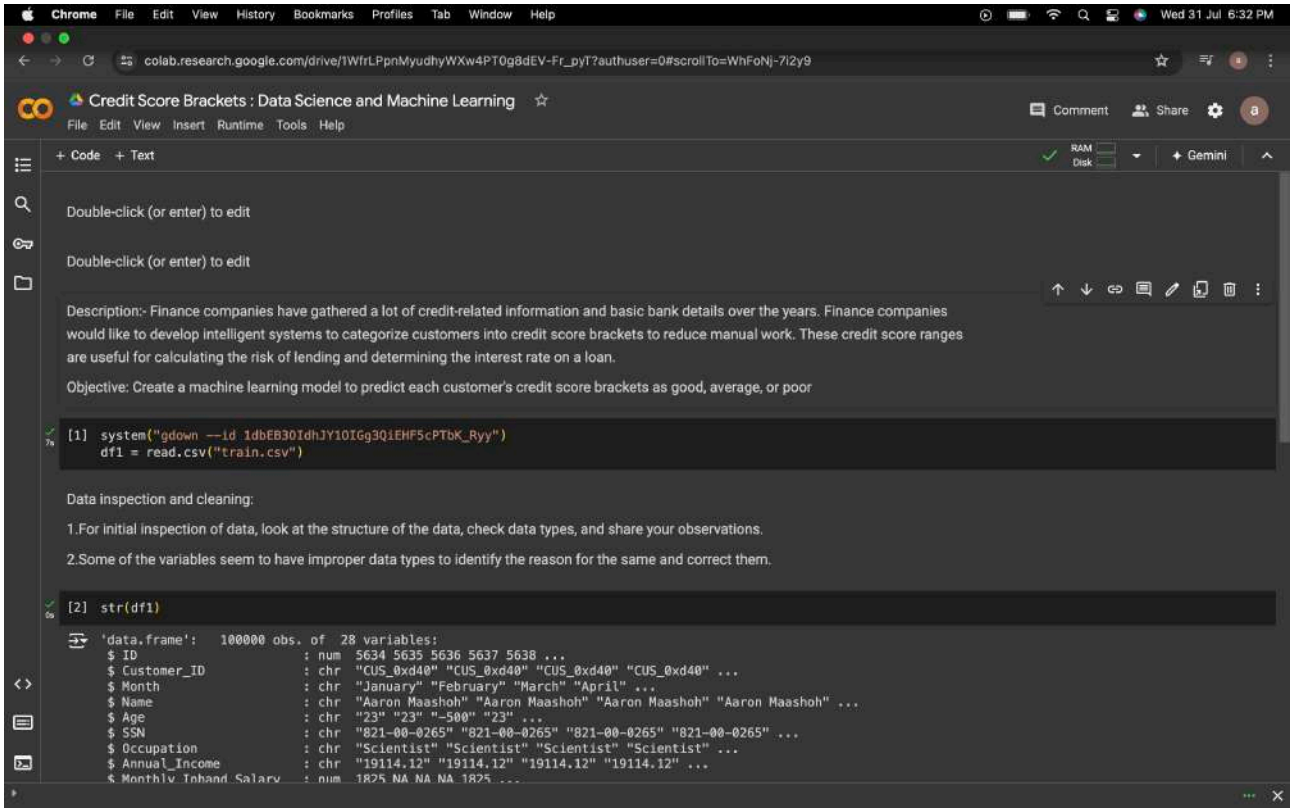
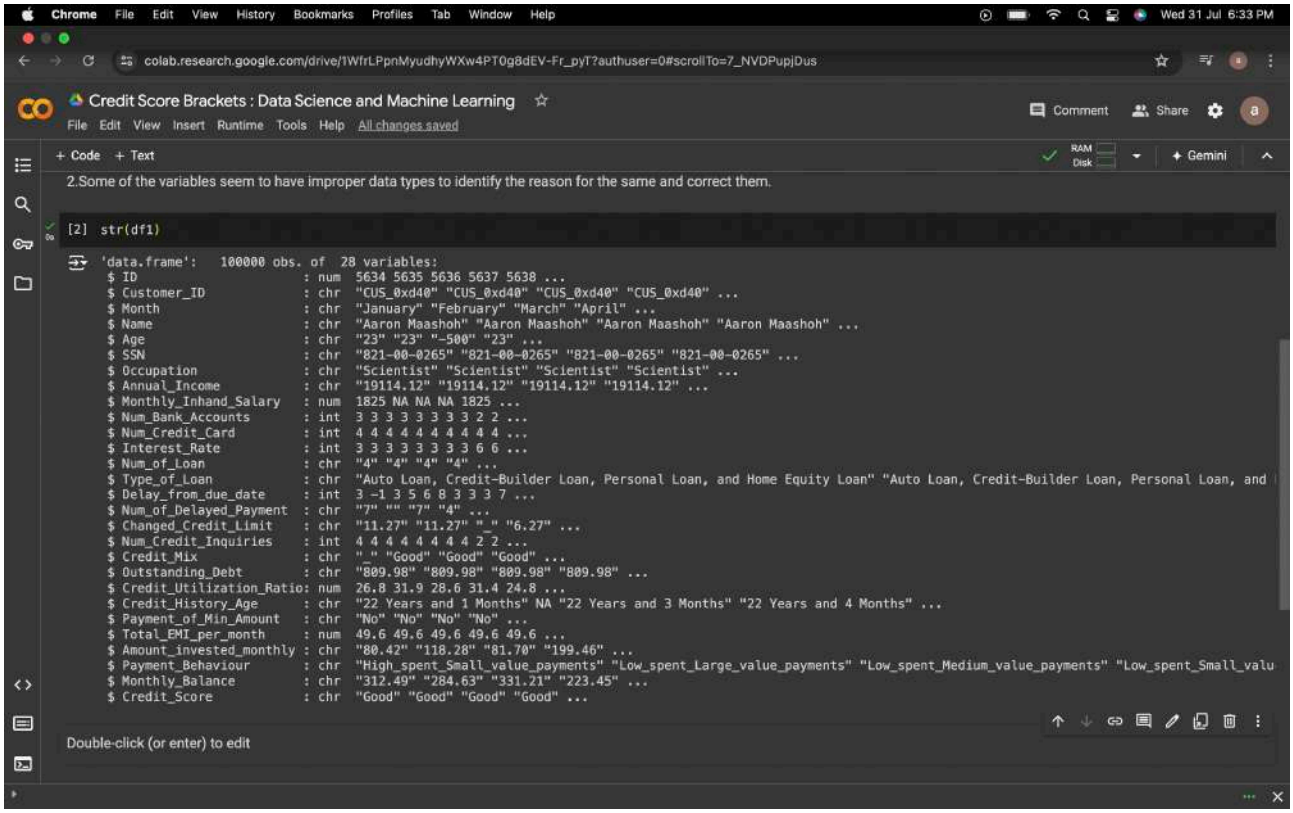
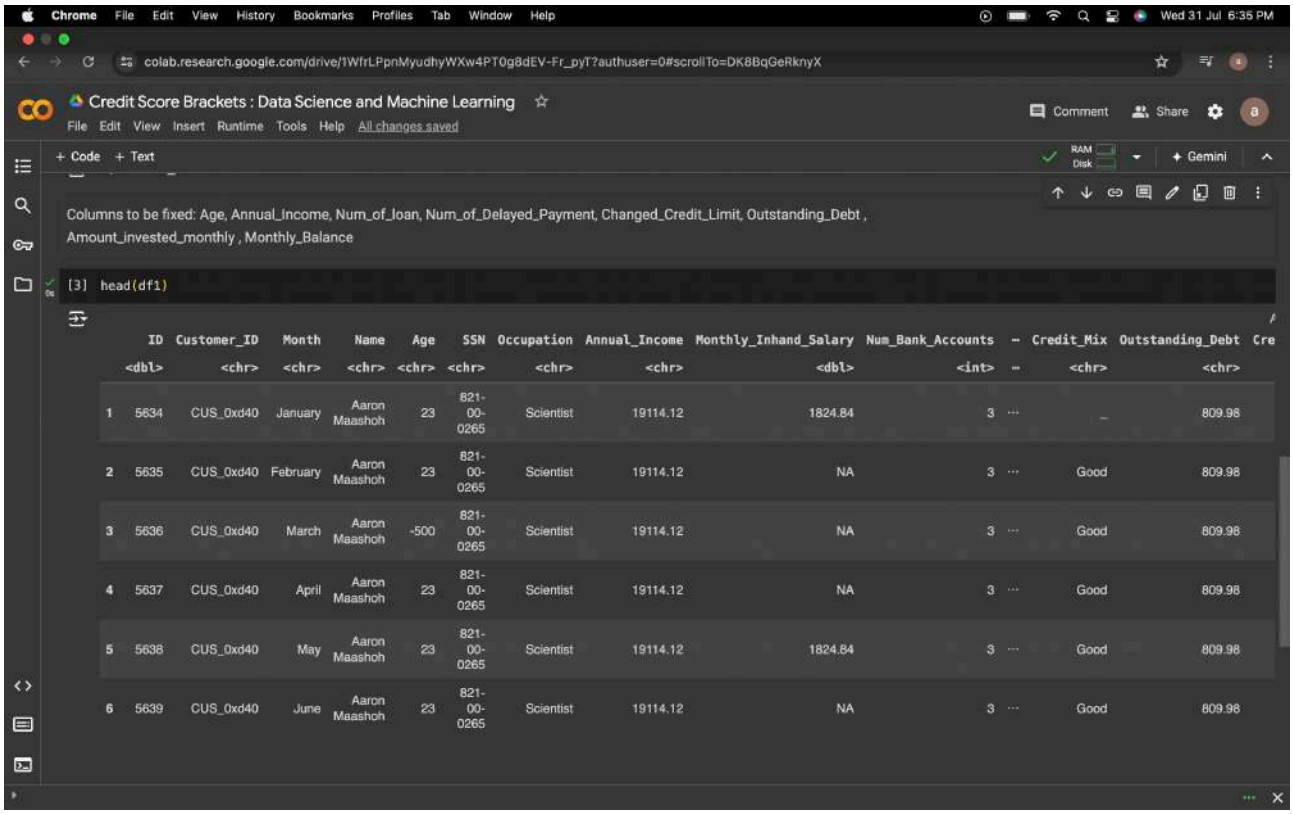
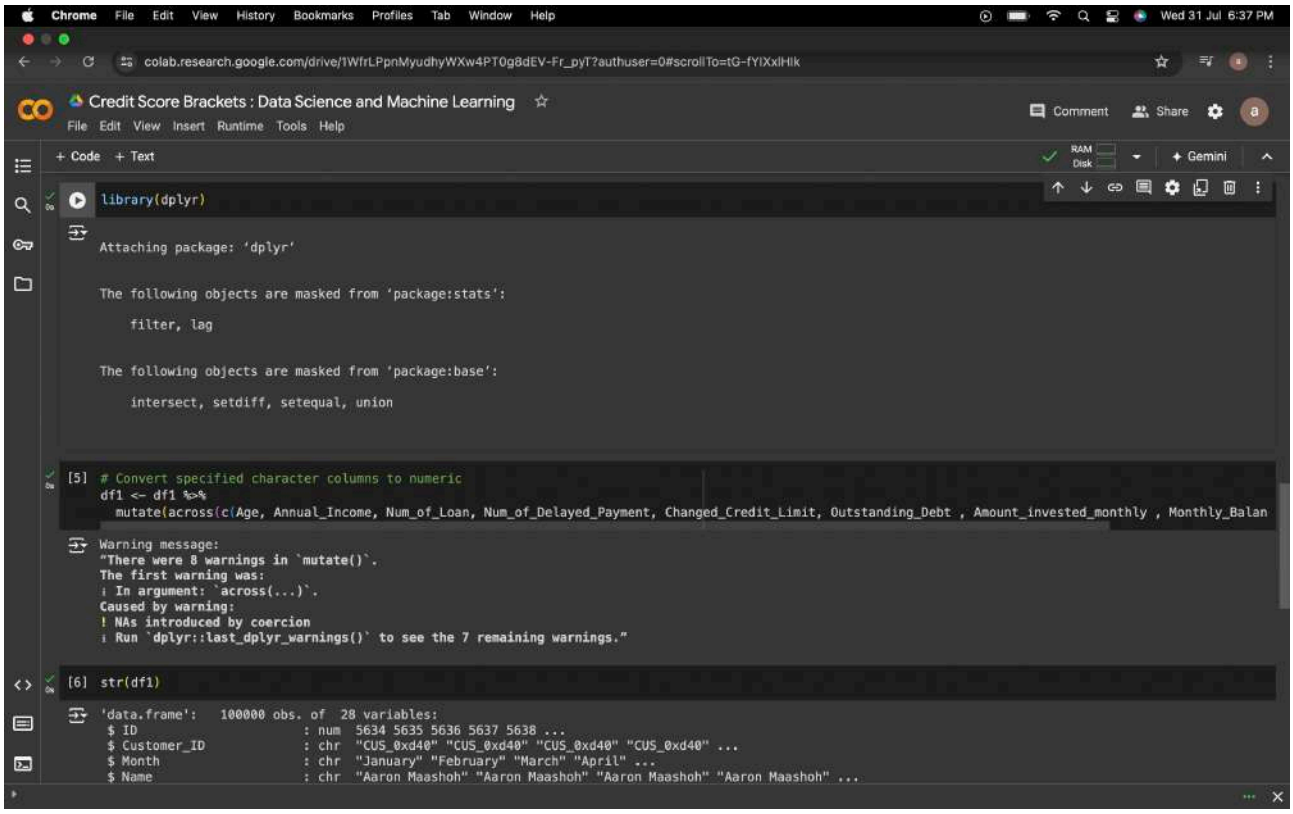
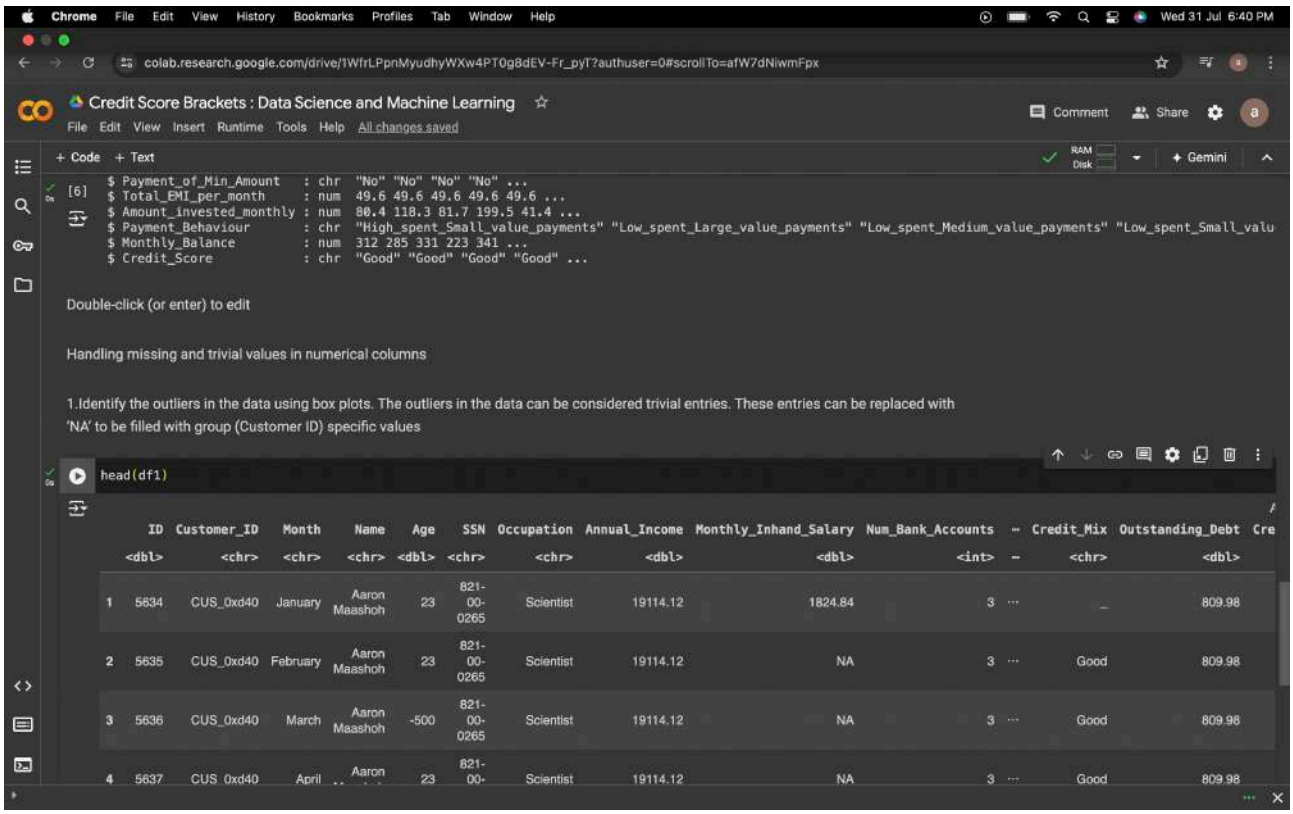
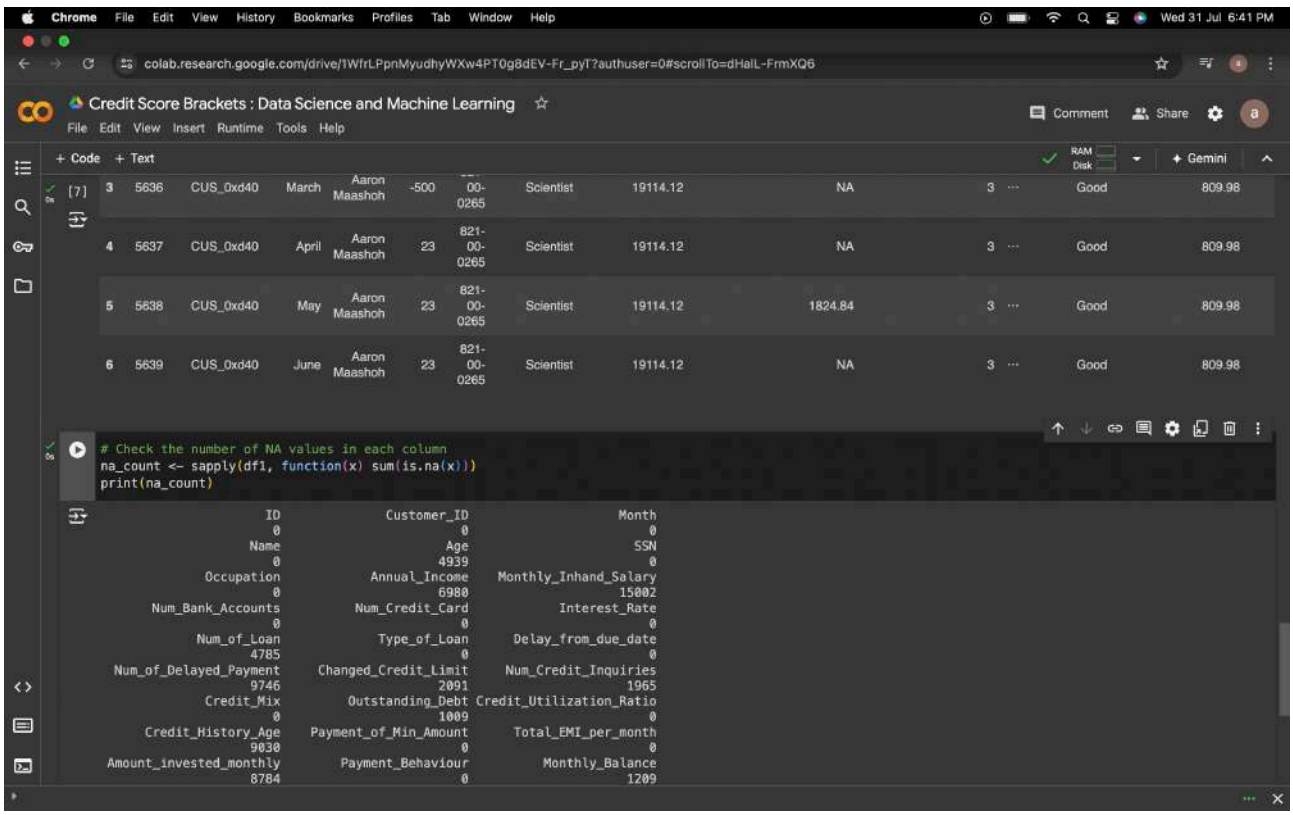

Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 6:32 PM
colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=WhFoNj-7i2y9
Credit Score Brackets : Data Science and Machine Learning
File Edit View Insert Runtime Tools Help
Comment
Share
+ Code + Text
RAM
Disk
Gemini
Double-click (or enter) to edit
Double-click (or enter) to edit
Description:- Finance companies have gathered a lot of credit-related information and basic bank details over the years. Finance companies would like to develop intelligent systems to categorize customers into credit score brackets to reduce manual work. These credit score ranges are useful for calculating the risk of lending and determining the interest rate on a loan.
Objective: Create a machine learning model to predict each customer's credit score brackets as good, average, or poor
[1] system("gdown --id 1dbEB3OIdhJY1OIGg3QiEHF5cPTbK_Ryy")
df1 = read.csv("train.csv")
Data inspection and cleaning:
1.For initial inspection of data, look at the structure of the data, check data types, and share your observations.
2.Some of the variables seem to have improper data types to identify the reason for the same and correct them.
[2] str(df1)
'data.frame': 100000 obs. of 28 variables:
$ ID : num 5634 5635 5636 5637 5638 ...
$ Customer_ID : chr "CUS_0xd40" "CUS_0xd40" "CUS_0xd40" "CUS_0xd40" ...
$ Month : chr "January" "February" "March" "April" ...
$ Name : chr "Aaron Maashoh" "Aaron Maashoh" "Aaron Maashoh" "Aaron Maashoh" ...
$ Age : chr "23" "23" "-500" "23" ...
$ SSN : chr "821-00-0265" "821-00-0265" "821-00-0265" "821-00-0265" ...
$ Occupation : chr "Scientist" "Scientist" "Scientist" "Scientist" ...
$ Annual_Income : chr "19114.12" "19114.12" "19114.12" "19114.12" ...
$ Monthly_Inhand_Salary : num 1825 NA NA NA 1825 ...


