CONTRIBUTOR', 'MOR BANK: POLITICAL CONTRIBUTOR', 'MOR BANK: HEALTH AND INSTITUTION CONTRIBUTOR', 'MOR BANK: GENERAL CONTRIBUTOR', 'MOR BANK: MISCELLANEOUS', 'MOR BANK: ODDS AND ENDS', 'MOR BANK: DEDUPED CATEGORY HIT COUNT', 'MOR BANK: NON-DEDUPED CATEGORY HIT COUNT', 'MORTGAGE-HOME PURCHASE: HOME PURCHASE PRICE', 'CHILDREN', 'CAPE: AGE: POP: MEDIAN AGE', 'CAPE: AGE: POP: % 0-17', 'CAPE: AGE: POP: % 18-99+', 'CAPE: AGE: POP: % 65-99+', 'CAPE: ETHNIC: POP: % WHITE ONLY', 'CAPE: ETHNIC: POP: % BLACK ONLY', 'CAPE: ETHNIC: POP: % ASIAN ONLY', 'CAPE: ETHNIC: POP: % HISPANIC', 'CAPE: DENSITY: PERSONS PER HH FOR POP IN HH', 'CAPE: HHSIZE: HH: AVERAGE HOUSEHOLD SIZE', 'CAPE: TYP: HH: % MARRIED COUPLE FAMILY', 'CAPE: CHILD: HH: % WITH PERSONS LT18', 'CAPE: CHILD: HH: % MARR COUPLE FAMW- PERSONS LT18', 'CAPE: CHILD: HH: % MARR COUPLE FAMW-O PERSONS LT18', 'CAPE: LANG: HH: % SPANISH SPEAKING', 'CAPE: EDUC: POP25+: MEDIAN EDUCATION ATTAINED', 'CAPE: HOMVAL: OOHU: MEDIAN HOME VALUE', 'CAPE: BUILT: HU: MEDIAN HOUSING UNIT AGE', 'CAPE: TENANCY: OCCHU: % OWNER OCCUPIED', 'CAPE: TENANCY: OCCHU: % RENTER OCCUPIED', 'CAPE: EDUC: ISPSA', 'CAPE: EDUC: ISPSA DECILE', 'CAPE: INC: FAMILY INC STATE DECILE', 'CAPE: INC: HH: MEDIAN FAMILY HOUSEHOLD INCOME', 'FREQUENCY', 'QUANTITY', 'TOTALSALES', 'FREQUENCY 2Y', 'QUANTITY 2Y', 'TOTALSALES 2Y', '1-Engagement Spend', '2-Wedding Bands Spend', '3-Fashion Diamonds Spend', '4-Fashion Jewelry Spend', '5-Close Out Spend', '6-Promotional Items Spend', '7-Cost Only Spend', '8-Marketing Premium SKUs Spend', '9-Repairs & Appraisals Spend', '10-Pre Owned Spend', '11-Watches Spend', '12-Misc Merchandise Spend', '15-Store Events Spend', '16-Single Stone Jewelry Spend', 'MDAYREV L1Y', 'MDAYREV L2Y', 'MDAYREV L3Y', 'MDAYREV L4Y', 'MDAYQTY L1Y', 'MDAYQTY L2Y', 'MDAYQTY L3Y', 'MDAYQTY L4Y',

```
'MDAYSHOPPER_L1Y', 'MDAYSHOPPER_L2Y', 'MDAYSHOPPER_L3Y', 'MDAYSHOPPER_L4Y']
Categorical Columns: ['STATE', 'PRIVATELABELTENDERFLAG', 'MOSAIC']
```

List the columns with duplicate a)Remove them b)Print before and after

2-	tocat Dafama	Demos		· Calumna.	
Je				cate Columns:	MENUDE IN MONIMIC
	COSTOMERID	SIAIE	LCPCOUNT	PRIVATELABELTENDERFLAG	TENURE_IN_MONTHS
d	5001	TX	1	N	-9.0
	_				
\					
0					
1	5002	ОН	0	Y	9.0

2	5003	3 TX	ζ	0			N			12.0	
3	5004	1 TN	J	0			N			-1.0	Δ
500		X	0	O			N		16.0	1.0	1
	CLOSES	TSTORE	DISTANCE	FEMALI	_	HS_	DIPLOMA	SOME	COLLEGE		\
0			NaN		0 NaN		0	^	0	• • •	
1				28943	0	Nal	N	0		0	
2			 NaN	(	0 NaN		0		0		
3			NaN	(	0 NaN		0		0		
4			NaN		0 NaN	- 1	0	- 0	0		
0	MDAYRE	0.0	MDAYREV_	_L4Y MI 0.0	DAYQTY_	_L1Y 0	MDAYQTY	_L2Y 0	MDAYQTY_	_L3Y	
1		0.0		0.0		0		0		0	
2		0.0		0.0		0		0		0	
3		0.0		0.0		0		0		0	
4	MDAMOR	0.0	MDANCHOL	0.0		0		0		0	\
0	MDAYQT	1_L41 0	MDAISHOI	_	r MDA. O	1 SHOI	PPER_L2Y 0	MDAY	SHOPPER_1	0	\
1		0			)		0			0	
2		0		(	0		0			0	
3		0			0		0			0	
4	MDAYSH	ODDFD	ΤΛV	(	)		0			0	
0	MDAISH	OFFER_	0								
1			0								
2			0								
3			0								
-	rows x	117 с	olumns]								
[ 0	20		0 1 0 ]								
72	taget	After 1	Removing	Duplica	te Col	limne	, <b>.</b>				
Ja											
	CUSTO	MERID :	STATE LO	CPCOUNT	PRIVAT	CELAE	BELTENDERI	FLAG	TENURE_I	N_MOI	NTHS
		5001	TX	1				N			-9.0
Dup	olicate	Colum	ns Ident:	riled:							
[ ]											
\											
0											
J											
1		5002	ОН	0				Y			9.0
2		5003	TX	0				N		1	2.0
3		5004	TN	0				N		-	1.0

4	5005	TX	0				N		1	6.0
	CLOSESTSTORE	DISTANCE	FEMALE	AGE	HS_	DIPLOMA	SOME	COLLEGE		\
0		NaN	0	NaN		0		0		
1		8.72	8943	0	NaN	•	0		0	
		• • •								
2		NaN	0	_		0		0		
3		NaN	0	NaN		0		0		
4		NaN	0	NaN		0		0		
	MDAYREV_L3Y	MDAYREV_	•	YQTY_	L1Y	MDAYQTY_	_L2Y	MDAYQTY_	-	/
0	0.0		0.0		0		0		0	
1	0.0		0.0		0		0		0	
2	0.0		0.0		0		0		0	
3	0.0		0.0		0		0		0	
4	0.0		0.0		0		0		0	,
_		MDAYSHOP	_	MDA	ZSHOP	PER_L2Y	MDAY	SHOPPER_I		/
0	0		0			0			0	
1	0		0			0			0	
2	0		0			0			0	
	0		0			0			0	
4	U	T // 37	0			0			0	
0	MDAYSHOPPER_	0								
1										
2		0								
2 3		0								
ა 4		0								
<u>-</u> [5	rows x 117 c	0								

List the constants columns a)Remove them b)Print before and after

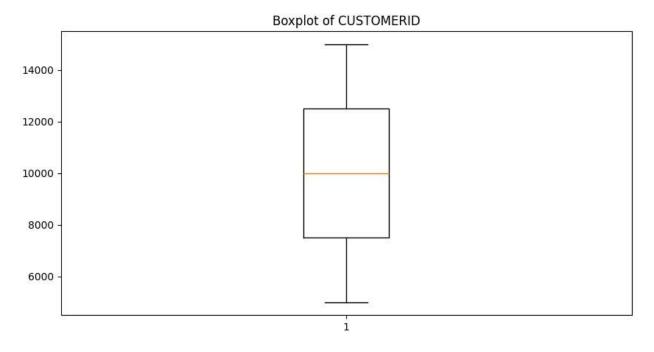
```
Dataset Before Removing constant Columns:
  CUSTOMERID STATE LCPCOUNT PRIVATELABELTENDERFLAG TENURE IN MONTHS
                                                             -9.0
        5001 TX 1
0
        5002 OH
                                                              9.0
        5003
               TX
                                                             12.0
        5004
               TN
                                                             -1.0
        5005 TX
                                               Ν
                                                             16.0
  CLOSESTSTOREDISTANCE FEMALE AGE HS DIPLOMA
                                              SOME COLLEGE
0
                  NaN 0 NaN
1
                  8.728943
                              0 NaN
                                                             0
2
                           0 NaN
                                           0
                                                         0
                  NaN
3
                          0 NaN
                                           0
                  NaN
                  NaN
                          0 NaN
                                           0
  MDAYREV L3Y MDAYREV L4Y MDAYQTY L1Y MDAYQTY L2Y MDAYQTY L3Y \
0
          0.0
                      0.0
                                    0
                      0.0
                                    0
1
          0.0
                                                            0
2
                      0.0
                                    0
                                                            0
          0.0
3
          0.0
                      0.0
                      0.0
                                    0
          0.0
  MDAYQTY L4Y MDAYSHOPPER L1Y MDAYSHOPPER L2Y MDAYSHOPPER L3Y \
0
           0
                            0
                                           0
1
            0
                            0
                                           0
                                                           0
2
            0
                            0
                                            0
                                                           0
3
            0
                            0
            0
  MDAYSHOPPER L4Y
0
1
                0
2
                0
3
[5 rows x 117 columns]
Constant Columns: ['MARRIED', 'CHILDREN', '7-Cost Only Spend']
```

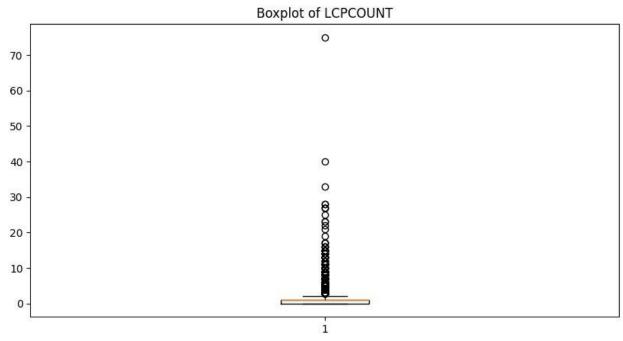
Dataset After Removing constants Columns:

