Correlation test

Imported dataset

```
ces2021_converted <- ces2021 %>% mutate(across(where(is.labelled), as_factor))
head(ces2021_converted)
```

```
## # A tibble: 6 x 1,059
##   cps21_StartDate     cps21_EndDate       Duration__in_seconds_
##   <dttm>              <dttm>                              <dbl>
## 1 2021-09-19 06:14:46 2021-09-19 06:28:25                   818
## 2 2021-09-15 15:23:33 2021-09-15 15:46:57                  1403
## 3 2021-08-20 09:44:55 2021-08-20 09:57:51                   775
## 4 2021-09-14 18:26:56 2021-09-14 18:40:41                   825
## 5 2021-09-17 17:24:04 2021-09-17 17:51:44                  1660
## 6 2021-09-09 14:13:01 2021-09-09 14:35:14                  1332
## # i 1,056 more variables: RecordedDate <dttm>, cps21_ResponseId <chr>,
## #   DistributionChannel <chr>, UserLanguage <chr>,
## #   cps21_consent_t_First_Click <dbl>, cps21_consent_t_Last_Click <dbl>,
## #   cps21_consent_t_Page_Submit <dbl>, cps21_consent_t_Click_Count <dbl>,
## #   cps21_consent <fct>, cps21_captcha_t_First_Click <dbl>,
## #   cps21_captcha_t_Last_Click <dbl>, cps21_captcha_t_Page_Submit <dbl>,
## #   cps21_captcha_t_Click_Count <dbl>, cps21_citizenship <fct>, ...
```

Feature Selected

Since the original dataset contains over 1000 columns, A subset of variables was selected as candidate features for building a predictive model.

```
mia_vars <- c("pes21_votechoice2021", "pes21_province", "cps21_age", "pes21_follow_pol", "pes21_rural_u
              "pes21_inequal", "pes21_abort2", "pes21_contact1", "Region", "cps21_marital",
              "cps21_imm_year", "cps21_bornin_canada", "cps21_rel_imp", "cps21_volunteer")

extra_vars <- c("Duration__in_seconds_","pes21_lived")

# Merge selected variables
feature_vars <- unique(c(mia_vars, extra_vars))

ces_selected <- ces2021_converted %>% select(all_of(feature_vars))
head(ces_selected)
```

```
## # A tibble: 6 x 16
##   pes21_votechoice2021         pes21_province   cps21_age pes21_follow_pol
##   <fct>                        <fct>                <dbl> <fct>
## 1 Conservative Party           Quebec                  57 Fairly closely
```

```
## 2 ndp                                  British Columbia        22 Not very closely
## 3 Don't know / Prefer not to answer British Columbia        28 Not very closely
## 4 Liberal Party                       Ontario                 29 Not at all
## 5 <NA>                                <NA>                    41 <NA>
## 6 Bloc Québécois                      Quebec                  63 Fairly closely
## # i 12 more variables: pes21_rural_urban <fct>, pes21_inequal <fct>,
## #   pes21_abort2 <fct>, pes21_contact1 <fct>, Region <chr>,
## #   cps21_marital <fct>, cps21_imm_year <fct>, cps21_bornin_canada <fct>,
## #   cps21_rel_imp <fct>, cps21_volunteer <fct>, Duration__in_seconds_ <dbl>,
## #   pes21_lived <fct>
```

Before modeling, we examine the distribution and potential correlations of these features to decide whether they should be included in the model.

Here is the code used to inspect the unique data entries for each selected feature. By reviewing these values, we can identify issues like missing data, inconsistent formatting,or unexpected categories — which indicates that data cleaning is needed before analysis.

```
# return unique entry for each feature
get_feature_levels <- function(data, column_name) {
  if (!column_name %in% names(data)) {
    stop("Column not found in dataset.")
  }
  unique_values <- unique(data[[column_name]])
  return(unique_values)
}
```

```
#get_feature_levels(ces2021_converted, "pes21_follow_pol")

for (var in feature_vars) {
  cat("\n---", var, "---\n")
  print(get_feature_levels(ces2021_converted, var))
}
```

```
##
## --- pes21_votechoice2021 ---
##  [1] Conservative Party                ndp
##  [3] Don't know / Prefer not to answer Liberal Party
##  [5] <NA>                              Bloc Québécois
##  [7] People's Party                    Green Party
##  [9] Another party (specify)           I spoiled my vote
## 9 Levels: Liberal Party Conservative Party ndp Bloc Québécois ... Don't know / Prefer not to answer
##
## --- pes21_province ---
##  [1] Quebec                 British Columbia
##  [3] Ontario                <NA>
##  [5] Alberta                Newfoundland and Labrador
##  [7] Saskatchewan           Manitoba
##  [9] New Brunswick          Yukon
## [11] Nova Scotia            Northwest Territories
## [13] Prince Edward Island   Nunavut
## 13 Levels: Alberta British Columbia Manitoba ... Yukon
##
## --- cps21_age ---
```

```
##   [1] 57 22 28 29 41 63 52 66 42 92 33 48 65 54 68 44 31 38 45 58 64 77 36 62 78
## [26] 72 81 24 46 60 40 59 56 25 49 30 69 53 26 34 43 76 75 80 27 47 35 82 32 73
## [51] 61 18 79 67 70 21 50 37 88 19 39 55 51 74 23 20 85 83 71 90 84 86 89 87 91
## [76] 95 93 96 97
##
## --- pes21_follow_pol ---
## [1] Fairly closely               Not very closely
## [3] Not at all                   <NA>
## [5] Very closely                 Don't know/ Prefer not to answer
## 5 Levels: Very closely Fairly closely Not very closely ... Don't know/ Prefer not to answer
##
## --- pes21_rural_urban ---
## [1] A small town (more than 1000 people but less than 15K)
## [2] A suburb of a large town or city
## [3] A large town or city (more than 50K people)
## [4] <NA>
## [5] A rural area or village (less than1000 people)
## [6] Don't know / Prefer not to answer
## [7] A middle-sized town (15K-50K people) not attached to a city
## 6 Levels: A rural area or village (less than1000 people) ...
##
## --- pes21_inequal ---
## [1] Probably yes                 Definitely yes
## [3] <NA>                         Definitely not
## [5] Probably not                 Not sure
## [7] Don't know/ Prefer not to answer
## 6 Levels: Definitely yes Probably yes Not sure Probably not ... Don't know/ Prefer not to answer
##
## --- pes21_abort2 ---
## [1] No                           <NA>
## [3] Yes                          In some circumstances
## [5] Don't know/ Prefer not to answer
## Levels: Yes In some circumstances No Don't know/ Prefer not to answer
##
## --- pes21_contact1 ---
## [1] No                           <NA>
## [3] Yes                          Don't know/ Prefer not to answer
## Levels: Yes No Don't know/ Prefer not to answer
##
## --- Region ---
## [1] "Quebec"      "West"        "Ontario"     "Atlantic"    "Territories"
##
## --- cps21_marital ---
## [1] Separated                    Never Married
## [3] Married                      Divorced
## [5] Living with a partner        Widowed
## [7] Don't know/ Prefer not to answer
## 7 Levels: Married Living with a partner Divorced Separated ... Don't know/ Prefer not to answer
##
## --- cps21_imm_year ---
##  [1] <NA>                        2001
##  [3] 1965                        1969
##  [5] 2011                        1966
##  [7] 1962                        2003
```

```
##  [9] 1996                                2009
## [11] 1957                                2017
## [13] 2012                                1982
## [15] 1994                                1968
## [17] 1964                                1946
## [19] 2013                                1974
## [21] 1963                                1970
## [23] 1999                                2008
## [25] 1988                                1997
## [27] 2010                                2007
## [29] 2005                                1950
## [31] 1992                                1971
## [33] 1979                                1960
## [35] 2019                                1995
## [37] 2002                                1948
## [39] 1984                                1998
## [41] 1973                                1987
## [43] 1976                                1991
## [45] 1981                                2014
## [47] 1967                                2000
## [49] 2016                                1978
## [51] 1975                                1972
## [53] 1990                                2004
## [55] 1989                                2006
## [57] 2015                                2018
## [59] 1955                                1986
## [61] 1983                                1980
## [63] 1977                                1993
## [65] 1985                                1953
## [67] 1952                                1949
## [69] 1959                                1951
## [71] 1961                                1956
## [73] 1958                                2020
## [75] Don't know/ Prefer not to answer 2021
## [77] 1954                                1947
## [79] 1938                                1945
## [81] 1920                                1934
## [83] 1942                                1931
## 103 Levels: 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 ... Don't know/ Prefer not to
##
## --- cps21_bornin_canada ---
## [1] Yes                              No
## [3] Don't know/ Prefer not to say
## Levels: Yes No Don't know/ Prefer not to say
##
## --- cps21_rel_imp ---
## [1] Not important at all             Somewhat important
## [3] <NA>                             Very important
## [5] Not very important               Don't know/ Prefer not to answer
## 5 Levels: Very important Somewhat important ... Don't know/ Prefer not to answer
##
## --- cps21_volunteer ---
## [1] Never                            A few times
## [3] More than five times             Don't know/ Prefer not to answer
```