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 6:33 PM
colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=7_NVDPupjDus
Credit Score Brackets : Data Science and Machine Learning
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Gemini
2.Some of the variables seem to have improper data types to identify the reason for the same and correct them.
[2] str(df1)
'data.frame': 100000 obs. of 28 variables:
$ ID : num 5634 5635 5636 5637 5638 ...
$ Customer_ID : chr "CUS_0xd40" "CUS_0xd40" "CUS_0xd40" "CUS_0xd40" ...
$ Month : chr "January" "February" "March" "April" ...
$ Name : chr "Aaron Maashoh" "Aaron Maashoh" "Aaron Maashoh" "Aaron Maashoh" ...
$ Age : chr "23" "23" "-500" "23" ...
$ SSN : chr "821-00-0265" "821-00-0265" "821-00-0265" "821-00-0265" ...
$ Occupation : chr "Scientist" "Scientist" "Scientist" "Scientist" ...
$ Annual_Income : chr "19114.12" "19114.12" "19114.12" "19114.12" ...
$ Monthly_Inhand_Salary : num 1825 NA NA NA 1825 ...
$ Num_Bank_Accounts : int 3 3 3 3 3 3 3 3 2 2 ...
$ Num_Credit_Card : int 4 4 4 4 4 4 4 4 4 4 ...
$ Interest_Rate : int 3 3 3 3 3 3 3 3 6 6 ...
$ Num_of_Loan : chr "4" "4" "4" "4" ...
$ Type_of_Loan : chr "Auto Loan, Credit-Builder Loan, Personal Loan, and Home Equity Loan" "Auto Loan, Credit-Builder Loan, Personal Loan, and
$ Delay_from_due_date : int 3 -1 3 5 6 8 3 3 3 7 ...
$ Num_of_Delayed_Payment : chr "7" "" "7" "4" ...
$ Changed_Credit_Limit : chr "11.27" "11.27" "_" "6.27" ...
$ Num_Credit_Inquiries : int 4 4 4 4 4 4 4 4 2 2 ...
$ Credit_Mix : chr "_" "Good" "Good" "Good" ...
$ Outstanding_Debt : chr "809.98" "809.98" "809.98" "809.98" ...
$ Credit_Utilization_Ratio: num 26.8 31.9 28.6 31.4 24.8 ...
$ Credit_History_Age : chr "22 Years and 1 Months" NA "22 Years and 3 Months" "22 Years and 4 Months" ...
$ Payment_of_Min_Amount : chr "No" "No" "No" "No" ...
$ Total_EMI_per_month : num 49.6 49.6 49.6 49.6 49.6 ...
$ Amount_invested_monthly : chr "80.42" "118.28" "81.70" "199.46" ...
$ Payment_Behaviour : chr "High_spent_Small_value_payments" "Low_spent_Large_value_payments" "Low_spent_Medium_value_payments" "Low_spent_Small_valu
$ Monthly_Balance : chr "312.49" "284.63" "331.21" "223.45" ...
$ Credit_Score : chr "Good" "Good" "Good" "Good" ...
Double-click (or enter) to edit


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 6:35 PM
colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=DK8BqGeRknyX
Credit Score Brackets : Data Science and Machine Learning
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Gemini
Columns to be fixed: Age, Annual_Income, Num_of_loan, Num_of_Delayed_Payment, Changed_Credit_Limit, Outstanding_Debt , Amount_invested_monthly , Monthly_Balance
[3] head(df1)
ID Customer_ID Month Name Age SSN Occupation Annual_Income Monthly_Inhand_Salary Num_Bank_Accounts ... Credit_Mix Outstanding_Debt Cre
<dbl> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <dbl> <int> ... <chr> <chr>
1 5634 CUS_0xd40 January Aaron Maashoh 23 821-00-0265 Scientist 19114.12 1824.84 3 ... _ 809.98
2 5635 CUS_0xd40 February Aaron Maashoh 23 821-00-0265 Scientist 19114.12 NA 3 ... Good 809.98
3 5636 CUS_0xd40 March Aaron Maashoh -500 821-00-0265 Scientist 19114.12 NA 3 ... Good 809.98
4 5637 CUS_0xd40 April Aaron Maashoh 23 821-00-0265 Scientist 19114.12 NA 3 ... Good 809.98
5 5638 CUS_0xd40 May Aaron Maashoh 23 821-00-0265 Scientist 19114.12 1824.84 3 ... Good 809.98
6 5639 CUS_0xd40 June Aaron Maashoh 23 821-00-0265 Scientist 19114.12 NA 3 ... Good 809.98


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 6:37 PM
colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=tG-fYIXxlHlk
Credit Score Brackets : Data Science and Machine Learning
File Edit View Insert Runtime Tools Help
Comment
Share
+ Code + Text
RAM
Disk
Gemini
library(dplyr)
Attaching package: 'dplyr'
The following objects are masked from 'package:stats':
filter, lag
The following objects are masked from 'package:base':
intersect, setdiff, setequal, union
[5] # Convert specified character columns to numeric
df1 <- df1 %>%
mutate(across(c(Age, Annual_Income, Num_of_Loan, Num_of_Delayed_Payment, Changed_Credit_Limit, Outstanding_Debt , Amount_invested_monthly , Monthly_Balan
Warning message:
"There were 8 warnings in `mutate()`.
The first warning was:
ℹ In argument: `across(...)`.
Caused by warning:
! NAs introduced by coercion
ℹ Run `dplyr::last_dplyr_warnings()` to see the 7 remaining warnings."
[6] str(df1)
'data.frame': 100000 obs. of 28 variables:
$ ID : num 5634 5635 5636 5637 5638 ...
$ Customer_ID : chr "CUS_0xd40" "CUS_0xd40" "CUS_0xd40" "CUS_0xd40" ...
$ Month : chr "January" "February" "March" "April" ...
$ Name : chr "Aaron Maashoh" "Aaron Maashoh" "Aaron Maashoh" "Aaron Maashoh" ...


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 6:40 PM
colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=afW7dNiwmFpx
Credit Score Brackets : Data Science and Machine Learning
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
RAM Disk
Gemini
[6]
$ Payment_of_Min_Amount : chr "No" "No" "No" "No" ...
$ Total_EMI_per_month : num 49.6 49.6 49.6 49.6 49.6 ...
$ Amount_invested_monthly : num 80.4 118.3 81.7 199.5 41.4 ...
$ Payment_Behaviour : chr "High_spent_Small_value_payments" "Low_spent_Large_value_payments" "Low_spent_Medium_value_payments" "Low_spent_Small_valu
$ Monthly_Balance : num 312 285 331 223 341 ...
$ Credit_Score : chr "Good" "Good" "Good" "Good" ...
Double-click (or enter) to edit
Handling missing and trivial values in numerical columns
1.Identify the outliers in the data using box plots. The outliers in the data can be considered trivial entries. These entries can be replaced with 'NA' to be filled with group (Customer ID) specific values
head(df1)
ID Customer_ID Month Name Age SSN Occupation Annual_Income Monthly_Inhand_Salary Num_Bank_Accounts ... Credit_Mix Outstanding_Debt Cre
<dbl> <chr> <chr> <chr> <dbl> <chr> <chr> <dbl> <dbl> <int> ... <chr> <dbl>
1 5634 CUS_0xd40 January Aaron Maashoh 23 821-00-0265 Scientist 19114.12 1824.84 3 ... _ 809.98
2 5635 CUS_0xd40 February Aaron Maashoh 23 821-00-0265 Scientist 19114.12 NA 3 ... Good 809.98
3 5636 CUS_0xd40 March Aaron Maashoh -500 821-00-0265 Scientist 19114.12 NA 3 ... Good 809.98
4 5637 CUS_0xd40 April Aaron 23 821-00- Scientist 19114.12 NA 3 ... Good 809.98


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 6:41 PM
colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=dHalL-FrmXQ6
Credit Score Brackets : Data Science and Machine Learning
File Edit View Insert Runtime Tools Help
Comment Share
+ Code + Text
RAM Disk
Gemini
[7]
3 5636 CUS_0xd40 March Aaron Maashoh -500 00-0265 Scientist 19114.12 NA 3 ... Good 809.98
4 5637 CUS_0xd40 April Aaron Maashoh 23 821-00-0265 Scientist 19114.12 NA 3 ... Good 809.98
5 5638 CUS_0xd40 May Aaron Maashoh 23 821-00-0265 Scientist 19114.12 1824.84 3 ... Good 809.98
6 5639 CUS_0xd40 June Aaron Maashoh 23 821-00-0265 Scientist 19114.12 NA 3 ... Good 809.98
# Check the number of NA values in each column
na_count <- sapply(df1, function(x) sum(is.na(x)))
print(na_count)
ID 0
Customer_ID 0
Month 0
Name 0
Age 4939
SSN 0
Occupation 0
Annual_Income 6980
Monthly_Inhand_Salary 15002
Num_Bank_Accounts 0
Num_Credit_Card 0
Interest_Rate 0
Num_of_Loan 4785
Type_of_Loan 0
Delay_from_due_date 0
Num_of_Delayed_Payment 9746
Changed_Credit_Limit 2091
Num_Credit_Inquiries 1965
Credit_Mix 0
Outstanding_Debt 1009
Credit_Utilization_Ratio 0
Credit_History_Age 9030
Payment_of_Min_Amount 0
Total_EMI_per_month 0
Amount_invested_monthly 8784
Payment_Behaviour 0
Monthly_Balance 1209

Chrome File Edit View History Bookmarks Profiles Tab Window Help Wed 31 Jul 6:43 PM

colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=dHalL-FrmXQ6

Credit Score Brackets : Data Science and Machine Learning

File Edit View Insert Runtime Tools Help All changes saved

Comment Share a

+ Code + Text RAM Disk Gemini

```
# Check the number of NA values in each column
na_count <- sapply(df1, function(x) sum(is.na(x)))
print(na_count)
```

```
                     ID             Customer_ID                   Month
                      0                       0                       0
                   Name                     Age                     SSN
                      0                    4939                       0
             Occupation          Annual_Income    Monthly_Inhand_Salary
                      0                    6980                   15002
        Num_Bank_Accounts        Num_Credit_Card           Interest_Rate
                      0                       0                       0
            Num_of_Loan           Type_of_Loan      Delay_from_due_date
                   4785                       0                       0
  Num_of_Delayed_Payment  Changed_Credit_Limit     Num_Credit_Inquiries
                   9746                    2091                    1965
             Credit_Mix       Outstanding_Debt Credit_Utilization_Ratio
                      0                    1009                       0
      Credit_History_Age  Payment_of_Min_Amount       Total_EMI_per_month
                   9030                       0                       0
Amount_invested_monthly       Payment_Behaviour         Monthly_Balance
                   8784                       0                    1209
            Credit_Score
                      0
```

```
[9] summary(df1$Monthly_Balance)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   0.01  270.11  336.73  402.55  470.26 1602.04    1209
```

```
[10] boxplot(df1$Monthly_Balance, main = "Box Plot of Values", ylab = "Value")

     # Identify outliers
     outliers <- boxplot.stats(df1$Monthly_Balance)$out
```

Chrome File Edit View History Bookmarks Profiles Tab Window Help Wed 31 Jul 6:43 PM

colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=Siavo-2rmt7W

Credit Score Brackets : Data Science and Machine Learning

File Edit View Insert Runtime Tools Help All changes saved

Comment Share a

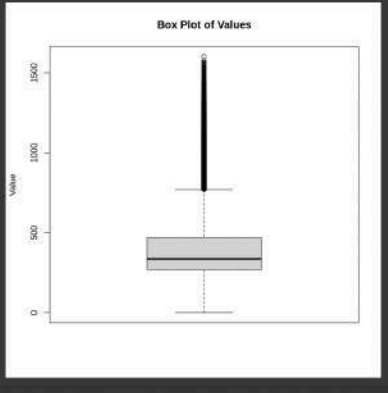
+ Code + Text RAM Disk Gemini

[9] summary(df1$Monthly_Balance)

   Min. 1st Qu. Median   Mean 3rd Qu.   Max.   NA's
   0.01 270.11 336.73 402.55 470.26 1602.04   1209

boxplot(df1$Monthly_Balance, main = "Box Plot of Values", ylab = "Value")

# Identify outliers
outliers <- boxplot.stats(df1$Monthly_Balance)$out



Chrome File Edit View History Bookmarks Profiles Tab Window Help Wed 31 Jul 6:44 PM

colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=hdrhTiuum-zv

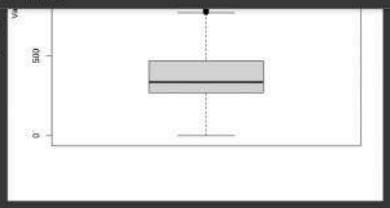
Credit Score Brackets : Data Science and Machine Learning

File Edit View Insert Runtime Tools Help

Comment Share a

+ Code + Text RAM Disk Gemini

[10]




```
outliers
```

1043.32 · 998.87 · 810.78 · 963.92 · 968.56 · 895.49 · 796.23 · 858.46 · 1038.57 · 899.2 · 963.25 · 1140.07 · 802.3 · 785.26 · 772.41 · 792.03 · 854.52 · 823.71 · 878.25 · 920.44 · 782.57 · 939.87 · 881.66 · 970.32 · 939.62 · 947.61 · 955.31 · 791.73 · 913.03 · 936.31 · 792.58 · 822.19 · 880.43 · 853.55 · 785.53 · 799.44 · 795.55 · 886.98 · 840.03 · 887.73 · 879.75 · 776.33 · 884.91 · 940.39 · 1157.86 · 789.44 · 1162.42 · 1184.9 · 1120.75 · 829.6 · 874.48 · 803.48 · 799.4 · 808.78 · 1023.89 · 942.74 · 1142.15 · 936.88 · 782.08 · 1026.92 · 1180.07 · 1078.49 · 878.78 · 1057.46 · 790.35 · 803.99 · 818.24 · 857.83 · 855.59 · 863.33 · 904.48 · 961.67 · 1081.29 · 1103.54 · 1061.92 · 1031.08 · 855.05 · 1145.99 · 815.35 · 938.08 · 1005.28 · 775.49 · 804.91 · 778.61 · 808.74 · 771.59 · 949.5 · 833.3 · 893.24 · 774.36 · 1041.08 · 859.97 · 926.74 · 922.06 · 946.54 · 821.35 · 882.42 · 779.63 · 956.67 · 992.22 · 1042.03 · 779.57 · 823.09 · 835.48 · 830.89 · 857.32 · 898.15 · 814.58 · 877.96 · 950.47 · 933.67 · 902.14 · 929.23 · 808.69 · 774.72 · 784.35 · 795.03 · 811.72 · 935.9 · 903.92 · 1007.85 · 939.02 · 1130.72 · 1143.58 · 1166.09 · 1214.32 · 1030.65 · 1051.26 · 1220.01 · 837.74 · 962.57 · 969.72 · 810.98 · 1020.69 · 844.5 · 898.33 · 955.9 · 822.7 · 817.17 · 790.59 · 818.37 · 814.86 · 899.8 · 1041.08 · 1008.33 · 1008.59 · 922.75 · 968.8 · 811.77 · 826.32 · 840.9 · 995.91 · 801.71 · 997.64 · 942.1 · 848.57 · 1046.54 · 944.4 · 1191.37 · 1108.42 · 1284.3 · 1419.32 · 1209.91 · 1029.72 · 882.05 · 846.11 · 794.57 · 959.56 · 958.56 · 1019.35 · 912.51 · 773.99 · 846.51 · 1102.1 · 799.03 · 846.95 · 807.35 · 791.02 · 812.66 · 876.87 · 832.57 · 783.75 · 901.98 · 774.2 · 983.71 · 869.36 · 960.08 · 791.65 · 1052.32 · 826.44 · 1071.97 · 1176.4 · 938.2 · 942.27 · 1018.67 · 1082.38 · 1114.89 · 1188.89 · 1001.75 · 1033.17 · ⋯ · 942.89 · 1047.33 · 1075.79 · 960.39 · 960.6 · 1118.62 · 812.81 · 859.62 · 819.76 · 805.71 · 786.19 · 803.97 · 784.36 · 876.33 · 820.53 · 844.3 · 773.79 · 786.57 · 874.2 · 802.81 · 846.87 · 1176.44 · 1229.54 · 1099.44 · 911.23 · 867.57 · 1092.16 · 889.72 · 1008.61 · 797.33 · 1084.09 · 1018.57 · 1092.82 · 1185.3 · 1270.03 · 820.54 · 1095.33 · 774.71 · 792.77 · 1042.58 · 1106.24 · 1270.86 · 784.13 · 1020.85 · 1066.61 · 1294.24 · 822.3 · 829.57 · 836.12 · 832.67 · 1063.2 · 1221.52 · 947.7 · 873.45 · 1344.34 · 968.04 · 801.88 · 918.81 · 848.66 · 866.45 · 894.33 · 891.23 · 817.7 · 847.6 · 808.25 · 812.88 · 895.91 · 1366.34 · 1383.14 · 1400.66 · 1151.63 · 1411.13 · 1076.86 · 1348.87 · 1272.45 · 1228.61 · 1299.6 · 1236.77 · 1307.77 · 801.68 · 807.55 · 937.89 · 944.48 · 951.75 · 828.16 · 897.45 · 788.18 · 914.07 · 899.98 · 916.92 · 840.57 · 1156.31 · 1266.51 · 981.46 · 1074.37 · 1112.97 · 1364.17 · 776.35 · 803.97 · 839.69 · 801.45 · 848.84 · 857.1 · 840.56 · 854.82 · 837.51 · 843.91 · 810.92 · 962.61 · 900.35 · 782.47 · 781.19 · 797.25 · 819.71 · 796.1 · 906.6 · 947.58 · 958.51 · 958.77 · 789.38 · 798.68 · 986.21 · 861.93 · 829 · 787.31 · 1005.11 · 914.98 · 862.37 · 1086.1 · 1207.08 · 773.38 · 1163.13 · 972.92 · 1205.75 · 919.85 · 943.6 · 1251.25 · 972.9 · 1123.25 · 971.73 · 1108.55 · 1178.06 · 1244.15 · 1377.66 · 1279.53 · 1295.27 · 1328.94 · 780.05 · 856.49 · 841.89 · 886.05 · 1145.74 · 1104.51 · 1052.74 · 1105.71 · 1034.44 · 907.9 · 1028.43 · 877.26 · 1059.41 · 1193.88 · 1121.26 · 917.23 · 879.82 · 844.25 · 855.81 · 937.27 · 853.54 · 860 · 882.22 · 835.18 · 1210.85 · 1261.36 · 1258.92 · 912.74 · 1096.89 · 883.06 · 903.63 · 801.19 · 801.2 · 839.69 · 875.23 · 836.36 · 835.82 · 896.19 · 811.82 · 842.59 · 904.52 · 810.51 · 801.75 · 792.42 · 874.91 · 808.33 · 823.29 · 819.57 · 805 · 864.47 · 833.09 · 830.79 · 869.91

