assignment2

April 26, 2024

0.1 Assignment 2

Name - Jay Shirgupe

PRN - 21510026

Batch - T7

[ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

[ ]:
```
train = pd.read_csv(
    '/home/oneautumleaf/Desktop/COLLEGE_ASSIGNMENTS/MLL/datasets/assignment2/
 ↪kdd+cup+1998+data/epsilon_mirror/cup98lrn/cup98LRN.txt')
test = pd.read_csv(
    '/home/oneautumleaf/Desktop/COLLEGE_ASSIGNMENTS/MLL/datasets/assignment2/
 ↪kdd+cup+1998+data/epsilon_mirror/cup98val/cup98VAL.txt')
```

```
/home/oneautumleaf/tmp/ipykernel_4960/2622488138.py:1: DtypeWarning: Columns (8)
have mixed types. Specify dtype option on import or set low_memory=False.
  train = pd.read_csv(
/home/oneautumleaf/tmp/ipykernel_4960/2622488138.py:3: DtypeWarning: Columns (8)
have mixed types. Specify dtype option on import or set low_memory=False.
  test = pd.read_csv(
```

[ ]:
```
train.head()
```

[ ]:

	ODATEDW	OSOURCE	TCODE	STATE	ZIP	MAILCODE	PVASTATE	DOB	NOEXCH	\
0	8901	GRI	0	IL	61081			3712	0	
1	9401	BOA	1	CA	91326			5202	0	
2	9001	AMH	1	NC	27017			0	0	
3	8701	BRY	0	CA	95953			2801	0	
4	8601		0	FL	33176			2001	0	

	RECINHSE	…	TARGET_D	HPHONE_D	RFA_2R	RFA_2F	RFA_2A	MDMAUD_R	MDMAUD_F	\
0		…	0.0	0	L	4	E	X	X	
1		…	0.0	0	L	2	G	X	X	
2		…	0.0	1	L	4	E	X	X	
3		…	0.0	1	L	4	E	X	X	
4	X	…	0.0	1	L	2	F	X	X	

```
  MDMAUD_A CLUSTER2 GEOCODE2
0        X     39.0        C
1        X      1.0        A
2        X     60.0        C
3        X     41.0        C
4        X     26.0        A

[5 rows x 481 columns]
```

[ ]:
```
test.head()
```

[ ]:
```
   ODATEDW OSOURCE  TCODE STATE     ZIP MAILCODE PVASTATE   DOB NOEXCH  \
0     9301     PV3      0    WI   53558                    1702      0
1     9501     CNT      0    TX   78501                    1501      0
2     9101     L01      1    CA   92392                    3001      0
3     9401     HAM      0    TX  77625-                    4909      0
4     9201     NAS      2    TX   76051                    5406      0

  RECINHSE  … CONTROLN HPHONE_D RFA_2R RFA_2F RFA_2A MDMAUD_R  MDMAUD_F  \
0        X  …   188946        1      L      1      G        X         X
1           …   126296        1      L      1      F        X         X
2           …   155244        1      L      1      G        X         X
3           …   123985        0      L      1      F        X         X
4           …   119118        1      L      3      E        X         X

  MDMAUD_A CLUSTER2 GEOCODE2
0        X     15.0        B
1        X     13.0        B
2        X     34.0        A
3        X     58.0        B
4        X      2.0        A

[5 rows x 479 columns]
```

1. Display shape of data frame

[ ]:
```
train.shape
```

[ ]:
```
(95412, 481)
```

[ ]:
```
test.shape
```

[ ]:
```
(96367, 479)
```

2. Display column names

[ ]:
```
len(train.columns)
```

[ ]: 481

[ ]:
```
train.columns
```

[ ]:
```
Index(['ODATEDW', 'OSOURCE', 'TCODE', 'STATE', 'ZIP', 'MAILCODE', 'PVASTATE',
       'DOB', 'NOEXCH', 'RECINHSE',
       …
       'TARGET_D', 'HPHONE_D', 'RFA_2R', 'RFA_2F', 'RFA_2A', 'MDMAUD_R',
       'MDMAUD_F', 'MDMAUD_A', 'CLUSTER2', 'GEOCODE2'],
      dtype='object', length=481)
```

[ ]:
```
len(test.columns)
```

[ ]: 479

[ ]:
```
test.columns
```

[ ]:
```
Index(['ODATEDW', 'OSOURCE', 'TCODE', 'STATE', 'ZIP', 'MAILCODE', 'PVASTATE',
       'DOB', 'NOEXCH', 'RECINHSE',
       …
       'CONTROLN', 'HPHONE_D', 'RFA_2R', 'RFA_2F', 'RFA_2A', 'MDMAUD_R',
       'MDMAUD_F', 'MDMAUD_A', 'CLUSTER2', 'GEOCODE2'],
      dtype='object', length=479)
```

[ ]:
```
for col in train.columns:
    if col not in test.columns:
        print(col)
```

```
TARGET_B
TARGET_D
```

3. Display 5 quantiles of the dataset

[ ]:
```
train.quantile([0, .25, .5, .75, 1], numeric_only=True)
```

[ ]:

	ODATEDW	TCODE	DOB	AGE	NUMCHLD	INCOME	WEALTH1	HIT	\
0.00	8306.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	
0.25	8801.0	0.0	201.0	48.0	1.0	2.0	3.0	0.0	
0.50	9201.0	1.0	2610.0	62.0	1.0	4.0	6.0	0.0	
0.75	9501.0	2.0	4601.0	75.0	2.0	5.0	8.0	3.0	
1.00	9701.0	72002.0	9710.0	98.0	7.0	7.0	9.0	241.0	

	MBCRAFT	MBGARDEN	…	FISTDATE	NEXTDATE	TIMELAG	AVGGIFT	\
0.00	0.0	0.0	…	0.0	7211.0	0.0	1.285714	
0.25	0.0	0.0	…	8810.0	8903.0	4.0	8.384615	
0.50	0.0	0.0	…	9201.0	9204.0	6.0	11.636364	
0.75	0.0	0.0	…	9409.0	9409.0	11.0	15.477955	
1.00	6.0	4.0	…	9603.0	9702.0	1088.0	1000.000000	

```
        CONTROLN  TARGET_B  TARGET_D  HPHONE_D  RFA_2F  CLUSTER2
0.00       1.00       0.0       0.0       0.0     1.0       1.0
0.25   47910.75       0.0       0.0       0.0     1.0      15.0
0.50   95681.50       0.0       0.0       1.0     2.0      32.0
0.75  143643.50       0.0       0.0       1.0     3.0      49.0
1.00  191779.00       1.0     200.0       1.0     4.0      62.0

[5 rows x 407 columns]
```

[ ]:
```
test.quantile([0, .25, .5, .75, 1], numeric_only=True)
```

[ ]:
```
      ODATEDW    TCODE     DOB   AGE  NUMCHLD  INCOME  WEALTH1    HIT  \
0.00   8301.0      0.0     0.0   1.0      1.0     1.0      0.0    0.0
0.25   8801.0      0.0   101.0  48.0      1.0     2.0      3.0    0.0
0.50   9201.0      1.0  2610.0  62.0      1.0     4.0      6.0    0.0
0.75   9501.0      2.0  4506.0  75.0      2.0     5.0      8.0    3.0
1.00   9701.0  39002.0  9705.0  98.0      7.0     7.0      9.0  242.0

      MBCRAFT  MBGARDEN  …  LASTGIFT  LASTDATE  FISTDATE  NEXTDATE  TIMELAG  \
0.00      0.0       0.0  …       0.0    9503.0       0.0    7312.0      0.0
0.25      0.0       0.0  …      10.0    9509.0    8811.0    8904.0      4.0
0.50      0.0       0.0  …      15.0    9512.0    9202.0    9204.0      6.0
0.75      0.0       0.0  …      20.0    9602.0    9409.0    9409.0     11.0
1.00      6.0       3.0  …   10000.0    9702.0    9603.0    9702.0   1060.0

         AVGGIFT  CONTROLN  HPHONE_D  RFA_2F  CLUSTER2
0.00    1.578947       3.0       0.0     1.0       1.0
0.25    8.400000   47965.0       0.0     1.0      14.0
0.50   11.666667   96125.0       1.0     1.0      32.0
0.75   15.500000  144034.0       1.0     3.0      49.0
1.00  650.000000  191776.0       1.0     4.0      62.0

[5 rows x 405 columns]
```

4. Display count of rows having null in any column

[ ]:
```
train.isnull()
```

[ ]:
```
       ODATEDW  OSOURCE  TCODE  STATE    ZIP  MAILCODE  PVASTATE    DOB  \
0        False    False  False  False  False     False     False  False
1        False    False  False  False  False     False     False  False
2        False    False  False  False  False     False     False  False
3        False    False  False  False  False     False     False  False
4        False    False  False  False  False     False     False  False
…        …      …    …      …      …      …         …       …      …
95407    False    False  False  False  False     False     False  False
```

```
95408    False    False  False  False  False     False     False  False
95409    False    False  False  False  False     False     False  False
95410    False    False  False  False  False     False     False  False
95411    False    False  False  False  False     False     False  False

       NOEXCH  RECINHSE  …  TARGET_D  HPHONE_D  RFA_2R  RFA_2F  RFA_2A  \
0       False     False  …     False     False   False   False   False
1       False     False  …     False     False   False   False   False
2       False     False  …     False     False   False   False   False
3       False     False  …     False     False   False   False   False
4       False     False  …     False     False   False   False   False
…       …       …  …       …       …       …     …     …
95407   False     False  …     False     False   False   False   False
95408   False     False  …     False     False   False   False   False
95409   False     False  …     False     False   False   False   False
95410   False     False  …     False     False   False   False   False
95411   False     False  …     False     False   False   False   False

       MDMAUD_R  MDMAUD_F  MDMAUD_A  CLUSTER2  GEOCODE2
0         False     False     False     False     False
1         False     False     False     False     False
2         False     False     False     False     False
3         False     False     False     False     False
4         False     False     False     False     False
…         …       …       …       …       …
95407     False     False     False     False     False
95408     False     False     False     False     False
95409     False     False     False     False     False
95410     False     False     False     False     False
95411     False     False     False     False     False

[95412 rows x 481 columns]
```

[ ]:
```
train.isnull().any(axis=0)
```

[ ]:
```
ODATEDW     False
OSOURCE     False
TCODE       False
STATE       False
ZIP         False
            …
MDMAUD_R    False
MDMAUD_F    False
MDMAUD_A    False
CLUSTER2     True
GEOCODE2     True
Length: 481, dtype: bool
```

[ ]:
```
print(f"Number of null columns: {train.isnull().any(axis=0).sum()}")
print(
    f"% of null columns: {train.isnull().any(axis=0).sum() / train.shape[1] *␣
 ↪100} %")


print(f"Number of null rows: {train.isnull().any(axis=1).sum()}")
print(
    f"% of null rows: {train.isnull().any(axis=1).sum() / train.shape[0] * 100}␣
 ↪%")
```

```
Number of null columns: 92
% of null columns: 19.12681912681913 %
Number of null rows: 95412
% of null rows: 100.0 %
```

[ ]:
```
train.isnull().any(axis=1)
```

[ ]:
```
0        True
1        True
2        True
3        True
4        True
         …
95407    True
95408    True
95409    True
95410    True
95411    True
Length: 95412, dtype: bool
```

[ ]:
```
train.isnull().any(axis=1).sum()
```

[ ]:
```
95412
```

[ ]:
```
print(f"Number of null columns: {test.isnull().any(axis=0).sum()}")
print(
    f"% of null columns: {test.isnull().any(axis=0).sum() / test.shape[1] *␣
 ↪100} %")


print(f"Number of null rows: {test.isnull().any(axis=1).sum()}")
print(
    f"% of null rows: {test.isnull().any(axis=1).sum() / test.shape[0] * 100}␣
 ↪%")
```

```
Number of null columns: 92
```

```
% of null columns: 19.206680584551147 %
Number of null rows: 96367
% of null rows: 100.0 %
```

5. Display first 10 rows

```
[ ]: train.head(10)
```

```
[ ]:    ODATEDW OSOURCE  TCODE STATE    ZIP MAILCODE PVASTATE   DOB NOEXCH  \
     0     8901     GRI      0    IL  61081                    3712      0
     1     9401     BOA      1    CA  91326                    5202      0
     2     9001     AMH      1    NC  27017                       0      0
     3     8701     BRY      0    CA  95953                    2801      0
     4     8601              0    FL  33176                    2001      0
     5     9401     CWR      0    AL  35603                       0      0
     6     8701     DRK      0    IN  46755                    6001      0
     7     9401     NWN      0    LA  70611                       0      0
     8     8801     LIS      1    IA  51033                       0      0
     9     9401     MSD      1    TN 37127-                    3211      0

       RECINHSE  … TARGET_D HPHONE_D RFA_2R RFA_2F RFA_2A MDMAUD_R  MDMAUD_F  \
     0           …      0.0        0      L      4      E        X         X
     1           …      0.0        0      L      2      G        X         X
     2           …      0.0        1      L      4      E        X         X
     3           …      0.0        1      L      4      E        X         X
     4        X  …      0.0        1      L      2      F        X         X
     5           …      0.0        0      L      1      F        X         X
     6           …      0.0        1      L      1      E        X         X
     7           …      0.0        1      L      3      E        X         X
     8           …      0.0        1      L      1      F        X         X
     9           …      0.0        0      L      1      F        X         X

       MDMAUD_A CLUSTER2 GEOCODE2
     0        X     39.0        C
     1        X      1.0        A
     2        X     60.0        C
     3        X     41.0        C
     4        X     26.0        A
     5        X     16.0        C
     6        X     53.0        D
     7        X     38.0        C
     8        X     57.0        D
     9        X     34.0        B

     [10 rows x 481 columns]
```

6. Display last 10 rows

[ ]:
```
train.tail(10)
```

[ ]:

	ODATEDW	OSOURCE	TCODE	STATE	ZIP	MAILCODE	PVASTATE	DOB	NOEXCH	\
95402	9501	HOS	0	AL	35653			0	0	
95403	9001	LHJ	0	MN	55932			4001	0	
95404	9401	L02	2	KY	41701-			4506	0	
95405	9201	HHH	28	CA	92581			0	0	
95406	9001	LHJ	0	FL	32776-			4803	0	
95407	9601	ASE	1	AK	99504			0	0	
95408	9601	DCD	1	TX	77379			5001	0	
95409	9501	MBC	1	MI	48910			3801	0	
95410	8601	PRV	0	CA	91320			4005	0	
95411	8801	MCC	2	NC	28409			1801	0	

	RECINHSE	…	TARGET_D	HPHONE_D	RFA_2R	RFA_2F	RFA_2A	MDMAUD_R	MDMAUD_F	\
95402		…	0.0	0	L	3	E	X	X	
95403		…	10.0	0	L	1	F	X	X	
95404		…	0.0	0	L	4	E	X	X	
95405		…	0.0	0	L	1	G	X	X	
95406		…	0.0	0	L	4	D	X	X	
95407		…	0.0	0	L	1	G	X	X	
95408		…	0.0	1	L	1	F	X	X	
95409		…	0.0	1	L	3	E	X	X	
95410	X	…	18.0	1	L	4	F	X	X	
95411	X	…	0.0	1	L	1	G	C	1	

	MDMAUD_A	CLUSTER2	GEOCODE2
95402	X	53.0	D
95403	X	56.0	D
95404	X	53.0	D
95405	X	35.0	A
95406	X	43.0	C
95407	X	12.0	C
95408	X	2.0	A
95409	X	34.0	B
95410	X	11.0	A
95411	C	12.0	C

[10 rows x 481 columns]

12. Replace/eliminate missing values

[ ]:
```
def calculate_null(df):
    df_null_rows = df[df.isnull().any(axis=1)]
    df_null_columns = df.columns[df.isnull().any(axis=0)]
    df_null_percentage = df[df_null_columns].isnull().mean()
    print(f"% null rows: {df_null_rows.shape[0] / df.shape[0] * 100}")
```

```
    print(f"% null columns: {df_null_columns.shape[0] / df.shape[1] * 100}")
```

[ ]:
```
calculate_null(train)
```

```
% null rows: 100.0
% null columns: 19.12681912681913
```

[ ]:
```
null_threshold = 0.2
filtered_columns = train.columns[train.isnull().mean() < null_threshold]
train_filtered = train[filtered_columns]
```

[ ]:
```
null_threshold = 0.2
filtered_columns = test.columns[test.isnull().mean() < null_threshold]
test_filtered = test[filtered_columns]
```

[ ]:
```
for column in train_filtered:
    if train_filtered[column].dtype == 'object':
        mode_value = train_filtered[column].mode().iloc[0]
        train_filtered[column] = train_filtered[column].fillna(mode_value)
    else:
        mean_value = train_filtered[column].mean()
        train_filtered[column] = train_filtered[column].fillna(mean_value)
```

```
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mean_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mean_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mean_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
```

```
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
```

```
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mean_value)
```

```
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mean_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

```
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
```

```
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mean_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
```

```
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
```

```
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
```

```
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mean_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mean_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

```
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mean_value)
/home/oneautumleaf/tmp/ipykernel_5699/2887327854.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  train_filtered[column] = train_filtered[column].fillna(mode_value)
```

[ ]:
```
for column in test_filtered:
    if test_filtered[column].dtype == 'object':
        mode_value = test_filtered[column].mode().iloc[0]
        test_filtered[column] = test_filtered[column].fillna(mode_value)
    else:
        mean_value = test_filtered[column].mean()
        test_filtered[column] = test_filtered[column].fillna(mean_value)
```

```
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mean_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mean_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mean_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
```

```
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
```

```
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
```

```
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mean_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mean_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

```
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
```

```
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mean_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
```

```
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
```

```
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
```

```
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mean_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

```
  test_filtered[column] = test_filtered[column].fillna(mean_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mean_value)
/home/oneautumleaf/tmp/ipykernel_5699/2190281181.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  test_filtered[column] = test_filtered[column].fillna(mode_value)
```

[ ]:
```
calculate_null(train_filtered)
```

```
% null rows: 0.0
% null columns: 0.0
```

[ ]:
```
calculate_null(test_filtered)
```

```
% null rows: 0.0
% null columns: 0.0
```

13. Change column name(s) to short/easy names if required.

[ ]:
```
for col in train_filtered.columns:
    print(col)
```

```
ODATEDW
OSOURCE
TCODE
STATE
ZIP
MAILCODE
PVASTATE
DOB
NOEXCH
RECINHSE
RECP3
RECPGVG
RECSWEEP
MDMAUD
DOMAIN
CLUSTER
AGEFLAG
HOMEOWNR
CHILD03
CHILD07
CHILD12
CHILD18
GENDER
HIT
DATASRCE
MALEMILI
MALEVET
VIETVETS
WWIIVETS
LOCALGOV
STATEGOV
FEDGOV
SOLP3
SOLIH
MAJOR
GEOCODE
COLLECT1
VETERANS
```

BIBLE
CATLG
HOMEE
PETS
CDPLAY
STEREO
PCOWNERS
PHOTO
CRAFTS
FISHER
GARDENIN
BOATS
WALKER
KIDSTUFF
CARDS
PLATES
LIFESRC
PEPSTRFL
POP901
POP902
POP903
POP90C1
POP90C2
POP90C3
POP90C4
POP90C5
ETH1
ETH2
ETH3
ETH4
ETH5
ETH6
ETH7
ETH8
ETH9
ETH10
ETH11
ETH12
ETH13
ETH14
ETH15
ETH16
AGE901
AGE902
AGE903
AGE904
AGE905
AGE906

AGE907
CHIL1
CHIL2
CHIL3
AGEC1
AGEC2
AGEC3
AGEC4
AGEC5
AGEC6
AGEC7
CHILC1
CHILC2
CHILC3
CHILC4
CHILC5
HHAGE1
HHAGE2
HHAGE3
HHN1
HHN2
HHN3
HHN4
HHN5
HHN6
MARR1
MARR2
MARR3
MARR4
HHP1
HHP2
DW1
DW2
DW3
DW4
DW5
DW6
DW7
DW8
DW9
HV1
HV2
HV3
HV4
HU1
HU2
HU3
HU4

HU5
HHD1
HHD2
HHD3
HHD4
HHD5
HHD6
HHD7
HHD8
HHD9
HHD10
HHD11
HHD12
ETHC1
ETHC2
ETHC3
ETHC4
ETHC5
ETHC6
HVP1
HVP2
HVP3
HVP4
HVP5
HVP6
HUR1
HUR2
RHP1
RHP2
RHP3
RHP4
HUPA1
HUPA2
HUPA3
HUPA4
HUPA5
HUPA6
HUPA7
RP1
RP2
RP3
RP4
MSA
ADI
DMA
IC1
IC2
IC3

IC4
IC5
IC6
IC7
IC8
IC9
IC10
IC11
IC12
IC13
IC14
IC15
IC16
IC17
IC18
IC19
IC20
IC21
IC22
IC23
HHAS1
HHAS2
HHAS3
HHAS4
MC1
MC2
MC3
TPE1
TPE2
TPE3
TPE4
TPE5
TPE6
TPE7
TPE8
TPE9
PEC1
PEC2
TPE10
TPE11
TPE12
TPE13
LFC1
LFC2
LFC3
LFC4
LFC5
LFC6

LFC7
LFC8
LFC9
LFC10
OCC1
OCC2
OCC3
OCC4
OCC5
OCC6
OCC7
OCC8
OCC9
OCC10
OCC11
OCC12
OCC13
EIC1
EIC2
EIC3
EIC4
EIC5
EIC6
EIC7
EIC8
EIC9
EIC10
EIC11
EIC12
EIC13
EIC14
EIC15
EIC16
OEDC1
OEDC2
OEDC3
OEDC4
OEDC5
OEDC6
OEDC7
EC1
EC2
EC3
EC4
EC5
EC6
EC7
EC8

SEC1
SEC2
SEC3
SEC4
SEC5
AFC1
AFC2
AFC3
AFC4
AFC5
AFC6
VC1
VC2
VC3
VC4
ANC1
ANC2
ANC3
ANC4
ANC5
ANC6
ANC7
ANC8
ANC9
ANC10
ANC11
ANC12
ANC13
ANC14
ANC15
POBC1
POBC2
LSC1
LSC2
LSC3
LSC4
VOC1
VOC2
VOC3
HC1
HC2
HC3
HC4
HC5
HC6
HC7
HC8
HC9

HC10
HC11
HC12
HC13
HC14
HC15
HC16
HC17
HC18
HC19
HC20
HC21
MHUC1
MHUC2
AC1
AC2
ADATE_2
ADATE_3
ADATE_4
ADATE_6
ADATE_7
ADATE_8
ADATE_9
ADATE_11
ADATE_12
ADATE_14
RFA_2
RFA_3
RFA_4
RFA_5
RFA_6
RFA_7
RFA_8
RFA_9
RFA_10
RFA_11
RFA_12
RFA_13
RFA_14
RFA_15
RFA_16
RFA_17
RFA_18
RFA_19
RFA_20
RFA_21
RFA_22
RFA_23

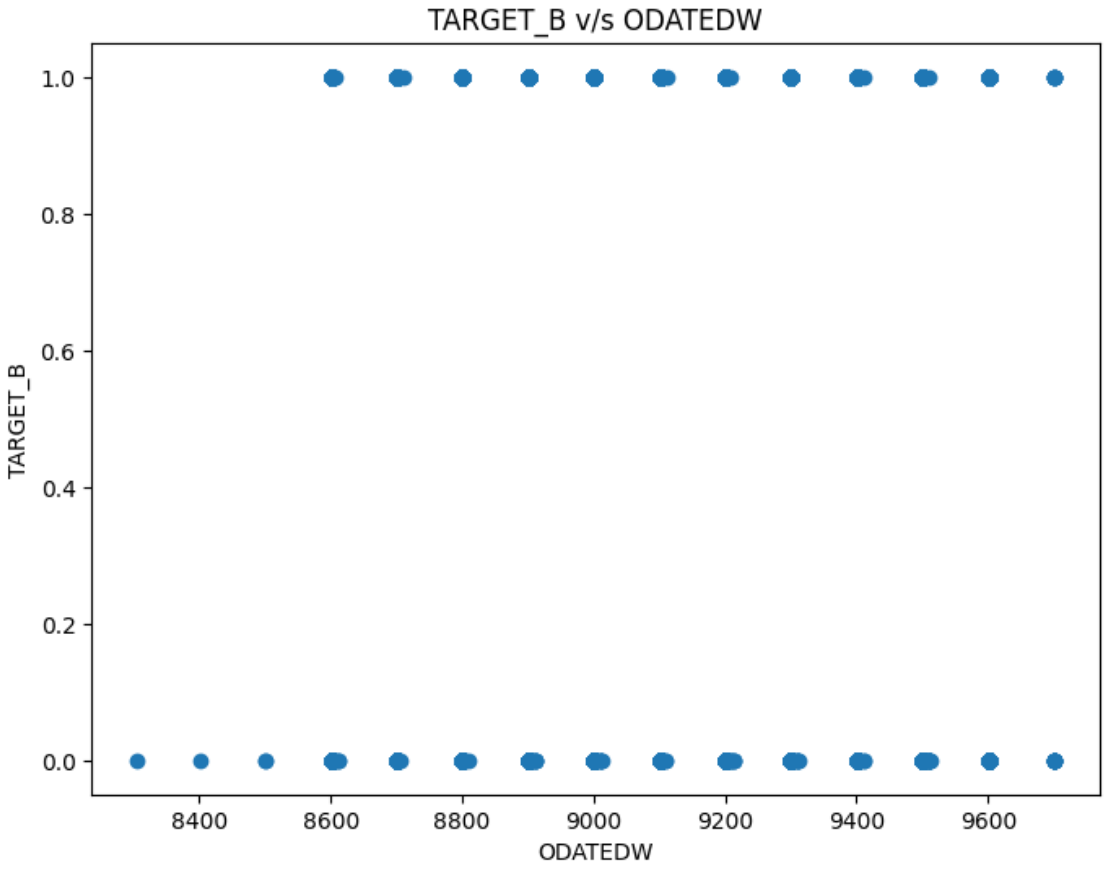
```
RFA_24
CARDPROM
MAXADATE
NUMPROM
CARDPM12
NUMPRM12
RAMNTALL
NGIFTALL
CARDGIFT
MINRAMNT
MINRDATE
MAXRAMNT
MAXRDATE
LASTGIFT
LASTDATE
FISTDATE
NEXTDATE
TIMELAG
AVGGIFT
CONTROLN
TARGET_B
TARGET_D
HPHONE_D
RFA_2R
RFA_2F
RFA_2A
MDMAUD_R
MDMAUD_F
MDMAUD_A
CLUSTER2
GEOCODE2
```

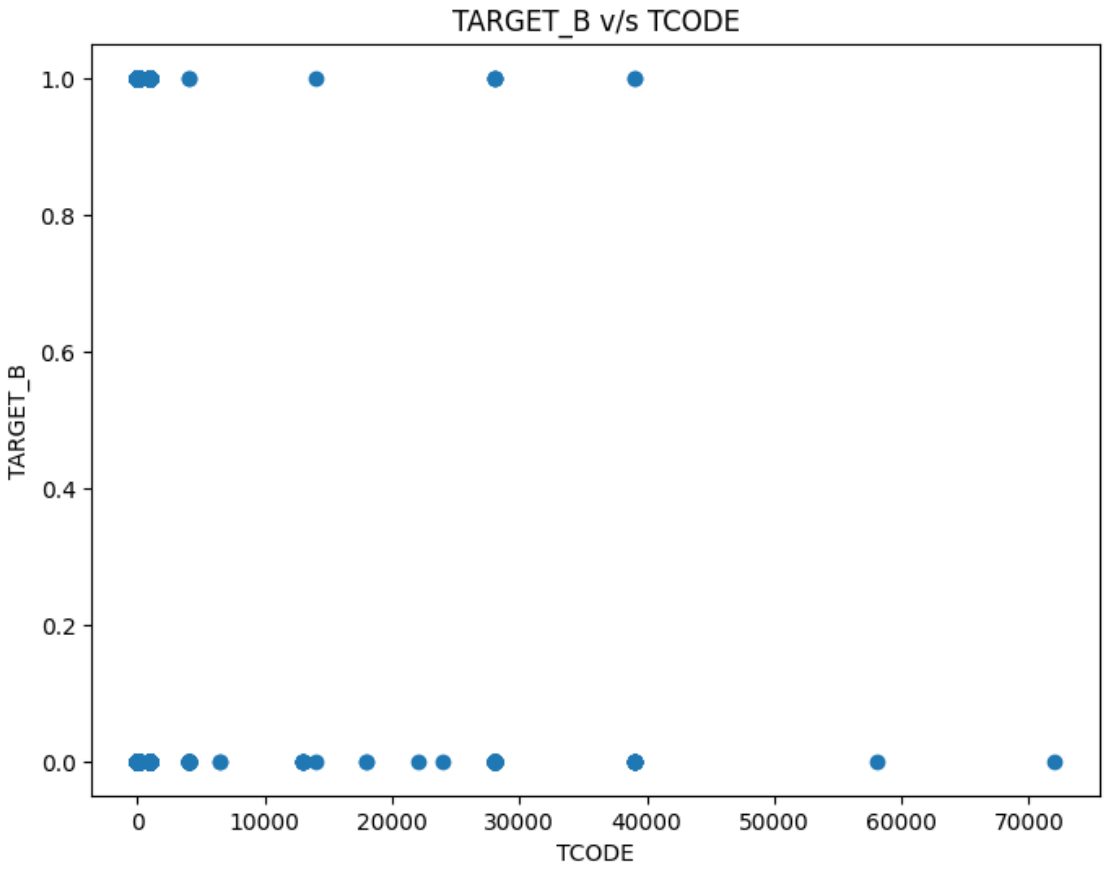
Creating a scatter plot of variables

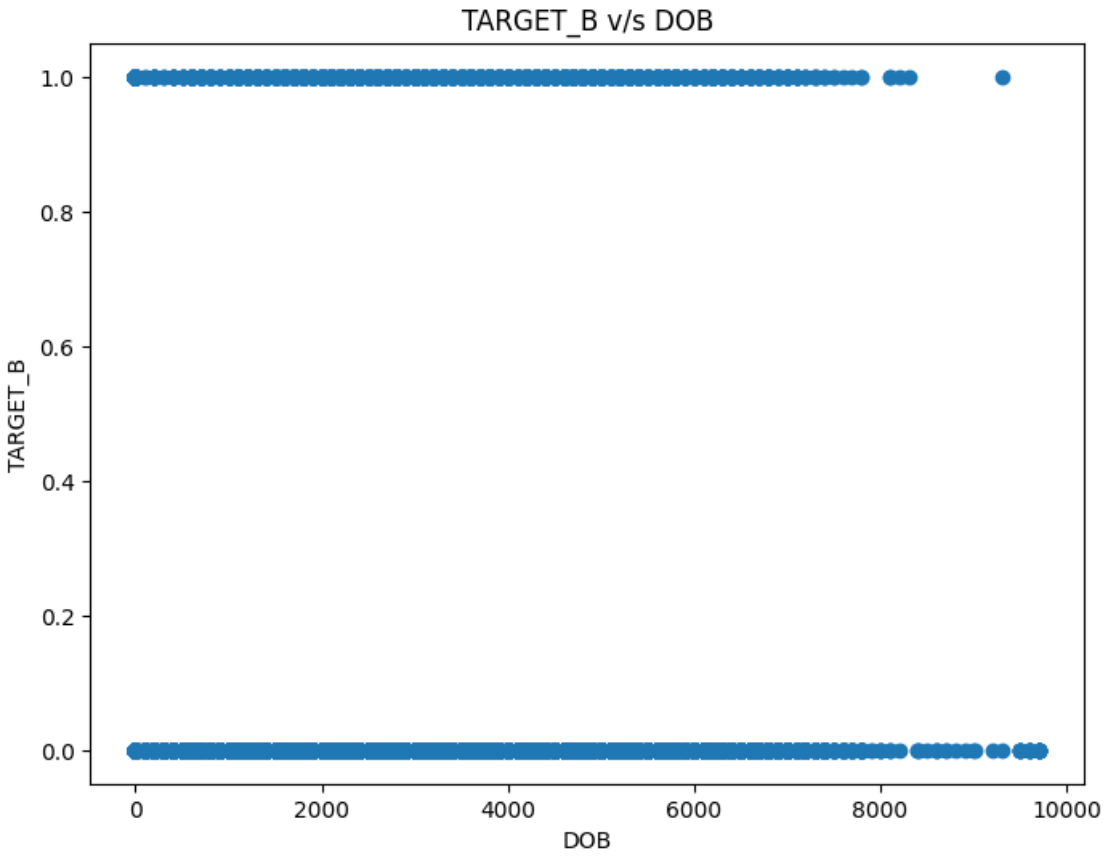
[ ]:
```
target_columns = ['TARGET_B', 'TARGET_D']
numeric_columns = []
for col in train_filtered.columns:
    if train_filtered[col].dtype != 'object':
        numeric_columns.append(col)
```

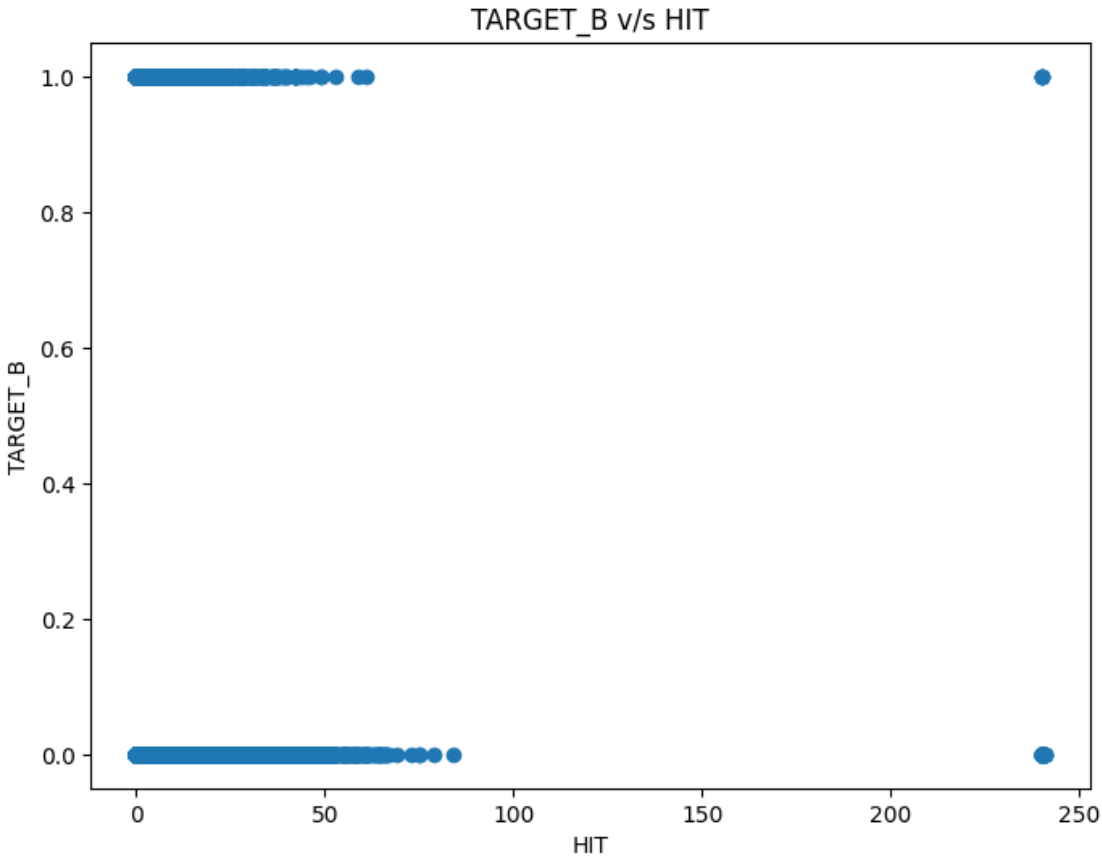
[ ]:
```
i = 0
for target_col in target_columns:
    for numeric_col in numeric_columns:
        plt.figure(figsize=(8, 6))
        plt.scatter(train_filtered[numeric_col], train_filtered[target_col])
        plt.xlabel(numeric_col)
        plt.ylabel(target_col)
        plt.title(f"{target_col} v/s {numeric_col}")
```