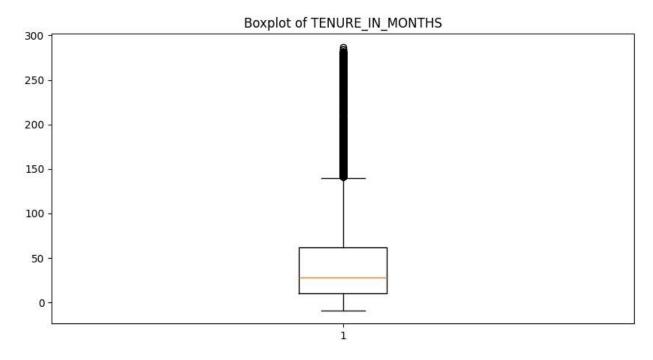
CUSTOMERID STATE LCPCOUNT PRIVATELABELTENDERFLAG TENURE\_IN\_MONTHS

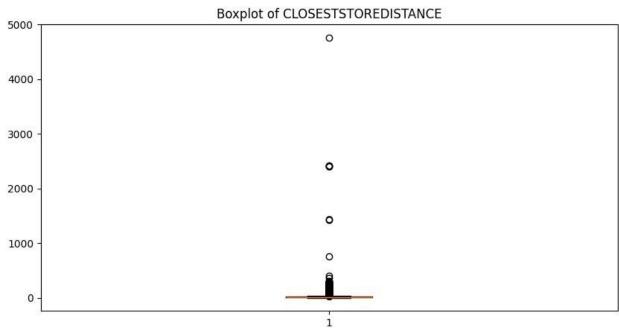
5001 TX 1 N -9.0

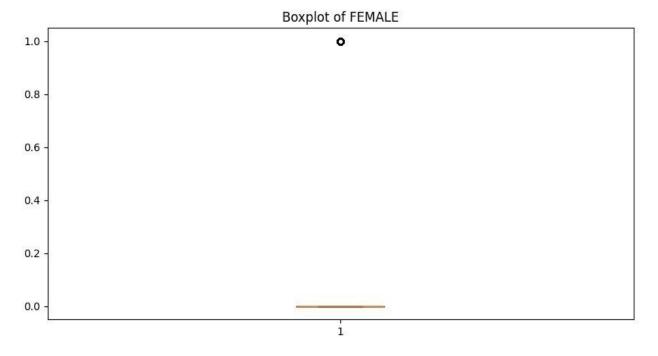
1	500	2 ОН	0				Y		9.0
2	5003	TX	0			N		12	. 0
	500A		•					4	0 4
3	5004	TN	0			N	1		.0 4
500	)5 TX CLOSESTSTC	0	FEMALE	AGE	пст	N DIPLOMA	SOME COI	16.0	\
0	CLOSESISIC	NaNCE NaN		NaN	по_1	O 0	SOME_COI	. O	• • \
1			28943	0	NaN	· ·	0	•	0
		•••					•		
2		NaN	0 1	NaN		0		0.	
3		NaN	0 1	NaN		0		0.	• •
4		NaN		NaN		0		0.	• •
	MDAYREV_L3		_	YTQYY_	-	MDAYQTY	<del>_</del>	AYQTY_L3	
0	0.		0.0		0		0		0
1 2	0.		0.0		0		0		0
3	0.		0.0		0		0		0
4	0.		0.0		0		0		0
	MDAYQTY L4		PPER L1Y	MDAY	SHOPE	PER L2Y	MDAYSHOR	PPER L3Y	. \
0	_	0	_ 0			_ 0		_ 0	
1		0	0			0		0	
2		0	0			0		0	
3		0	0			0		0	
4	MDAVCHODDE	0	0			0		0	
0	MDAYSHOPPE	0 0							
1		0							
2		0							
3		0							
4		0							
[5	rows x 117	columns]							