Chrome File Edit View History Bookmarks Profiles Tab Window Help Wed 31 Jul 6:45 PM

colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=k2wJH2a0nU3A

Credit Score Brackets : Data Science and Machine Learning

File Edit View Insert Runtime Tools Help

Comment Share a

+ Code + Text RAM Disk Gemini

[11]

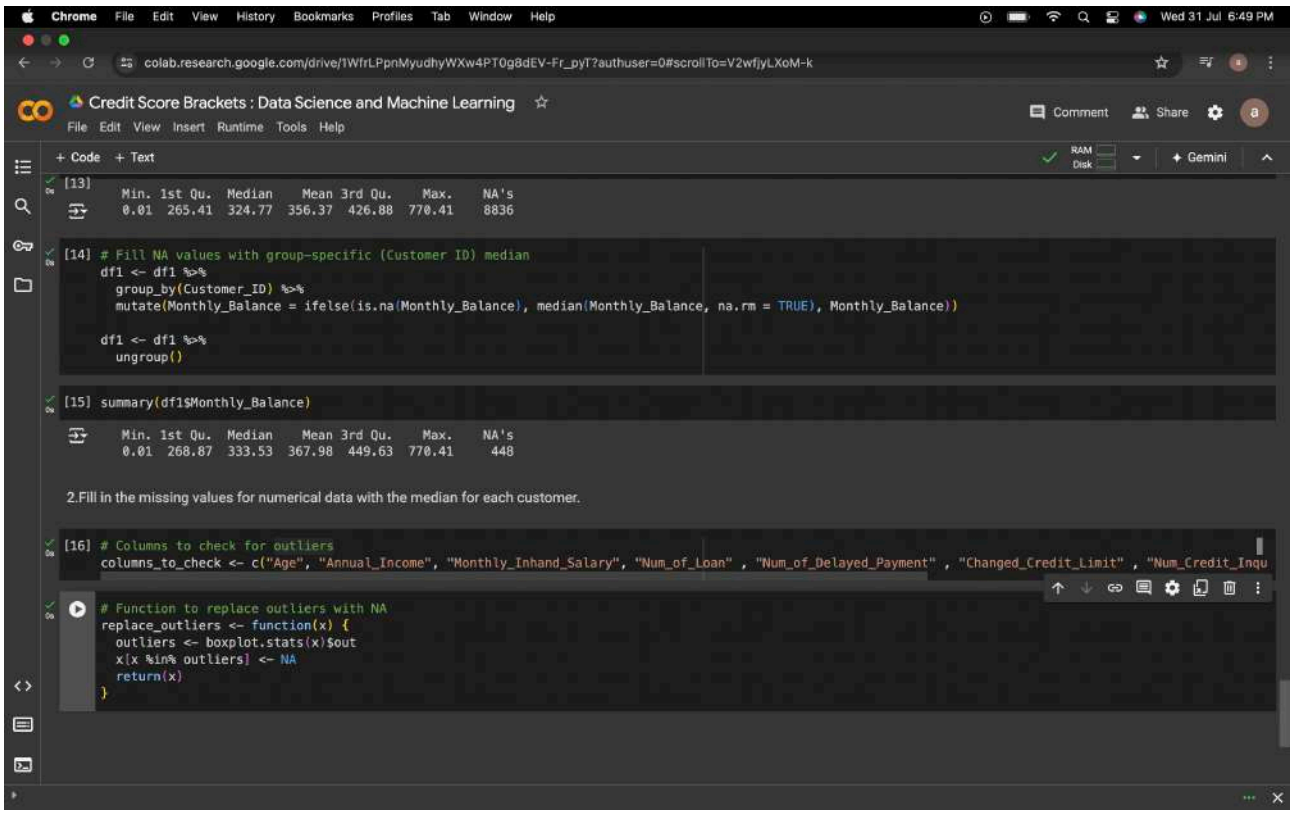
814.86 · 899.8 · 1041.08 · 1008.33 · 1008.59 · 922.75 · 968.8 · 811.77 · 826.32 · 840.9 · 995.91 · 801.71 · 997.64 · 942.1 · 848.57 · 1046.54 · 944.4 · 1191.37 · 1108.42 · 1284.3 · 1419.32 · 1209.91 · 1029.72 · 882.05 · 846.11 · 794.57 · 959.56 · 958.56 · 1019.35 · 912.51 · 773.99 · 846.51 · 1102.1 · 799.03 · 846.95 · 807.35 · 791.02 · 812.66 · 876.87 · 832.57 · 783.75 · 901.98 · 774.2 · 983.71 · 869.36 · 960.08 · 791.65 · 1052.32 · 826.44 · 1071.97 · 1176.4 · 938.2 · 942.27 · 1018.67 · 1082.38 · 1114.89 · 1188.89 · 1001.75 · 1033.17 · ⋯ · 942.89 · 1047.33 · 1075.79 · 960.39 · 960.6 · 1118.62 · 812.81 · 859.62 · 819.76 · 805.71 · 786.19 · 803.97 · 784.36 · 876.33 · 820.53 · 844.3 · 773.79 · 786.57 · 874.2 · 802.81 · 846.87 · 1176.44 · 1229.54 · 1099.44 · 911.23 · 867.57 · 1092.16 · 889.72 · 1008.61 · 797.33 · 1084.09 · 1018.57 · 1092.82 · 1185.3 · 1270.03 · 820.54 · 1095.33 · 774.71 · 792.77 · 1042.58 · 1106.24 · 1270.86 · 784.13 · 1020.85 · 1066.61 · 1294.24 · 822.3 · 829.57 · 836.12 · 832.67 · 1063.2 · 1221.52 · 947.7 · 873.45 · 1344.34 · 968.04 · 801.88 · 918.81 · 848.66 · 866.45 · 894.33 · 891.23 · 817.7 · 847.6 · 808.25 · 812.88 · 895.91 · 1366.34 · 1383.14 · 1400.66 · 1151.63 · 1411.13 · 1076.86 · 1348.87 · 1272.45 · 1228.61 · 1299.6 · 1236.77 · 1307.77 · 801.68 · 807.55 · 937.89 · 944.48 · 951.75 · 828.16 · 897.45 · 788.18 · 914.07 · 899.98 · 916.92 · 840.57 · 1156.31 · 1266.51 · 981.46 · 1074.37 · 1112.97 · 1364.17 · 776.35 · 803.97 · 839.69 · 801.45 · 848.84 · 857.1 · 840.56 · 854.82 · 837.51 · 843.91 · 810.92 · 962.61 · 900.35 · 782.47 · 781.19 · 797.25 · 819.71 · 796.1 · 906.6 · 947.58 · 958.51 · 958.77 · 789.38 · 798.68 · 986.21 · 861.93 · 829 · 787.31 · 1005.11 · 914.98 · 862.37 · 1086.1 · 1207.08 · 773.38 · 1163.13 · 972.92 · 1205.75 · 919.85 · 943.6 · 1251.25 · 972.9 · 1123.25 · 971.73 · 1108.55 · 1178.06 · 1244.15 · 1377.66 · 1279.53 · 1295.27 · 1328.94 · 780.05 · 856.49 · 841.89 · 886.05 · 1145.74 · 1104.51 · 1052.74 · 1105.71 · 1034.44 · 907.9 · 1028.43 · 877.26 · 1059.41 · 1193.88 · 1121.26 · 917.23 · 879.82 · 844.25 · 855.81 · 937.27 · 853.54 · 860 · 882.22 · 835.18 · 1210.85 · 1261.36 · 1258.92 · 912.74 · 1096.89 · 883.06 · 903.63 · 801.19 · 801.2 · 839.69 · 875.23 · 836.36 · 835.82 · 896.19 · 811.82 · 842.59 · 904.52 · 810.51 · 801.75 · 792.42 · 874.91 · 808.33 · 823.29 · 819.57 · 805 · 864.47 · 833.09 · 830.79 · 869.91

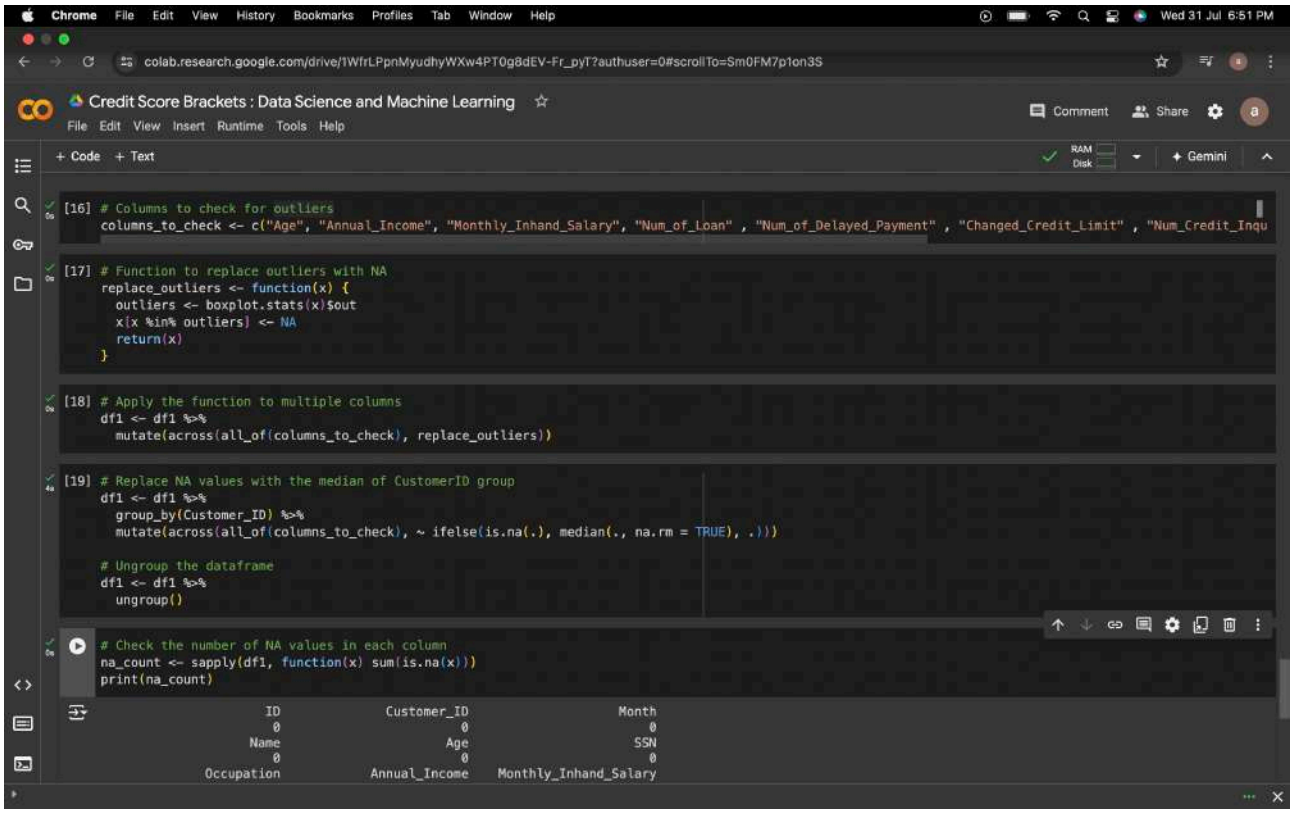
[12]
```
# Replace outliers with NA
df1$Monthly_Balance[df1$Monthly_Balance %in% outliers] <- NA
```

[13]
```
summary(df1$Monthly_Balance)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   0.01  265.41  324.77  356.37  426.88  770.41    8836
```

```
# Fill NA values with group-specific (Customer ID) median
df1 <- df1 %>%
  group_by(Customer_ID) %>%
  mutate(Monthly_Balance = ifelse(is.na(Monthly_Balance), median(Monthly_Balance, na.rm = TRUE), Monthly_Balance))

df1 <- df1 %>%
  ungroup()
```


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 6:49 PM
colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=V2wfjyLXoM-k
Credit Score Brackets : Data Science and Machine Learning
File Edit View Insert Runtime Tools Help
Comment
Share
+ Code + Text
RAM
Disk
Gemini
[13]
Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
0.01 265.41 324.77 356.37 426.88 770.41 8836
[14] # Fill NA values with group-specific (Customer ID) median
df1 <- df1 %>%
group_by(Customer_ID) %>%
mutate(Monthly_Balance = ifelse(is.na(Monthly_Balance), median(Monthly_Balance, na.rm = TRUE), Monthly_Balance))
df1 <- df1 %>%
ungroup()
[15] summary(df1$Monthly_Balance)
Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
0.01 268.87 333.53 367.98 449.63 770.41 448
2.Fill in the missing values for numerical data with the median for each customer.
[16] # Columns to check for outliers
columns_to_check <- c("Age", "Annual_Income", "Monthly_Inhand_Salary", "Num_of_Loan" , "Num_of_Delayed_Payment" , "Changed_Credit_Limit" , "Num_Credit_Inqu
# Function to replace outliers with NA
replace_outliers <- function(x) {
outliers <- boxplot.stats(x)$out
x[x %in% outliers] <- NA
return(x)
}


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 6:51 PM
colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=Sm0FM7p1on3S
Credit Score Brackets : Data Science and Machine Learning
File Edit View Insert Runtime Tools Help
Comment
Share
+ Code + Text
RAM
Disk
Gemini
[16] # Columns to check for outliers
columns_to_check <- c("Age", "Annual_Income", "Monthly_Inhand_Salary", "Num_of_Loan" , "Num_of_Delayed_Payment" , "Changed_Credit_Limit" , "Num_Credit_Inqu
[17] # Function to replace outliers with NA
replace_outliers <- function(x) {
outliers <- boxplot.stats(x)$out
x[x %in% outliers] <- NA
return(x)
}
[18] # Apply the function to multiple columns
df1 <- df1 %>%
mutate(across(all_of(columns_to_check), replace_outliers))
[19] # Replace NA values with the median of CustomerID group
df1 <- df1 %>%
group_by(Customer_ID) %>%
mutate(across(all_of(columns_to_check), ~ ifelse(is.na(.), median(., na.rm = TRUE), .)))
# Ungroup the dataframe
df1 <- df1 %>%
ungroup()
# Check the number of NA values in each column
na_count <- sapply(df1, function(x) sum(is.na(x)))
print(na_count)
ID Customer_ID Month
0 0 0
Name Age SSN
0 0 0
Occupation Annual_Income Monthly_Inhand_Salary

Chrome File Edit View History Bookmarks Profiles Tab Window Help Wed 31 Jul 6:52 PM

colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=OR89Oj7Do3vM

Credit Score Brackets : Data Science and Machine Learning

File Edit View Insert Runtime Tools Help

Comment Share

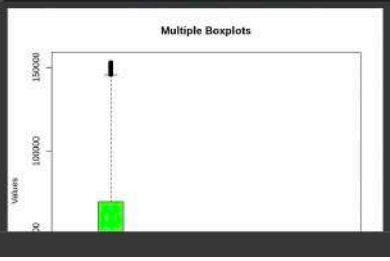
+ Code + Text

RAM Disk Gemini

[20]

```
Num_Bank_Accounts          0
Num_Credit_Card            0
Interest_Rate              0
Num_of_Loan                0
Type_of_Loan               0
Delay_from_due_date        0
Num_of_Delayed_Payment     0
Changed_Credit_Limit     224
Num_Credit_Inquiries       0
Credit_Mix                 0
Outstanding_Debt        5264
Credit_Utilization_Ratio   0
Credit_History_Age      9030
Payment_of_Min_Amount      0
Total_EMI_per_month        0
Amount_invested_monthly   16
Payment_Behaviour          0
Monthly_Balance          448
Credit_Score               0
```

```
data <- df1[, columns_to_check]

boxplot(data,
        main = "Multiple Boxplots",
        xlab = "Variables",
        ylab = "Values",
        col = c("red", "green", "blue"))
```



Chrome File Edit View History Bookmarks Profiles Tab Window Help Wed 31 Jul 6:53 PM

colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=OR89Oj7Do3vM

Credit Score Brackets : Data Science and Machine Learning

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

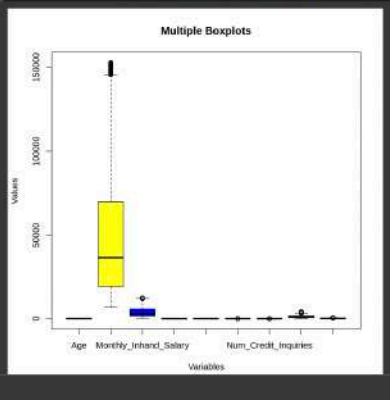
+ Code + Text

RAM Disk Gemini

[20]

```
Num_Bank_Accounts          0
Num_Credit_Card            0
Interest_Rate              0
Num_of_Loan                0
Type_of_Loan               0
Delay_from_due_date        0
Num_of_Delayed_Payment     0
Changed_Credit_Limit     224
Num_Credit_Inquiries       0
Credit_Mix                 0
Outstanding_Debt        5264
Credit_Utilization_Ratio   0
Credit_History_Age      9030
Payment_of_Min_Amount      0
Total_EMI_per_month        0
Amount_invested_monthly   16
Payment_Behaviour          0
Monthly_Balance          448
Credit_Score               0
```

```
data <- df1[, columns_to_check]

boxplot(data,
        main = "Multiple Boxplots",
        xlab = "Variables",
        ylab = "Values",
        col = c("red", "yellow", "blue"))
```



Chrome File Edit View History Bookmarks Profiles Tab Window Help Wed 31 Jul 6:54 PM

colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=-CtqQwfDpOfb

Credit Score Brackets : Data Science and Machine Learning

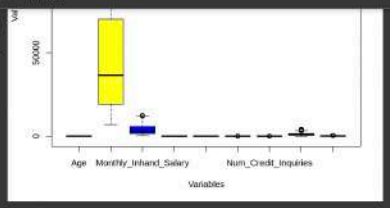
File Edit View Insert Runtime Tools Help