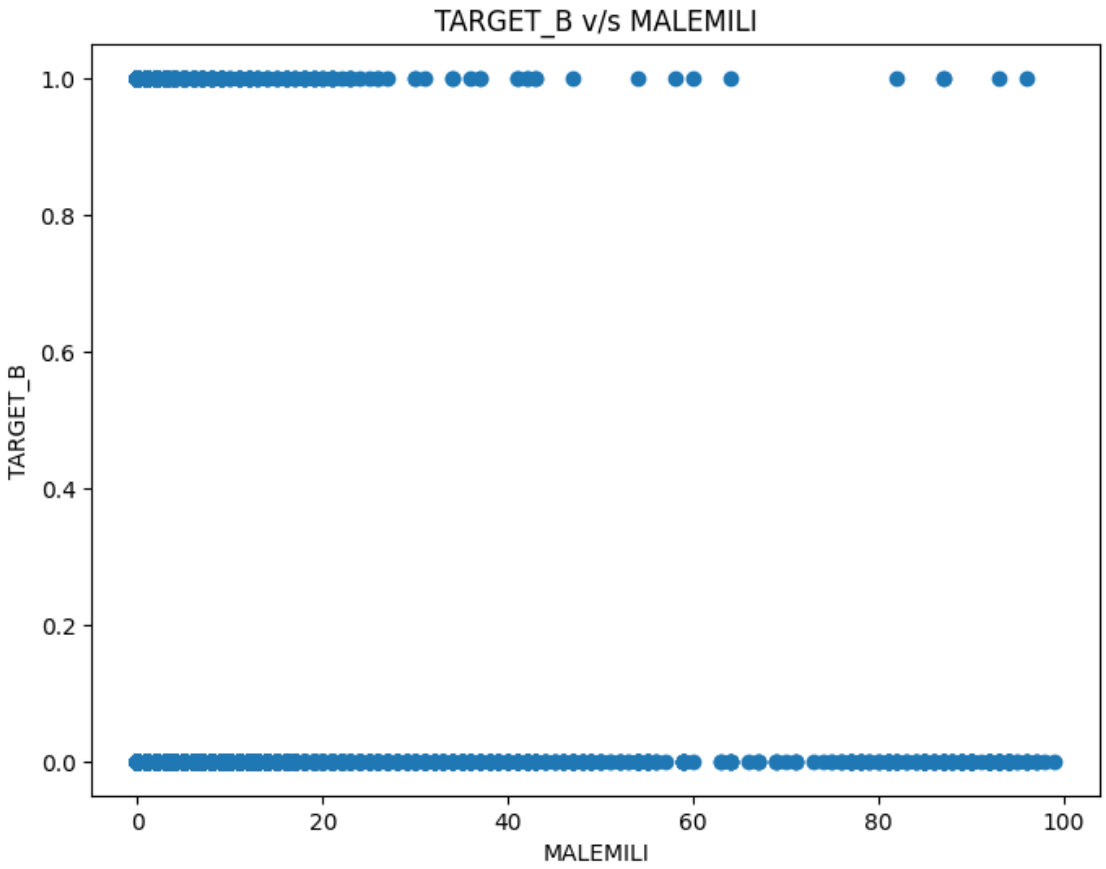
```
        i += 1
        if i > 10:
            break
```

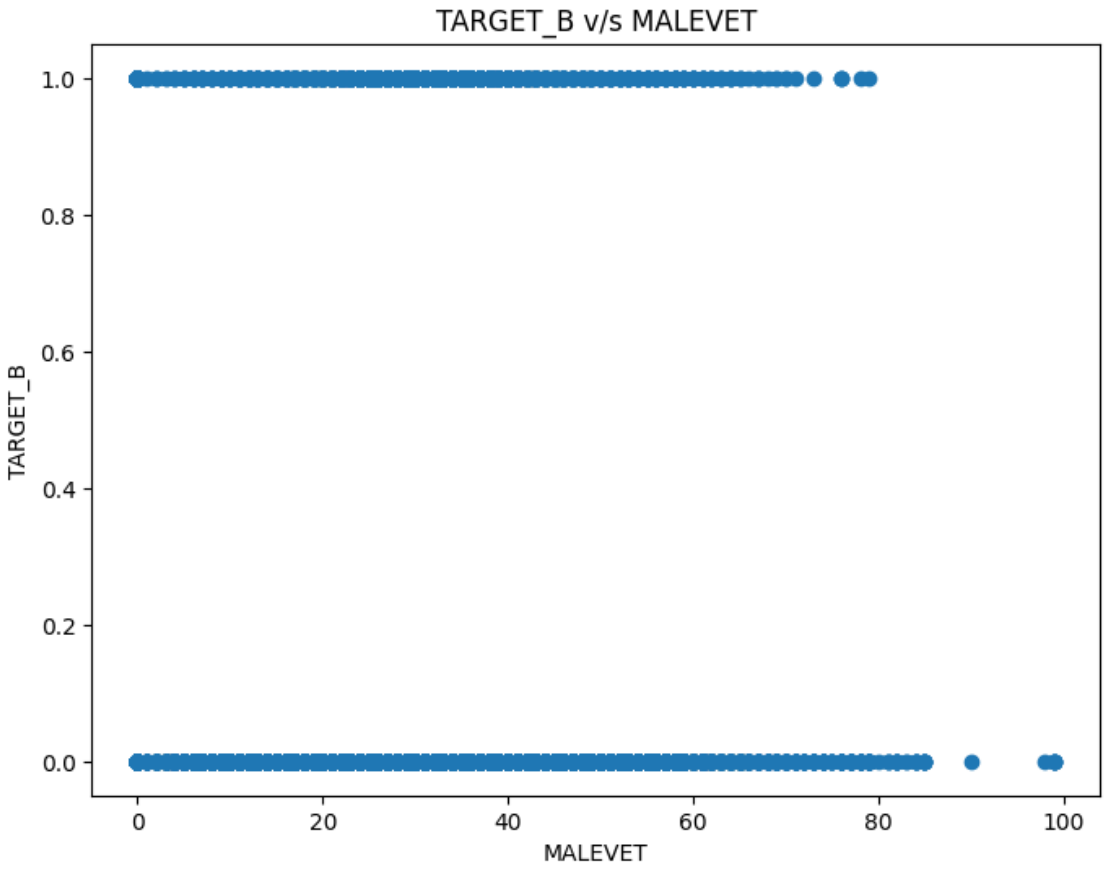


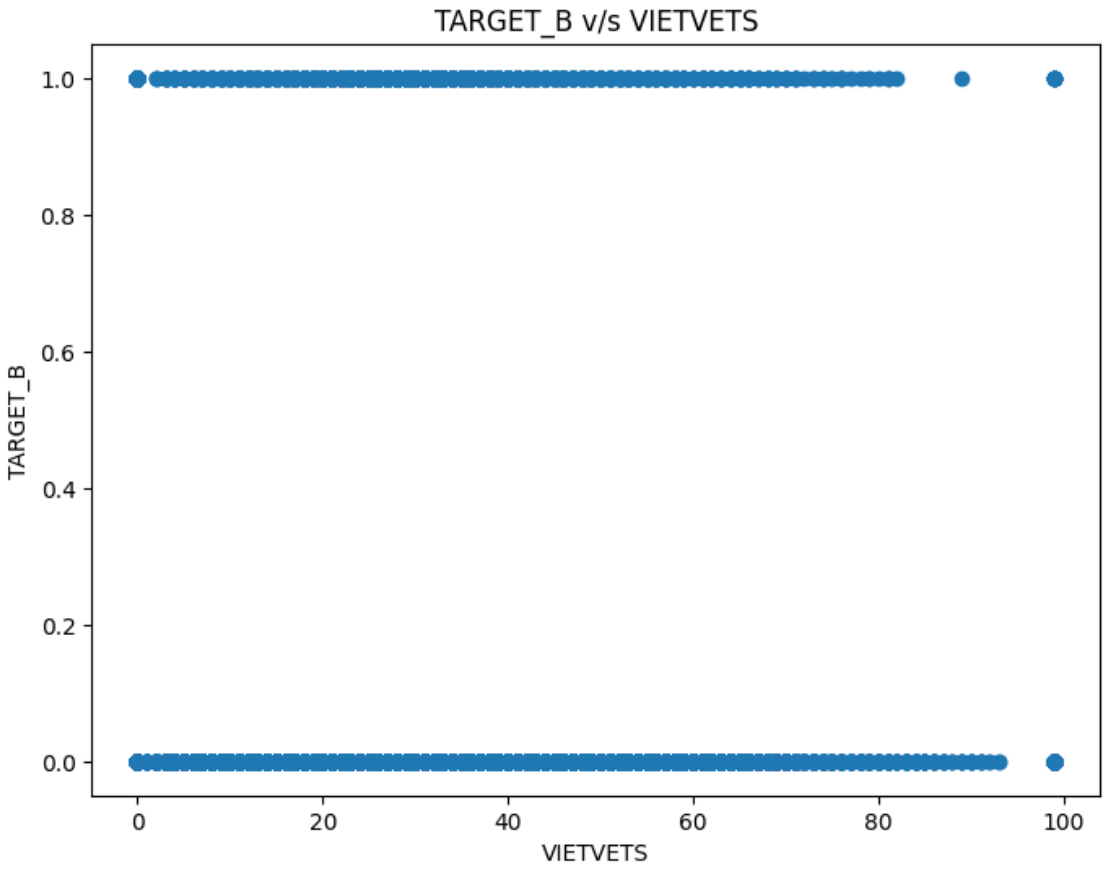

TARGET_B v/s TCODE
TARGET_B
TCODE
1.0
0.8
0.6
0.4
0.2
0.0
0
10000
20000
30000
40000
50000
60000
70000

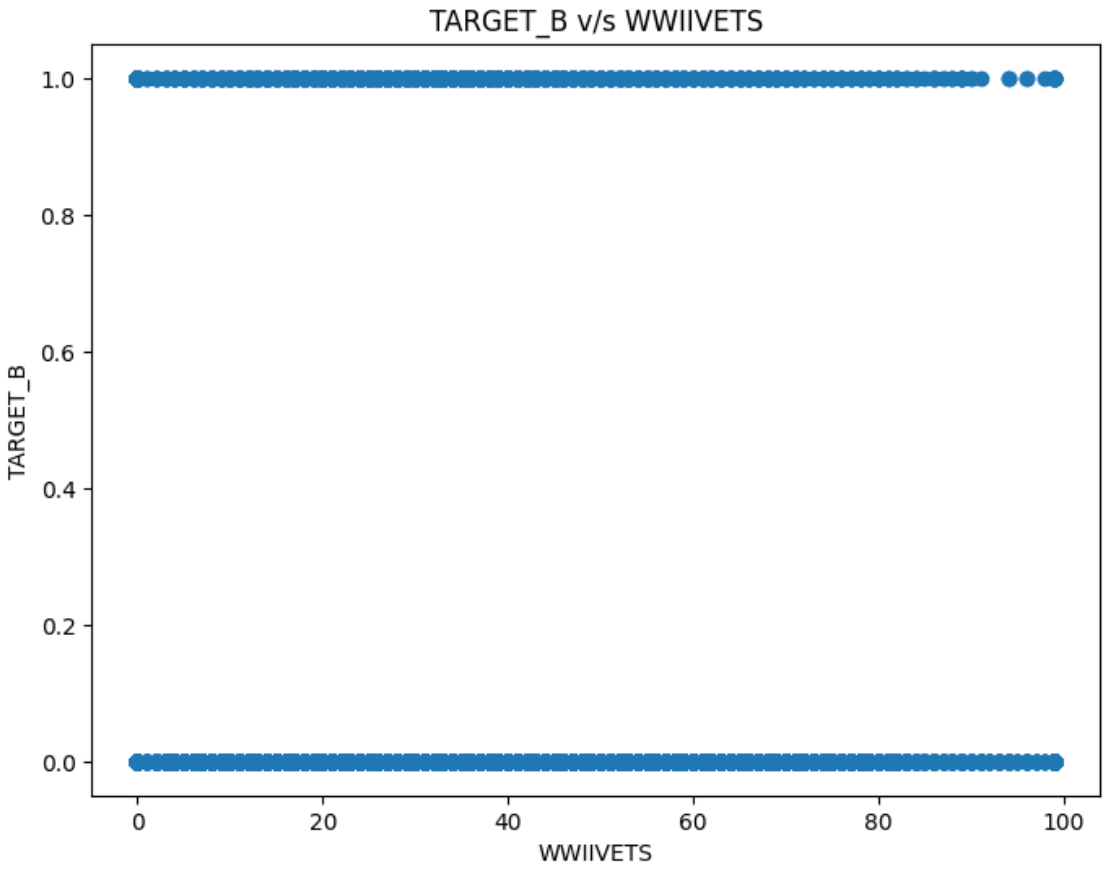

TARGET_B v/s DOB
TARGET_B
1.0
0.8
0.6
0.4
0.2
0.0
0
2000
4000
6000
8000
10000
DOB

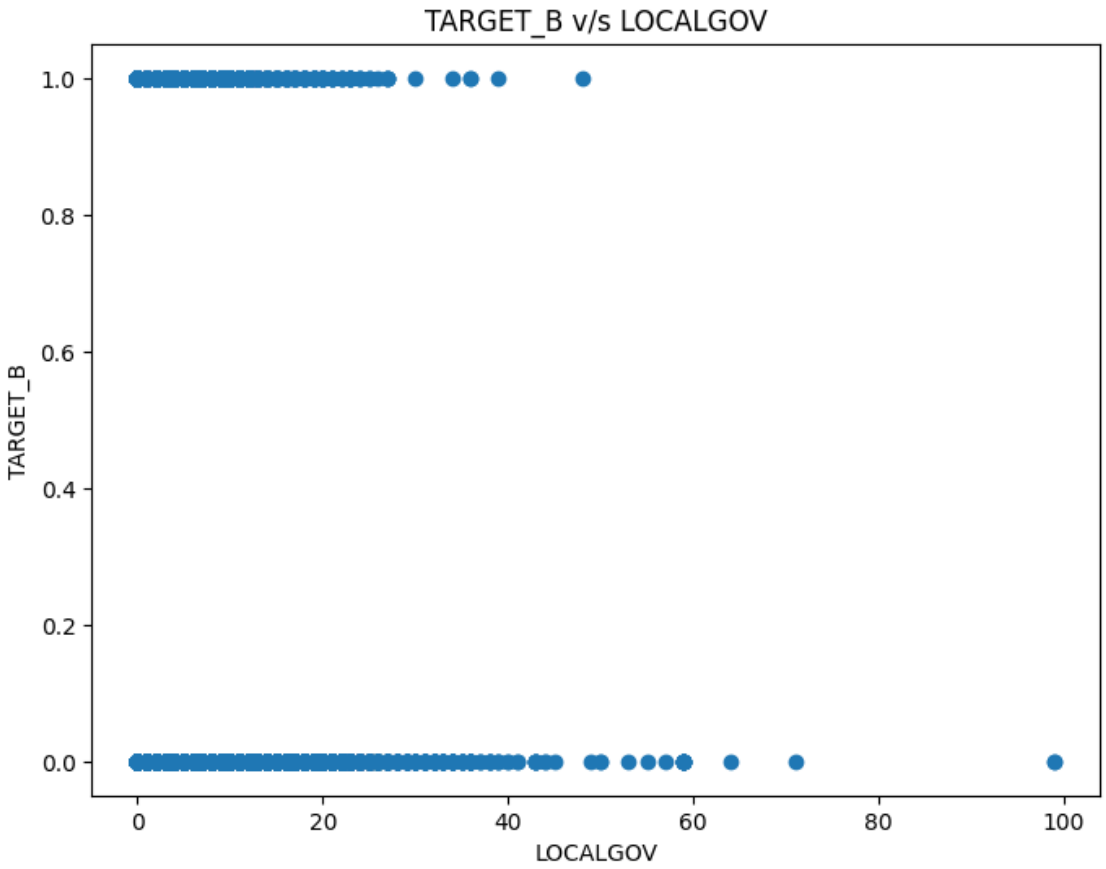

TARGET_B v/s HIT
TARGET_B
HIT
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
250

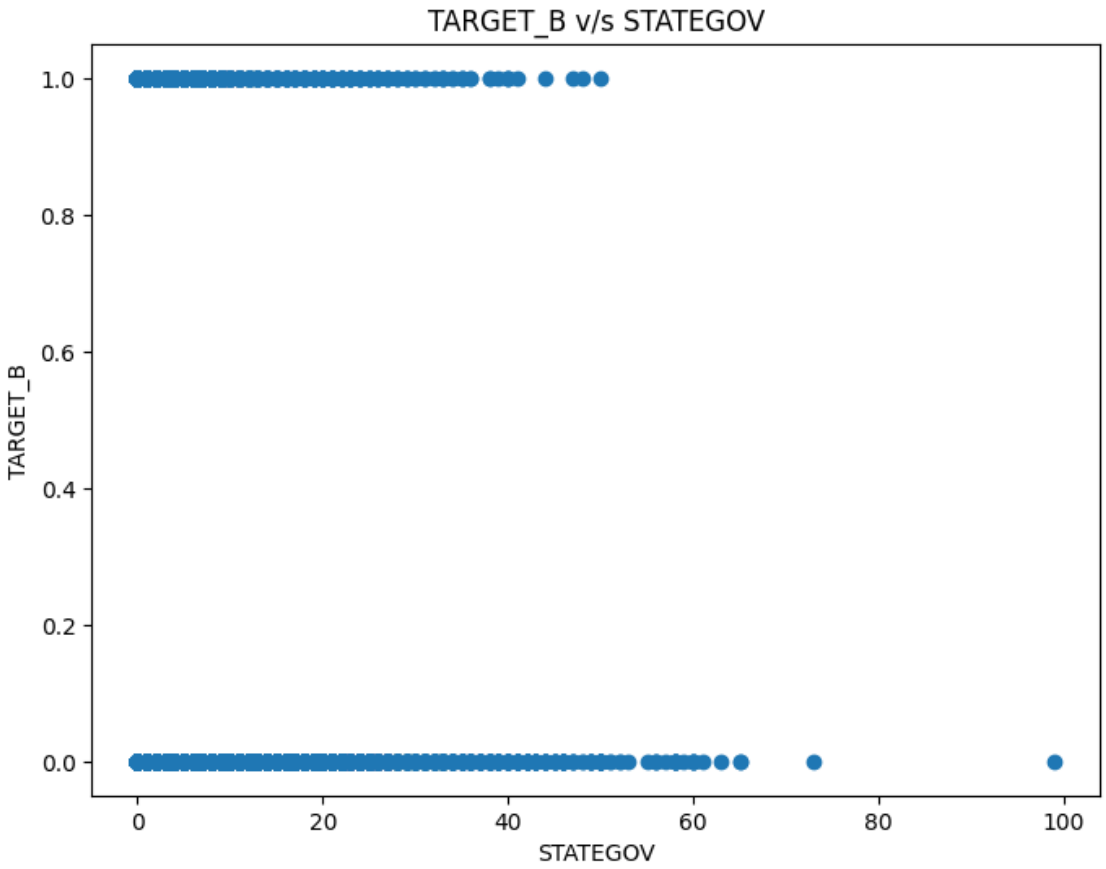

TARGET_B v/s MALEMILI
TARGET_B
MALEMILI
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100

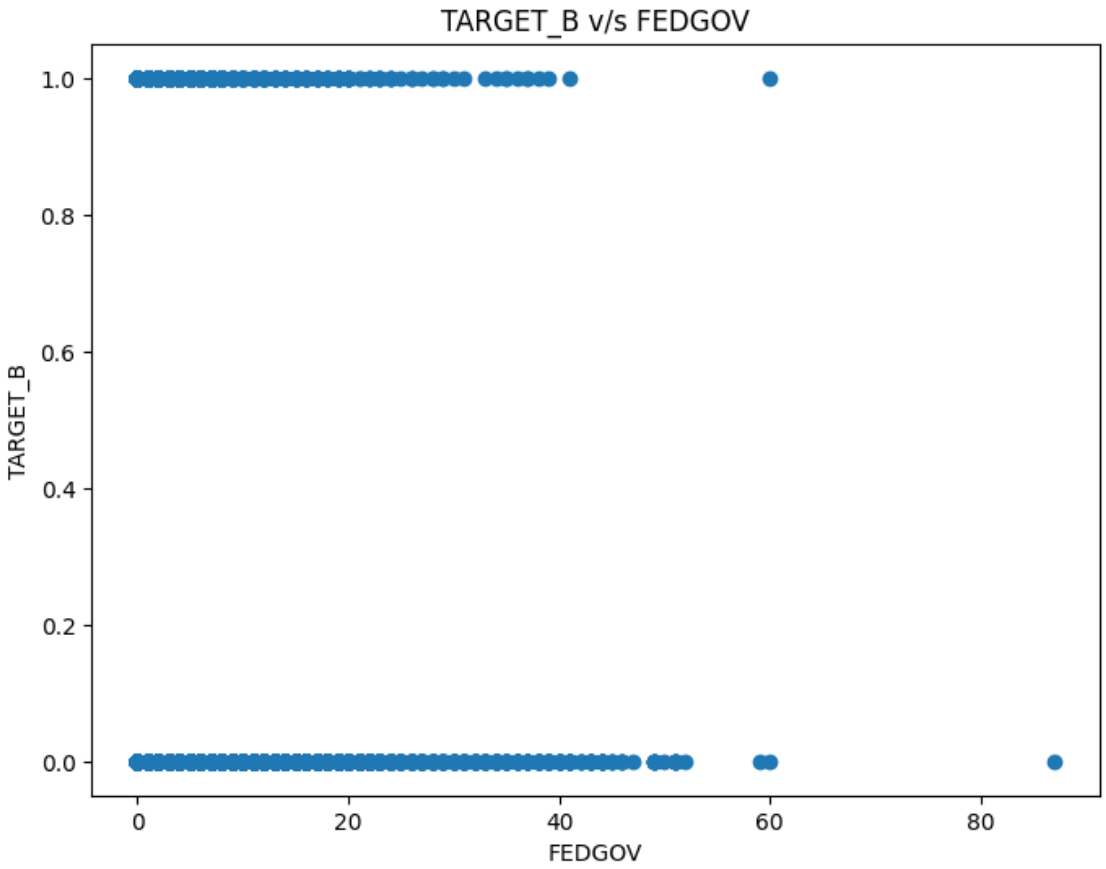

TARGET_B v/s MALEVET
TARGET_B
MALEVET
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100

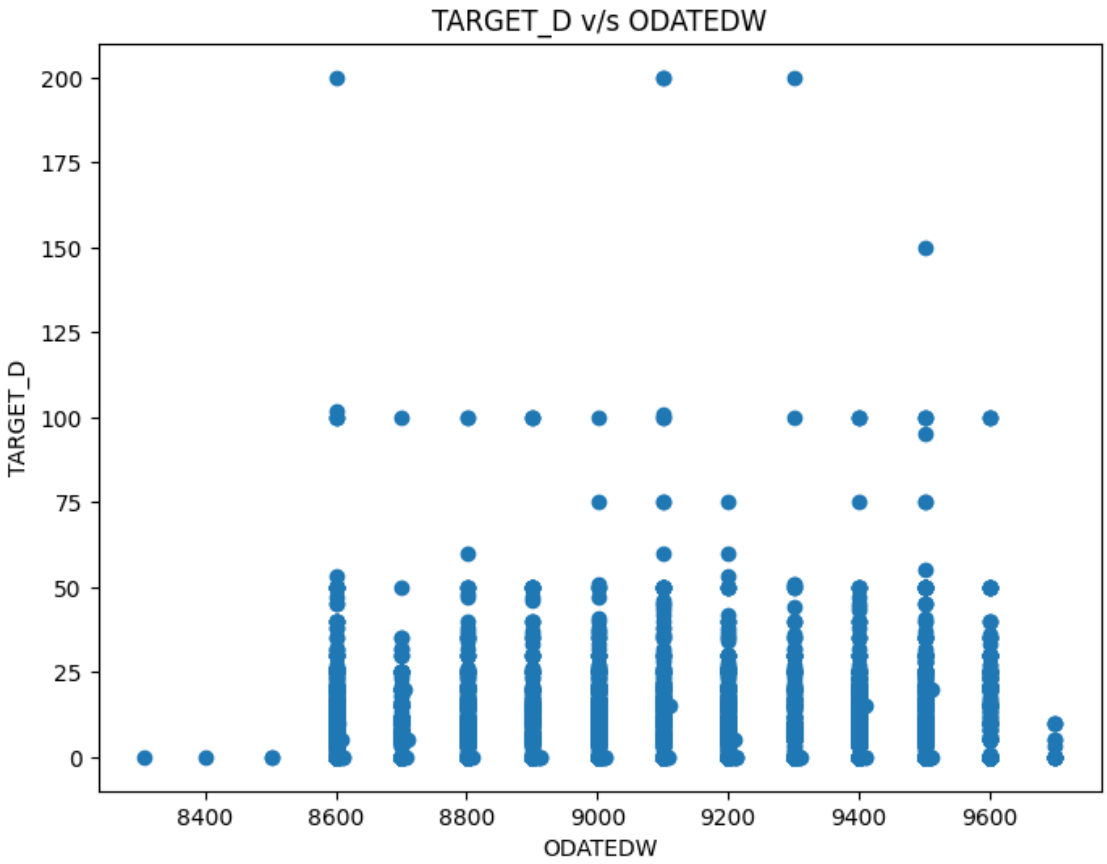

TARGET_B v/s VIETVETS
TARGET_B
VIETVETS
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100


TARGET_B v/s WWIIVETS
TARGET_B
WWIIVETS
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100


TARGET_B v/s LOCALGOV
TARGET_B
LOCALGOV
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100


TARGET_B v/s STATEGOV
TARGET_B
STATEGOV
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100


TARGET_B v/s FEDGOV
TARGET_B
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
FEDGOV



14. Drop unessential columns (feature selection).

[ ]:
```
train_filtered = train.drop(['LOCALGOV', 'STATEGOV', 'FEDGOV'], axis=1)
test_filtered = test.drop(['LOCALGOV', 'STATEGOV', 'FEDGOV'], axis=1)
```

[ ]:
```
train_filtered.columns.shape
```

[ ]: (478,)

[ ]:
```
test_filtered.columns.shape
```

[ ]: (476,)

0.1.1 Normalize and standardize all the columns

[ ]:
```
from sklearn.preprocessing import StandardScaler, MinMaxScaler
```

[ ]:
```
minmax_scaler = MinMaxScaler()
standard_scaler = StandardScaler()
```

[ ]:
```
train_filtered.mean()
```

```
/home/oneautumleaf/tmp/ipykernel_3555/1501147903.py:1: FutureWarning: The
default value of numeric_only in DataFrame.mean is deprecated. In a future
version, it will default to False. In addition, specifying 'numeric_only=None'
is deprecated. Select only valid columns or specify the value of numeric_only to
silence this warning.
  train_filtered.mean()
```

[ ]:
```
ODATEDW     9141.363256
TCODE         54.223117
DOB         2723.602933
AGE           61.611649
NUMCHLD        1.527773
               …
TARGET_B       0.050759
TARGET_D       0.793073
HPHONE_D       0.500618
RFA_2F         1.910053
CLUSTER2      31.533711
Length: 404, dtype: float64
```

[ ]:
```
train_filtered.std()
```

```
/home/oneautumleaf/tmp/ipykernel_3555/2403525054.py:1: FutureWarning: The
default value of numeric_only in DataFrame.std is deprecated. In a future
version, it will default to False. In addition, specifying 'numeric_only=None'
is deprecated. Select only valid columns or specify the value of numeric_only to
silence this warning.
  train_filtered.std()
```

[ ]:
```
ODATEDW      343.454752
TCODE        953.844476
DOB         2132.241295
AGE           16.664199
NUMCHLD        0.806861
               …
TARGET_B       0.219506
TARGET_D       4.429725
HPHONE_D       0.500002
RFA_2F         1.072749
CLUSTER2      18.764614
Length: 404, dtype: float64
```

[ ]:
```
numeric_columns = train_filtered.select_dtypes(include='number').columns
train_filtered[numeric_columns] = minmax_scaler.fit_transform(
    train_filtered[numeric_columns])
```

```
train_filtered[numeric_columns] = standard_scaler.fit_transform(
    train_filtered[numeric_columns])
```

[ ]:
```
train_filtered.mean()
```

```
/home/oneautumleaf/tmp/ipykernel_3555/1501147903.py:1: FutureWarning: The
default value of numeric_only in DataFrame.mean is deprecated. In a future
version, it will default to False. In addition, specifying 'numeric_only=None'
is deprecated. Select only valid columns or specify the value of numeric_only to
silence this warning.
  train_filtered.mean()
```

[ ]:
```
ODATEDW    -4.888277e-16
TCODE       9.383346e-18
DOB        -3.663973e-17
AGE         3.614759e-16
NUMCHLD     4.216445e-17
               …
TARGET_B    7.193899e-17
TARGET_D   -7.447100e-18
HPHONE_D    2.919263e-17
RFA_2F      5.756609e-17
CLUSTER2    2.053773e-16
Length: 404, dtype: float64
```

[ ]:
```
train_filtered.std()
```

```
/home/oneautumleaf/tmp/ipykernel_3555/2403525054.py:1: FutureWarning: The
default value of numeric_only in DataFrame.std is deprecated. In a future
version, it will default to False. In addition, specifying 'numeric_only=None'
is deprecated. Select only valid columns or specify the value of numeric_only to
silence this warning.
  train_filtered.std()
```

[ ]:
```
ODATEDW     1.000005
TCODE       1.000005
DOB         1.000005
AGE         1.000007
NUMCHLD     1.000040
              …
TARGET_B    1.000005
TARGET_D    1.000005
HPHONE_D    1.000005
RFA_2F      1.000005
CLUSTER2    1.000005
Length: 404, dtype: float64
```

[ ]:
```
test_filtered.mean()
```

```
/home/oneautumleaf/tmp/ipykernel_3555/2421697746.py:1: FutureWarning: The
default value of numeric_only in DataFrame.mean is deprecated. In a future
version, it will default to False. In addition, specifying 'numeric_only=None'
is deprecated. Select only valid columns or specify the value of numeric_only to
silence this warning.
  test_filtered.mean()
```

[ ]:
```
ODATEDW      9142.930879
TCODE          50.261656
DOB          2716.428892
AGE            61.646625
NUMCHLD         1.533814
                 …
AVGGIFT        13.351390
CONTROLN    96000.714871
HPHONE_D        0.502299
RFA_2F          1.906192
CLUSTER2       31.537319
Length: 402, dtype: float64
```

[ ]:
```
test_filtered.std()
```

```
/home/oneautumleaf/tmp/ipykernel_3555/3047912150.py:1: FutureWarning: The
default value of numeric_only in DataFrame.std is deprecated. In a future
version, it will default to False. In addition, specifying 'numeric_only=None'
is deprecated. Select only valid columns or specify the value of numeric_only to
silence this warning.
  test_filtered.std()
```

[ ]:
```
ODATEDW       341.841721
TCODE         849.009266
DOB          2131.905959
AGE            16.666172
NUMCHLD         0.814021
                 …
AVGGIFT         9.470951
CONTROLN    55438.541281
HPHONE_D        0.499997
RFA_2F          1.072086
CLUSTER2       18.795616
Length: 402, dtype: float64
```

[ ]:
```
numeric_columns = test_filtered.select_dtypes(include='number').columns
test_filtered[numeric_columns] = minmax_scaler.fit_transform(
    test_filtered[numeric_columns])
```

```
test_filtered[numeric_columns] = standard_scaler.fit_transform(
    test_filtered[numeric_columns])
```

[ ]:
```
test_filtered.mean()
```

```
/home/oneautumleaf/tmp/ipykernel_3555/2421697746.py:1: FutureWarning: The
default value of numeric_only in DataFrame.mean is deprecated. In a future
version, it will default to False. In addition, specifying 'numeric_only=None'
is deprecated. Select only valid columns or specify the value of numeric_only to
silence this warning.
  test_filtered.mean()
```

[ ]:
```
ODATEDW     3.177155e-16
TCODE       4.423980e-18
DOB        -4.840571e-17
AGE        -2.469924e-16
NUMCHLD     1.281996e-17
               …
AVGGIFT    -2.248856e-17
CONTROLN   -1.781389e-16
HPHONE_D    4.129048e-18
RFA_2F     -2.211990e-17
CLUSTER2   -4.433186e-16
Length: 402, dtype: float64
```

[ ]:
```
test_filtered.std()
```

```
/home/oneautumleaf/tmp/ipykernel_3555/3047912150.py:1: FutureWarning: The
default value of numeric_only in DataFrame.std is deprecated. In a future
version, it will default to False. In addition, specifying 'numeric_only=None'
is deprecated. Select only valid columns or specify the value of numeric_only to
silence this warning.
  test_filtered.std()
```

[ ]:
```
ODATEDW     1.000005
TCODE       1.000005
DOB         1.000005
AGE         1.000007
NUMCHLD     1.000041
              …
AVGGIFT     1.000005
CONTROLN    1.000005
HPHONE_D    1.000005
RFA_2F      1.000005
CLUSTER2    1.000005
Length: 402, dtype: float64
```

17. Display unique values in each column

[ ]:
```
for column in train_filtered.columns:
    print(f"{column}: {len(train_filtered[column].unique())}")
```

```
ODATEDW: 54
OSOURCE: 896
TCODE: 55
STATE: 57
ZIP: 19938
MAILCODE: 2
PVASTATE: 3
DOB: 947
NOEXCH: 6
RECINHSE: 2
RECP3: 2
RECPGVG: 2
RECSWEEP: 2
MDMAUD: 28
DOMAIN: 17
CLUSTER: 54
AGE: 97
AGEFLAG: 3
HOMEOWNR: 3
CHILD03: 4
CHILD07: 4
CHILD12: 4
CHILD18: 4
NUMCHLD: 8
INCOME: 8
GENDER: 7
WEALTH1: 11
HIT: 75
MBCRAFT: 8
MBGARDEN: 6
MBBOOKS: 11
MBCOLECT: 8
MAGFAML: 10
MAGFEM: 7
MAGMALE: 6
PUBGARDN: 7
PUBCULIN: 8
PUBHLTH: 11
PUBDOITY: 10
PUBNEWFN: 11
PUBPHOTO: 4
PUBOPP: 11
DATASRCE: 4
MALEMILI: 95
```

MALEVET: 89
VIETVETS: 95
WWIIVETS: 100
SOLP3: 5
SOLIH: 8
MAJOR: 2
WEALTH2: 11
GEOCODE: 8
COLLECT1: 2
VETERANS: 2
BIBLE: 2
CATLG: 2
HOMEE: 2
PETS: 2
CDPLAY: 2
STEREO: 2
PCOWNERS: 2
PHOTO: 2
CRAFTS: 2
FISHER: 2
GARDENIN: 2
BOATS: 2
WALKER: 2
KIDSTUFF: 2
CARDS: 2
PLATES: 2
LIFESRC: 4
PEPSTRFL: 2
POP901: 9906
POP902: 4786
POP903: 5698
POP90C1: 100
POP90C2: 100
POP90C3: 100
POP90C4: 81
POP90C5: 81
ETH1: 100
ETH2: 100
ETH3: 85
ETH4: 96
ETH5: 100
ETH6: 21
ETH7: 61
ETH8: 60
ETH9: 61
ETH10: 33
ETH11: 34
ETH12: 43

ETH13: 98
ETH14: 45
ETH15: 78
ETH16: 66
AGE901: 74
AGE902: 67
AGE903: 59
AGE904: 66
AGE905: 62
AGE906: 55
AGE907: 64
CHIL1: 85
CHIL2: 70
CHIL3: 82
AGEC1: 98
AGEC2: 74
AGEC3: 57
AGEC4: 43
AGEC5: 43
AGEC6: 63
AGEC7: 77
CHILC1: 62
CHILC2: 46
CHILC3: 62
CHILC4: 56
CHILC5: 97
HHAGE1: 99
HHAGE2: 85
HHAGE3: 99
HHN1: 98
HHN2: 85
HHN3: 96
HHN4: 86
HHN5: 69
HHN6: 53
MARR1: 95
MARR2: 51
MARR3: 60
MARR4: 100
HHP1: 393
HHP2: 377
DW1: 100
DW2: 100
DW3: 70
DW4: 100
DW5: 100
DW6: 100
DW7: 97

DW8: 86
DW9: 91
HV1: 4434
HV2: 4623
HV3: 14
HV4: 14
HU1: 100
HU2: 100
HU3: 93
HU4: 94
HU5: 100
HHD1: 95
HHD2: 100
HHD3: 99
HHD4: 90
HHD5: 100
HHD6: 100
HHD7: 66
HHD8: 17
HHD9: 64
HHD10: 83
HHD11: 82
HHD12: 66
ETHC1: 52
ETHC2: 95
ETHC3: 100
ETHC4: 45
ETHC5: 78
ETHC6: 53
HVP1: 100
HVP2: 100
HVP3: 100
HVP4: 100
HVP5: 100
HVP6: 100
HUR1: 96
HUR2: 100
RHP1: 82
RHP2: 81
RHP3: 27
RHP4: 25
HUPA1: 96
HUPA2: 100
HUPA3: 100
HUPA4: 97
HUPA5: 77
HUPA6: 100
HUPA7: 48

RP1: 100
RP2: 100
RP3: 100
RP4: 100
MSA: 299
ADI: 205
DMA: 207
IC1: 1134
IC2: 1213
IC3: 1091
IC4: 1156
IC5: 21514
IC6: 99
IC7: 65
IC8: 60
IC9: 66
IC10: 67
IC11: 50
IC12: 36
IC13: 31
IC14: 71
IC15: 96
IC16: 78
IC17: 72
IC18: 71
IC19: 73
IC20: 54
IC21: 43
IC22: 36
IC23: 75
HHAS1: 100
HHAS2: 72
HHAS3: 99
HHAS4: 93
MC1: 99
MC2: 99
MC3: 98
TPE1: 99
TPE2: 70
TPE3: 72
TPE4: 67
TPE5: 50
TPE6: 27
TPE7: 18
TPE8: 83
TPE9: 56
PEC1: 82
PEC2: 99

TPE10: 67
TPE11: 62
TPE12: 59
TPE13: 100
LFC1: 100
LFC2: 100
LFC3: 99
LFC4: 100
LFC5: 98
LFC6: 98
LFC7: 100
LFC8: 98
LFC9: 96
LFC10: 91
OCC1: 70
OCC2: 61
OCC3: 35
OCC4: 60
OCC5: 63
OCC6: 27
OCC7: 40
OCC8: 63
OCC9: 69
OCC10: 61
OCC11: 61
OCC12: 48
OCC13: 44
EIC1: 75
EIC2: 54
EIC3: 51
EIC4: 73
EIC5: 45
EIC6: 34
EIC7: 43
EIC8: 71
EIC9: 50
EIC10: 44
EIC11: 50
EIC12: 41
EIC13: 56
EIC14: 69
EIC15: 58
EIC16: 51
OEDC1: 54
OEDC2: 63
OEDC3: 54
OEDC4: 66
OEDC5: 91