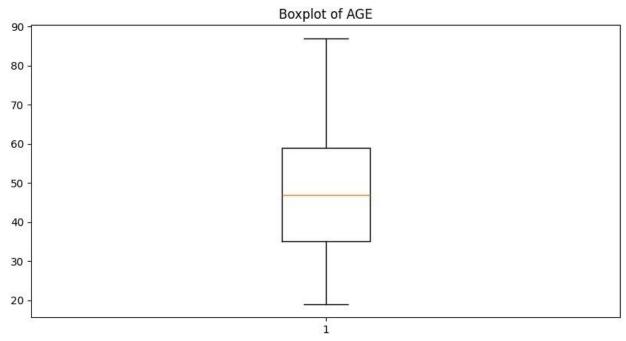


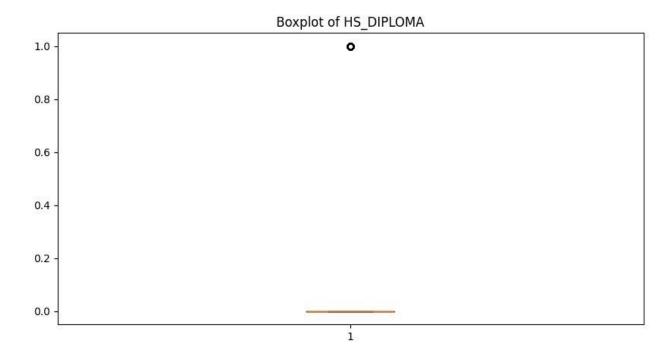


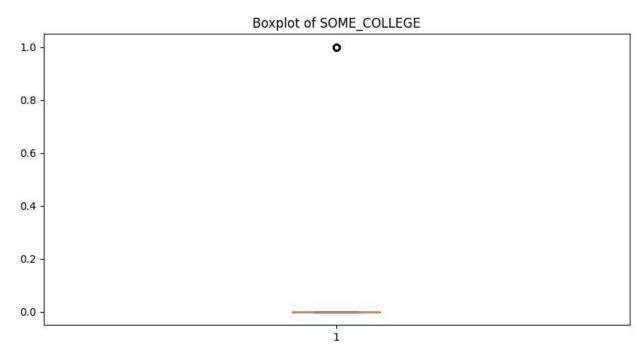


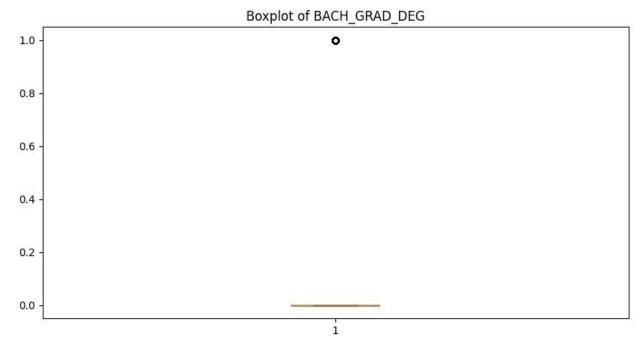


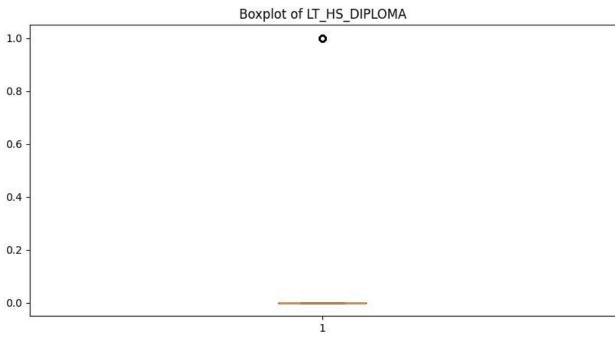


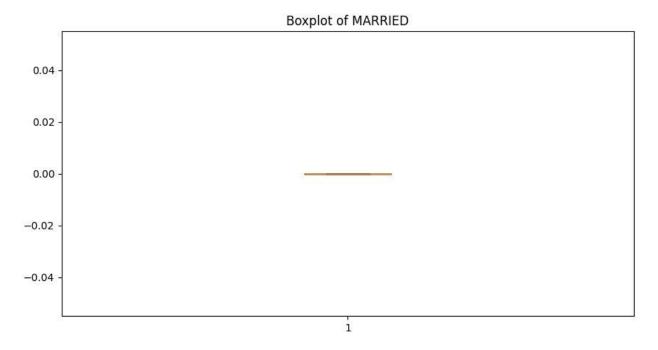


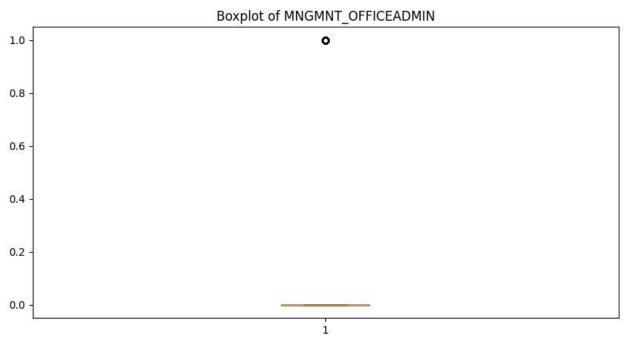


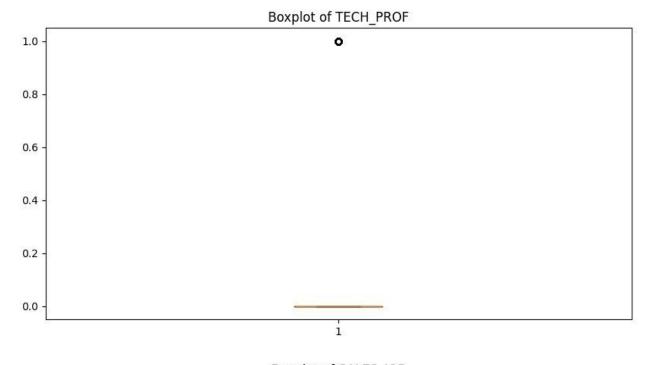


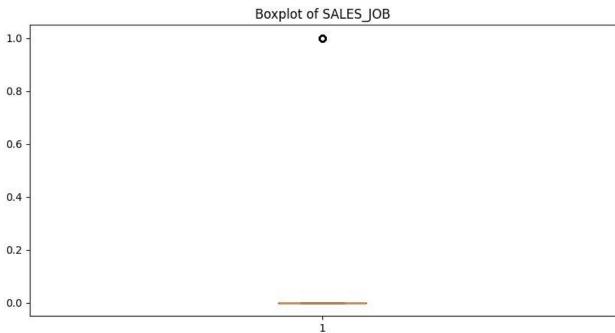


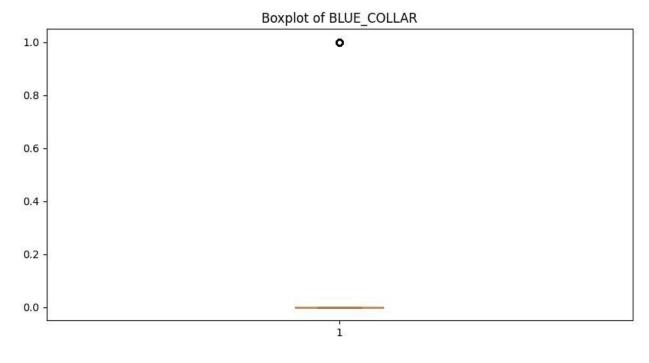


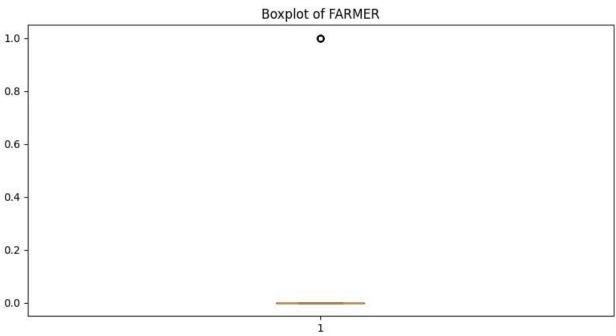




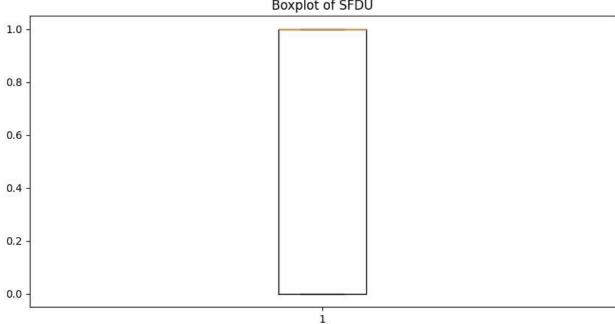


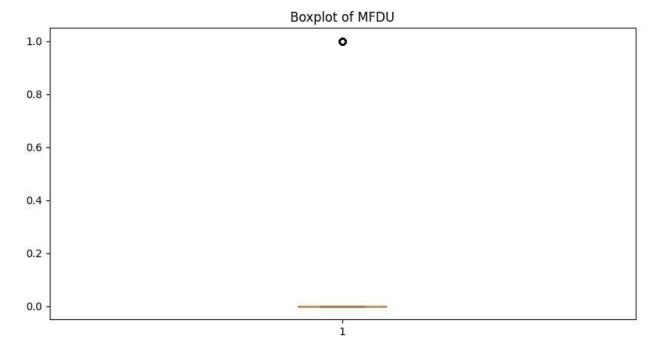


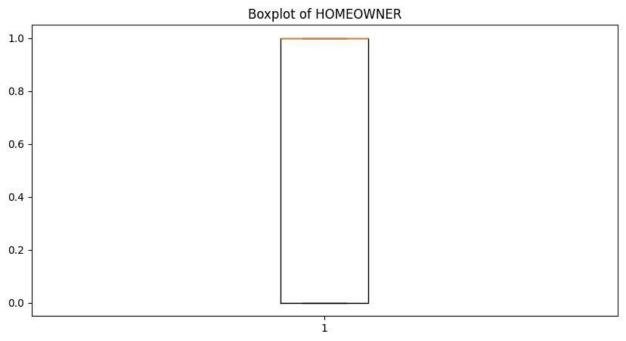


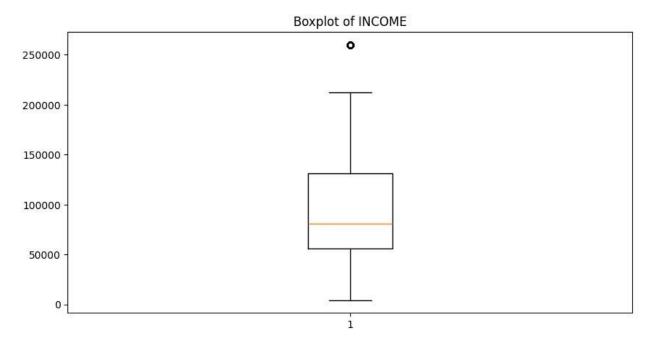


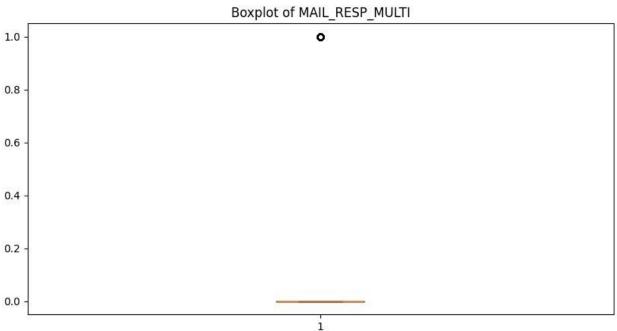




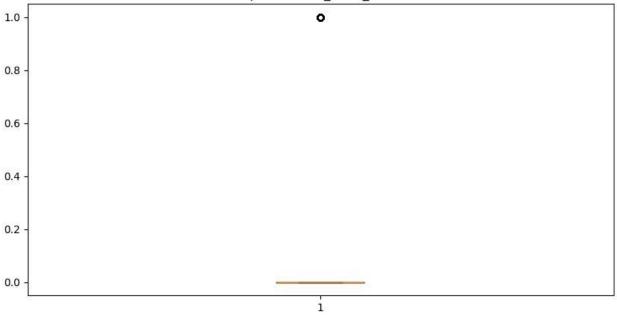




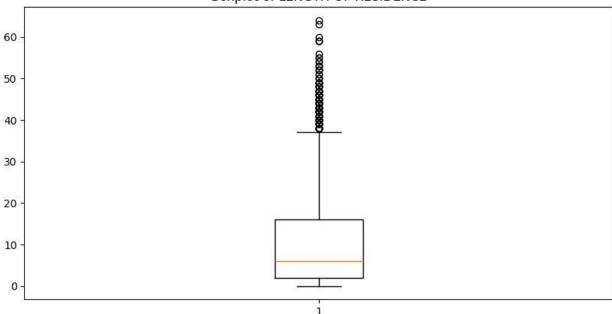




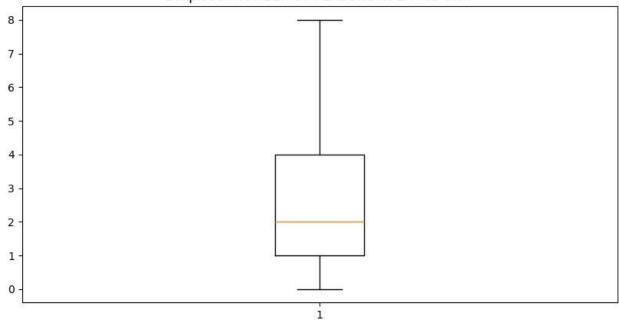




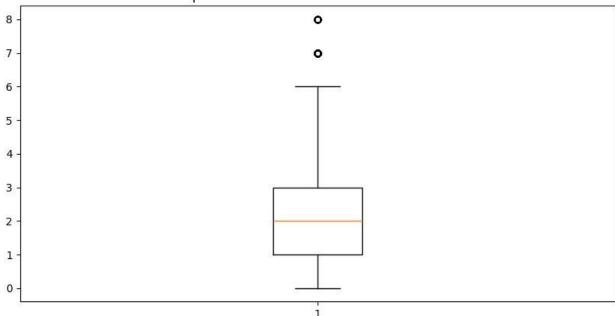
## Boxplot of LENGTH OF RESIDENCE

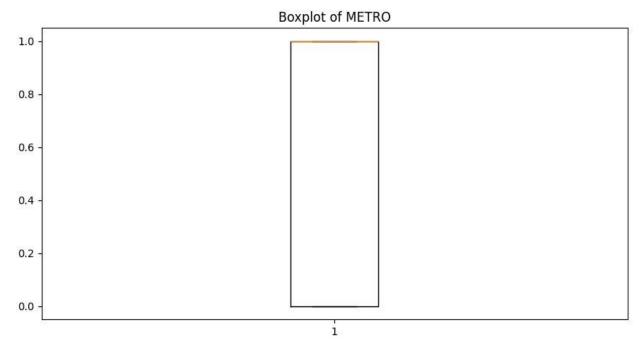


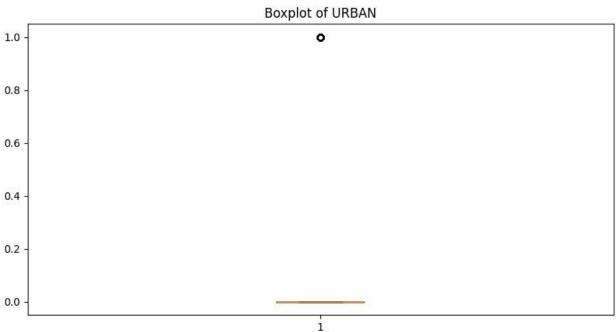
#### Boxplot of NUMBER OF PERSONS IN LIVING UNIT



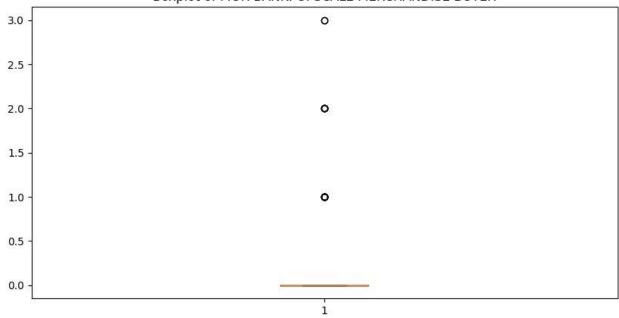
## Boxplot of NUMBER OF ADULTS IN LIVING UNIT



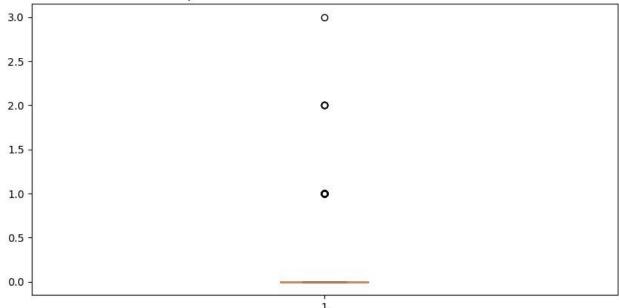




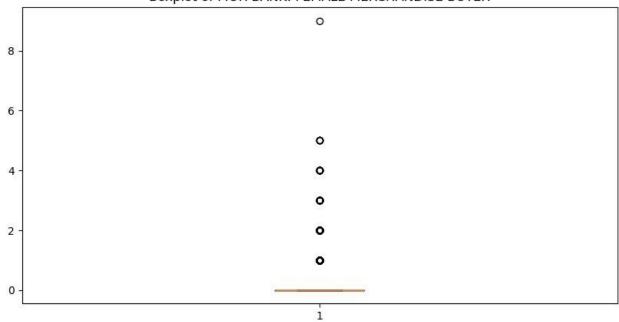




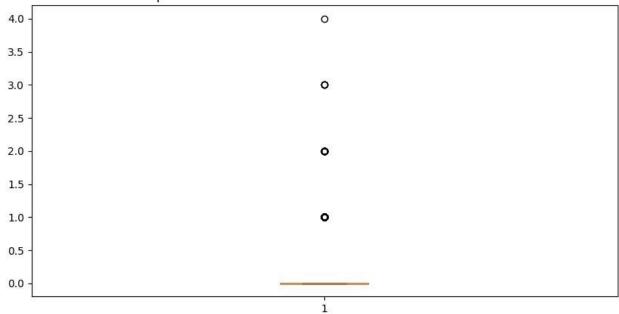
# Boxplot of MOR BANK: MALE MERCHANDISE BUYER



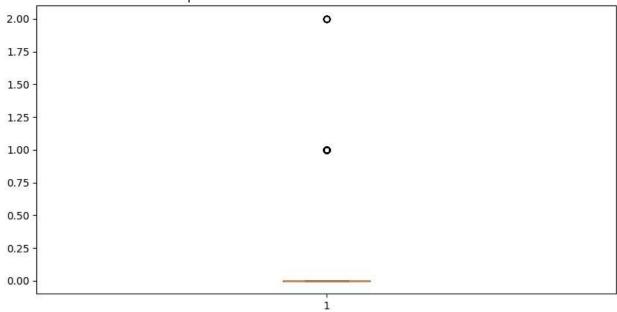
#### Boxplot of MOR BANK: FEMALE MERCHANDISE BUYER



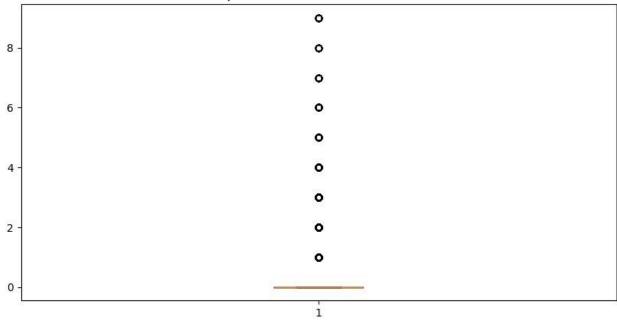
#### Boxplot of MOR BANK: CRAFTS-HOBBY MERCHANDISE BUYER