Comment Share

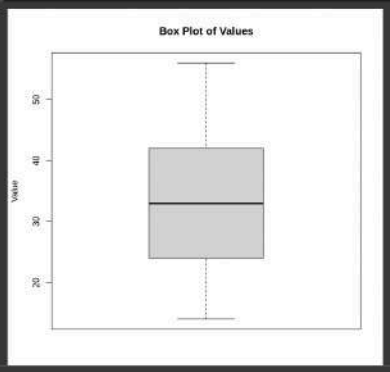
+ Code + Text

RAM Disk Gemini

[22]




```
boxplot(df1$Age, main = "Box Plot of Values", ylab = "Value")
```



Chrome File Edit View History Bookmarks Profiles Tab Window Help Wed 31 Jul 6:56 PM

colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=-CtqQwfDpOfb

Credit Score Brackets : Data Science and Machine Learning

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk Gemini

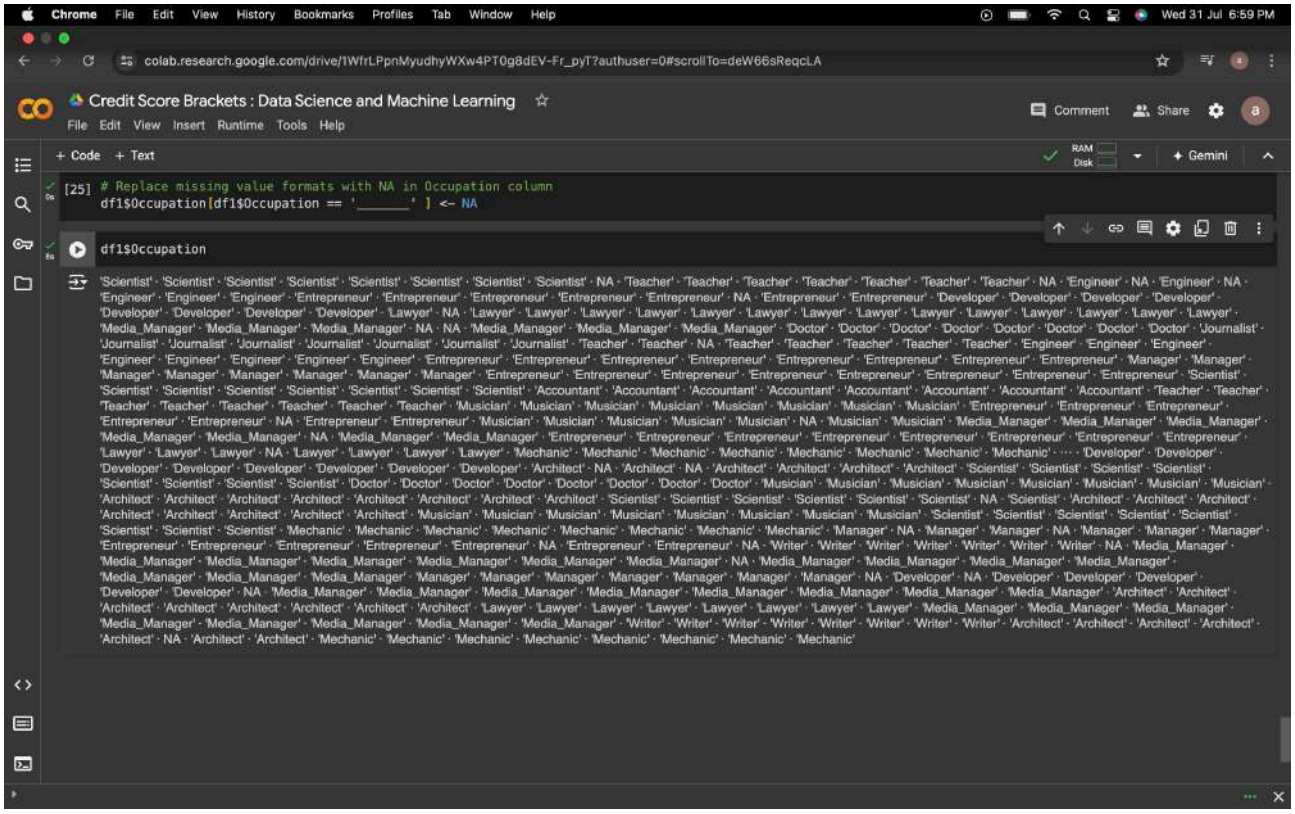
Handling missing and trivial values in character columns

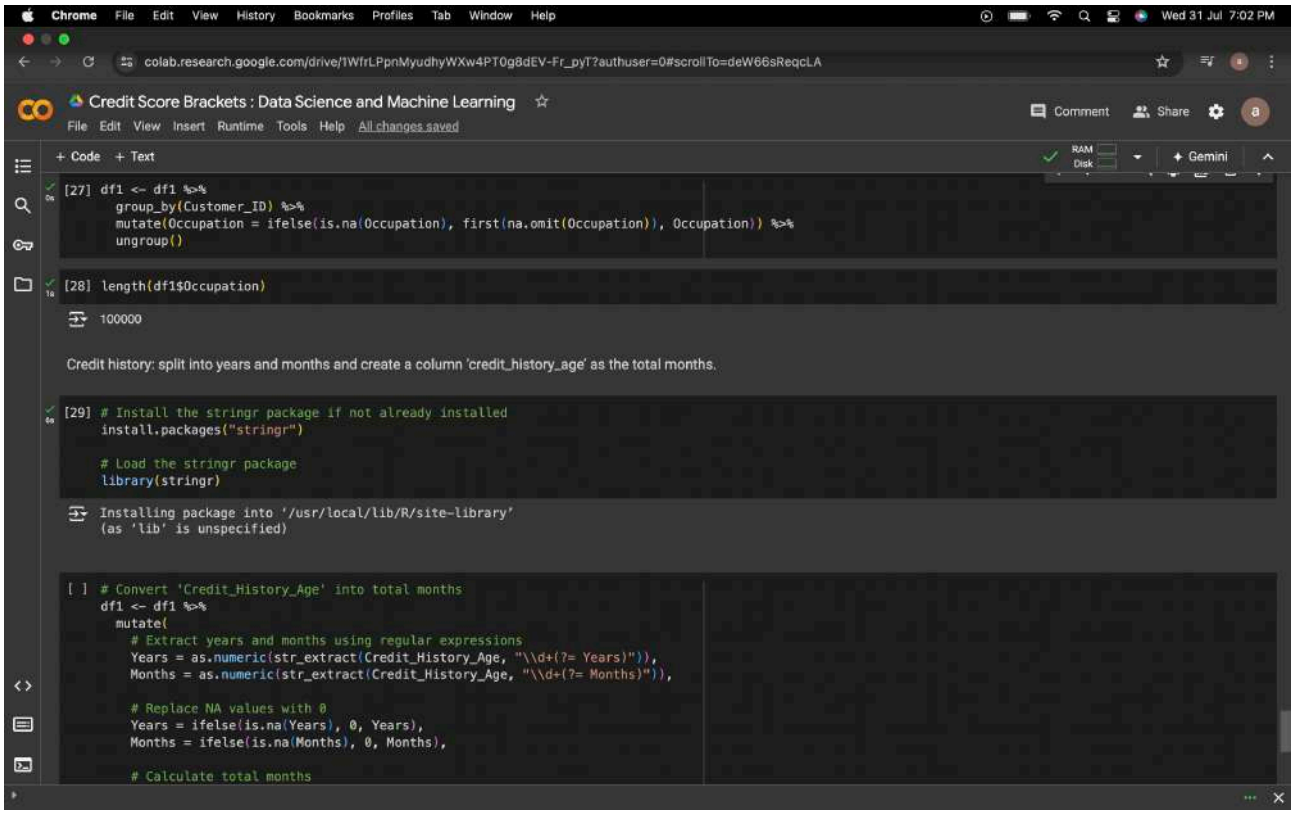
Character missing values may also be represented as an empty string or blank space or any other format. Identify the formats in each variable and replace them.

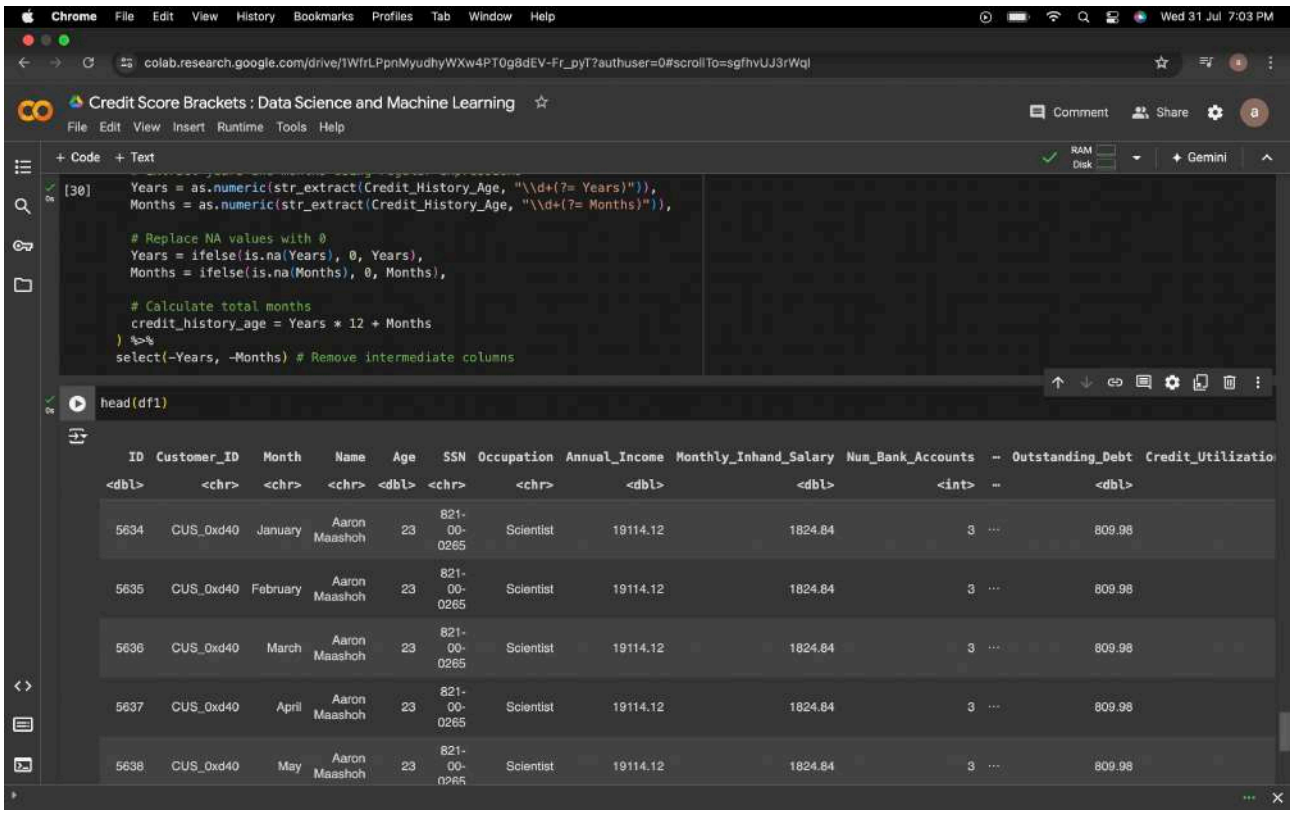
Hint: Identify the missing value format in occupation and replace it with the occupation based on customer id.

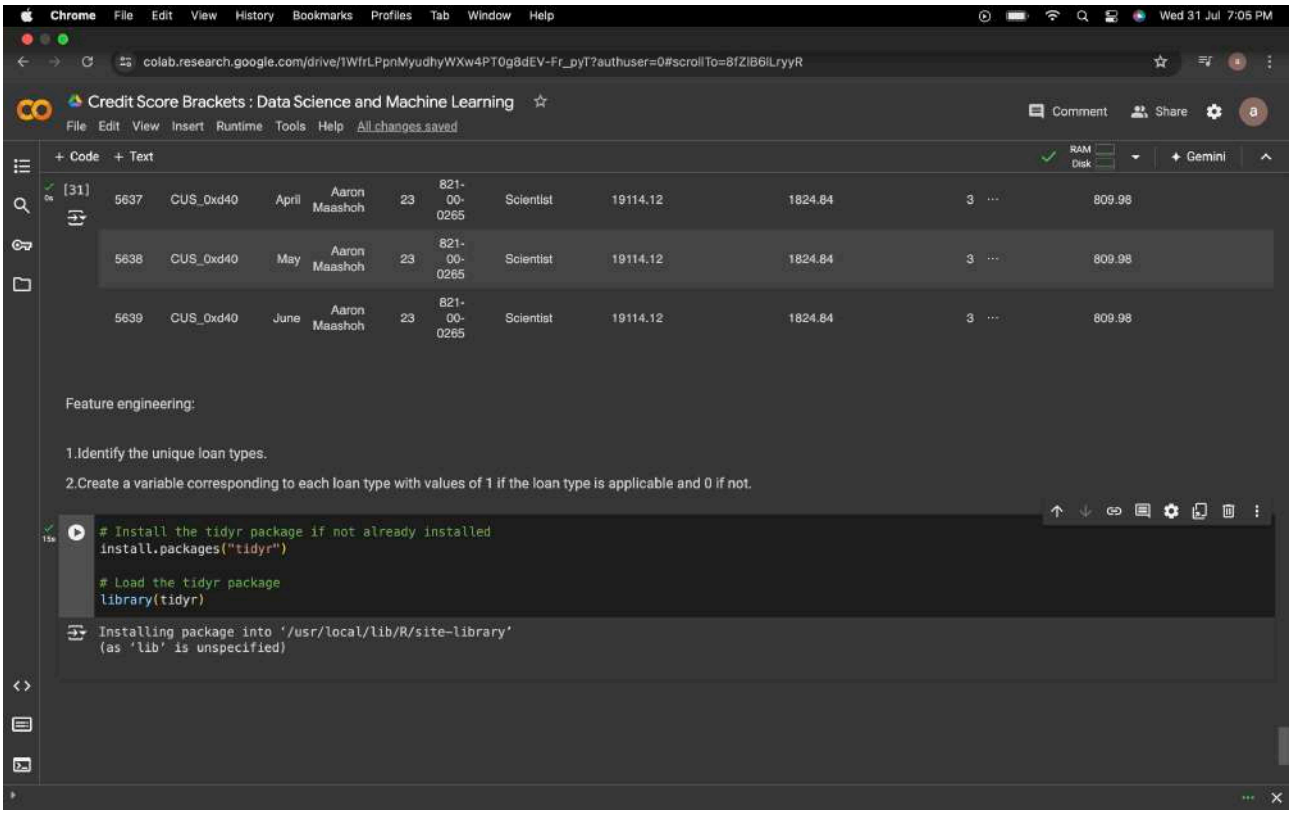
```
[24] df1$Occupation
```

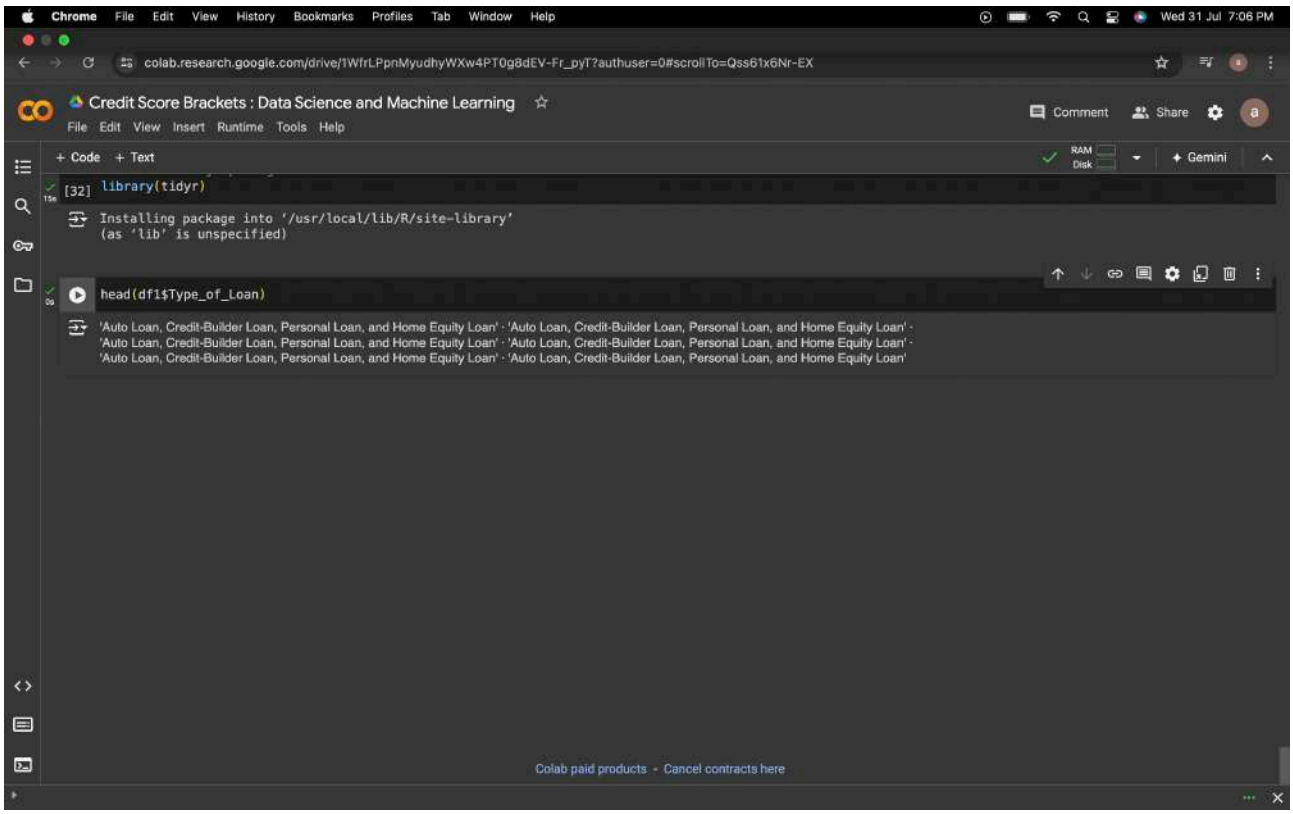
'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Scientist' '_______' 'Teacher' 'Teacher' 'Teacher' 'Teacher' 'Teacher' 'Teacher' 'Teacher' '_______' 'Engineer' '_______' 'Engineer' '_______' 'Engineer' 'Engineer' 'Engineer' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' '_______' 'Entrepreneur' 'Entrepreneur' 'Developer' 'Developer' 'Developer' 'Developer' 'Developer' 'Developer' 'Developer' 'Developer' 'Lawyer' '_______' 'Lawyer' 'Lawyer' 'Lawyer' 'Lawyer' 'Lawyer' 'Lawyer' 'Lawyer' 'Lawyer' 'Lawyer' 'Lawyer' 'Lawyer' 'Media_Manager' 'Media_Manager' 'Media_Manager' '_______' '_______' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Doctor' 'Doctor' 'Doctor' 'Doctor' 'Doctor' 'Doctor' 'Doctor' 'Doctor' 'Doctor' 'Journalist' 'Journalist' 'Journalist' 'Journalist' 'Journalist' 'Journalist' 'Journalist' 'Teacher' 'Teacher' '_______' 'Teacher' 'Teacher' 'Teacher' 'Teacher' 'Teacher' 'Engineer' 'Engineer' 'Engineer' 'Engineer' 'Engineer' 'Engineer' 'Engineer' 'Engineer' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Manager' 'Manager' 'Manager' 'Manager' 'Manager' 'Manager' 'Manager' 'Manager' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Accountant' 'Accountant' 'Accountant' 'Accountant' 'Accountant' 'Accountant' 'Accountant' 'Accountant' 'Teacher' 'Teacher' 'Teacher' 'Teacher' 'Teacher' 'Teacher' 'Teacher' 'Teacher' 'Musician' 'Musician' 'Musician' 'Musician' 'Musician' 'Musician' 'Musician' 'Musician' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' '_______' 'Entrepreneur' 'Entrepreneur' 'Musician' 'Musician' 'Musician' 'Musician' 'Musician' '_______' 'Musician' 'Musician' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Media_Manager' '_______' 'Media_Manager' 'Media_Manager' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Lawyer' 'Lawyer' 'Lawyer' '_______' 'Lawyer' 'Lawyer' 'Lawyer' 'Lawyer' 'Mechanic' 'Mechanic' 'Mechanic' 'Mechanic' 'Mechanic' 'Mechanic' 'Mechanic' 'Mechanic' 'Mechanic' ... 'Developer' 'Developer' 'Developer' 'Developer' 'Developer' 'Developer' 'Developer' 'Developer' 'Architect' '_______' 'Architect' '_______' 'Architect' 'Architect' 'Architect' 'Architect' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Doctor' 'Doctor' 'Doctor' 'Doctor' 'Doctor' 'Doctor' 'Doctor' 'Musician' 'Musician' 'Musician' 'Musician' 'Musician' 'Musician' 'Musician' 'Musician' 'Architect' 'Architect' 'Architect' 'Architect' 'Architect' 'Architect' 'Architect' 'Architect' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Scientist' '_______' 'Scientist' 'Architect' 'Architect' 'Architect' 'Architect' 'Architect' 'Architect' 'Architect' 'Architect' 'Musician' 'Musician' 'Musician' 'Musician' 'Musician' 'Musician' 'Musician' 'Musician' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Scientist' 'Mechanic' 'Mechanic' 'Mechanic' 'Mechanic' 'Mechanic' 'Mechanic' 'Mechanic' 'Mechanic' 'Manager' '_______' 'Manager' 'Manager' '_______' 'Manager' 'Manager' 'Manager' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' 'Entrepreneur' '_______' 'Entrepreneur' 'Entrepreneur' 'Writer' 'Writer' 'Writer' 'Writer' 'Writer' 'Writer' 'Writer' '_______' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Media_Manager' '_______' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Manager' 'Manager' 'Manager' 'Manager' 'Manager' 'Manager' '_______' 'Developer' '_______' 'Developer' 'Developer' 'Developer' 'Developer' 'Developer' '_______' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Architect' 'Architect' 'Architect' 'Architect' 'Architect' 'Architect' 'Architect' 'Architect' 'Lawyer' 'Lawyer' 'Lawyer' 'Lawyer' 'Lawyer' 'Lawyer' 'Lawyer' 'Lawyer' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Media_Manager' 'Writer' 'Writer' 'Writer' 'Writer' 'Writer' 'Writer' 'Writer' 'Writer' 'Architect' 'Architect' 'Architect' 'Architect' 'Architect' '_______' 'Architect' 'Architect' 'Mechanic' 'Mechanic' 'Mechanic' 'Mechanic' 'Mechanic' 'Mechanic' 'Mechanic' 'Mechanic'

