Target Variable

Removing N,U and X

Columns to keep

C\_MNTH (removing UU and XX)

**C\_MNTH** count

*<chr>* *<int>*

1 1 24703

2 10 24617

3 11 23497

4 12 23255

5 2 22286

6 3 19581

7 4 17493

8 5 20259

9 6 23849

10 7 24804

11 8 24627

12 9 23328

**C\_WDAY**

|  |
| --- |
| C\_WDAY count  *<chr>* *<int>*  1 1 37786  2 2 39399  3 3 39684  4 4 41057  5 5 45610  6 6 37163  7 7 31600  8 U 2 |
|  |
| |  | | --- | | > | |

**C\_HOUR**

count(data$C\_HOUR)

x freq

1 0 3353

2 1 2543

3 10 12235

4 11 14413

5 12 17601

6 13 17179

7 14 18929

8 15 23124

9 16 24904

10 17 23709

11 18 17087

12 19 12636

13 2 2079

14 20 9822

15 21 8759

16 22 6757

17 23 5105

18 3 1832

19 4 1472

20 5 2533

21 6 6129

22 7 11156

23 8 14972

24 9 11823

25 UU 2149

**C\_VEHS**

> count(data$C\_VEHS)

x freq

1 1 58234

2 2 165012

3 3 34636

4 4 9175

5 5 2522

6 6 904

7 7 342

8 8 223

9 9 76

10 10 174

11 11 52

12 12 76

13 13 98

14 14 29

15 15 73

16 16 72

17 17 26

18 18 57

19 20 40

20 23 25

21 24 25

22 25 45

23 27 28

24 28 34

25 33 38

26 38 78

27 45 53

28 59 63

29 70 91

**C\_CONF** (requires grouping)

|  |
| --- |
| count(data$C\_CONF)  x freq  1 1 2836  2 2 11421  3 21 79451  4 22 8964  5 23 3336  6 24 2548  7 25 871  8 3 6646  9 31 8589  10 32 2994  11 33 20271  12 34 1590  13 35 37133  14 36 23896  15 4 9646  16 41 3771  17 5 826  18 6 23555  19 QQ 15579  20 UU 8378 |
|  |
| |  | | --- | |  | |

**C\_RCFG** (group a few, ie 11,12 and and drop QQ UU XX)

C\_RCFG count

*<chr>* *<int>*

1 1 113317

2 10 81

3 2 121015

4 3 9829

5 4 794

6 5 2258

7 6 355

8 7 42

9 8 887

10 9 297

11 QQ 8608

12 UU 14818

**C\_WTHR**

C\_WTHR count

*<chr>* *<int>*

1 1 193680

2 2 24822

3 3 22824

4 4 18369

5 5 1860

6 6 5617

7 7 969

8 Q 616

9 U 3544

C\_**RSUR**

|  |
| --- |
| C\_RSUR count  *<chr>* *<int>*  1 1 180433  2 2 42929  3 3 13111  4 4 4411  5 5 17730  6 6 847  7 7 176  8 8 43  9 9 62  10 Q 10241  11 U 2318 |
|  |
| |  | | --- | | > | |

**C\_RALN**

count(data$C\_RALN)