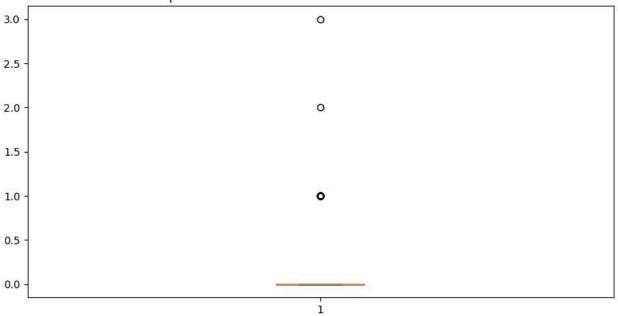




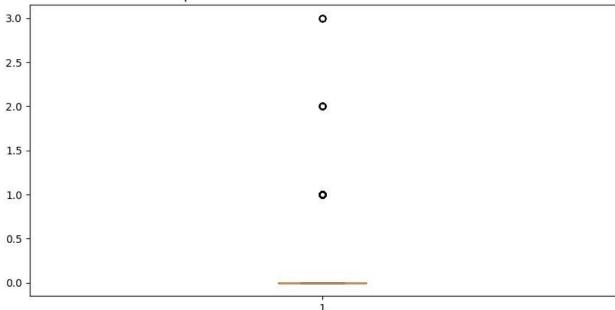
#### Boxplot of MOR BANK: BOOK BUYER



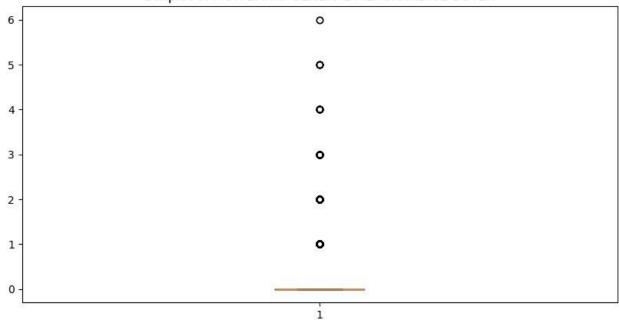




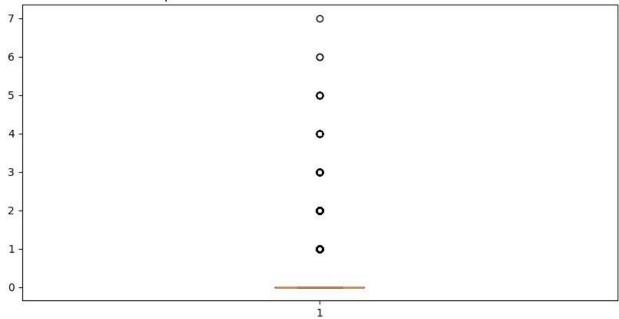
#### Boxplot of MOR BANK: GIFTS AND GADGETS BUYER



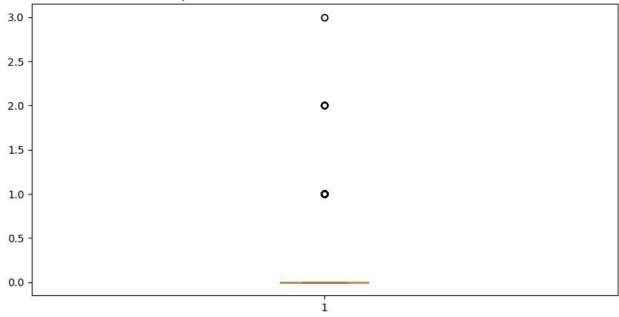




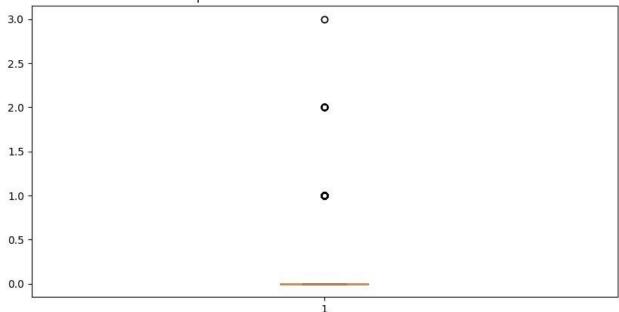
#### Boxplot of MOR BANK: FAMILY AND GENERAL MAGAZINE



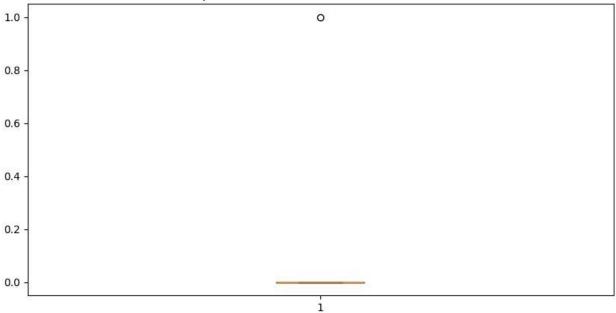




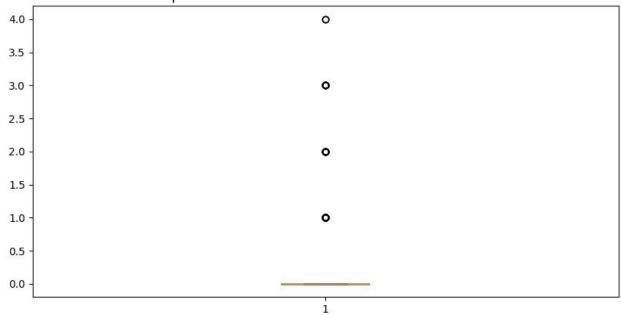
#### Boxplot of MOR BANK: MALE SPORTS MAGAZINE



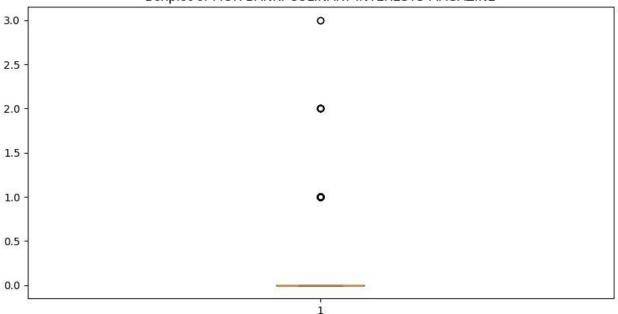




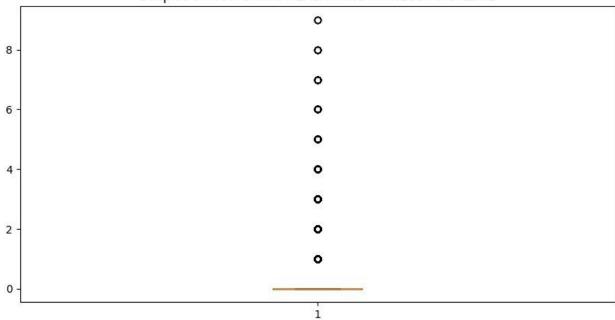
#### Boxplot of MOR BANK: GARDENING-FARMING MAGAZINE



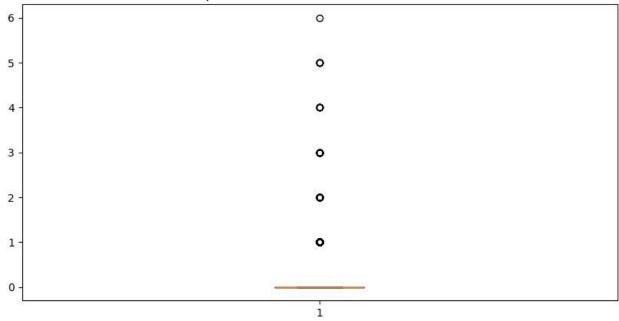




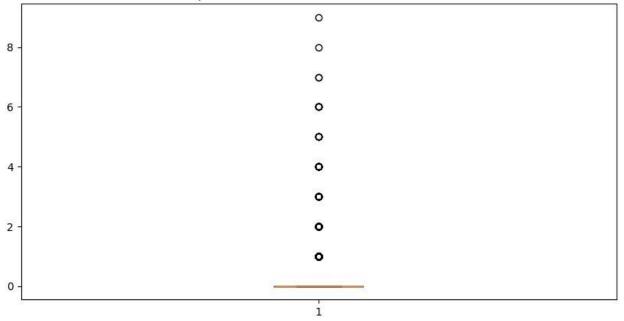
#### Boxplot of MOR BANK: HEALTH AND FITNESS MAGAZINE



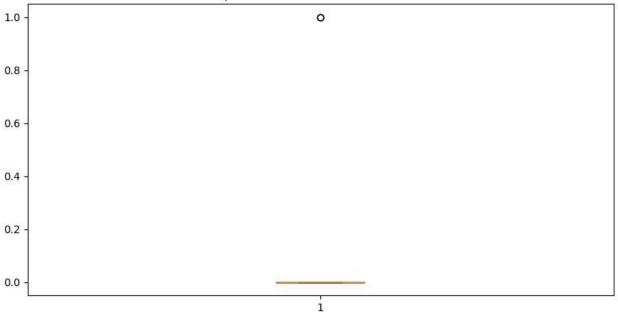




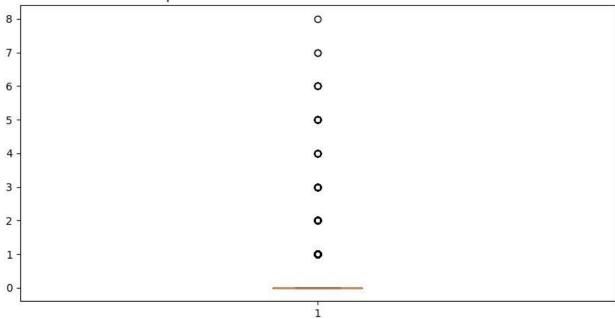
#### Boxplot of MOR BANK: NEWS AND FINANCIAL



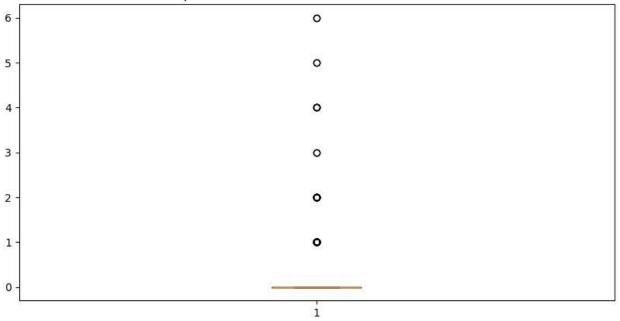




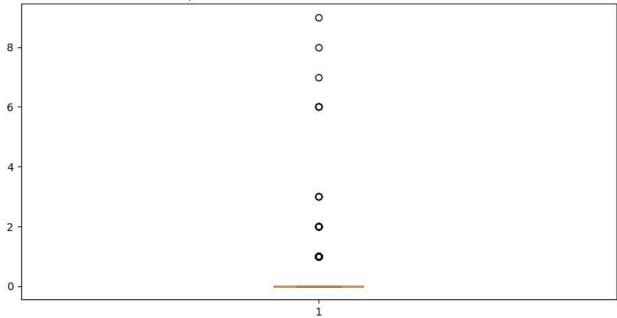
#### Boxplot of MOR BANK: OPPORTUNITY SEEKERS AND CE