4

```
## [5] Just once
## 5 Levels: Never Just once A few times ... Don't know/ Prefer not to answer
##
## --- Duration__in_seconds_ ---
##    [1]     818    1403     775     825    1660    1332    1240    1594     945
##   [10]    4966    1084    1137    1054    1315    1500    1202    3016    2058
##   [19]    1512    1402    1003    1114    1464    2032  487407  262897    1549
##   [28]    1895    1278    6006     827    1146     883    1621    1063     832
##   [37]    1878    2145    1427    2579   13114    1280    1773    1030    1196
##   [46]    1430     721    1345    1258    6775    5551     691     800    1805
##   [55]     987    4412     849     835     933    1534    3422    1057    1426
##   [64]    2238    1442     884    2278     857    1200    1457     991     960
##   [73]    1538    1108    1675     794    1216    1293    1455    4257    3880
##   [82]    1211     830    1837    1518     967     859    1180    6811    2868
##   [91]    2207     863    1025    1600     864     769    1330    1504    1454
##  [100]     838     914    2219    4180    1680    2939    1392    1116     833
##  [109]    1160     690    2280    1667    1232    2424    1870    2226     892
##  [118]     774     604    1233    2011    8259    9533    1244    2422    1068
##  [127]    1126    1106    5988    1201    1263    1489     820    1702    4426
##  [136]  153231     717    1090    2239    1010    2241    2373    2320    1145
##  [145]     927     936    1079    2473    1435    1024    2713    3609    1125
##  [154]    1027     805    1596    5420     625    1302    1840    1100     727
##  [163]    1740    1186    1643    8888     698    1506    1784    1015    1163
##  [172]    1034     557     495    2132    1320     997     813    1299     757
##  [181]    1593  623804    1952     755    3767    1277    1571    1491     890
##  [190]     861    1806    1363    1791    1215     878    1988    2204    3777
##  [199]    1009    1612    1046    1101    1266    1769     837     646    1067
##  [208]    1443     910     836    2510    1123    1220     941    1418    1407
##  [217]    1018    1043    2242     653    2076    2237     971    2187    1210
##  [226]    2764   60988    2050    1154   72178    1075    1486    1092    1185
##  [235]     957     741    4623    1781    3197    1036     384    1283     731
##  [244]    1042    1590     875    1229    1031    1181    1391    1972  450692
##  [253]    1014    1273     929    2155     968    2060    1608    5094    2731
##  [262]    1183    1404    1381     886    1742    1366    2583    1924    2304
##  [271]    1704    1778    1094    2705     982    1155    1115    1401    8374
##  [280]     713     831     871    2584    2206    1329    2042    1337    6785
##  [289]     751    1228     564    2224     888   82982    1149    1714     897
##  [298]    1287  355040    1509    2149    3638    2931    1572    1150   13565
##  [307]     962    1615    1662     766    1252    1522    1019     694    1164
##  [316]    1247     925    1802     689    1820     706    1013    1045    1792
##  [325]    1409    1963    1390     949   10847     829    1497    2209     787
##  [334]    1767     583    1022    1937    1635    1602  422114    1033    2562
##  [343]    1637    1618    1206   99289    4807    1735    2307    1279    1624
##  [352]     728     606     942    1113    1086    1323     560    2190    4617
##  [361]    1603    2258     744    1351   34293     669  185385    1441    1771
##  [370]    1313     714    1026     932     644    1671    1871     976    1705
##  [379]     848    1221    1088    1801    2268    1759    6513    2248    1550
##  [388]    1967    1242    1066    3167    1096    1017    1948    1259    1444
##  [397]    3212    1829     746    2181    1488    1977    1002    2486    1359
##  [406]    2077    1669     708    1659    2325    1632    1478    1262    1120
##  [415]    4218    1541     958     665    1673    1078    1197    6837    1917
##  [424]   79172    1949    2179    1222    1502    1310    1059    1241    1466
##  [433]    2968    2087    1982   16930    1495   14796    1468     928    1004
##  [442]     951    1682    1419    1965  438638    1032    1324  358063    1889
```

```
## [451]    2537    995   1039    812   1073    956   2014    784   1364
## [460]     663   1341    985   2402   2282   3920   1158    926   1065
## [469]     984   1873   6143   5969   5231    667   1052   2123   2220
## [478]    1411    715   1738   1111   1605    544   1393  10331    826
## [487]    1474   1182   5225 618455   4431    869   1319    530   1281
## [496]  381051   1213    763   2546   1388  18691   1668   1334   2176
## [505]     992    915   8929    804    994   3307   1234   1053    845
## [514]     953    660   1000   1521   1118    657   1348   1803    566
## [523]    1119   5726   2928   2624 104366   1817    795   1087   1001
## [532]   10523   1903   1162  83493   2001    841   2619   1611   1844
## [541]    1375   1267    748   3159    622 533094   2762   1567   1777
## [550]    7397   1246   2530    817   1194   2326   1657    645   2981
## [559]    2308   1695  59167    963    712    762   1723   1718   1269
## [568]    1753   1165   1372   1297    916   1333   1205   1479   2161
## [577]   60708   1161   3766    980    989   2166   2323   8196   1892
## [586]    3063   1498    474   1744   1253   1472  13017    814  90884
## [595]    1354    670   4520   1973 603591   1049    906  61824   2993
## [604]    1438    822   1248   1005   2653    738   1508   1727   1422
## [613]     771    700   2637   3065   2150   1151   2732    904   1899
## [622]    1733   2989   1144   1685    631   1312   1195   1765   1307
## [631]    3424   2707   3026   1490   2247   1311   1344   1260   1072
## [640]   51075   1270   1245    816   2018   1085   1231   1896   2807
## [649]     772    559   2385   1289    573   1055   1342   7906 260425
## [658]    2574   1365    659    862   1880   1851    948    408    613
## [667]     924   1141   1552   1811    807   1290   1192   3514   1074
## [676]    1569   9902   4841    990    887    783   1130    902    768
## [685]    1135   1178   3129   2553    535    710   1451   1080   2901
## [694]     677    999   2094   4170   1547   1188    750 271948   3437
## [703]     501   1168   6637  18334   1453   2027   1712    834   1303
## [712]    1400    868   1048   2165   1127   1517   1408   2623   6855
## [721]     596   1627   2172   1638   1109 268851    939   1879   1256
## [730]    2065    684   7219   1604   1741   1535   5509   1860   1217
## [739]    1238   3054    943    658   1021    961   2499   1450 188304
## [748]     776   2594   1774    589   4658   1016    900   1566   1243
## [757]     782   1999   2310    672   1867    865   1226    723    802
## [766]   14415   1420    699   1519   1429  20333   1551   1335   1089
## [775]     581   1469 244206   1630    514   6768   2985   2750   2292
## [784]    1436   2382   1687   5877    729    846    788    496   1715
## [793]    2647   2781   1539    647   5515   1546    789    796   2019
## [802]    1511    930    978    998   4278   1555   1857   2723   7136
## [811]    3360    545    770   2613    632   1460   3656    898    719
## [820]  260611   1826    445   2730   9302   3570    940    697   1579
## [829]    1665   2004   1679   1574   2607   3537   1598   2759   1224
## [838]    1542   2136   1218    603   2966   1437   3624    695   5204
## [847]    1869    921   2708   1095    687   1007    491   1575   1763
## [856]     964   1230   1620   1828    981   2240   1128   2115   1537
## [865]     850   3352   1264   1729   1176    955    923   1107   1306
## [874]    2012   1110   1271   2944   1609    901   1990   1225   2016
## [883]    1790   2121   1288   1098   1686    934    952   3601   9716
## [892]    1462   2550    597   1842   1559   1720   1103   1235   1473
## [901]    2479   1847   2340   1995   1832   1585   3079    851   2610
## [910]    1294    797   1138   1421   1167   1915   3871   4217   1475
## [919]    1902   2937   2471   1800   1369   2355   2300   1316    801
## [928]    2841    730   1214   4625   1818   3472   1394   1835   1700
```

```
