Nando Dyas Arya

Dashboard 1

FINAL PROJECT CELERATES

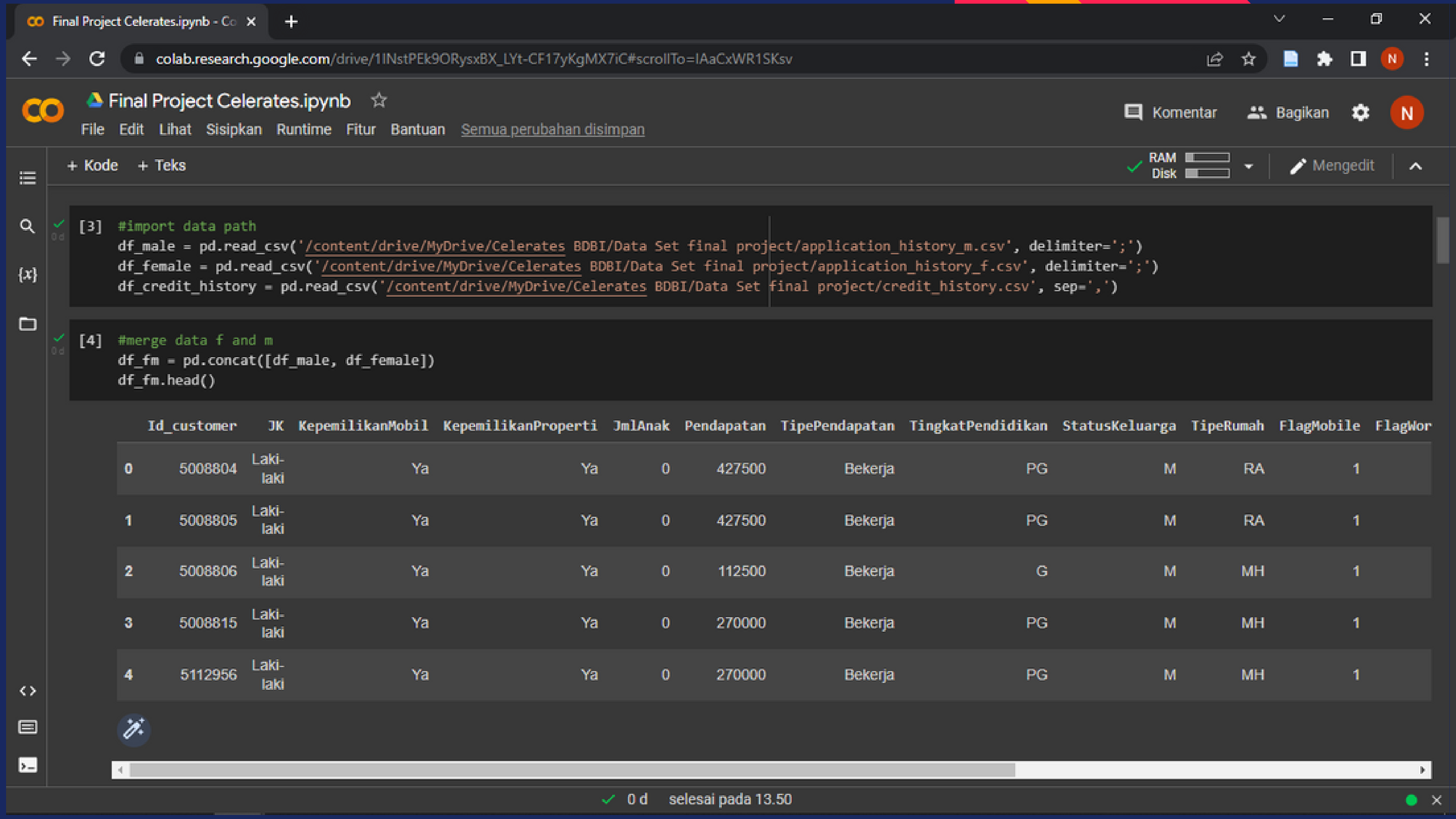
Discussion Points

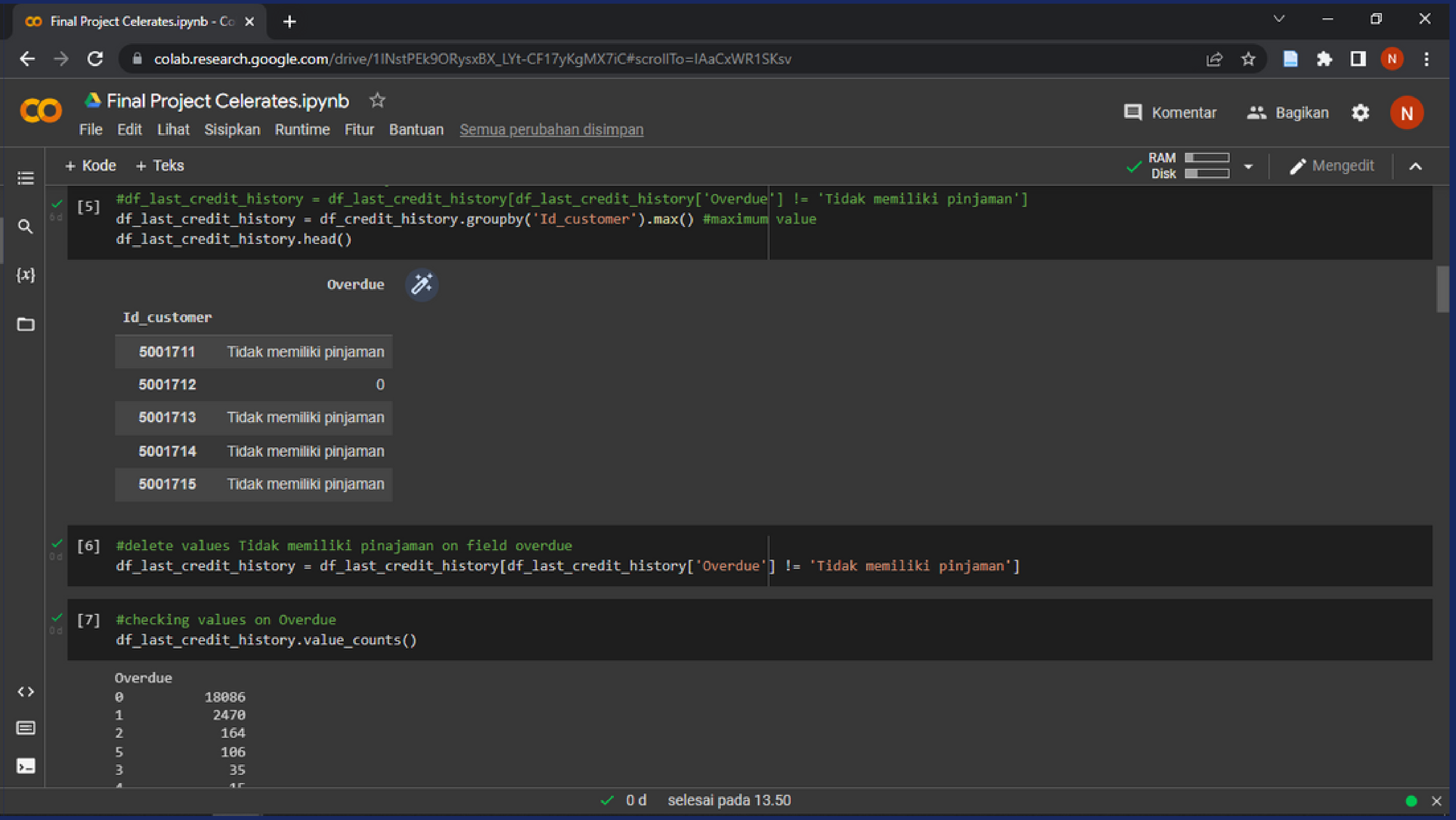
- Data Preprocessing
- Modelling
- Data Testing