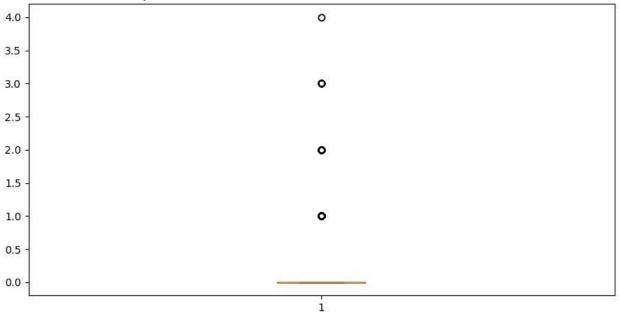




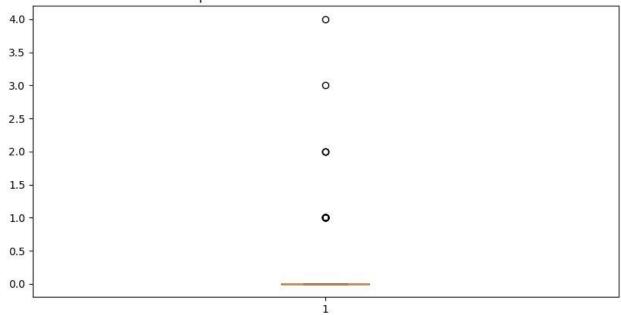
#### Boxplot of MOR BANK: POLITICAL CONTRIBUTOR



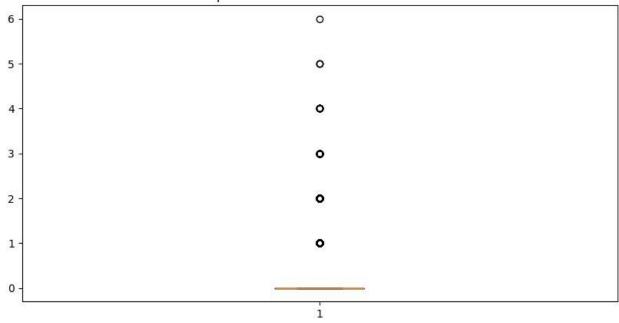




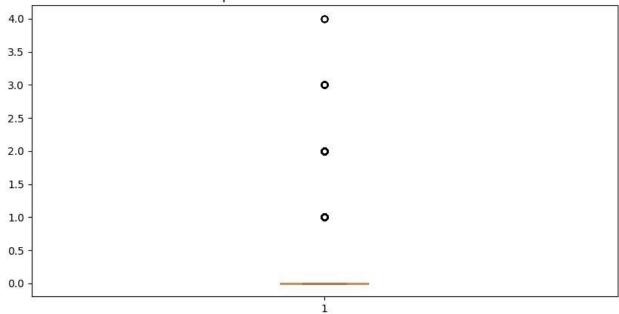
#### Boxplot of MOR BANK: GENERAL CONTRIBUTOR



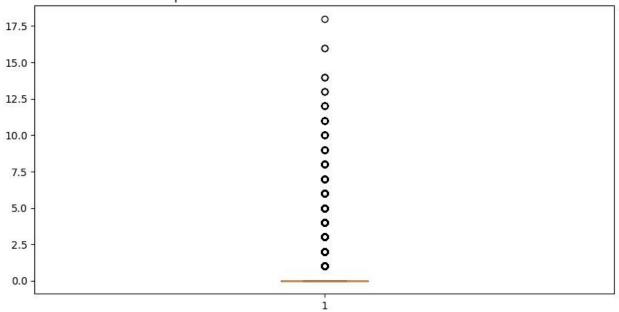
#### Boxplot of MOR BANK: MISCELLANEOUS



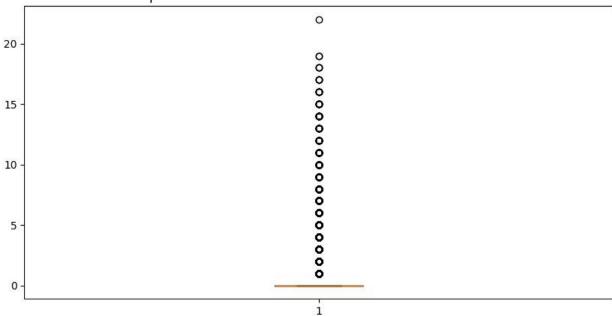
#### Boxplot of MOR BANK: ODDS AND ENDS



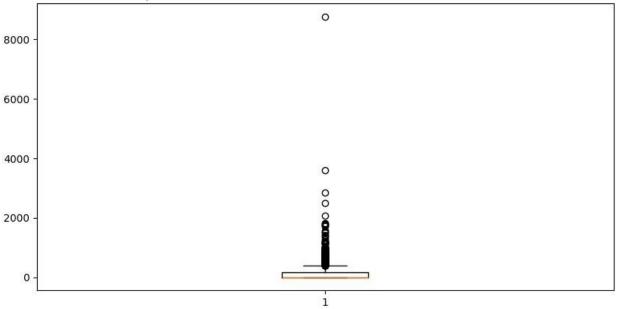
#### Boxplot of MOR BANK: DEDUPED CATEGORY HIT COUNT



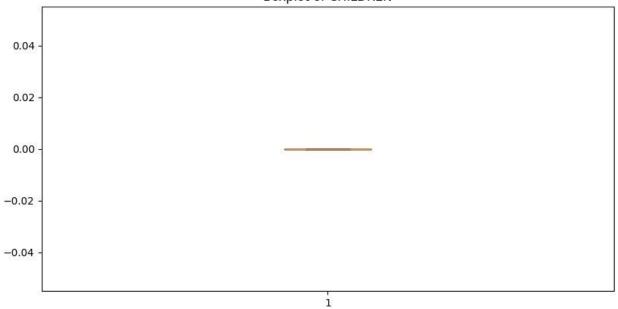
## Boxplot of MOR BANK: NON-DEDUPED CATEGORY HIT COUNT



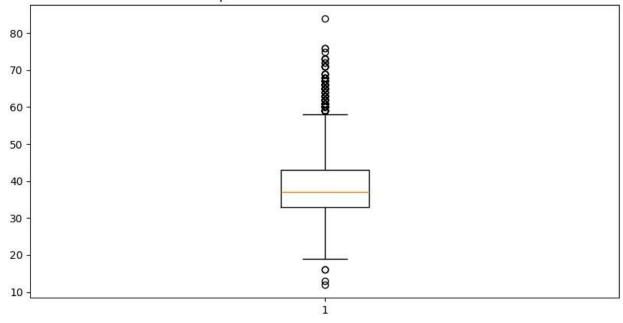


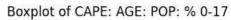


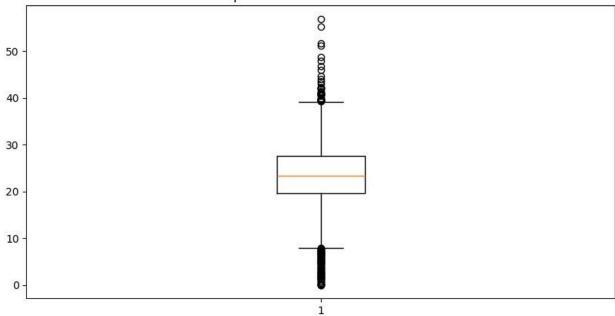
# Boxplot of CHILDREN



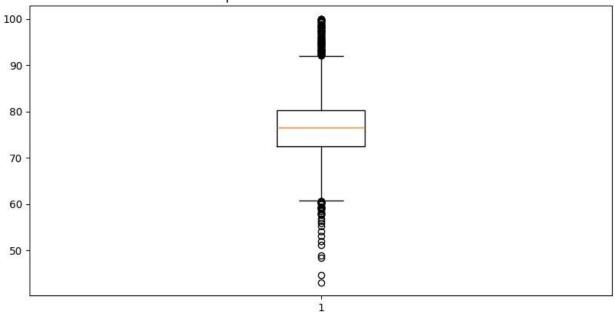
Boxplot of CAPE: AGE: POP: MEDIAN AGE



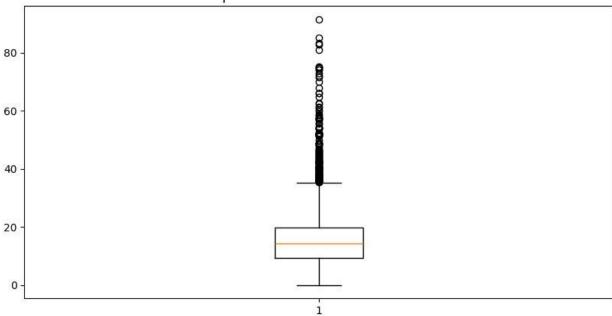




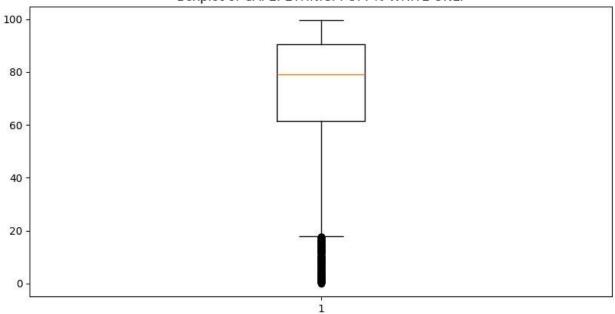
Boxplot of CAPE: AGE: POP: % 18-99+



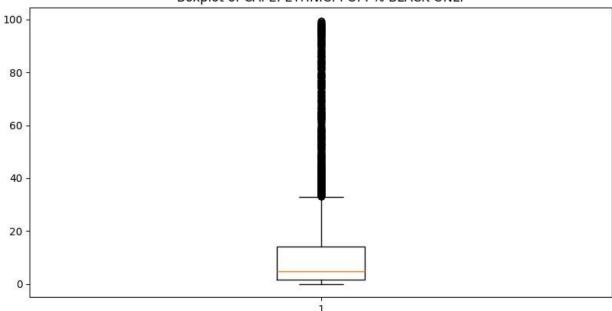




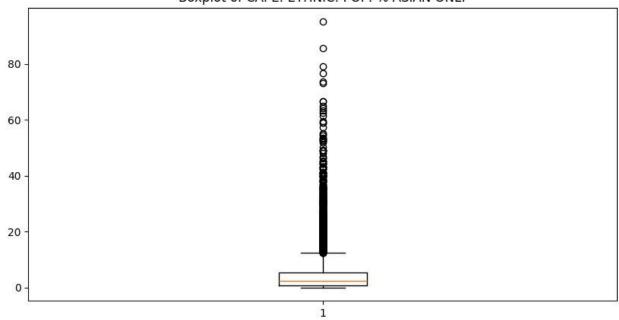
Boxplot of CAPE: ETHNIC: POP: % WHITE ONLY



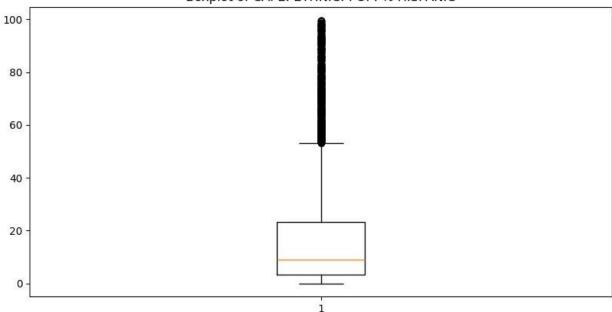




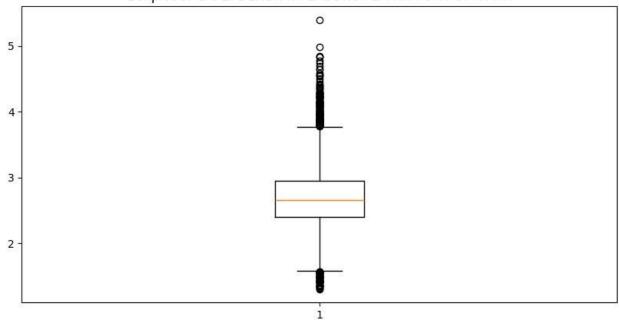
Boxplot of CAPE: ETHNIC: POP: % ASIAN ONLY