OEDC6: 63
OEDC7: 25
EC1: 79
EC2: 76
EC3: 63
EC4: 80
EC5: 69
EC6: 36
EC7: 61
EC8: 71
SEC1: 83
SEC2: 97
SEC3: 19
SEC4: 54
SEC5: 97
AFC1: 81
AFC2: 96
AFC3: 36
AFC4: 57
AFC5: 88
AFC6: 27
VC1: 95
VC2: 89
VC3: 100
VC4: 97
ANC1: 68
ANC2: 54
ANC3: 31
ANC4: 82
ANC5: 26
ANC6: 14
ANC7: 43
ANC8: 42
ANC9: 52
ANC10: 60
ANC11: 33
ANC12: 30
ANC13: 19
ANC14: 26
ANC15: 17
POBC1: 88
POBC2: 100
LSC1: 100
LSC2: 100
LSC3: 71
LSC4: 79
VOC1: 93
VOC2: 100

VOC3: 78
HC1: 32
HC2: 53
HC3: 83
HC4: 100
HC5: 100
HC6: 100
HC7: 100
HC8: 100
HC9: 87
HC10: 53
HC11: 100
HC12: 92
HC13: 100
HC14: 98
HC15: 16
HC16: 92
HC17: 100
HC18: 100
HC19: 100
HC20: 61
HC21: 77
MHUC1: 22
MHUC2: 6
AC1: 37
AC2: 39
ADATE_2: 2
ADATE_3: 3
ADATE_4: 9
ADATE_5: 2
ADATE_6: 3
ADATE_7: 4
ADATE_8: 6
ADATE_9: 4
ADATE_10: 3
ADATE_11: 5
ADATE_12: 5
ADATE_13: 4
ADATE_14: 3
ADATE_15: 2
ADATE_16: 4
ADATE_17: 4
ADATE_18: 10
ADATE_19: 4
ADATE_20: 3
ADATE_21: 3
ADATE_22: 6
ADATE_23: 4

ADATE_24: 3
RFA_2: 14
RFA_3: 71
RFA_4: 64
RFA_5: 41
RFA_6: 109
RFA_7: 106
RFA_8: 109
RFA_9: 107
RFA_10: 94
RFA_11: 101
RFA_12: 107
RFA_13: 87
RFA_14: 95
RFA_15: 34
RFA_16: 123
RFA_17: 118
RFA_18: 122
RFA_19: 108
RFA_20: 80
RFA_21: 102
RFA_22: 117
RFA_23: 87
RFA_24: 97
CARDPROM: 59
MAXADATE: 5
NUMPROM: 165
CARDPM12: 19
NUMPRM12: 63
RDATE_3: 15
RDATE_4: 22
RDATE_5: 6
RDATE_6: 18
RDATE_7: 10
RDATE_8: 15
RDATE_9: 11
RDATE_10: 9
RDATE_11: 13
RDATE_12: 11
RDATE_13: 15
RDATE_14: 13
RDATE_15: 17
RDATE_16: 19
RDATE_17: 12
RDATE_18: 15
RDATE_19: 14
RDATE_20: 11
RDATE_21: 13

RDATE_22: 14
RDATE_23: 18
RDATE_24: 15
RAMNT_3: 30
RAMNT_4: 33
RAMNT_5: 10
RAMNT_6: 41
RAMNT_7: 76
RAMNT_8: 109
RAMNT_9: 91
RAMNT_10: 93
RAMNT_11: 83
RAMNT_12: 115
RAMNT_13: 78
RAMNT_14: 103
RAMNT_15: 77
RAMNT_16: 110
RAMNT_17: 75
RAMNT_18: 106
RAMNT_19: 83
RAMNT_20: 70
RAMNT_21: 78
RAMNT_22: 93
RAMNT_23: 77
RAMNT_24: 76
RAMNTALL: 2094
NGIFTALL: 89
CARDGIFT: 37
MINRAMNT: 191
MINRDATE: 146
MAXRAMNT: 275
MAXRDATE: 150
LASTGIFT: 231
LASTDATE: 24
FISTDATE: 177
NEXTDATE: 189
TIMELAG: 69
AVGGIFT: 7713
CONTROLN: 95412
TARGET_B: 2
TARGET_D: 71
HPHONE_D: 2
RFA_2R: 1
RFA_2F: 4
RFA_2A: 4
MDMAUD_R: 5
MDMAUD_F: 4
MDMAUD_A: 5

```
CLUSTER2: 63
GEOCODE2: 6
```

[ ]:
```
for column in test_filtered.columns:
    print(f"{column}: {len(test_filtered[column].unique())}")
```

```
ODATEDW: 56
OSOURCE: 898
TCODE: 54
STATE: 59
ZIP: 19887
MAILCODE: 2
PVASTATE: 3
DOB: 941
NOEXCH: 6
RECINHSE: 2
RECP3: 2
RECPGVG: 2
RECSWEEP: 2
MDMAUD: 30
DOMAIN: 17
CLUSTER: 54
AGE: 94
AGEFLAG: 3
HOMEOWNR: 3
CHILD03: 4
CHILD07: 4
CHILD12: 4
CHILD18: 4
NUMCHLD: 8
INCOME: 8
GENDER: 5
WEALTH1: 11
HIT: 82
MBCRAFT: 8
MBGARDEN: 5
MBBOOKS: 11
MBCOLECT: 8
MAGFAML: 11
MAGFEM: 6
MAGMALE: 6
PUBGARDN: 8
PUBCULIN: 6
PUBHLTH: 11
PUBDOITY: 11
PUBNEWFN: 11
PUBPHOTO: 4
PUBOPP: 11
```

DATASRCE: 4
MALEMILI: 95
MALEVET: 91
VIETVETS: 95
WWIIVETS: 100
SOLP3: 5
SOLIH: 9
MAJOR: 2
WEALTH2: 11
GEOCODE: 8
COLLECT1: 2
VETERANS: 2
BIBLE: 2
CATLG: 2
HOMEE: 2
PETS: 2
CDPLAY: 2
STEREO: 2
PCOWNERS: 2
PHOTO: 2
CRAFTS: 2
FISHER: 2
GARDENIN: 2
BOATS: 2
WALKER: 2
KIDSTUFF: 2
CARDS: 2
PLATES: 2
LIFESRC: 4
PEPSTRFL: 2
POP901: 9817
POP902: 4803
POP903: 5714
POP90C1: 100
POP90C2: 100
POP90C3: 100
POP90C4: 84
POP90C5: 84
ETH1: 100
ETH2: 100
ETH3: 84
ETH4: 95
ETH5: 100
ETH6: 21
ETH7: 59
ETH8: 61
ETH9: 59
ETH10: 34

ETH11: 34
ETH12: 46
ETH13: 97
ETH14: 42
ETH15: 79
ETH16: 63
AGE901: 73
AGE902: 67
AGE903: 59
AGE904: 66
AGE905: 62
AGE906: 55
AGE907: 60
CHIL1: 88
CHIL2: 69
CHIL3: 80
AGEC1: 95
AGEC2: 68
AGEC3: 57
AGEC4: 40
AGEC5: 44
AGEC6: 63
AGEC7: 78
CHILC1: 63
CHILC2: 49
CHILC3: 67
CHILC4: 55
CHILC5: 93
HHAGE1: 99
HHAGE2: 85
HHAGE3: 100
HHN1: 97
HHN2: 85
HHN3: 96
HHN4: 85
HHN5: 69
HHN6: 54
MARR1: 95
MARR2: 53
MARR3: 62
MARR4: 99
HHP1: 399
HHP2: 385
DW1: 100
DW2: 100
DW3: 71
DW4: 100
DW5: 100

DW6: 100
DW7: 96
DW8: 90
DW9: 87
HV1: 4416
HV2: 4575
HV3: 14
HV4: 14
HU1: 100
HU2: 100
HU3: 95
HU4: 96
HU5: 100
HHD1: 96
HHD2: 99
HHD3: 100
HHD4: 91
HHD5: 100
HHD6: 100
HHD7: 66
HHD8: 16
HHD9: 63
HHD10: 83
HHD11: 82
HHD12: 59
ETHC1: 52
ETHC2: 95
ETHC3: 100
ETHC4: 46
ETHC5: 80
ETHC6: 51
HVP1: 100
HVP2: 100
HVP3: 100
HVP4: 100
HVP5: 100
HVP6: 100
HUR1: 94
HUR2: 100
RHP1: 81
RHP2: 79
RHP3: 26
RHP4: 24
HUPA1: 98
HUPA2: 100
HUPA3: 100
HUPA4: 98
HUPA5: 83

HUPA6: 100
HUPA7: 50
RP1: 100
RP2: 100
RP3: 100
RP4: 100
MSA: 301
ADI: 205
DMA: 207
IC1: 1120
IC2: 1224
IC3: 1094
IC4: 1152
IC5: 21679
IC6: 100
IC7: 66
IC8: 59
IC9: 63
IC10: 71
IC11: 54
IC12: 40
IC13: 31
IC14: 71
IC15: 95
IC16: 71
IC17: 70
IC18: 72
IC19: 74
IC20: 60
IC21: 45
IC22: 39
IC23: 78
HHAS1: 100
HHAS2: 75
HHAS3: 100
HHAS4: 91
MC1: 99
MC2: 100
MC3: 94
TPE1: 99
TPE2: 69
TPE3: 74
TPE4: 71
TPE5: 49
TPE6: 29
TPE7: 17
TPE8: 84
TPE9: 55

PEC1: 83
PEC2: 99
TPE10: 72
TPE11: 65
TPE12: 63
TPE13: 99
LFC1: 99
LFC2: 100
LFC3: 99
LFC4: 100
LFC5: 98
LFC6: 98
LFC7: 100
LFC8: 99
LFC9: 96
LFC10: 92
OCC1: 70
OCC2: 63
OCC3: 38
OCC4: 61
OCC5: 59
OCC6: 25
OCC7: 37
OCC8: 71
OCC9: 73
OCC10: 56
OCC11: 56
OCC12: 42
OCC13: 44
EIC1: 79
EIC2: 55
EIC3: 51
EIC4: 74
EIC5: 41
EIC6: 34
EIC7: 41
EIC8: 69
EIC9: 51
EIC10: 45
EIC11: 51
EIC12: 40
EIC13: 58
EIC14: 64
EIC15: 61
EIC16: 58
OEDC1: 56
OEDC2: 64
OEDC3: 58

OEDC4: 64
OEDC5: 89
OEDC6: 58
OEDC7: 24
EC1: 78
EC2: 76
EC3: 66
EC4: 76
EC5: 63
EC6: 38
EC7: 64
EC8: 69
SEC1: 76
SEC2: 94
SEC3: 17
SEC4: 56
SEC5: 98
AFC1: 84
AFC2: 96
AFC3: 42
AFC4: 66
AFC5: 91
AFC6: 29
VC1: 96
VC2: 85
VC3: 100
VC4: 97
ANC1: 62
ANC2: 53
ANC3: 29
ANC4: 82
ANC5: 21
ANC6: 15
ANC7: 41
ANC8: 46
ANC9: 48
ANC10: 57
ANC11: 31
ANC12: 31
ANC13: 21
ANC14: 27
ANC15: 16
POBC1: 91
POBC2: 100
LSC1: 100
LSC2: 100
LSC3: 76
LSC4: 78

VOC1: 95
VOC2: 100
VOC3: 80
HC1: 32
HC2: 53
HC3: 83
HC4: 100
HC5: 100
HC6: 100
HC7: 100
HC8: 100
HC9: 89
HC10: 55
HC11: 100
HC12: 94
HC13: 100
HC14: 98
HC15: 19
HC16: 93
HC17: 100
HC18: 100
HC19: 100
HC20: 53
HC21: 72
MHUC1: 22
MHUC2: 6
AC1: 35
AC2: 46
ADATE_2: 2
ADATE_3: 3
ADATE_4: 10
ADATE_5: 2
ADATE_6: 3
ADATE_7: 4
ADATE_8: 5
ADATE_9: 4
ADATE_10: 3
ADATE_11: 5
ADATE_12: 5
ADATE_13: 4
ADATE_14: 3
ADATE_15: 2
ADATE_16: 4
ADATE_17: 4
ADATE_18: 10
ADATE_19: 4
ADATE_20: 3
ADATE_21: 3

ADATE_22: 6
ADATE_23: 4
ADATE_24: 3
RFA_2: 14
RFA_3: 67
RFA_4: 67
RFA_5: 41
RFA_6: 108
RFA_7: 104
RFA_8: 105
RFA_9: 109
RFA_10: 94
RFA_11: 105
RFA_12: 108
RFA_13: 88
RFA_14: 99
RFA_15: 34
RFA_16: 126
RFA_17: 121
RFA_18: 125
RFA_19: 113
RFA_20: 82
RFA_21: 105
RFA_22: 118
RFA_23: 88
RFA_24: 97
CARDPROM: 59
MAXADATE: 5
NUMPROM: 162
CARDPM12: 21
NUMPRM12: 60
RDATE_3: 17
RDATE_4: 25
RDATE_5: 7
RDATE_6: 18
RDATE_7: 11
RDATE_8: 17
RDATE_9: 8
RDATE_10: 8
RDATE_11: 10
RDATE_12: 10
RDATE_13: 16
RDATE_14: 12
RDATE_15: 17
RDATE_16: 19
RDATE_17: 12
RDATE_18: 16
RDATE_19: 14

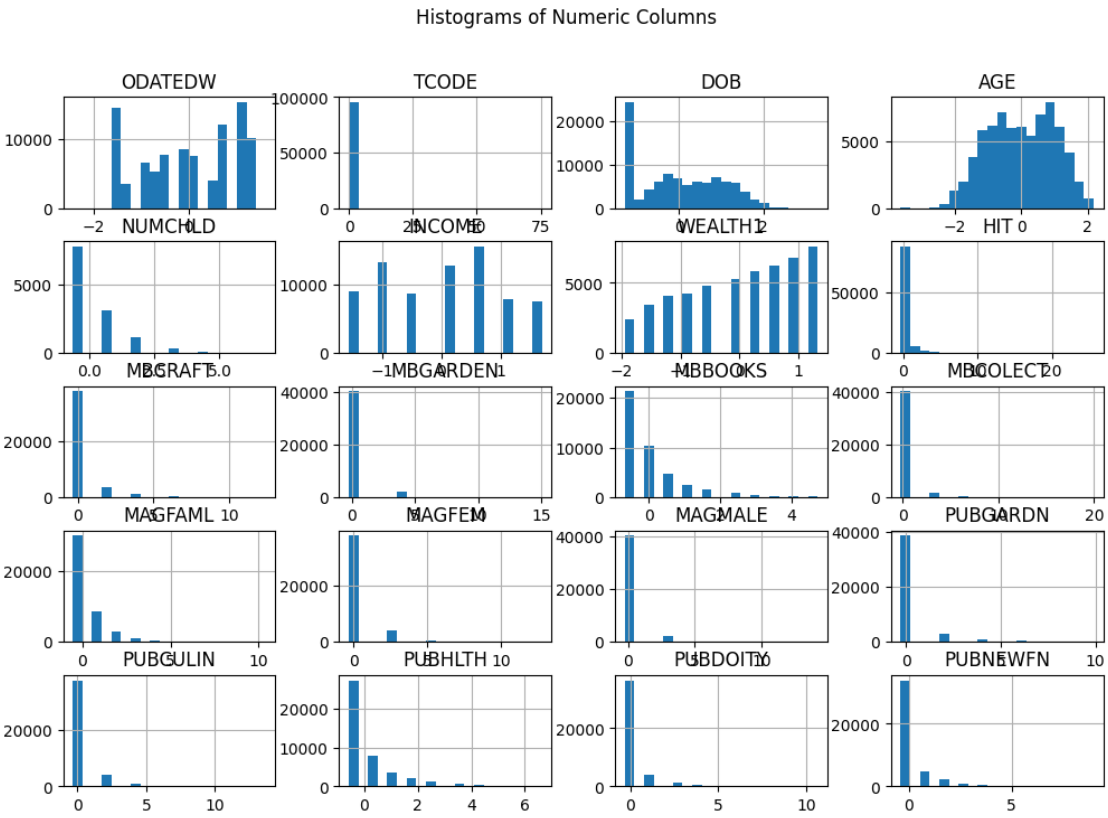
RDATE_20: 11
RDATE_21: 13
RDATE_22: 13
RDATE_23: 17
RDATE_24: 15
RAMNT_3: 32
RAMNT_4: 29
RAMNT_5: 9
RAMNT_6: 39
RAMNT_7: 82
RAMNT_8: 106
RAMNT_9: 95
RAMNT_10: 80
RAMNT_11: 95
RAMNT_12: 130
RAMNT_13: 79
RAMNT_14: 94
RAMNT_15: 77
RAMNT_16: 99
RAMNT_17: 66
RAMNT_18: 86
RAMNT_19: 89
RAMNT_20: 75
RAMNT_21: 79
RAMNT_22: 93
RAMNT_23: 70
RAMNT_24: 91
RAMNTALL: 2129
NGIFTALL: 90
CARDGIFT: 35
MINRAMNT: 208
MINRDATE: 146
MAXRAMNT: 278
MAXRDATE: 146
LASTGIFT: 238
LASTDATE: 24
FISTDATE: 175
NEXTDATE: 181
TIMELAG: 65
AVGGIFT: 7735
CONTROLN: 96367
HPHONE_D: 2
RFA_2R: 1
RFA_2F: 4
RFA_2A: 4
MDMAUD_R: 5
MDMAUD_F: 4
MDMAUD_A: 5

```
CLUSTER2: 63
GEOCODE2: 6
```

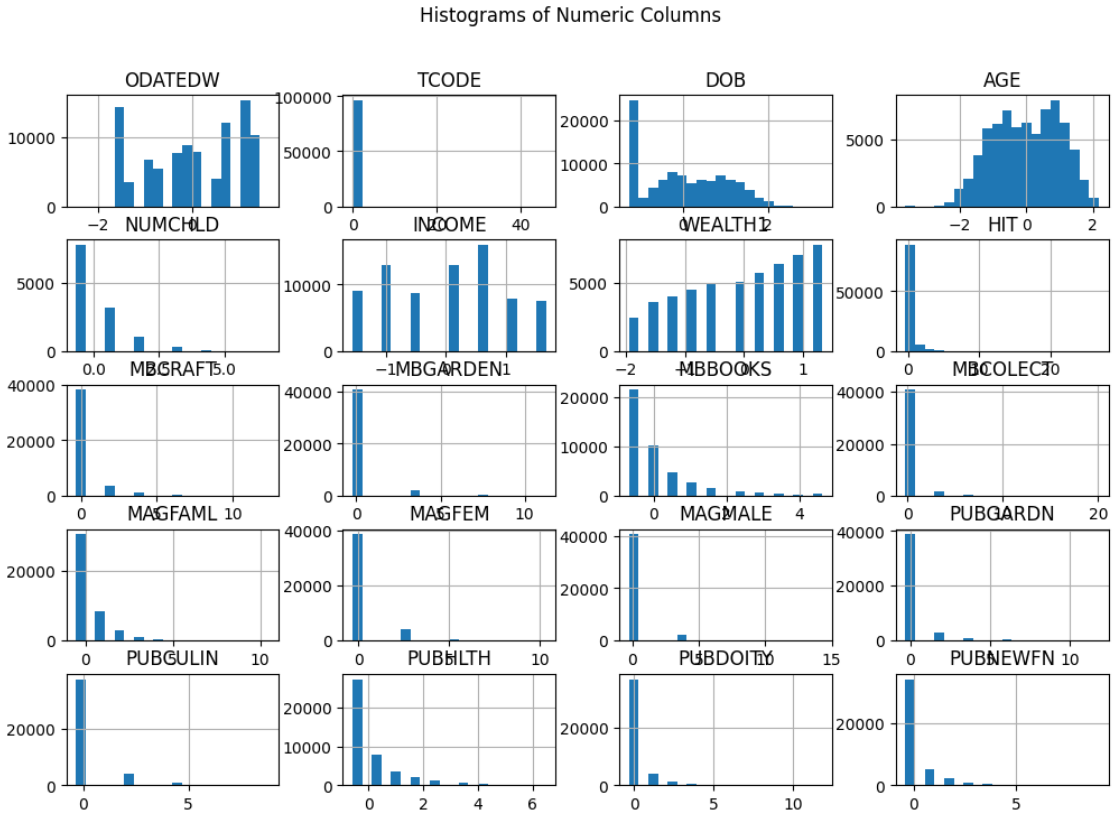
0.1.2 Histogram and Bar chart

[ ]:
```python
def showHistogram(df):
    numeric_columns = df.select_dtypes(include='number').columns
    numeric_columns = numeric_columns[:min(20, len(numeric_columns))]
    df[numeric_columns].hist(bins=20, figsize=(12, 8))
    plt.suptitle('Histograms of Numeric Columns')
    plt.show()


def showBarGrap(df):
    categorical_columns = df.select_dtypes(include='object').columns
    categorical_columns = categorical_columns[:min(
        20, len(categorical_columns))]

    df[numeric_columns].hist(bins=20, figsize=(12, 8))
    df[categorical_columns].value_counts().plot(kind='bar', color='skyblue')

    plt.show()
```

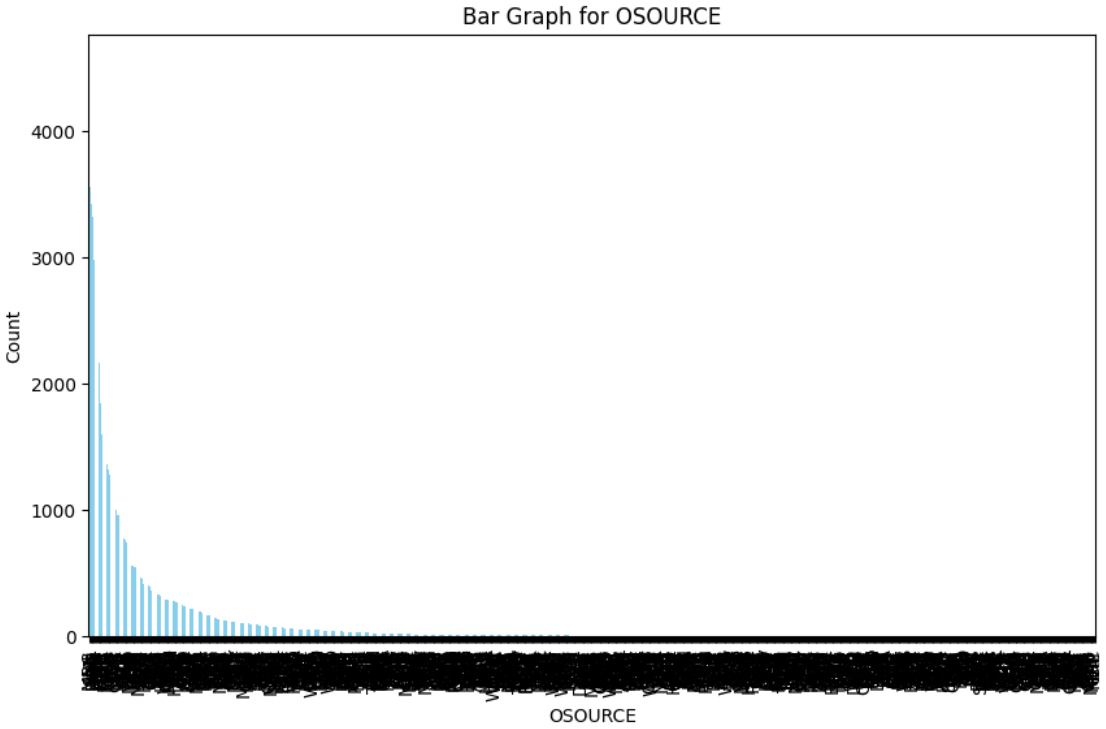
[ ]:
```python
showHistogram(train_filtered)
```

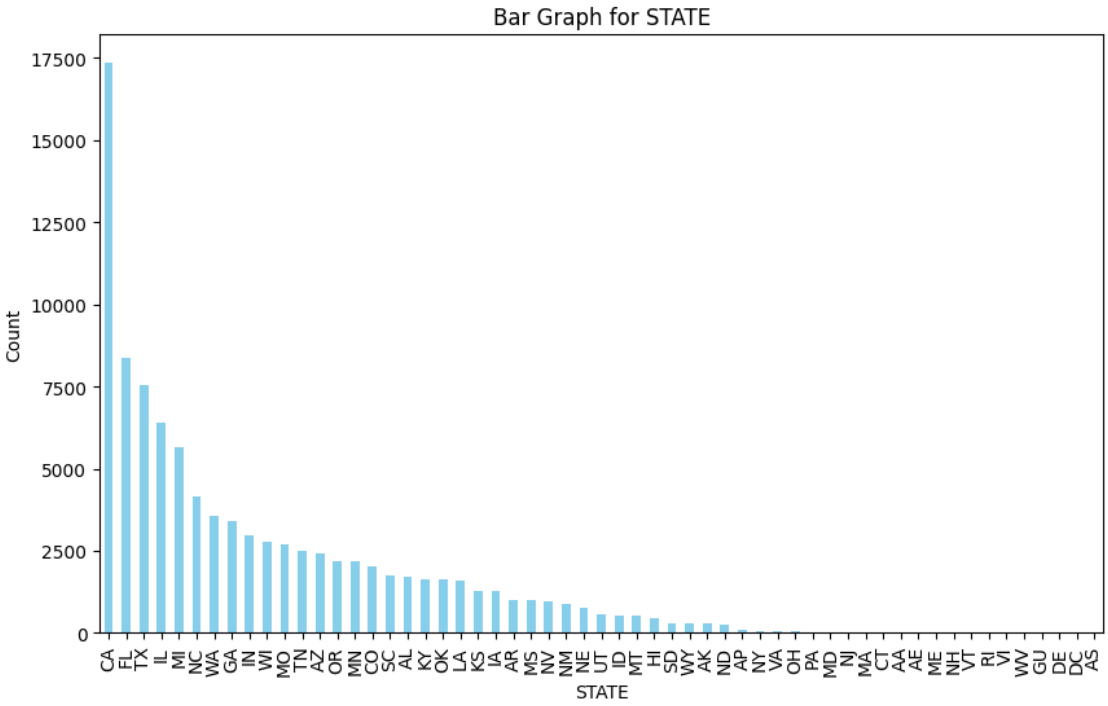


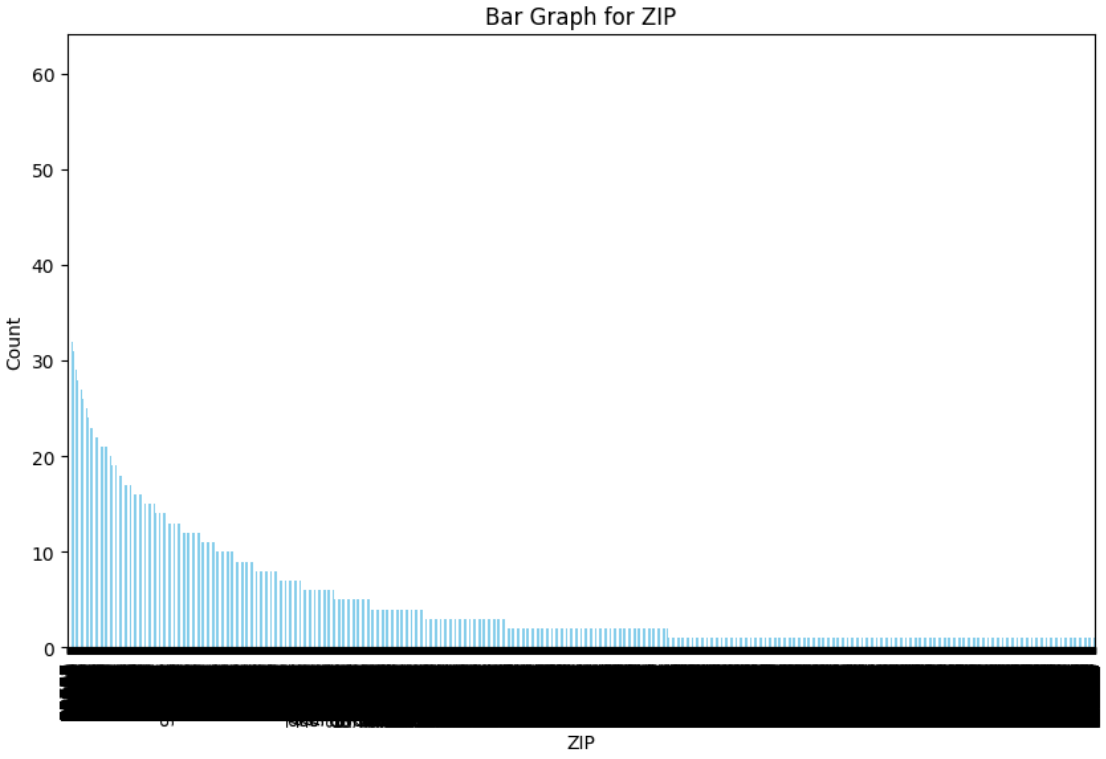
[ ]:
```
showHistogram(test_filtered)
```

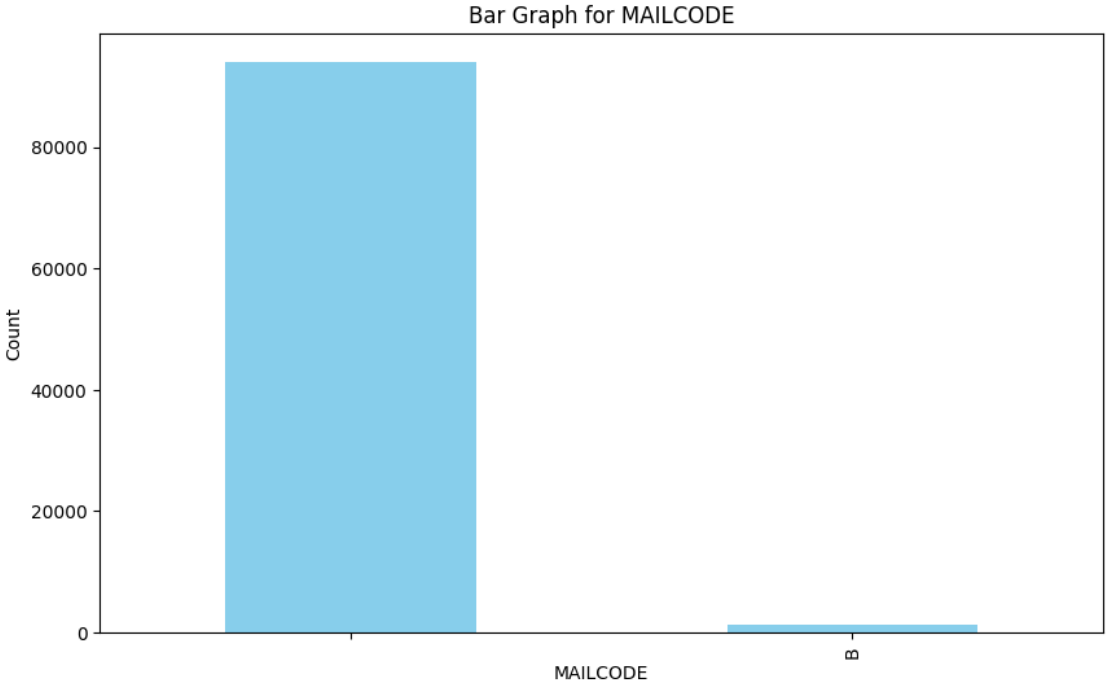


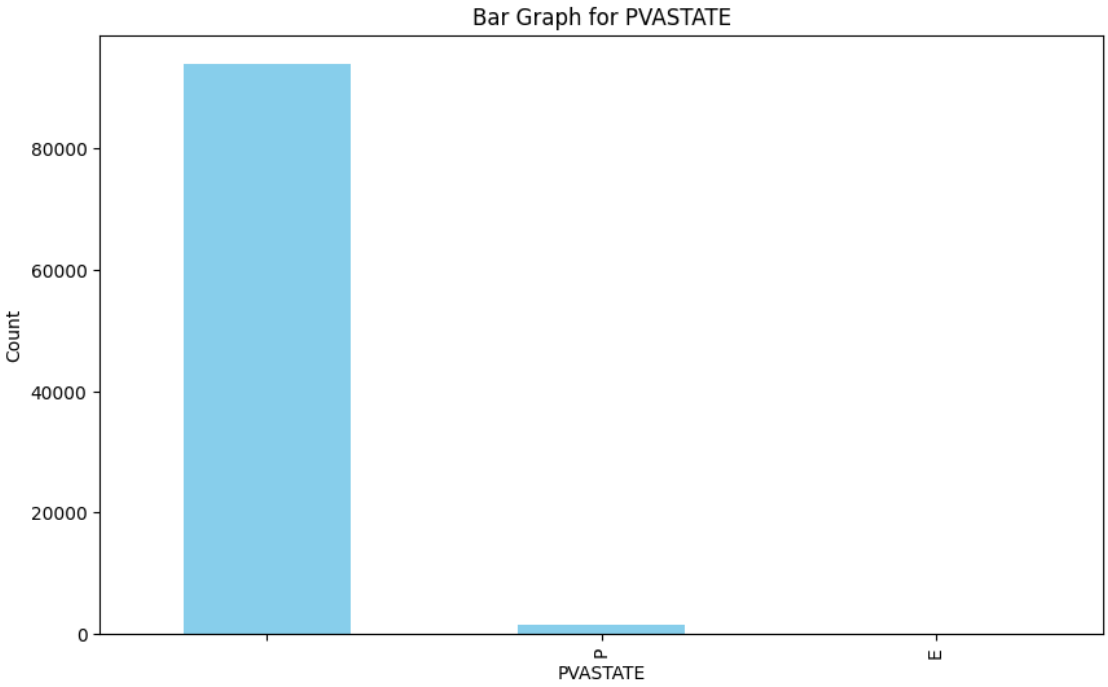
[ ]:
```
showBarGrap(train_filtered)
```

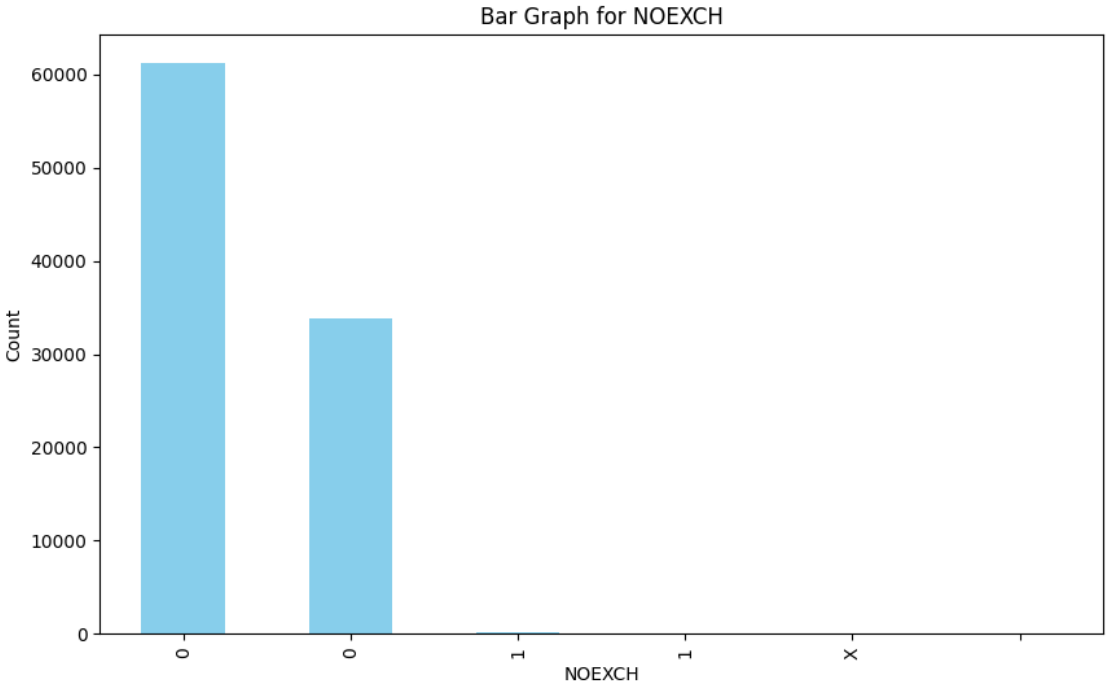

Bar Graph for OSOURCE
Count
4000
3000
2000
1000
0
OSOURCE

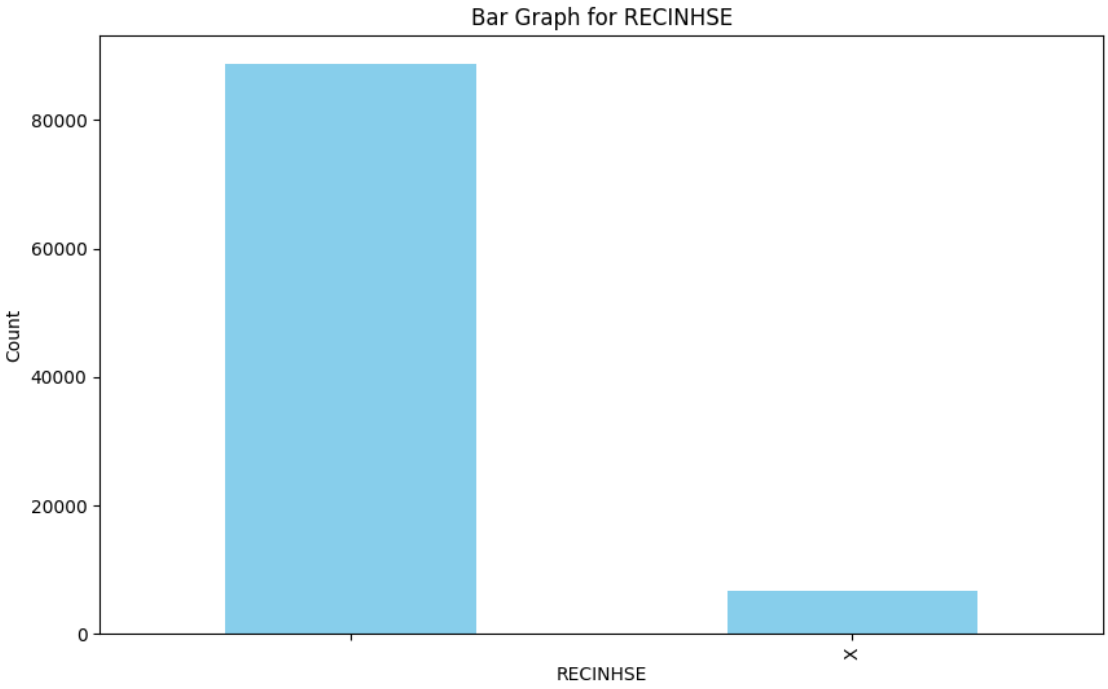

Bar Graph for STATE
Count
17500
15000
12500
10000
7500
5000
2500
0
CA FL TX IL MI NC WA GA IN WI MO TN AZ OR MN CO SC AL KY OK LA KS IA AR MS NV NM NE UT ID MT HI SD WY AK ND AP NY VA OH PA MD NJ MA CT AA AE ME NH VT RI VI WV GU DE DC AS
STATE