TODAY'S MAIN TOPICS

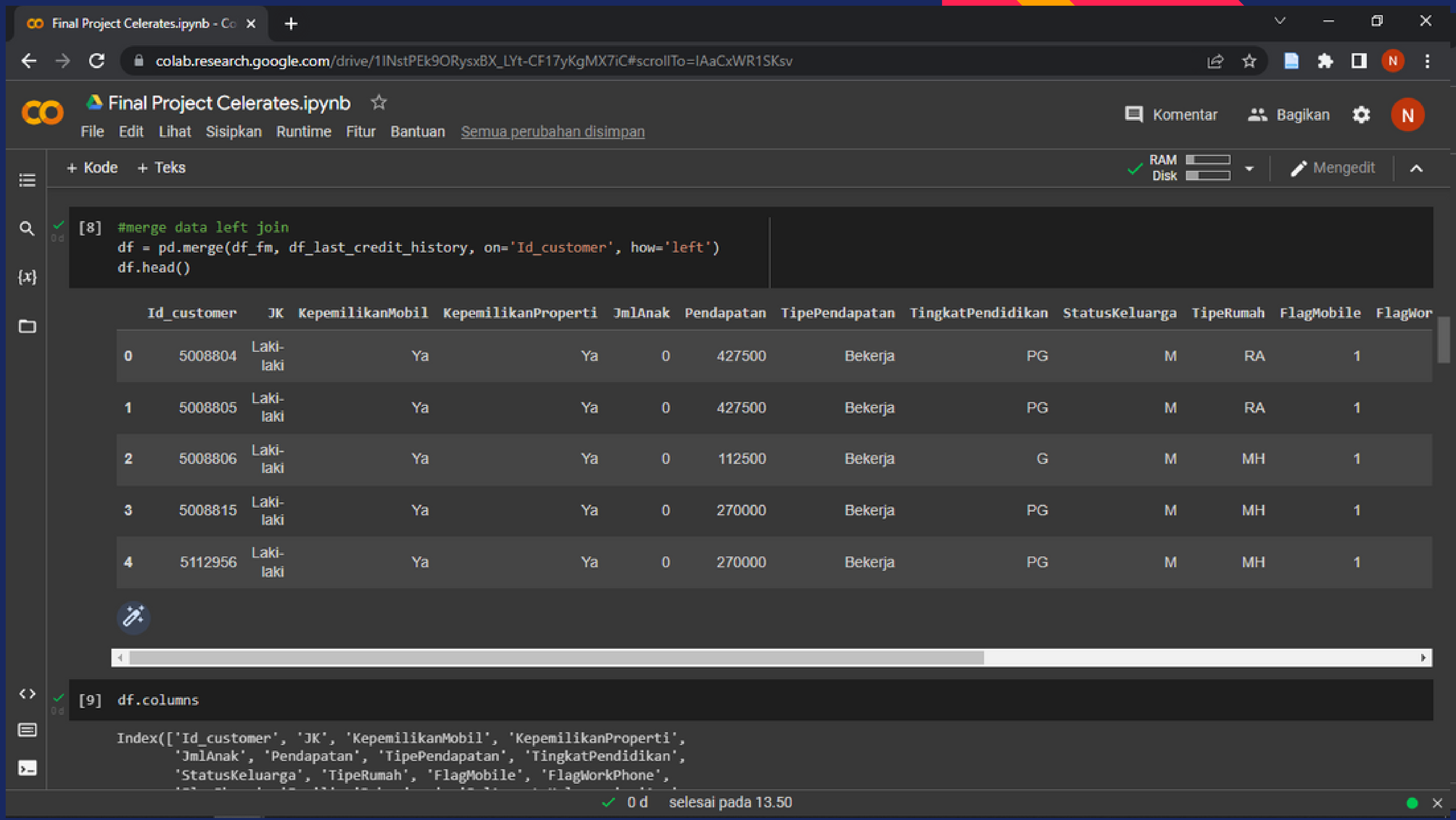
FIRST TOPIC PART 1

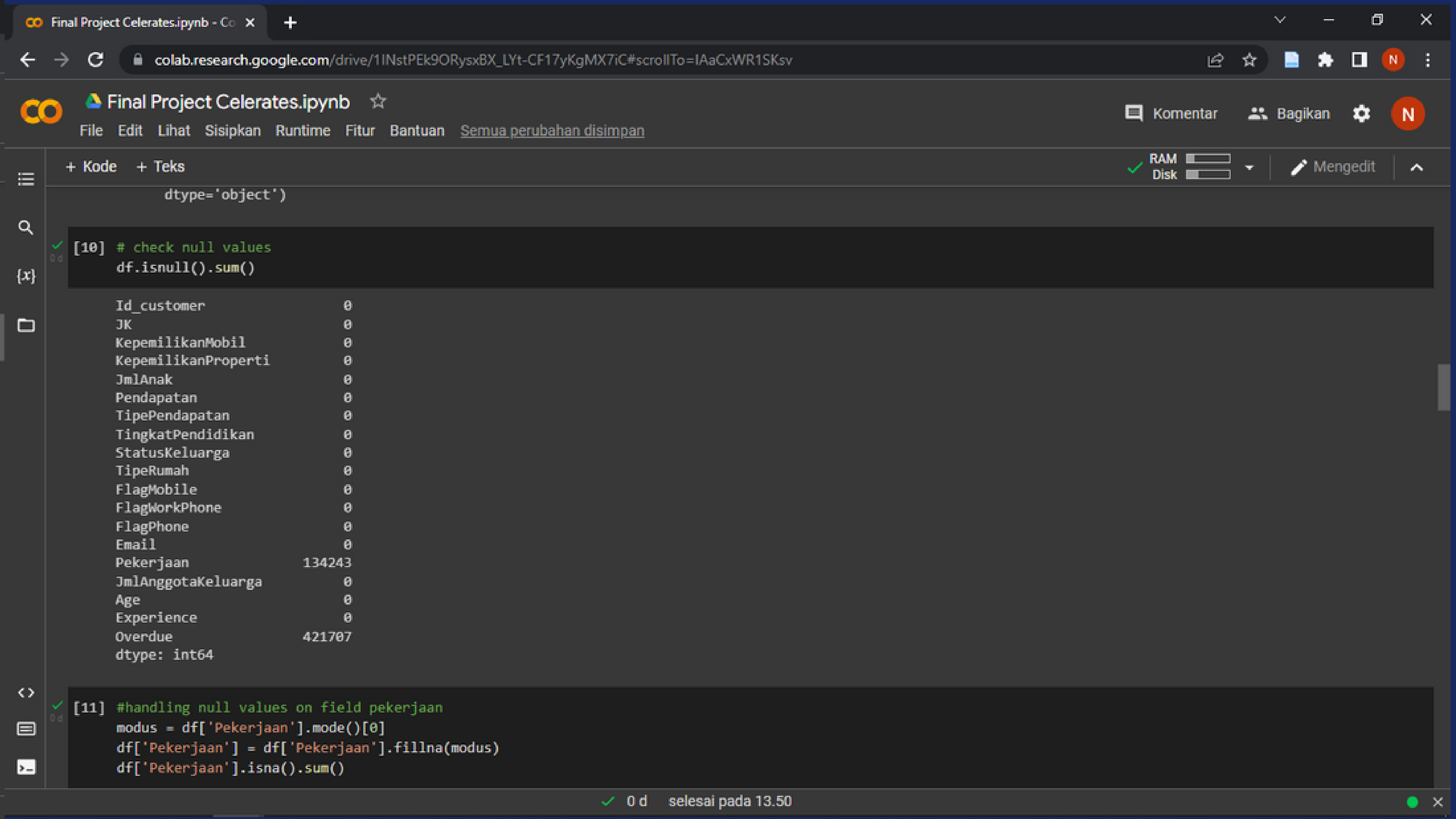
Data Preprocessing





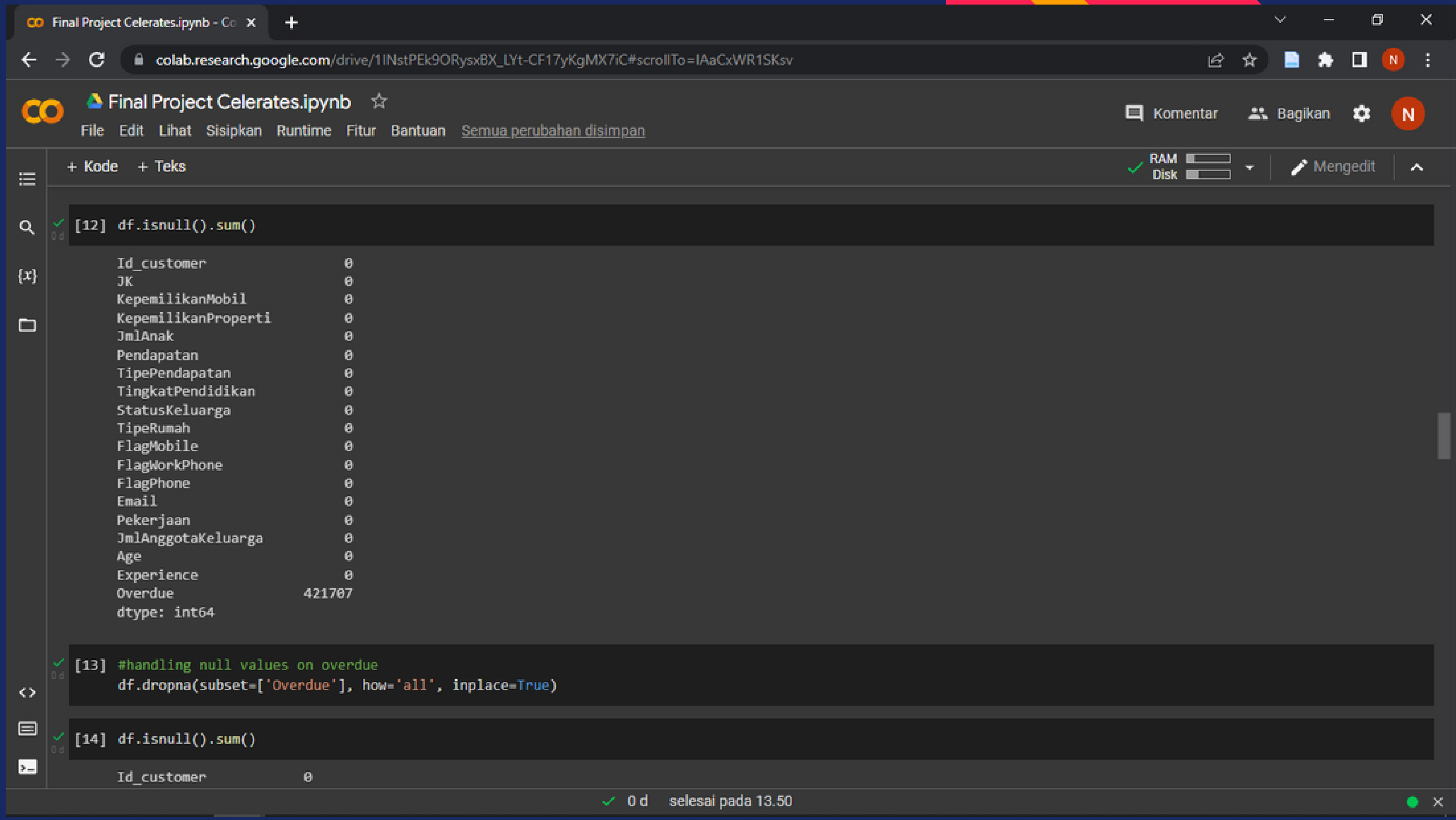
Read data and get max values from credit history