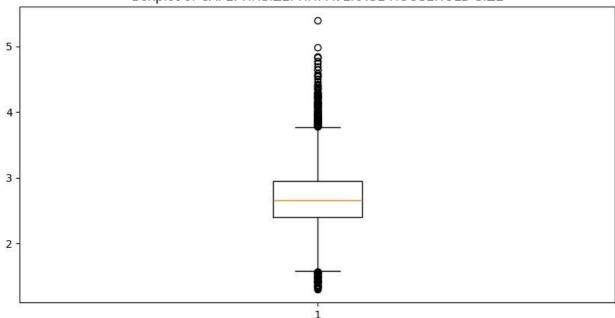




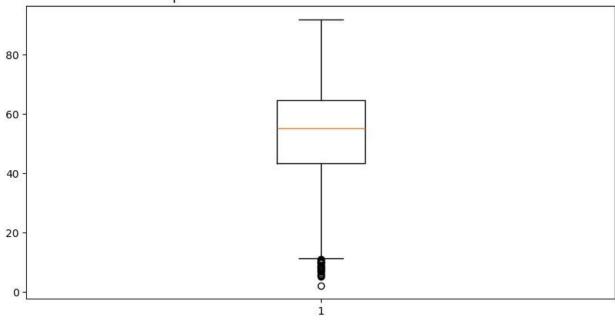
Boxplot of CAPE: DENSITY: PERSONS PER HH FOR POP IN HH

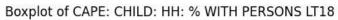


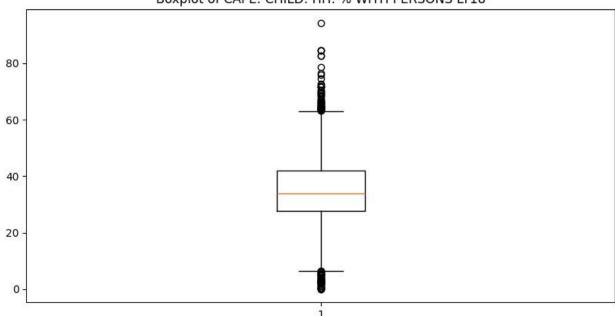
Boxplot of CAPE: HHSIZE: HH: AVERAGE HOUSEHOLD SIZE



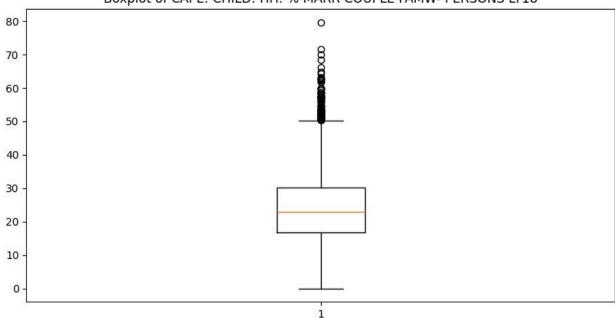
Boxplot of CAPE: TYP: HH: % MARRIED COUPLE FAMILY



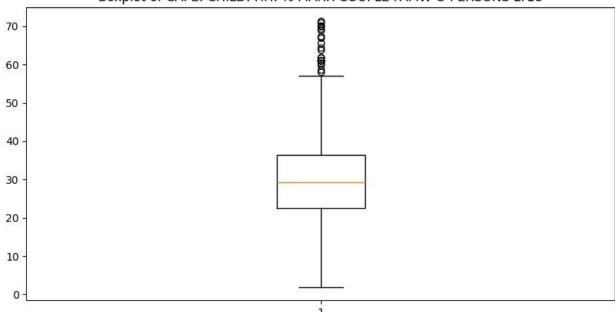




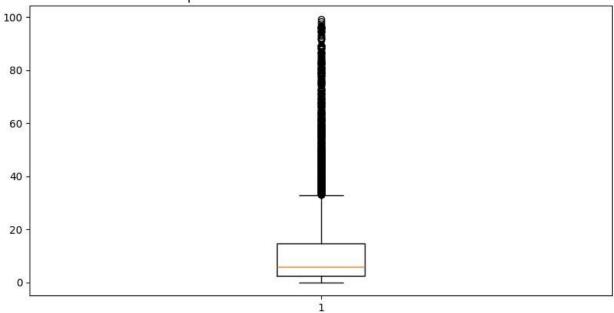


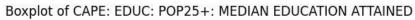


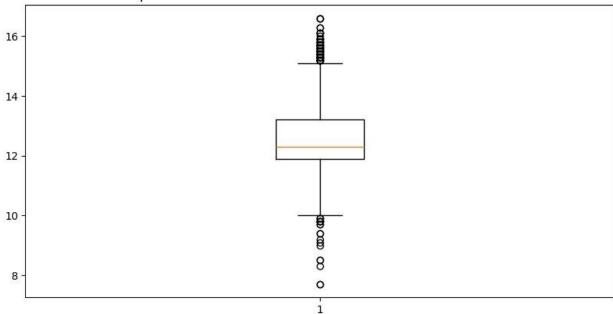
#### Boxplot of CAPE: CHILD: HH: % MARR COUPLE FAMW-O PERSONS LT18

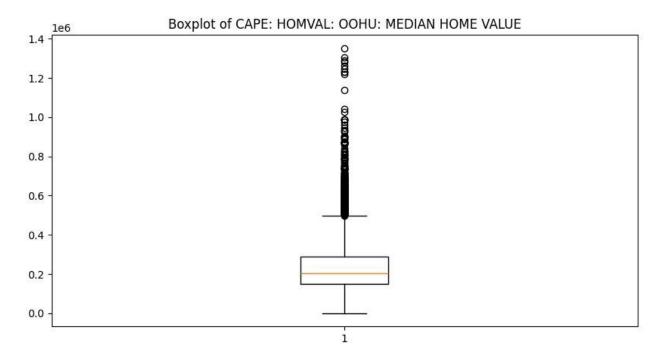


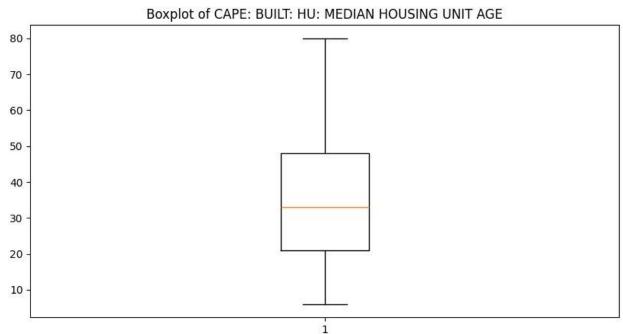
Boxplot of CAPE: LANG: HH: % SPANISH SPEAKING



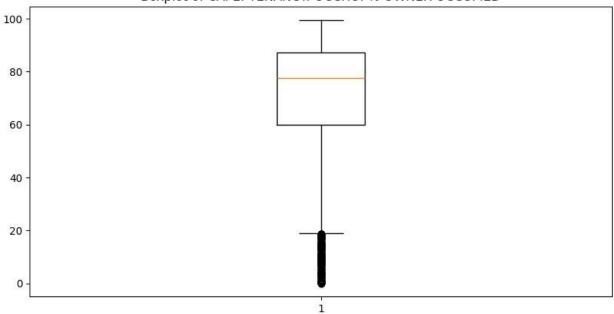


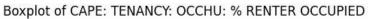


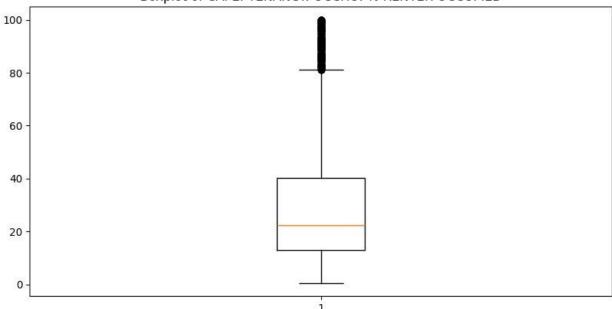




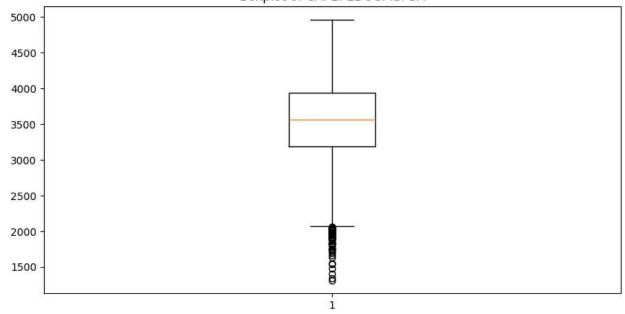
Boxplot of CAPE: TENANCY: OCCHU: % OWNER OCCUPIED



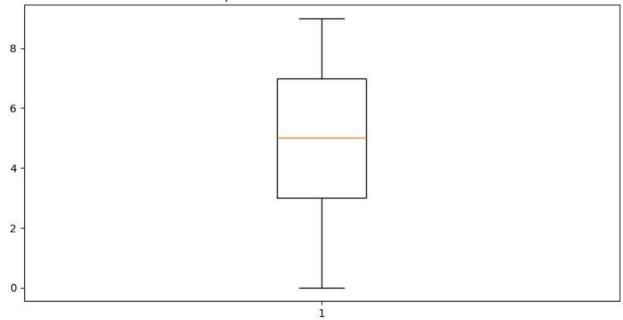




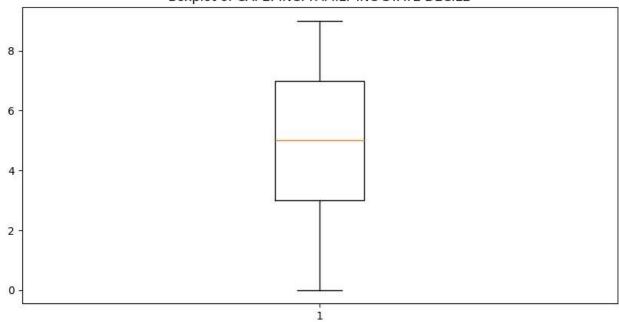
Boxplot of CAPE: EDUC: ISPSA



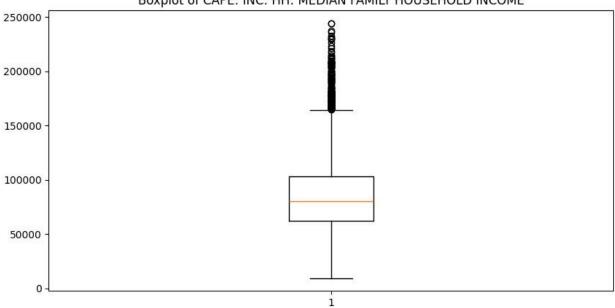
## Boxplot of CAPE: EDUC: ISPSA DECILE

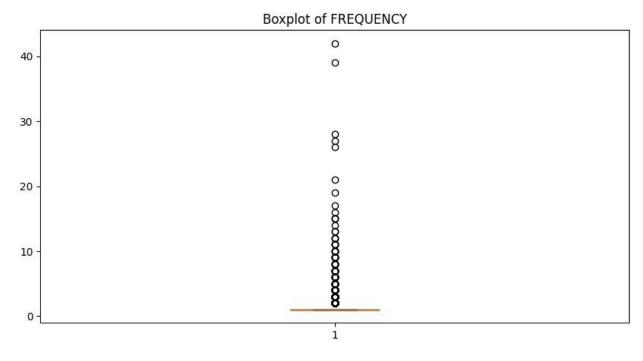


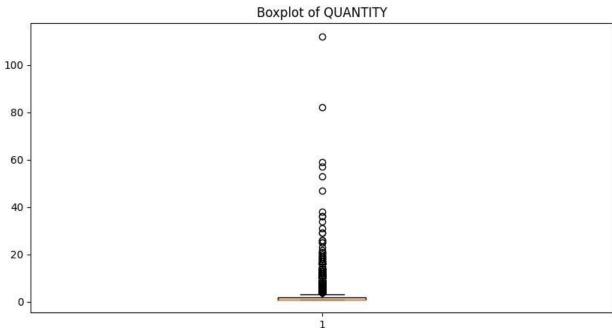
Boxplot of CAPE: INC: FAMILY INC STATE DECILE