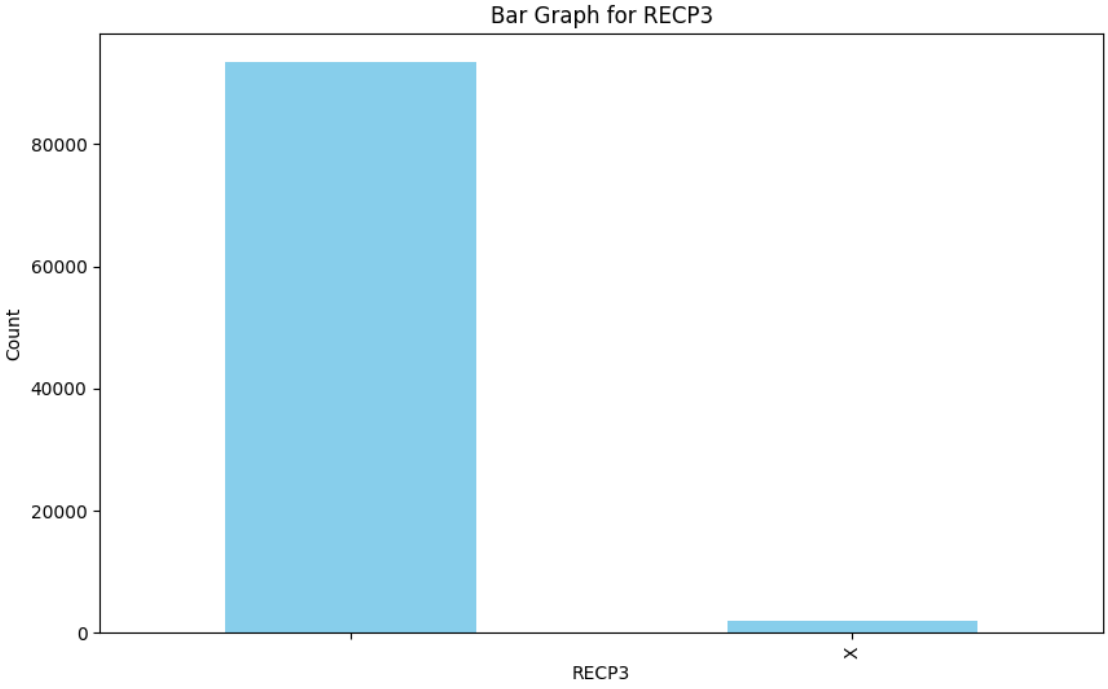

Bar Graph for ZIP
Count
60
50
40
30
20
10
0
ZIP

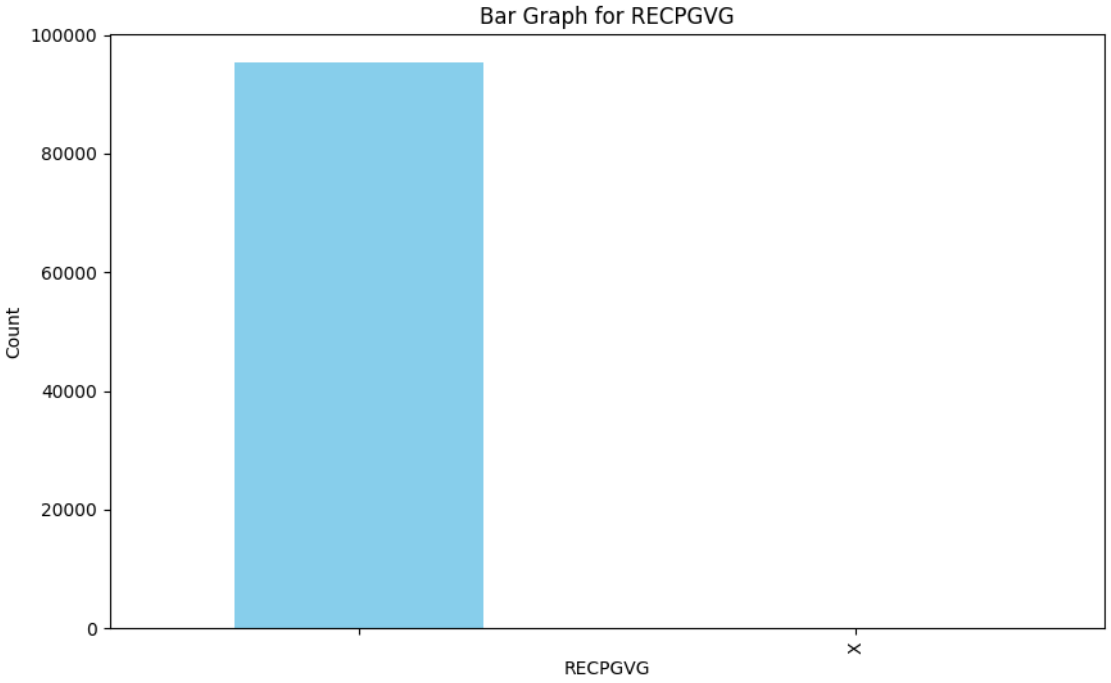

Bar Graph for MAILCODE
Count
80000
60000
40000
20000
0
B
MAILCODE

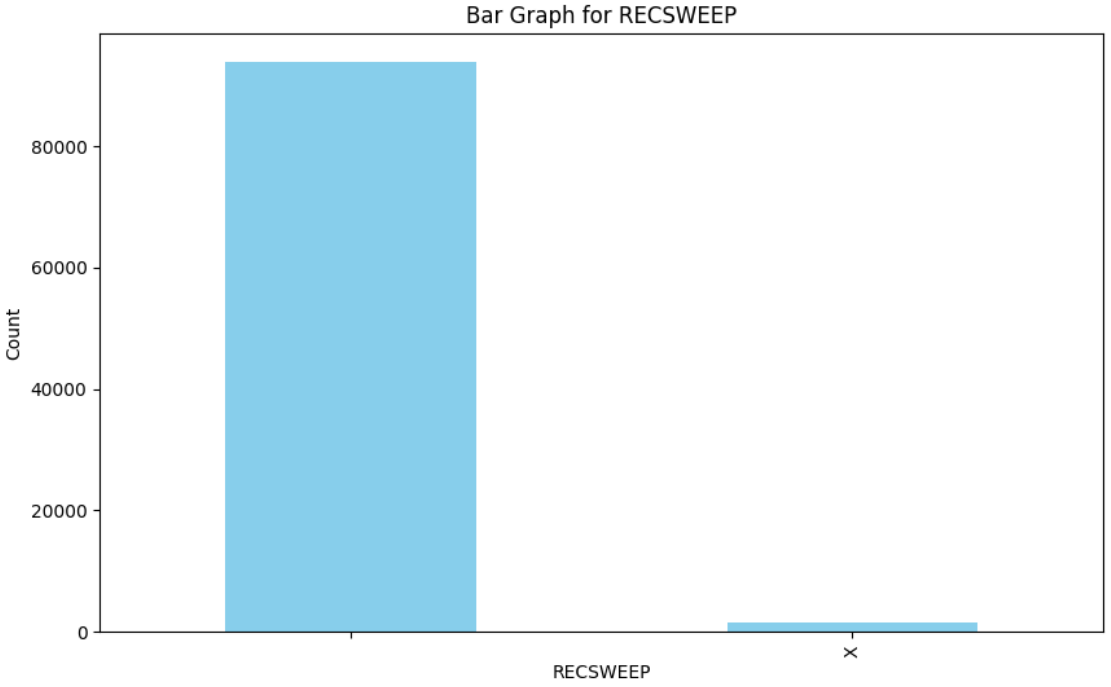

Bar Graph for PVASTATE
Count
80000
60000
40000
20000
0
P
E
PVASTATE

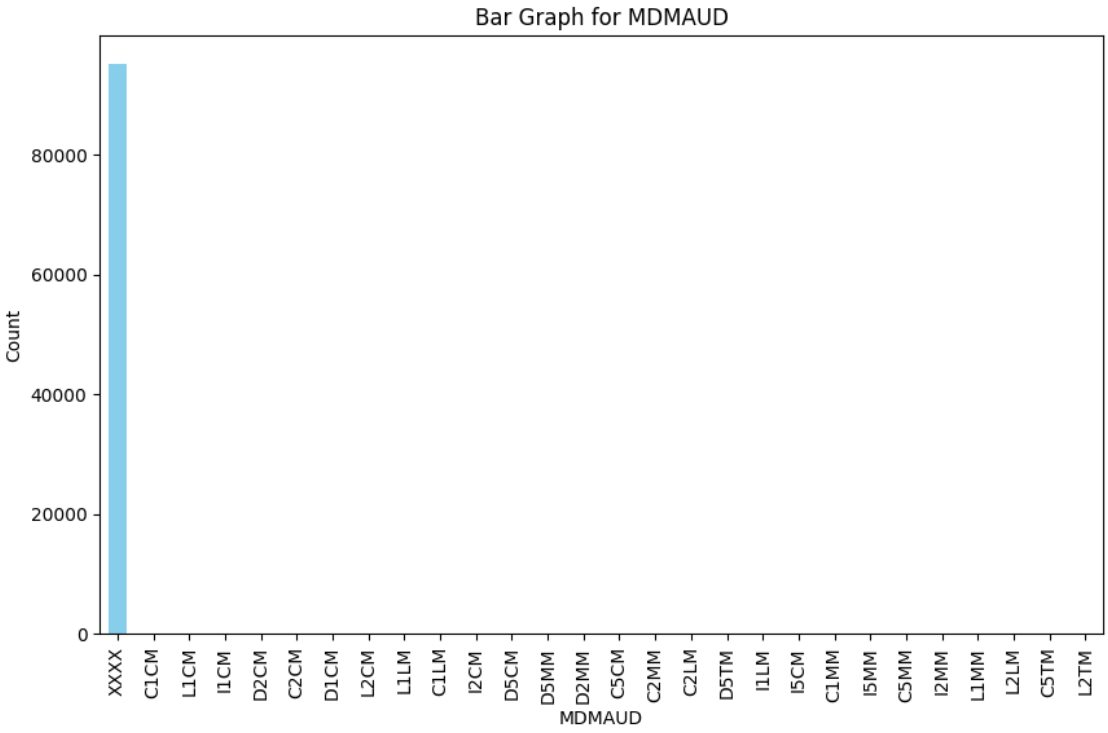

Bar Graph for NOEXCH
Count
60000
50000
40000
30000
20000
10000
0
0
0
1
1
X
NOEXCH

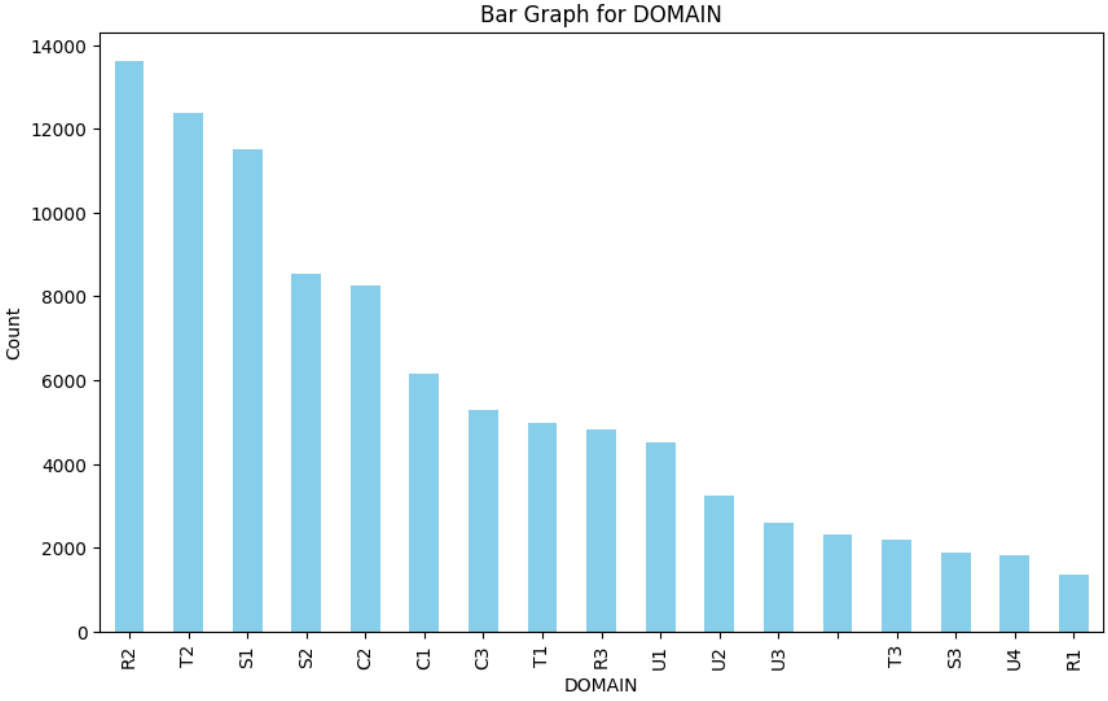

Bar Graph for RECINHSE
Count
80000
60000
40000
20000
0
RECINHSE
X

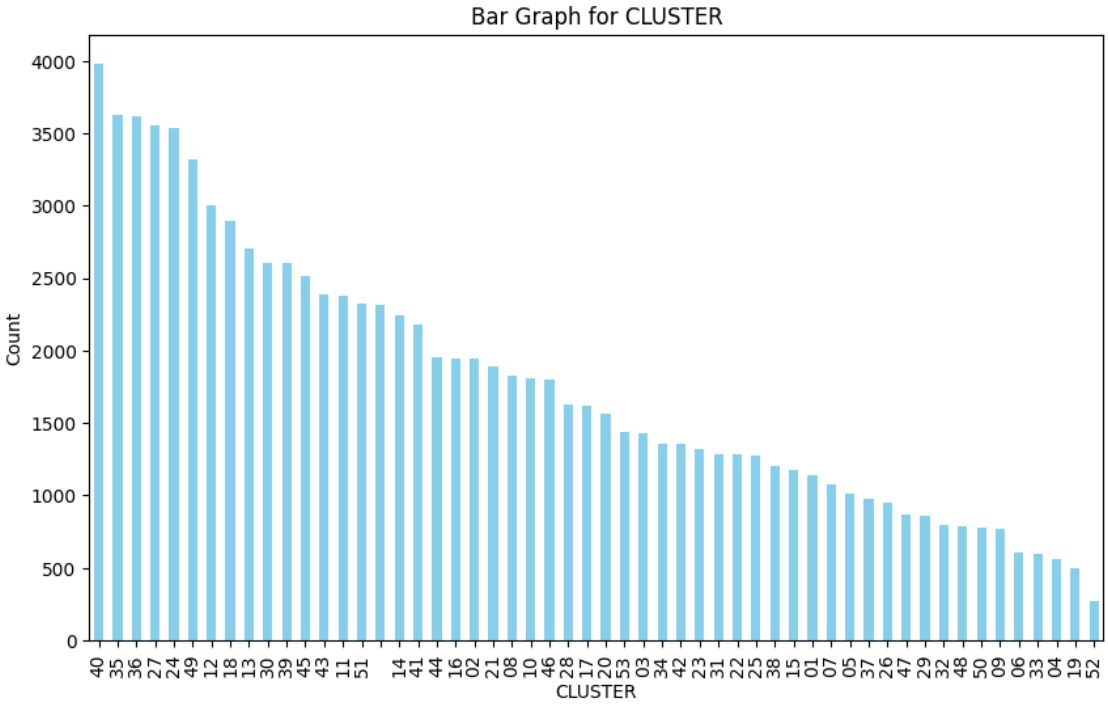

Bar Graph for RECP3
Count
80000
60000
40000
20000
0
RECP3
X

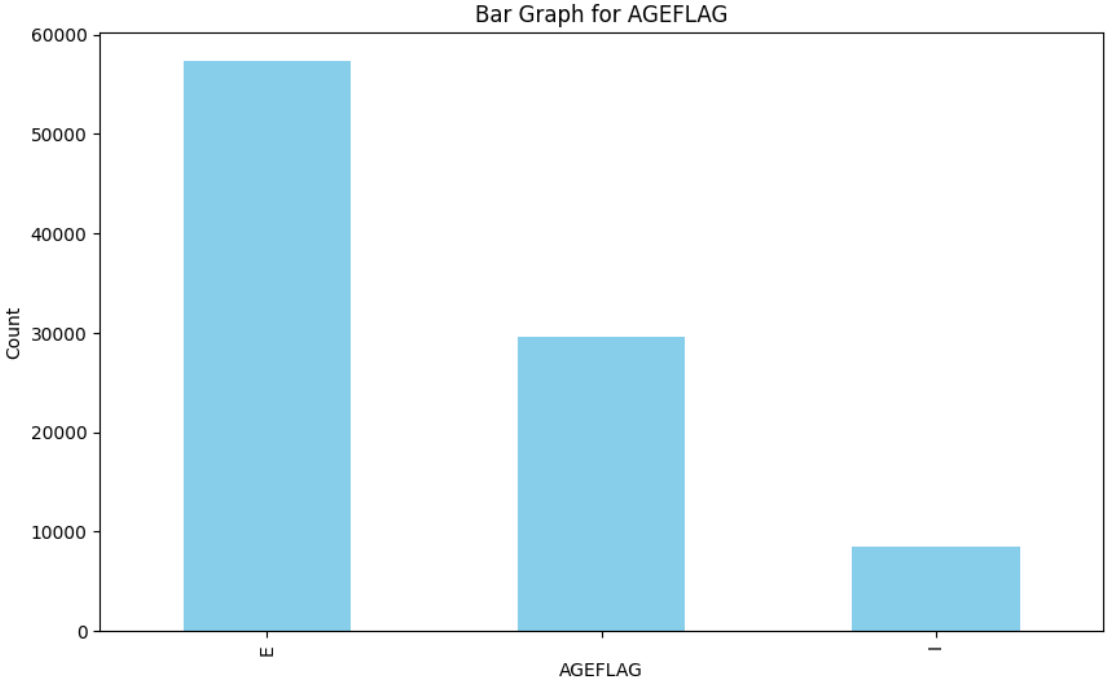

Bar Graph for RECPGVG
100000
80000
60000
40000
20000
0
Count
X
RECPGVG

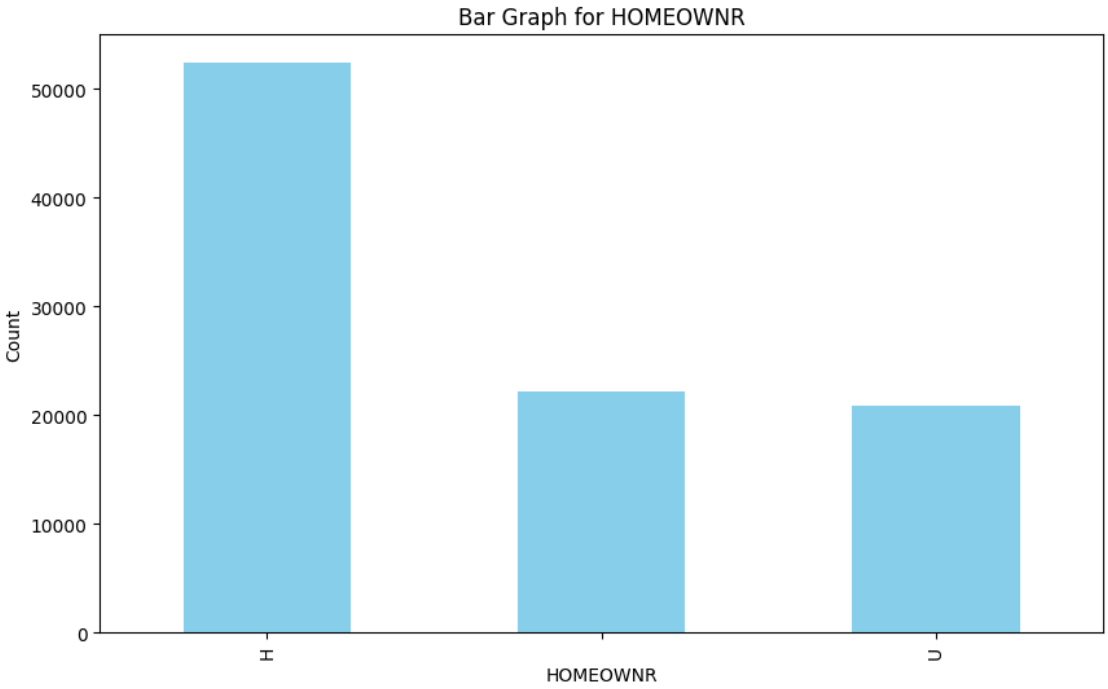

Bar Graph for RECSWEEP
80000
60000
40000
20000
0
Count
X
RECSWEEP

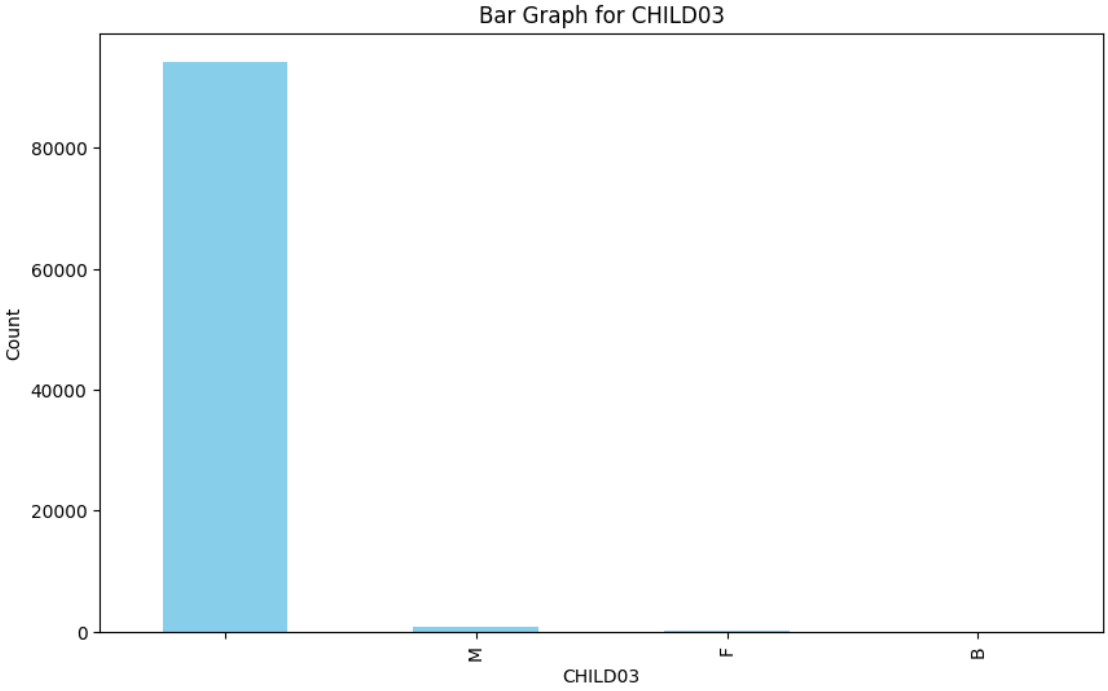

Bar Graph for MDMAUD
Count
0
20000
40000
60000
80000
XXXX
C1CM
L1CM
I1CM
D2CM
C2CM
D1CM
L2CM
L1LM
C1LM
I2CM
D5CM
D5MM
D2MM
C5CM
C2MM
C2LM
D5TM
I1LM
I5CM
C1MM
I5MM
C5MM
I2MM
L1MM
L2LM
C5TM
L2TM
MDMAUD

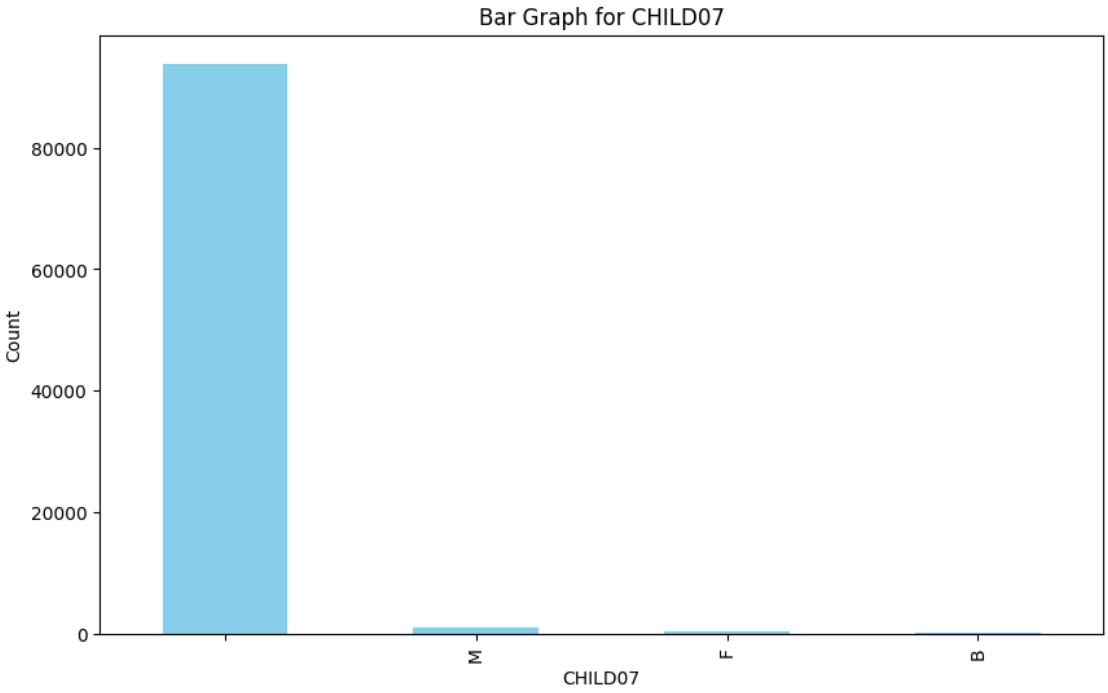

Bar Graph for DOMAIN
Count
0
2000
4000
6000
8000
10000
12000
14000
R2
T2
S1
S2
C2
C1
C3
T1
R3
U1
U2
U3
T3
S3
U4
R1
DOMAIN

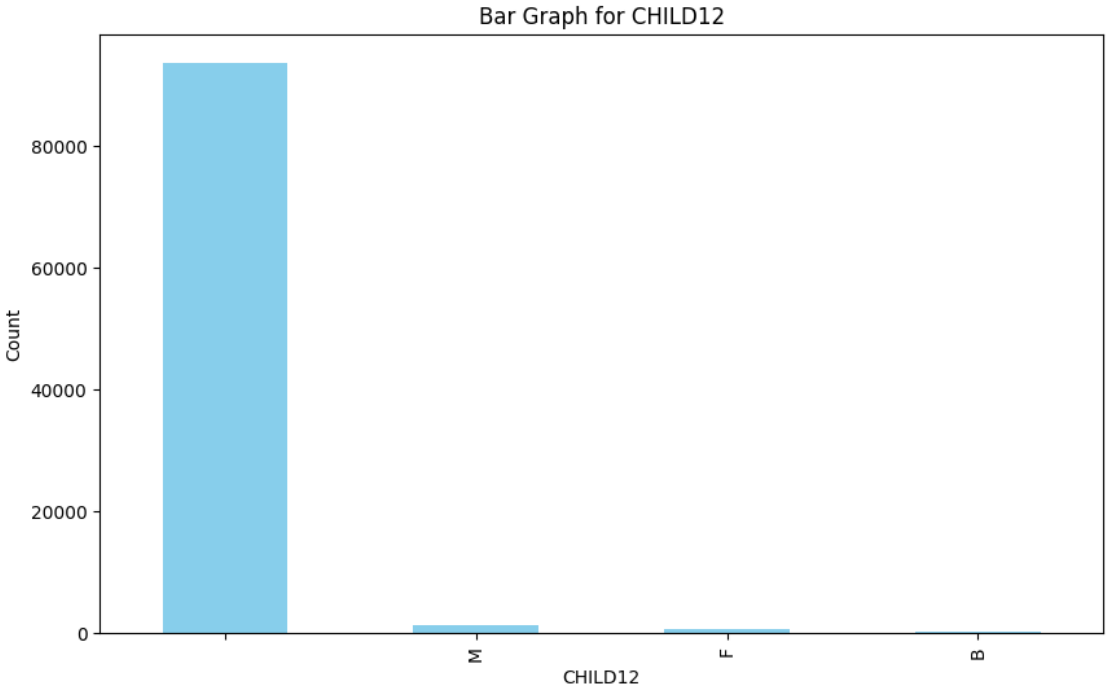

Bar Graph for CLUSTER
Count
CLUSTER
0
500
1000
1500
2000
2500
3000
3500
4000
40 35 36 27 24 49 12 18 13 30 39 45 43 11 51 14 41 44 16 02 21 08 10 46 28 17 20 53 03 34 42 23 31 22 25 38 15 01 07 05 37 26 47 29 32 48 50 09 06 33 04 19 52