## [937]    10263     650     648  302364   15969    1047    1255    1124    1731
## [946]     1484    1412    1129    1734     533     946    1646    5851    1148
## [955]     1379    1023    2003     969    1321     821   56036    3891    1809
## [964]     1179    2816    2540  110522    2803    1198    1986    1117  134072
## [973]     7390     977    1642     975    2930    6358    1076    1357    1501
## [982]     1104     779    2070     761    1573     891    2350     679    1384
## [991]     2948    2267    2097     972    2007    1492    1955    2197    1477
## [1000]    1562     720     701  250347    5016     877    3172    1209    1595
## [1009]     661    1439    1905    2393    1318     636    1981    2039     617
## [1018]    1751     918    7412    4683     760     853    2378    1169     734
## [1027]    2687    2526    3558    5542    1304    5759     809    2618    4091
## [1036]    1029    1236     592     944     437    4674     793    4117    1894
## [1045]     594     554   16881   85374   11267 1226493    1625    5465     683
## [1054]    2180    2765    1251     803    3612    1056    1159    1449    1346
## [1063]    2305    1368    1423    1386    2518    1932  626130    1563    2037
## [1072]    2677     652    1696     947     860     856     664    1543  114450
## [1081]  275316    1152    1577    2063    1285    1863    1175    1028    1868
## [1090]   15875     686    3680    1051    1219    1713    1922   32575     704
## [1099]    1071    4256    2053    1698    2147   84654    1476    2446     874
## [1108]     595    2720    2850    1139    4110    6026    2496    1516    2023
## [1117]    3138    2312    3309    1122    5009    1143  116704    1050     893
## [1126]    1644    2843     681    1153    1980    7453    2374     889    3139
## [1135]  420489    2109    2146     764    1105    2327    2137    3813    1257
## [1144]    1171    3132     580    1377    2029    7946    4738    1780    1398
## [1153]    3169    2675    2318    1373    1524    1964    1020  603145    2976
## [1162]  119888    3046    1174     620    1992    1136     765   66715   84671
## [1171]    1770    1035     819    1292    1732    1483  352570    1347    1852
## [1180]    1823    3184     705    1921    1317    1721   18376    1487     553
## [1189]     911    1752    2074    1588    2979  156910    2069    2275  125869
## [1198]     724    1544    7706    2144    1887    1358    2551    1950    1505
## [1207]    1653    2754    1414    2167    2022    1900    3860   11095     903
## [1216]     880  251062    2071    2995    2083    2617    1841    1918    1343
## [1225]    1592    3355     815    2992     577    1623    1265     828     623
## [1234]    2820    2934    1807   41462     740     703    1652     983    1578
## [1243]    3861     599    1237    1530    1614    1755   82548    1626    1140
## [1252]    1591    1757    6560    2205     688   42462     702     896     931
## [1261]    1295     565    1525    1452    4177    1424    1338     641     799
## [1270]  353113     610     791    1954    2038     907    5444    1406    1284
## [1279]    1434    2853    3841    2650    1395     447    1931    1296     630
## [1288]    1666    6491    5791    1709     895    3768    1748    1274    1061
## [1297]     798     469    1885    1722    1786    1131    1112    1083    1749
## [1306]    2153    2105    1743    1831     780    1861    2567    1848    1097
## [1315]    2982     743     973   38306    3215    6877    1374     711    2104
## [1324]    1301    2552    4456    1622    2173     637    1884    1645    1275
## [1333]     872    2778     696    1041    1371    1874    1447     996     578
## [1342]    2409    1456    2404   14439    2427    1410     536    1699    2164
## [1351]    1650    1470    2458    1998    1204     651    2043     806    2245
## [1360]     894     823    1619    2188     913    1607    1397    1846     486
## [1369]    3539    1300    2112    2980  103566   13803     467     899    1739
## [1378]   16758   71064    8459    1190    2821    5045    4535    3357    2337
## [1387]    1305    1545    1006    4258    3094    1132    1651    1060    2158
## [1396]    1914     917    9693    2020    2461    1223    1471    1583     674
## [1405]    1133    1907     876    2093     649    2718    3312    1926    2391
## [1414]    2170    2425    3879    1694     588    1370    2436    2008    1923
```

```
## [1423]     937    1647     842    2331    2088   17569    1966    1816    4127
## [1432]     920    1440    3043    2440    1328    1810    1661     873   38129
## [1441]    1523  426112    2673   13567  125930    1367    1970    1843    5595
## [1450]    1187    1389    1639    3416    2210  110251    9176    7318     627
## [1459]  122091    1356    3265    2889    1350    2279    2329     777    3822
## [1468]     522    1037    2644    1268    1581     611     747    2469    1589
## [1477]    2192     965    1913    1725    1093    4306    2017    1703    1170
## [1486]     844    2464    2277     511    2695  172448    1382    4048    1189
## [1495]    2052    1554  121162   15588    4528  685869    1610    1062    2384
## [1504]     537     905    5085    1298    5681    1082    2142    1799    2048
## [1513]  272010    4241    1886    1282    2255     600    2912    1385    2081
## [1522]   98681   18665     654    1785    1683     792    2000    2075    2113
## [1531]     843    1728    2201    4526     735     682    1584   18306    2699
## [1540]    2057    1309     935    1250    2491    1177   19896    2243    1812
## [1549]  185335    2120     673    1875    1011   11402    1664    1322    2171
## [1558]    1681    1121     824   28685    3702    8598    3569     678    3479
## [1567]     582     749    2160     922    1510    3637    2040    9332     400
## [1576]    1909    5179   16727    1789   69264    1339    1361    2168    1989
## [1585]    2493    1207    4601    1629     909   12009    4458    2531    1503
## [1594]    1286   27364    1362     628    1212    3085   40718     642    6581
## [1603]    2506     366    1953    1227    1883    6470    1428     575    1893
## [1612]    1761    4083    2974    2692    2134    2717    1939    3241     979
## [1621]    2090     919    2285    8616    1383    9584     737    1962     759
## [1630]    1927    4512     531    7049     718     638    1766    2127    2565
## [1639]    2366    2559    8630   14916     810    1862    2639    4006    1446
## [1648]    1081    1597   10400    1527    2846    1697    1533    1564    1387
## [1657]    1459  340819    2873    4013    1764    2177    1494  521870     509
## [1666]    2784    1413    1142    6250     464     988  293178    1940    1099
## [1675]    5201    3773    1044    1091    1606    2033     839    1969    1985
## [1684]     752    2649     626    6485    1975    3365    1945    2005    1193
## [1693]    1824    1425    3288    4092    2576   10159     453    6572   14425
## [1702]    1991    1737    1881    4513    1943   10030    1458     567    2124
## [1711]     547    8845    2680  500220    9439    1467    4577  101158    2507
## [1720]    8184     882    2749    2449   46047    1513    3334    3644    3239
## [1729]    4883   10124  155486    3141   28399    1480    1070    1654     655
## [1738]    1947   19745     966    1676    1775    1331    2420     778    1134
## [1747]    1724    1496    1677  206129    2480 1396648   14367     524    2672
## [1756]   11555    1485    2861    3699    1796    1891     587    7056    2046
## [1765]  209289    2586    2435    3787    2198    4934    2195    3527    1726
## [1774]   27609    2714    1716    2101    2118  843127     986    2959     785
## [1783]     954     605    6061    3025    5546    3206    2616    2935    4100
## [1792]   82913    1514     790     938    1758     506  241315    4994    1325
## [1801]    2232   77284    1327     881    1959    2024    9239     362    2163
## [1810]    3124     612    1872    1858    1166     619    1445    2287    2703
## [1819]    1730    1532     614    9618    2135    2281    7249     483    1314
## [1828]    2524    1576   18218    3765    1580     870    5860    1961    1855
## [1837]  427789    1553    2899    2394    3120    1788    2061    2191    2478
## [1846]  132668    2362    1191     773    7569    2265     529  519377    2028
## [1855] 1208397   19967    2752    2554    1352    5114    2334    2872    2244
## [1864]    1866   29408     733    1707   64275   11299     950  343835    4121
## [1873]   23552    9242     758  215063    2474    1971     754     867    1203
## [1882]     571    2072    1649  176059  625235    1261    1663    1984    5369
## [1891]    6453    2659   25157    1208    1706    2002    7950   22960    1808
## [1900]     736    1906     484    2489    1901     526    3271    1515    2834
```