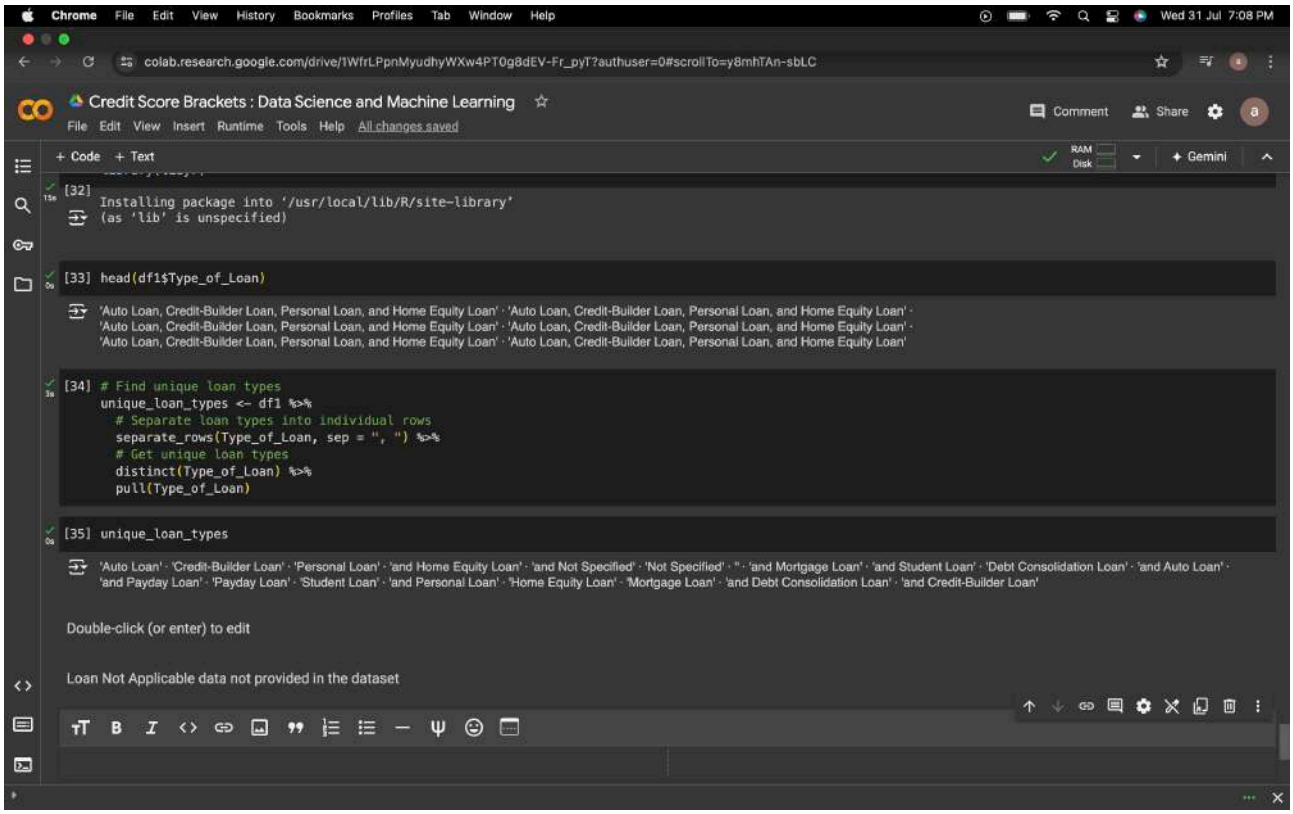

Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 6:59 PM
colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=deW66sReqcLA
Credit Score Brackets : Data Science and Machine Learning
File Edit View Insert Runtime Tools Help
Comment
Share
+ Code + Text
RAM
Disk
Gemini
[25] # Replace missing value formats with NA in Occupation column
df1$Occupation[df1$Occupation == '_______'] <- NA
df1$Occupation


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 7:02 PM
colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=deW66sReqcLA
Credit Score Brackets : Data Science and Machine Learning
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Gemini
[27] df1 <- df1 %>%
group_by(Customer_ID) %>%
mutate(Occupation = ifelse(is.na(Occupation), first(na.omit(Occupation)), Occupation)) %>%
ungroup()
[28] length(df1$Occupation)
100000
Credit history: split into years and months and create a column 'credit_history_age' as the total months.
[29] # Install the stringr package if not already installed
install.packages("stringr")
# Load the stringr package
library(stringr)
Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)
[ ] # Convert 'Credit_History_Age' into total months
df1 <- df1 %>%
mutate(
# Extract years and months using regular expressions
Years = as.numeric(str_extract(Credit_History_Age, "\\d+(?= Years)")),
Months = as.numeric(str_extract(Credit_History_Age, "\\d+(?= Months)")),
# Replace NA values with 0
Years = ifelse(is.na(Years), 0, Years),
Months = ifelse(is.na(Months), 0, Months),
# Calculate total months


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 7:03 PM
colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=sgfhvUJ3rWql
Credit Score Brackets : Data Science and Machine Learning
File Edit View Insert Runtime Tools Help
Comment
Share
+ Code + Text
RAM Disk
Gemini
[30]
Years = as.numeric(str_extract(Credit_History_Age, "\\d+(?= Years)")),
Months = as.numeric(str_extract(Credit_History_Age, "\\d+(?= Months)")),
# Replace NA values with 0
Years = ifelse(is.na(Years), 0, Years),
Months = ifelse(is.na(Months), 0, Months),
# Calculate total months
credit_history_age = Years * 12 + Months
) %>%
select(-Years, -Months) # Remove intermediate columns
head(df1)
ID Customer_ID Month Name Age SSN Occupation Annual_Income Monthly_Inhand_Salary Num_Bank_Accounts ... Outstanding_Debt Credit_Utilizatio
<dbl> <chr> <chr> <chr> <dbl> <chr> <chr> <dbl> <dbl> <int> ... <dbl>
5634 CUS_0xd40 January Aaron Maashoh 23 821-00-0265 Scientist 19114.12 1824.84 3 ... 809.98
5635 CUS_0xd40 February Aaron Maashoh 23 821-00-0265 Scientist 19114.12 1824.84 3 ... 809.98
5636 CUS_0xd40 March Aaron Maashoh 23 821-00-0265 Scientist 19114.12 1824.84 3 ... 809.98
5637 CUS_0xd40 April Aaron Maashoh 23 821-00-0265 Scientist 19114.12 1824.84 3 ... 809.98
5638 CUS_0xd40 May Aaron Maashoh 23 821-00-0265 Scientist 19114.12 1824.84 3 ... 809.98


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 7:05 PM
colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=8fZlB6lLryyR
Credit Score Brackets : Data Science and Machine Learning
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM Disk
Gemini
[31]
5637 CUS_0xd40 April Aaron Maashoh 23 821-00-0265 Scientist 19114.12 1824.84 3 ... 809.98
5638 CUS_0xd40 May Aaron Maashoh 23 821-00-0265 Scientist 19114.12 1824.84 3 ... 809.98
5639 CUS_0xd40 June Aaron Maashoh 23 821-00-0265 Scientist 19114.12 1824.84 3 ... 809.98
Feature engineering:
1.Identify the unique loan types.
2.Create a variable corresponding to each loan type with values of 1 if the loan type is applicable and 0 if not.
# Install the tidyr package if not already installed
install.packages("tidyr")
# Load the tidyr package
library(tidyr)
Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 7:06 PM
colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=Qss61x6Nr-EX
Credit Score Brackets : Data Science and Machine Learning
File Edit View Insert Runtime Tools Help
Comment
Share
+ Code + Text
RAM
Disk
Gemini
[32] library(tidyr)
Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)
head(df1$Type_of_Loan)
'Auto Loan, Credit-Builder Loan, Personal Loan, and Home Equity Loan' · 'Auto Loan, Credit-Builder Loan, Personal Loan, and Home Equity Loan' ·
'Auto Loan, Credit-Builder Loan, Personal Loan, and Home Equity Loan' · 'Auto Loan, Credit-Builder Loan, Personal Loan, and Home Equity Loan' ·
'Auto Loan, Credit-Builder Loan, Personal Loan, and Home Equity Loan' · 'Auto Loan, Credit-Builder Loan, Personal Loan, and Home Equity Loan'
Colab paid products - Cancel contracts here


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 7:08 PM
colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=y8mhTAn-sbLC
Credit Score Brackets : Data Science and Machine Learning
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Gemini
[32] Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)
[33] head(df1$Type_of_Loan)
'Auto Loan, Credit-Builder Loan, Personal Loan, and Home Equity Loan' · 'Auto Loan, Credit-Builder Loan, Personal Loan, and Home Equity Loan' ·
'Auto Loan, Credit-Builder Loan, Personal Loan, and Home Equity Loan' · 'Auto Loan, Credit-Builder Loan, Personal Loan, and Home Equity Loan' ·
'Auto Loan, Credit-Builder Loan, Personal Loan, and Home Equity Loan' · 'Auto Loan, Credit-Builder Loan, Personal Loan, and Home Equity Loan'
[34] # Find unique loan types
unique_loan_types <- df1 %>%
# Separate loan types into individual rows
separate_rows(Type_of_Loan, sep = ", ") %>%
Get unique loan types
distinct(Type_of_Loan) %>%
pull(Type_of_Loan)
[35] unique_loan_types
'Auto Loan' · 'Credit-Builder Loan' · 'Personal Loan' · 'and Home Equity Loan' · 'and Not Specified' · 'Not Specified' · '' · 'and Mortgage Loan' · 'and Student Loan' · 'Debt Consolidation Loan' · 'and Auto Loan' ·
'and Payday Loan' · 'Payday Loan' · 'Student Loan' · 'and Personal Loan' · 'Home Equity Loan' · 'Mortgage Loan' · 'and Debt Consolidation Loan' · 'and Credit-Builder Loan'
Double-click (or enter) to edit
Loan Not Applicable data not provided in the dataset

Chrome File Edit View History Bookmarks Profiles Tab Window Help Wed 31 Jul 7:10 PM

colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=MBjLkWq-syK4

Credit Score Brackets : Data Science and Machine Learning

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

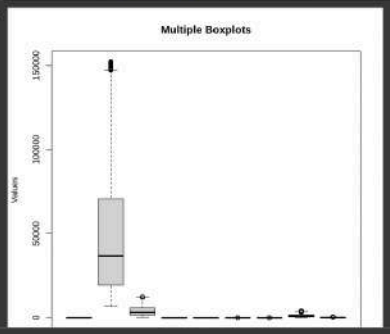
RAM Disk Gemini

Drop all rows with any missing values

```
[36] df1 <- na.omit(df1)
```

Exploratory data analysis:

```
data <- df1[, columns_to_check]

boxplot(data,
        main = "Multiple Boxplots",
        xlab = "Variables",
        ylab = "Values")
```



Chrome File Edit View History Bookmarks Profiles Tab Window Help Wed 31 Jul 7:11 PM

colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=MBjLkWq-syK4

Credit Score Brackets : Data Science and Machine Learning

File Edit View Insert Runtime Tools Help

Comment Share

+ Code + Text

RAM Disk Gemini

1.Study the probability distribution of numerical variables such as age, annual income, monthly balance, and monthly investments

```
[38] str(df1)
```

```
tibble [83,935 × 29] (S3: tbl_df/tbl/data.frame)
 $ ID                      : num [1:83935] 5634 5636 5637 5638 5639 ...
 $ Customer_ID             : chr [1:83935] "CUS_0xd40" "CUS_0xd40" "CUS_0xd40" "CUS_0xd40" ...
 $ Month                   : chr [1:83935] "January" "March" "April" "May" ...
 $ Name                    : chr [1:83935] "Aaron Maashoh" "Aaron Maashoh" "Aaron Maashoh" "Aaron Maashoh" ...
 $ Age                     : num [1:83935] 23 23 23 23 23 23 28 28 28 28 ...
 $ SSN                     : chr [1:83935] "821-00-0265" "821-00-0265" "821-00-0265" "821-00-0265" ...
 $ Occupation              : chr [1:83935] "Scientist" "Scientist" "Scientist" "Scientist" ...
 $ Annual_Income           : num [1:83935] 19114 19114 19114 19114 19114 ...
 $ Monthly_Inhand_Salary   : num [1:83935] 1825 1825 1825 1825 1825 ...
 $ Num_Bank_Accounts       : int [1:83935] 3 3 3 3 3 3 2 2 2 2 ...
 $ Num_Credit_Card         : int [1:83935] 4 4 4 4 4 4 4 4 1385 4 ...
 $ Interest_Rate           : int [1:83935] 3 3 3 3 3 3 6 6 6 6 ...
 $ Num_of_Loan             : num [1:83935] 4 4 4 4 4 4 1 1 1 1 ...
 $ Type_of_Loan            : chr [1:83935] "Auto Loan, Credit-Builder Loan, Personal Loan, and Home Equity Loan" "Auto Loan, Credit-Builder Loan, Personal L
 $ Delay_from_due_date     : int [1:83935] 3 3 5 6 8 3 3 7 3 3 ...
 $ Num_of_Delayed_Payment  : num [1:83935] 7 7 4 6 4 6 4 1 -1 1 ...
 $ Changed_Credit_Limit    : num [1:83935] 11.27 11.27 6.27 11.27 9.27 ...
 $ Num_Credit_Inquiries    : num [1:83935] 4 4 4 4 4 4 2 2 2 2 ...
 $ Credit_Mix              : chr [1:83935] "_" "Good" "Good" "Good" ...
 $ Outstanding_Debt        : num [1:83935] 810 810 810 810 810 ...
 $ Credit_Utilization_Ratio: num [1:83935] 26.8 28.6 31.4 24.8 27.3 ...
 $ Credit_History_Age      : chr [1:83935] "22 Years and 1 Months" "22 Years and 3 Months" "22 Years and 4 Months" "22 Years and 5 Months" ...
 $ Payment_of_Min_Amount   : chr [1:83935] "No" "No" "No" "No" ...
 $ Total_EMI_per_month     : num [1:83935] 49.6 49.6 49.6 49.6 49.6 ...
 $ Amount_invested_monthly : num [1:83935] 80.4 81.7 199.5 41.4 62.4 ...
 $ Payment_Behaviour       : chr [1:83935] "High_spent_Small_value_payments" "Low_spent_Medium_value_payments" "Low_spent_Small_value_payments" "High_spent_
 $ Monthly_Balance         : num [1:83935] 312 331 223 341 340 ...
 $ Credit_Score            : chr [1:83935] "Good" "Good" "Good" "Good" ...
 $ credit_history_age      : num [1:83935] 265 267 268 269 270 271 319 320 321 322 ...
 - attr(*, "na.action")= 'omit' Named int [1:16065] 2 8 20 41 43 46 57 73 78 88 ...
  ..- attr(*, "names")= chr [1:16065] "2" "8" "20" "41" ...
```

Chrome File Edit View History Bookmarks Profiles Tab Window Help Wed 31 Jul 7:15 PM

colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=LSofghXMtVqJ

Credit Score Brackets : Data Science and Machine Learning

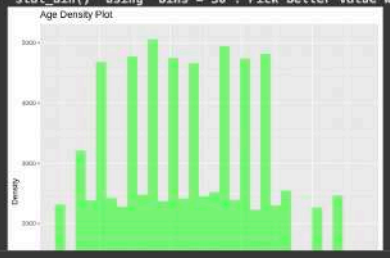
File Edit View Insert Runtime Tools Help All changes saved