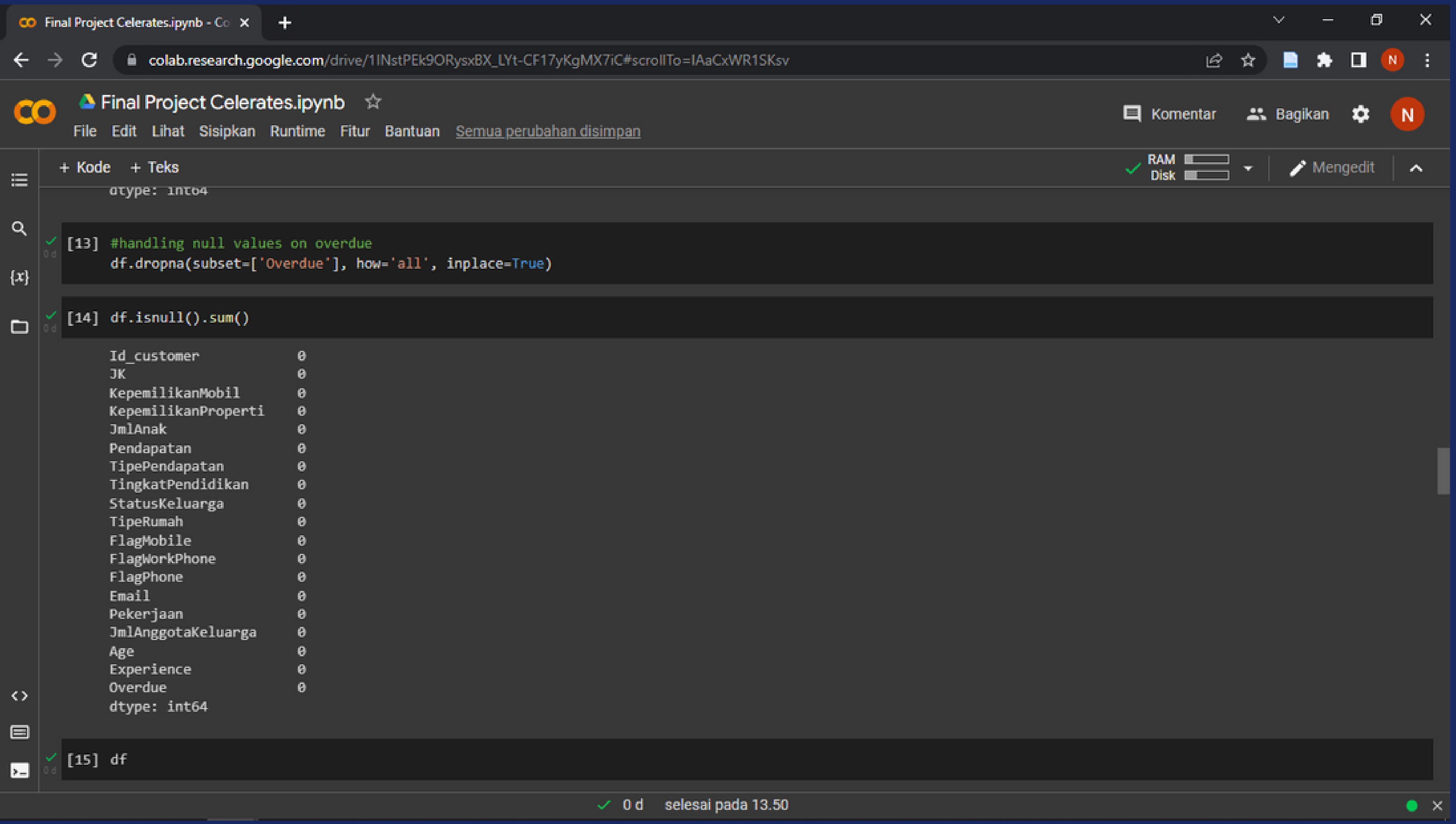




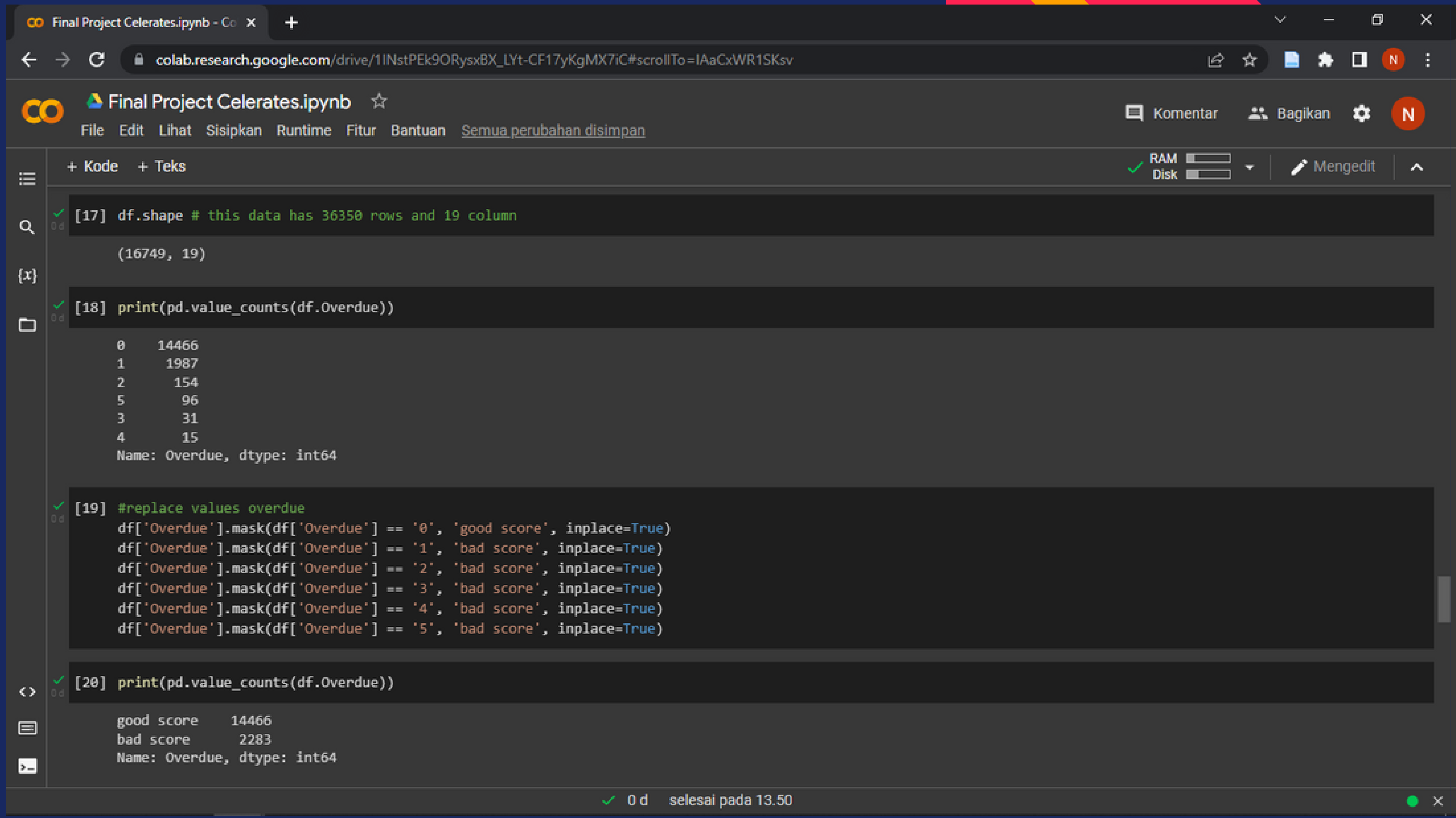
Join male and female data to credit history

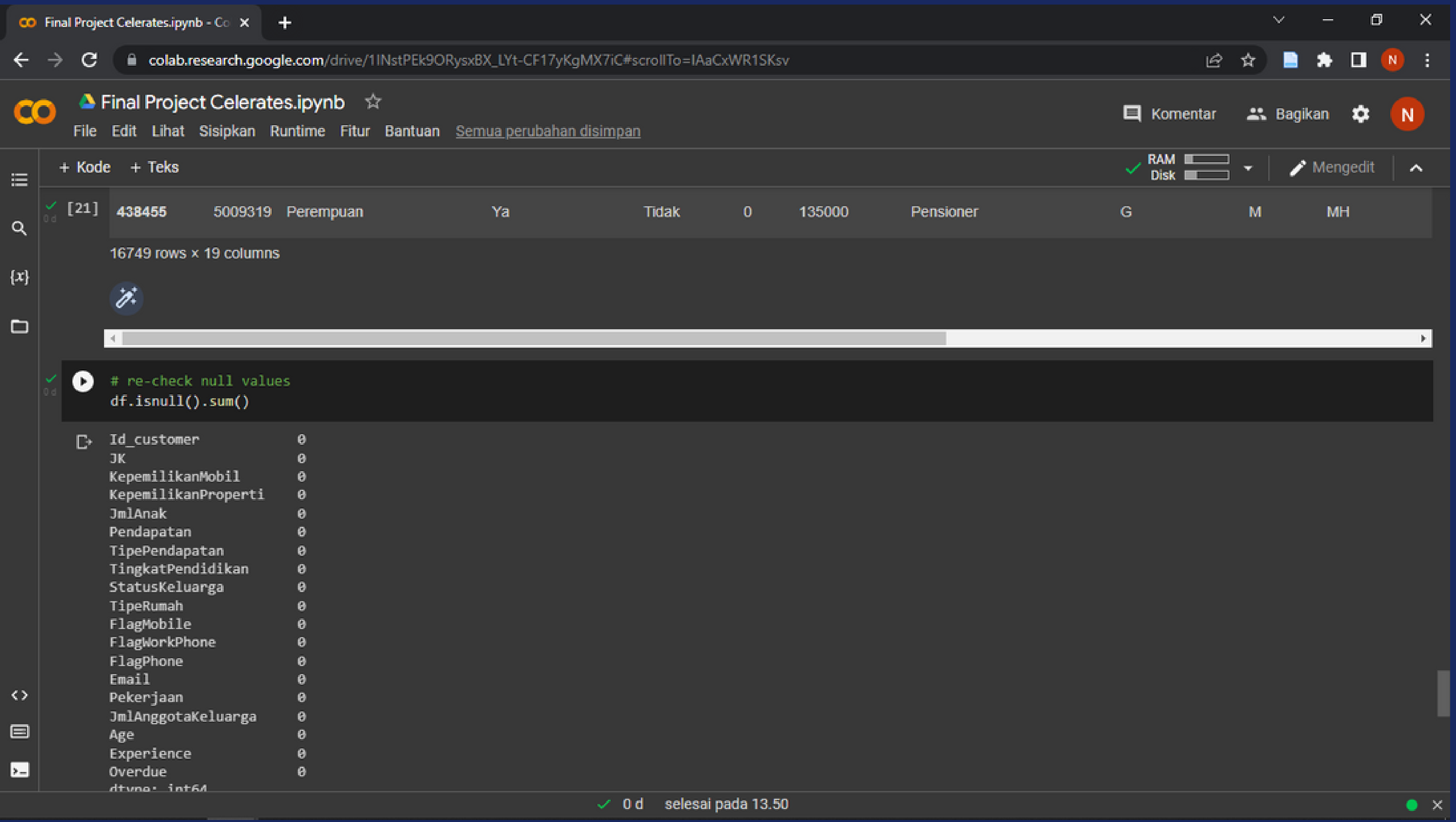
Handling null values on field pekerjaan





Removes null values in overdue as they are not used in creating the model





Manipulate data values in overdue and check if there is null data before exploration

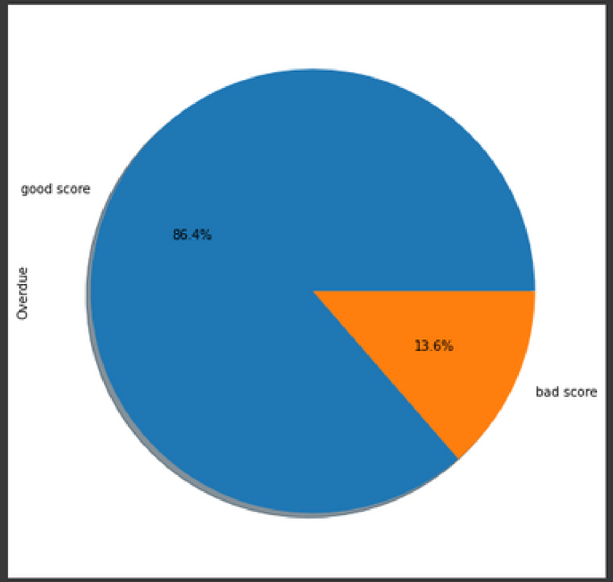
FIRST TOPIC PART 2

Exploring

Final Project Celerates.ipynb

File Edit Lihat Sisipkan Runtime Fitur Bantuan Semua perubahan disimpan

```
df['Overdue'].value_counts().plot.pie(x=[0.1,0.1,0.1],autopct='%1.1f%%'
                                      ,shadow=True, figsize=(10,8))
plt.show()
```

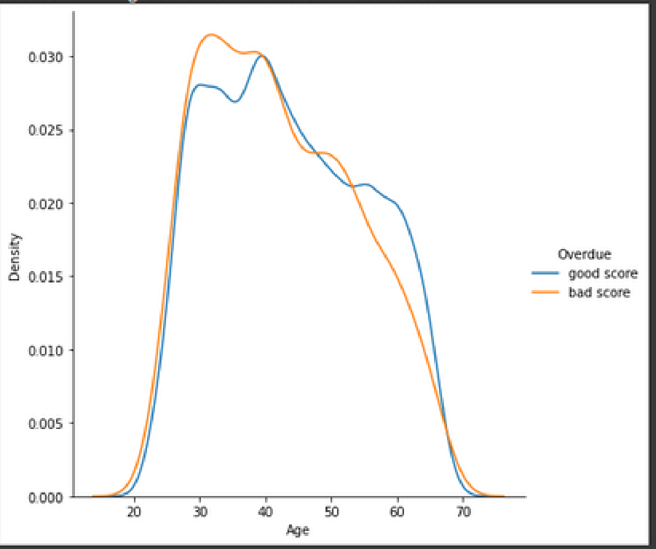


Final Project Celerates.ipynb

File Edit Lihat Sisipkan Runtime Fitur Bantuan Semua perubahan disimpan

```
[25] sns.FacetGrid(df, hue="Overdue", size=6) \
       .map(sns.kdeplot, "Age") \
       .add_legend()
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/axisgrid.py:337: UserWarning: The `size` parameter has been renamed to `height`; please update your code.
  warnings.warn(msg, UserWarning)
<seaborn.axisgrid.FacetGrid at 0x7f4b3e03d880>
```



Final Project Celerates.ipynb

File Edit Lihat Sisipkan Runtime Fitur Bantuan Semua perubahan disimpan

```
/usr/local/lib/python3.8/dist-packages/pandas/io/formats/style.py:2813: RuntimeWarning: All-NaN slice encountered
  smin = np.nanmin(gmap) if vmin is None else vmin
/usr/local/lib/python3.8/dist-packages/pandas/io/formats/style.py:2814: RuntimeWarning: All-NaN slice encountered
  smax = np.nanmax(gmap) if vmax is None else vmax
```