8

```
## [1909]    1776    8240     416    4112    2381    1012    1540    1405     609
## [1918]    1710    1038    6351    3243    3363  337639   32964    3285    2719
## [1927]  709893    3576     811    2924   18913    3776    1601     753    2212
## [1936]    2103     707    1481    1102    3198    5901    2829    1561    1157
## [1945]   15723    1184   13596     908    5122    1416    8673    1465    1854
## [1954]    2079    1493   10144    2141    2905    1570    2091    1463  578537
## [1963]     879    1956    1996     852     510   45425    3646    1747    1291
## [1972]    1834    2078    1548    1838    2189   46684   20340    1565   17379
## [1981]     568     786    3140    4318  145994     993     569   76641   93214
## [1990]     513    1674    1173   29483     912    1378    2561   25940    3866
## [1999]    1856    2214    2107    2062   57507    2303    2795    3467    2704
## [2008]    2286    2612    4991   13723    2006    2296    2122    2035    2262
## [2017]    3394    2412    2870    2575    5787     756    1058     739    1557
## [2026]     546    6384    3671    7040    5378    3060     745    4071    2211
## [2035]   11630     540    2297    1920    1040    2257    2497    1711    2045
## [2044]    2521     550    1693    2766    3290    1951   74359     959    5821
## [2053]    1249    1536    2419    1276   13785  193845  160394    1399    2929
## [2062]    2298    1968    2080   30882    4017  426736     398  279352    2400
## [2071]    3634    8650    3287    1736    3613   12280    2371    2230    4237
## [2080]     602    2092     742    7411    1641    2138    2343    4097   11213
## [2089]    1850    1077     656    1461     538    1865    2346     732    1272
## [2098]   50699   71259    1942    2640    4789    5158    9134     847    1877
## [2107]    2174    2401  105062    1415    2148    2116    1448    5498    5410
## [2116]    1782    2208  106587   11289    2439    5775    1746     866    5896
## [2125]   13556  116038    1560    3074   22041    7476    2962    1890   96075
## [2134]    1946    8421    5135    4275    1827    2817     635   14884    2883
## [2143]    1888   89080    2444    3123    2735    5552    2357    2025    2922
## [2152]  605876    2733    2399   15490    6467    4295    4001    1432    3617
## [2161]    1526    3597    1308  201000    2380    1064    8119    2691    2259
## [2170]    1935    2964     558     709    3551    2154   12844    1836    1156
## [2179]    1717    7338  878510    4334   11776    2178     885   21944    1254
## [2188]    4178    2067    1587    3829   79457    4335    3324    1690  247735
## [2197]    2648   31807    1904    1326    4504  354072     668    2119     781
## [2206]    2235     624    7264     570    4581    2395    2470  172109    1353
## [2215]    1616    1582    2202    3921    2693    2358    1688   17341  194922
## [2224]    2855     676  765223    4518   21988    3096    3832    5839    4159
## [2233]    2217    7500    1719    3269    2958     601    1830    1678     680
## [2242]    2363    3080    3907   18315    1772    1814   11111   12097    1568
## [2251]    3027    2010    4656    2487  613307    1528    1631    2560     556
## [2260]    3678    2504     643    1691    2502    2110    5232    2064    2106
## [2269]    1929    6120     516    2095   70384   26599    4273    2998    2047
## [2278]  261877   13289    2250    3946    3543  815986    4123    5216    3274
## [2287]    2284  505670    3802    2709    1380    2632   87620    4239    4176
## [2296]   32593    7753    3535     541   25144     974   14008     435    3304
## [2305]     858    7317     633    1431    2082    2794    2670    2605    2664
## [2314]    3103     479   92510    8778    4424    2706  243320    5155    6027
## [2323]    3227    3482    1762    2349   87838    2620    1520   80402    1839
## [2332]   47721    6897    2596    1376   18455    1684   19947    1433    1960
## [2341]    1845   74391    2919    3598    3337    1882    2372   17311    4551
## [2350]    2854     542    1239    6076    1634    1987  151709     767    1336
## [2359]    2949    2342    3556    3936     586    3248   42431    2254    1640
## [2368]    2428    3421    1994  103290    4584    2743    1976    2221   84737
## [2377]    2811  241547    2266     722    1656    2182 1276346    4833    4647
## [2386]    2408    2514  520260    3724    2950    3364    2222    2223    2495
```

```
## [2395]    1360   465736   176445     2418     6363   428086     3249     1978     2415
## [2404]  225721     2313     1529     3192    10681   118108   227798     1147     2558
## [2413]   37647     1876     9095     3896     3566      457     1658    85314     2585
## [2422]    3500     2880     2508     5716   174060     2044      532   167375   193102
## [2431]    7492     1815   782757     2354     4817     1199   100509     2085     1708
## [2440]    1008      671     3011     7174      662     2034      970   453253    86777
## [2449]    3987     2288     2633     4016     4274   161190     1349     2978     3277
## [2458]    1655    15164     3775     3982     1933      840     2984     6172     3906
## [2467]   30005    29675   549393     3604     2193      639     1974     5652     1172
## [2476]    5553     2451     4152    42023      855     4214     2611     2523     2055
## [2485]    3116    10978     2443     2667     2264     5653      854     2651     6718
## [2494]    5349     2798    11273      725     3812     3039     3066     4150    12418
## [2503]    2751     3952     2203     3910     2662     3160     2143    14093      593
## [2512]    4887     6127   371981     1586      492     5278     3610     1938    18703
## [2521]    2185     5827   793699     3297     7482     1499      525     2272   522246
## [2530]    7045     2453     1958     4082     1768     5893     4736    32274     9081
## [2539]    8533     7552      373     1916     7229      476     2068     2956    11782
## [2548]  121797     2216      616     4376     1482   142763   183971    14258      466
## [2557]    1340     2702     2903     3356     1795     9836     3717    10141     3443
## [2566]   14479     2054     1760     1531     2184     1750    87453   415687     8857
## [2575]    2525     2513   153158     3675    33684      579     2839     6463    15784
## [2584]    5355   237257    13647   297165     1355    21622     2406     1833    20531
## [2593]   46139     1701     2631    29890     4410      555     2768    10418     4379
## [2602]    2621     2344    11222     2293     2108     2536    66086   758490    10539
## [2611]     528     2421     2228    29590     5092     2515    14685      383     2535
## [2620]    2270     6990     9799     3326    19126     2595      552   670248     8174
## [2629]    2396     2289      518     1754     1396     5477   845075    10168   245264
## [2638]    3055     2857     1745     3228     3893     2973     3361     2674     6989
## [2647]    1672     3393     4805     3379     2465     2852     3587     3590    26451
## [2656]    2098      666     4157     3450     2538     3608     4539     1558     2943
## [2665]    3191     1689  1395835     3253     5642   141680     4296     3384     2338
## [2674]    8302     1822     1934     2597     2893     2319      726      692     8585
## [2683]    3378   190806    79799     3441     2634     2891     2301      549     3168
## [2692]    2139      478     2505   264379    19390     2089    15441     6890      365
## [2701]    6230    38223     3968     5731     3469     2455     2643     2918     2842
## [2710]    3823   345313     3752     8110   236488     6144    11910    28095   411859
## [2719]    2111     2194     2447      716     5866    28810      487    57968     6529
## [2728]    4699   186683     4191    72687     4058     3070     2599    13463     6888
## [2737]    3836     8881      615      563      585     1957      505     2086     1911
## [2746]    2999     3842   184572     3020    14134    77898   329101    16729    36166
## [2755]    2251   116401      685    25633   173919    10301    50940    19182     3205
## [2764]    1628     4616     4929    79918     2858      543     5481   937459     2472
## [2773]  159884     9307     3261     6562     4554     1859     5650      521     5695
## [2782]    5547     2370     2324     3603      808     2333      502     1925     6211
## [2791]     621   255136    39617   112259     2263     2592     1507     8402   277374
## [2800]   67245     4221     4606     2801     3348     2271    33540     3518     2140
## [2809]   57380     3452     4213     3022     2761    11556     2578     6618    51208
## [2818]    3991     2152     1613   149212     3193      608   503599     7823    14075
## [2827]    4428     2757    88619     2628    20511     4210     2879     7402   169183
## [2836]    1979    14201     1633     3218     3652    94243     2671     1069     2383
## [2845]    7374     4107     6397     1804     3262      519     4940    12363    76083
## [2854]    1897     2233   110897     3069     4413   125650   996186     4167    80049
## [2863]  186565     4591     3427      500     1819     2896     5863     6721      675
## [2872]   46005     2555     6973   165596     3064     4118     3526    12461     2739
```

