Laporan Kerja Individu

Final Project - Data Analytics

Program Zenius Studi Independen Bersertifikat - Angkatan 4

Nama Lengkap: Himma Faicha Hubbiya
Nomor ID live class: 136
Nomor Kelompok: 12
Mentor: Muhammad Verly

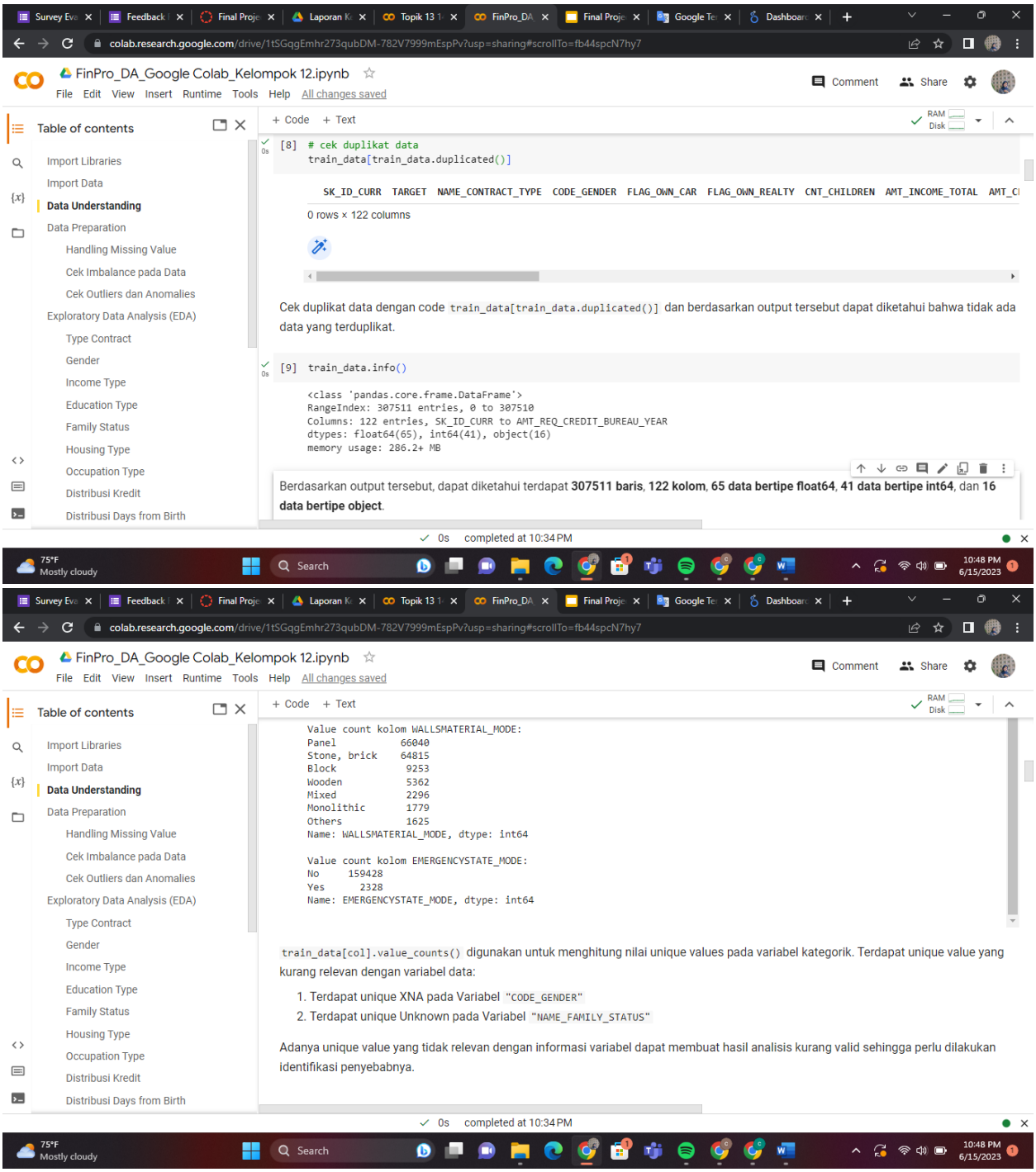
Deskripsi peran

- Berpartisipasi dalam diskusi kelompok.
- Membantu dalam Data Understanding dan menambahkan summary.
- Membantu dalam Visualisasi Missing Value dan menambahkan summary.
- Membantu dalam Handling Missing Value pada data inrelevan dan menambahkan summary.
- Membantu dalam Outlier and Anomalies Checking dan menambahkan summary.
- Membantu dalam Exploratory Data Analysis (EDA) dan menambahkan summary
- Membantu menyusun PPT
- Membuat link Dashboard Looker Studio 2
- Membuat dan mengatur susunan data pada dashboard.