Comment Share a

+ Code + Text RAM Disk Gemini

```
[39] library(ggplot2)

ggplot(df1, aes(x = Age)) +
  geom_histogram(fill = "Green", alpha = 0.5) +
  labs(title = "Age Density Plot", x = "Age", y = "Density")

ggplot(df1, aes(x = Annual_Income)) +
  geom_density(fill = "lightgreen", alpha = 0.5) +
  labs(title = "Annual Income Density Plot", x = "Annual Income", y = "Density")

ggplot(df1, aes(x = Monthly_Inhand_Salary)) +
  geom_density(fill = "lightcoral", alpha = 0.5) +
  labs(title = "Monthly Balance Density Plot", x = "Monthly Balance", y = "Density")

ggplot(df1, aes(x = Amount_invested_monthly)) +
  geom_density(fill = "lightgoldenrod", alpha = 0.5) +
  labs(title = "Monthly Investments Density Plot", x = "Monthly Investments", y = "Density")
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Chrome File Edit View History Bookmarks Profiles Tab Window Help Wed 31 Jul 7:16 PM

colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=LSofghXMtVqJ

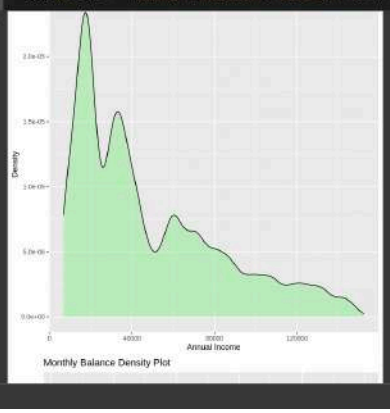
Credit Score Brackets : Data Science and Machine Learning

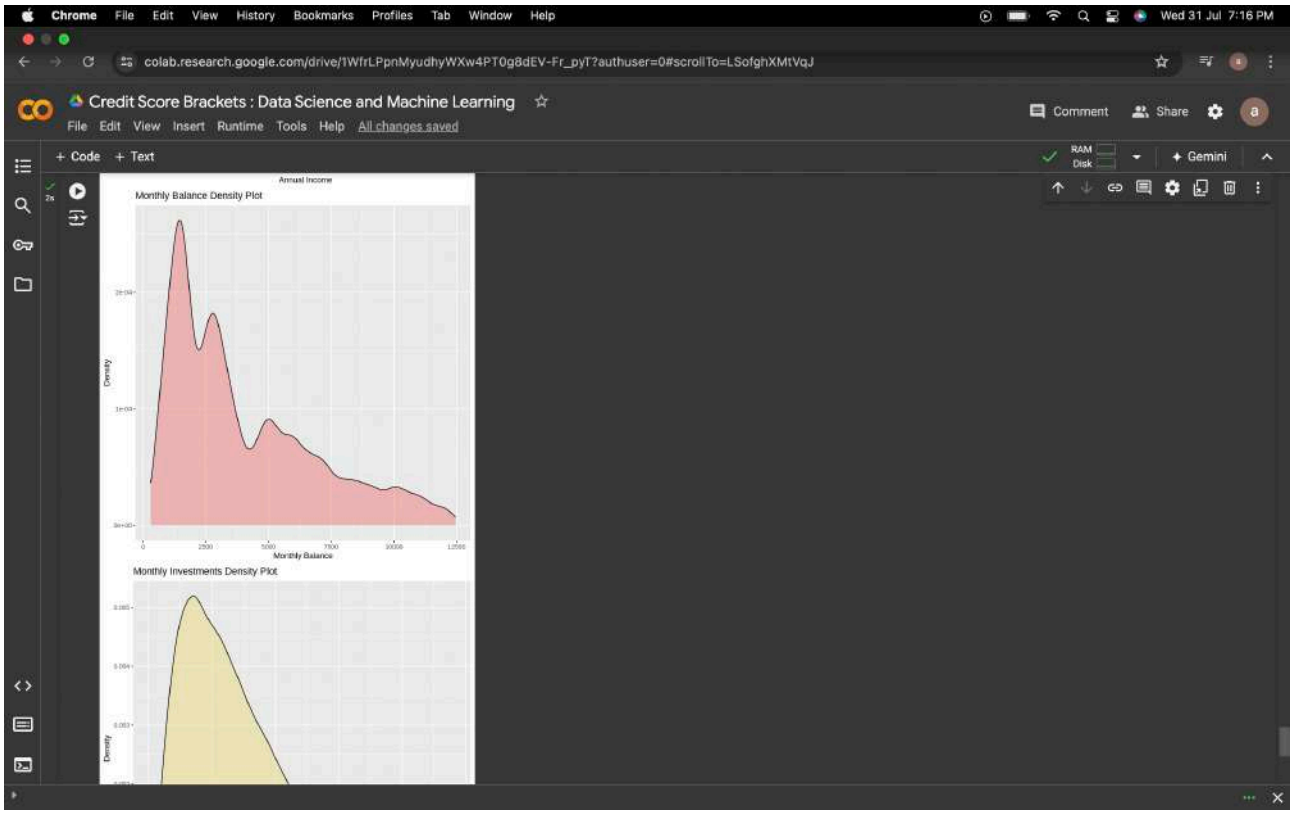
File Edit View Insert Runtime Tools Help All changes saved

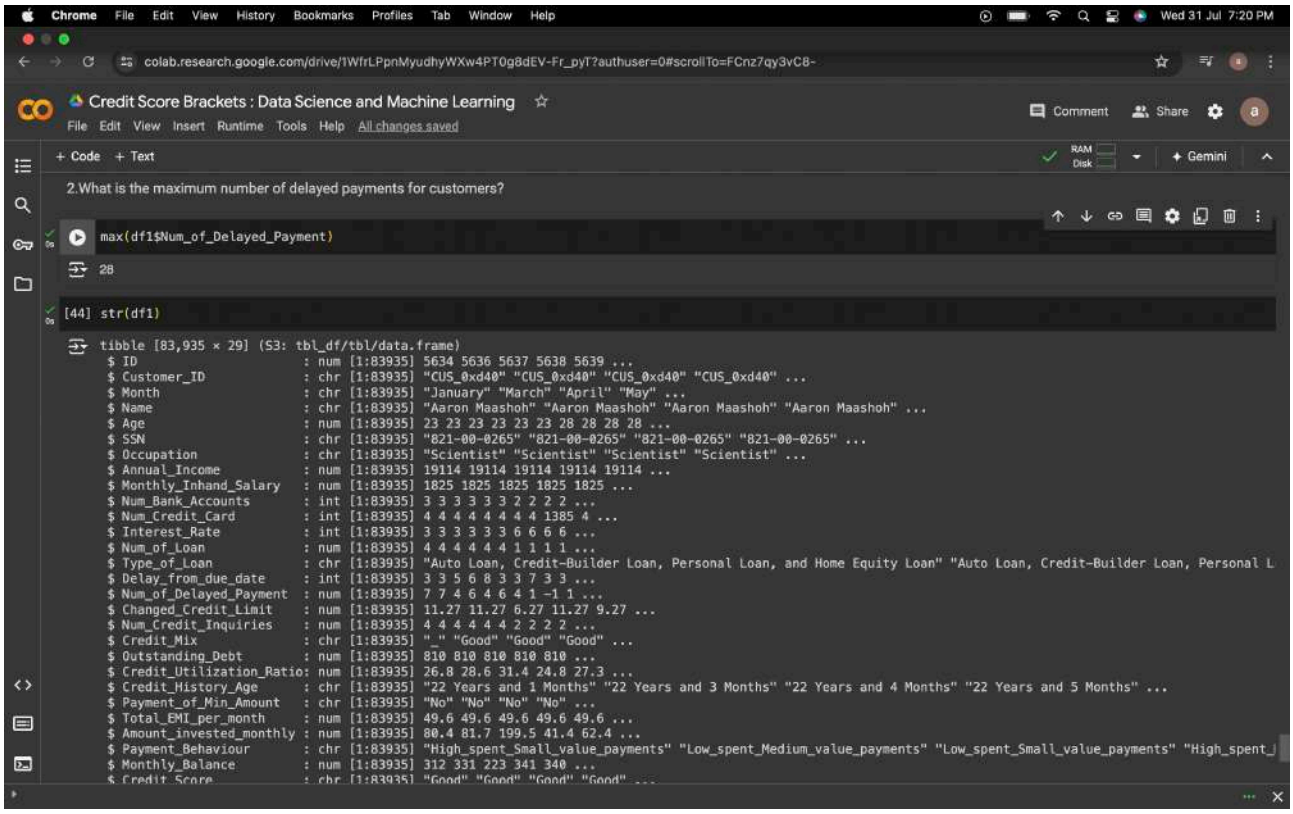
Comment Share a

+ Code + Text RAM Disk Gemini

```
  labs(title = "Age Density Plot", x = "Age", y = "Density")

ggplot(df1, aes(x = Annual_Income)) +
  geom_density(fill = "lightgreen", alpha = 0.5) +
  labs(title = "Annual Income Density Plot", x = "Annual Income", y = "Density")

ggplot(df1, aes(x = Monthly_Inhand_Salary)) +
  geom_density(fill = "lightcoral", alpha = 0.5) +
  labs(title = "Monthly Balance Density Plot", x = "Monthly Balance", y = "Density")

ggplot(df1, aes(x = Amount_invested_monthly)) +
  geom_density(fill = "lightgoldenrod", alpha = 0.5) +
  labs(title = "Monthly Investments Density Plot", x = "Monthly Investments", y = "Density")
```




Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 7:16 PM
colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=LSofghXMtVqJ
Credit Score Brackets : Data Science and Machine Learning
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Gemini
Annual Income
Monthly Balance Density Plot
Density
Monthly Balance
Monthly Investments Density Plot
Density


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 7:20 PM
colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=FCnz7qy3vC8-
Credit Score Brackets : Data Science and Machine Learning
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Gemini
2.What is the maximum number of delayed payments for customers?
max(df1$Num_of_Delayed_Payment)
28
[44] str(df1)
tibble [83,935 × 29] (S3: tbl_df/tbl/data.frame)
$ ID : num [1:83935] 5634 5636 5637 5638 5639 ...
$ Customer_ID : chr [1:83935] "CUS_0xd40" "CUS_0xd40" "CUS_0xd40" "CUS_0xd40" ...
$ Month : chr [1:83935] "January" "March" "April" "May" ...
$ Name : chr [1:83935] "Aaron Maashoh" "Aaron Maashoh" "Aaron Maashoh" "Aaron Maashoh" ...
$ Age : num [1:83935] 23 23 23 23 23 23 28 28 28 28 ...
$ SSN : chr [1:83935] "821-00-0265" "821-00-0265" "821-00-0265" "821-00-0265" ...
$ Occupation : chr [1:83935] "Scientist" "Scientist" "Scientist" "Scientist" ...
$ Annual_Income : num [1:83935] 19114 19114 19114 19114 19114 ...
$ Monthly_Inhand_Salary : num [1:83935] 1825 1825 1825 1825 1825 ...
$ Num_Bank_Accounts : int [1:83935] 3 3 3 3 3 3 2 2 2 2 ...
$ Num_Credit_Card : int [1:83935] 4 4 4 4 4 4 4 4 1385 4 ...
$ Interest_Rate : int [1:83935] 3 3 3 3 3 3 6 6 6 6 ...
$ Num_of_Loan : num [1:83935] 4 4 4 4 4 4 1 1 1 1 ...
$ Type_of_Loan : chr [1:83935] "Auto Loan, Credit-Builder Loan, Personal Loan, and Home Equity Loan" "Auto Loan, Credit-Builder Loan, Personal L
$ Delay_from_due_date : int [1:83935] 3 5 6 8 3 3 7 3 3 ...
$ Num_of_Delayed_Payment : num [1:83935] 7 7 4 6 4 6 4 1 -1 1 ...
$ Changed_Credit_Limit : num [1:83935] 11.27 11.27 6.27 11.27 9.27 ...
$ Num_Credit_Inquiries : num [1:83935] 4 4 4 4 4 4 2 2 2 2 ...
$ Credit_Mix : chr [1:83935] "_" "Good" "Good" "Good" ...
$ Outstanding_Debt : num [1:83935] 810 810 810 810 810 ...
$ Credit_Utilization_Ratio: num [1:83935] 26.8 28.6 31.4 24.8 27.3 ...
$ Credit_History_Age : chr [1:83935] "22 Years and 1 Months" "22 Years and 3 Months" "22 Years and 4 Months" "22 Years and 5 Months" ...
$ Payment_of_Min_Amount : chr [1:83935] "No" "No" "No" "No" ...
$ Total_EMI_per_month : num [1:83935] 49.6 49.6 49.6 49.6 49.6 ...
$ Amount_invested_monthly : num [1:83935] 80.4 81.7 199.5 41.4 62.4 ...
$ Payment_Behaviour : chr [1:83935] "High_spent_Small_value_payments" "Low_spent_Medium_value_payments" "Low_spent_Small_value_payments" "High_spent_
$ Monthly_Balance : num [1:83935] 312 331 223 341 340 ...
$ Credit_Score : chr [1:83935] "Good" "Good" "Good" "Good" ...

Chrome File Edit View History Bookmarks Profiles Tab Window Help Wed 31 Jul 7:21 PM

colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=aOFjcrCvvUlf

Credit Score Brackets : Data Science and Machine Learning

File Edit View Insert Runtime Tools Help All changes saved

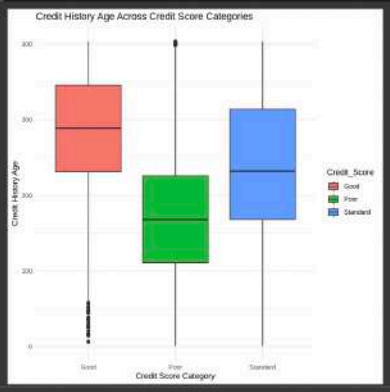
Comment Share a

+ Code + Text RAM Disk Gemini

```
$ Total_EMI_per_month     : num [1:83935] 49.6 49.6 49.6 49.6 49.6 ...
$ Amount_invested_monthly : num [1:83935] 80.4 81.7 199.5 41.4 62.4 ...
$ Payment_Behaviour       : chr [1:83935] "High_spent_Small_value_payments" "Low_spent_Medium_value_payments" "Low_spent_Small_value_payments" "High_spent_
$ Monthly_Balance         : num [1:83935] 312 331 223 341 340 ...
$ Credit_Score            : chr [1:83935] "Good" "Good" "Good" "Good" ...
$ credit_history_age      : num [1:83935] 265 267 268 269 270 271 319 320 321 322 ...
- attr(*, "na.action")= 'omit' Named int [1:16065] 2 8 20 41 43 46 57 73 78 88 ...
 ..- attr(*, "names")= chr [1:16065] "2" "8" "20" "41" ...
```

[44]