```
## [2881]    2261    4304     497    7224    1813    2481   10377    2430    2947
## [2890]    3632    2682    2793    7661    5980  605293     499    4900    8022
## [2899]    2157   23223    1756    7748    2477    2353    1556    2390    2990
## [2908]    7438  163861   47216   18875  626772    5699   11130    2490   20719
## [2917]   10075    4889    7304   15025   14839    2253    2710    5320   46460
## [2926]    2249    7046    3918   13614  953863    3583   10310    2544     419
## [2935]    2299    2936  505149    1779    3513   54211    6670    4803    9654
## [2944]     618    2131     640    2021    2126    4495  269245   88209   10672
## [2953]    9042    4685    1692    2030  252007    2336   19054   51871    4752
## [2962]    8927    8571    3629    4868    3532    2543    2129    3259   16295
## [2971]    4473    5241    2379  134173    3633   10456    2438    6361    3594
## [2980]    5148   20263   13741     406   19742    3317    2862    3463    9317
## [2989]    2921    2260    4629    4531    2328    1898   33003   11970   55891
## [2998]    3035   20756    3171    2392   26117   64461    2311    2849    2125
## [3007]    2051    2800    2398    1417  411055   24098    6283    2423     598
## [3016]   80762    2236  378256   78486    3042    3781    2448    4247    6612
## [3025]    4843   24158     629    1928    3018    3396    4276   12463  203979
## [3034]    2734  171761    2656    2866    2952   10125   91662    3214  104156
## [3043]    2890  160071    3044    2975    2330    2727   22932  102967    8409
## [3052]    2200    1944    8325    4810     562    4635  478065    2361    4828
## [3061]    2360    1617    2983  164803   18739    3010   18393    2367  121738
## [3070]    2445    4235    3400    7925    3325    7131   10873    3100    2698
## [3079]   77844    7339   14199   45166    3977    5590    5875    2317    1648
## [3088]    2456   11181    6060     504   22191   21654    4440    2654    2833
## [3097]    2796  532107    2532   11848    3272    2696   10549    6436  876490
## [3106]    5227    5884    2721    3076  196020    5409    4893    2697    3444
## [3115]    3208    3824    1794   75523   17715   92083     584    4468    3732
## [3124]    2227     480    2397    6517    2056     590    2923    5503   46712
## [3133]    5331    3505    1849    2283   80554    6067    2970    2926  270863
## [3142]  167290  536211    3105     512    7640    3588    4640    5660    6011
## [3151]    3978    2690    4136   11599  990859    2347   67624  410845    2669
## [3160]    2454    7087   48627  428925    1599    2881    4299    2933    3996
## [3169]    3294  168403  264307    2741   19013   54972    2162   12399    3808
## [3178]     460    2589 1450429     561    2099    2748  244549    5073    2756
## [3187]    4812     607    1798    5550    3095    7784    2341    3772  143085
## [3196]    5555    3001    2429     421    2503    2810    3577    6160    1941
## [3205]    3466    9706  505250    4621   76772  174600  201291    3779  717889
## [3214]    4021   73397    4471    3716   47568    2133    3517    2791   12448
## [3223]    2468    2405    8959     468   22793   36421     370    3886    2196
## [3232]   14749   14656    8344    2815    5776   80291   11938  129970    1864
## [3241]    3091   16917   18117    3559   77530   17510    2442    5577    3884
## [3250]    2780   10025     572    4098    2799    1983    6971    3425   88134
## [3259]   16439    6739    2533    4815    2556    5362    8979   21802    3133
## [3268]    3210    2877   27499    2865    6523   14603    3005    4876  551885
## [3277]    3957    2015    2660     693  174703   24959    7002    2114  200358
## [3286]  272823    3654    2365  170866     634   93378   56022  114755     436
## [3295]    3714   69102   75717    1908    3853    4461    2606    2906    4285
## [3304]   76622    2302    2482   21584    2960    2600   40093    2117    7421
## [3313]  194711  346602    2622    1797   14195    3816    3037    3881    4796
## [3322]    3726   77648    2683  169494   11020    2049    2269    3087     488
## [3331]   25192    3075    3820     517  702049   47334   59459    3460  187193
## [3340]    2084    7939    2322   11145    2096   14124    4234    3195    5134
## [3349]    2688  179844  256714 1294172   10513    2356    8296    4710    4670
## [3358]    2294    2351    2917    3322    3072    7962    3516   16975   10788
```

```
## [3367]    6982   24156   67598  452698    1783    3528    3771   24008    2376
## [3376]    5976   13863    4638    3711    5483    5857   24332    4270   18165
## [3385]   63970   20365    3762     430   83465  974478    3015    3730    2884
## [3394]   11610     471    3305    3062    3014    2573    3757    2527     534
## [3403]  788019    4886   10813   13671     458   20751    1912   85309    9559
## [3412]   13135    3182    6378    4821    7017  406916    2608    4673    2927
## [3421]    5878    3506    2845    6343    3125    4878    6197    3146    3170
## [3430]    4064    7495    1930    5318    6829    3591   12428   34829    4668
## [3439]    2577   10755    3428    3938     446    4983    3464   16982    9882
## [3448]  274806   55556    8789     397  260830    2339   44569    3676    4317
## [3457]    9728   61770  117934    4279   32193    7698    5445   41489     498
## [3466]    2630    3389    3002    4675    4104    3061    2128    2593    1636
## [3475]    8838  515167  191564    2932    4452    5407   45309    4179    1993
## [3484]    3534    2520    2215   77113    6794   12059   14831   10340   71454
## [3493]  689781    5314  166931   15701    3254    2073    4022   85049    2967
## [3502]    2066   20873    3489     433    2309    3409    2199   81963   80789
## [3511] 1031201  365994    2548    3000    8419     448    1670   53947    2059
## [3520]    2860  355177  163767    1787     472   98107    3176    3084   13431
## [3529] 1011352    2036    3828   16285    5918    8130    3118    3988    4423
## [3538]  169870   10529    7568    8773   11667    9565  556655    5003    2364
## [3547]   14909    9630   35580    2566    3640    3173  147122    1825   12235
## [3556]    3057    2802     481   16829    2386    5011    7543    3417  262612
## [3565]    9624    3999    2387   97385     442    5334    2186     515     574
## [3574]   80918     426    5064    5017    2913   10279    2925  156310    2314
## [3583]   10840    8750    6460  146485    4692   69083   12138    2588    3661
## [3592]    2410  138922   13623  114701    6388   13589   11895   22669  130369
## [3601]    2638    3754  559072   20448    7636   10474    3806   89613    2229
## [3610]    2666   53059    8840    3047   19777    5115  802166    3308   16817
## [3619]  266512    2156    3251    4813    3501   13869    4448    4012    7651
## [3628]  416686   42298  535364   28222    3451   24049     395    4145   88448
## [3637]   27940    5375  168619   29684  406673    7933    2876    8388  152732
## [3646]    2500   85271   23925    5816   20711    