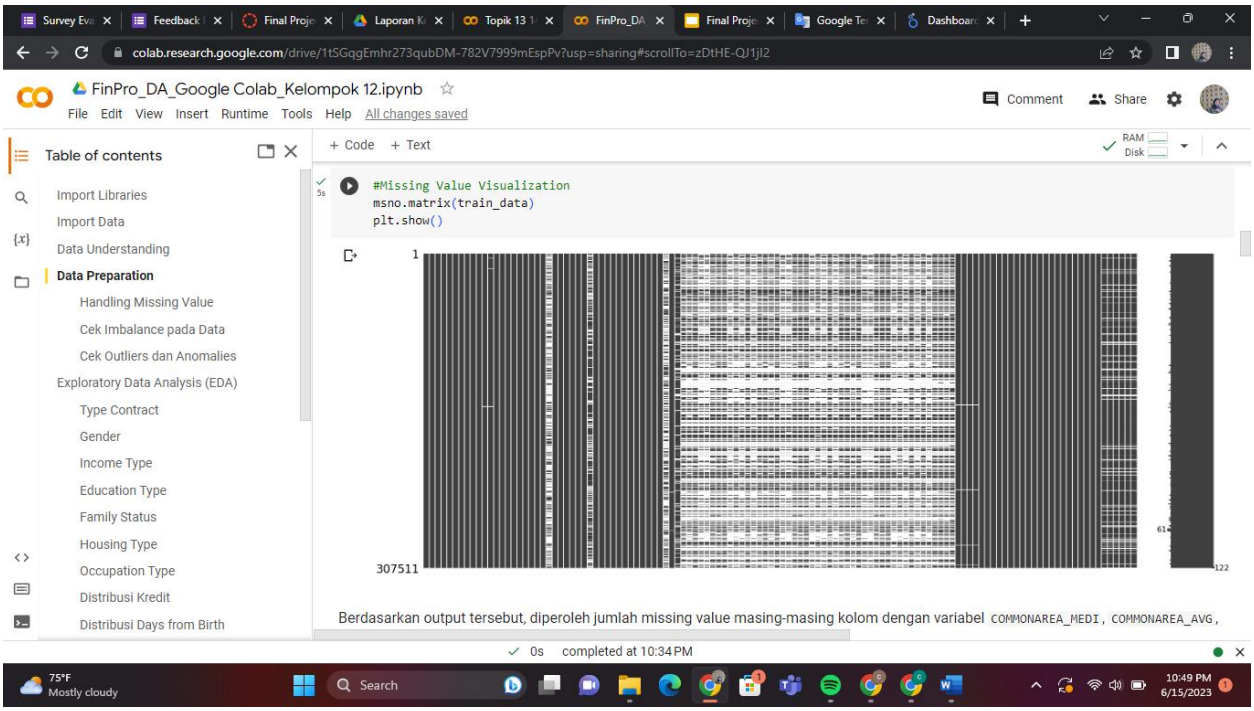
Lampiran Hasil Kerja

1. Link Google Colab :
   https://colab.research.google.com/drive/1tSGqgEmhr273qubDM-782V7999mEspPv?usp=sharing

