Name : Kanyana Samantha

Exploring the Dataset

Exploring the Dataset

```
> head(Data)
  Age Gender Total_Bilirubin Direct_Bilirubin Alkaline_Phosphotase ALamino_Aminotransferase Aspartate_Aminotransferase Total_Protein
1  65 Female             0.7             0.1               187             16             18             6.8
2  62   Male             10.9             5.5               699             64            100             7.5
3  62   Male             7.3             4.1               490             60             68             7.0
4  58   Male             1.0             0.3               195             14             20             6.8
5  72   Male             3.9             2.0               268             50             112            5.1
6  62 Female             1.3             0.3               303             18             12             7.3
  Albumin Albumin_and_Globulin_Ratio Selector
1     3.3                       0.9        1
2     3.2                       0.7        1
3     3.3                       0.9        1
4     3.4                       1.0        1
5     2.7                       0.5        2
6     3.3                       0.9        1
> tail(Data)
    Age Gender Total_Bilirubin Direct_Bilirubin Alkaline_Phosphotase ALamino_Aminotransferase Aspartate_Aminotransferase Total_Protein
579  40   Male             0.6             0.1               098             35             31             6.0
580  52   Male             0.8             0.2               245             48             49             6.4
581  31   Male             1.3             0.5               184             29             32             6.8
582  35   Male             1.0             0.3               318             79             29             6.4
583  39   Male             0.9             0.3               444             59             73             7.3
584  62 Female             0.7             0.2               463             62             80             5.9
    Albumin Albumin_and_Globulin_Ratio Selector
579     3.1                       0.9        1
580     3.2                       0.8        1
581     3.4                       0.9        1
582     3.2                       1.0        1
583     3.3                       0.9        1
584     3.0                       0.8        2
```

Checking the Variables

```
> colnames(Data)
 [1] "Age"                       "Gender"                    "Total_Bilirubin"           "Direct_Bilirubin"
 [5] "Alkaline_Phosphotase"      "ALamino_Aminotransferase"  "Aspartate_Aminotransferase" "Total_Protein"
 [9] "Albumin"                   "Albumin_and_Globulin_Ratio" "Selector"
> rownames(Data)
  [1] "1"   "2"   "3"   "4"   "5"   "6"   "7"   "8"   "9"   "10"  "11"  "12"  "13"  "14"  "15"  "16"  "17"  "18"  "19"  "20"
 [21] "21"  "22"  "23"  "24"  "25"  "26"  "27"  "28"  "29"  "30"  "31"  "32"  "33"  "34"  "35"  "36"  "37"  "38"  "39"  "40"
 [41] "41"  "42"  "43"  "44"  "45"  "46"  "47"  "48"  "49"  "50"  "51"  "52"  "53"  "54"  "55"  "56"  "57"  "58"  "59"  "60"
 [61] "61"  "62"  "63"  "64"  "65"  "66"  "67"  "68"  "69"  "70"  "71"  "72"  "73"  "74"  "75"  "76"  "77"  "78"  "79"  "80"
 [81] "81"  "82"  "83"  "84"  "85"  "86"  "87"  "88"  "89"  "90"  "91"  "92"  "93"  "94"  "95"  "96"  "97"  "98"  "99"  "100"
 [101] "101" "102" "103" "104" "105" "106" "107" "108" "109" "110" "111" "112" "113" "114" "115" "116" "117" "118" "119" "120"
 [121] "121" "122" "123" "124" "125" "126" "127" "128" "129" "130" "131" "132" "133" "134" "135" "136" "137" "138" "139" "140"
 [141] "141" "142" "143" "144" "145" "146" "147" "148" "149" "150" "151" "152" "153" "154" "155" "156" "157" "158" "159" "160"
 [161] "161" "162" "163" "164" "165" "166" "167" "168" "169" "170" "171" "172" "173" "174" "175" "176" "177" "178" "179" "180"
 [181] "181" "182" "183" "184" "185" "186" "187" "188" "189" "190" "191" "192" "193" "194" "195" "196" "197" "198" "199" "200"
 [201] "201" "202" "203" "204" "205" "206" "207" "208" "209" "210" "211" "212" "213" "214" "215" "216" "217" "218" "219" "220"
 [221] "221" "222" "223" "224" "225" "226" "227" "228" "229" "230" "231" "232" "233" "234" "235" "236" "237" "238" "239" "240"
 [241] "241" "242" "243" "244" "245" "246" "247" "248" "249" "250" "251" "252" "253" "254" "255" "256" "257" "258" "259" "260"
 [261] "261" "262" "263" "264" "265" "266" "267" "268" "269" "270" "271" "272" "273" "274" "275" "276" "277" "278" "279" "280"
 [281] "281" "282" "283" "284" "285" "286" "287" "288" "289" "290" "291" "292" "293" "294" "295" "296" "297" "298" "299" "300"
 [301] "301" "302" "303" "304" "305" "306" "307" "308" "309" "310" "311" "312" "313" "314" "315" "316" "317" "318" "319" "320"
 [321] "321" "322" "323" "324" "325" "326" "327" "328" "329" "330" "331" "332" "333" "334" "335" "336" "337" "338" "339" "340"
 [341] "341" "342" "343" "344" "345" "346" "347" "348" "349" "350" "351" "352" "353" "354" "355" "356" "357" "358" "359" "360"
 [361] "361" "362" "363" "364" "365" "366" "367" "368" "369" "370" "371" "372" "373" "374" "375" "376" "377" "378" "379" "380"
 [381] "381" "382" "383" "384" "385" "386" "387" "388" "389" "390" "391" "392" "393" "394" "395" "396" "397" "398" "399" "400"
 [401] "401" "402" "403" "404" "405" "406" "407" "408" "409" "410" "411" "412" "413" "414" "415" "416" "417" "418" "419" "420"
 [421] "421" "422" "423" "424" "425" "426" "427" "428" "429" "430" "431" "432" "433" "434" "435" "436" "437" "438" "439" "440"
 [441] "441" "442" "443" "444" "445" "446" "447" "448" "449" "450" "451" "452" "453" "454" "455" "456" "457" "458" "459" "460"
 [461] "461" "462" "463" "464" "465" "466" "467" "468" "469" "470" "471" "472" "473" "474" "475" "476" "477" "478" "479" "480"
 [481] "481" "482" "483" "484" "485" "486" "487" "488" "489" "490" "491" "492" "493" "494" "495" "496" "497" "498" "499" "500"
 [501] "501" "502" "503" "504" "505" "506" "507" "508" "509" "510" "511" "512" "513" "514" "515" "516" "517" "518" "519" "520"
 [521] "521" "522" "523" "524" "525" "526" "527" "528" "529" "530" "531" "532" "533" "534" "535" "536" "537" "538" "539" "540"
 [541] "541" "542" "543" "544" "545" "546" "547" "548" "549" "550" "551" "552" "553" "554" "555" "556" "557" "558" "559" "560"
 [561] "561" "562" "563" "564" "565" "566" "567" "568" "569" "570" "571" "572" "573" "574" "575" "576" "577" "578" "579" "580"
 [581] "581" "582" "583" "584"
```