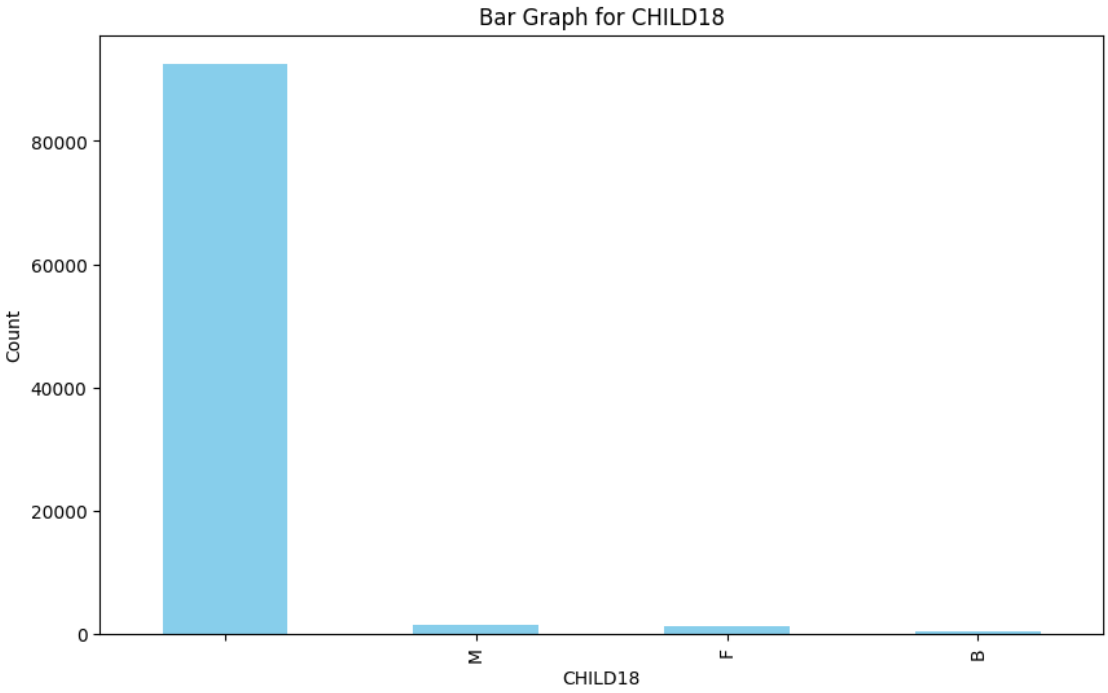

Bar Graph for AGEFLAG
Count
AGEFLAG
0
10000
20000
30000
40000
50000
60000
E
I

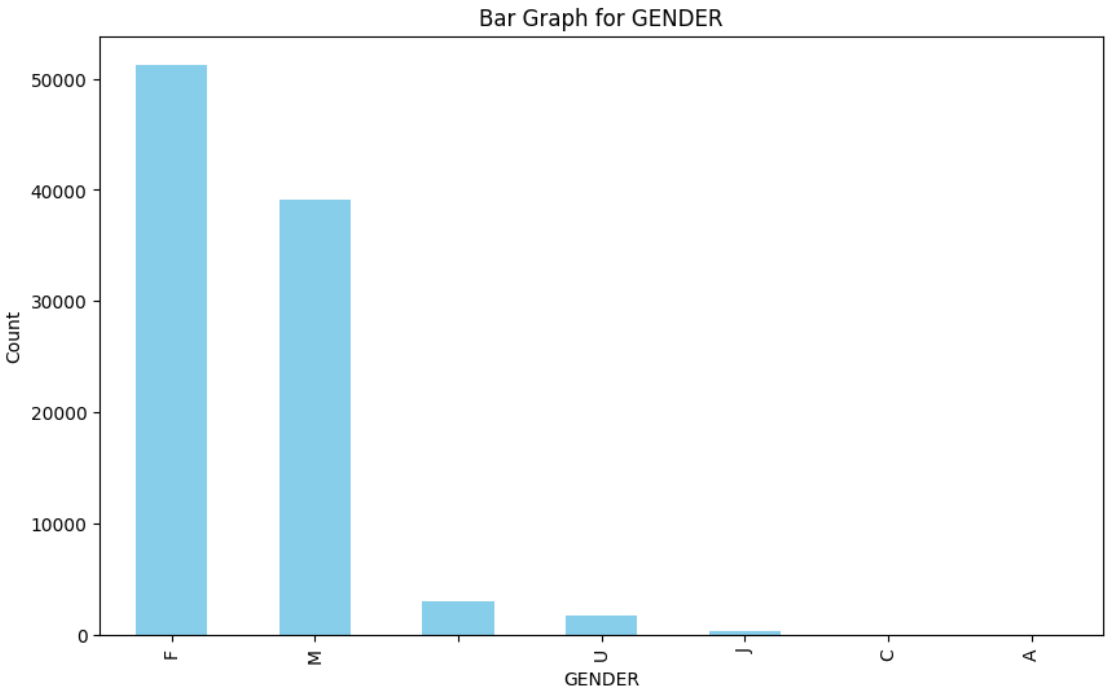

Bar Graph for HOMEOWNR
Count
50000
40000
30000
20000
10000
0
H
U
HOMEOWNR

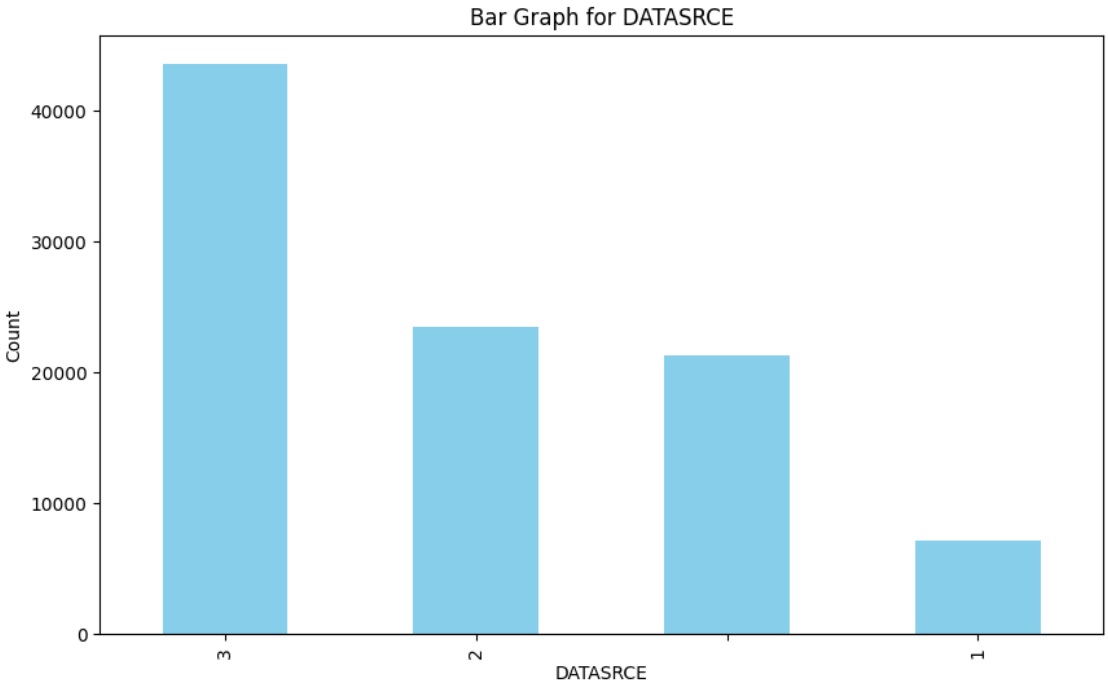

Bar Graph for CHILD03
Count
80000
60000
40000
20000
0
M
F
B
CHILD03

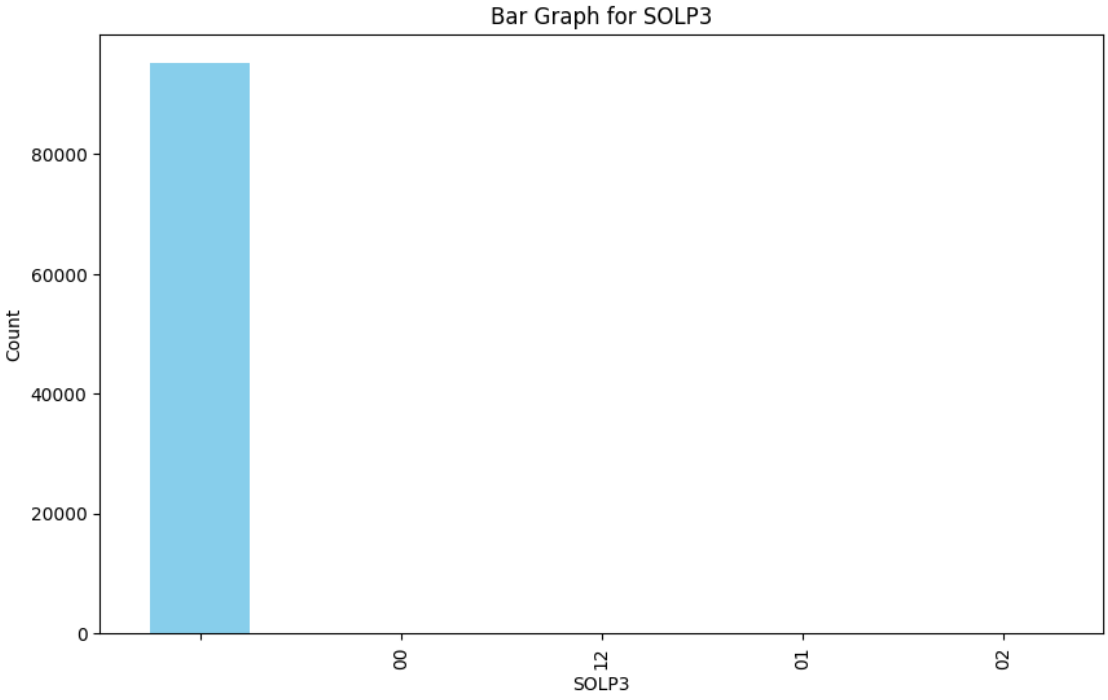

Bar Graph for CHILD07
Count
80000
60000
40000
20000
0
M
F
B
CHILD07

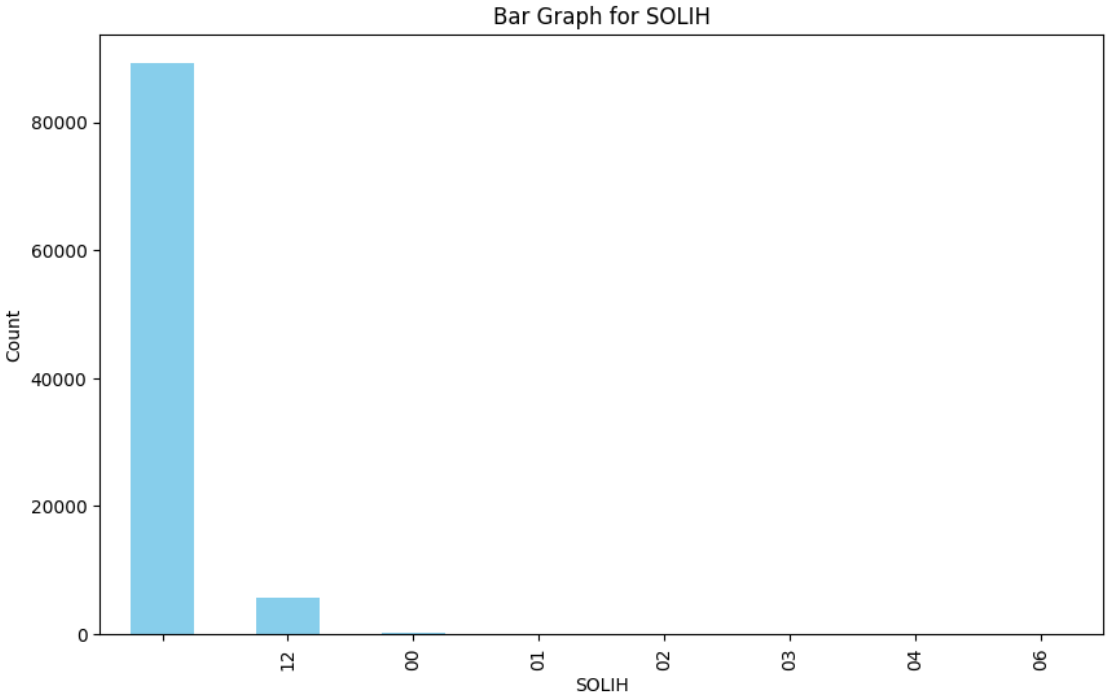

Bar Graph for CHILD12
Count
80000
60000
40000
20000
0
M
F
B
CHILD12

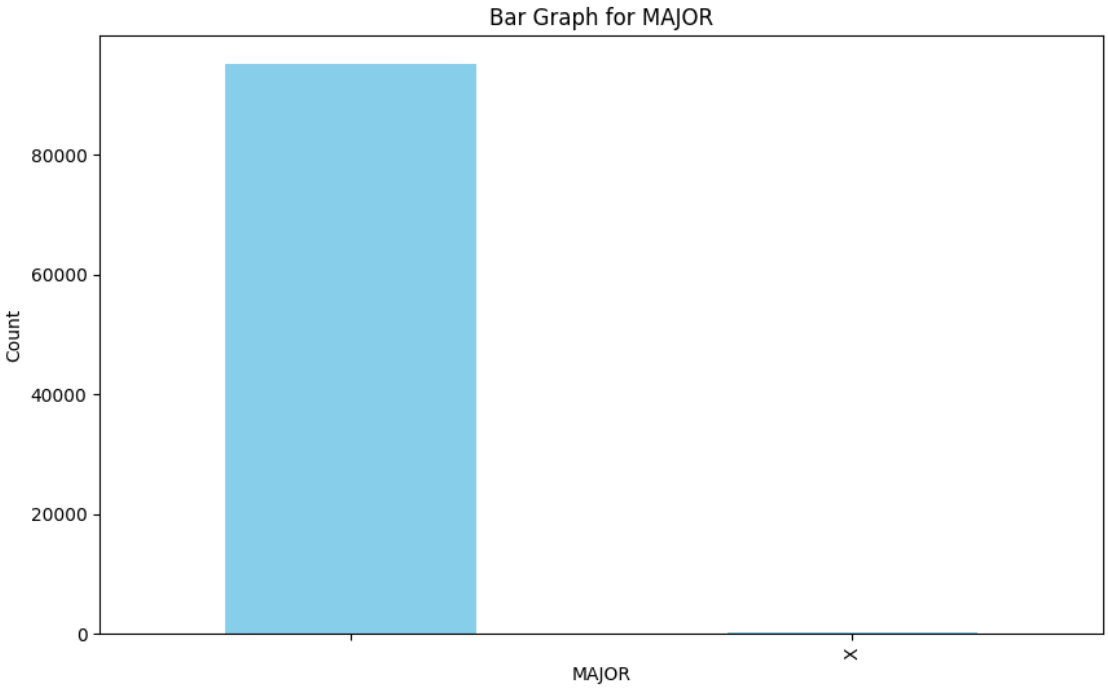

Bar Graph for CHILD18
Count
80000
60000
40000
20000
0
M
F
B
CHILD18

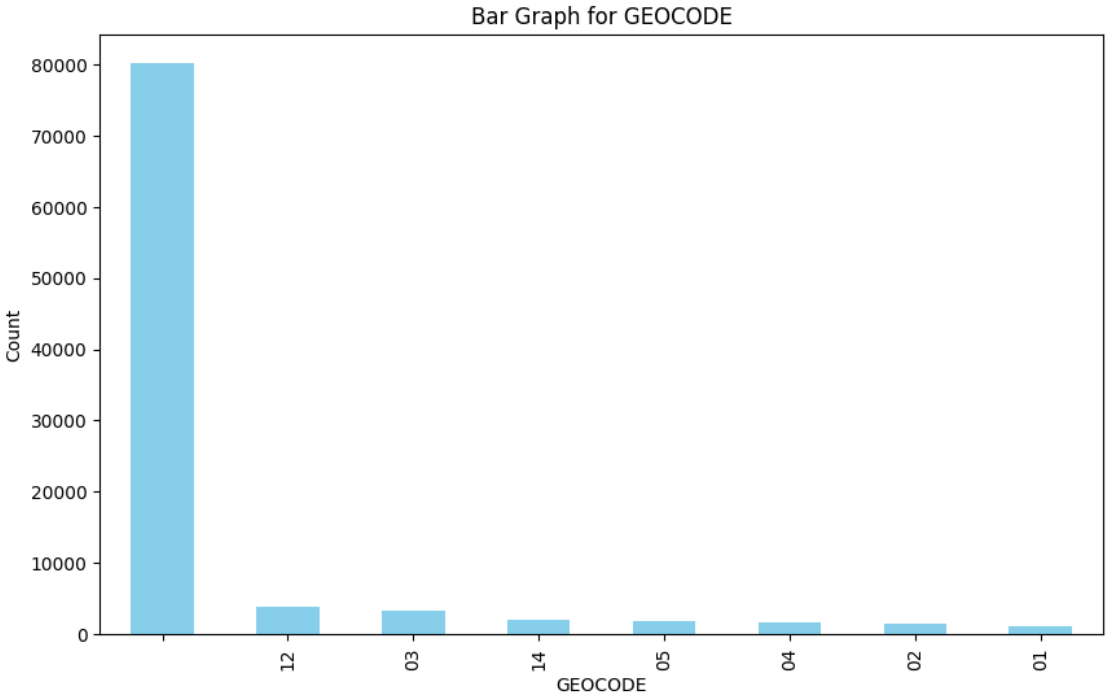

Bar Graph for GENDER
Count
50000
40000
30000
20000
10000
0
F
M
U
J
C
A
GENDER

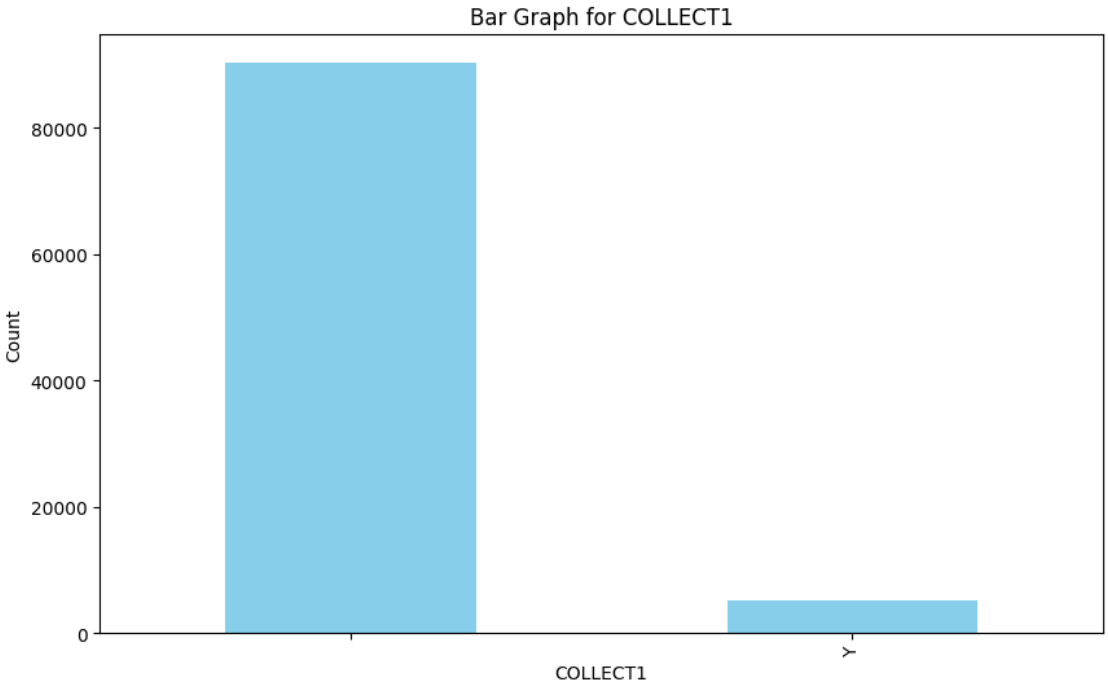

Bar Graph for DATASRCE
Count
40000
30000
20000
10000
0
3
2
1
DATASRCE

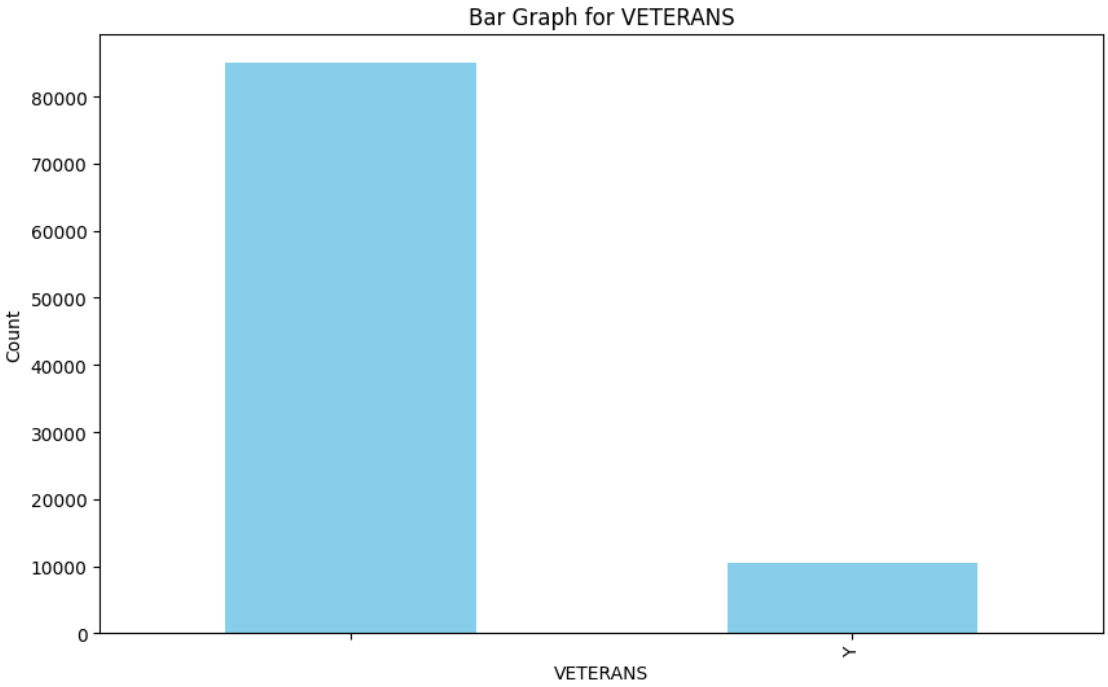

Bar Graph for SOLP3
Count
80000
60000
40000
20000
0
00
12
01
02
SOLP3

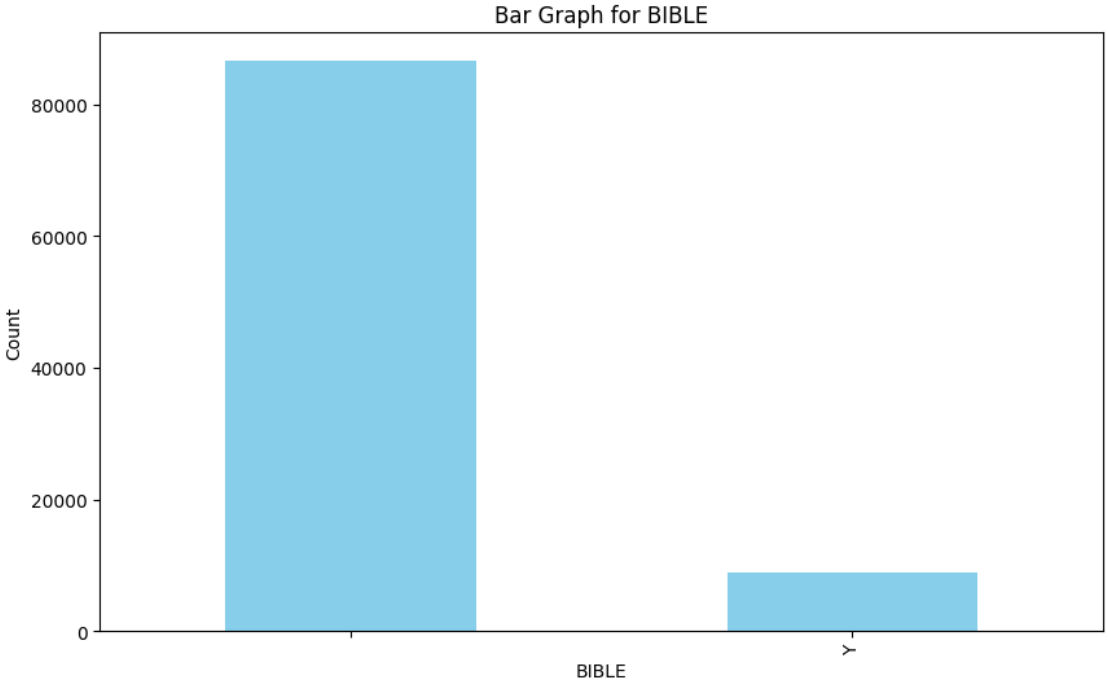

Bar Graph for SOLIH
Count
80000
60000
40000
20000
0
12
00
01
02
03
04
06
SOLIH

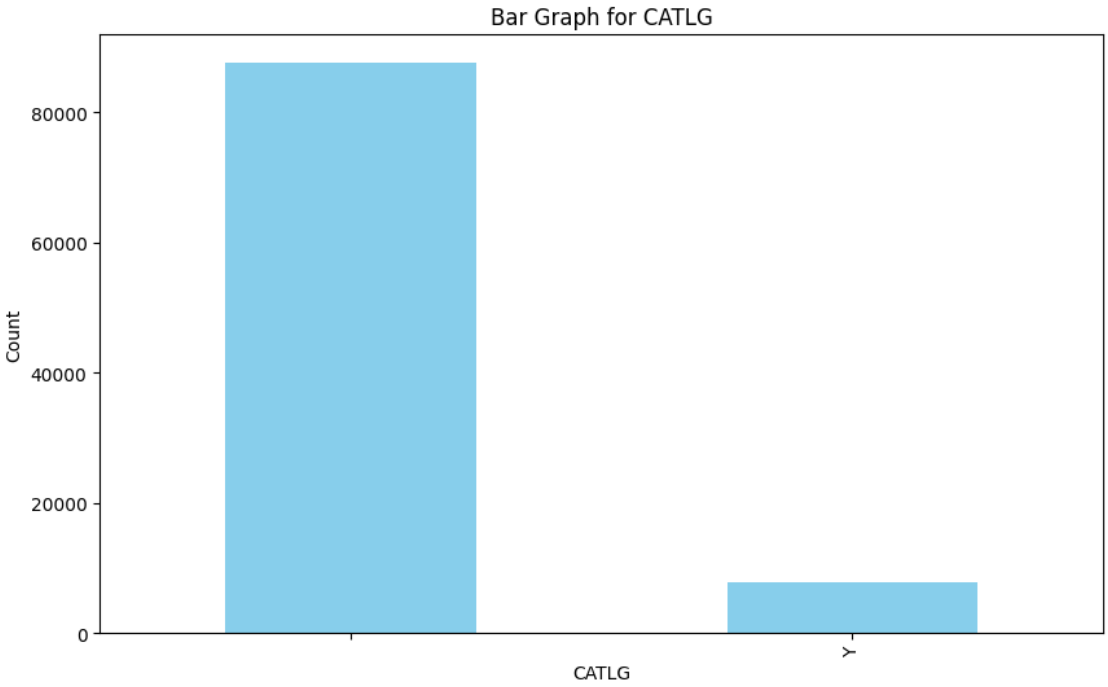

Bar Graph for MAJOR
Count
80000
60000
40000
20000
0
X
MAJOR

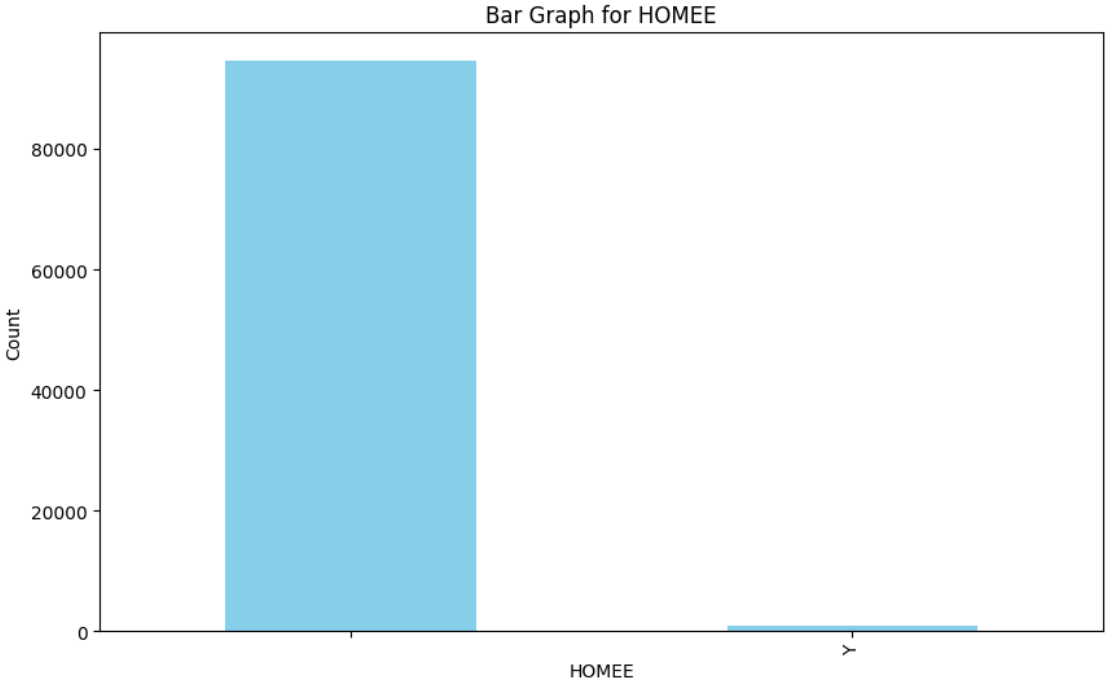

Bar Graph for GEOCODE
Count
80000
70000
60000
50000
40000
30000
20000
10000
0
12
03
14
05
04
02
01
GEOCODE

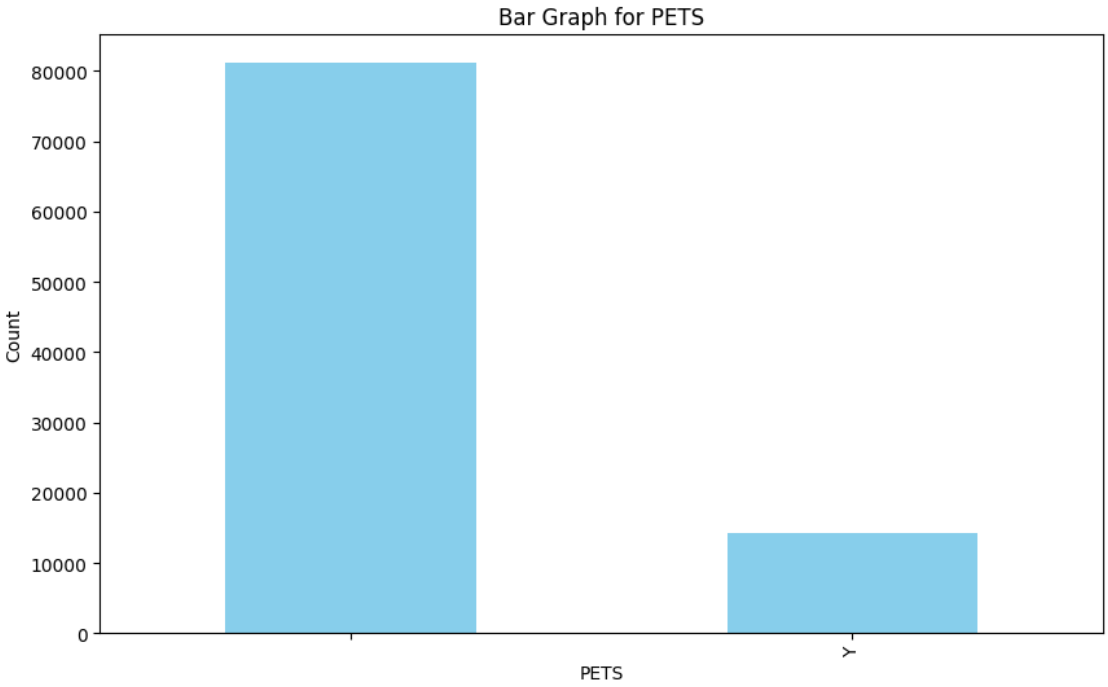

Bar Graph for COLLECT1
Count
80000
60000
40000
20000
0
Y
COLLECT1

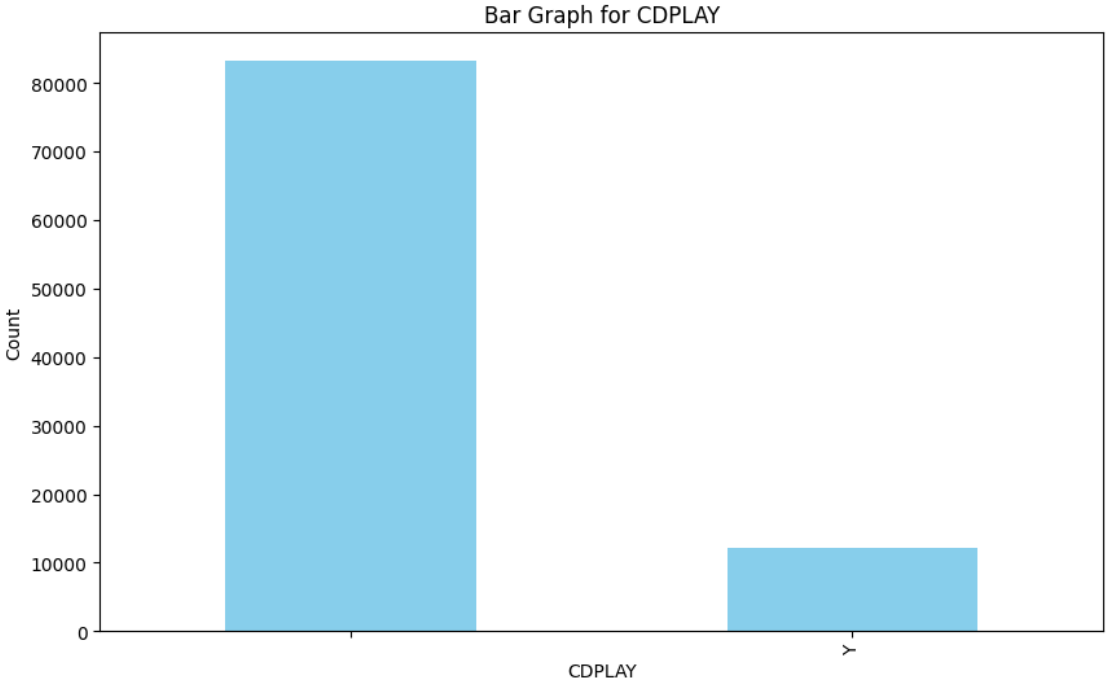

Bar Graph for VETERANS
Count
80000
70000
60000
50000
40000
30000
20000
10000
0
Y
VETERANS

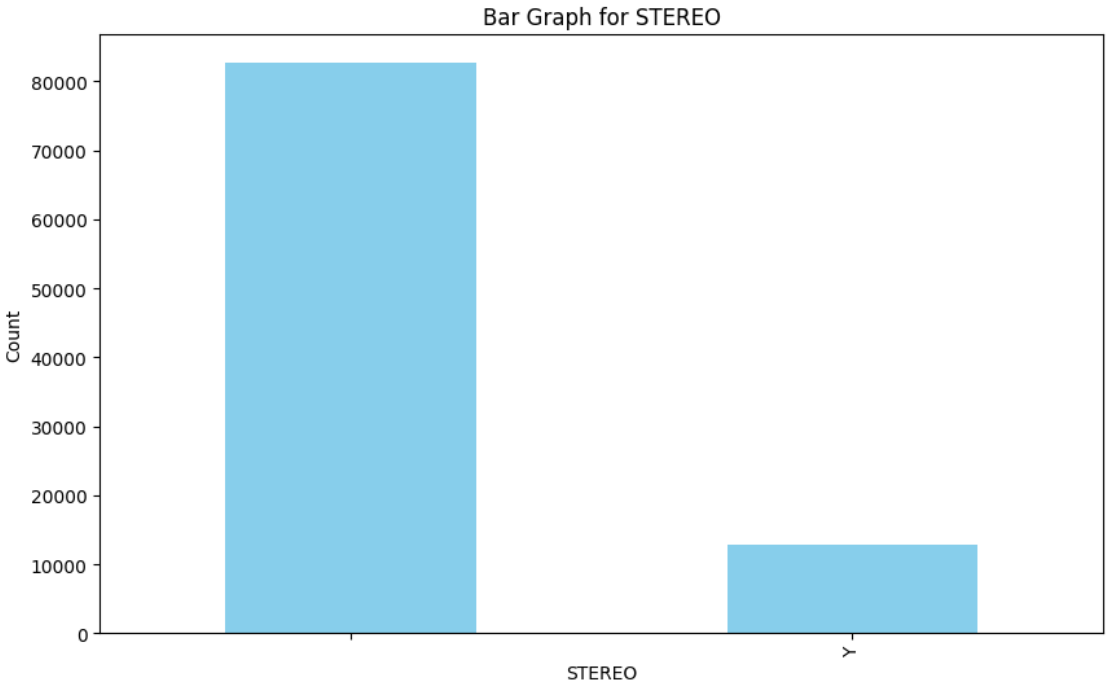

Bar Graph for BIBLE
Count
80000
60000
40000
20000
0
Y
BIBLE

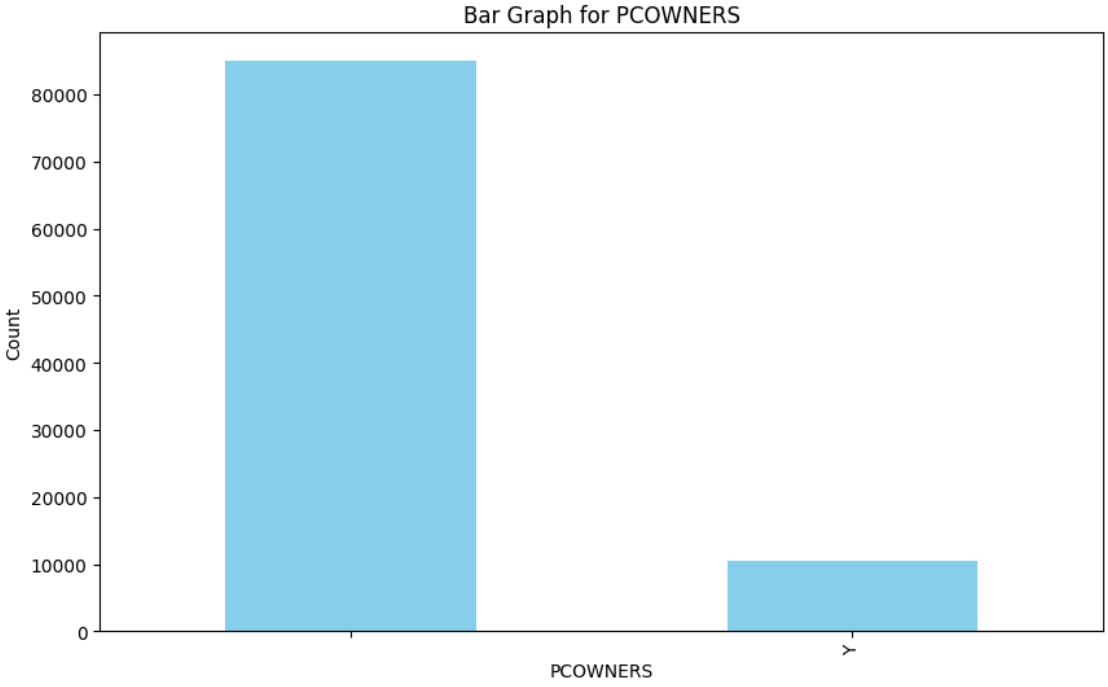

Bar Graph for CATLG
Count
80000
60000
40000
20000
0
Y
CATLG

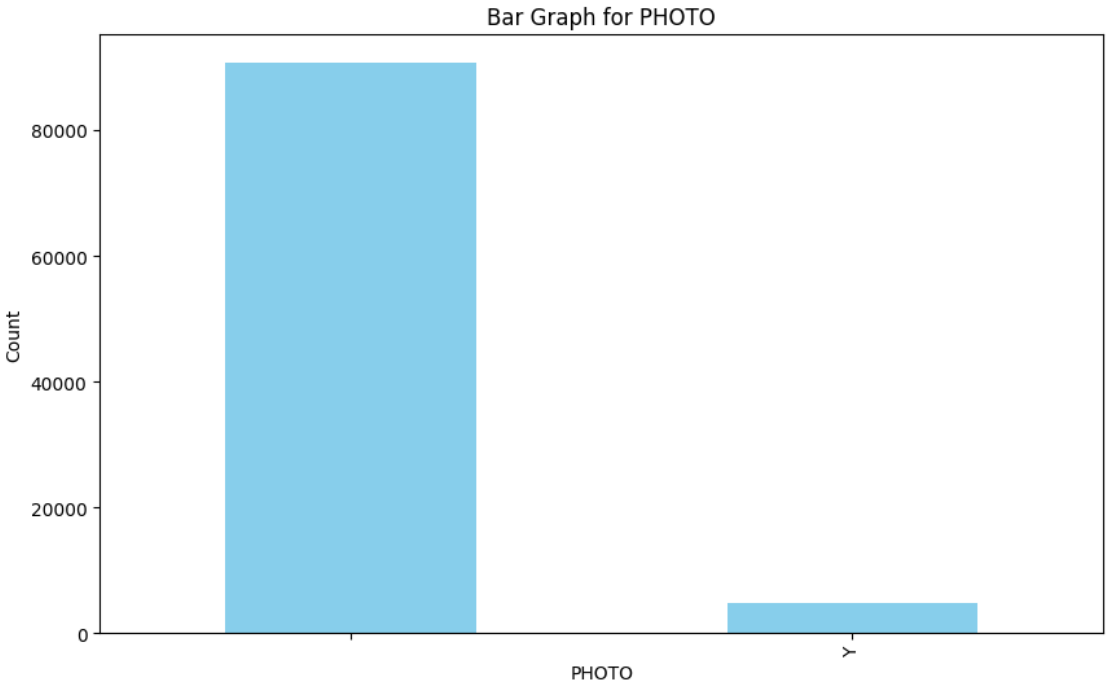

Bar Graph for HOMEE
Count
80000
60000
40000
20000
0
Y
HOMEE

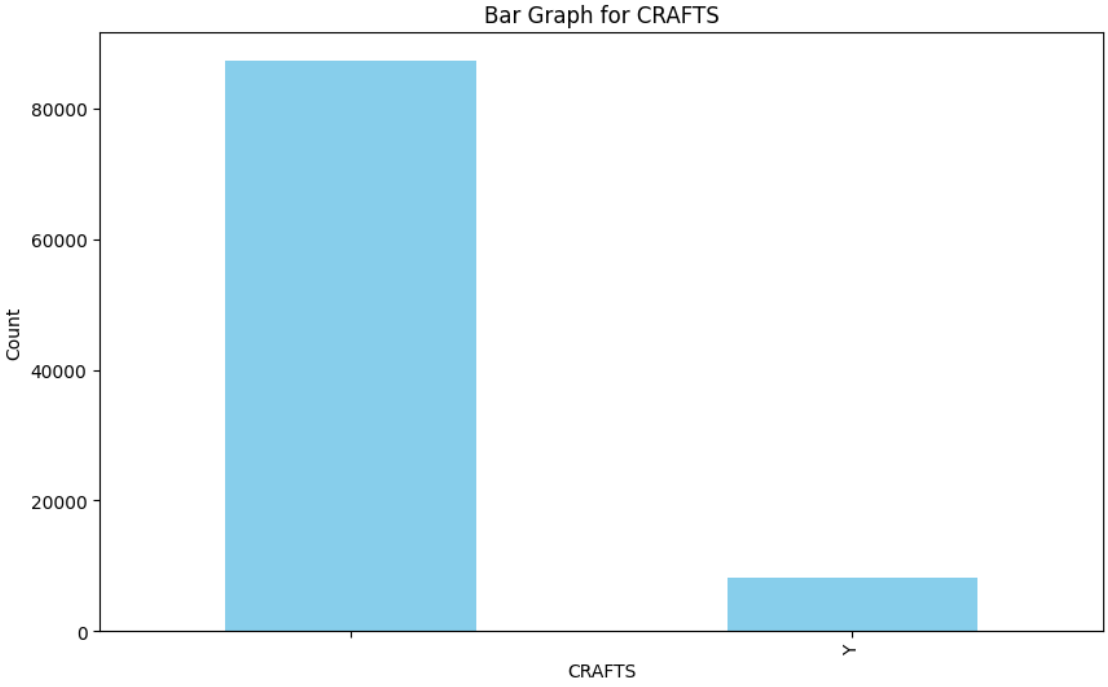

Bar Graph for PETS
Count
80000
70000
60000
50000
40000
30000
20000
10000
0
Y
PETS