```
ggplot(df1, aes(x = Credit_Score, y = credit_history_age, fill = Credit_Score)) +
  geom_boxplot() +
  labs(title = "Credit History Age Across Credit Score Categories", x = "Credit Score Category", y = "Credit History Age") +
  theme_minimal()
```



Chrome File Edit View History Bookmarks Profiles Tab Window Help Wed 31 Jul 7:21 PM

colab.research.google.com/drive/1WfrLPpnMyudhyWXw4PT0g8dEV-Fr_pyT?authuser=0#scrollTo=aOFjcrCvvUlf

Credit Score Brackets : Data Science and Machine Learning

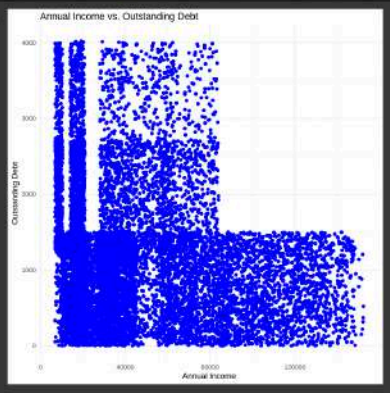
File Edit View Insert Runtime Tools Help

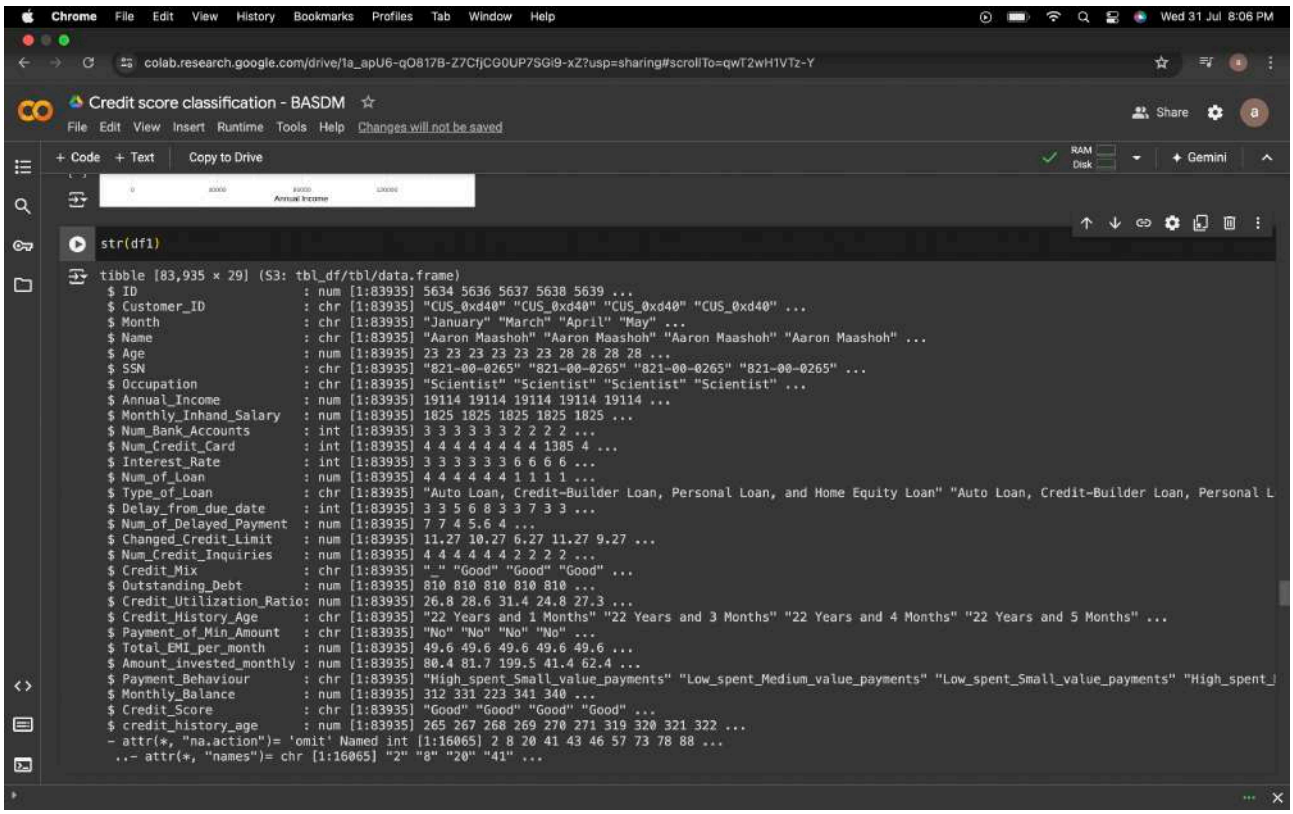
Comment Share a

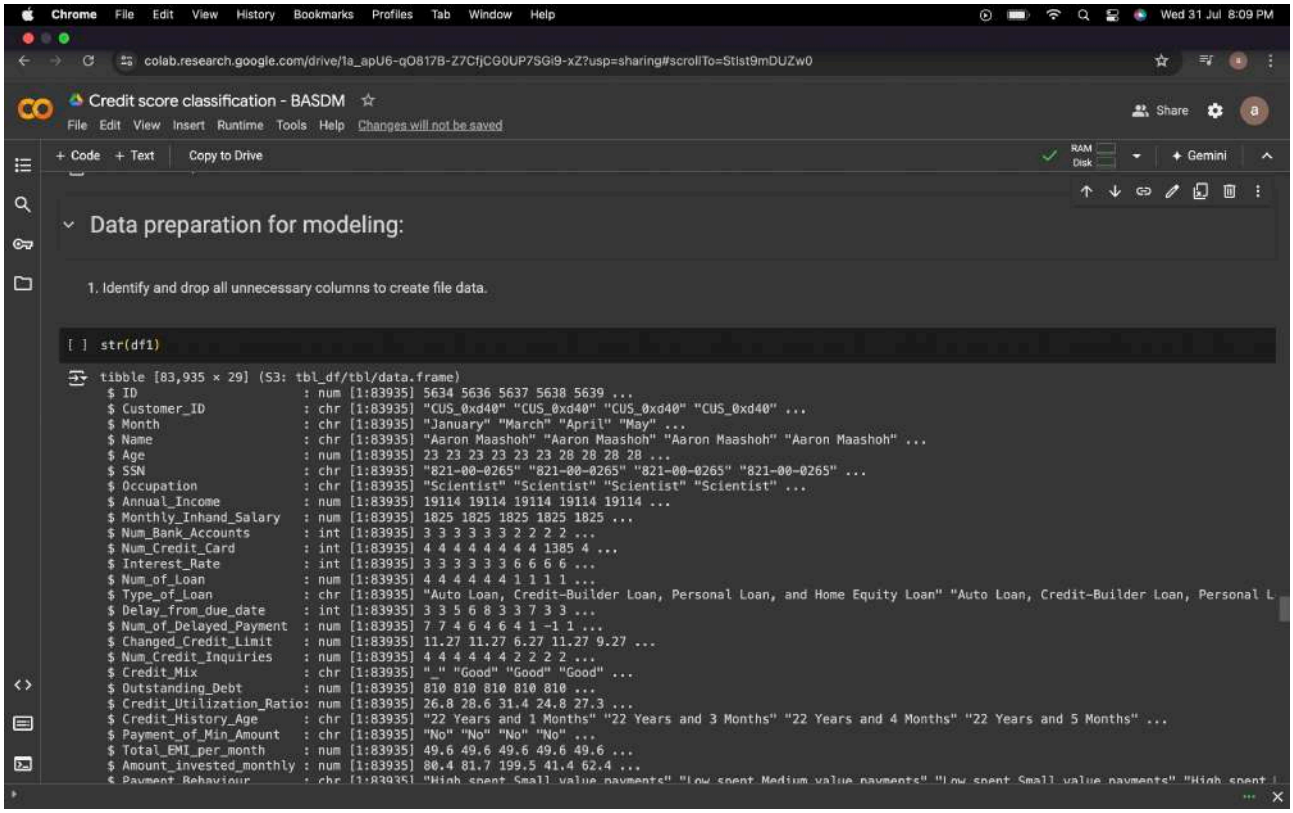
+ Code + Text RAM Disk Gemini

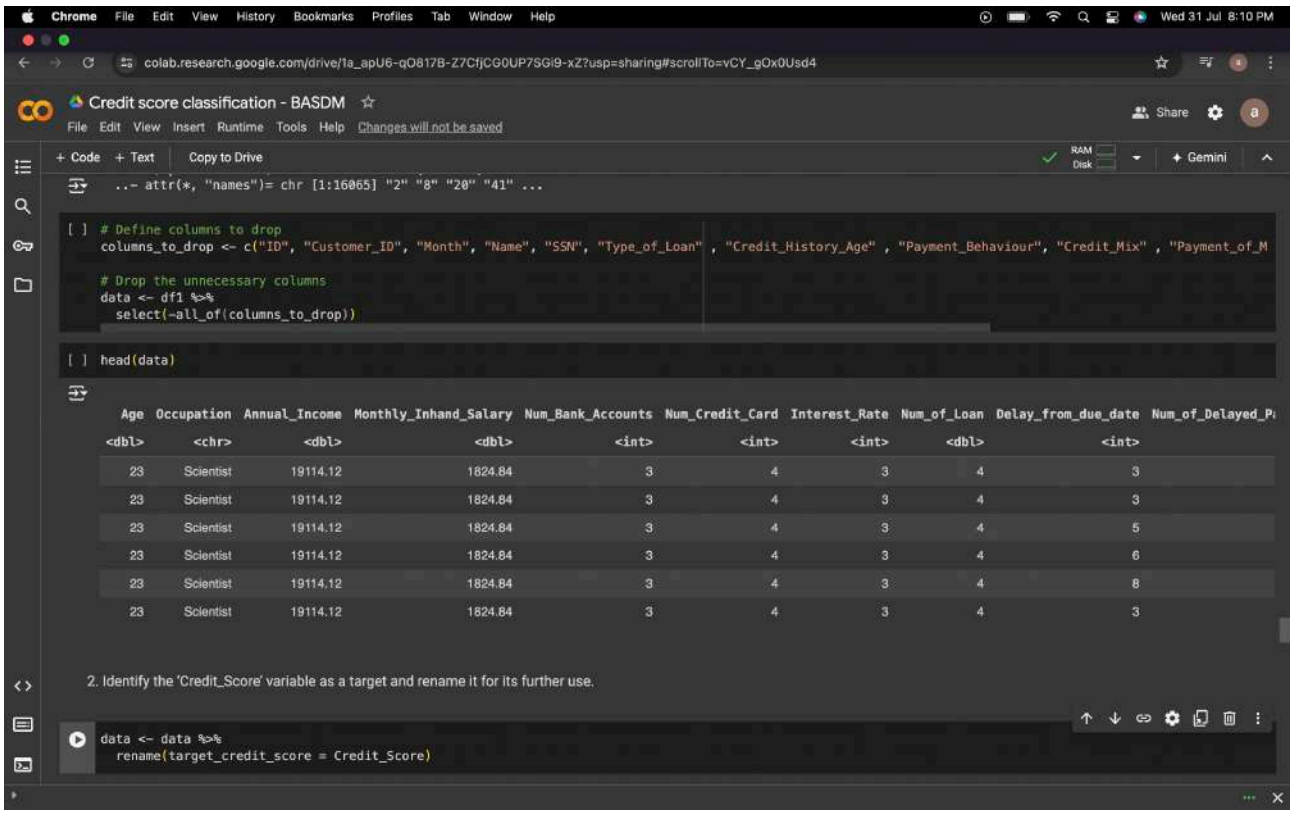
[46]

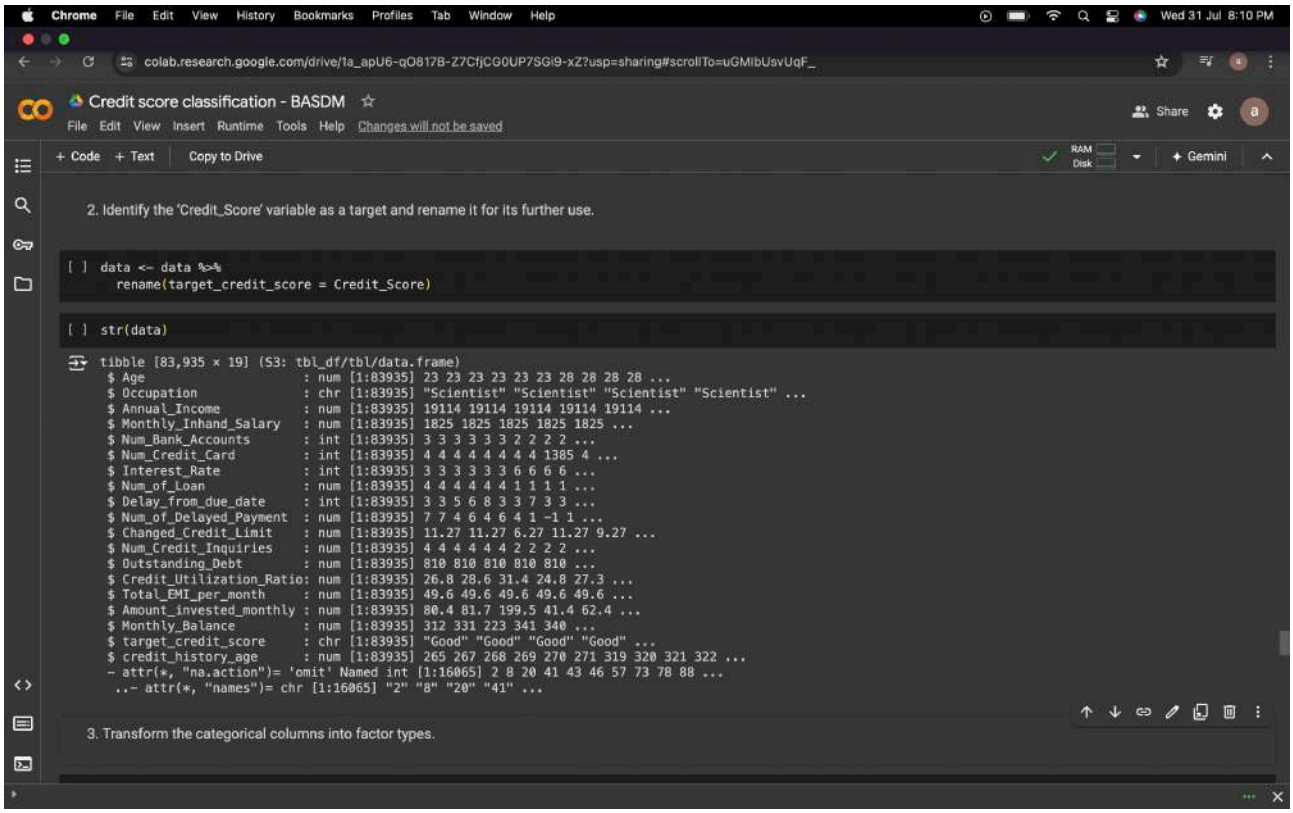
```
ggplot(df1, aes(x = Annual_Income, y = Outstanding_Debt)) +
    geom_point(color = "blue") +
    labs(title = "Annual Income vs. Outstanding Debt",
         x = "Annual Income",
         y = "Outstanding Debt") +
    theme_minimal()
```

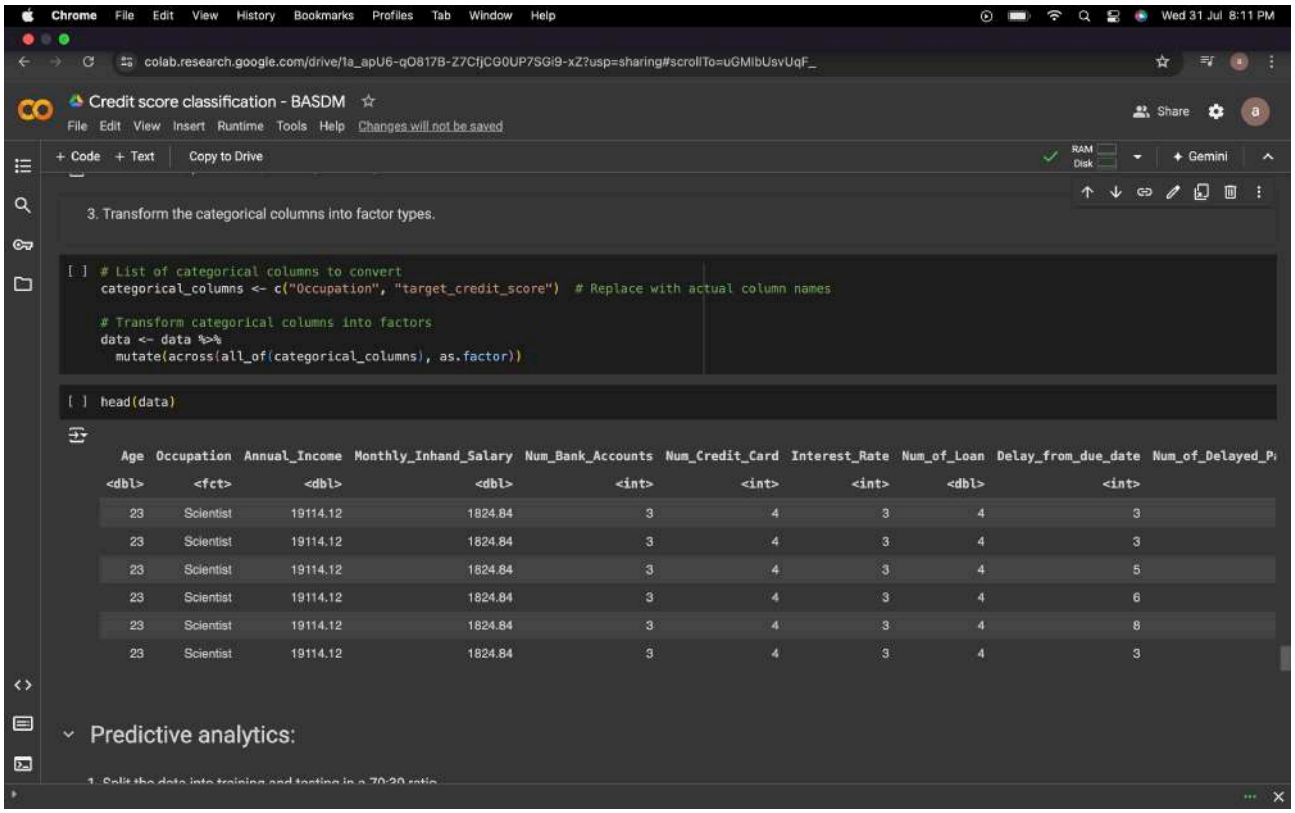


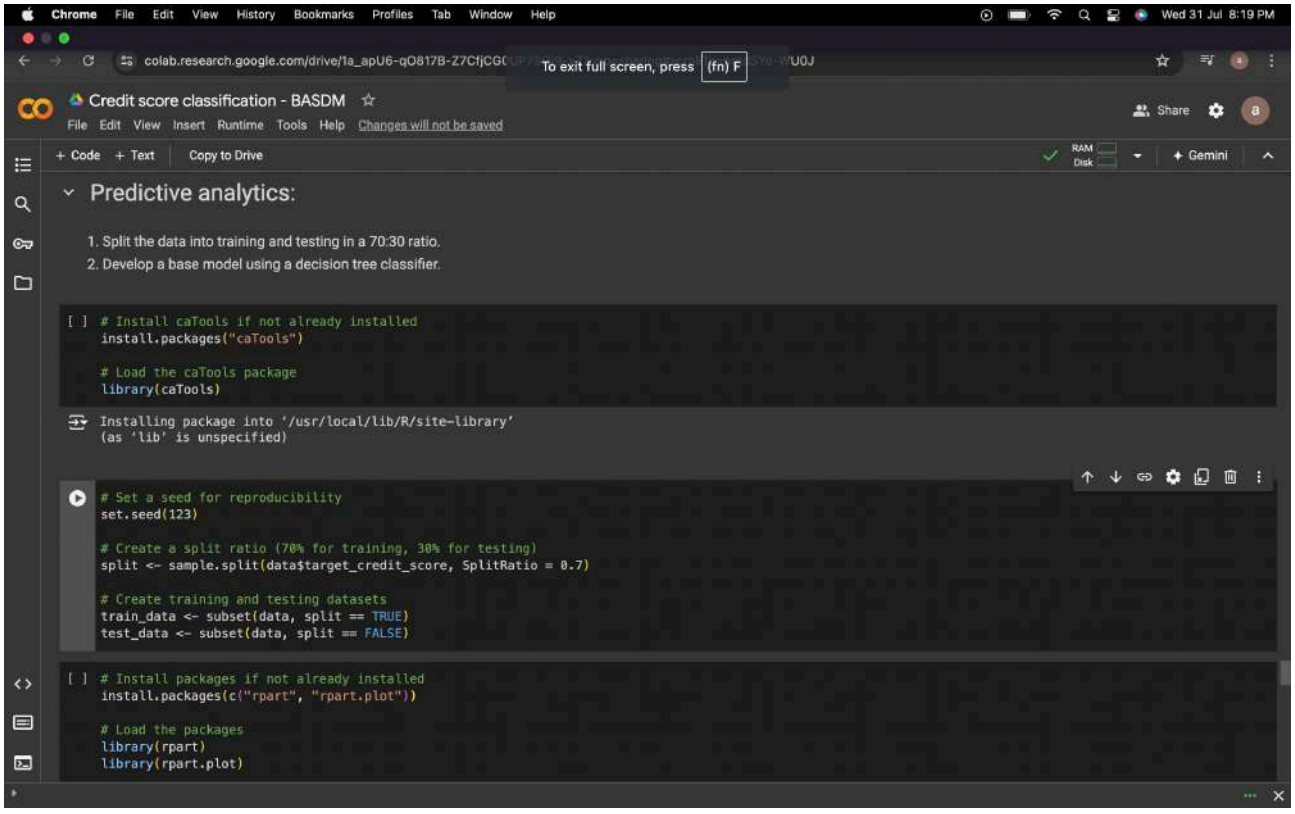

Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 8:06 PM
colab.research.google.com/drive/1a_apU6-qO817B-Z7CfjCG0UP7SGi9-xZ?usp=sharing#scrollTo=qwT2wH1VTz-Y
Credit score classification - BASDM
File Edit View Insert Runtime Tools Help Changes will not be saved
Share
+ Code + Text Copy to Drive
RAM Disk
Gemini
str(df1)
tibble [83,935 × 29] (S3: tbl_df/tbl/data.frame)
$ ID : num [1:83935] 5634 5636 5637 5638 5639 ...
$ Customer_ID : chr [1:83935] "CUS_0xd40" "CUS_0xd40" "CUS_0xd40" "CUS_0xd40" ...
$ Month : chr [1:83935] "January" "March" "April" "May" ...
$ Name : chr [1:83935] "Aaron Maashoh" "Aaron Maashoh" "Aaron Maashoh" "Aaron Maashoh" ...
$ Age : num [1:83935] 23 23 23 23 23 23 28 28 28 28 ...
$ SSN : chr [1:83935] "821-00-0265" "821-00-0265" "821-00-0265" "821-00-0265" ...
$ Occupation : chr [1:83935] "Scientist" "Scientist" "Scientist" "Scientist" ...
$ Annual_Income : num [1:83935] 19114 19114 19114 19114 19114 ...
$ Monthly_Inhand_Salary : num [1:83935] 1825 1825 1825 1825 1825 ...
$ Num_Bank_Accounts : int [1:83935] 3 3 3 3 3 3 2 2 2 2 ...
$ Num_Credit_Card : int [1:83935] 4 4 4 4 4 4 4 4 1385 4 ...
$ Interest_Rate : int [1:83935] 3 3 3 3 3 3 6 6 6 6 ...
$ Num_of_Loan : num [1:83935] 4 4 4 4 4 4 1 1 1 1 ...
$ Type_of_Loan : chr [1:83935] "Auto Loan, Credit-Builder Loan, Personal Loan, and Home Equity Loan" "Auto Loan, Credit-Builder Loan, Personal L
$ Delay_from_due_date : int [1:83935] 3 3 5 6 8 3 3 7 3 3 ...
$ Num_of_Delayed_Payment : num [1:83935] 7 7 4 5.6 4 ...
$ Changed_Credit_Limit : num [1:83935] 11.27 10.27 6.27 11.27 9.27 ...
$ Num_Credit_Inquiries : num [1:83935] 4 4 4 4 4 4 2 2 2 2 ...
$ Credit_Mix : chr [1:83935] "_" "Good" "Good" "Good" ...
$ Outstanding_Debt : num [1:83935] 810 810 810 810 810 ...
$ Credit_Utilization_Ratio: num [1:83935] 26.8 28.6 31.4 24.8 27.3 ...
$ Credit_History_Age : chr [1:83935] "22 Years and 1 Months" "22 Years and 3 Months" "22 Years and 4 Months" "22 Years and 5 Months" ...
$ Payment_of_Min_Amount : chr [1:83935] "No" "No" "No" "No" ...
$ Total_EMI_per_month : num [1:83935] 49.6 49.6 49.6 49.6 49.6 ...
$ Amount_invested_monthly : num [1:83935] 80.4 81.7 199.5 41.4 62.4 ...
$ Payment_Behaviour : chr [1:83935] "High_spent_Small_value_payments" "Low_spent_Medium_value_payments" "Low_spent_Small_value_payments" "High_spent_
$ Monthly_Balance : num [1:83935] 312 331 223 341 340 ...
$ Credit_Score : chr [1:83935] "Good" "Good" "Good" "Good" ...
$ credit_history_age : num [1:83935] 265 267 268 269 270 271 319 320 321 322 ...
- attr(*, "na.action")= 'omit' Named int [1:16065] 2 8 20 41 43 46 57 73 78 88 ...
..- attr(*, "names")= chr [1:16065] "2" "8" "20" "41" ...


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 8:09 PM
colab.research.google.com/drive/1a_apU6-qO817B-Z7CfjCG0UP7SGi9-xZ?usp=sharing#scrollTo=Stist9mDUZw0
Credit score classification - BASDM
File Edit View Insert Runtime Tools Help Changes will not be saved
Share
+ Code + Text Copy to Drive
RAM Disk
Gemini
Data preparation for modeling:
1. Identify and drop all unnecessary columns to create file data.
[ ] str(df1)
tibble [83,935 × 29] (S3: tbl_df/tbl/data.frame)
$ ID : num [1:83935] 5634 5636 5637 5638 5639 ...
$ Customer_ID : chr [1:83935] "CUS_0xd40" "CUS_0xd40" "CUS_0xd40" "CUS_0xd40" ...
$ Month : chr [1:83935] "January" "March" "April" "May" ...
$ Name : chr [1:83935] "Aaron Maashoh" "Aaron Maashoh" "Aaron Maashoh" "Aaron Maashoh" ...
$ Age : num [1:83935] 23 23 23 23 23 23 28 28 28 28 ...
$ SSN : chr [1:83935] "821-00-0265" "821-00-0265" "821-00-0265" "821-00-0265" ...
$ Occupation : chr [1:83935] "Scientist" "Scientist" "Scientist" "Scientist" ...
$ Annual_Income : num [1:83935] 19114 19114 19114 19114 19114 ...
$ Monthly_Inhand_Salary : num [1:83935] 1825 1825 1825 1825 1825 ...
$ Num_Bank_Accounts : int [1:83935] 3 3 3 3 3 3 2 2 2 2 ...
$ Num_Credit_Card : int [1:83935] 4 4 4 4 4 4 4 4 1385 4 ...
$ Interest_Rate : int [1:83935] 3 3 3 3 3 3 6 6 6 6 ...
$ Num_of_Loan : num [1:83935] 4 4 4 4 4 4 1 1 1 1 ...
$ Type_of_Loan : chr [1:83935] "Auto Loan, Credit-Builder Loan, Personal Loan, and Home Equity Loan" "Auto Loan, Credit-Builder Loan, Personal L
$ Delay_from_due_date : int [1:83935] 3 3 5 6 8 3 3 7 3 3 ...
$ Num_of_Delayed_Payment : num [1:83935] 7 7 4 6 4 6 4 1 -1 1 ...
$ Changed_Credit_Limit : num [1:83935] 11.27 11.27 6.27 11.27 9.27 ...
$ Num_Credit_Inquiries : num [1:83935] 4 4 4 4 4 4 2 2 2 2 ...
$ Credit_Mix : chr [1:83935] "_" "Good" "Good" "Good" ...
$ Outstanding_Debt : num [1:83935] 810 810 810 810 810 ...
$ Credit_Utilization_Ratio: num [1:83935] 26.8 28.6 31.4 24.8 27.3 ...
$ Credit_History_Age : chr [1:83935] "22 Years and 1 Months" "22 Years and 3 Months" "22 Years and 4 Months" "22 Years and 5 Months" ...
$ Payment_of_Min_Amount : chr [1:83935] "No" "No" "No" "No" ...
$ Total_EMI_per_month : num [1:83935] 49.6 49.6 49.6 49.6 49.6 ...
$ Amount_invested_monthly : num [1:83935] 80.4 81.7 199.5 41.4 62.4 ...


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 8:10 PM
colab.research.google.com/drive/1a_apU6-qO817B-Z7CfjCG0UP7SGi9-xZ?usp=sharing#scrollTo=vCY_gOx0Usd4
Credit score classification - BASDM
File Edit View Insert Runtime Tools Help Changes will not be saved
Share
+ Code + Text Copy to Drive
RAM Disk
Gemini
..- attr(*, "names")= chr [1:16065] "2" "8" "20" "41" ...
# Define columns to drop
columns_to_drop <- c("ID", "Customer_ID", "Month", "Name", "SSN", "Type_of_Loan" , "Credit_History_Age" , "Payment_Behaviour", "Credit_Mix" , "Payment_of_M
# Drop the unnecessary columns
data <- df1 %>%
select(-all_of(columns_to_drop))
head(data)
Age Occupation Annual_Income Monthly_Inhand_Salary Num_Bank_Accounts Num_Credit_Card Interest_Rate Num_of_Loan Delay_from_due_date Num_of_Delayed_P
<dbl> <chr> <dbl> <dbl> <int> <int> <int> <dbl> <int>
23 Scientist 19114.12 1824.84 3 4 3 4 3
23 Scientist 19114.12 1824.84 3 4 3 4 3
23 Scientist 19114.12 1824.84 3 4 3 4 5
23 Scientist 19114.12 1824.84 3 4 3 4 6
23 Scientist 19114.12 1824.84 3 4 3 4 8
23 Scientist 19114.12 1824.84 3 4 3 4 3
2. Identify the 'Credit_Score' variable as a target and rename it for its further use.
data <- data %>%
rename(target_credit_score = Credit_Score)


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 8:10 PM
colab.research.google.com/drive/1a_apU6-qO817B-Z7CfjCG0UP7SGi9-xZ?usp=sharing#scrollTo=uGMlbUsvUqF_
Credit score classification - BASDM
File Edit View Insert Runtime Tools Help Changes will not be saved
Share
+ Code + Text Copy to Drive
RAM Disk
Gemini
2. Identify the 'Credit_Score' variable as a target and rename it for its further use.
data <- data %>%
rename(target_credit_score = Credit_Score)
str(data)
tibble [83,935 × 19] (S3: tbl_df/tbl/data.frame)
$ Age : num [1:83935] 23 23 23 23 23 28 28 28 28 ...
$ Occupation : chr [1:83935] "Scientist" "Scientist" "Scientist" "Scientist" ...
$ Annual_Income : num [1:83935] 19114 19114 19114 19114 19114 ...
$ Monthly_Inhand_Salary : num [1:83935] 1825 1825 1825 1825 1825 ...
$ Num_Bank_Accounts : int [1:83935] 3 3 3 3 3 2 2 2 2 ...
$ Num_Credit_Card : int [1:83935] 4 4 4 4 4 4 4 4 1385 4 ...
$ Interest_Rate : int [1:83935] 3 3 3 3 3 6 6 6 6 ...
$ Num_of_Loan : num [1:83935] 4 4 4 4 4 4 1 1 1 1 ...
$ Delay_from_due_date : int [1:83935] 3 3 5 6 8 3 3 7 3 3 ...
$ Num_of_Delayed_Payment : num [1:83935] 7 7 4 6 4 6 4 1 -1 1 ...
$ Changed_Credit_Limit : num [1:83935] 11.27 11.27 6.27 11.27 9.27 ...
$ Num_Credit_Inquiries : num [1:83935] 4 4 4 4 4 4 2 2 2 2 ...
$ Outstanding_Debt : num [1:83935] 810 810 810 810 810 ...
$ Credit_Utilization_Ratio: num [1:83935] 26.8 28.6 31.4 24.8 27.3 ...
$ Total_EMI_per_month : num [1:83935] 49.6 49.6 49.6 49.6 49.6 ...
$ Amount_invested_monthly : num [1:83935] 80.4 81.7 199.5 41.4 62.4 ...
$ Monthly_Balance : num [1:83935] 312 331 223 341 340 ...
$ target_credit_score : chr [1:83935] "Good" "Good" "Good" "Good" ...
$ credit_history_age : num [1:83935] 265 267 268 269 270 271 319 320 321 322 ...
- attr(*, "na.action")= 'omit' Named int [1:16065] 2 8 20 41 43 46 57 73 78 88 ...
..- attr(*, "names")= chr [1:16065] "2" "8" "20" "41" ...
3. Transform the categorical columns into factor types.


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 8:11 PM
colab.research.google.com/drive/1a_apU6-qO817B-Z7CfjCG0UP7SGi9-xZ?usp=sharing#scrollTo=uGMibUsvUqF_
Credit score classification - BASDM
File Edit View Insert Runtime Tools Help Changes will not be saved
Share
+ Code + Text Copy to Drive
RAM Disk
Gemini
3. Transform the categorical columns into factor types.
# List of categorical columns to convert
categorical_columns <- c("Occupation", "target_credit_score") # Replace with actual column names
# Transform categorical columns into factors
data <- data %>%
mutate(across(all_of(categorical_columns), as.factor))
head(data)
Age Occupation Annual_Income Monthly_Inhand_Salary Num_Bank_Accounts Num_Credit_Card Interest_Rate Num_of_Loan Delay_from_due_date Num_of_Delayed_P
<dbl> <fct> <dbl> <dbl> <int> <int> <int> <dbl> <int>
23 Scientist 19114.12 1824.84 3 4 3 4 3
23 Scientist 19114.12 1824.84 3 4 3 4 3
23 Scientist 19114.12 1824.84 3 4 3 4 5
23 Scientist 19114.12 1824.84 3 4 3 4 6
23 Scientist 19114.12 1824.84 3 4 3 4 8
23 Scientist 19114.12 1824.84 3 4 3 4 3
Predictive analytics:


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 8:19 PM
To exit full screen, press (fn) F
Credit score classification - BASDM
File Edit View Insert Runtime Tools Help Changes will not be saved
Share
+ Code + Text Copy to Drive
RAM Disk
Gemini
Predictive analytics:
1. Split the data into training and testing in a 70:30 ratio.
2. Develop a base model using a decision tree classifier.
# Install caTools if not already installed
install.packages("caTools")
# Load the caTools package
library(caTools)
Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)
# Set a seed for reproducibility
set.seed(123)
# Create a split ratio (70% for training, 30% for testing)
split <- sample.split(data$target_credit_score, SplitRatio = 0.7)
# Create training and testing datasets
train_data <- subset(data, split == TRUE)
test_data <- subset(data, split == FALSE)
# Install packages if not already installed
install.packages(c("rpart", "rpart.plot"))
# Load the packages
library(rpart)
library(rpart.plot)

Chrome File Edit View History Bookmarks Profiles Tab Window Help Wed 31 Jul 8:21 PM

colab.research.google.com/drive/1a_apU6-qO817B-Z7CfjCG0UP7SGi9-xZ?usp=sharing#scrollTo=7OdhkbEUWimx

Credit score classification - BASDM

File Edit View Insert Runtime Tools Help Changes will not be saved

Share

+ Code + Text Copy to Drive

RAM Disk

Gemini

```r
# Create a decision tree model
decision_tree_model <- rpart(target_credit_score ~ ., data = train_data, method = "class")

# View the model summary
summary(decision_tree_model)
```

```
Call:
rpart(formula = target_credit_score ~ ., data = train_data, method = "class")
  n= 58754

          CP nsplit rel error    xerror       xstd
1 0.17774190      0 1.0000000 1.0000000 0.004435436
2 0.05232451      1 0.8222581 0.8222581 0.004320167
3 0.01119143      2 0.7699336 0.7699336 0.004261672
4 0.01040252      4 0.7475507 0.7516237 0.004238416
5 0.01000000      9 0.6934282 0.7198840 0.004194576

Variable importance
      Outstanding_Debt          Interest_Rate           Num_of_Loan
                    25                     16                    12
   Delay_from_due_date     credit_history_age   Num_Credit_Inquiries
                    11                     11                    10
      Num_Credit_Card Num_of_Delayed_Payment   Changed_Credit_Limit
                     7                      5                     2
     Num_Bank_Accounts
                     1

Node number 1: 58754 observations,    complexity param=0.1777419
  predicted class=Standard  expected loss=0.4638493  P(node) =1
    class counts: 10636 16617 31501
   probabilities: 0.181 0.283 0.536
  left son=2 (16695 obs) right son=3 (42059 obs)
  Primary splits:
      Outstanding_Debt     < 1500.005 to the right, improve=4426.298, (0 missing)
      Interest_Rate        < 20.5     to the right, improve=3495.451, (0 missing)
      Num_Credit_Inquiries < 5.75     to the left,  improve=2649.592, (0 missing)
      Delay_from_due_date  < 30.5     to the right, improve=2357.879, (0 missing)
      Num_of_Loan          < 4.5      to the left,  improve=2116.517, (0 missing)
```

Chrome File Edit View History Bookmarks Profiles Tab Window Help Wed 31 Jul 8:22 PM

colab.research.google.com/drive/1a_apU6-qO817B-Z7CfjCG0UP7SGi9-xZ?usp=sharing#scrollTo=Zrh0CbFWWIyw

Credit score classification - BASDM

File Edit View Insert Runtime Tools Help Changes will not be saved

Share

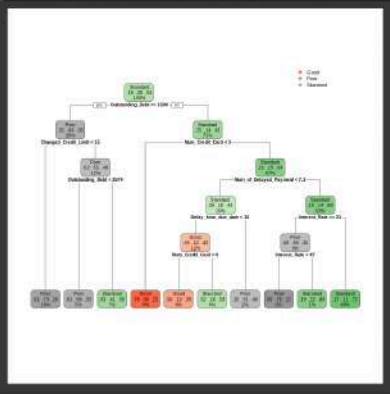
+ Code + Text Copy to Drive

RAM Disk

Gemini

```
      Num_Credit_Inquiries < 5.5      to the right, improve= 66.33288, (0 missing)
  Surrogate splits:
      Outstanding_Debt      < 2699.89  to the left,  agree=0.724, adj=0.351, (0 split)
      credit_history_age    < 63.5     to the right, agree=0.632, adj=0.136, (0 split)
      Num_of_Loan           < 7.5      to the left,  agree=0.594, adj=0.046, (0 split)
      Monthly_Inhand_Salary < 477.965  to the right, agree=0.577, adj=0.006, (0 split)
      Interest_Rate         < 10.5     to the right, agree=0.576, adj=0.004, (0 split)

Node number 3: 42059 observations,    complexity param=0.05232451
```

```r
# Plot the decision tree
rpart.plot(decision_tree_model, extra = 104)
```



Chrome File Edit View History Bookmarks Profiles Tab Window Help Wed 31 Jul 8:23 PM

colab.research.google.com/drive/1a_apU6-qO817B-Z7CfjCG0UP7SGi9-xZ?usp=sharing#scrollTo=Zrh0CbFWWIyw

Credit score classification - BASDM

File Edit View Insert Runtime Tools Help Changes will not be saved

Share

+ Code + Text Copy to Drive

RAM Disk

Gemini

3. Optimize the model by pruning. (Hint: Try to find the optimized control hyperparameters (minsplit, minibucket, cp, and others) and develop a pruned tree using them.
4. Evaluate the model using the confusion matrix and share your interpretation.

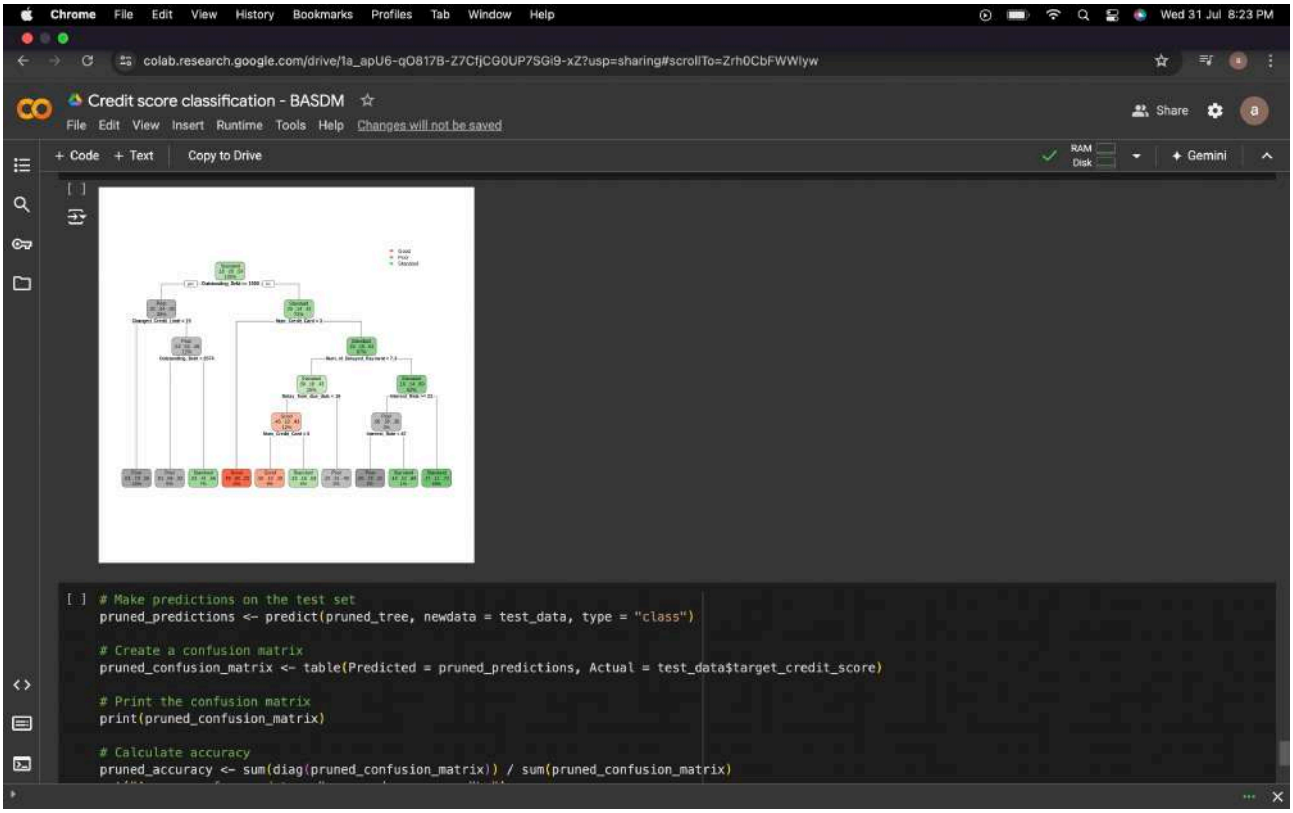
```
[ ] # Print the complexity parameter table
    print(decision_tree_model$cptable)

    # Find the optimal cp value (e.g., cp with minimum xerror)
    optimal_cp <- decision_tree_model$cptable[which.min(decision_tree_model$cptable[,"xerror"]),"CP"]
```

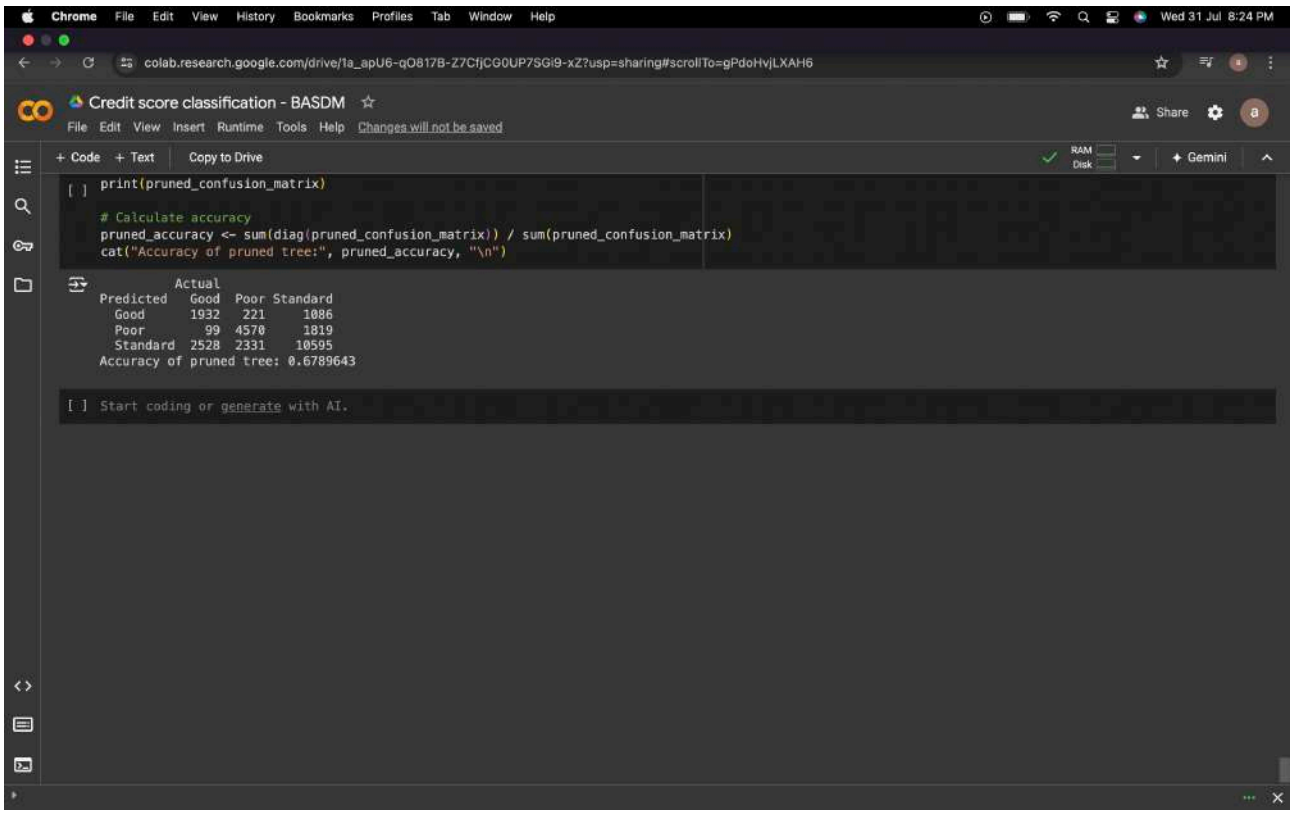
```
          CP nsplit rel error    xerror       xstd
1 0.17774190      0 1.0000000 1.0000000 0.004435436
2 0.05232451      1 0.8222581 0.8222581 0.004320167
3 0.01119143      2 0.7699336 0.7699336 0.004261672
4 0.01040252      4 0.7475507 0.7516237 0.004238416
5 0.01000000      9 0.6934282 0.7198840 0.004194576
```

```
[ ] optimal_cp
```

```
0.01
```

```
[ ] # Prune the tree using the optimal cp value
    pruned_tree <- prune(decision_tree_model, cp = optimal_cp)

    # Plot the pruned tree
    rpart.plot(pruned_tree, extra = 104)
```




Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed 31 Jul 8:24 PM
colab.research.google.com/drive/1a_apU6-qO817B-Z7CfjCG0UP7SGi9-xZ?usp=sharing#scrollTo=gPdoHvjLXAH6
Credit score classification - BASDM
File Edit View Insert Runtime Tools Help Changes will not be saved
Share
+ Code + Text Copy to Drive
RAM
Disk
Gemini
[ ] print(pruned_confusion_matrix)
# Calculate accuracy
pruned_accuracy <- sum(diag(pruned_confusion_matrix)) / sum(pruned_confusion_matrix)
cat("Accuracy of pruned tree:", pruned_accuracy, "\n")
Actual
Predicted Good Poor Standard
Good 1932 221 1086
Poor 99 4570 1819
Standard 2528 2331 10595
Accuracy of pruned tree: 0.6789643
[ ] Start coding or generate with AI.