Checking the Data Types

```
> str(Data)
'data.frame':	584 obs. of  11 variables:
 $ Age                       : int 65 62 62 58 72 62 116 62 18 85 ...
 $ Gender                    : chr "Female" "Male" "Male" "Male" "Male" "Female" "Male" "Male" "Male" "Male" ...
 $ Total_Bilirubin           : chr "0.7" "10.9" "7.3" "1.0" "3.9" "1.3" "0.6" "1.8" "1.3" "0.9" ...
 $ Direct_Bilirubin          : chr "0.1" "5.5" "4.1" "0.3" "2.0" "0.3" "0.1" "0.8" "0.6" "0.3" ...
 $ Alkaline_Phosphotase      : chr "187" "699" "490" "195" "268" "303" "098" "208" "100" "175" ...
 $ ALamino_Aminotransferase  : chr "16" "64" "60" "14" "50" "18" "35" "29" "24" "35" ...
 $ Aspartate_Aminotransferase : chr "18" "100" "68" "20" "112" "12" "31" "32" "31" "62" ...
 $ Total_Protein             : chr "6.8" "7.5" "7.0" "6.8" "5.1" "7.3" "6.0" "6.4" "7.5" "7.0" ...
 $ Albumin                   : chr "3.3" "3.2" "3.3" "3.4" "2.7" "3.3" "3.1" "2.0" "3.2" "3.3" ...
 $ Albumin_and_Globulin_Ratio : chr "0.9" "0.7" "0.9" "1.0" "0.5" "0.9" "0.9" "0.9" "0.7" "1.0" ...
 $ Selector                  : int 1 1 1 1 2 1 1 1 1 1 ...
```

Checking the Data Types

```
> str(Data)
'data.frame':	584 obs. of  11 variables:
 $ Age                       : int 65 62 62 58 72 62 116 62 18 85 ...
 $ Gender                    : chr "Female" "Male" "Male" "Male" "Male" "Female" "Male" "Male" "Male" "Male" ...
 $ Total_Bilirubin           : chr "0.7" "10.9" "7.3" "1.0" "3.9" "1.3" "0.6" "1.8" "1.3" "0.9" ...
 $ Direct_Bilirubin          : chr "0.1" "5.5" "4.1" "0.3" "2.0" "0.3" "0.1" "0.8" "0.6" "0.3" ...
 $ Alkaline_Phosphotase      : chr "187" "699" "490" "195" "268" "303" "098" "208" "100" "175" ...
 $ ALamino_Aminotransferase  : chr "16" "64" "60" "14" "50" "18" "35" "29" "24" "35" ...
 $ Aspartate_Aminotransferase : chr "18" "100" "68" "20" "112" "12" "31" "32" "31" "62" ...
 $ Total_Protein             : chr "6.8" "7.5" "7.0" "6.8" "5.1" "7.3" "6.0" "6.4" "7.5" "7.0" ...
 $ Albumin                   : chr "3.3" "3.2" "3.3" "3.4" "2.7" "3.3" "3.1" "2.0" "3.2" "3.3" ...
 $ Albumin_and_Globulin_Ratio : chr "0.9" "0.7" "0.9" "1.0" "0.5" "0.9" "0.9" "0.9" "0.7" "1.0" ...
 $ Selector                  : int 1 1 1 1 2 1 1 1 1 1 ...
```