3298    2100    3392    1793
## [3655]    4698   66098   29562    5151    4543     485    3040    7525    4069
## [3664]    5983    2498    3620  176533   21226   69221    3737  173509    2646
## [3673]   16262  162433    4003    2626    3093    3221    4392    3268    2627
## [3682]   64329    4375     551    2744    3342    4026   68283    4310    2273
## [3691]    2590    3117    5226    9532   36226    3115   14407  372561    2570
## [3700]    6084    4253    5382    5235    3353    2359    2403    6187  762478
## [3709]   14408    2591    3689   11611   10433   11570    3008    4422    2867
## [3718]    4522     372  299721    9733     523    3503   12836  108015  683237
## [3727]    9837    2041    2256    3552    3369   10639    4460   21469   73587
## [3736]     527  253721    5636   16425  184415    9933    4269   14341   20858
## [3745]   17268    2994    7754    3703    4005   32383    3887    3490    6435
## [3754]   11850    4174    2450   12847    1910  120576    4861    6978  272305
## [3763]    2306    4166    2779  101384    9903   29011   17195    4989    3257
## [3772]    2785  572764    2657    7456     539   70144   44521    1997    3593
## [3781]     422    2986 1046904    4610    2102    5192    4355  241171    3581
## [3790]    3295   26991   11858  233126   13472    6686    3545    2969    2725
## [3799]    4664    3592    3468    3693   15418    3815    7600    4593  576440
## [3808]     591   10306    3137  354204    2601    7546    2411    5182    6404
## [3817]   17699    4142    2940    4840    3522   25027   14011    3486   41090
## [3826]    2462   10164    7563    2175   34794    2511    7271   74885    4920
## [3835]    2747    4595     459    2782    8407   11437    5692  268664  260443
## [3844]    7160    8495 1575155   18473    4057    2459    8322    7540    4328
```

```
## [3853]   56788     2641   179704    16318    79715    33335   798729     2542   182817
## [3862]    5572     4130     8060    15803    87639   444219   424690     5502   910179
## [3871]  127722     2942     5329     6111   171293    11560     3916     8128     2885
## [3880]  174501   136848     4418   297634   108217     2466     2822    15515     3423
## [3889]  161205   165187     2159     4437     2434     9619  1455759     3319     3036
## [3898]     396   277449     3650      369    39643   103666     3021     1821   271536
## [3907]    1936    13985     3341   270497     3705     5075      477    17120    13697
## [3916]   26254   718087    10494    10782   249492     2679     9994     3229     2745
## [3925]    4592     5684   248989   247434    10597    25040     3256   180738     2729
## [3934]    6899     3927   205593    90355     2835    26419     2183     3962     3586
## [3943]   41430     3669   516678    14769    47367     8132    15988   204688   255034
## [3952]    2352   325972     4141     9865     6091    41864     6882     3235    99578
## [3961]   10318     6687     2996     5972   180223     6175     3034     3540     3366
## [3970]    2676    10004     5669     2291    14454     4147     2316     5865     6380
## [3979]  602459     5525      420     5793    11882    18796      452     6432     3662
## [3988]    8034   197652    28974    34711    12629    39893     4951   519011     8294
## [3997]    3411    85250     3135     4476     5224     2026     4351     9779     3708
## [4006]    6328   425990   165286      410    11029     2541   189863     4944     2851
## [4015]    2345     5373     3098     8160     2909     2681     3979     5036    71977
## [4024]   10586    14398     5357    38970    71759     2572     4350     3701     2348
## [4033]    2517    82636     3653    17867    84186     3289    80015    10346    11477
## [4042]    2485     3998     7221    93962      508    11559     4615     3349     4175
## [4051]  454767     6891    27305     3258     2824     8835     2528     5021     7585
## [4060]    2700     3376     2972     6728     2629     3331     2965     3719    85870
## [4069]    2888     8843    35056     5634   509429     4042     3744     2740     3136
## [4078]  384530     3834     7721     4445     9664     6411   191060    10014   431632
## [4087]    3626  1071775     3234     4434      386   480628    11117    35900      443
## [4096]   10880     3795    10725    40751   445016     8827   248744    12854     3928
## [4105]    5538     8610     6122     6245     4389     6309    90865     4041     3301
## [4114]    7391    16095    76856     4053    98901    12243     5426     3023     5656
## [4123]   21199     3092     3279     2997     2218     5057     2790   118781    37873
## [4132]    3024    21885   102475     3475     8707     3330     3350    18546     7928
## [4141]    2441    59806     3006     2234     7715     3831    11354     6313    28059
## [4150]  166791     4941     2295     4139    12011     5880    49024     3067     6119
## [4159]    4148     3161     3399    53837    13038     2642   346897     3827     9898
## [4168]    2494    72147     4155     8361     1919   123957    10443    97097    11688
## [4177]    7396   571362     4349     5871   109331     3264   165489     2712     6354
## [4186]    6049     7444    13548     2213    14450   871701   527260    28845   169440
## [4195]   12472     4000    26657     2529     7111     9263     3338     6659    34234
## [4204]    5654    84473     7470   169590     4427    85581     2413     7212   101072
## [4213]    2437    77823     3149     5018     4369    15696   415311     3630     3202
## [4222]    5455     3077     3833     3439   512663    24523    80363  1209959     7541
## [4231]  189657     3247    11792     3370     3984     7059    11892     4835     3224
## [4240]    9004    14599     3107     3687     3238     4733     2954     6504   266911
## [4249]  188129    16149   296263   160463    34949     4019   108692     5157     9523
## [4258]    4342     3209     6357   181485     2911    84236   169975    17884     2788
## [4267]   26116     4203   630100     4998     3226     5452     4183    13278     4277
## [4276]   13693     5014    16347   276538   123501   522313     9591     2315     6452
## [4285]    5132     3942     5646    93673   104208     9944     2767    41934    35514
## [4294]   11472     3641      417     3190     3255    22656     5688    16084     5647
## [4303]    4661     4686     9983    79375    87790     2321   340068   521096     4255
## [4312]    9443  1285866    11961     3858     2655     3003    23699     4404     2804
## [4321]    8757     2792     9145     6336     6556     7664    22039     3788     5765
## [4330]    2452     2665     5015     3408     4094     8811    11217     4453     6095
```

```
## [4339]    40901     6222     3252     3792   312133     7406     2661    11999    26395
## [4348]    11589   344385     2864     5790    17361    68806     5013     4609     4291
## [4357]   171250    24933     4579    22028     7366   310807     3201    33756   172514
## [4366]     2658    10193    10645     3745    74397     7024     4957   183329     5823
## [4375]   140110   425604     4488      489     4959     8956     2904     3013     3030
## [4384]     9234  1121638     3332  1073367     4946   416745     2894   769461  1308713
## [4393]     4844    16174     2569     2869   292460     9855     5033     2806     4354
## [4402]     4467    68573   102528    50698     2938     3111     3477     9718     4146
## [4411]    12789     3733    33188   424855     2635    48758     5607   500010      402
## [4420]     6505     4066     5414     2169     5194   146201    70711   660054     4746
## [4429]    58870     6428   342109    10297    81228     3397     2961    91946     4628
## [4438]   801796    16242     8472     4529     3761     2830     4120     2878      520
## [4447]    81200     6963     4236    65861     2694     7429
##
## --- pes21_lived ---
## [1] More than 10 years                Less than 1 year
## [3] <NA>                              3-10 years
## [5] 1-3 years                         Don't know/ Prefer not to answer
## 5 Levels: Less than 1 year 1-3 years 3-10 years ... Don't know/ Prefer not to answer
```