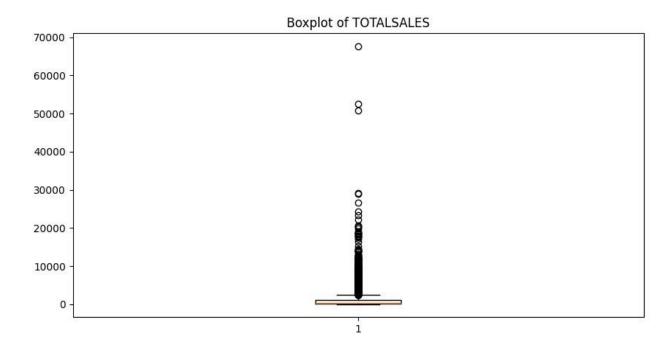


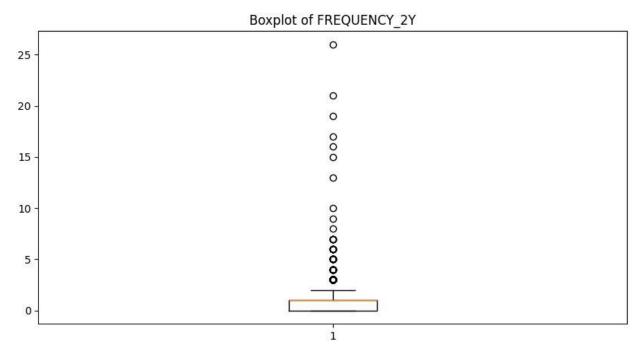


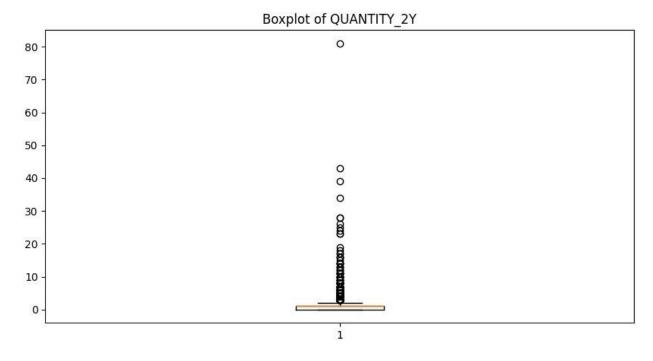


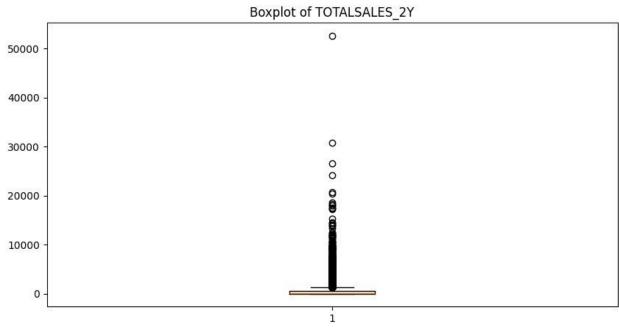


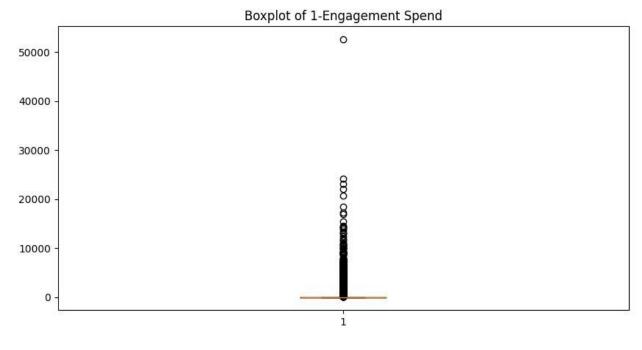


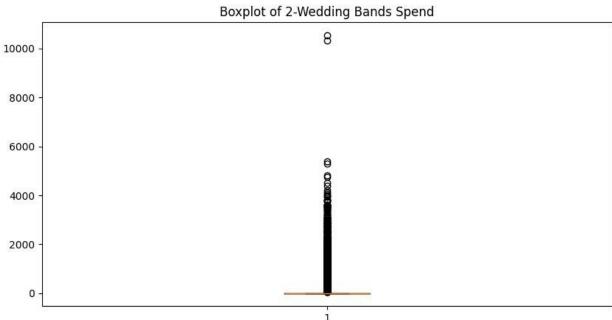


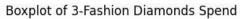


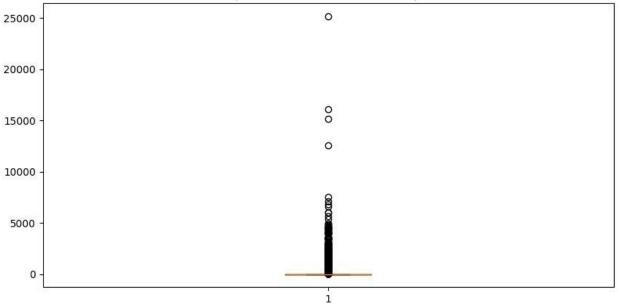




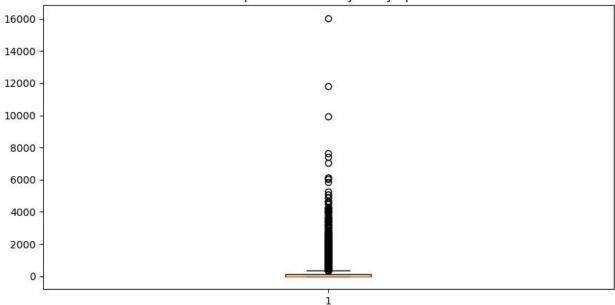


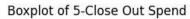


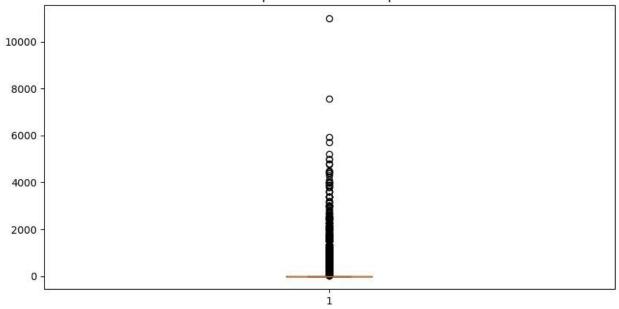




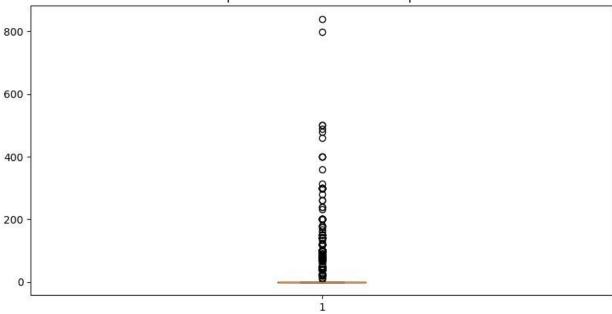
#### Boxplot of 4-Fashion Jewelry Spend



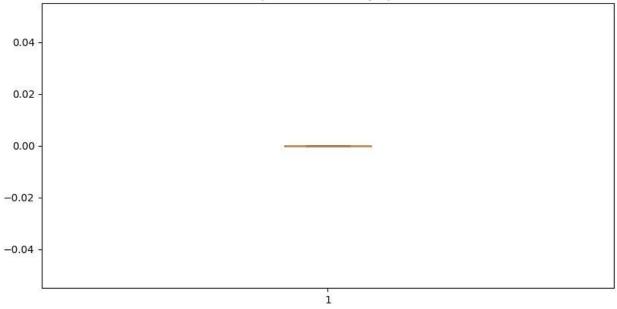




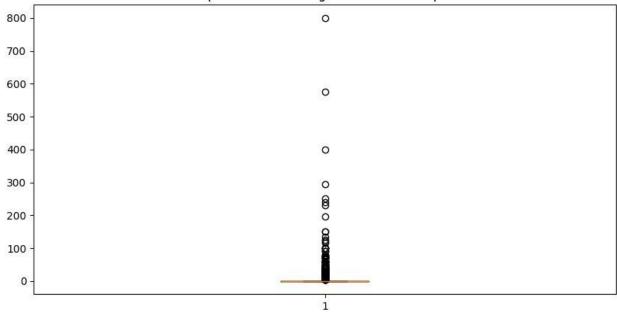
### Boxplot of 6-Promotional Items Spend

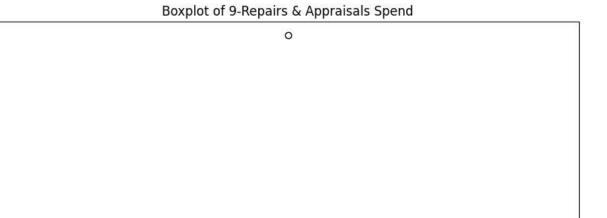




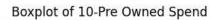


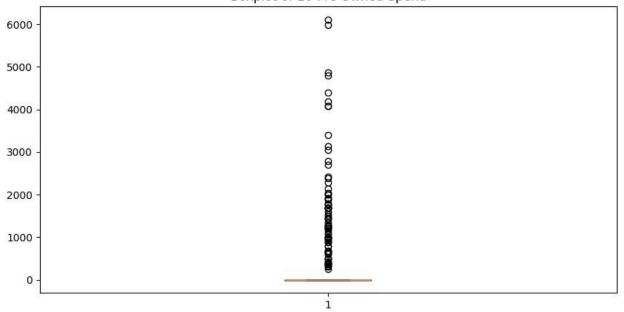
#### Boxplot of 8-Marketing Premium SKUs Spend