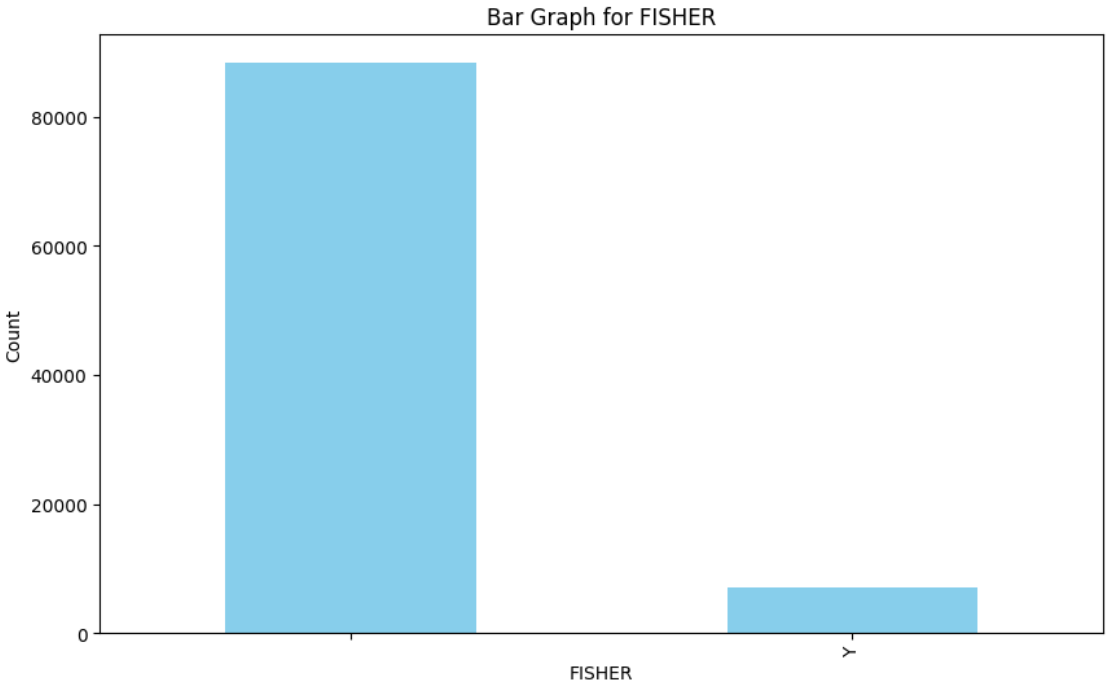

Bar Graph for CDPLAY
Count
80000
70000
60000
50000
40000
30000
20000
10000
0
Y
CDPLAY

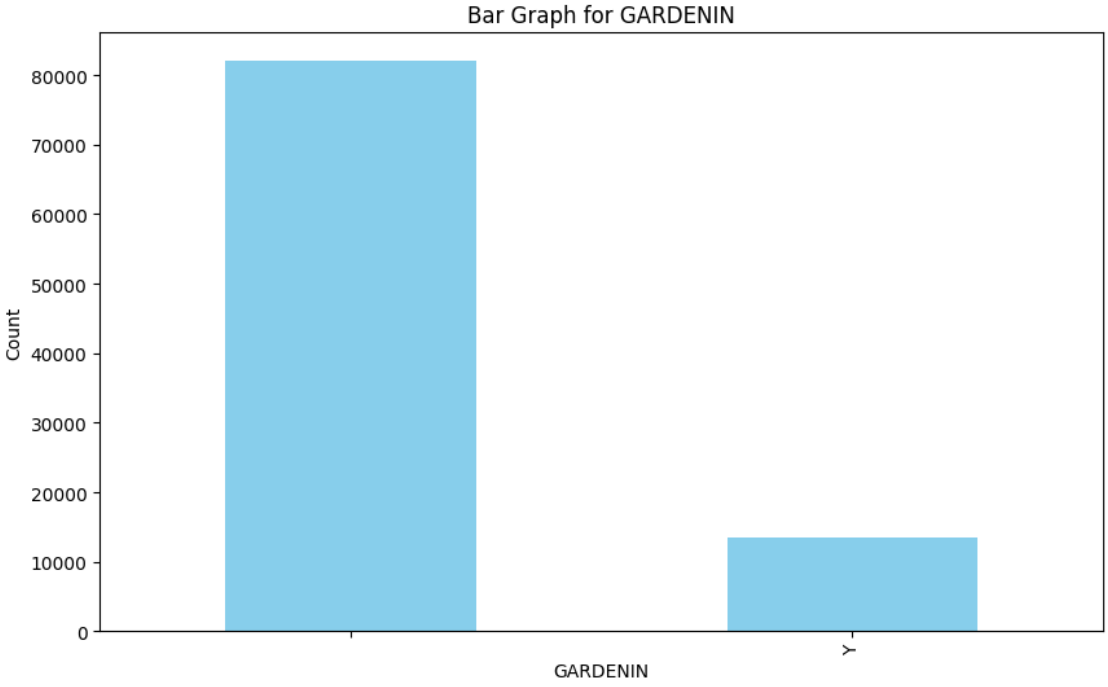

Bar Graph for STEREO
Count
80000
70000
60000
50000
40000
30000
20000
10000
0
Y
STEREO

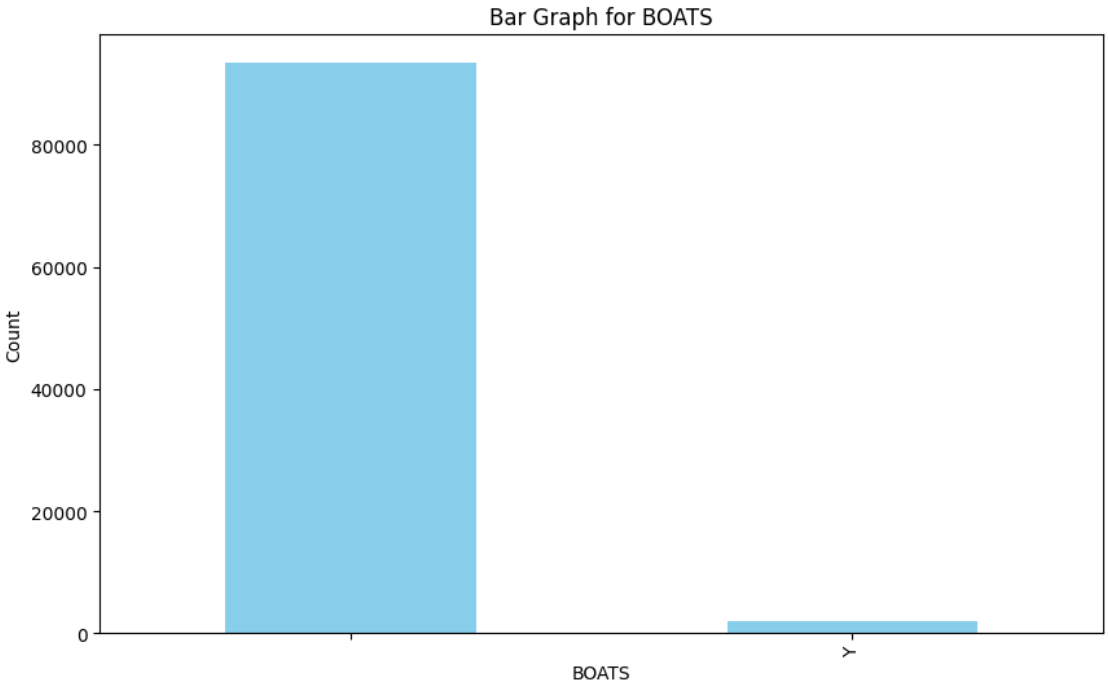

Bar Graph for PCOWNERS
Count
80000
70000
60000
50000
40000
30000
20000
10000
0
Y
PCOWNERS

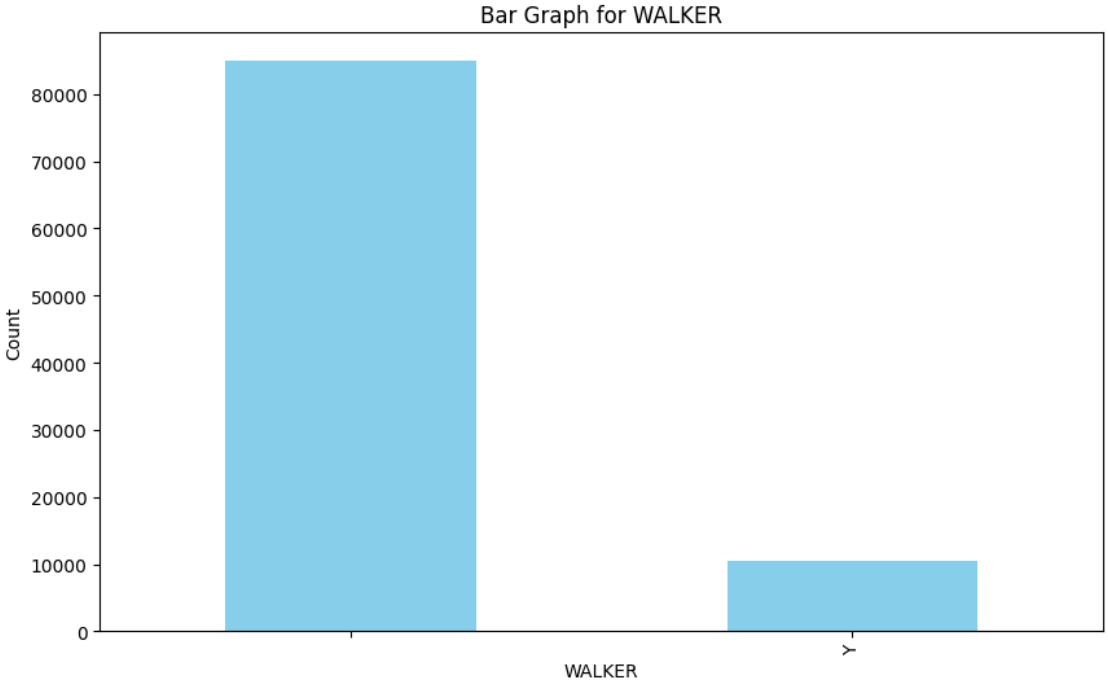

Bar Graph for PHOTO
Count
80000
60000
40000
20000
0
Y
PHOTO

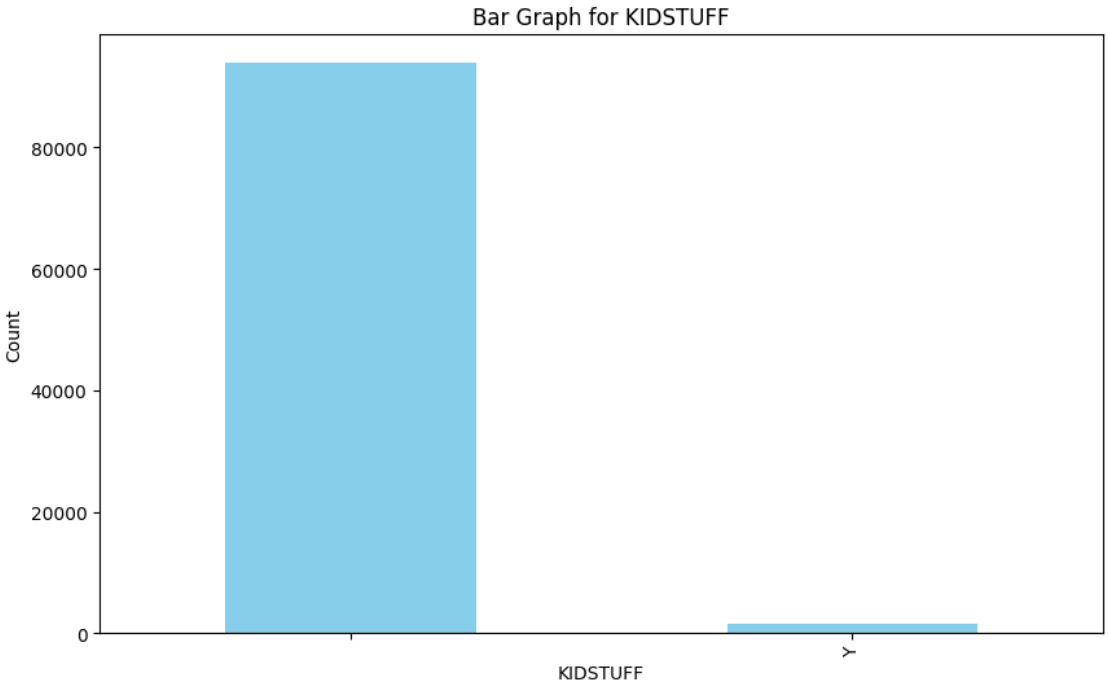

Bar Graph for CRAFTS
Count
80000
60000
40000
20000
0
Y
CRAFTS

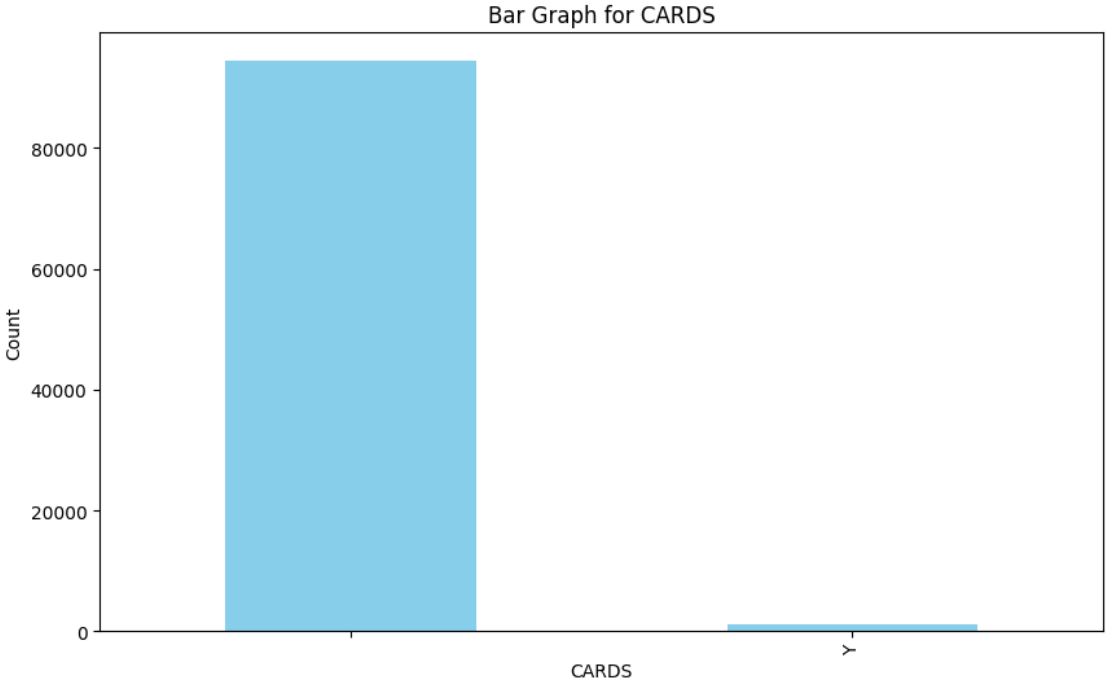

Bar Graph for FISHER
Count
80000
60000
40000
20000
0
Y
FISHER

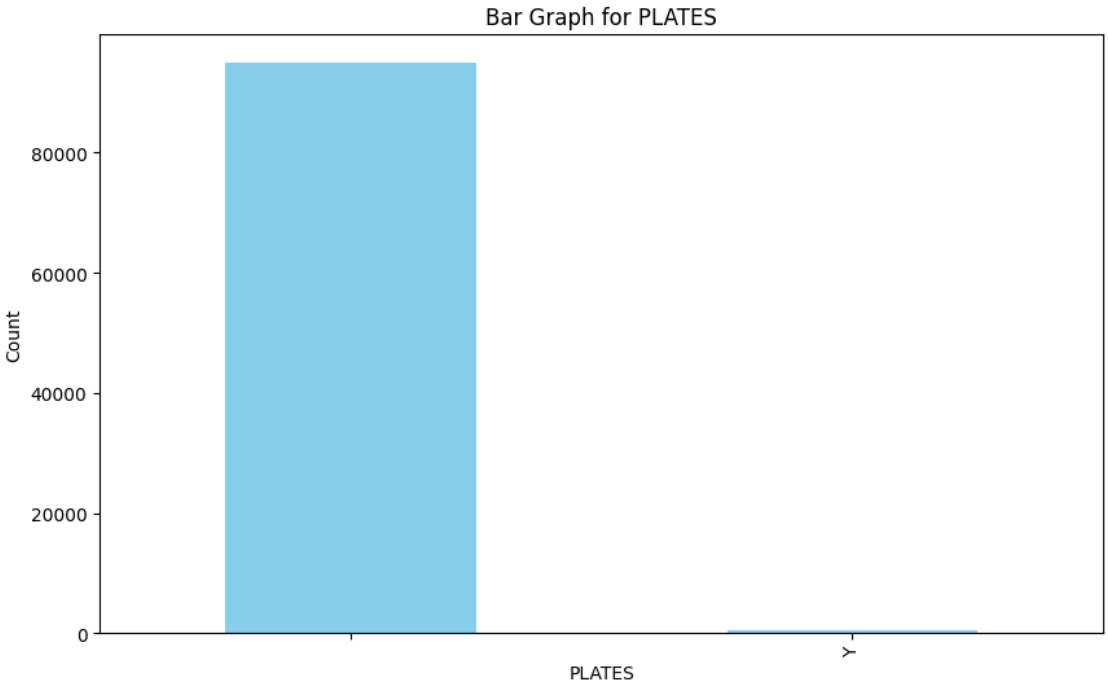

Bar Graph for GARDENIN
Count
80000
70000
60000
50000
40000
30000
20000
10000
0
Y
GARDENIN

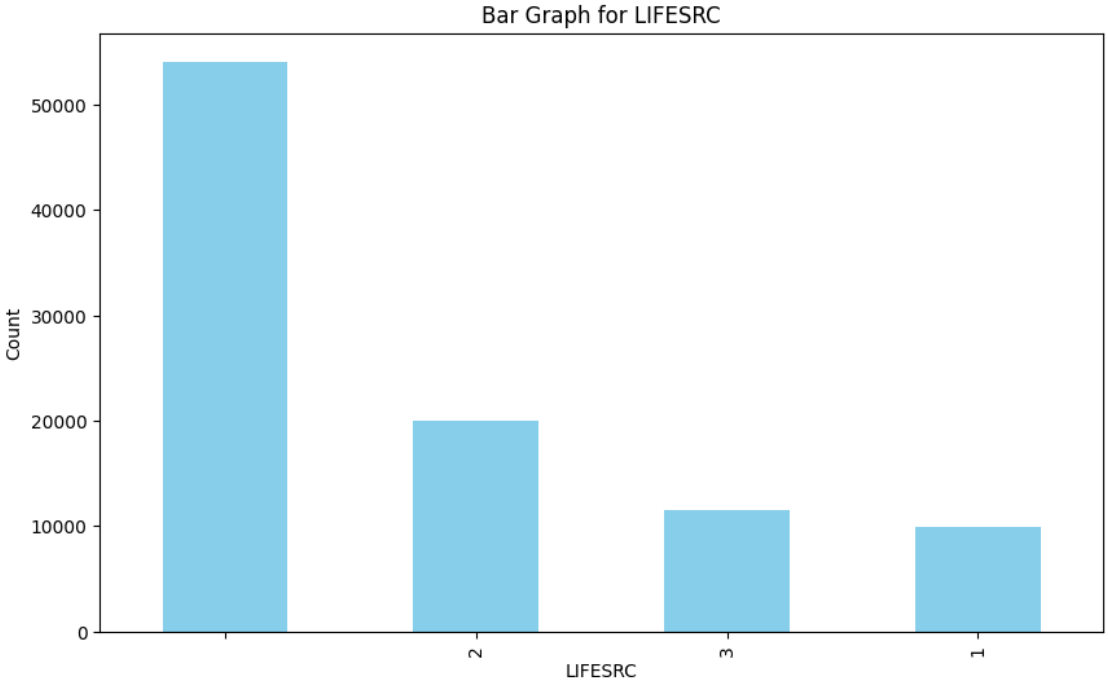

Bar Graph for BOATS
Count
80000
60000
40000
20000
0
Y
BOATS

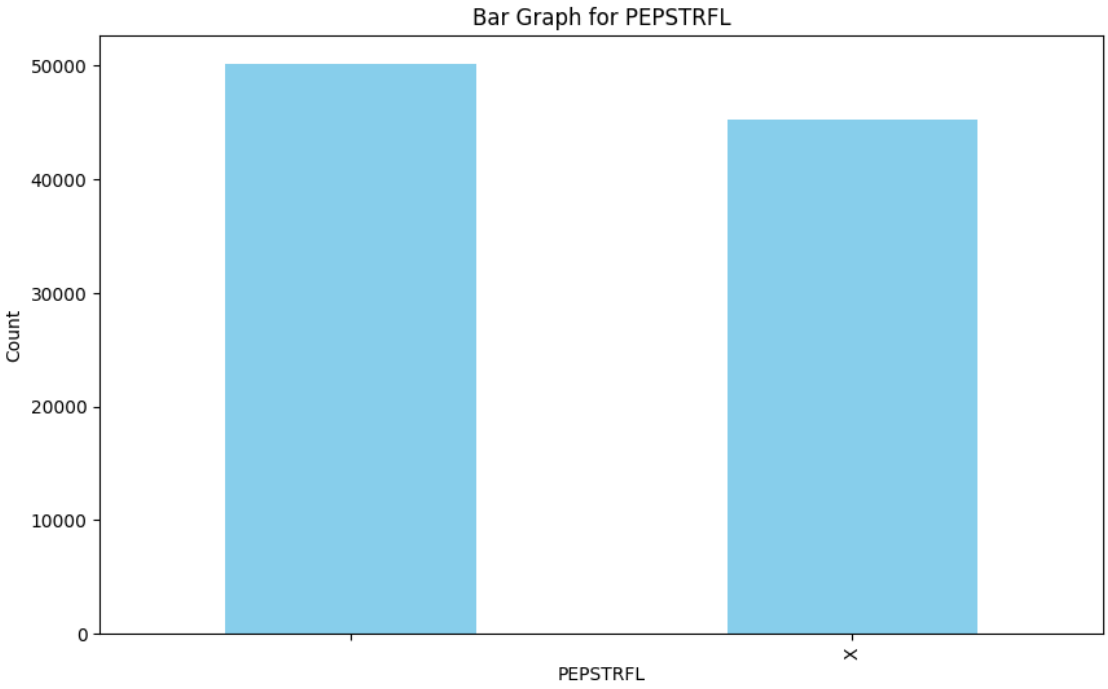

Bar Graph for WALKER
Count
80000
70000
60000
50000
40000
30000
20000
10000
0
Y
WALKER

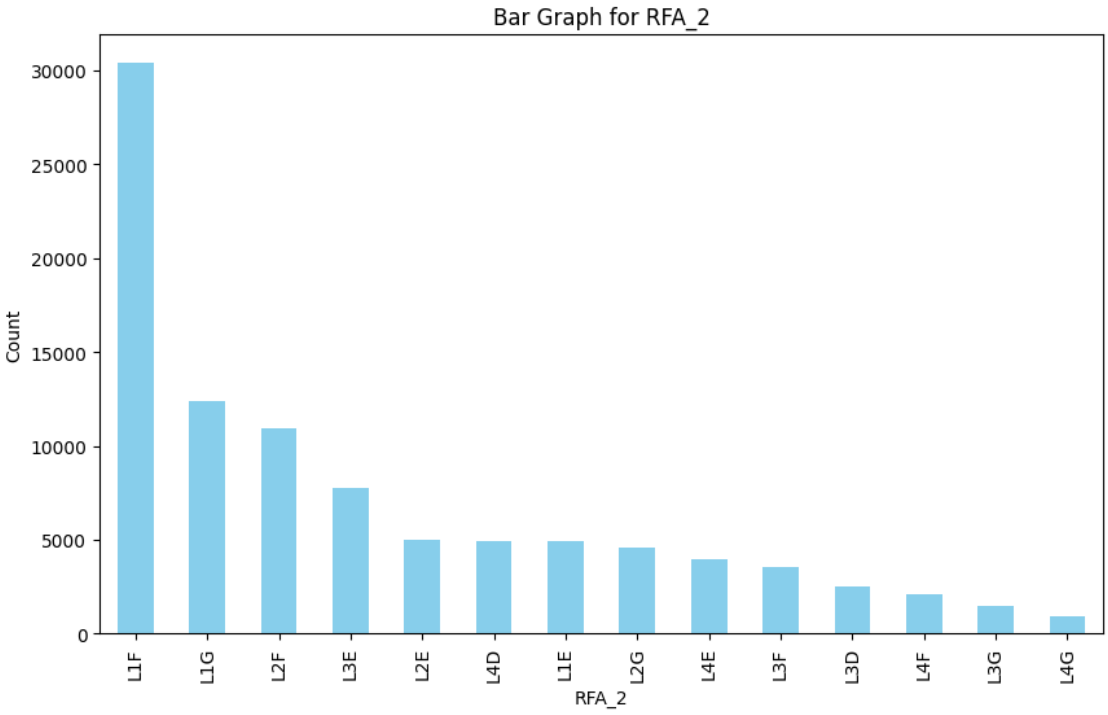

Bar Graph for KIDSTUFF
Count
80000
60000
40000
20000
0
Y
KIDSTUFF

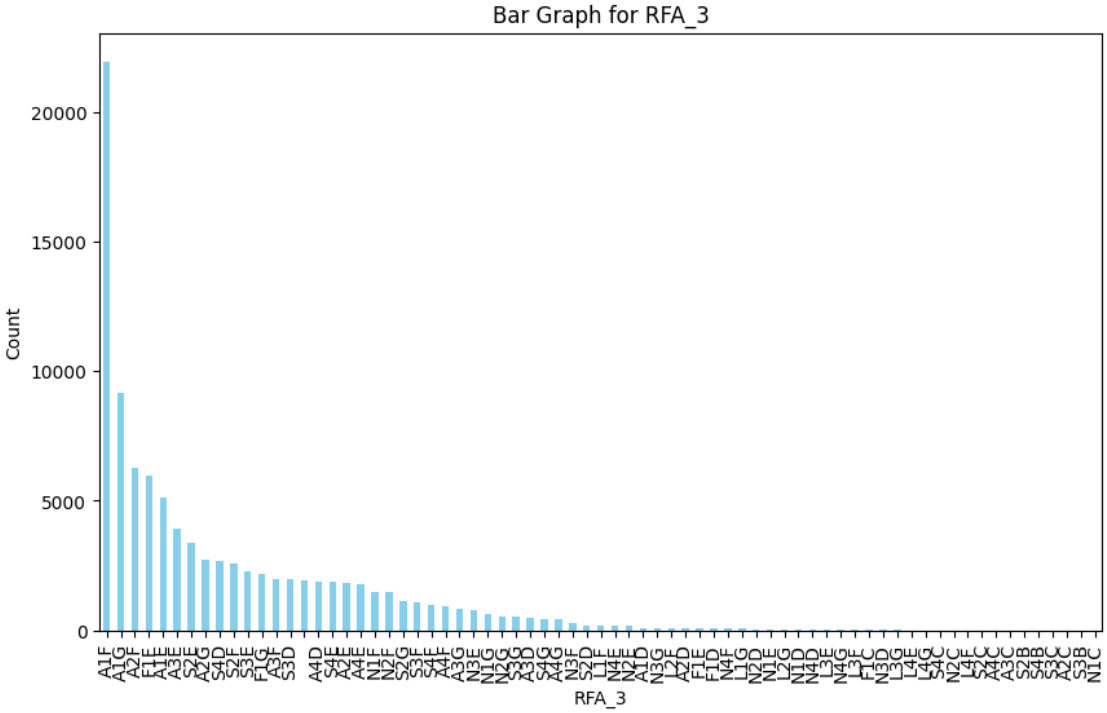

Bar Graph for CARDS
Count
80000
60000
40000
20000
0
Y
CARDS

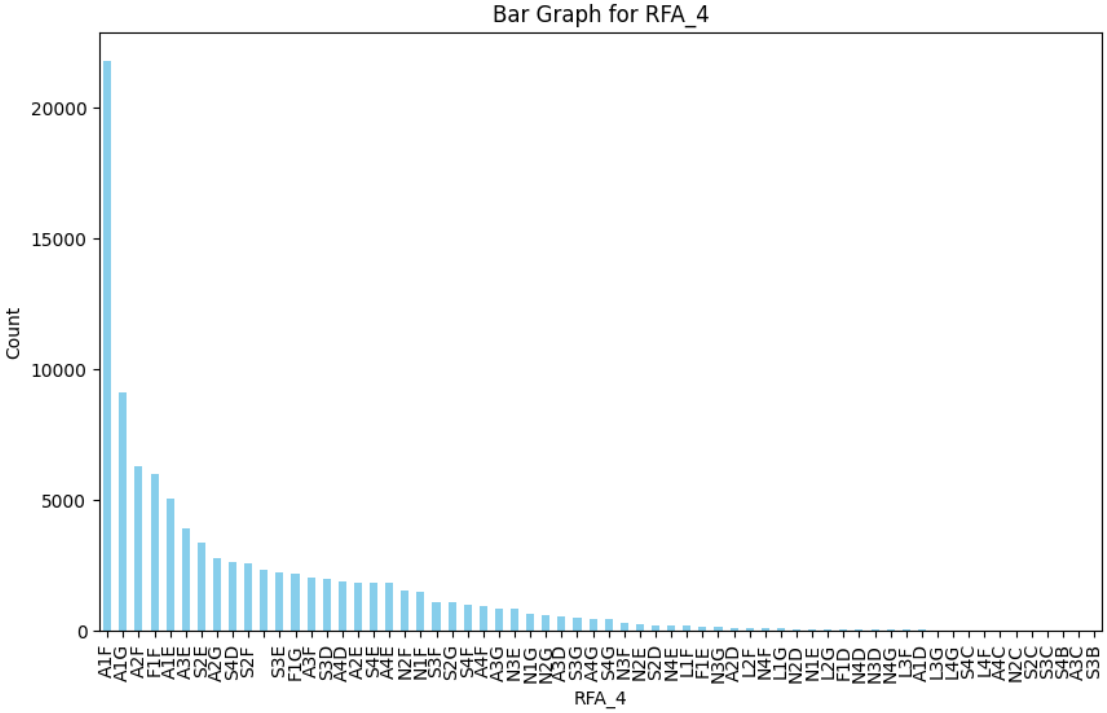

Bar Graph for PLATES
Count
80000
60000
40000
20000
0
Y
PLATES

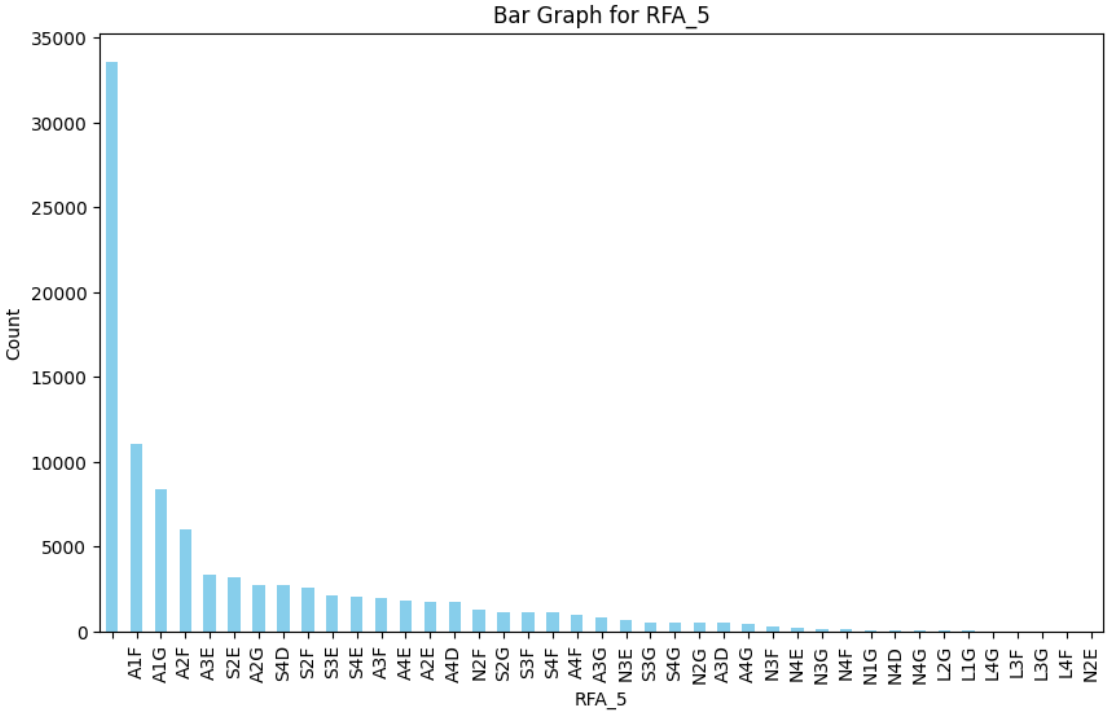

Bar Graph for LIFESRC
Count
50000
40000
30000
20000
10000
0
2
3
1
LIFESRC

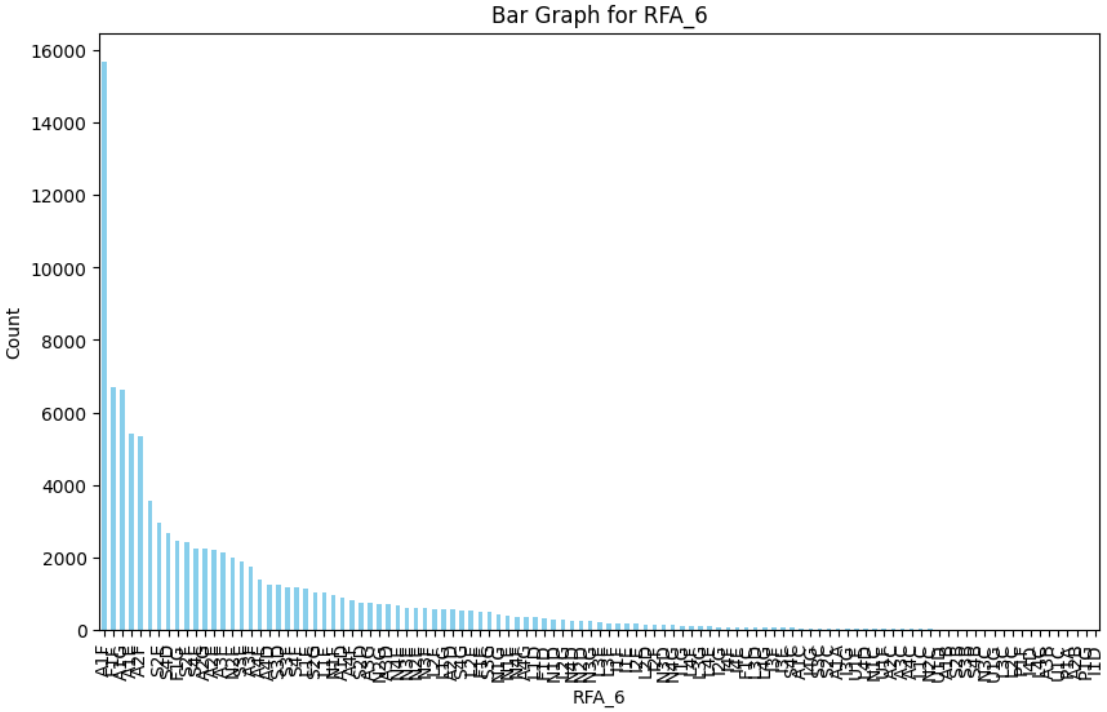

Bar Graph for PEPSTRFL
Count
50000
40000
30000
20000
10000
0
X
PEPSTRFL

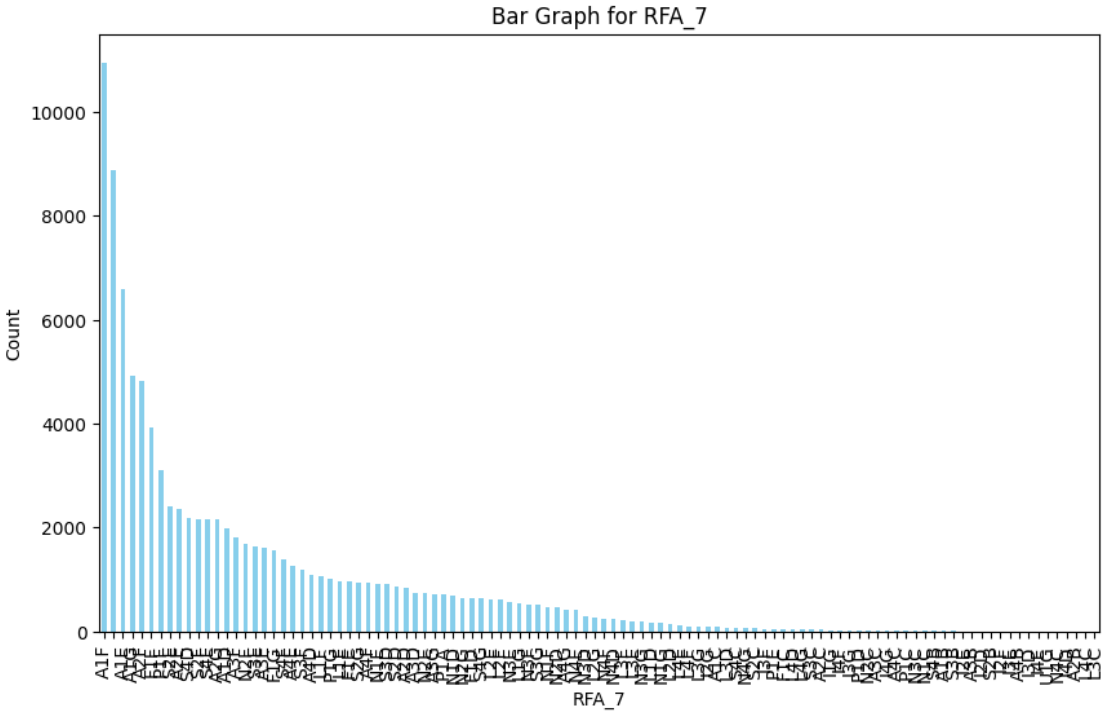

Bar Graph for RFA_2
Count
30000
25000
20000
15000
10000
5000
0
L1F
L1G
L2F
L3E
L2E
L4D
L1E
L2G
L4E
L3F
L3D
L4F
L3G
L4G
RFA_2