Since these features are of different types and most of them are non-numeric. We cannot apply a single, unified statistical method. Instead, we need to adopt different analysis strategies based on the nature of each variable.

Pre-step for Correlation test

We are beginning with data cleaning as the first step, focusing on handling ambiguous responses such as NA and "don't know". In parallel, we aim to identify patterns of political apathy, which may be reflected through missing values, neutral responses, or lack of engagement.

```r
# Replace all "Don't know / Prefer not to answer" with NA (applies to character or factor columns)
replace_dontknow_with_na <- function(col) {
  if (is.character(col) || is.factor(col)) {
    col <- as.character(col)
    col[col %in% c("Don't know / Prefer not to answer", "Don't know / Prefer not to answer")] <- NA
    return(as.factor(col))
  } else {
    return(col)
  }
}

# Apply the function to all columns
ces_cleaned <- ces_selected %>% mutate(across(everything(), replace_dontknow_with_na))
head(ces_cleaned)
```

```
## # A tibble: 6 x 16
##   pes21_votechoice2021 pes21_province   cps21_age pes21_follow_pol
##   <fct>                <fct>                <dbl> <fct>
## 1 Conservative Party   Quebec                  57 Fairly closely
## 2 ndp                  British Columbia        22 Not very closely
## 3 <NA>                 British Columbia        28 Not very closely
## 4 Liberal Party        Ontario                 29 Not at all
## 5 <NA>                 <NA>                    41 <NA>
```

```
6 Bloc Québécois   Quebec   63 Fairly closely
i 12 more variables: pes21_rural_urban <fct>, pes21_inequal <fct>,
pes21_abort2 <fct>, pes21_contact1 <fct>, Region <fct>,
cps21_marital <fct>, cps21_imm_year <fct>, cps21_bornin_canada <fct>,
cps21_rel_imp <fct>, cps21_volunteer <fct>, Duration__in_seconds_ <dbl>,
pes21_lived <fct>
```

Based on survey duration time, we identified some responses as unreliable. These cases are also labeled as politically disengaged. Since they may bias the model, we temporarily remove them from the dataset before modeling.

```
library(ggplot2)
ggplot(ces2021, aes(x = Duration__in_seconds_)) +
  geom_histogram(binwidth = 10) +
  xlim(0, 2400) +
  labs(title = "Distribution of Survey Duration",
       x = "Duration (seconds)", y = "Count")
```

```
## Warning: Removed 3210 rows containing non-finite outside the scale range
## ('stat_bin()').
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range
## ('geom_bar()').
```


Distribution of Survey Duration
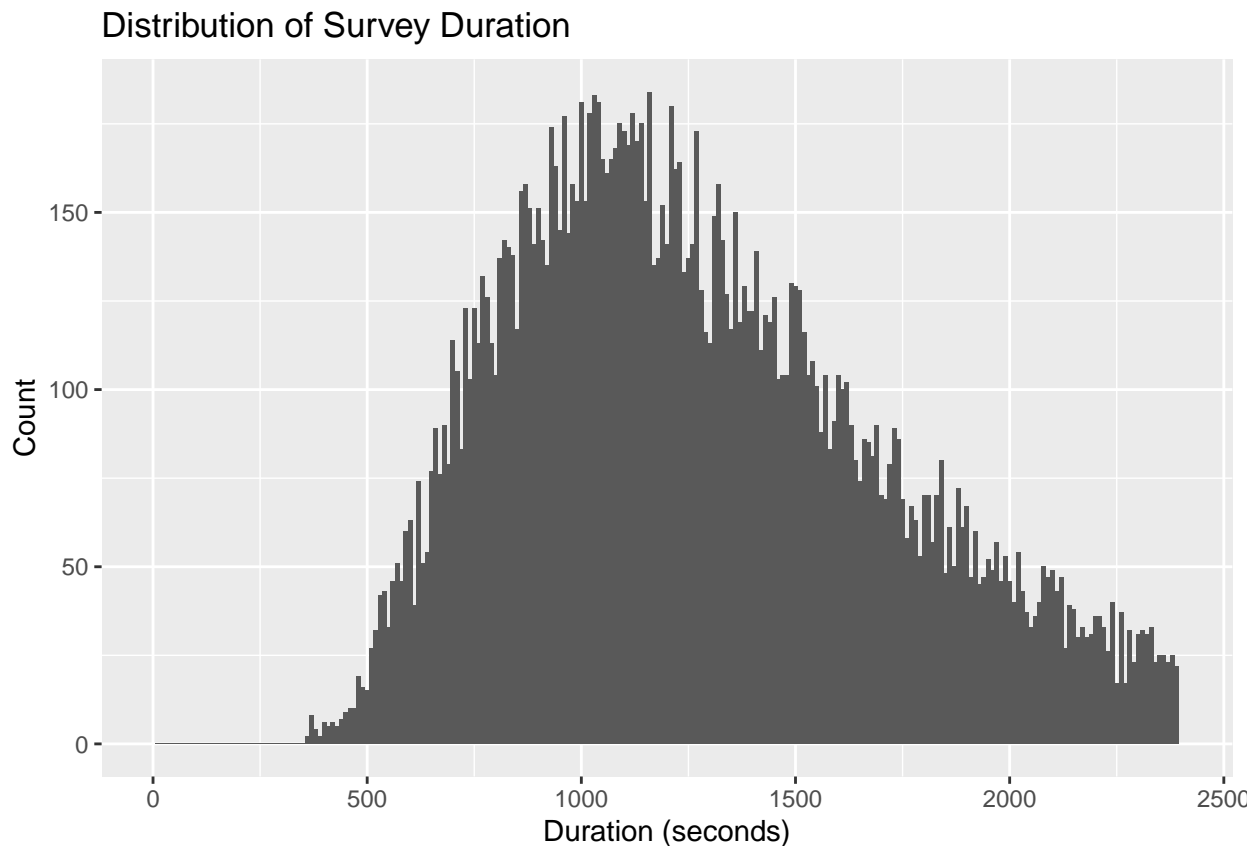
```
summary(ces2021$Duration__in_seconds_)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.     Max.
##     362     995    1325    8710    1875 1575155
```

The summary statistics of the `Duration` variable are as follows:

- **Minimum**: 362 seconds (~6 minutes)

- **1st Quartile (Q1)**: 995 seconds (~16.6 minutes)

- **Median**: 1325 seconds (~22 minutes)

- **Mean**: 8710 seconds (significantly inflated by outliers)

- **3rd Quartile (Q3)**: 1875 seconds (~31.3 minutes)

- **Maximum**: 1,575,155 seconds (> 400 hours)

According to the summary, These values suggest that while most respondents completed the survey in under 30 minutes, there are a few extreme outliers with excessively long durations that strongly distort the mean.

The **minimum value of 362 seconds** and the **Q1 value of 995 seconds** suggest that any respondent completing the survey in under 10 minutes may not have engaged meaningfully with the content. Similarly, values above 48 hour are highly suspicious and may indicate participants who were inactive for long periods.

Therefore, a threshold of **600 seconds (10 minutes)** was chosen to identify "too fast" respondents, while an upper cap of **172800 seconds (48 hour)** was applied to identify "too slow" responses.

```
classify_engagement <- function(data, duration_col = "Duration__in_seconds_",
                                 fast_threshold = 600, slow_threshold = 172800) {
  data <- data %>%
    mutate(
      engagement_group = case_when(
        .data[[duration_col]] < fast_threshold ~ "too_fast",
        .data[[duration_col]] > slow_threshold ~ "too_slow",
        TRUE ~ "normal"
      )
    )

  low_engagement_data <- data %>% filter(engagement_group != "normal")
  main_cleaned_data <- data %>% filter(engagement_group == "normal")

  return(list(
    full_with_labels = data,
    low_engagement = low_engagement_data,
    main_clean = main_cleaned_data
  ))
}
```

```
result <- classify_engagement(ces_cleaned)

ces_selected_engag <- result$full_with_labels
```

```r
low_engagement_data <- result$low_engagement # for low engagement analysis
main_cleaned_data <- result$main_clean # for correlation test
```

Next, we divide the selected features into two groups based on their data types: - Categorical features will be tested using Cramér's V - Ordinal or continuous features will be tested using the Kruskal-Wallis test This helps us evaluate the strength of correlation between each feature and the target variable.

```r
# Initialize lists for variable classification
list_chi <- c()
list_kruskal <- c()

# Target variable (e.g. party vote choice)
target_var <- "pes21_votechoice2021"
target <- main_cleaned_data[[target_var]]
is_target_cat <- is.factor(target) || is.character(target)

# Loop through feature variables
if (is_target_cat) {
  for (var in feature_vars) {
    x <- main_cleaned_data[[var]]

    if (is.factor(x) || is.character(x)) {
      list_chi <- c(list_chi, var)
    } else if (is.numeric(x) || is.ordered(x)) {
      list_kruskal <- c(list_kruskal, var)
    }
  }
}

# Print results
cat("Variables for Cramér's V (categorical):\n")
```

Variables for Cramér's V (categorical):

```r
print(list_chi)
```

```
##  [1] "pes21_votechoice2021" "pes21_province"       "pes21_follow_pol"
##  [4] "pes21_rural_urban"    "pes21_inequal"        "pes21_abort2"
##  [7] "pes21_contact1"       "Region"               "cps21_marital"
## [10] "cps21_imm_year"       "cps21_bornin_canada"  "cps21_rel_imp"
## [13] "cps21_volunteer"      "pes21_lived"
```

```r
cat("\nVariables for Kruskal-Wallis (numeric or ordered):\n")
```

```
##
## Variables for Kruskal-Wallis (numeric or ordered):
```

```r
print(list_kruskal)
```

```
## [1] "cps21_age"              "Duration__in_seconds_"
```

Correlation Test

For the features in list_chi, we compute their correlation with the target variable (party vote choice) using Cramér's V. The calculation uses the cramerV() function from the rcompanion package, which automatically removes observations with missing values (NA).

```r
library(rcompanion)  # cramerV
cramer_results <- data.frame(Variable = character(),
                             CramersV = numeric(),
                             stringsAsFactors = FALSE)

for (var in list_chi) {
  tbl <- table(main_cleaned_data[[target_var]], main_cleaned_data[[var]])
  if (min(dim(tbl)) > 1) {
    result <- cramerV(tbl, bias.correct = TRUE)
    cramer_results <- rbind(cramer_results, data.frame(Variable = var, CramersV = result))
  }
}

# print result
cramer_results <- cramer_results[order(-cramer_results$CramersV), ]
print(cramer_results, row.names = FALSE)
```

```
##               Variable CramersV
##   pes21_votechoice2021  1.00000
##                 Region  0.28970
##          pes21_province  0.22470
##           pes21_inequal  0.17570
##             pes21_abort2  0.16080
##            cps21_rel_imp  0.11440
##     cps21_bornin_canada  0.10060
##            cps21_marital  0.09504
##       pes21_rural_urban  0.07665
##         pes21_follow_pol  0.05808
##              pes21_lived  0.04878
##            pes21_contact1  0.03479
##          cps21_volunteer  0.02414
##            cps21_imm_year     NaN
```

Higher Cramér's V values indicate stronger associations with the target variable. Variables such as Region and Province showed relatively strong correlations with vote choice, while others like Volunteer activity and Immigration year had weaker or missing correlations.

During the Cramér's V analysis, we found that 'cps21_imm_year' returned NaN. It might because the 'cps21_imm_year' variable has many unique values (immigration years). To address this, we converted 'cps21_imm_year' into 'years since immigration' by subtracting it from 2021. This transformed variable is numeric and can be meaningfully analyzed using the Kruskal–Wallis test.

```r
# convert new variable
main_cleaned_data$imm_duration <- 2021 - as.numeric(main_cleaned_data$cps21_imm_year)

# add to list_kruskal
list_kruskal <- c(list_kruskal, "imm_duration")
```

Then we applied the Kruskal–Wallis test to evaluate whether the distributions of features in list_kruskal differ significantly across vote choice categories.

```
kruskal_results <- data.frame(Variable = character(),
                              KruskalP = numeric(),
                              stringsAsFactors = FALSE)


for (var in list_kruskal) {
  df <- na.omit(main_cleaned_data[, c(var, target_var)])
  formula <- as.formula(paste(var, "~", target_var))
  result <- kruskal.test(formula, data = df)

  kruskal_results <- rbind(kruskal_results,
                           data.frame(Variable = var,
                                      KruskalP = result$p.value))
}

# print result
kruskal_results <- kruskal_results[order(kruskal_results$KruskalP), ]
print(kruskal_results)
```

```
##                 Variable      KruskalP
## 1              cps21_age 5.913515e-119
## 2 Duration__in_seconds_  2.867399e-34
## 3           imm_duration  3.078417e-07
```

Since small p-values indicate strong evidence of differences between groups. The features 'cps21_age', 'Duration_*in_seconds*', and 'imm_duration' all represented that these features are highly associated with voting behavior and may be valuable for predictive modeling.

Result

We will use the following features in the prediction model:

```
# filter variables with Cramér's V > 0.1
selected_cramer_vars <- cramer_results %>%
  filter(CramersV > 0.1) %>%
  pull(Variable)

# filter variables with kruskal < 0.05
selected_kruskal_vars <- kruskal_results %>%
  filter(KruskalP < 0.05) %>%
  pull(Variable)

selected_model_vars <- unique(c(selected_cramer_vars, selected_kruskal_vars))
print(selected_model_vars)
```

```
##  [1] "pes21_votechoice2021"  "Region"                "pes21_province"
##  [4] "pes21_inequal"         "pes21_abort2"          "cps21_rel_imp"
##  [7] "cps21_bornin_canada"   "cps21_age"             "Duration__in_seconds_"
## [10] "imm_duration"
```