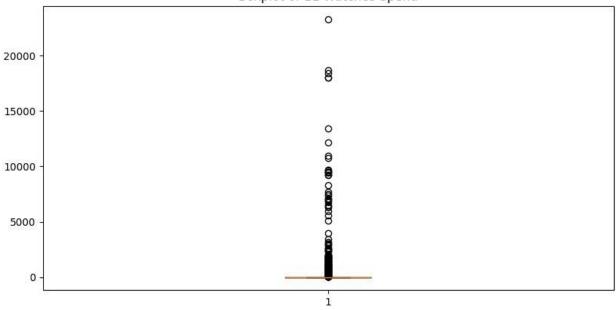


100 -

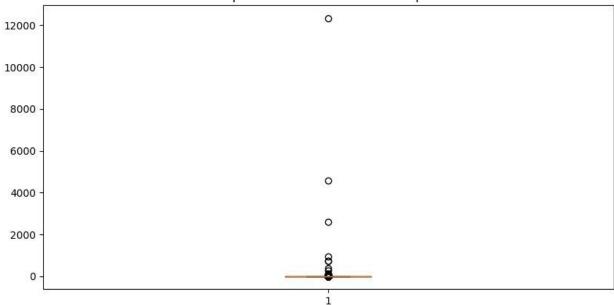




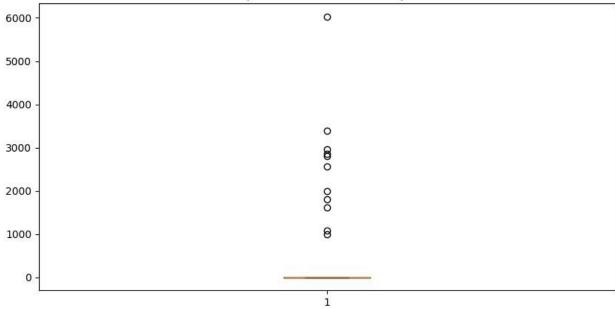




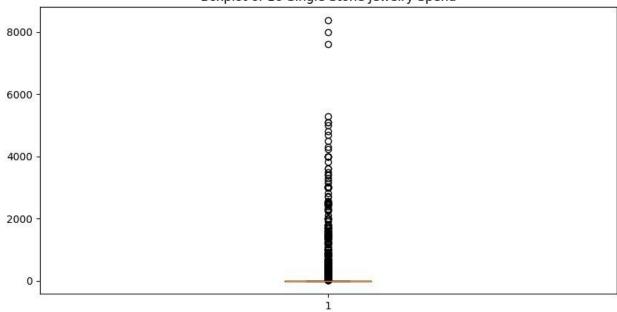
### Boxplot of 12-Misc Merchandise Spend

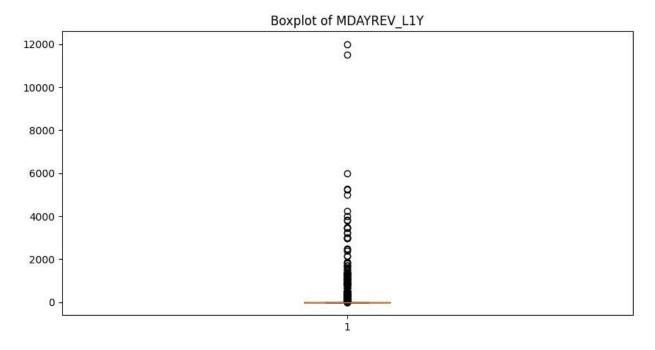


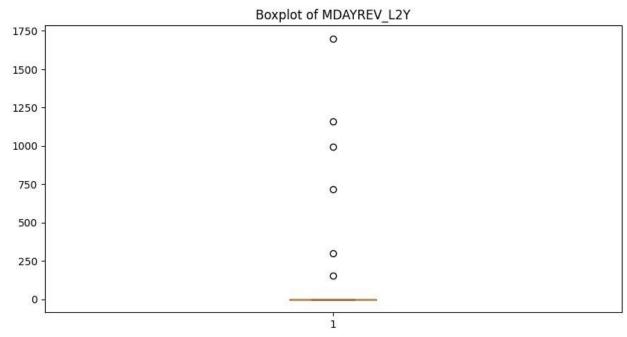


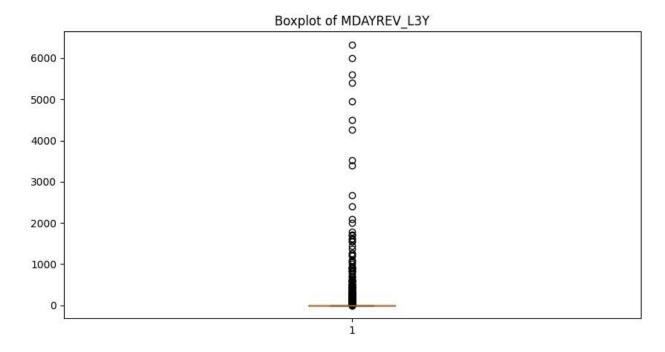


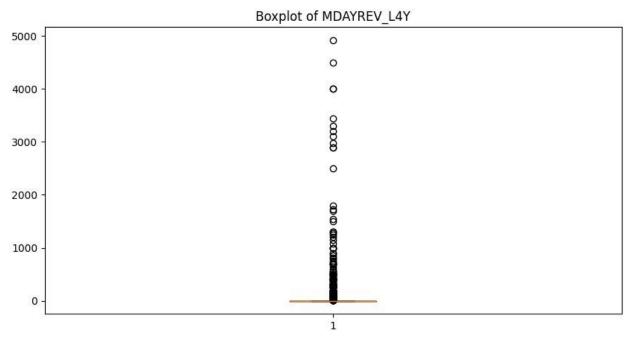
### Boxplot of 16-Single Stone Jewelry Spend

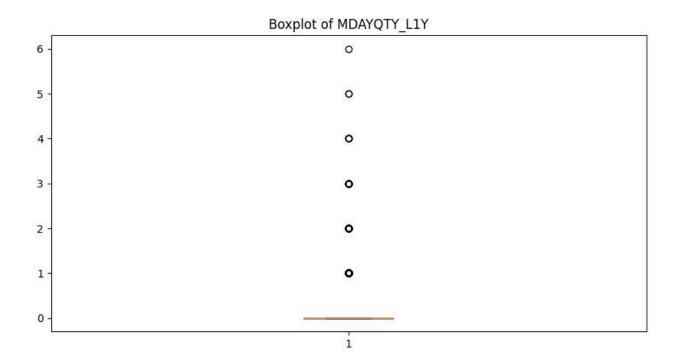


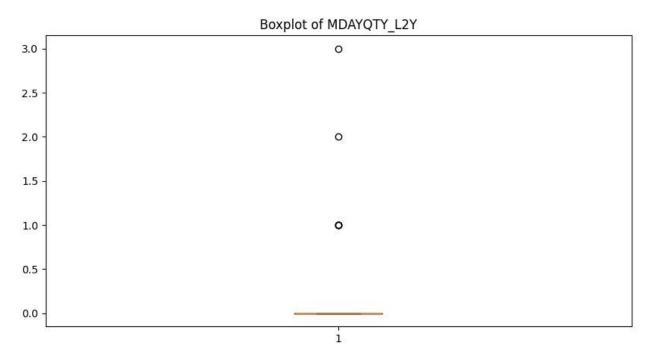


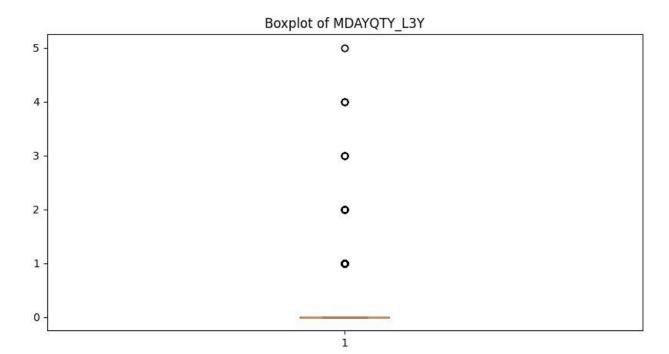


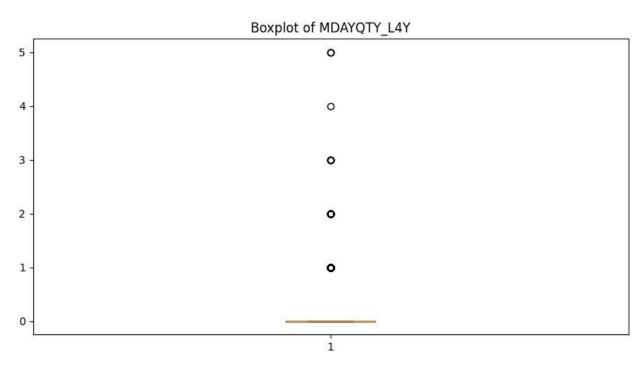


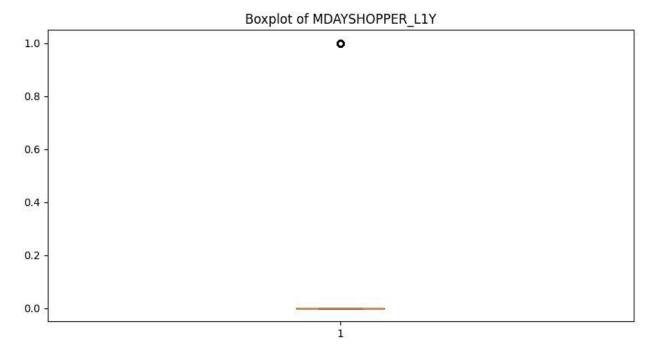


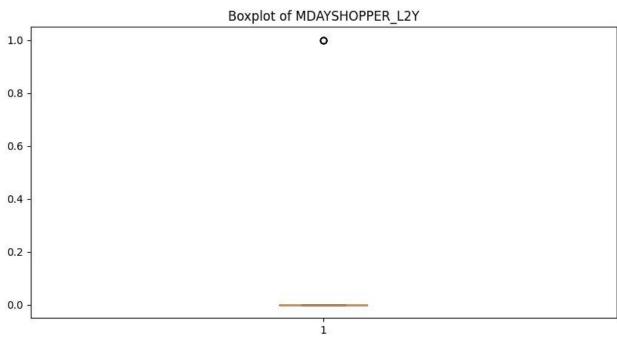


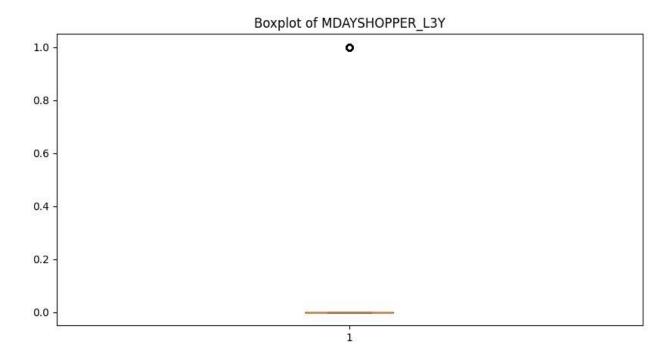


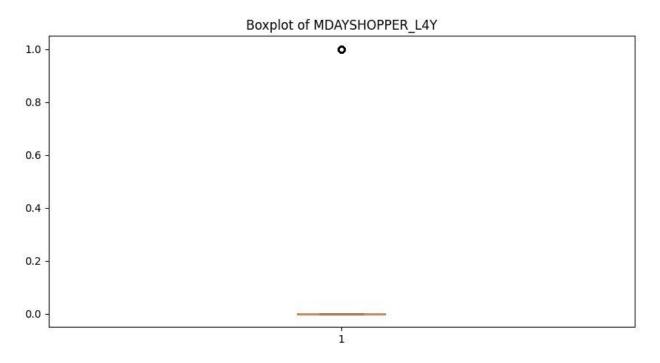












#### Plot distribution for 6 columns