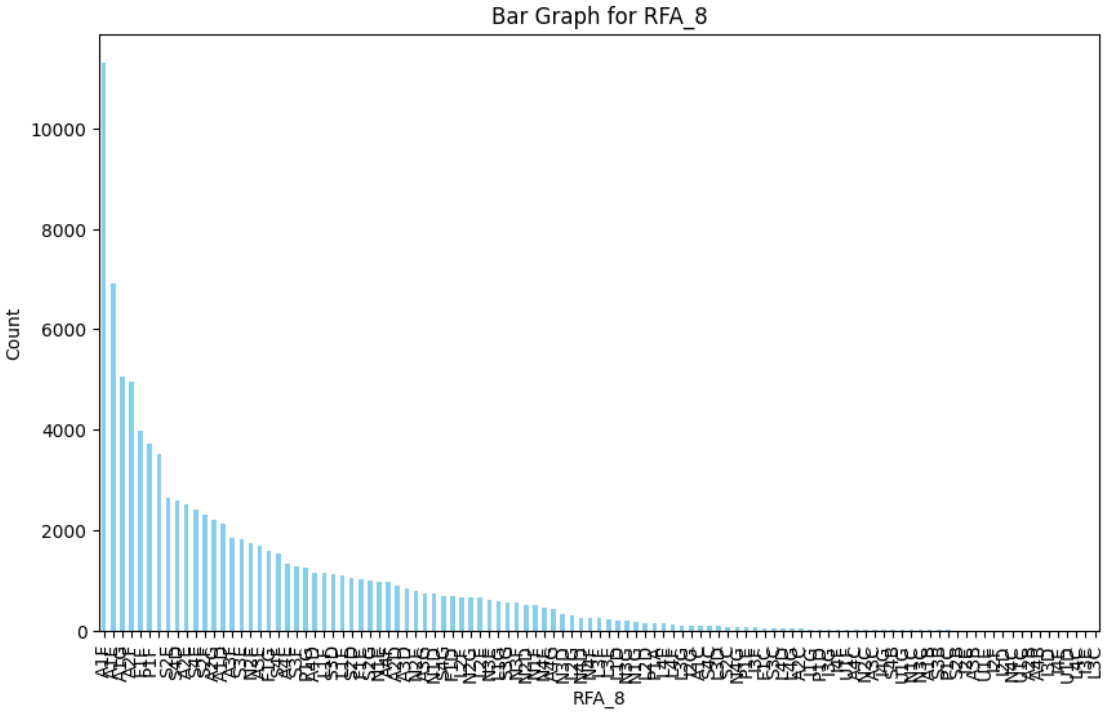

Bar Graph for RFA_3
Count
RFA_3
0
5000
10000
15000
20000
A1F A1G A2F F1E A1E A3E S2E A2G S4D S2F S3E F1G A3F S3D A4D S4E A2E A4E N1F N2F S2G S3F S4F A3G N3E N1G N2G S3G A3D S4G A4G N3F S2D L1F N4E N2E A1D N3G L2F A2D F1E F1D N4F L1G N2D N1E L2G N1D N4D L3E N4G L3F F1C N3D L3G L4E L4G S4C N2C S2C L4F S4F A4C A3C S2B S4B S3C A2C S3B N1C

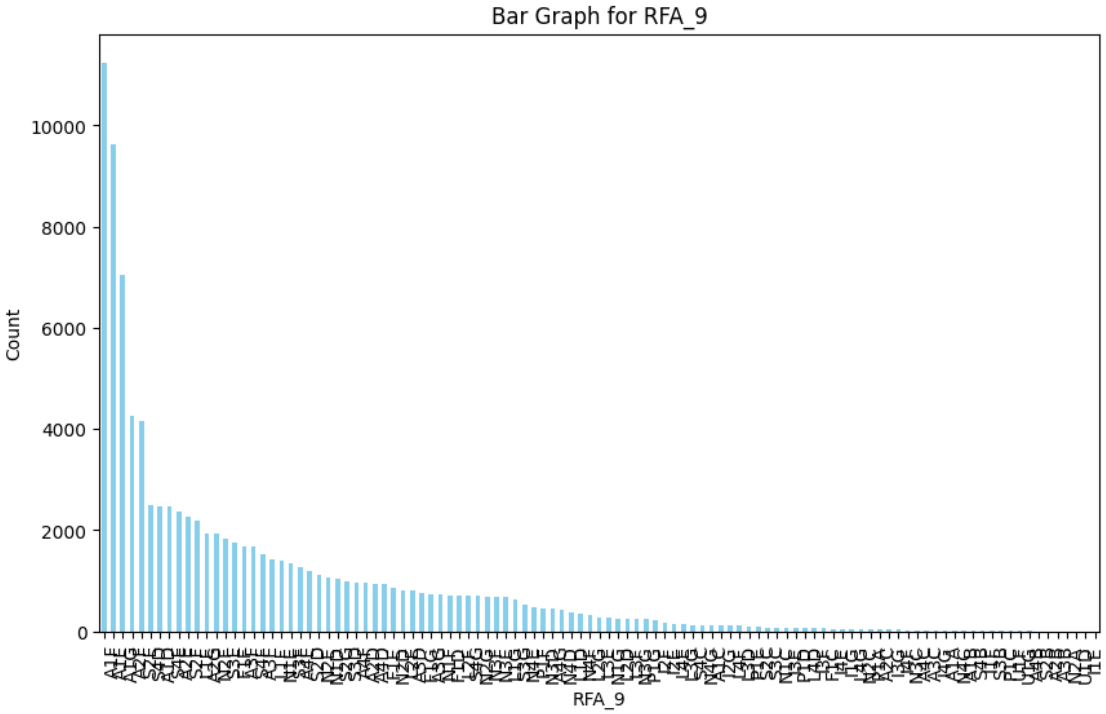

Bar Graph for RFA_4
Count
RFA_4
0
5000
10000
15000
20000
A1F A1G A2F F1E A1E A3E S2E A2G S4D S2F S3E F1G A3F S3D A4D A2E S4E A4E N2F N1F S3F S2G S4F A4F A3G N3E N1G N2G A3D S3G A4G S4G N3F N2E S2D N4E L1F F1E N3G A2D L2F N4F L1G N2D N1E L2G F1D N4D N3D N4G L3F A1D L3G L4G S4C L4F A4C N2C S2C S3C S4B A3C S3B

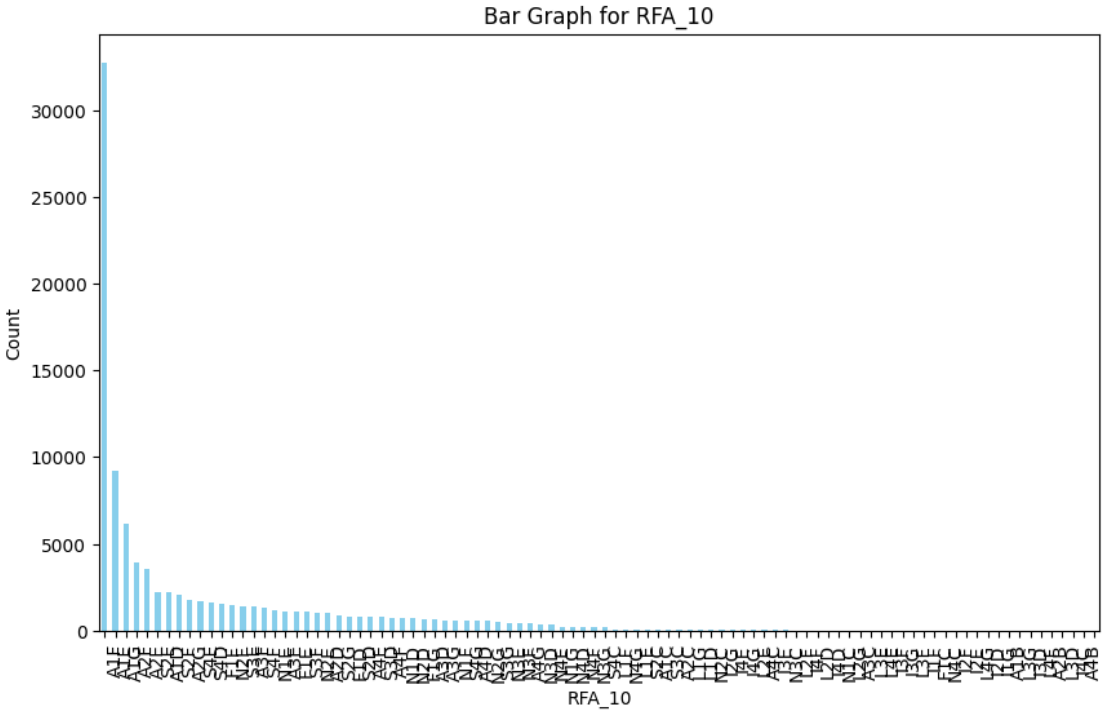

Bar Graph for RFA_5
Count
RFA_5
35000
30000
25000
20000
15000
10000
5000
0
A1F
A1G
A2F
A3E
S2E
A2G
S4D
S2F
S3E
S4E
A3F
A4E
A2E
A4D
N2F
S2G
S3F
S4F
A4F
A3G
N3E
S3G
S4G
N2G
A3D
A4G
N3F
N4E
N3G
N4F
N1G
N4D
N4G
L2G
L1G
L4G
L3F
L3G
L4F
N2E

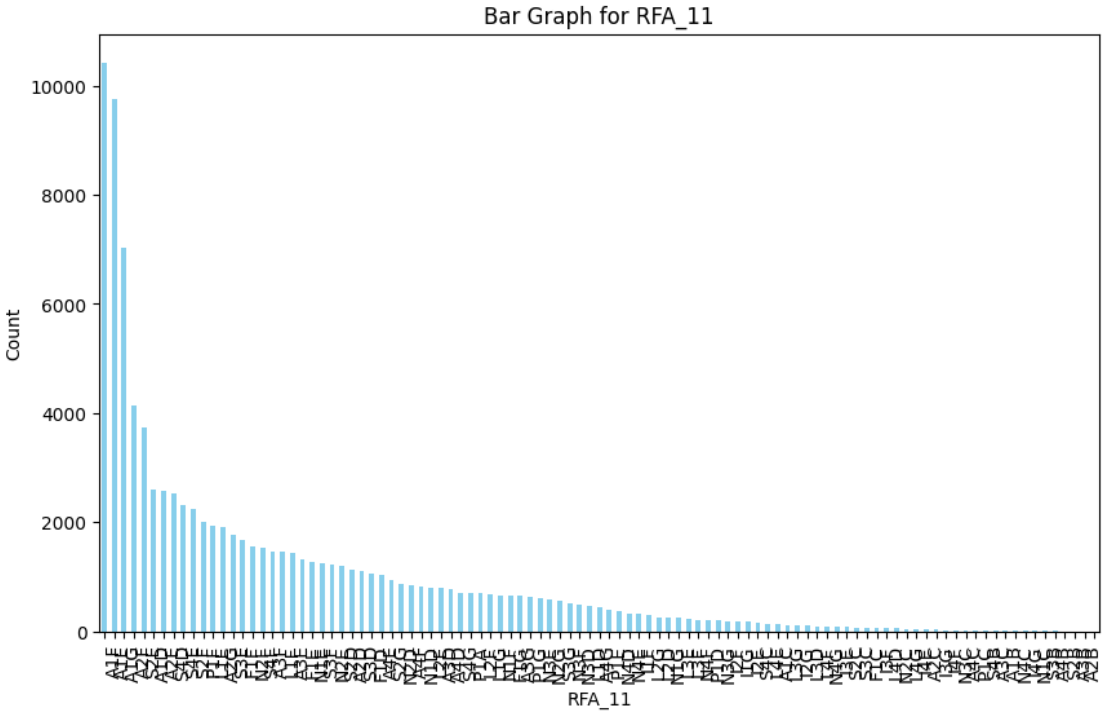

Bar Graph for RFA_6
Count
RFA_6
16000
14000
12000
10000
8000
6000
4000
2000
0

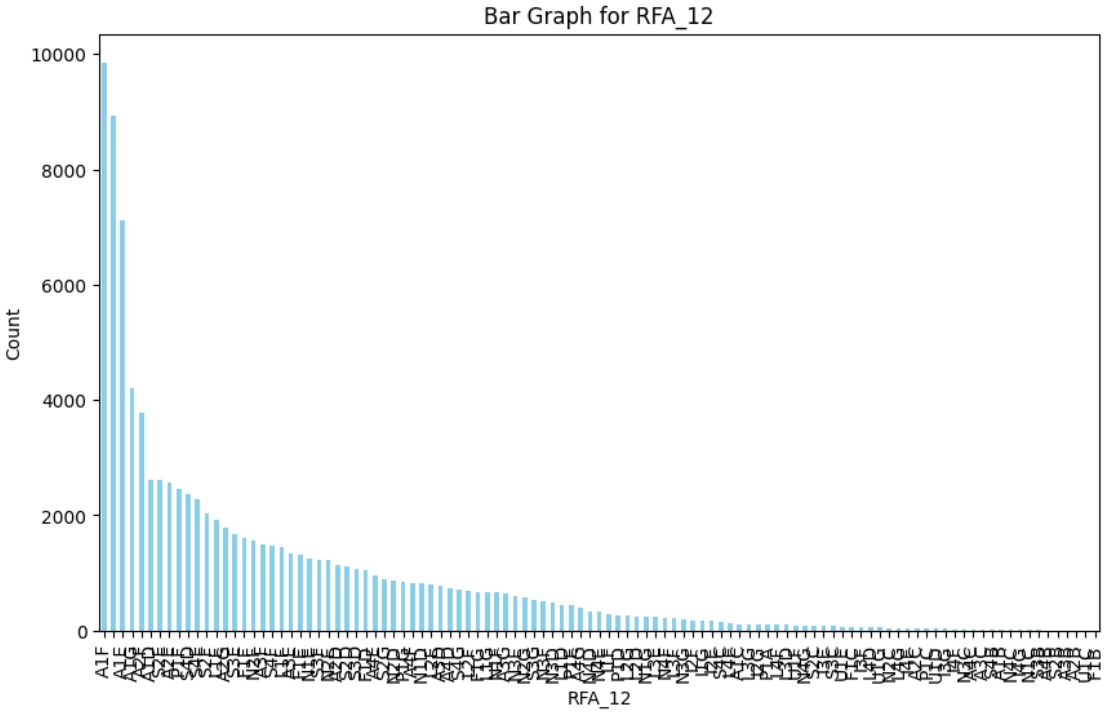

Bar Graph for RFA_7
Count
RFA_7

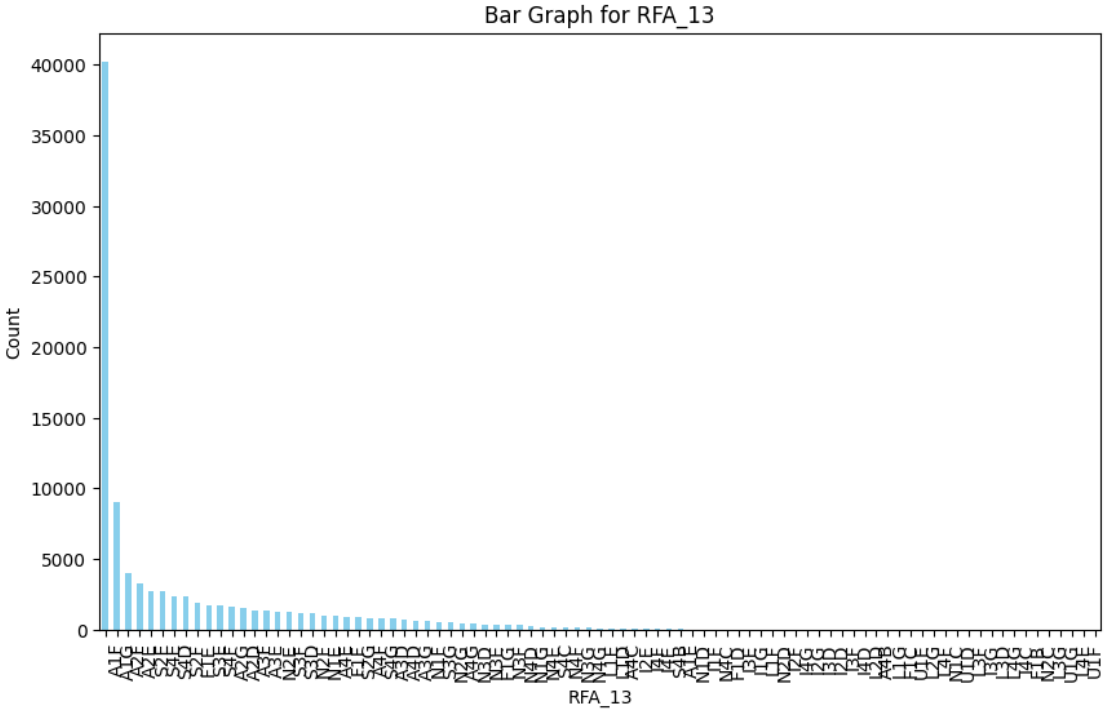

Bar Graph for RFA_8
Count
RFA_8

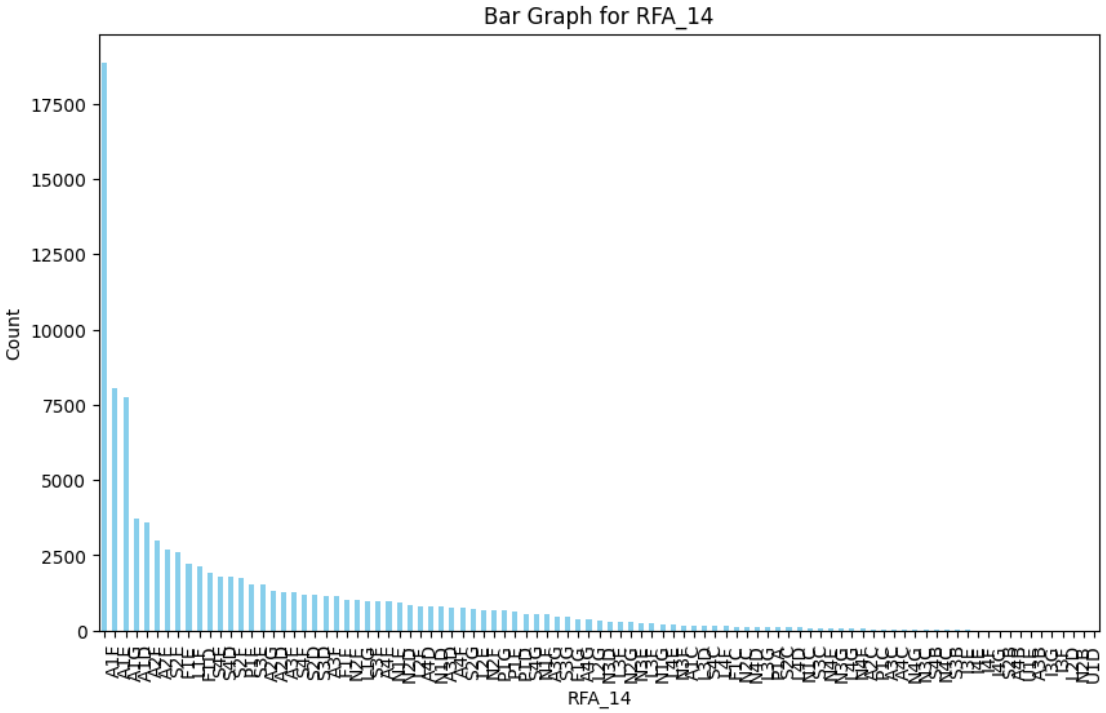

Bar Graph for RFA_9
Count
RFA_9

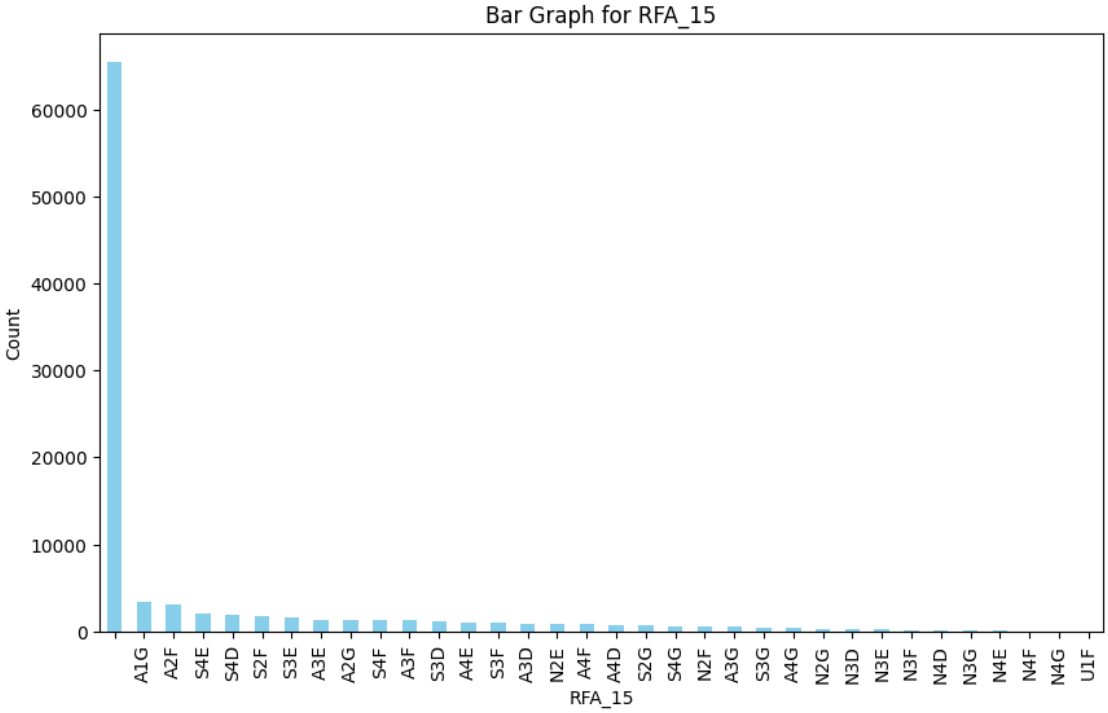

Bar Graph for RFA_10
Count
RFA_10

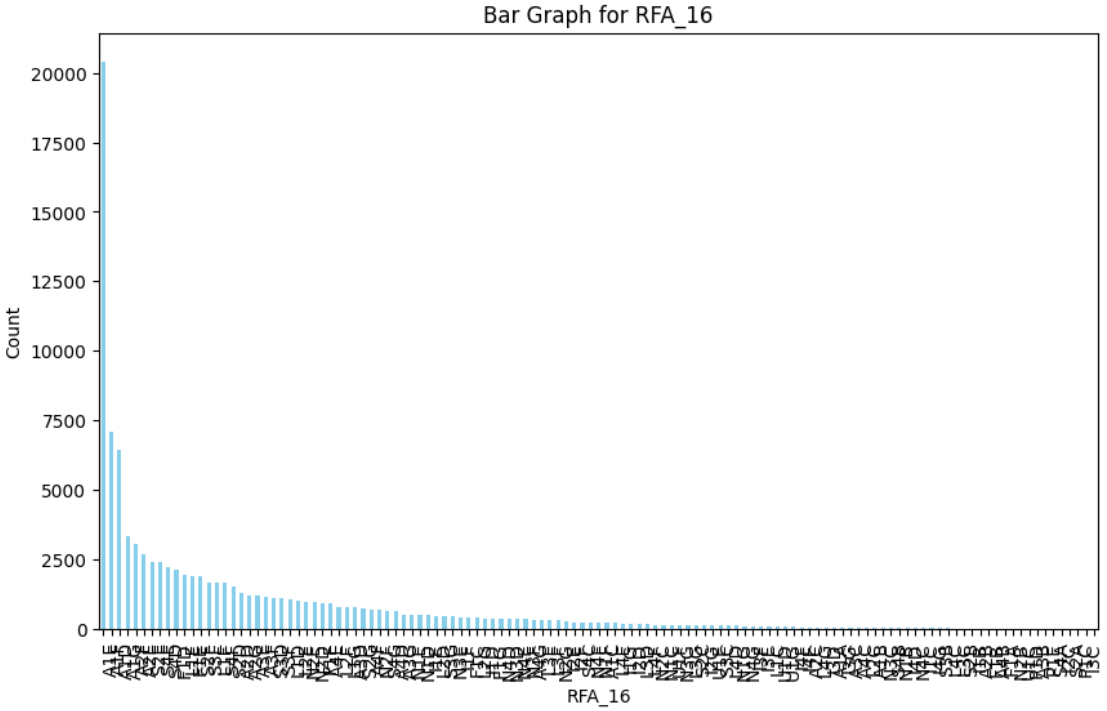

Bar Graph for RFA_11
Count
RFA_11
0
2000
4000
6000
8000
10000

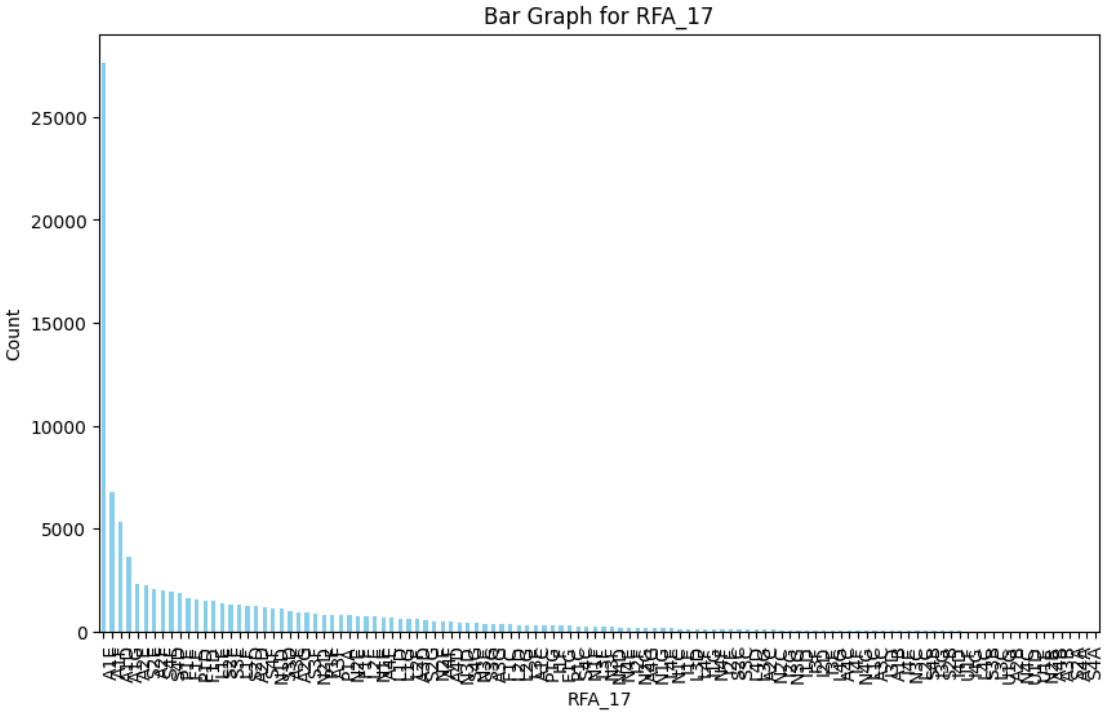

Bar Graph for RFA_12
Count
RFA_12
0
2000
4000
6000
8000
10000

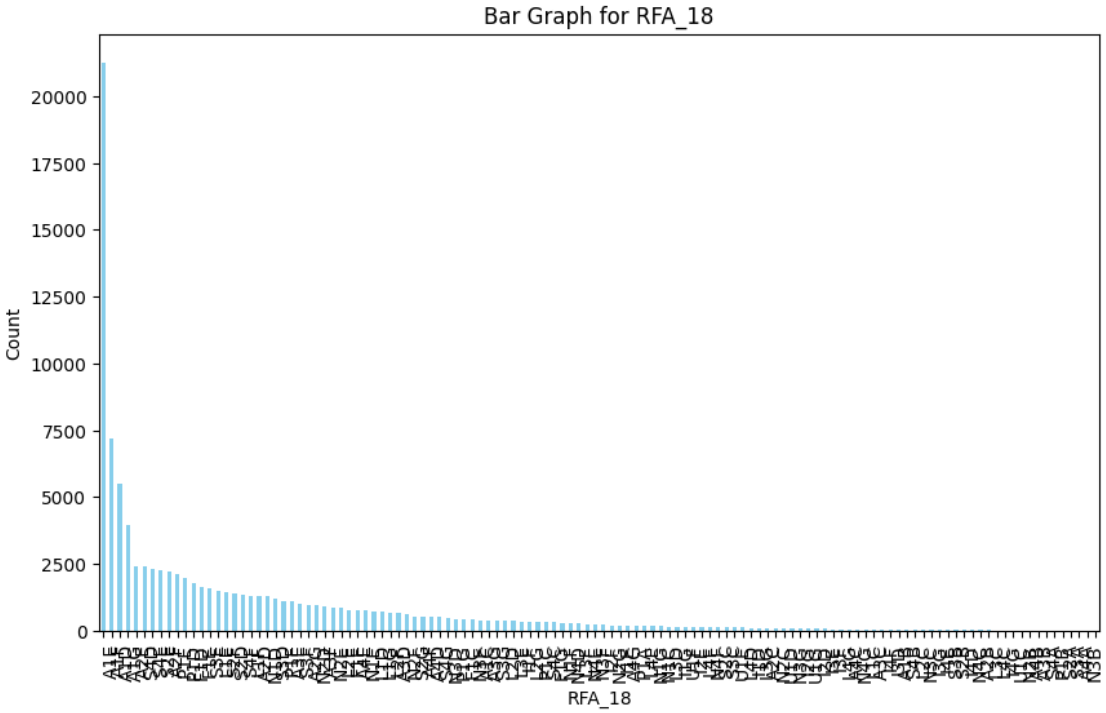

Bar Graph for RFA_13
Count
RFA_13

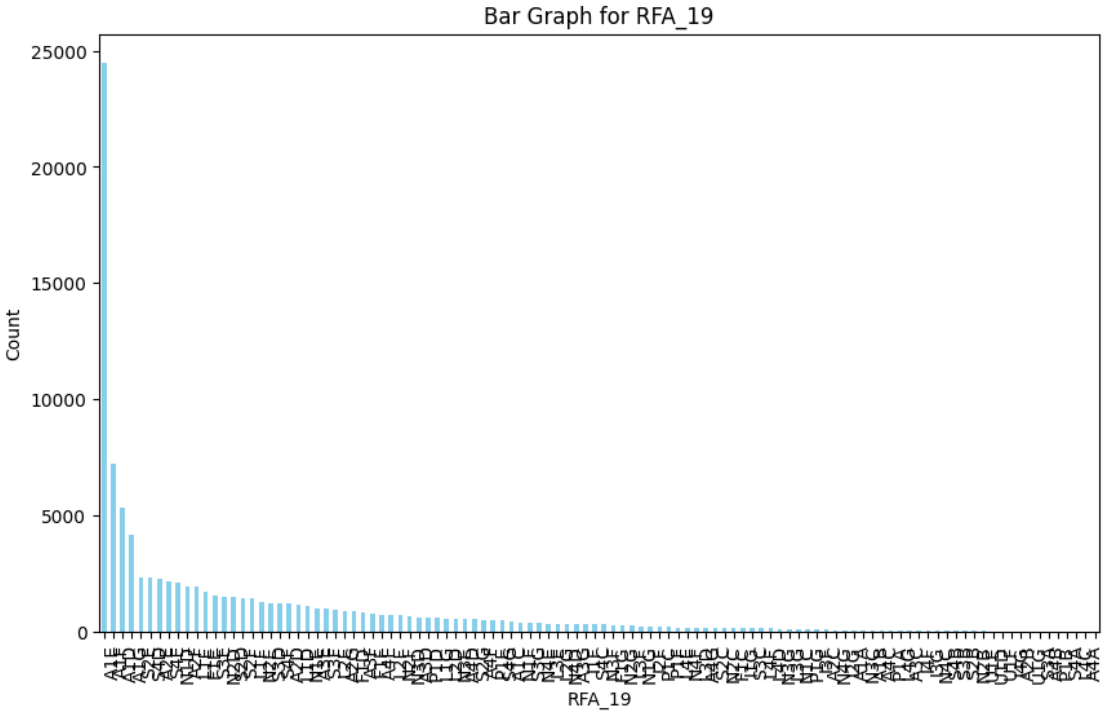

Bar Graph for RFA_14
Count
RFA_14

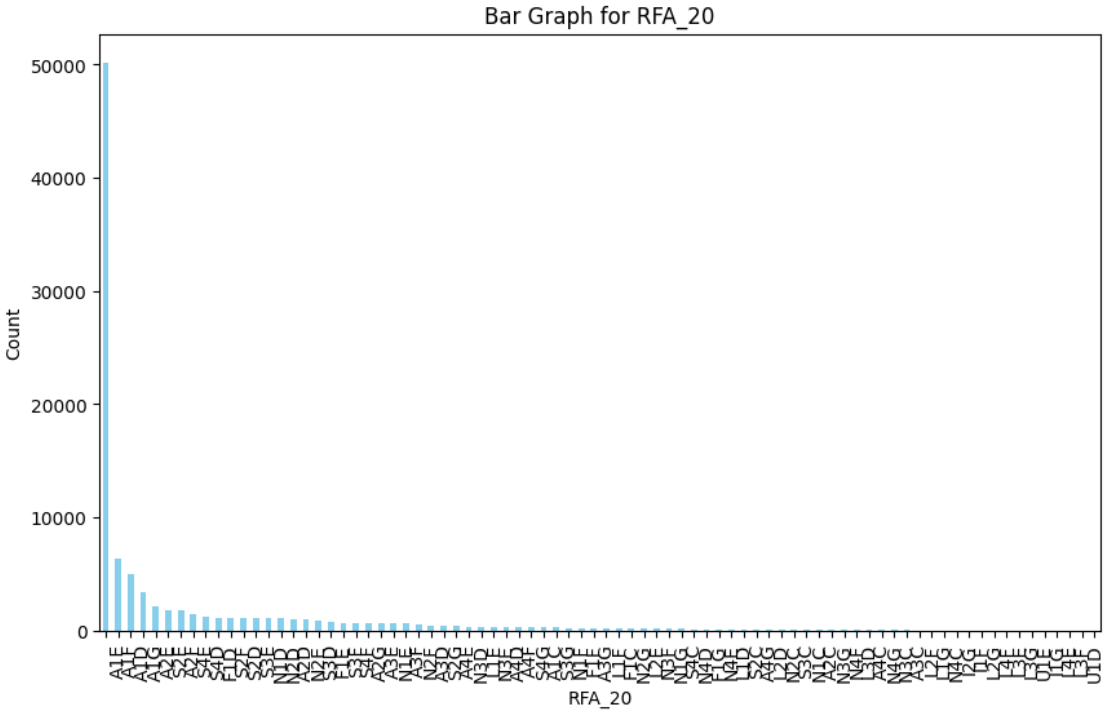

Bar Graph for RFA_15
Count
RFA_15
60000
50000
40000
30000
20000
10000
0
A1G
A2F
S4E
S4D
S2F
S3E
A3E
A2G
S4F
A3F
S3D
A4E
S3F
A3D
N2E
A4F
A4D
S2G
S4G
N2F
A3G
S3G
A4G
N2G
N3D
N3E
N3F
N4D
N3G
N4E
N4F
N4G
U1F