	Id_customer	JmlAnak	FlagMobile	FlagWorkPhone	FlagPhone	Email	JmlAnggotaKeluarga	Age	Experience
Id_customer	1.00	0.03	nan	0.08	0.01	-0.05	0.02	-0.05	-0.05
JmlAnak	0.03	1.00	nan	0.07	-0.01	0.02	0.89	-0.33	-0.23
FlagMobile	nan	nan	nan	nan	nan	nan	nan	nan	nan
FlagWorkPhone	0.08	0.07	nan	1.00	0.32	-0.03	0.07	-0.18	-0.24
FlagPhone	0.01	-0.01	nan	0.32	1.00	0.02	-0.01	0.02	-0.01
Email	-0.05	0.02	nan	-0.03	0.02	1.00	0.02	-0.11	-0.08
JmlAnggotaKeluarga	0.02	0.89	nan	0.07	-0.01	0.02	1.00	-0.29	-0.22
Age	-0.05	-0.33	nan	-0.18	0.02	-0.11	-0.29	1.00	0.63
Experience	-0.05	-0.23	nan	-0.24	-0.01	-0.08	-0.22	0.63	1.00

Imbalance Data Check and Handling

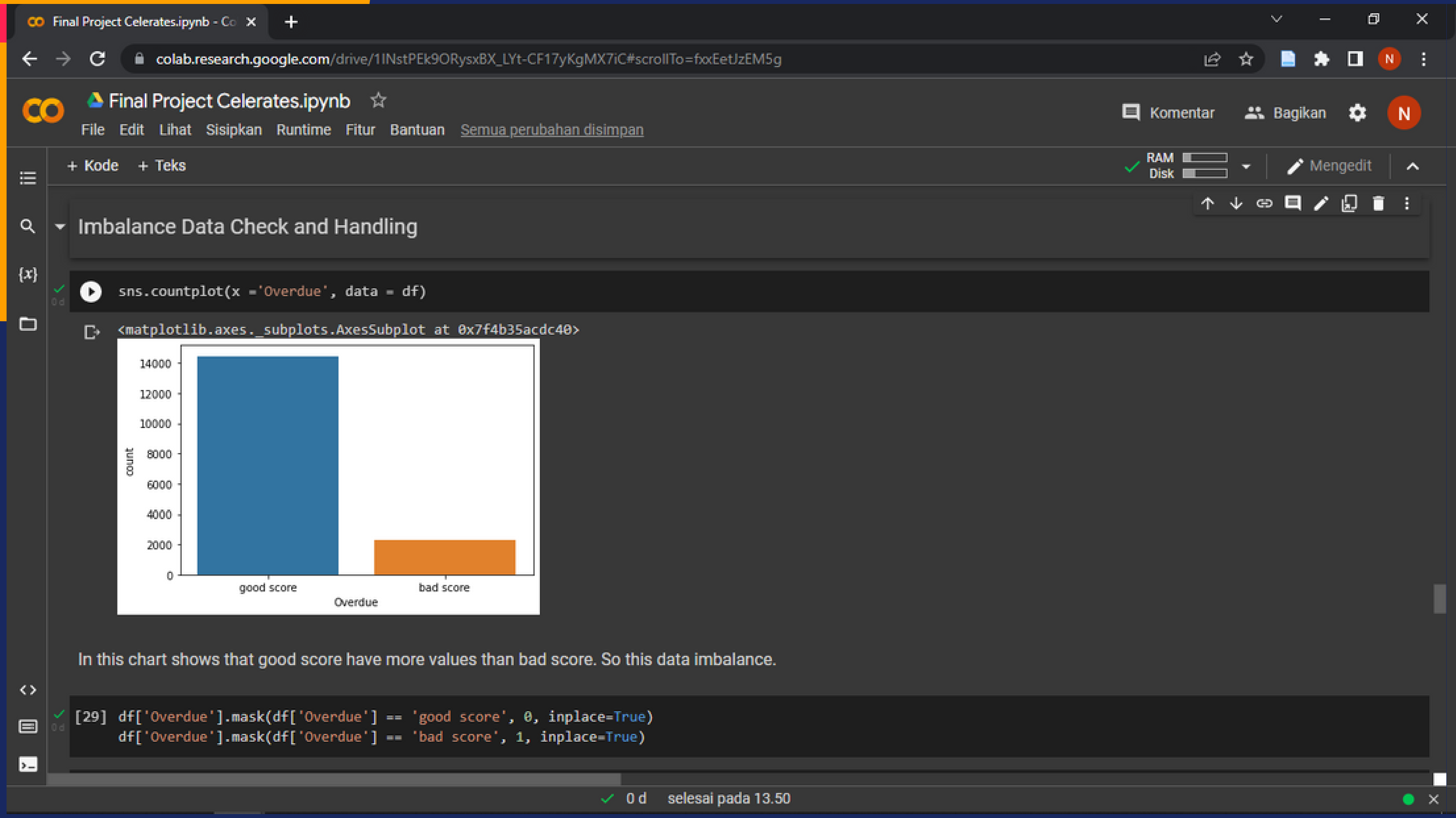
[ ] 15 sel tersembunyi

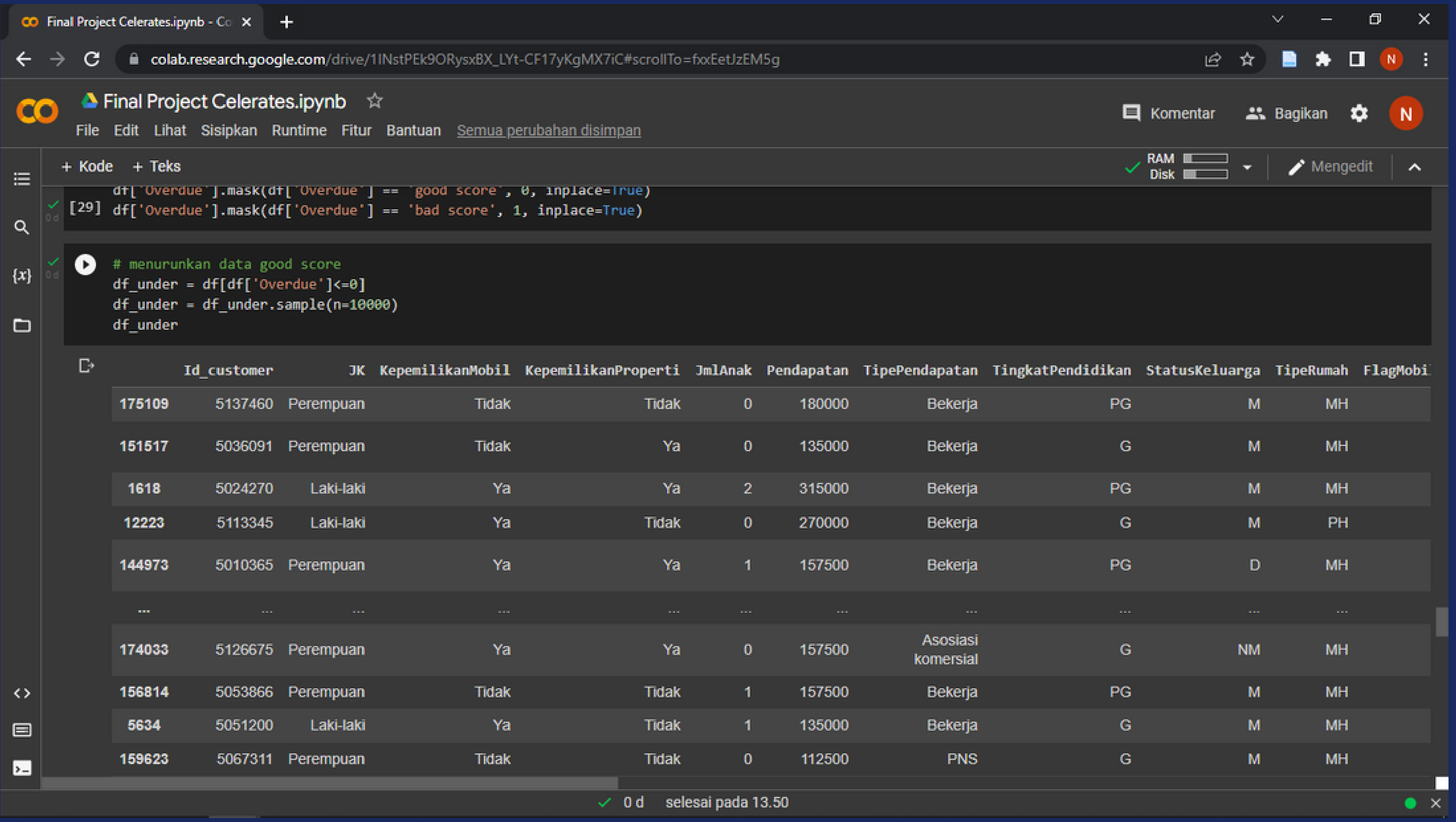
Modelling

0 d selesai pada 13.50

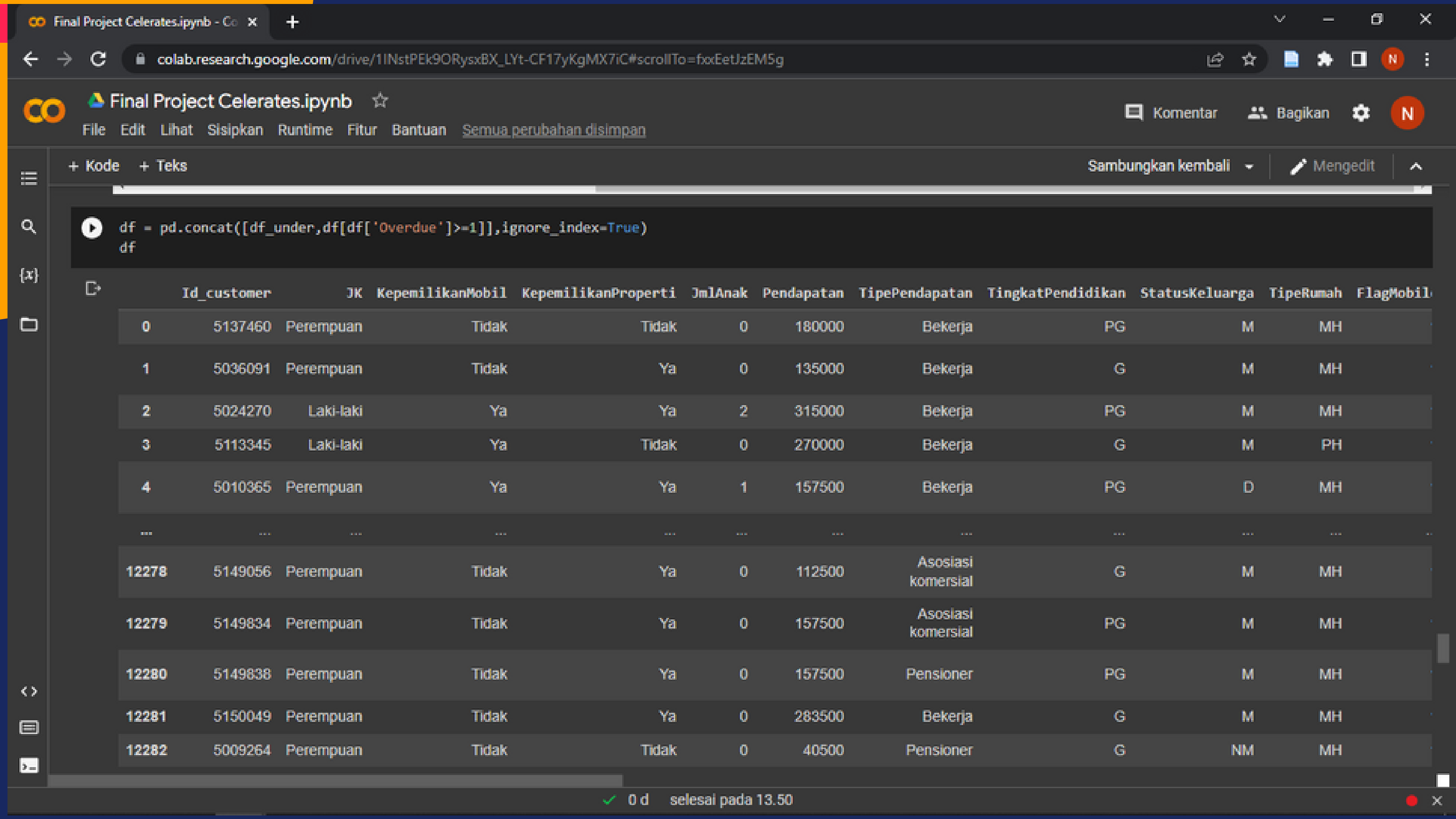
FIRST TOPIC PART 3

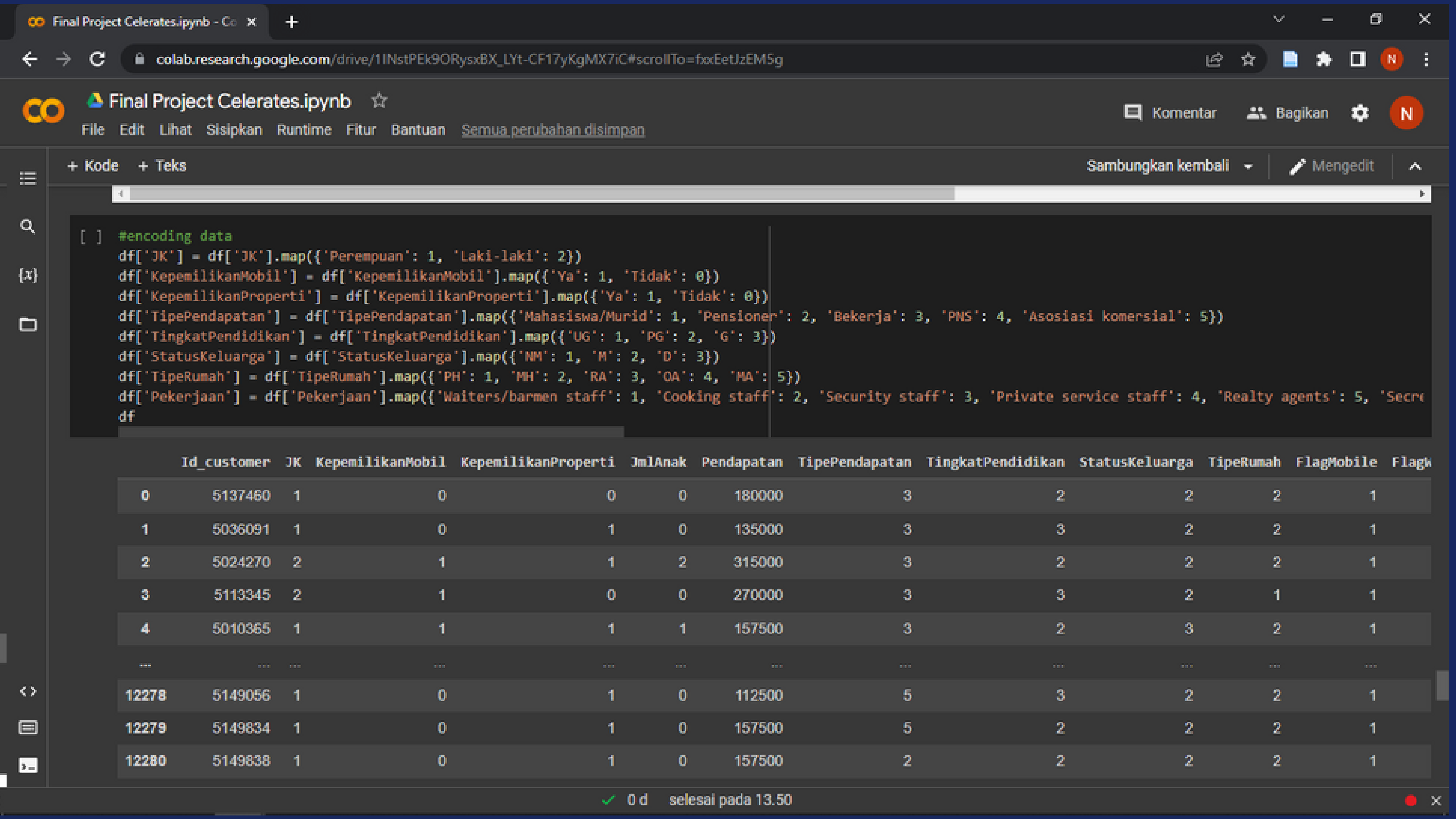
Imbalance data Cheking and Handling



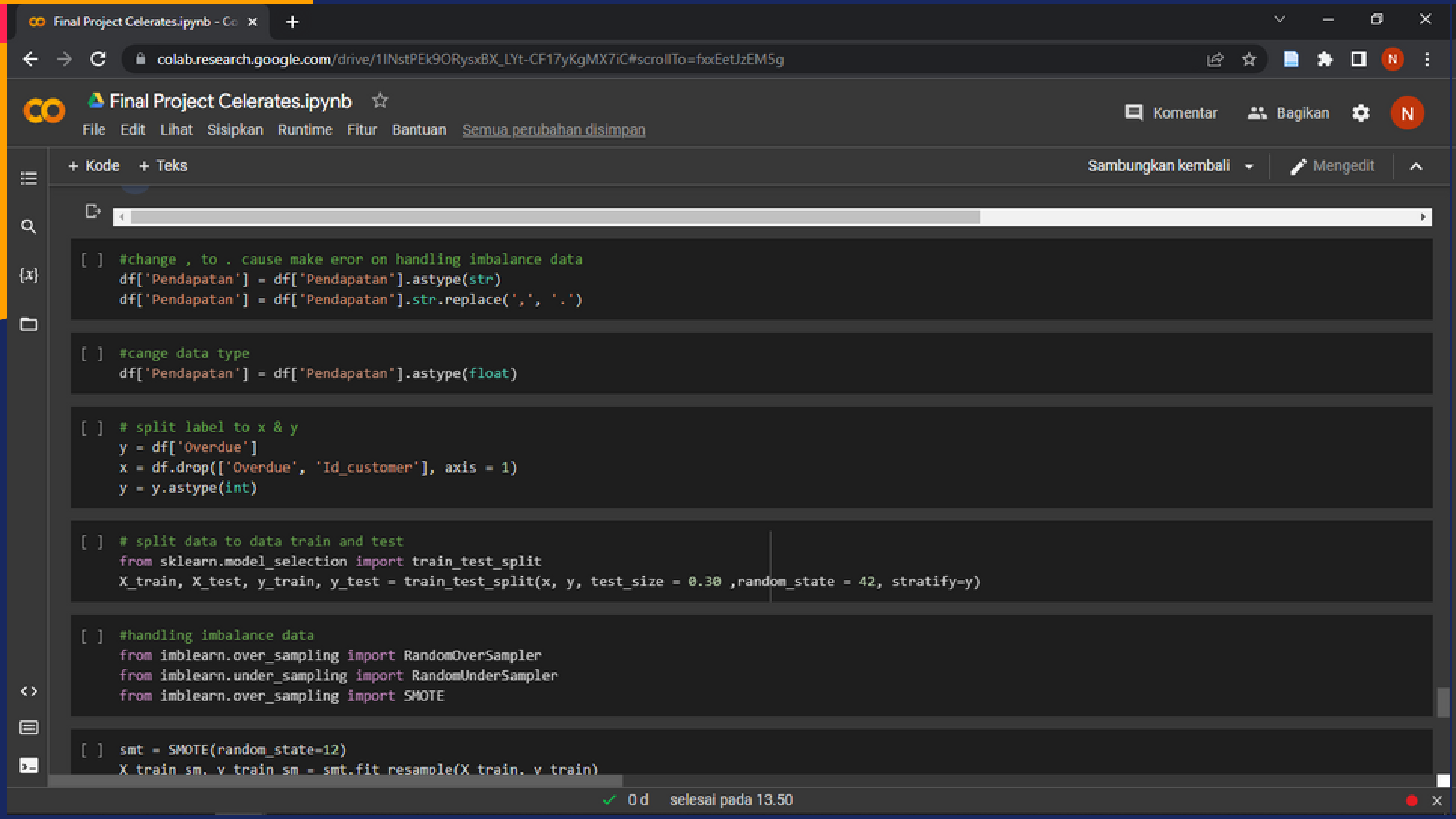


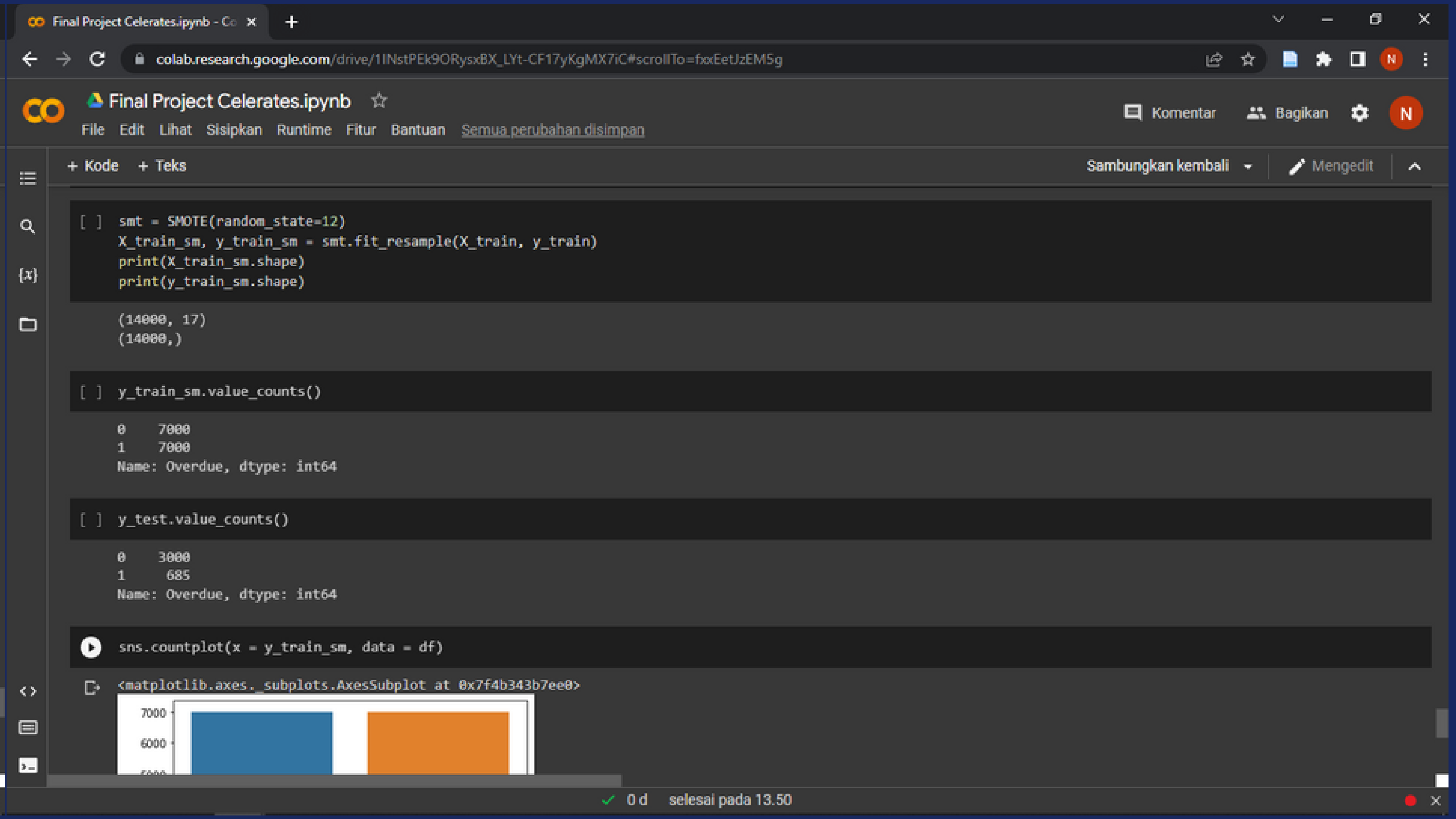
there is imbalance data in the overdue field and undersampling techniques are carried out

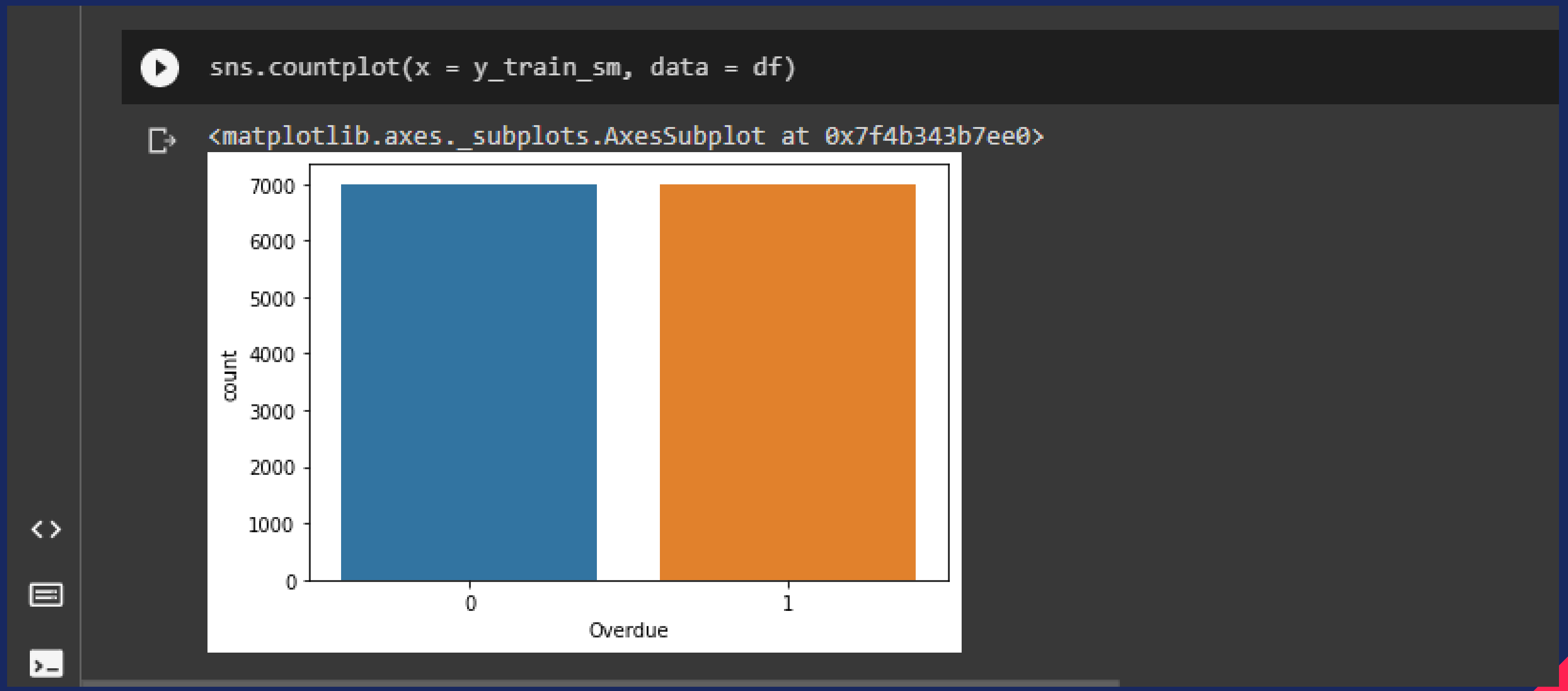




Concat 10k data with credit good scoring (0) with bad scoring (1) and encode the data before building the model







Handling oversampling with SMOTE technique.

MODELlING

- Decision Tree
- SVM
- Naive Bayes
- Random Forest
- Neural Network

IN THE FIVE MODELLING

Decision Tree

```
matrix = confusion_matrix(y_test, prediksi)
matrix
```

```
array([[2583,  417],
       [ 343,  342]])
```

```
accuracy = accuracy_score(y_test, prediksi)
accuracy
```

```
0.7937584803256446
```

```
print(classification_report(y_test, prediksi))
```

```
              precision    recall  f1-score   support

           0       0.88      0.86      0.87      3000
           1       0.45      0.50      0.47       685

    accuracy                           0.79      3685
   macro avg       0.67      0.68      0.67      3685
weighted avg       0.80      0.79      0.80      3685
```

SVM

```
confusion_matrix(y_test, prediksi1)
```

```
array([[2542,  458],
       [ 344,  341]])
```

```
accuracy_score(y_test, prediksi1)
```

```
0.7823609226594301
```

```
print(classification_report(y_test, prediksi1))
```

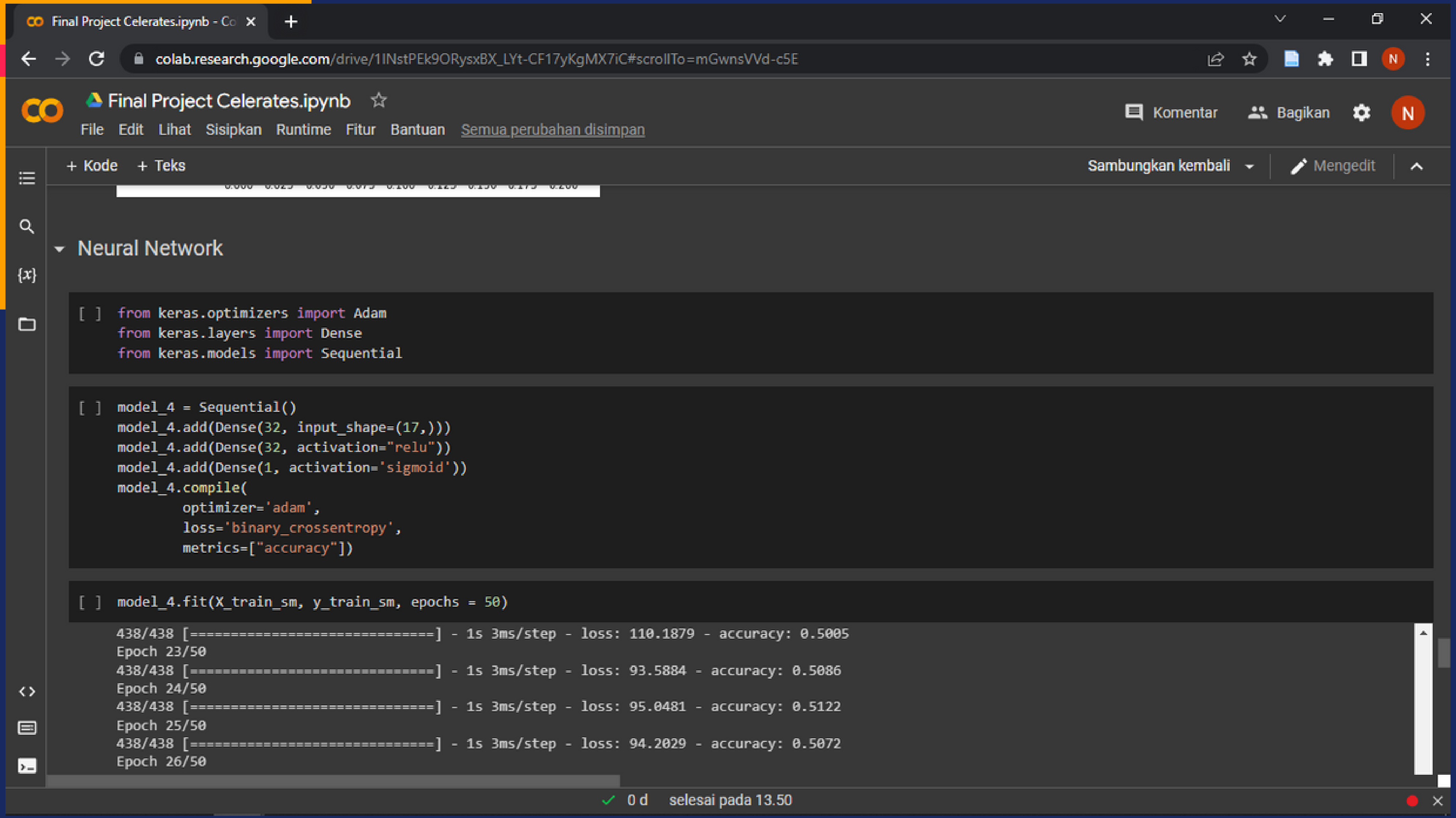
```
              precision    recall  f1-score   support

           0       0.88      0.85      0.86      3000
           1       0.43      0.50      0.46       685

    accuracy                           0.78      3685
   macro avg       0.65      0.67      0.66      3685
weighted avg       0.80      0.78      0.79      3685
```

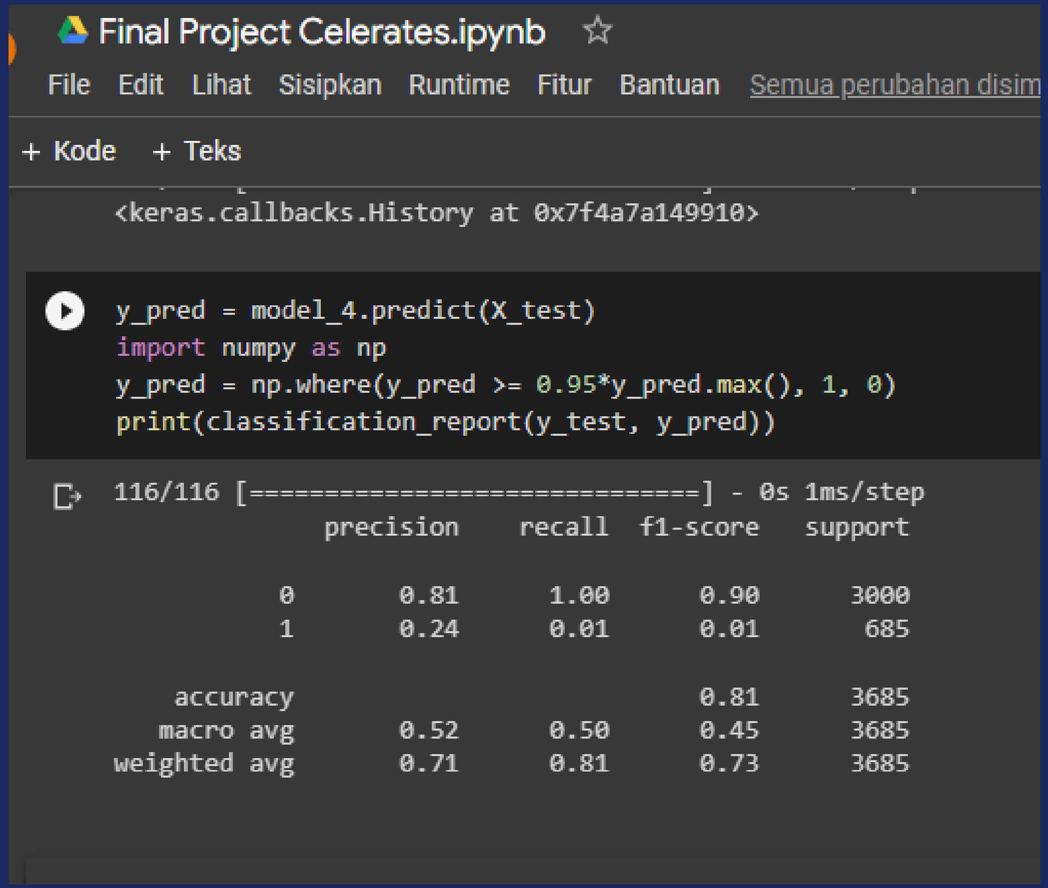
Naive Bayes

```
[ ]  confusion_matrix(y_test, prediksi2)

     array([[ 726, 2274],
            [ 136,  549]])

[ ]  accuracy_score(y_test, prediksi2)

     0.34599728629579374

[ ]  print(classification_report(y_test, prediksi2))

                   precision    recall  f1-score   support

                0       0.84      0.24      0.38      3000
                1       0.19      0.80      0.31       685

         accuracy                           0.35      3685
        macro avg       0.52      0.52      0.34      3685
     weighted avg       0.72      0.35      0.36      3685
```

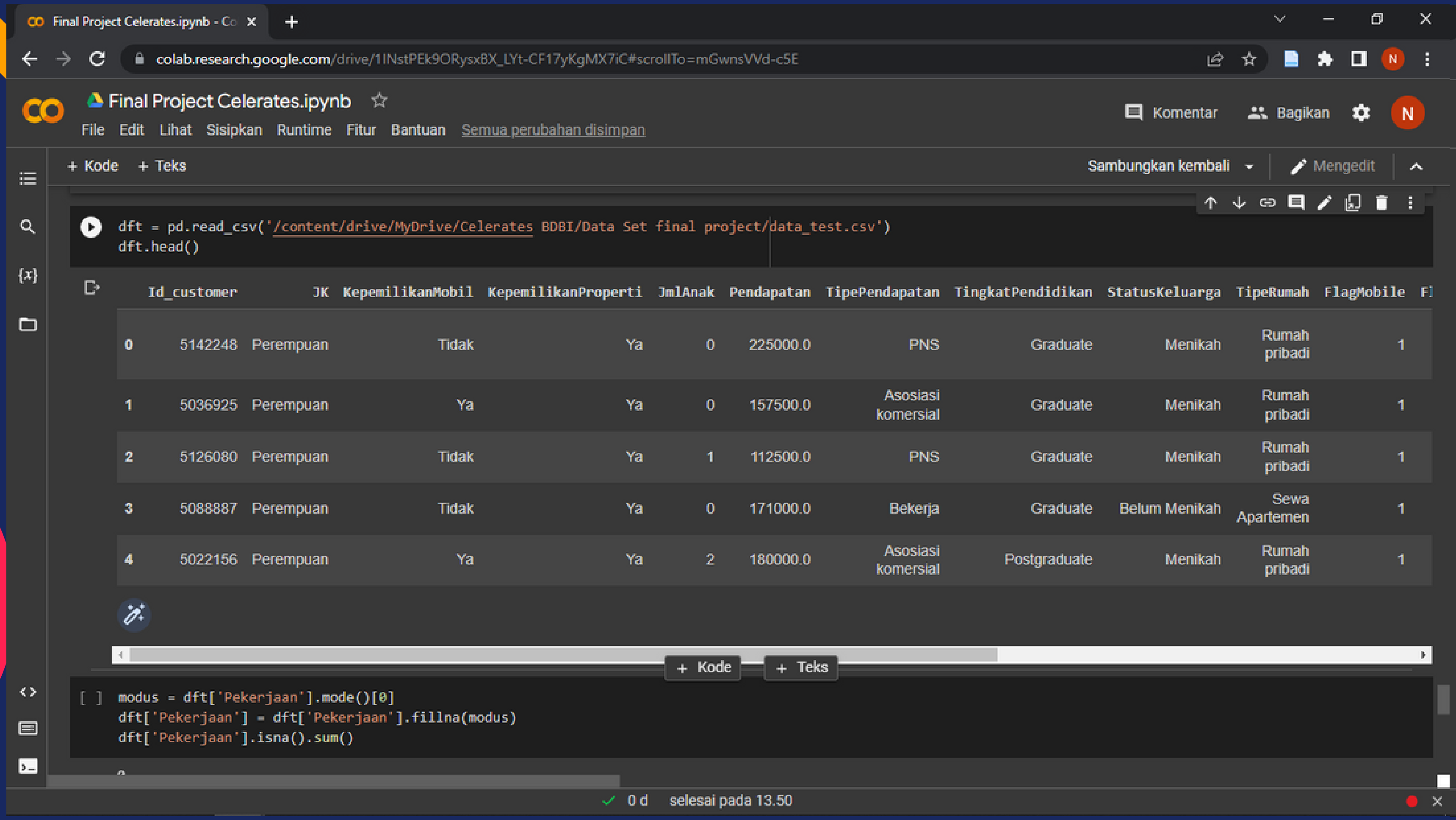
Random Forest

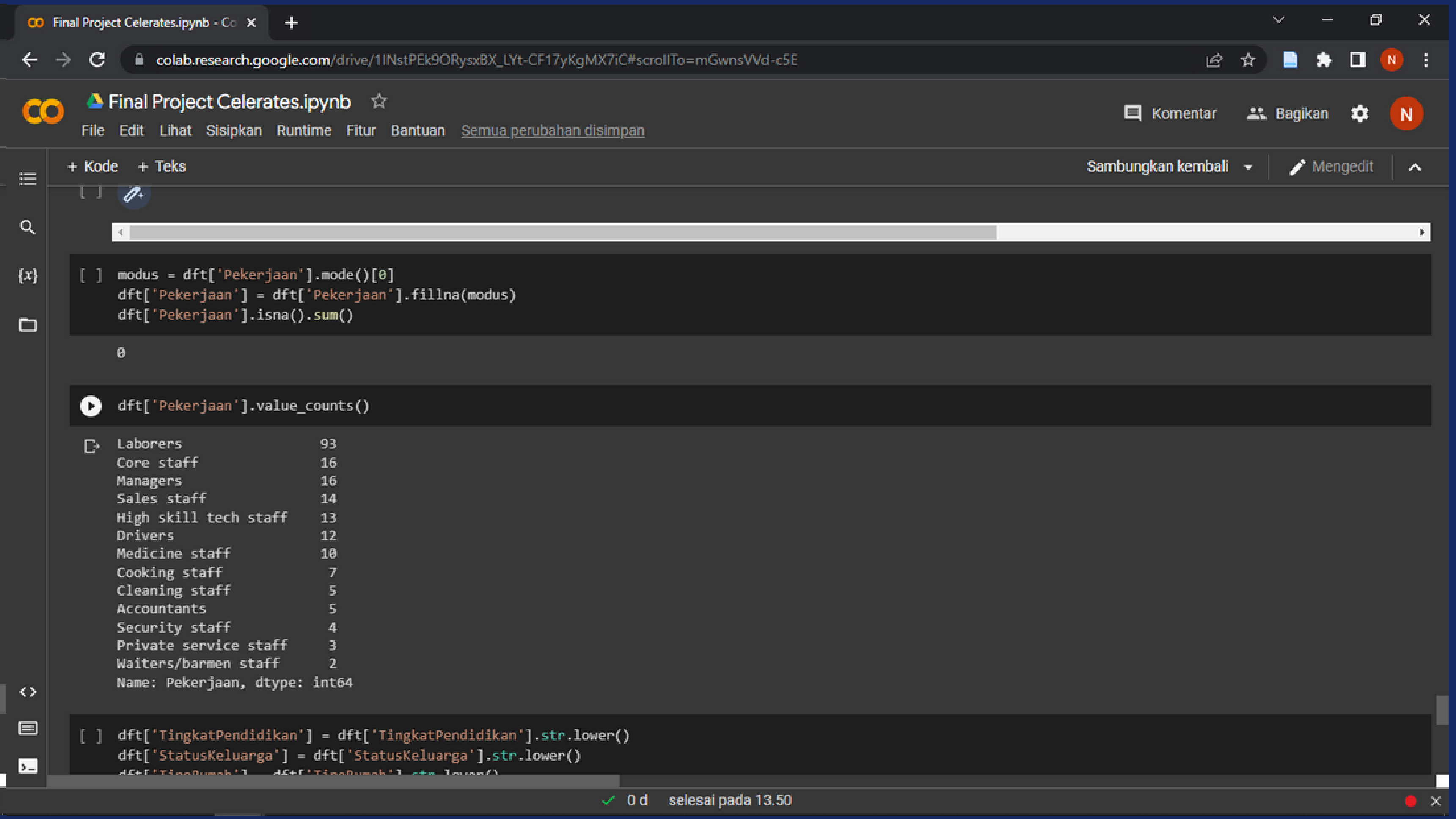
```
prediksi3 = model_3.predict(X_test)
print(prediksi3.shape)
print(y_test.shape)

(3685,)
(3685,)

[ ]  confusion_matrix(y_test, prediksi3)

     array([[2653,  347],
            [ 351,  334]])

[ ]  accuracy_score(y_test, prediksi3)

     0.8105834464043419

[ ]  print(classification_report(y_test, prediksi3))

                   precision    recall  f1-score   support

                0       0.88      0.88      0.88      3000
                1       0.49      0.49      0.49       685

         accuracy                           0.81      3685
        macro avg       0.69      0.69      0.69      3685
     weighted avg       0.81      0.81      0.81      3685
```


Final Project Celerates.ipynb
File Edit Lihat Sisipkan Runtime Fitur Bantuan Semua perubahan disimpan
Komentar
Bagikan
+ Kode + Teks
Sambungkan kembali
Mengedit
Neural Network
from keras.optimizers import Adam
from keras.layers import Dense
from keras.models import Sequential
model_4 = Sequential()
model_4.add(Dense(32, input_shape=(17,)))
model_4.add(Dense(32, activation="relu"))
model_4.add(Dense(1, activation='sigmoid'))
model_4.compile(
optimizer='adam',
loss='binary_crossentropy',
metrics=["accuracy"])
model_4.fit(X_train_sm, y_train_sm, epochs = 50)
438/438 [==============================] - 1s 3ms/step - loss: 110.1879 - accuracy: 0.5005
Epoch 23/50
438/438 [==============================] - 1s 3ms/step - loss: 93.5884 - accuracy: 0.5086
Epoch 24/50
438/438 [==============================] - 1s 3ms/step - loss: 95.0481 - accuracy: 0.5122
Epoch 25/50
438/438 [==============================] - 1s 3ms/step - loss: 94.2029 - accuracy: 0.5072
Epoch 26/50
0 d selesai pada 13.50

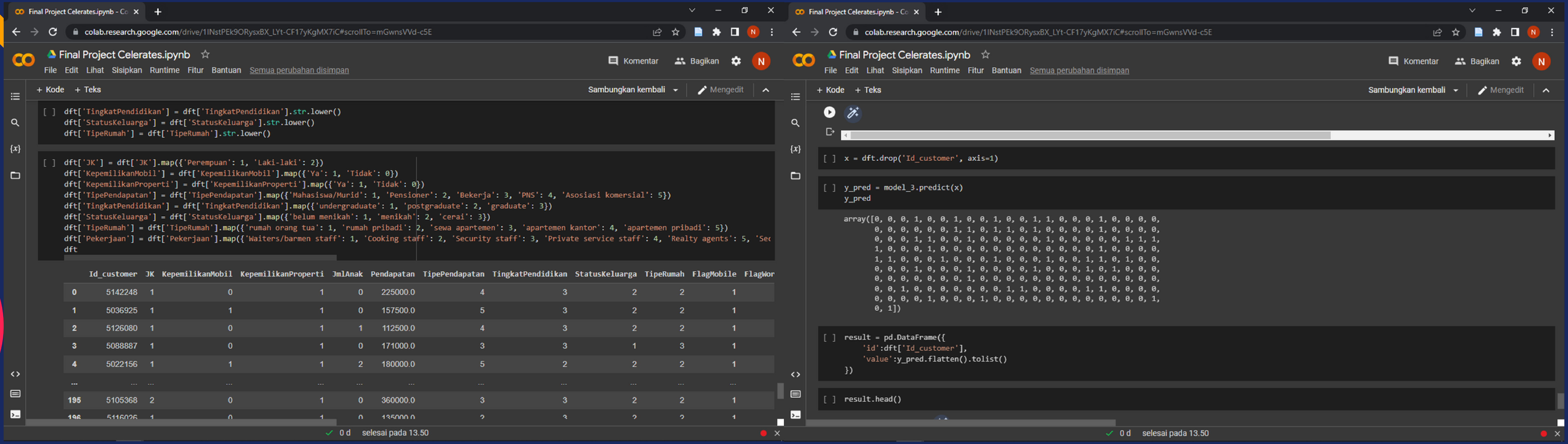

Final Project Celerates.ipynb
File Edit Lihat Sisipkan Runtime Fitur Bantuan Semua perubahan disim
+ Kode + Teks
<keras.callbacks.History at 0x7f4a7a149910>
y_pred = model_4.predict(X_test)
import numpy as np
y_pred = np.where(y_pred >= 0.95*y_pred.max(), 1, 0)
print(classification_report(y_test, y_pred))
116/116 [==============================] - 0s 1ms/step
precision recall f1-score support
0 0.81 1.00 0.90 3000
1 0.24 0.01 0.01 685
accuracy 0.81 3685
macro avg 0.52 0.50 0.45 3685
weighted avg 0.71 0.81 0.73 3685

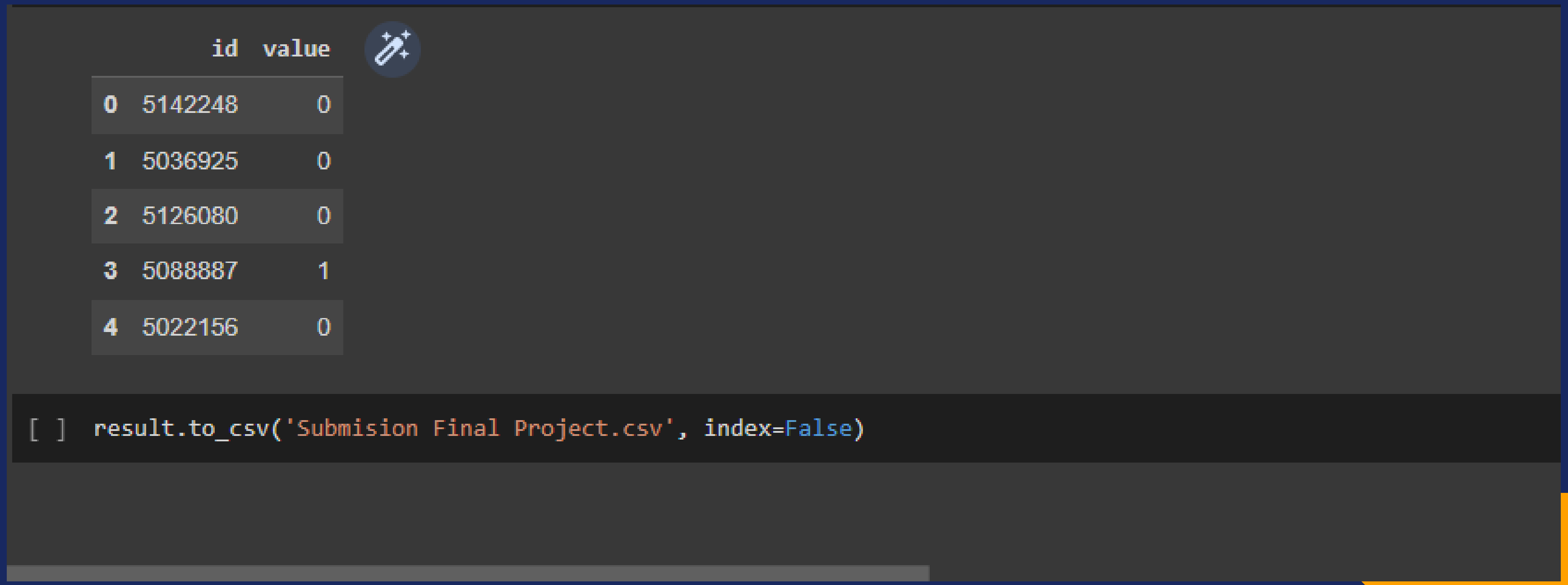
Data Test Prediction





Read data and fill null data on pekerjaan





Encode the data test and predict with model_3 (Random Forest) and export the result.

Credit Risk Anlaysis

Annual income
26100 6750000

Income category (All)

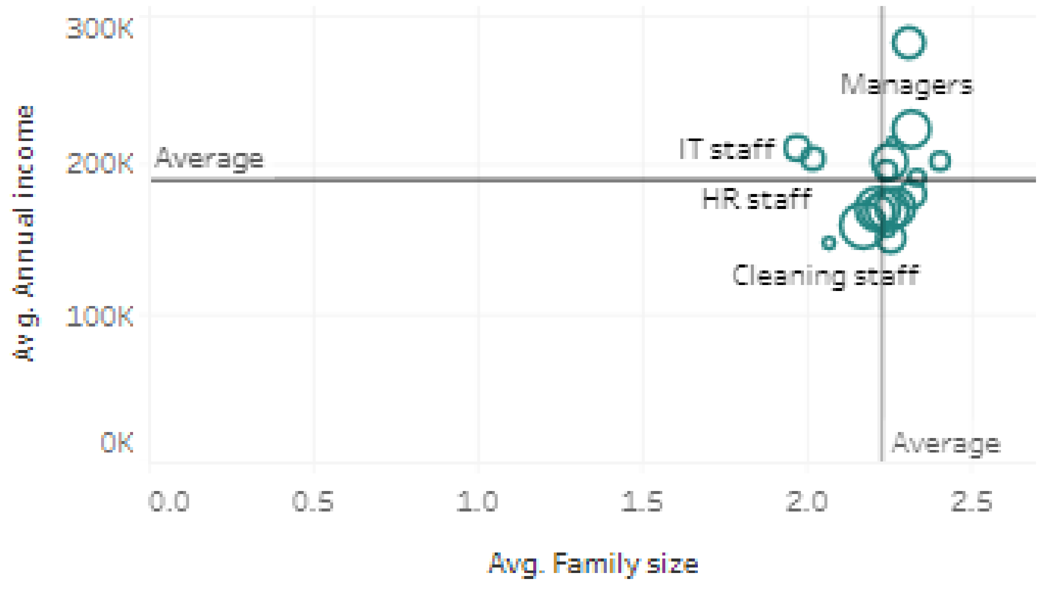
Gender (All)

Credit Status (All)

Total Client Number	Avg Annual Income	Avg Family Size
438,310	187,492	2.194

Reset Filters

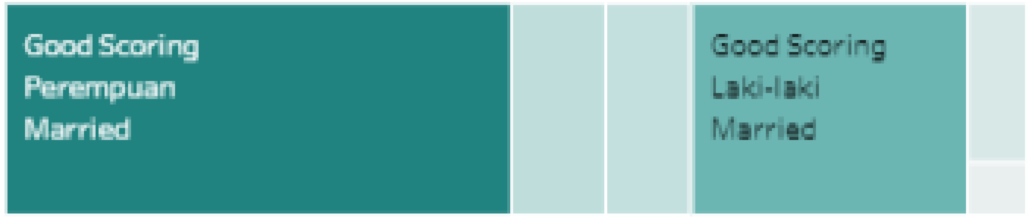
Click Here!



Regulatory Understanding

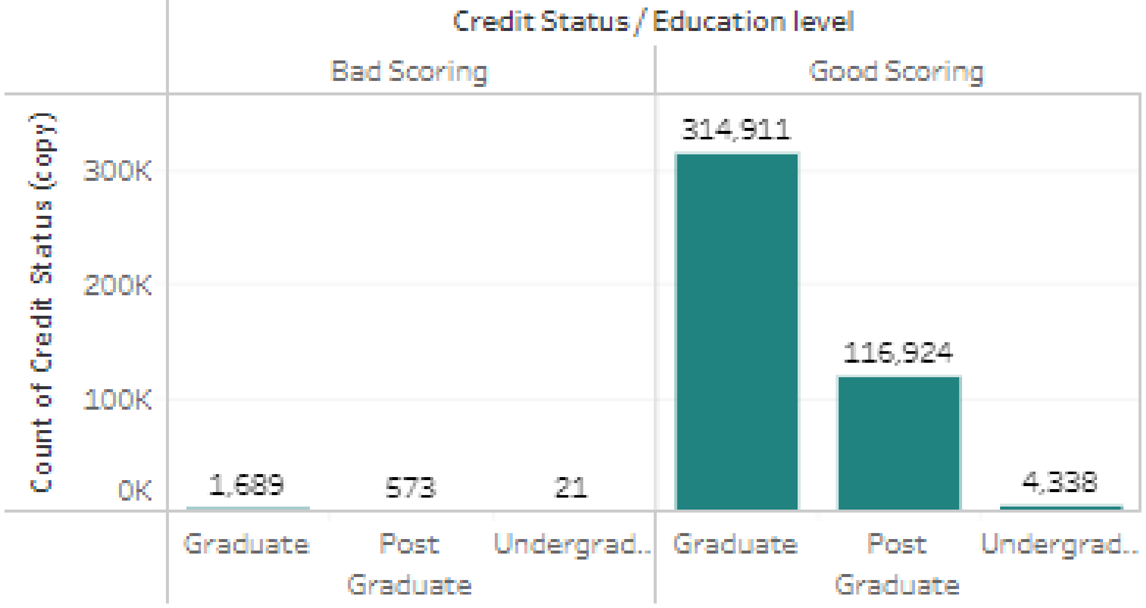
In granting credit, the systems and procedures for giving attention are not only considered, but also the management of credit. Education has a significant effect on how customers understand the contents of existing regulations.

So before giving credit, an understanding is made in advance so that credit does not experience problems from either the bank or the customer.



Problem Credit

Providing working credit capital is a bank business with a high risk. The risk in credit is the occurrence of credit problems so the background of the customer needs to be considered, such as average income and number of families. Which is a factor in whether the customer can return the credit or not.



Track Record

The provision of existing credit is also one of the considerations of whether the customer can repay the credit (good score) or not (bad score). Men and women have their own needs and status and also have relationships that can affect responsibility. The principles that must be applied in granting credit are the 5 C principles, namely character, capacity, capital, economic conditions, and collateral.

Thank You..

FOR INQUIRIES AND CONCERNS

NEXT DASHBOARD ON THIS LINK

https://public.tableau.com/views/FinalProjectCelerates_16720424300480/Dashboard1?:language=en-US&:display_count=n&:origin=viz_share_link