Fixing the Data Types

```
> Data$Gender = factor(Data$Gender)
> Data$Total_Bilirubin = as.numeric(Data$Total_Bilirubin)
> Data$Direct_Bilirubin = as.numeric(Data$Direct_Bilirubin)
> Data$Alkaline_Phosphotase = as.numeric(Data$Alkaline_Phosphotase)
> Data$ALamino_Aminotransferase = as.numeric(Data$ALamino_Aminotransferase)
> Data$Aspartate_Aminotransferase = as.numeric(Data$Aspartate_Aminotransferase)
> Data$Total_Protein = as.numeric(Data$Total_Protein)
> Data$Albumin = as.numeric(Data$Albumin)
> Data$Albumin_and_Globulin_Ratio = as.numeric(Data$Albumin_and_Globulin_Ratio)
> Data$Selector = factor(Data$Selector)
> str(Data)
'data.frame':	584 obs. of  11 variables:
 $ Age                       : int 65 62 62 58 72 62 116 62 18 85 ...
 $ Gender                    : Factor w/ 2 levels "Female","Male": 1 2 2 2 2 1 2 2 2 2 ...
 $ Total_Bilirubin           : num 0.7 10.9 7.3 1.0 3.9 1.3 0.6 1.8 1.3 0.9 ...
 $ Direct_Bilirubin          : num 0.1 5.5 4.1 0.3 2.0 0.3 0.1 0.8 0.6 0.3 ...
 $ Alkaline_Phosphotase      : num 187 699 490 195 268 303 098 208 100 175 ...
 $ ALamino_Aminotransferase  : num 16 64 60 14 50 18 35 29 24 35 ...
 $ Aspartate_Aminotransferase : num 18 100 68 20 112 12 31 32 31 62 ...
 $ Total_Protein             : num 6.8 7.5 7.0 6.8 5.1 7.3 6.0 6.4 7.5 7.0 ...
 $ Albumin                   : num 3.3 3.2 3.3 3.4 2.7 3.3 3.1 2.0 3.2 3.3 ...
 $ Albumin_and_Globulin_Ratio : num 0.9 0.7 0.9 1.0 0.5 0.9 0.9 0.9 0.7 1.0 ...
 $ Selector                  : Factor w/ 2 levels 1,2: 1 1 1 1 2 1 1 1 1 1 ...
```

Checking for Duplicates

```
> Data[duplicated(Data),]
    Age Gender Total_Bilirubin Direct_Bilirubin Alkaline_Phosphotase ALamino_Aminotransferase Aspartate_Aminotransferase Total_Protein
579  40   Male             0.6             0.1               098             35             31             6.0
580  52   Male             0.8             0.2               245             48             49             6.4
581  31   Male             1.3             0.5               184             29             32             6.8
582  35   Male             1.0             0.3               318             79             29             6.4
583  39   Male             0.9             0.3               444             59             73             7.3
584  62 Female             0.7             0.2               463             62             80             5.9
    Albumin Albumin_and_Globulin_Ratio Selector
579     3.1                       0.9        1
580     3.2                       0.8        1
581     3.4                       0.9        1
582     3.2                       1.0        1
583     3.3                       0.9        1
584     3.0                       0.8        2
```

Checking

```
> Data[duplicated(Data),]
    Age Gender Total_Bilirubin Direct_Bilirubin Alkaline_Phosphotase ALamino_Aminotransferase Aspartate_Aminotransferase Total_Protein
579  40   Male             0.6             0.1               098             35             31             6.0
580  52   Male             0.8             0.2               245             48             49             6.4
581  31   Male             1.3             0.5               184             29             32             6.8
582  35   Male             1.0             0.3               318             79             29             6.4
583  39   Male             0.9             0.3               444             59             73             7.3
584  62 Female             0.7             0.2               463             62             80             5.9
    Albumin Albumin_and_Globulin_Ratio Selector
579     3.1                       0.9        1
580     3.2                       0.8        1
581     3.4                       0.9        1
582     3.2                       1.0        1
583     3.3                       0.9        1
584     3.0                       0.8        2
```

Checking Duplicates