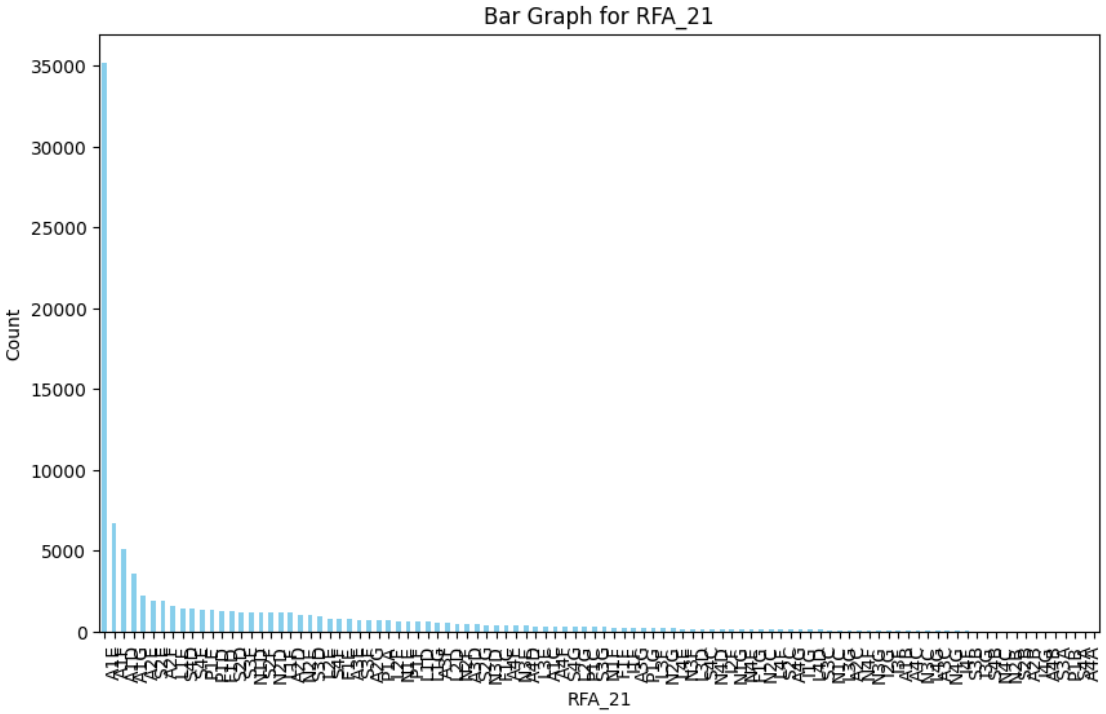

Bar Graph for RFA_16
Count
RFA_16
20000
17500
15000
12500
10000
7500
5000
2500
0

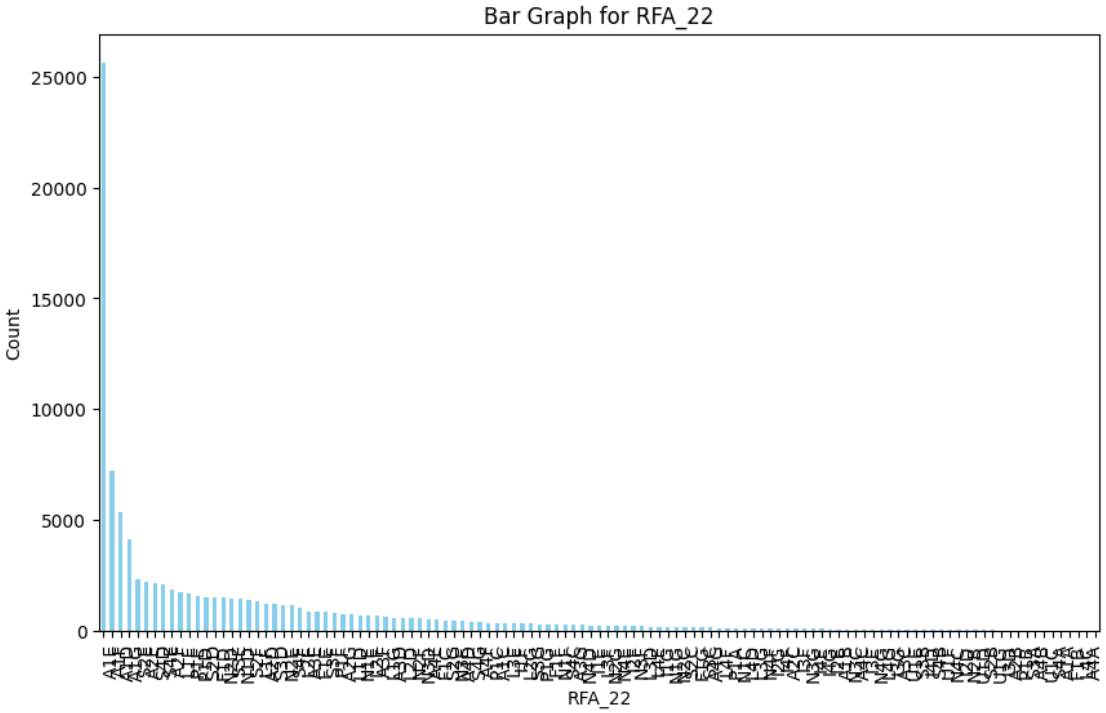

Bar Graph for RFA_17
Count
25000
20000
15000
10000
5000
0
RFA_17

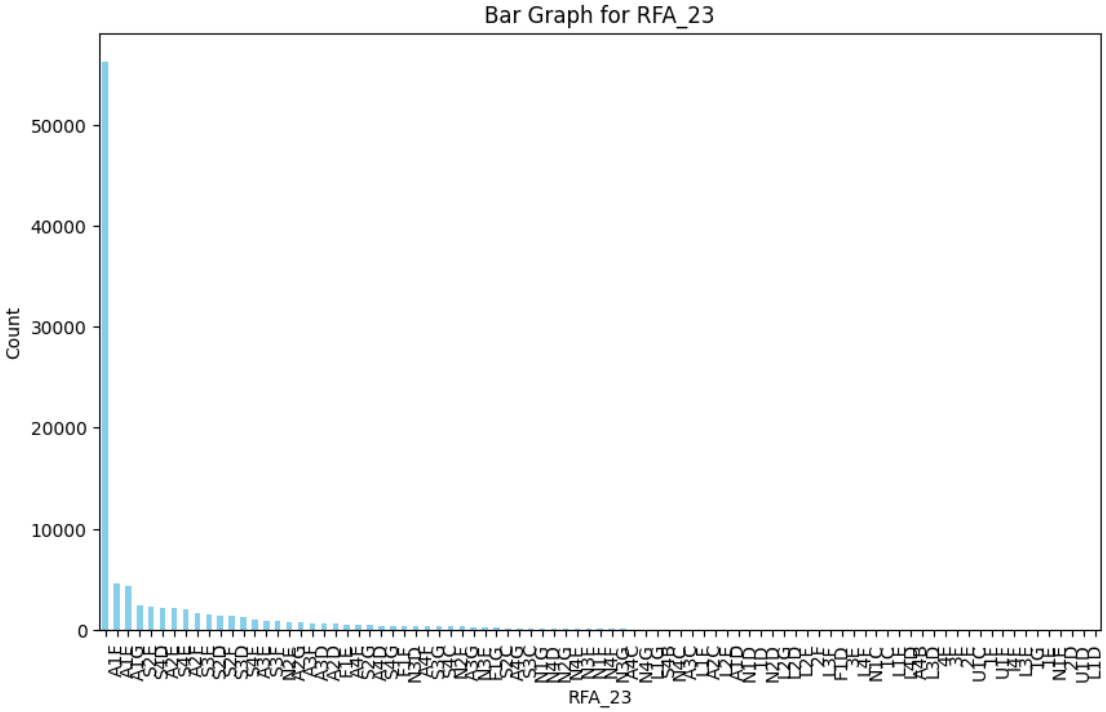

Bar Graph for RFA_18
Count
20000
17500
15000
12500
10000
7500
5000
2500
0
RFA_18

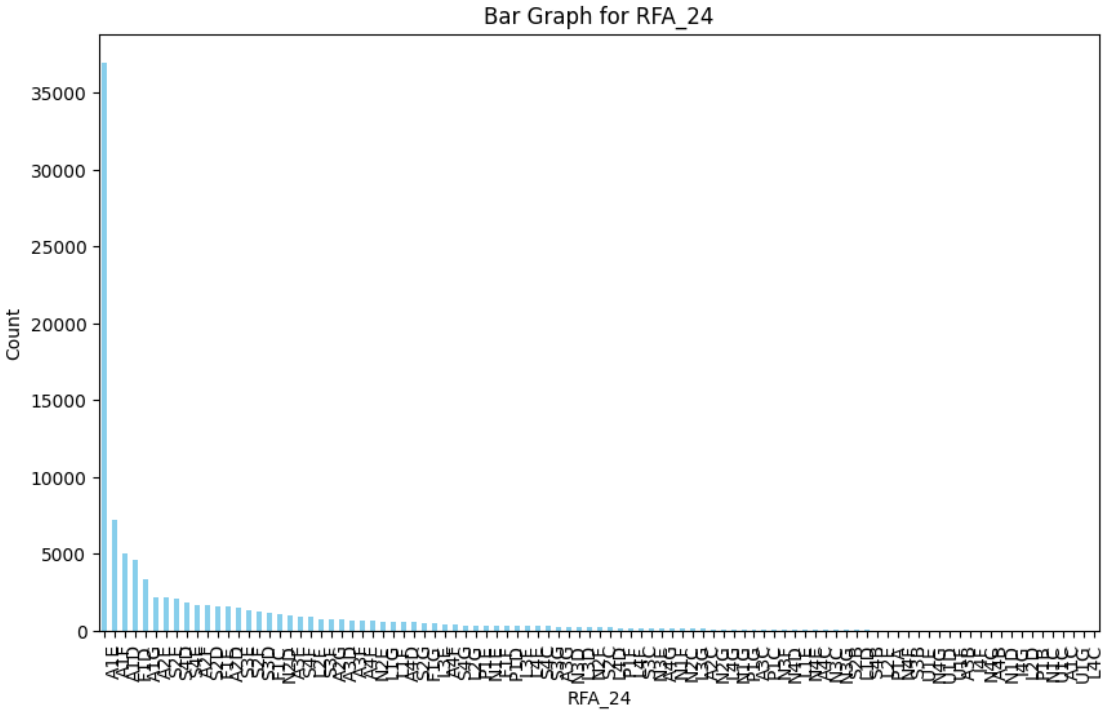

Bar Graph for RFA_19
Count
25000
20000
15000
10000
5000
0
RFA_19

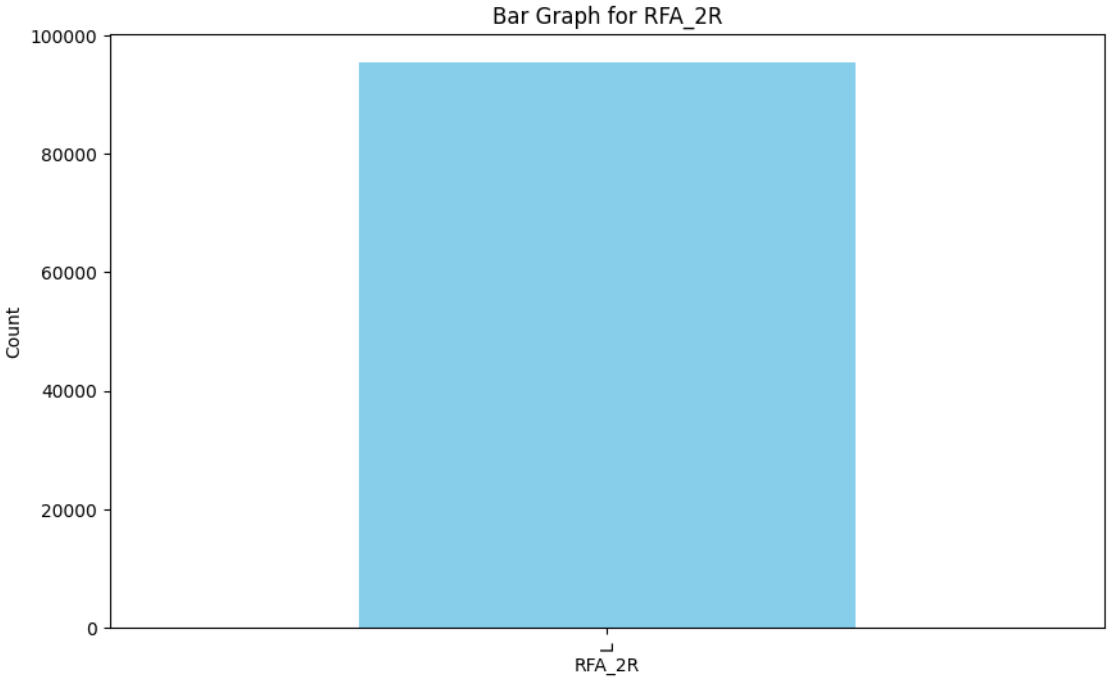

Bar Graph for RFA_20
Count
50000
40000
30000
20000
10000
0
RFA_20

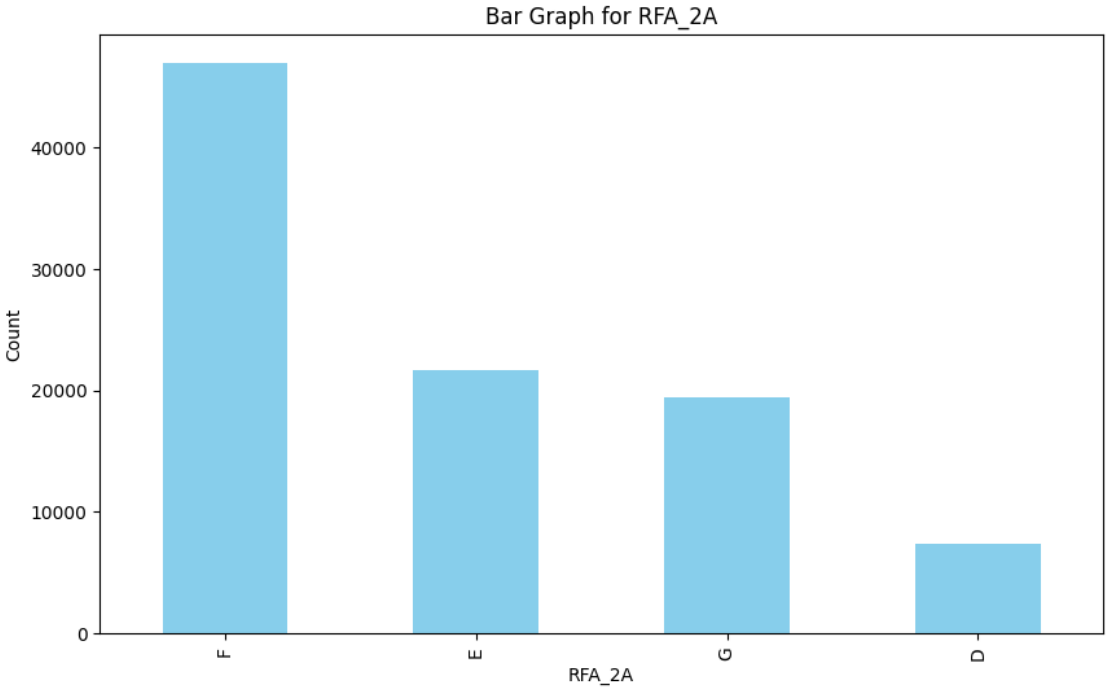

Bar Graph for RFA_21
Count
RFA_21

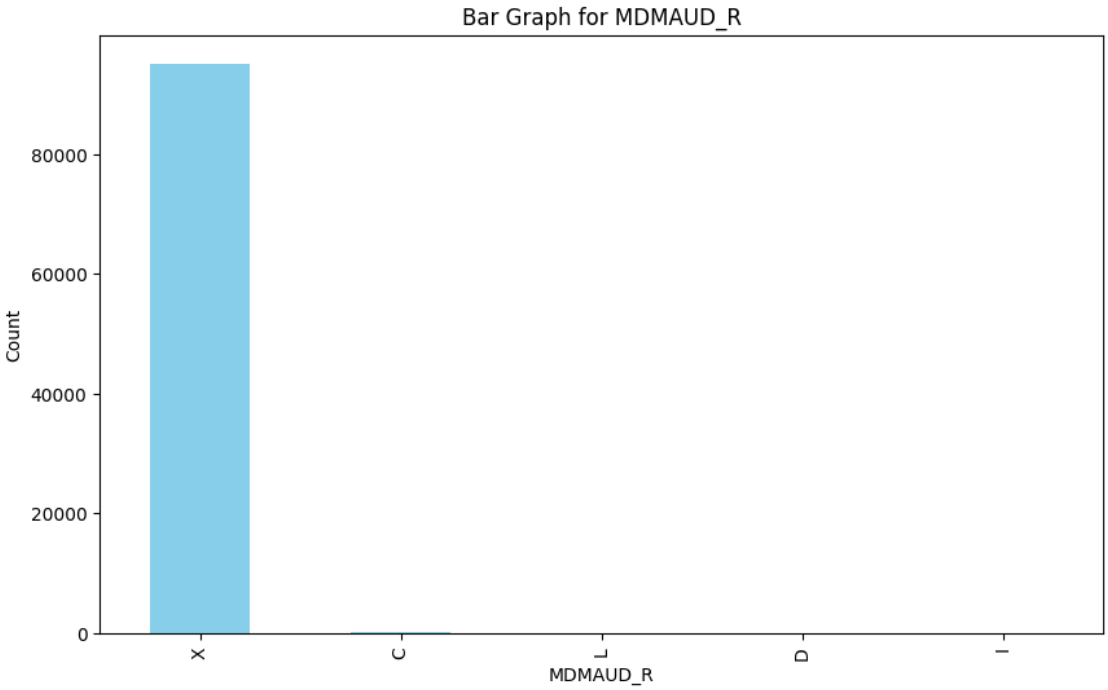

Bar Graph for RFA_22
Count
RFA_22

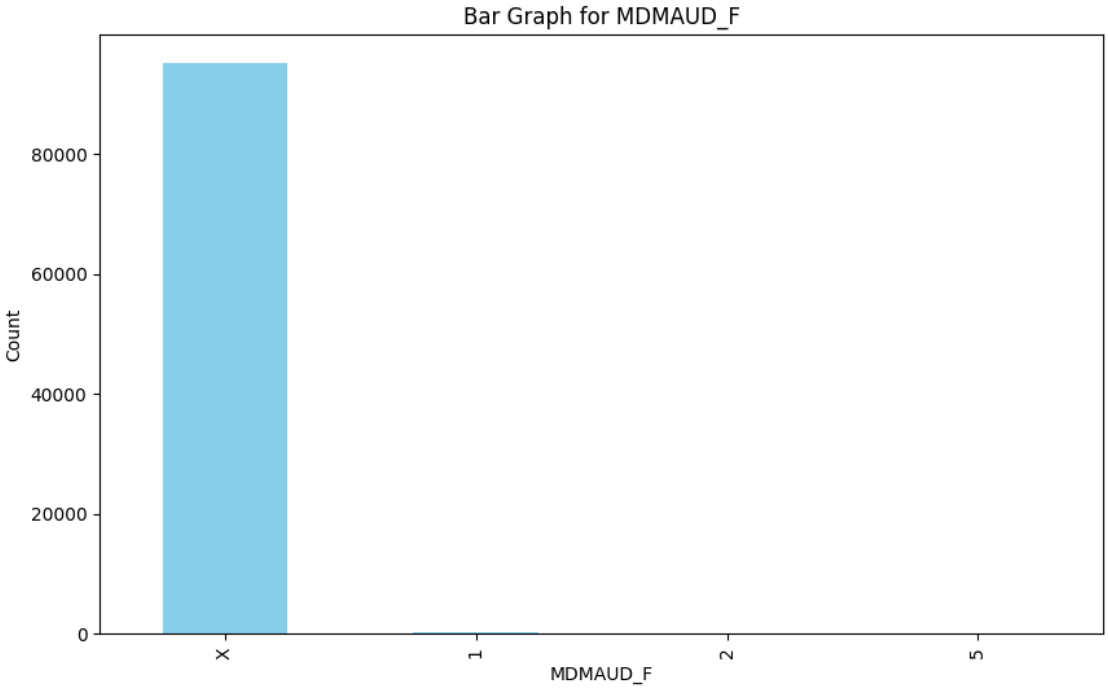

Bar Graph for RFA_23
Count
RFA_23

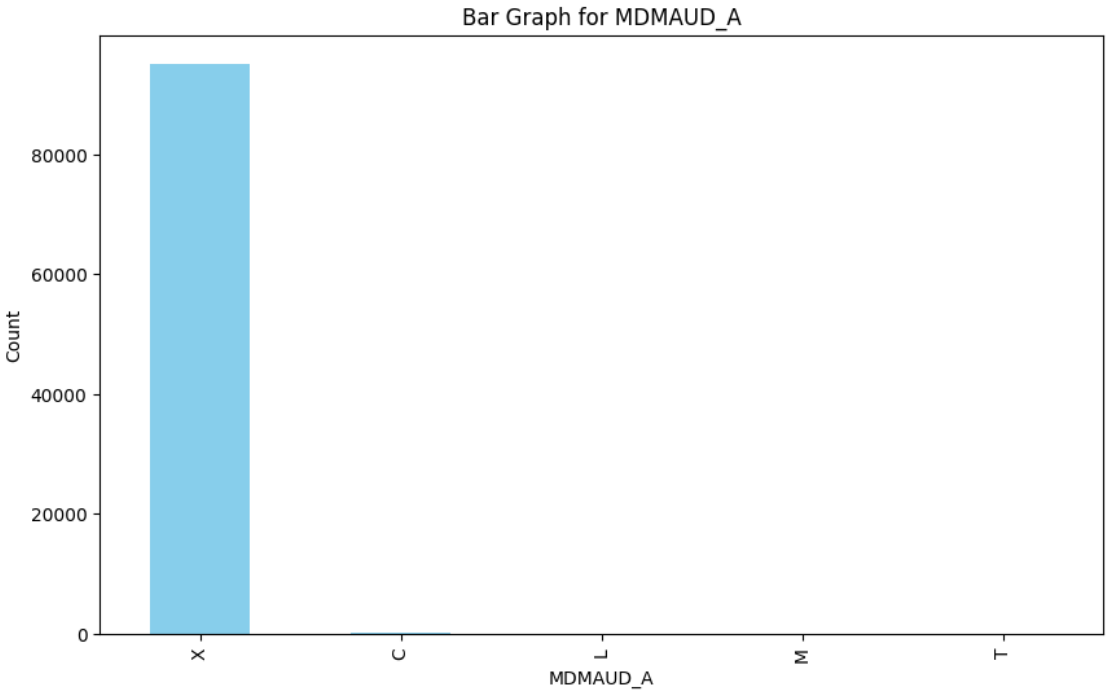

Bar Graph for RFA_24
Count
RFA_24

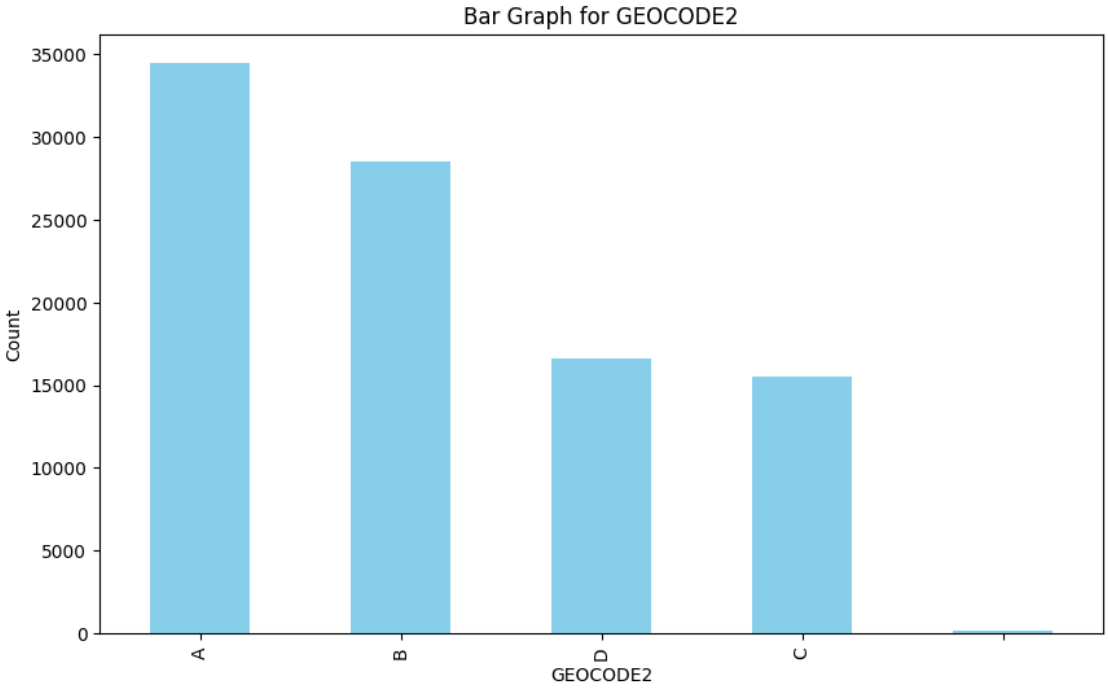

Bar Graph for RFA_2R
100000
80000
60000
40000
20000
0
Count
L
RFA_2R


Bar Graph for RFA_2A
40000
30000
20000
10000
0
Count
F
E
G
D
RFA_2A


Bar Graph for MDMAUD_R
Count
80000
60000
40000
20000
0
X
C
L
D
I
MDMAUD_R


Bar Graph for MDMAUD_F
Count
80000
60000
40000
20000
0
X
1
2
5
MDMAUD_F





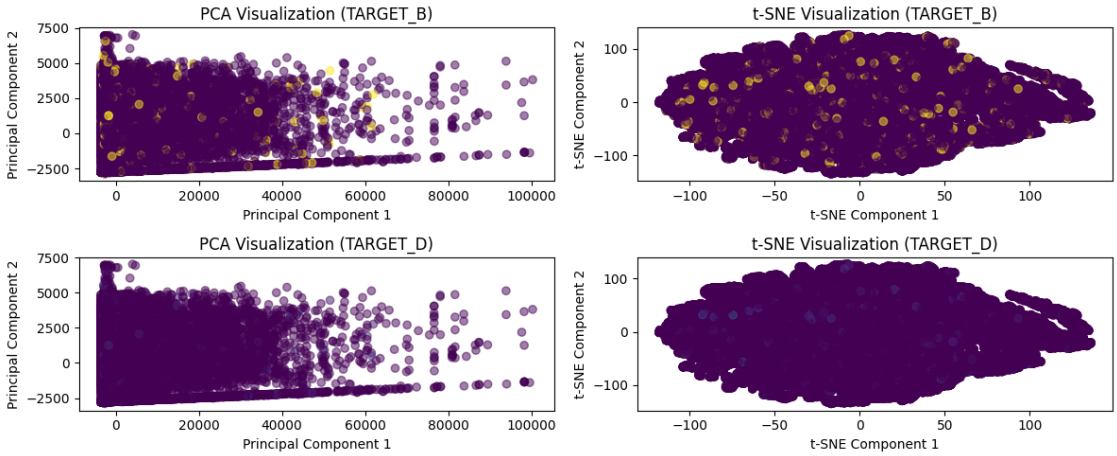
[ ]:
```
showBarGrap(test_filtered)
```

0.1.3 Dimensionality Reduction

[ ]:
```
from sklearn.decomposition import PCA
from sklearn.manifold import TSNE
def dim_reduction(df):
    numeric_columns = df.select_dtypes(include='number').columns
    numeric_columns = numeric_columns[:20]
    features = df.drop(['TARGET_B', 'TARGET_D'], axis=1)
    features = features[numeric_columns]
    target_b = df['TARGET_B']
    target_d = df['TARGET_D']

    pca = PCA(n_components=2)
    pca_result = pca.fit_transform(features)

    tsne = TSNE(n_components=2, random_state=42)
    tsne_result = tsne.fit_transform(features)

    plt.figure(figsize=(12, 5))

    plt.subplot(2, 2, 1)
    plt.scatter(pca_result[:, 0], pca_result[:, 1], c=target_b, cmap='viridis',␣
 ↪alpha=0.5)
    plt.title('PCA Visualization (TARGET_B)')
    plt.xlabel('Principal Component 1')
    plt.ylabel('Principal Component 2')

    # Visualize t-SNE result
    plt.subplot(2, 2, 2)
    plt.scatter(tsne_result[:, 0], tsne_result[:, 1], c=target_b,␣
 ↪cmap='viridis', alpha=0.5)
    plt.title('t-SNE Visualization (TARGET_B)')
    plt.xlabel('t-SNE Component 1')
    plt.ylabel('t-SNE Component 2')


    plt.subplot(2, 2, 3)
    plt.scatter(pca_result[:, 0], pca_result[:, 1], c=target_d, cmap='viridis',␣
 ↪alpha=0.5)
    plt.title('PCA Visualization (TARGET_D)')
    plt.xlabel('Principal Component 1')
    plt.ylabel('Principal Component 2')

    # Visualize t-SNE result
    plt.subplot(2, 2, 4)
    plt.scatter(tsne_result[:, 0], tsne_result[:, 1], c=target_d,␣
 ↪cmap='viridis', alpha=0.5)
```

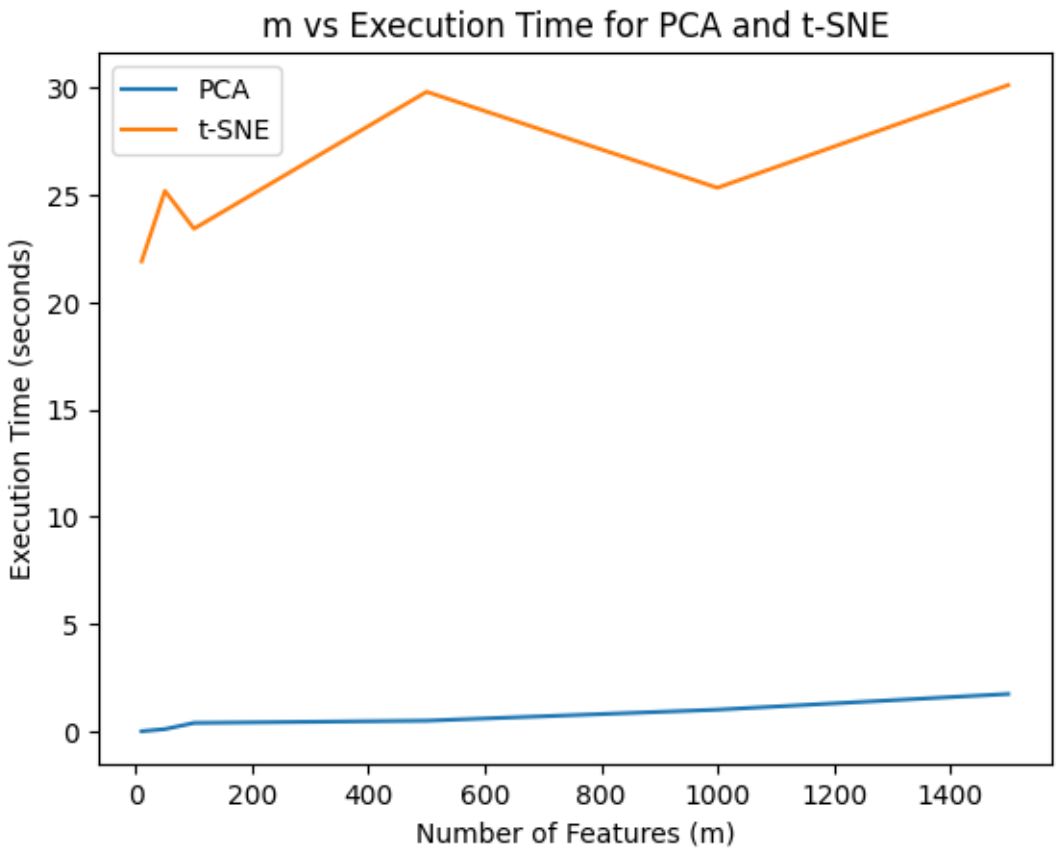
```
    plt.title('t-SNE Visualization (TARGET_D)')
    plt.xlabel('t-SNE Component 1')
    plt.ylabel('t-SNE Component 2')

    plt.tight_layout()
    plt.show()
```

[ ]:
```
dim_reduction(train_filtered)
```



[ ]:
```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
from sklearn.manifold import TSNE
import timeit

# Function to generate random dataset
def generate_dataset(n, m):
    return np.random.randn(n, m)

# Function to perform PCA and measure execution time
def perform_pca(data):
    pca = PCA(n_components=2)
    return pca.fit_transform(data)

# Function to perform t-SNE and measure execution time
def perform_tsne(data):
    tsne = TSNE(n_components=2)
    return tsne.fit_transform(data)

# Define different values of m
m_values = [10, 50, 100, 500, 1000, 1500]
```

```
# Initialize lists to store execution times for PCA and t-SNE
pca_execution_times = []
tsne_execution_times = []

# Define the number of data points (n)
n = 10000

# Iterate over different values of m
for m in m_values:
    # Generate a random dataset with n data points and m features
    dataset = generate_dataset(n, m)

    # Measure execution time for PCA
    pca_time = timeit.timeit(lambda: perform_pca(dataset), number=1)
    pca_execution_times.append(pca_time)

    # Measure execution time for t-SNE
    tsne_time = timeit.timeit(lambda: perform_tsne(dataset), number=1)
    tsne_execution_times.append(tsne_time)

# Plotting m vs execution time for PCA and t-SNE
plt.plot(m_values, pca_execution_times, label='PCA')
plt.plot(m_values, tsne_execution_times, label='t-SNE')
plt.xlabel('Number of Features (m)')
plt.ylabel('Execution Time (seconds)')
plt.title('m vs Execution Time for PCA and t-SNE')
plt.legend()
plt.show()
```



[ ]:

[ ]:

[ ]:
```
columnNames = [col for col in train if train[col].nunique() > 1]
```

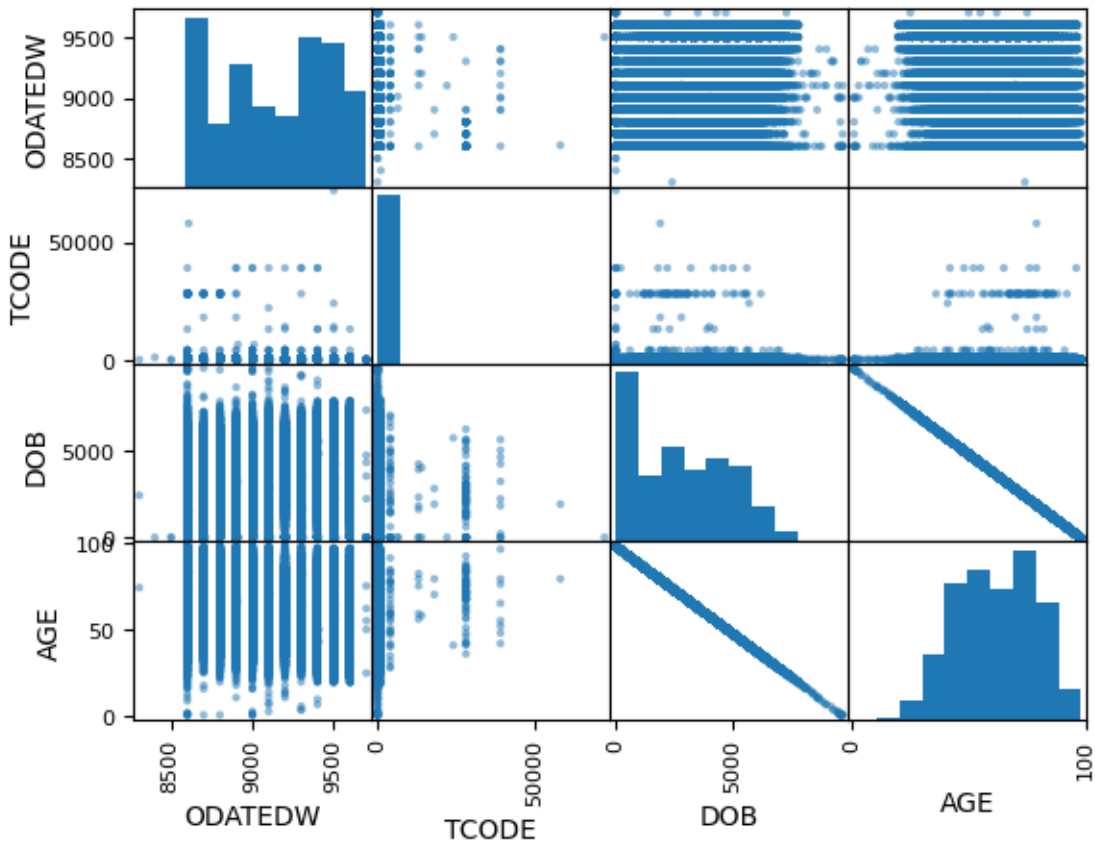
[ ]:
```
len(columnNames)
```

[ ]: 478

[ ]:
```
columnNames = columnNames[:20]
```

[ ]:
```
train_reduced = train[columnNames]
```

[ ]:
```
ax = pd.plotting.scatter_matrix(train_reduced)
```


ODATEDW
TCODE
DOB
AGE
9500
9000
8500
50000
0
5000
100
50
0

[ ]: