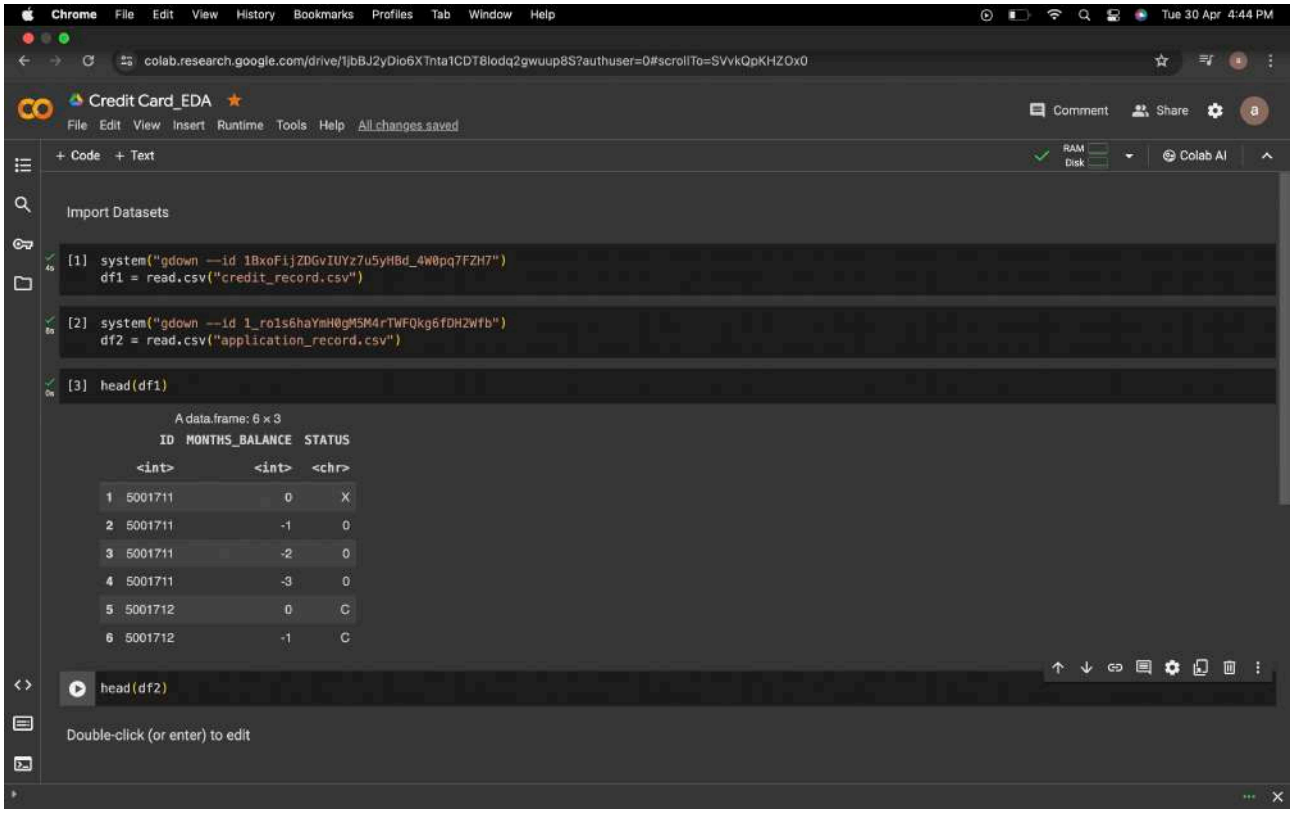
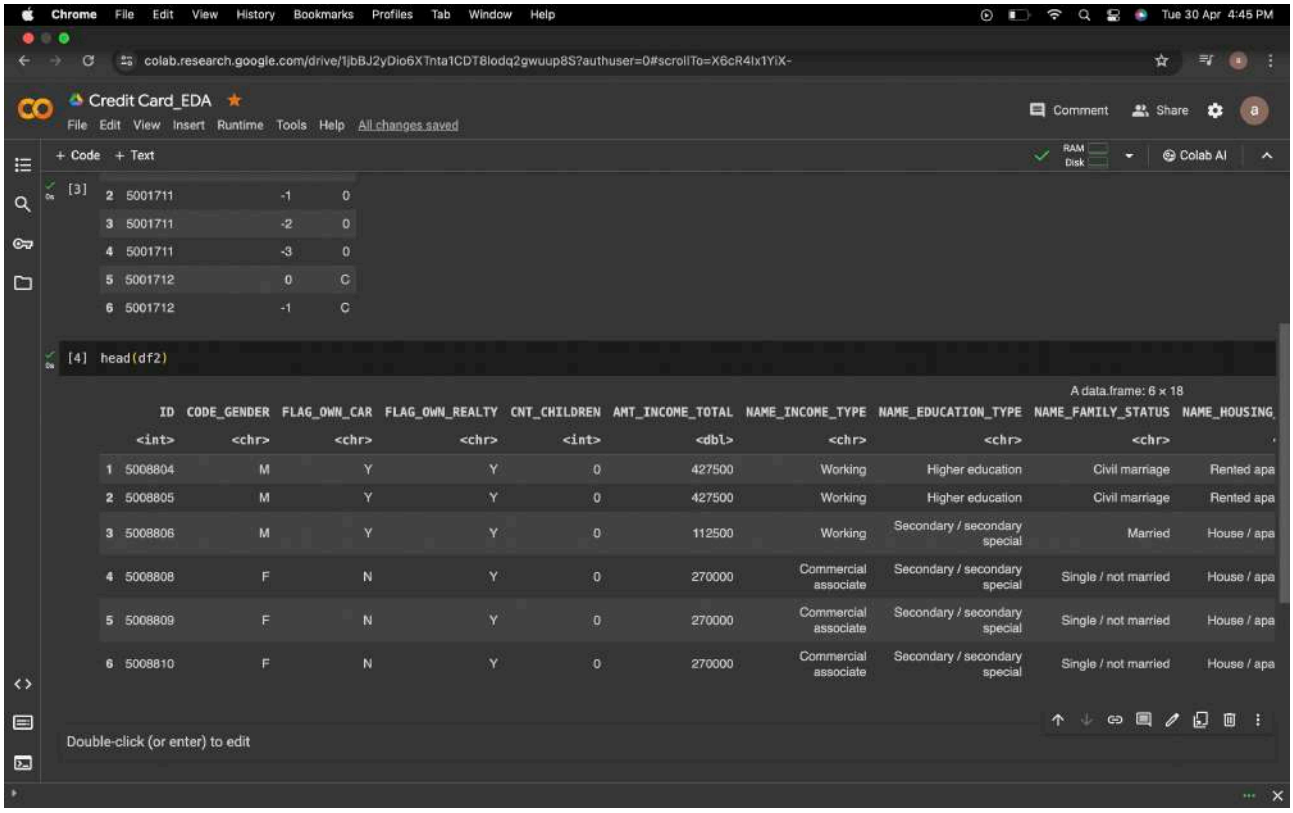
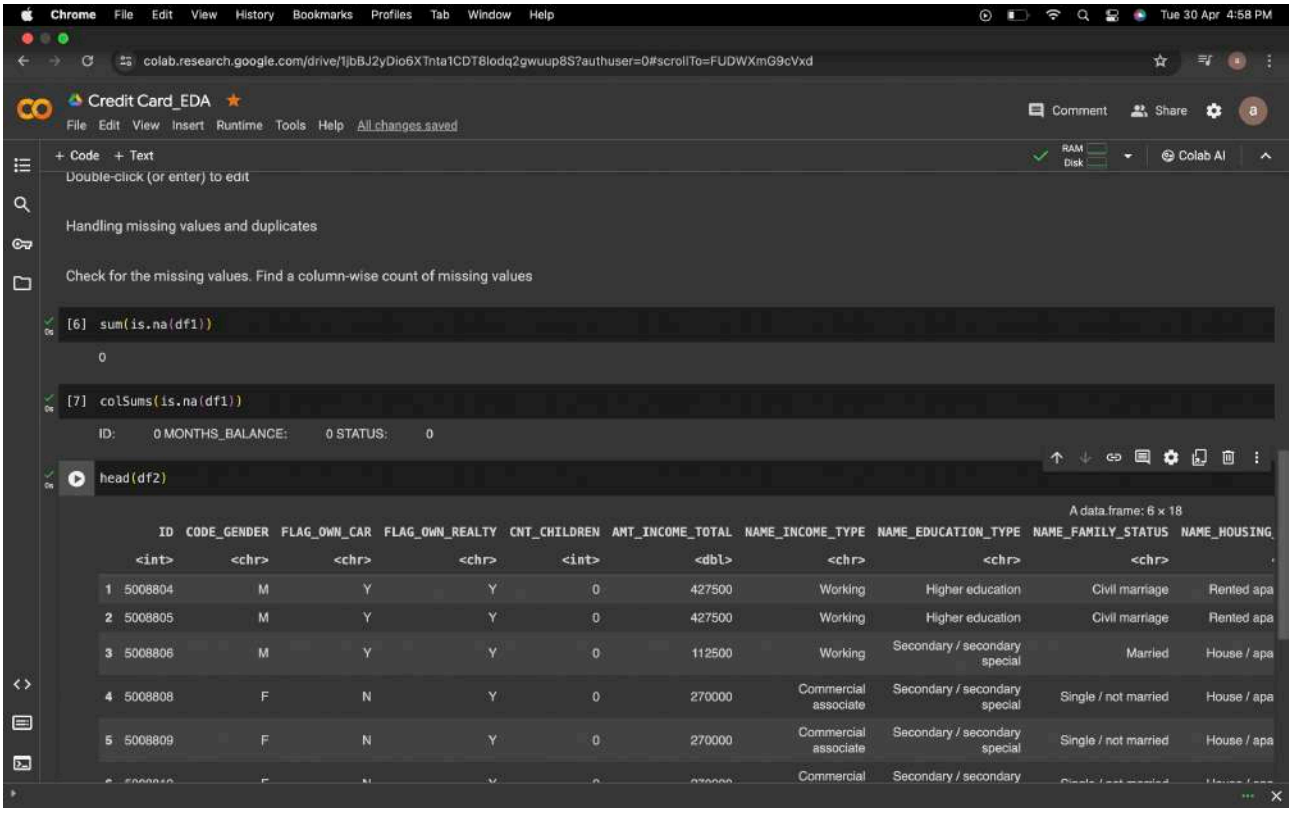
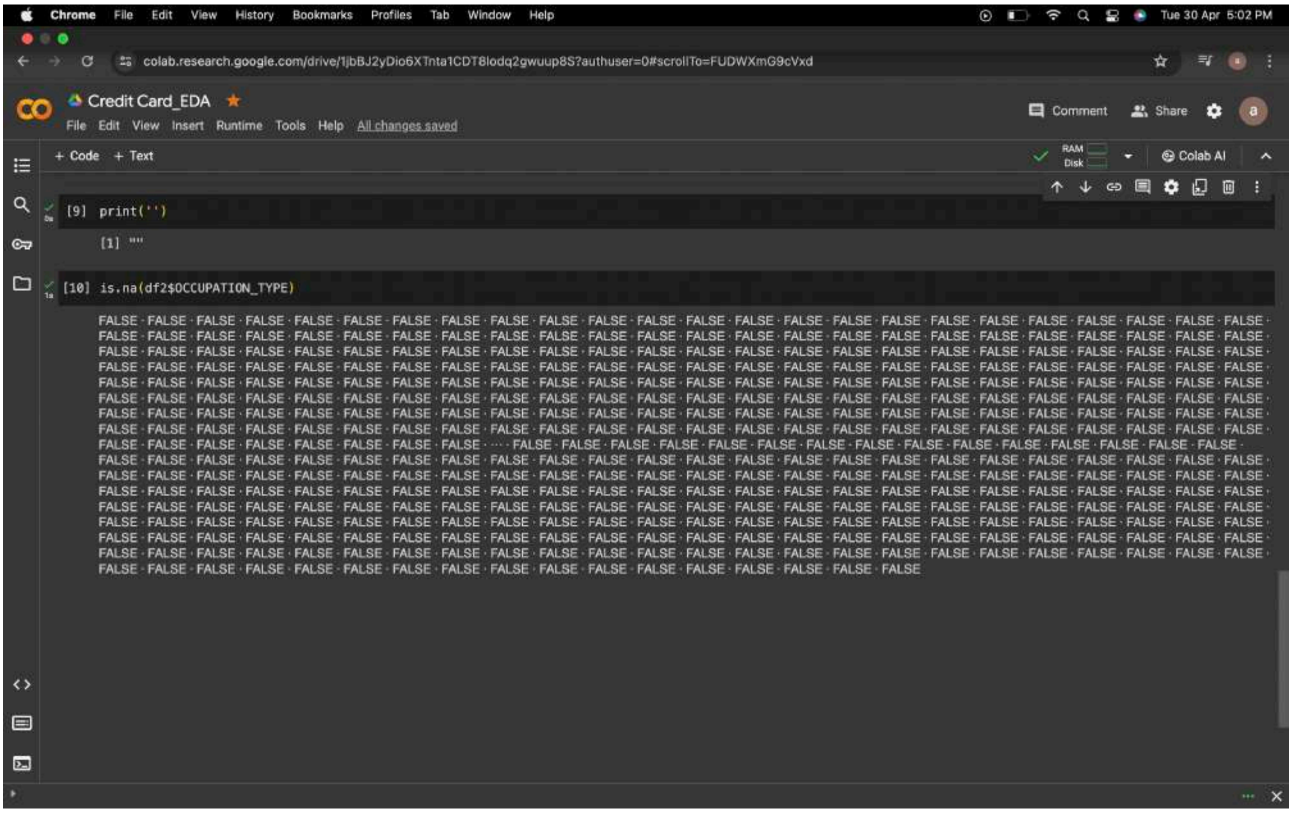
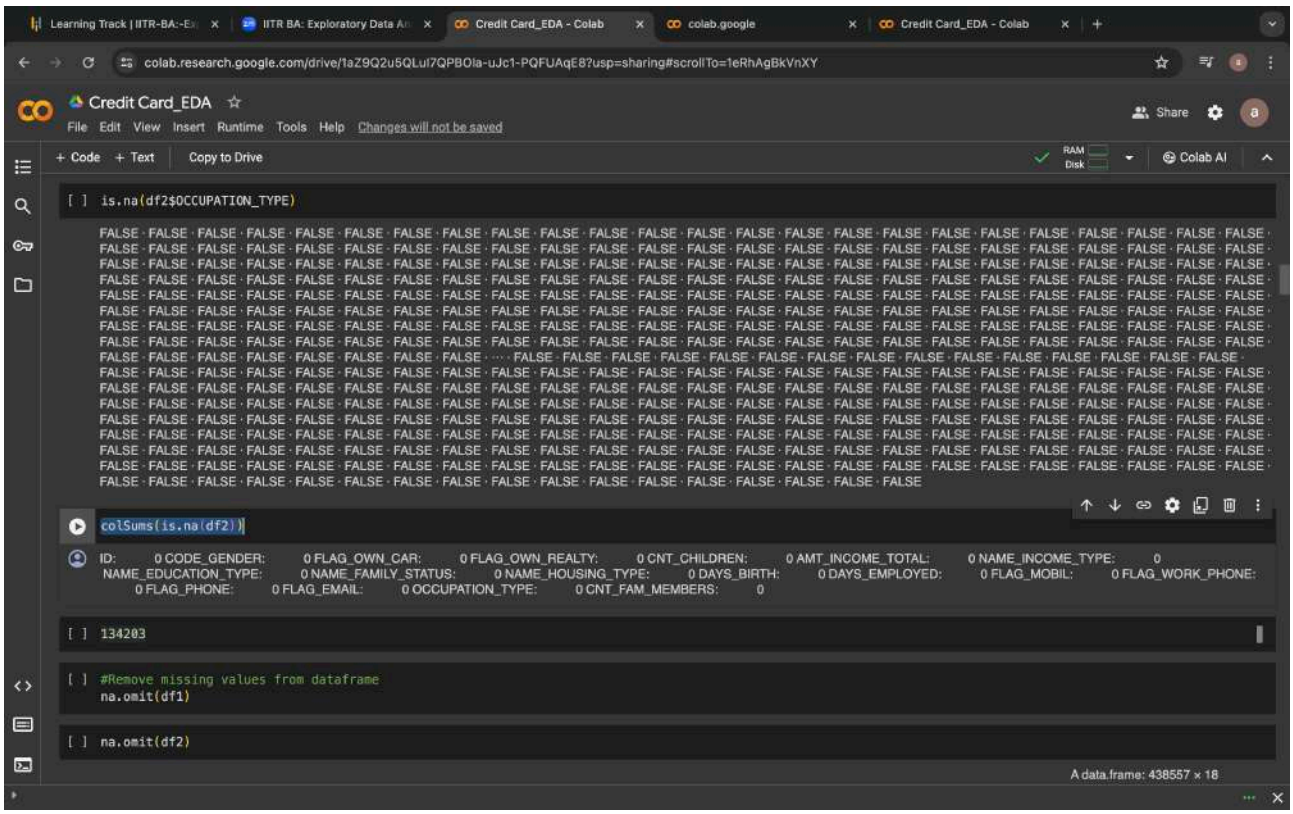
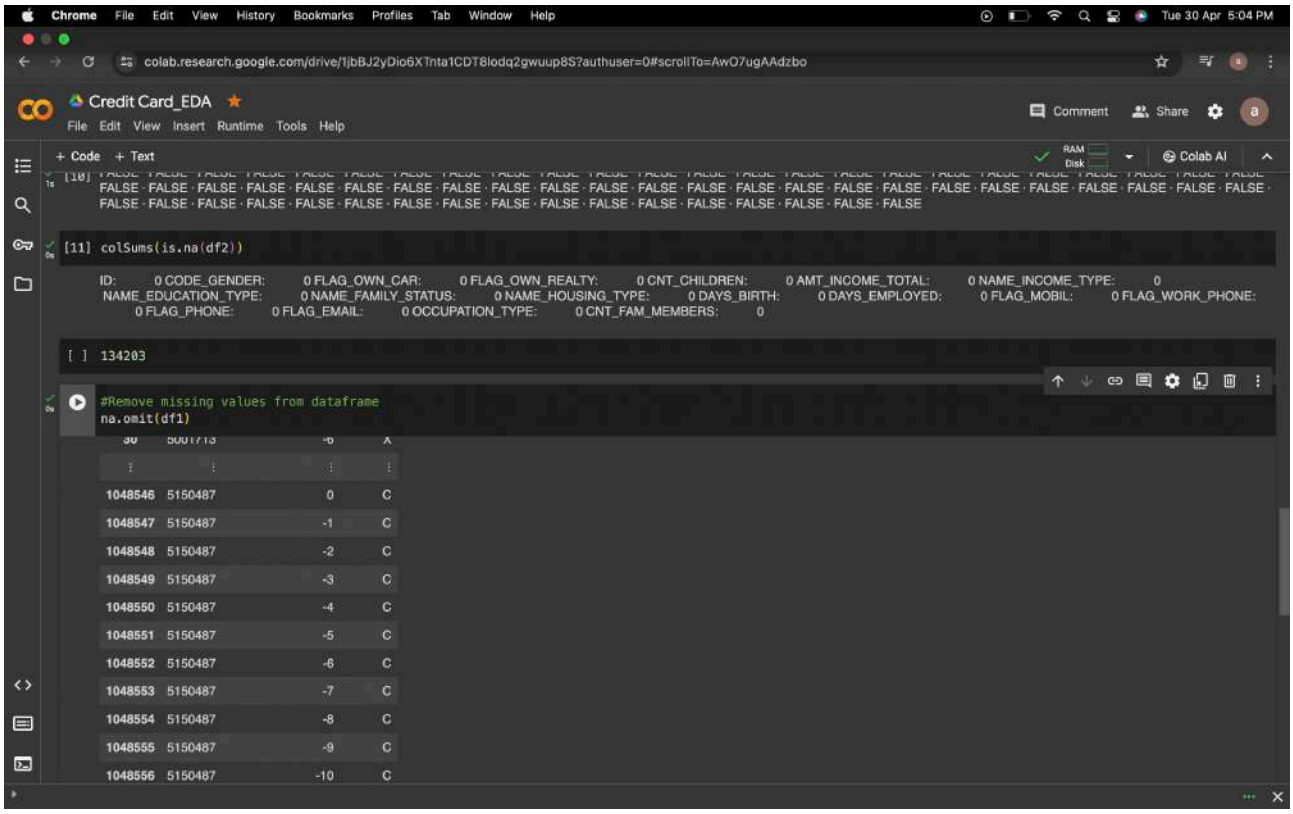

Chrome File Edit View History Bookmarks Profiles Tab Window Help
Tue 30 Apr 4:44 PM
colab.research.google.com/drive/1jbBJ2yDio6XTnta1CDT8lodq2gwuup8S?authuser=0#scrollTo=SVvkQpKHZOx0
Credit Card_EDA
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Colab AI
Import Datasets
[1] system("gdown --id 1BxoFijZDGvIUYz7u5yHBd_4W0pq7FZH7")
df1 = read.csv("credit_record.csv")
[2] system("gdown --id 1_ro1s6haYmH0gM5M4rTWFQkg6fDH2Wfb")
df2 = read.csv("application_record.csv")
[3] head(df1)
A data.frame: 6 × 3
ID MONTHS_BALANCE STATUS
<int> <int> <chr>
1 5001711 0 X
2 5001711 -1 0
3 5001711 -2 0
4 5001711 -3 0
5 5001712 0 C
6 5001712 -1 C
head(df2)
Double-click (or enter) to edit

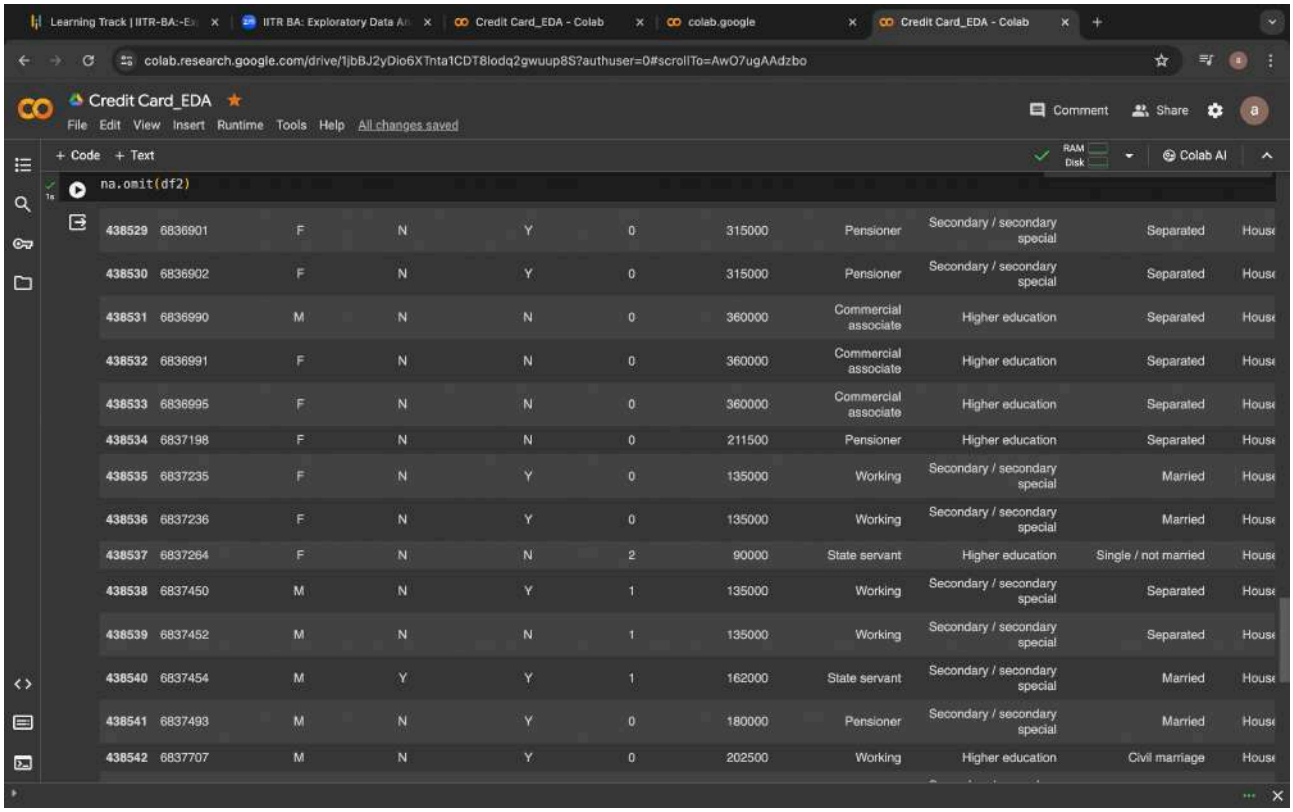

Chrome File Edit View History Bookmarks Profiles Tab Window Help
Tue 30 Apr 4:45 PM
colab.research.google.com/drive/1jbBJ2yDio6XTnta1CDT8lodq2gwuup8S?authuser=0#scrollTo=X6cR4Ix1YiX-
Credit Card_EDA
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Colab AI
[3]
2 5001711 -1 0
3 5001711 -2 0
4 5001711 -3 0
5 5001712 0 C
6 5001712 -1 C
[4] head(df2)
A data.frame: 6 × 18
ID CODE_GENDER FLAG_OWN_CAR FLAG_OWN_REALTY CNT_CHILDREN AMT_INCOME_TOTAL NAME_INCOME_TYPE NAME_EDUCATION_TYPE NAME_FAMILY_STATUS NAME_HOUSING
<int> <chr> <chr> <chr> <int> <dbl> <chr> <chr> <chr>
1 5008804 M Y Y 0 427500 Working Higher education Civil marriage Rented apa
2 5008805 M Y Y 0 427500 Working Higher education Civil marriage Rented apa
3 5008806 M Y Y 0 112500 Working Secondary / secondary special Married House / apa
4 5008808 F N Y 0 270000 Commercial associate Secondary / secondary special Single / not married House / apa
5 5008809 F N Y 0 270000 Commercial associate Secondary / secondary special Single / not married House / apa
6 5008810 F N Y 0 270000 Commercial associate Secondary / secondary special Single / not married House / apa
Double-click (or enter) to edit


Credit Card_EDA
Double-click (or enter) to edit
Handling missing values and duplicates
Check for the missing values. Find a column-wise count of missing values
[6] sum(is.na(df1))
0
[7] colSums(is.na(df1))
ID: 0 MONTHS_BALANCE: 0 STATUS: 0
head(df2)
A data.frame: 6 × 18
ID CODE_GENDER FLAG_OWN_CAR FLAG_OWN_REALTY CNT_CHILDREN AMT_INCOME_TOTAL NAME_INCOME_TYPE NAME_EDUCATION_TYPE NAME_FAMILY_STATUS NAME_HOUSING
1 5008804 M Y Y 0 427500 Working Higher education Civil marriage Rented apa
2 5008805 M Y Y 0 427500 Working Higher education Civil marriage Rented apa
3 5008806 M Y Y 0 112500 Working Secondary / secondary special Married House / apa
4 5008808 F N Y 0 270000 Commercial associate Secondary / secondary special Single / not married House / apa
5 5008809 F N Y 0 270000 Commercial associate Secondary / secondary special Single / not married House / apa

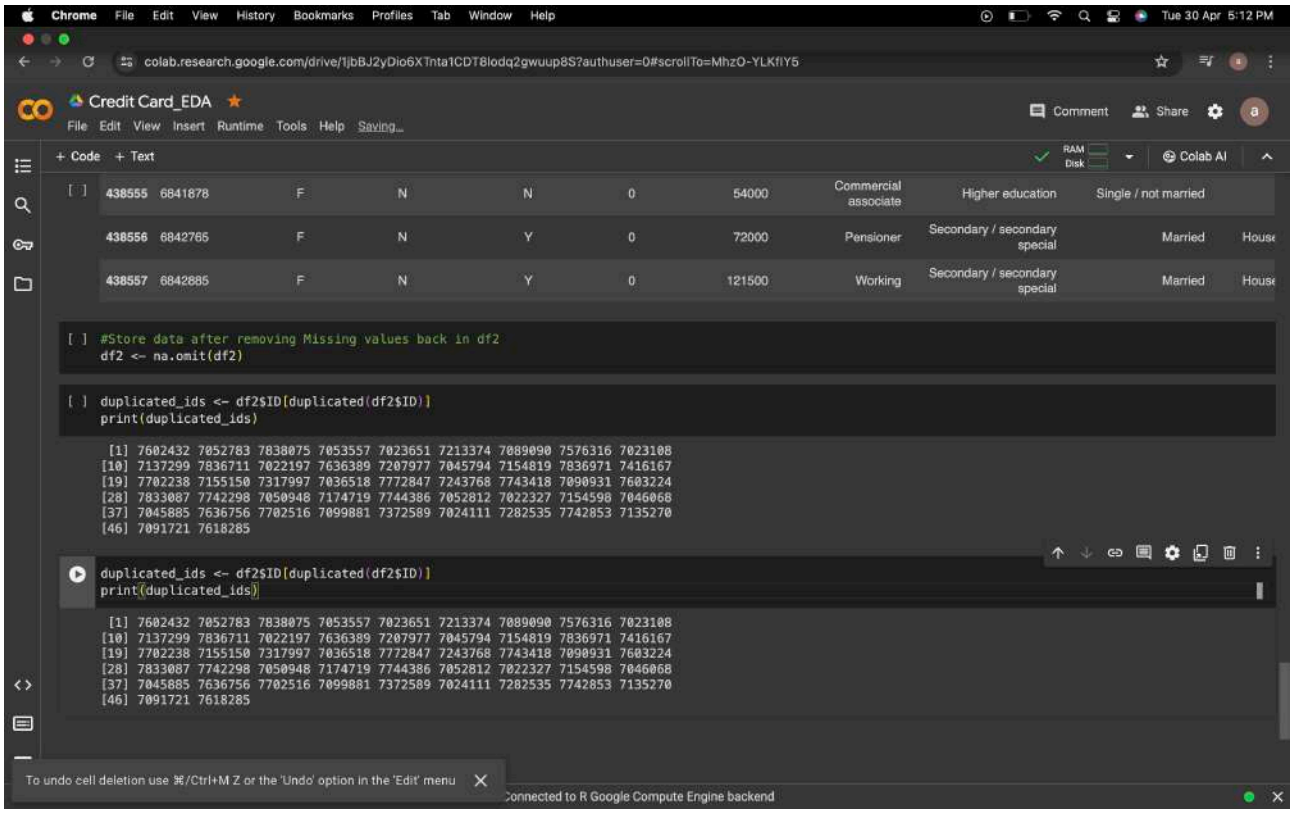

Credit Card_EDA
[9] print('')
[1] ""
[10] is.na(df2$OCCUPATION_TYPE)
FALSE · ...
FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · ···· FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE
FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE


Credit Card_EDA - Colab
colab.research.google.com/drive/1aZ9Q2u5QLul7QPBOla-uJc1-PQFUAqE8?usp=sharing#scrollTo=1eRhAgBkVnXY
Credit Card_EDA
File Edit View Insert Runtime Tools Help Changes will not be saved
Share
+ Code + Text Copy to Drive
Colab AI
[ ] is.na(df2$OCCUPATION_TYPE)
colSums(is.na(df2))
ID: 0 CODE_GENDER: 0 FLAG_OWN_CAR: 0 FLAG_OWN_REALTY: 0 CNT_CHILDREN: 0 AMT_INCOME_TOTAL: 0 NAME_INCOME_TYPE: 0
NAME_EDUCATION_TYPE: 0 NAME_FAMILY_STATUS: 0 NAME_HOUSING_TYPE: 0 DAYS_BIRTH: 0 DAYS_EMPLOYED: 0 FLAG_MOBIL: 0 FLAG_WORK_PHONE:
0 FLAG_PHONE: 0 FLAG_EMAIL: 0 OCCUPATION_TYPE: 0 CNT_FAM_MEMBERS: 0
[ ] 134203
[ ] #Remove missing values from dataframe
na.omit(df1)
[ ] na.omit(df2)
A data.frame: 438557 × 18


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Tue 30 Apr 5:04 PM
colab.research.google.com/drive/1jbBJ2yDio6XTnta1CDT8lodq2gwuup8S?authuser=0#scrollTo=AwO7ugAAdzbo
Credit Card_EDA
File Edit View Insert Runtime Tools Help
Comment
Share
+ Code + Text
Colab AI
[11] colSums(is.na(df2))
ID: 0 CODE_GENDER: 0 FLAG_OWN_CAR: 0 FLAG_OWN_REALTY: 0 CNT_CHILDREN: 0 AMT_INCOME_TOTAL: 0 NAME_INCOME_TYPE: 0
NAME_EDUCATION_TYPE: 0 NAME_FAMILY_STATUS: 0 NAME_HOUSING_TYPE: 0 DAYS_BIRTH: 0 DAYS_EMPLOYED: 0 FLAG_MOBIL: 0 FLAG_WORK_PHONE:
0 FLAG_PHONE: 0 FLAG_EMAIL: 0 OCCUPATION_TYPE: 0 CNT_FAM_MEMBERS: 0
[ ] 134203
#Remove missing values from dataframe
na.omit(df1)
1048546 5150487 0 C
1048547 5150487 -1 C
1048548 5150487 -2 C
1048549 5150487 -3 C
1048550 5150487 -4 C
1048551 5150487 -5 C
1048552 5150487 -6 C
1048553 5150487 -7 C
1048554 5150487 -8 C
1048555 5150487 -9 C
1048556 5150487 -10 C



438529	6836901	F	N	Y	0	315000	Pensioner	Secondary / secondary special	Separated	House
438530	6836902	F	N	Y	0	315000	Pensioner	Secondary / secondary special	Separated	House
438531	6836990	M	N	N	0	360000	Commercial associate	Higher education	Separated	House
438532	6836991	F	N	N	0	360000	Commercial associate	Higher education	Separated	House
438533	6836995	F	N	N	0	360000	Commercial associate	Higher education	Separated	House
438534	6837198	F	N	N	0	211500	Pensioner	Higher education	Separated	House
438535	6837235	F	N	Y	0	135000	Working	Secondary / secondary special	Married	House
438536	6837236	F	N	Y	0	135000	Working	Secondary / secondary special	Married	House
438537	6837264	F	N	N	2	90000	State servant	Higher education	Single / not married	House
438538	6837450	M	N	Y	1	135000	Working	Secondary / secondary special	Separated	House
438539	6837452	M	N	N	1	135000	Working	Secondary / secondary special	Separated	House
438540	6837454	M	Y	Y	1	162000	State servant	Secondary / secondary special	Married	House
438541	6837493	M	N	Y	0	180000	Pensioner	Secondary / secondary special	Married	House
438542	6837707	M	N	Y	0	202500	Working	Higher education	Civil marriage	House



438555	6841878	F	N	N	0	54000	Commercial associate	Higher education	Single / not married	
438556	6842765	F	N	Y	0	72000	Pensioner	Secondary / secondary special	Married	House
438557	6842885	F	N	Y	0	121500	Working	Secondary / secondary special	Married	House

```
#Store data after removing Missing values back in df2
df2 <- na.omit(df2)
```

```
duplicated_ids <- df2$ID[duplicated(df2$ID)]
print(duplicated_ids)
```

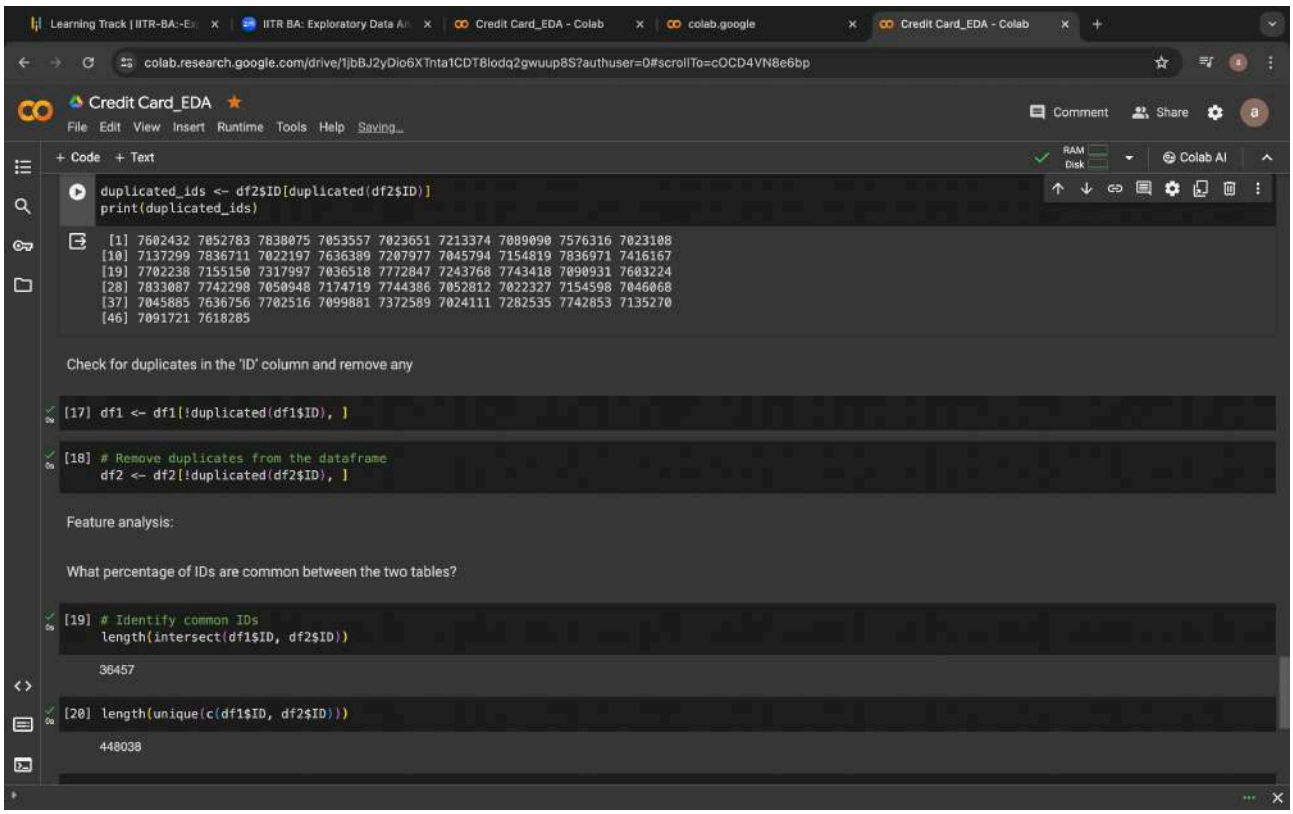
```
 [1] 7602432 7052783 7838075 7053557 7023651 7213374 7089090 7576316 7023108
[10] 7137299 7836711 7022197 7636389 7207977 7045794 7154819 7836971 7416167
[19] 7702238 7155150 7317997 7036518 7772847 7243768 7743418 7090931 7603224
[28] 7833087 7742298 7050948 7174719 7744386 7052812 7022327 7154598 7046068
[37] 7045885 7636756 7702516 7099881 7372589 7024111 7282535 7742853 7135270
[46] 7091721 7618285
```

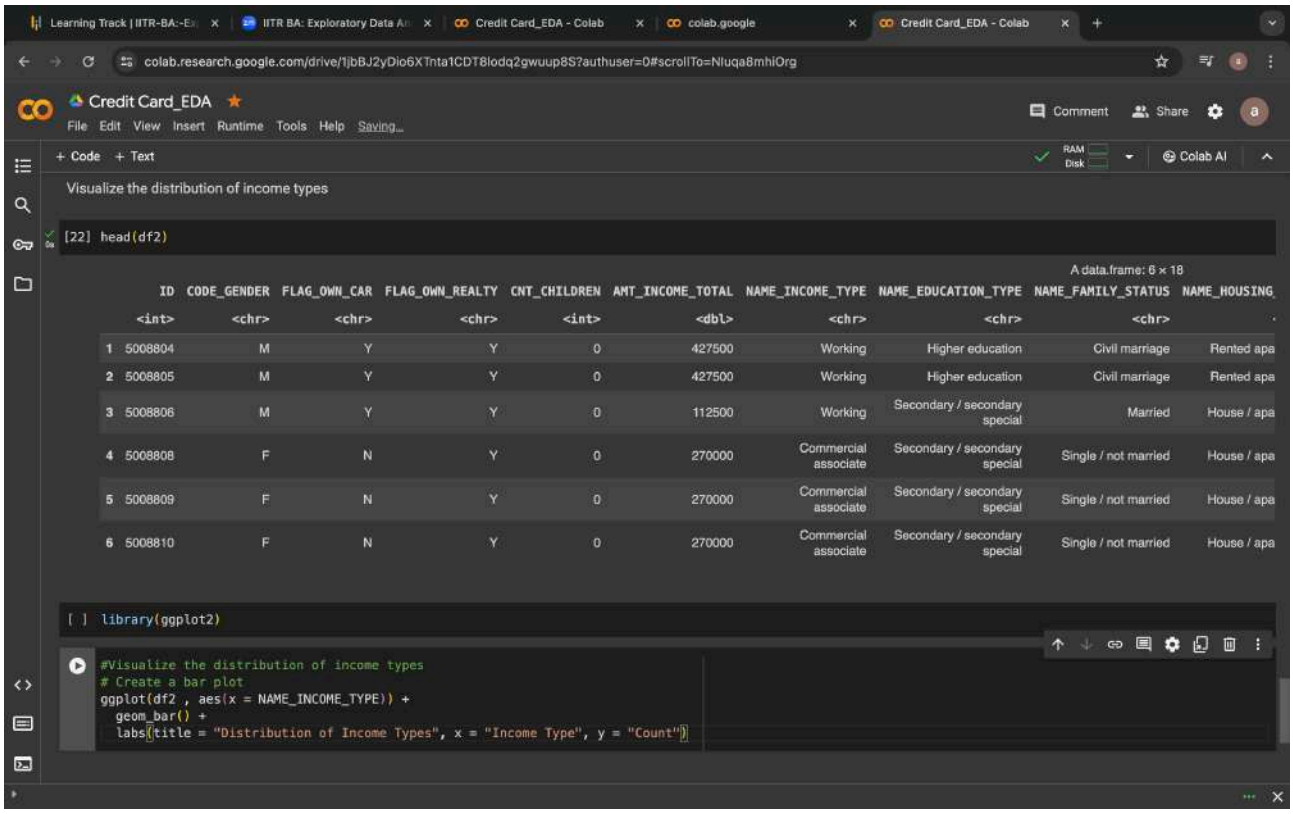
```
duplicated_ids <- df2$ID[duplicated(df2$ID)]
print(duplicated_ids)
```

```
 [1] 7602432 7052783 7838075 7053557 7023651 7213374 7089090 7576316 7023108
[10] 7137299 7836711 7022197 7636389 7207977 7045794 7154819 7836971 7416167
[19] 7702238 7155150 7317997 7036518 7772847 7243768 7743418 7090931 7603224
[28] 7833087 7742298 7050948 7174719 7744386 7052812 7022327 7154598 7046068
[37] 7045885 7636756 7702516 7099881 7372589 7024111 7282535 7742853 7135270
[46] 7091721 7618285
```

To undo cell deletion use ⌘/Ctrl+M Z or the 'Undo' option in the 'Edit' menu

Connected to R Google Compute Engine backend


Credit Card_EDA
File Edit View Insert Runtime Tools Help Saving...
Comment
Share
+ Code + Text
duplicated_ids <- df2$ID[duplicated(df2$ID)]
print(duplicated_ids)
[1] 7602432 7052783 7838075 7053557 7023651 7213374 7089090 7576316 7023108
[10] 7137299 7836711 7022197 7636389 7207977 7045794 7154819 7836971 7416167
[19] 7702238 7155150 7317997 7036518 7772847 7243768 7743418 7090931 7603224
[28] 7833087 7742298 7050948 7174719 7744386 7052812 7022327 7154598 7046068
[37] 7045885 7636756 7702516 7099881 7372589 7024111 7282535 7742853 7135270
[46] 7091721 7618285
Check for duplicates in the 'ID' column and remove any
[17] df1 <- df1[!duplicated(df1$ID),]
[18] # Remove duplicates from the dataframe
df2 <- df2[!duplicated(df2$ID),]
Feature analysis:
What percentage of IDs are common between the two tables?
[19] # Identify common IDs
length(intersect(df1$ID, df2$ID))
36457
[20] length(unique(c(df1$ID, df2$ID)))
448038


Credit Card_EDA
File Edit View Insert Runtime Tools Help Saving...
Comment
Share
+ Code + Text
Visualize the distribution of income types
[22] head(df2)
A data.frame: 6 × 18
ID CODE_GENDER FLAG_OWN_CAR FLAG_OWN_REALTY CNT_CHILDREN AMT_INCOME_TOTAL NAME_INCOME_TYPE NAME_EDUCATION_TYPE NAME_FAMILY_STATUS NAME_HOUSING
<int> <chr> <chr> <chr> <int> <dbl> <chr> <chr> <chr>
1 5008804 M Y Y 0 427500 Working Higher education Civil marriage Rented apa
2 5008805 M Y Y 0 427500 Working Higher education Civil marriage Rented apa
3 5008806 M Y Y 0 112500 Working Secondary / secondary special Married House / apa
4 5008808 F N Y 0 270000 Commercial associate Secondary / secondary special Single / not married House / apa
5 5008809 F N Y 0 270000 Commercial associate Secondary / secondary special Single / not married House / apa
6 5008810 F N Y 0 270000 Commercial associate Secondary / secondary special Single / not married House / apa
[ ] library(ggplot2)
#Visualize the distribution of income types
Create a bar plot
ggplot(df2 , aes(x = NAME_INCOME_TYPE)) +
geom_bar() +
labs(title = "Distribution of Income Types", x = "Income Type", y = "Count")

Chrome File Edit View History Bookmarks Profiles Tab Window Help Tue 30 Apr 5:24 PM

colab.research.google.com/drive/1jbBJ2yDio6XTnta1CDT8lodq2gwuup8S?authuser=0#scrollTo=Nluqa8mhiOrg

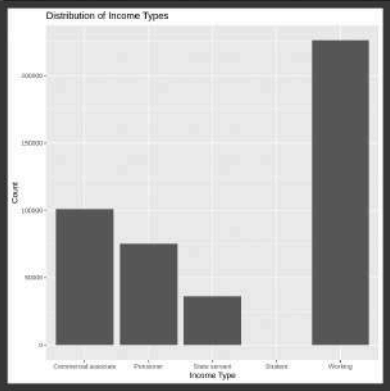
Credit Card_EDA

File Edit View Insert Runtime Tools Help All changes saved

Comment Share a

+ Code + Text RAM Disk Colab AI

```
ggplot(df2 , aes(x = NAME_INCOME_TYPE)) +
  geom_bar() +
  labs(title = "Distribution of Income Types", x = "Income Type", y = "Count")
```



Chrome File Edit View History Bookmarks Profiles Tab Window Help Tue 30 Apr 5:26 PM

colab.research.google.com/drive/1jbBJ2yDio6XTnta1CDT8lodq2gwuup8S?authuser=0#scrollTo=u6WtXSPTi0z8

Credit Card_EDA

File Edit View Insert Runtime Tools Help All changes saved

Comment Share a

+ Code + Text RAM Disk Colab AI

[25]


```
[26] table(df2$NAME_INCOME_TYPE)
```

```
     Commercial associate                 Pensioner              State servant
                   100744                     75488                     36185
                  Student                   Working
                       17                    226076
```

Find the age in years and study the age distribution

```
[27] 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
```

```
# Convert DAYS_BIRTH to age in years
df2 <- mutate(df2, Age = round(-DAYS_BIRTH / 365, 1))
```

Chrome File Edit View History Bookmarks Profiles Tab Window Help Tue 30 Apr 5:27 PM

colab.research.google.com/drive/1jbBJ2yDio6XTnta1CDT8lodq2gwuup8S?authuser=0#scrollTo=xrb0kYUojBXB

Credit Card_EDA

File Edit View Insert Runtime Tools Help All changes saved

Comment Share a

+ Code + Text

RAM Disk Colab AI

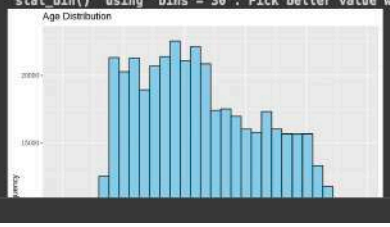
```
[27] library(dplyr)
```

```
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
```

```
[28] # Convert DAYS_BIRTH to age in years
df2 <- mutate(df2, Age = round(-DAYS_BIRTH / 365, 1))
```

```
ggplot(df2 , aes(x = Age)) +
  geom_histogram(fill = "skyblue", color = "black") +
  labs(title = "Age Distribution", x = "Age (years)", y = "Frequency")
```

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



Chrome File Edit View History Bookmarks Profiles Tab Window Help Tue 30 Apr 5:29 PM

colab.research.google.com/drive/1jbBJ2yDio6XTnta1CDT8lodq2gwuup8S?authuser=0#scrollTo=llkKY79XjbJ_

Credit Card_EDA

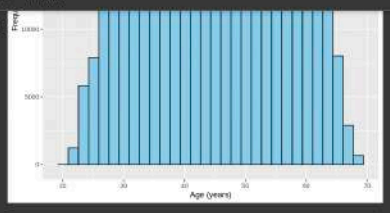
File Edit View Insert Runtime Tools Help All changes saved

Comment Share a

+ Code + Text

RAM Disk Colab AI

```
[29]
```



Study the gender distribution in applicants' data

```
[30] gender_distribution <- table(df2$CODE_GENDER)
```

```
gender_percentage <- prop.table(gender_distribution) * 100
print(gender_percentage)
```

```
       F        M
67.13781 32.86219
```

Chrome File Edit View History Bookmarks Profiles Tab Window Help Tue 30 Apr 5:29 PM

colab.research.google.com/drive/1jbBJ2yDio6XTnta1CDT8lodq2gwuup8S?authuser=0#scrollTo=Nvb4lUv-jkVp

Credit Card_EDA

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

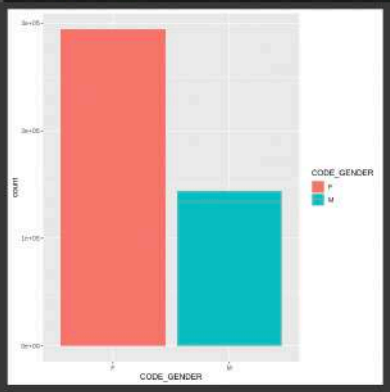
+ Code + Text

RAM Disk Colab AI

```
[31] gender_percentage <- prop.table(gender_distribution) * 100
     print(gender_percentage)
```

```
       F        M
67.13781 32.86219
```

```
ggplot(df2, aes(x = CODE_GENDER, fill = CODE_GENDER)) + geom_bar()
```



Chrome File Edit View History Bookmarks Profiles Tab Window Help Tue 30 Apr 5:31 PM

colab.research.google.com/drive/1jbBJ2yDio6XTnta1CDT8lodq2gwuup8S?authuser=0#scrollTo=pBjz8y-Wj3J2

Credit Card_EDA

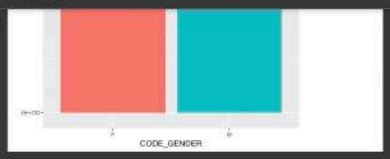
File Edit View Insert Runtime Tools Help Saving...

Comment Share

+ Code + Text

RAM Disk Colab AI

[32]



Study the average annual income across different education levels

```
head(df2)
```

A data.frame: 6 × 19

	ID	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	NAME_INCOME_TYPE	NAME_EDUCATION_TYPE	NAME_FAMILY_STATUS	NAME_HOUSING
	<int>	<chr>	<chr>	<chr>	<int>	<dbl>	<chr>	<chr>	<chr>	
1	5008804	M	Y	Y	0	427500	Working	Higher education	Civil marriage	Rented apa
2	5008805	M	Y	Y	0	427500	Working	Higher education	Civil marriage	Rented apa
3	5008806	M	Y	Y	0	112500	Working	Secondary / secondary special	Married	House / apa
4	5008808	F	N	Y	0	270000	Commercial associate	Secondary / secondary special	Single / not married	House / apa
5	5008809	F	N	Y	0	270000	Commercial associate	Secondary / secondary special	Single / not married	House / apa
6	5008810	F	N	Y	0	270000	Commercial associate	Secondary / secondary special	Single / not married	House / apa

Chrome File Edit View History Bookmarks Profiles Tab Window Help Tue 30 Apr 5:32 PM

colab.research.google.com/drive/1jbBJ2yDio6XTnta1CDT8lodq2gwuup8S?authuser=0#scrollTo=6q0_qaTnkL1W

Credit Card_EDA

File Edit View Insert Runtime Tools Help All changes saved

Comment Share a

+ Code + Text RAM Disk Colab AI

[33]

3	5008806	M	Y	Y	0	112500	Working	Secondary / secondary special	Married	House / apa
4	5008808	F	N	Y	0	270000	Commercial associate	Secondary / secondary special	Single / not married	House / apa
5	5008809	F	N	Y	0	270000	Commercial associate	Secondary / secondary special	Single / not married	House / apa
6	5008810	F	N	Y	0	270000	Commercial associate	Secondary / secondary special	Single / not married	House / apa

```
[34] average_income <- df2 %>%
       group_by(NAME_EDUCATION_TYPE) %>%
       summarise(avg_income = mean(AMT_INCOME_TOTAL))
```

```
print(average_income)
```

```
# A tibble: 5 × 2
  NAME_EDUCATION_TYPE          avg_income
  <chr>                             <dbl>
1 Academic degree                 240692.
2 Higher education                226110.
3 Incomplete higher               207330.
4 Lower secondary                 143934.
5 Secondary / secondary special   172057.
```

Chrome File Edit View History Bookmarks Profiles Tab Window Help Tue 30 Apr 5:33 PM

colab.research.google.com/drive/1jbBJ2yDio6XTnta1CDT8lodq2gwuup8S?authuser=0#scrollTo=_ewRwVLxkYq_

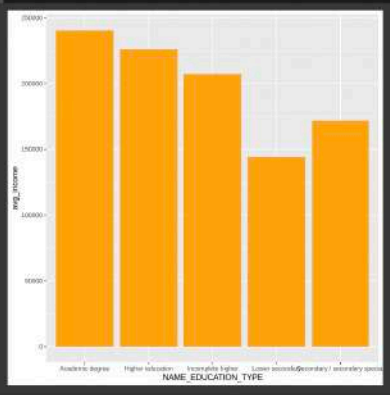
Credit Card_EDA

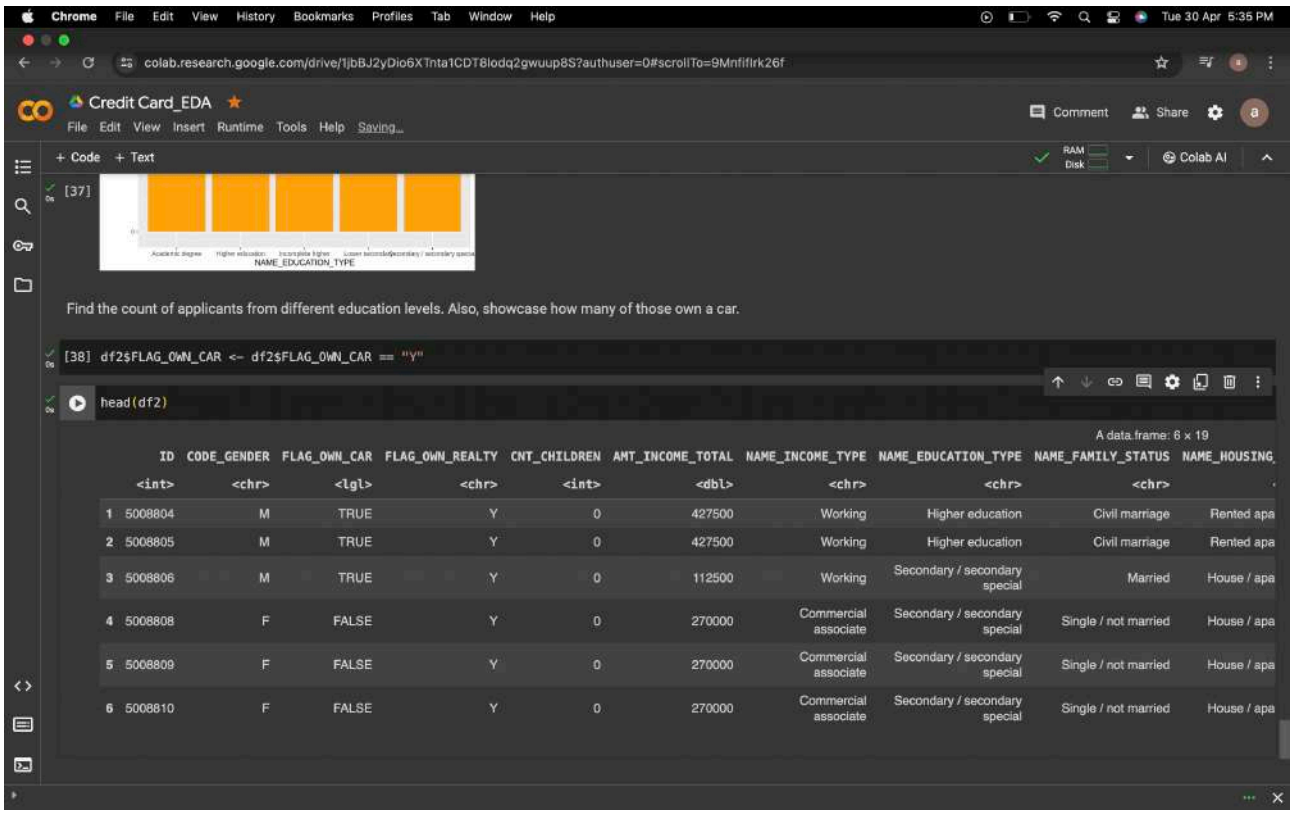
File Edit View Insert Runtime Tools Help All changes saved

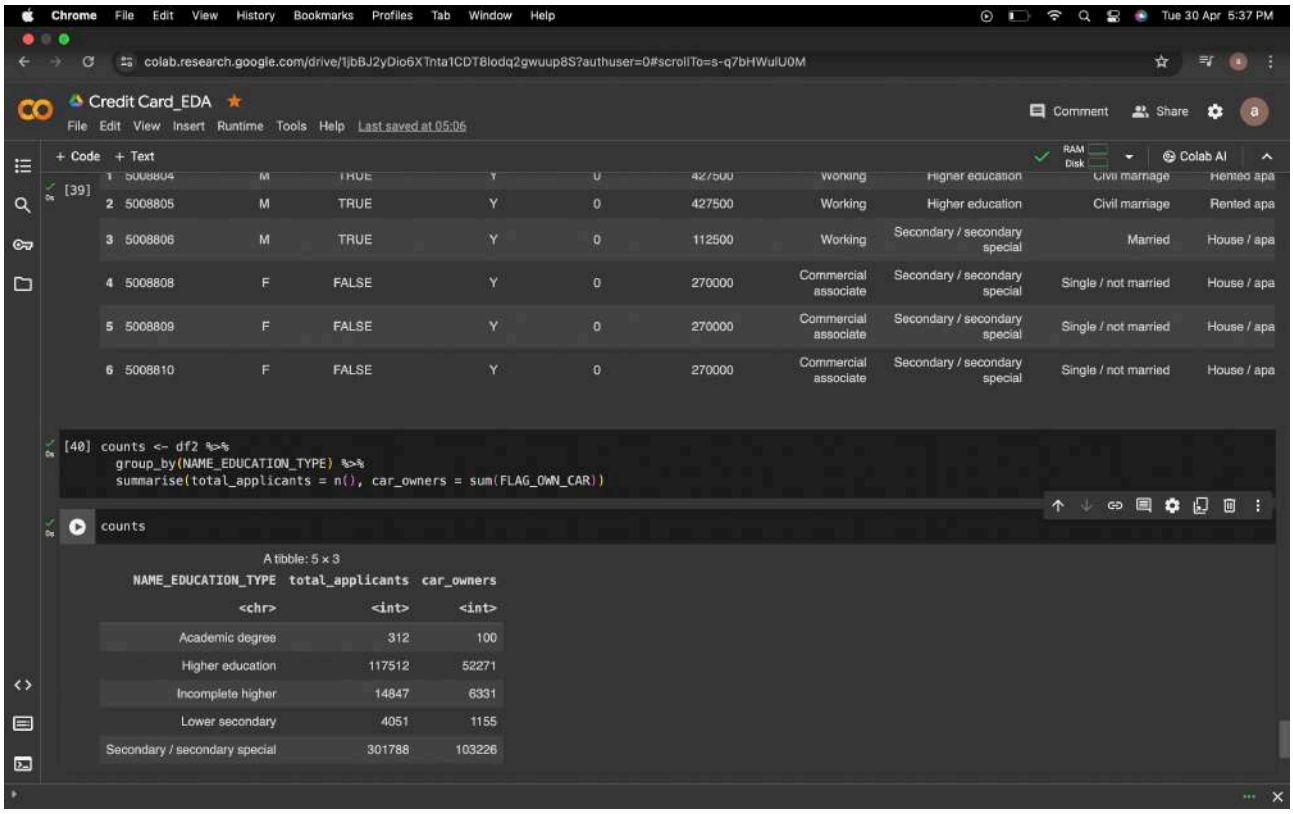
Comment Share a

+ Code + Text RAM Disk Colab AI

```
ggplot(data = average_income, aes(x = NAME_EDUCATION_TYPE, y = avg_income)) +
  geom_bar(stat = "identity", fill = "orange")
```

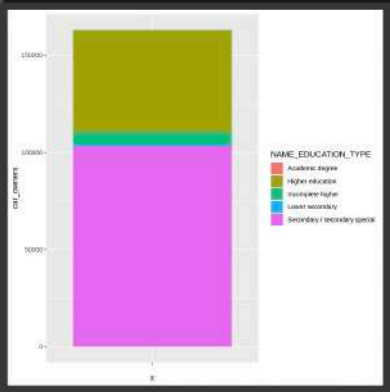



Chrome File Edit View History Bookmarks Profiles Tab Window Help
Tue 30 Apr 5:35 PM
colab.research.google.com/drive/1jbBJ2yDio6XTnta1CDT8lodq2gwuup8S?authuser=0#scrollTo=9MnfiflrK26f
Credit Card_EDA
File Edit View Insert Runtime Tools Help Saving...
Comment
Share
+ Code + Text
RAM
Disk
Colab AI
[37]
NAME_EDUCATION_TYPE
Find the count of applicants from different education levels. Also, showcase how many of those own a car.
[38] df2$FLAG_OWN_CAR <- df2$FLAG_OWN_CAR == "Y"
head(df2)
A data.frame: 6 x 19
ID CODE_GENDER FLAG_OWN_CAR FLAG_OWN_REALTY CNT_CHILDREN AMT_INCOME_TOTAL NAME_INCOME_TYPE NAME_EDUCATION_TYPE NAME_FAMILY_STATUS NAME_HOUSING
<int> <chr> <lgl> <chr> <int> <dbl> <chr> <chr> <chr>
1 5008804 M TRUE Y 0 427500 Working Higher education Civil marriage Rented apa
2 5008805 M TRUE Y 0 427500 Working Higher education Civil marriage Rented apa
3 5008806 M TRUE Y 0 112500 Working Secondary / secondary special Married House / apa
4 5008808 F FALSE Y 0 270000 Commercial associate Secondary / secondary special Single / not married House / apa
5 5008809 F FALSE Y 0 270000 Commercial associate Secondary / secondary special Single / not married House / apa
6 5008810 F FALSE Y 0 270000 Commercial associate Secondary / secondary special Single / not married House / apa


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Tue 30 Apr 5:37 PM
colab.research.google.com/drive/1jbBJ2yDio6XTnta1CDT8lodq2gwuup8S?authuser=0#scrollTo=s-q7bHWulU0M
Credit Card_EDA
File Edit View Insert Runtime Tools Help Last saved at 05:06
Comment
Share
+ Code + Text
RAM
Disk
Colab AI
[39]
2 5008805 M TRUE Y 0 427500 Working Higher education Civil marriage Rented apa
3 5008806 M TRUE Y 0 112500 Working Secondary / secondary special Married House / apa
4 5008808 F FALSE Y 0 270000 Commercial associate Secondary / secondary special Single / not married House / apa
5 5008809 F FALSE Y 0 270000 Commercial associate Secondary / secondary special Single / not married House / apa
6 5008810 F FALSE Y 0 270000 Commercial associate Secondary / secondary special Single / not married House / apa
[40] counts <- df2 %>%
group_by(NAME_EDUCATION_TYPE) %>%
summarise(total_applicants = n(), car_owners = sum(FLAG_OWN_CAR))
counts
A tibble: 5 x 3
NAME_EDUCATION_TYPE total_applicants car_owners
<chr> <int> <int>
Academic degree 312 100
Higher education 117512 52271
Incomplete higher 14847 6331
Lower secondary 4051 1155
Secondary / secondary special 301788 103226

Create a stacked column chart to visualize the above

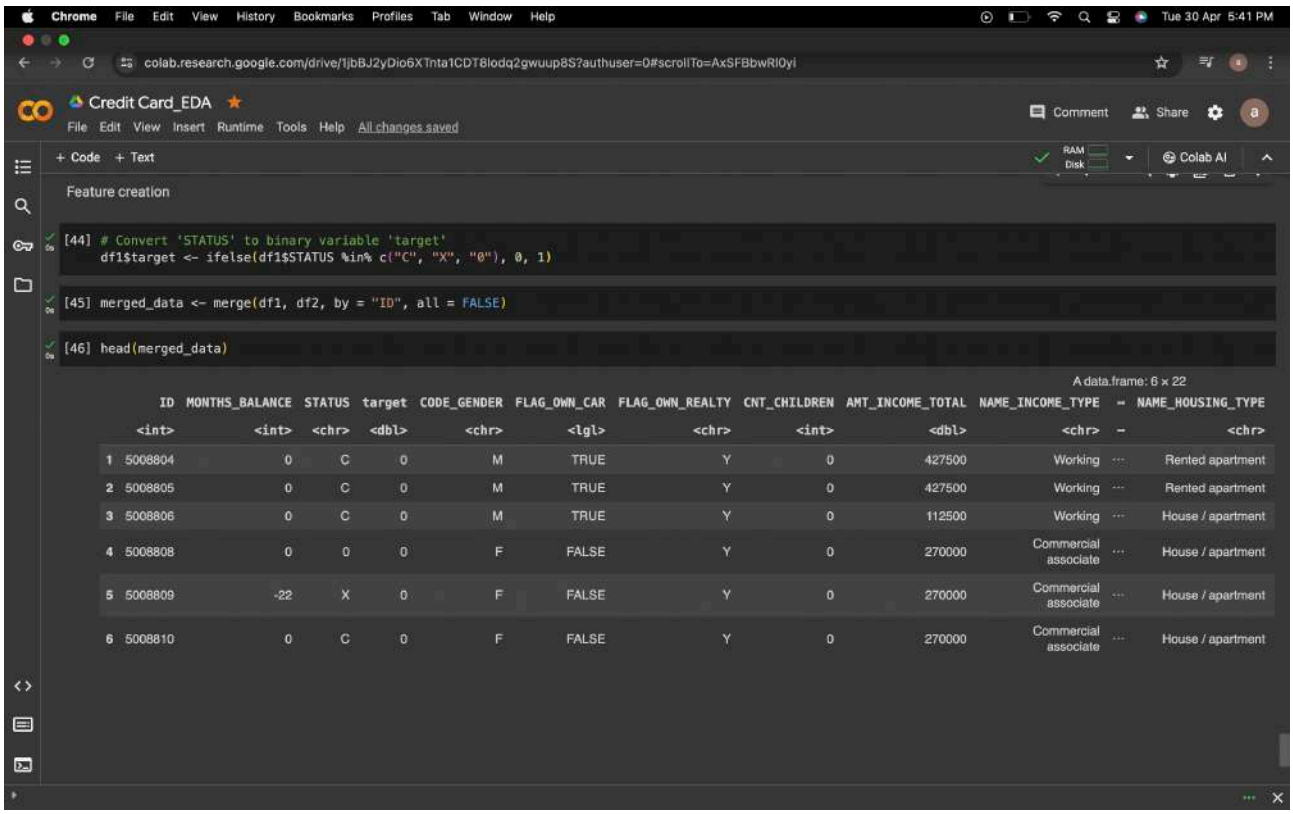
```
ggplot(counts, aes(x="", y=car_owners, fill = NAME_EDUCATION_TYPE)) + geom_bar(stat = "identity")
```




```
[43] head(df1)
```

A data.frame: 6 x 3

	ID	MONTHS_BALANCE	STATUS
	<int>	<int>	<chr>
1	5001711	0	X
5	5001712	0	C
24	5001713	0	X
46	5001714	0	X
61	5001715	0	X
121	5001717	0	C



Chrome File Edit View History Bookmarks Profiles Tab Window Help — Tue 30 Apr 5:41 PM

colab.research.google.com/drive/1jbBJ2yDio6XTnta1CDT8lodq2gwuup8S?authuser=0#scrollTo=AxSFBbwRl0yi

Credit Card_EDA

File Edit View Insert Runtime Tools Help All changes saved

Comment Share a

+ Code + Text

RAM Disk Colab AI

Feature creation

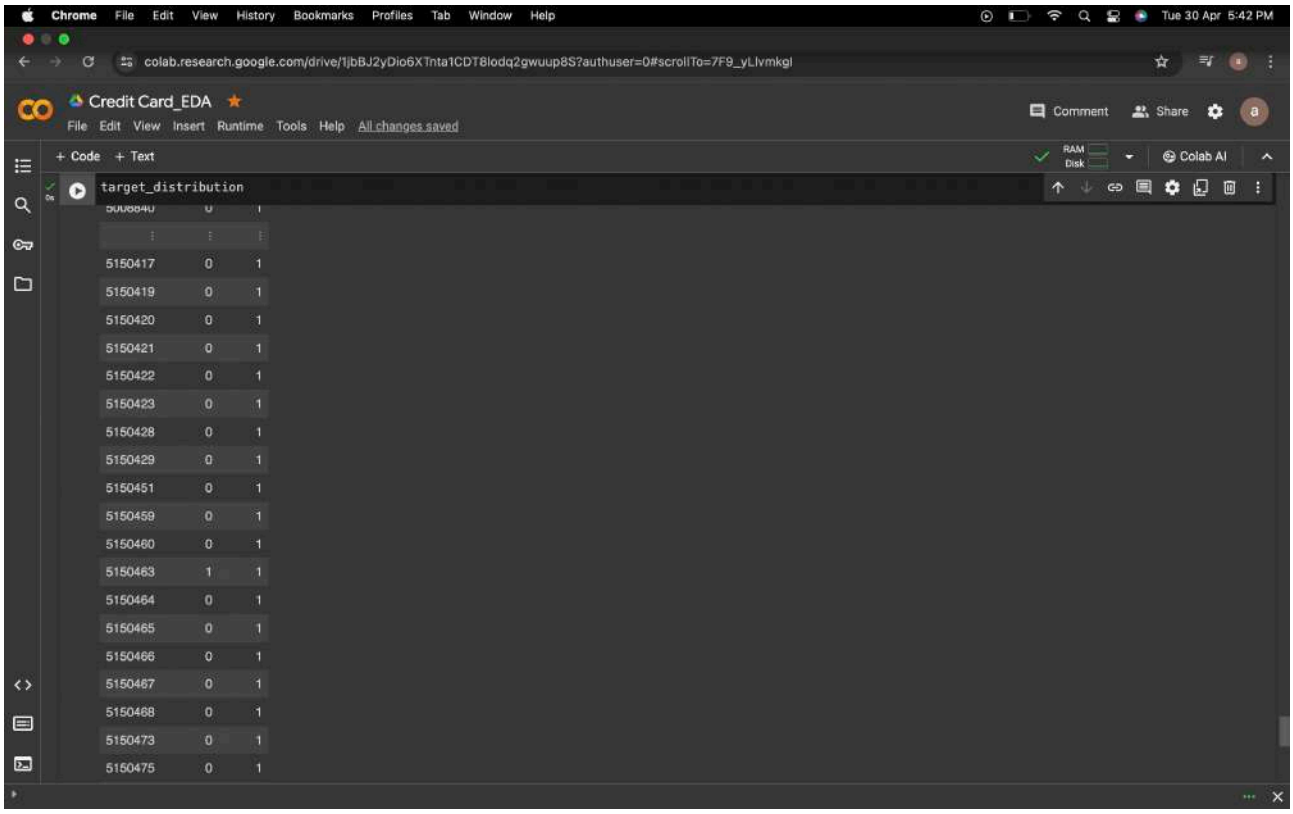
```
[44] # Convert 'STATUS' to binary variable 'target'
     df1$target <- ifelse(df1$STATUS %in% c("C", "X", "0"), 0, 1)
```

```
[45] merged_data <- merge(df1, df2, by = "ID", all = FALSE)
```

```
[46] head(merged_data)
```

A data.frame: 6 × 22

	ID	MONTHS_BALANCE	STATUS	target	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	NAME_INCOME_TYPE	…	NAME_HOUSING_TYPE
	<int>	<int>	<chr>	<dbl>	<chr>	<lgl>	<chr>	<int>	<dbl>	<chr>	…	<chr>
1	5008804	0	C	0	M	TRUE	Y	0	427500	Working	…	Rented apartment
2	5008805	0	C	0	M	TRUE	Y	0	427500	Working	…	Rented apartment
3	5008806	0	C	0	M	TRUE	Y	0	112500	Working	…	House / apartment
4	5008808	0	0	0	F	FALSE	Y	0	270000	Commercial associate	…	House / apartment
5	5008809	-22	X	0	F	FALSE	Y	0	270000	Commercial associate	…	House / apartment
6	5008810	0	C	0	F	FALSE	Y	0	270000	Commercial associate	…	House / apartment



Chrome File Edit View History Bookmarks Profiles Tab Window Help — Tue 30 Apr 5:42 PM

colab.research.google.com/drive/1jbBJ2yDio6XTnta1CDT8lodq2gwuup8S?authuser=0#scrollTo=7F9_yLIvmkgI

Credit Card_EDA

File Edit View Insert Runtime Tools Help All changes saved

Comment Share a

+ Code + Text

RAM Disk Colab AI

```
target_distribution
```

5008840	0	1
⋮	⋮	⋮
5150417	0	1
5150419	0	1
5150420	0	1
5150421	0	1
5150422	0	1
5150423	0	1
5150428	0	1
5150429	0	1
5150451	0	1
5150459	0	1
5150460	0	1
5150463	1	1
5150464	0	1
5150465	0	1
5150466	0	1
5150467	0	1
5150468	0	1
5150473	0	1
5150475	0	1

Chrome File Edit View History Bookmarks Profiles Tab Window Help Tue 30 Apr 5:45 PM

colab.research.google.com/drive/1jbBJ2yDio6XTnta1CDT8lodq2gwuup8S?authuser=0#scrollTo=WC0R1PqQm7MR

Credit Card_EDA

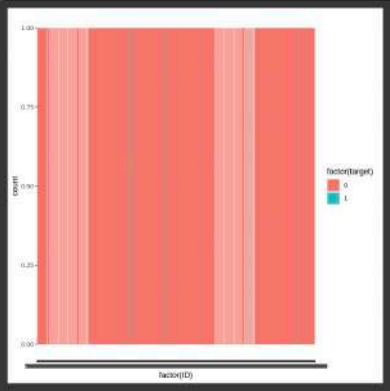
File Edit View Insert Runtime Tools Help All changes saved

Comment Share a

+ Code + Text

RAM Disk Colab AI

```
ggplot(data = target_distribution, aes(x = factor(ID), y = count, fill = factor(target))) +
  geom_bar(stat = "identity", position = "dodge")
```




```
[50] # Convert categorical columns into factors
categorical_cols <- c("CODE_GENDER", "FLAG_OWN_CAR", "FLAG_OWN_REALTY", "NAME_INCOME_TYPE", "NAME_EDUCATION_TYPE", "NAME_FAMILY_STATUS", "NAME_HOUSING_TYPE
merged_data[categorical_cols] <- lapply(merged_data[categorical_cols], as.factor)
```

Chrome File Edit View History Bookmarks Profiles Tab Window Help Tue 30 Apr 5:48 PM

colab.research.google.com/drive/1jbBJ2yDio6XTnta1CDT8lodq2gwuup8S?authuser=0#scrollTo=4YruLt0ZncTm

Credit Card_EDA

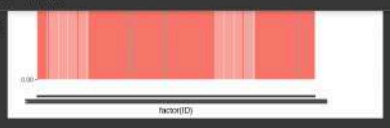
File Edit View Insert Runtime Tools Help All changes saved

Comment Share a

+ Code + Text

RAM Disk Colab AI

[49]



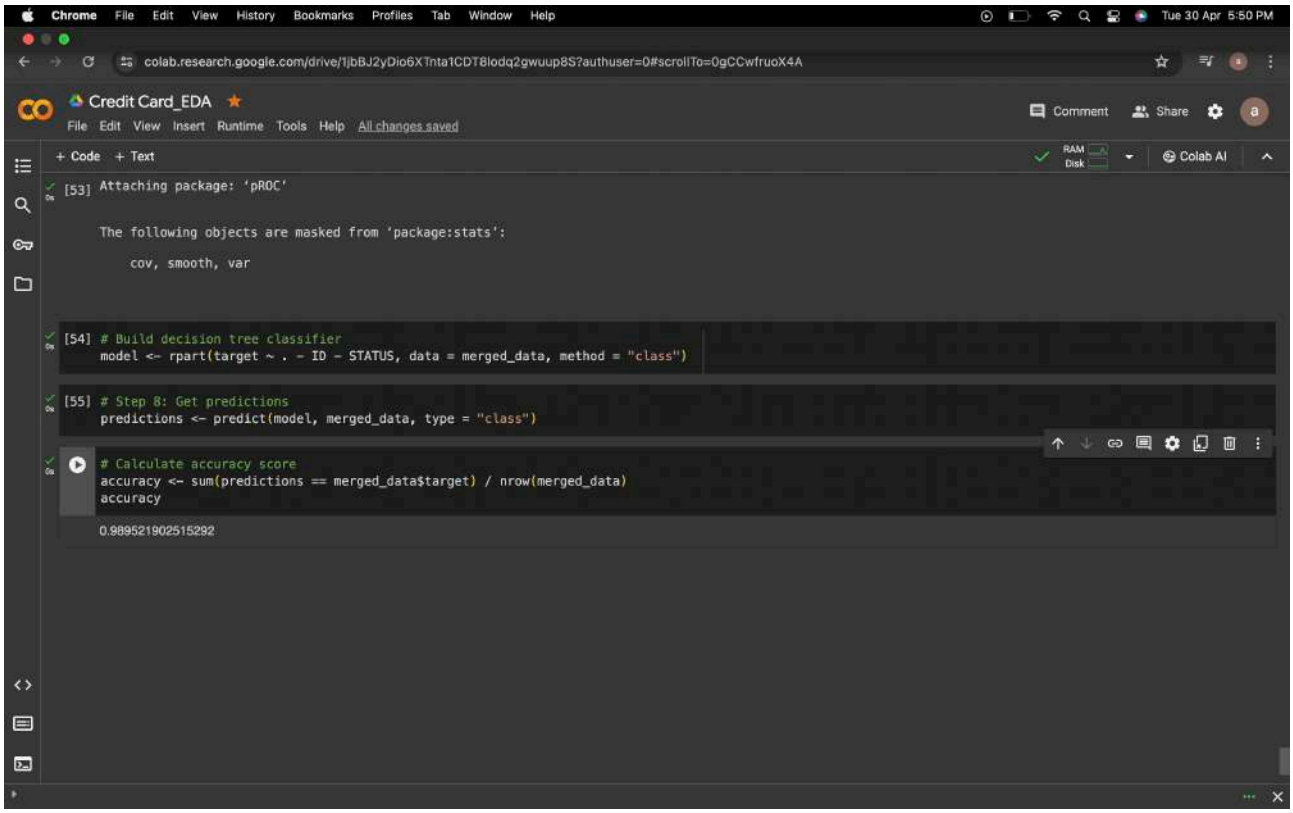

```
[51] # Convert categorical columns into factors
categorical_cols <- c("CODE_GENDER", "FLAG_OWN_CAR", "FLAG_OWN_REALTY", "NAME_INCOME_TYPE", "NAME_EDUCATION_TYPE", "NAME_FAMILY_STATUS", "NAME_HOUSING_TYPE
merged_data[categorical_cols] <- lapply(merged_data[categorical_cols], as.factor)
```

```
[52] install.packages('pROC')
```

```
Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)

also installing the dependencies 'plyr', 'Rcpp'
```

```
library(pROC)
library(rpart)
```

```
Type 'citation("pROC")' for a citation.


Attaching package: 'pROC'


The following objects are masked from 'package:stats':

    cov, smooth, var
```


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Tue 30 Apr 5:50 PM
colab.research.google.com/drive/1jbBJ2yDio6XTnta1CDT8lodq2gwuup8S?authuser=0#scrollTo=0gCCwfruoX4A
Credit Card_EDA
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Colab AI
[53] Attaching package: 'pROC'
The following objects are masked from 'package:stats':
cov, smooth, var
[54] # Build decision tree classifier
model <- rpart(target ~ . - ID - STATUS, data = merged_data, method = "class")
[55] # Step 8: Get predictions
predictions <- predict(model, merged_data, type = "class")
# Calculate accuracy score
accuracy <- sum(predictions == merged_data$target) / nrow(merged_data)
accuracy
0.989521902515292