x freq

1 1 205590

2 2 21533

3 3 16498

4 4 8050

5 5 3467

6 6 2842

7 Q 560

8 U 13761

**C\_TRAF (consider grouping)**

C\_TRAF count

*<chr>* *<int>*

1 1 77410

2 10 107

3 11 291

4 12 547

5 13 151

6 15 147

7 16 78

8 17 176

9 18 135069

10 2 993

11 3 31351

12 4 4687

13 5 100

14 6 2208

15 7 52

16 8 873

17 9 23

18 QQ 4332

19 UU 13706

V\_TYPE

V\_TYPE count

*<chr>* *<int>*

1 1 225381

2 10 39

3 11 2770

4 14 6032

5 16 412

6 17 4687

7 18 133

8 19 118

9 20 406

10 21 73

11 22 152

12 23 142

13 5 2384

14 6 4492

15 7 4621

16 8 3346

17 9 1364

18 NN 13093

19 QQ 1037

20 UU 1619

**V\_YEAR**

|  |
| --- |
| count(data$V\_YEAR )  x freq  1 1918 1  2 1920 24  3 1927 2  4 1938 2  5 1947 3  6 1950 2  7 1951 2  8 1954 2  9 1955 2  10 1956 2  11 1957 1  12 1958 3  13 1959 1  14 1960 4  15 1962 5  16 1963 3  17 1964 5  18 1965 10  19 1966 10  20 1967 8  21 1968 8  22 1969 6  23 1970 9  24 1971 7  25 1972 9  26 1973 7  27 1974 8  28 1975 12  29 1976 23  30 1977 14  31 1978 24  32 1979 27  33 1980 34  34 1981 45  35 1982 34  36 1983 54  37 1984 56  38 1985 74  39 1986 57  40 1987 116  41 1988 116  42 1989 135  43 1990 208  44 1991 228  45 1992 292  46 1993 323  47 1994 478  48 1995 604  49 1996 719  50 1997 1151  51 1998 1616  52 1999 2055  53 2000 2943  54 2001 3404  55 2002 4992  56 2003 7049  57 2004 7375  58 2005 10014  59 2006 11380  60 2007 13697  61 2008 13988  62 2009 12492  63 2010 14457  64 2011 12368  65 2012 14340  66 2013 15620  67 2014 15965  68 2015 16584  69 2016 16064  70 2017 18193  71 2018 16952  72 2019 9693  73 2020 789  74 NNNN 14153  75 UUUU 11148 |
|  |
| |  | | --- | | > | |

**P\_SEX**

|  |
| --- |
| P\_SEX count  *<chr>* *<int>*  1 F 115007  2 M 140920  3 N 582  4 U 15792 |
|  |
| |  | | --- | | > | |

**P\_AGE**

|  |
| --- |
| count(data$P\_AGE)  x freq  1 1 2899  2 10 1129  3 11 1147  4 12 1161  5 13 1161  6 14 1570  7 15 1836  8 16 3002  9 17 5104  10 18 5626  11 19 5755  12 2 1062  13 20 5682  14 21 5724  15 22 5613  16 23 5653  17 24 5535  18 25 5450  19 26 5383  20 27 5234  21 28 4891  22 29 4770  23 3 1022  24 30 4719  25 31 4582  26 32 4403  27 33 4450  28 34 4417  29 35 4492  30 36 4071  31 37 4140  32 38 4256  33 39 4128  34 4 1011  35 40 3956  36 41 3899  37 42 3814  38 43 3809  39 44 3816  40 45 3552  41 46 3672  42 47 3688  43 48 3934  44 49 3760  45 5 1074  46 50 3620  47 51 3672  48 52 3640  49 53 3703  50 54 3876  51 55 3733  52 56 3631  53 57 3559  54 58 3619  55 59 3305  56 6 1129  57 60 3171  58 61 2845  59 62 2853  60 63 2619  61 64 2468  62 65 2414  63 66 2213  64 67 2044  65 68 1880  66 69 1896  67 7 1094  68 70 1824  69 71 1740  70 72 1565  71 73 1443  72 74 1299  73 75 1321  74 76 1119  75 77 1074  76 78 964  77 79 910  78 8 1096  79 80 837  80 81 774  81 82 613  82 83 595  83 84 542  84 85 467  85 86 402  86 87 370  87 88 267  88 89 220  89 9 1184  90 90 184  91 91 123  92 92 94  93 93 70  94 94 48  95 95 21  96 96 17  97 97 18  98 98 24  99 99 14  100 NN 582  101 UU 17444 |
|  |
| |  | | --- | | > | |

**P\_PSN**

|  |
| --- |
| P\_PSN count  *<chr>* *<int>*  1 11 185435  2 12 2696  3 13 33280  4 21 8889  5 22 2771  6 23 10607  7 31 340  8 32 939  9 33 429  10 96 2990  11 97 51  12 98 201  13 99 10779  14 NN 630  15 QQ 5942  16 UU 6322 |
|  |
| |  | | --- | | > | |

**P\_USER**

|  |
| --- |
| P\_USER count  *<chr>* *<int>*  1 1 173210  2 2 60918  3 3 12685  4 4 4687  5 5 6032  6 U 14769 |
|  |
| |  | | --- | | > | |

Columns to drop

C\_YEAR

C\_SEV

V\_ID

P\_ID

P\_SAFE