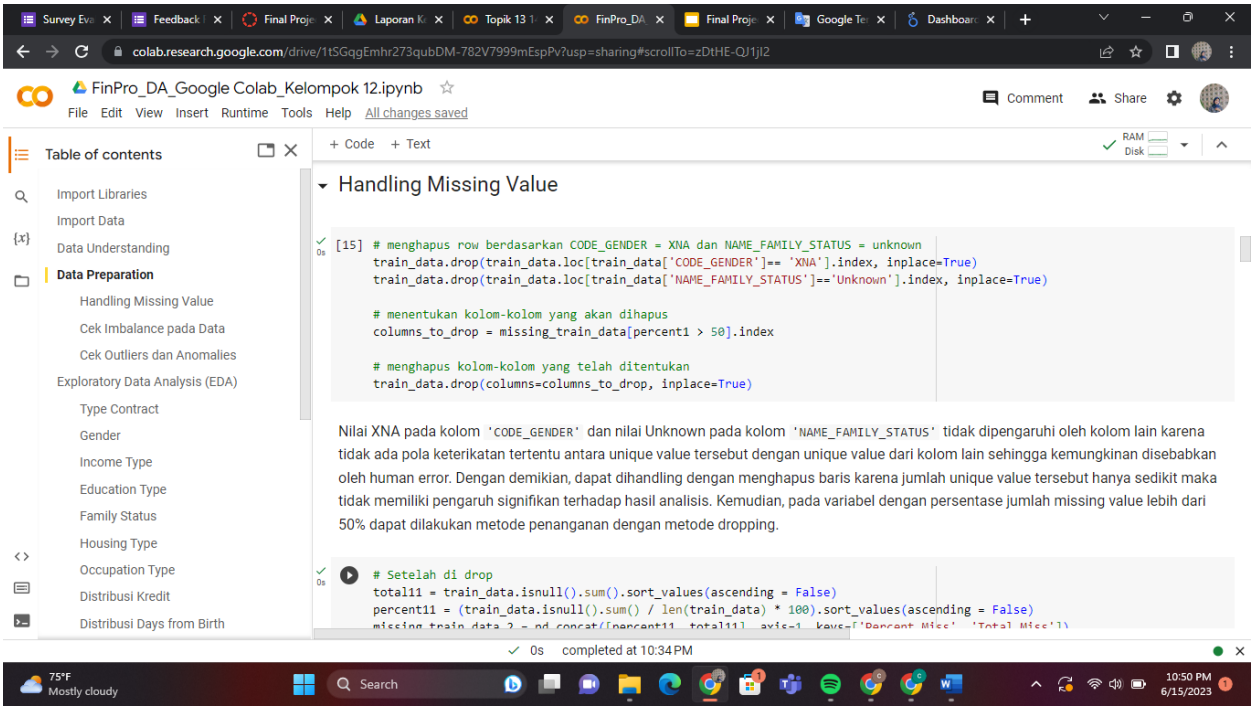
- Membantu dalam Data Understanding dan menambahkan summary.



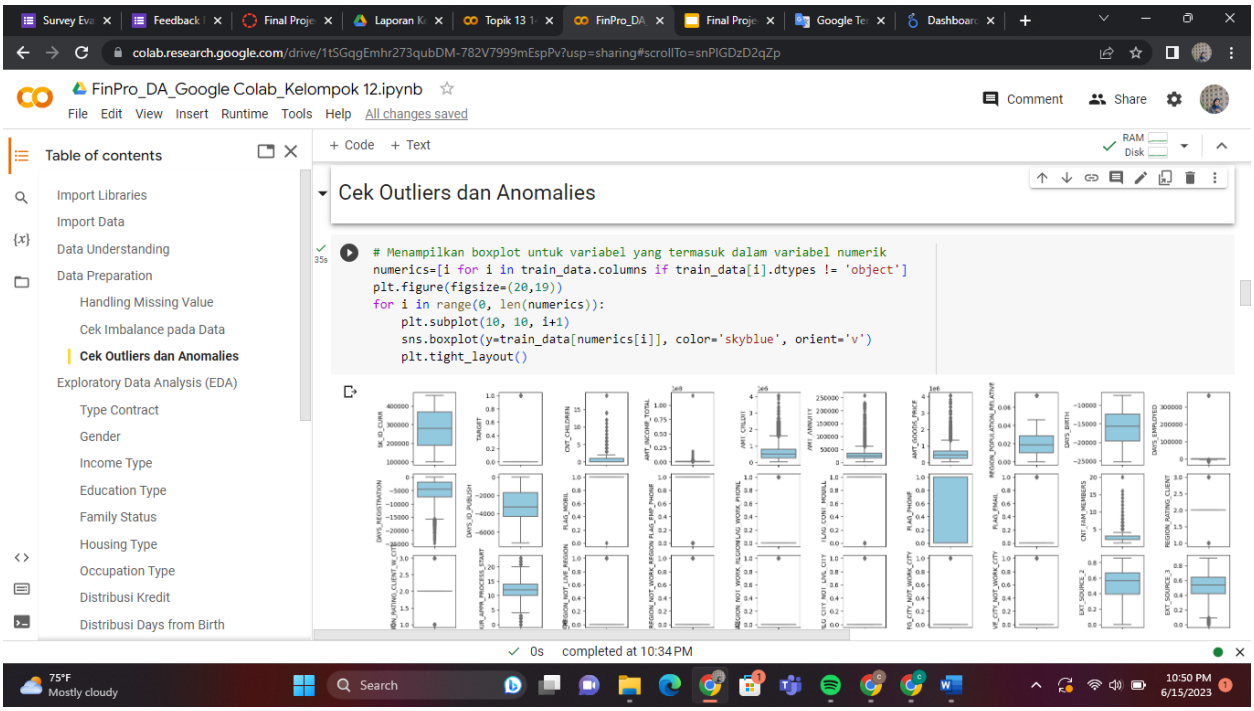
- Membantu dalam Visualisasi Missing Value dan menambahkan summary.



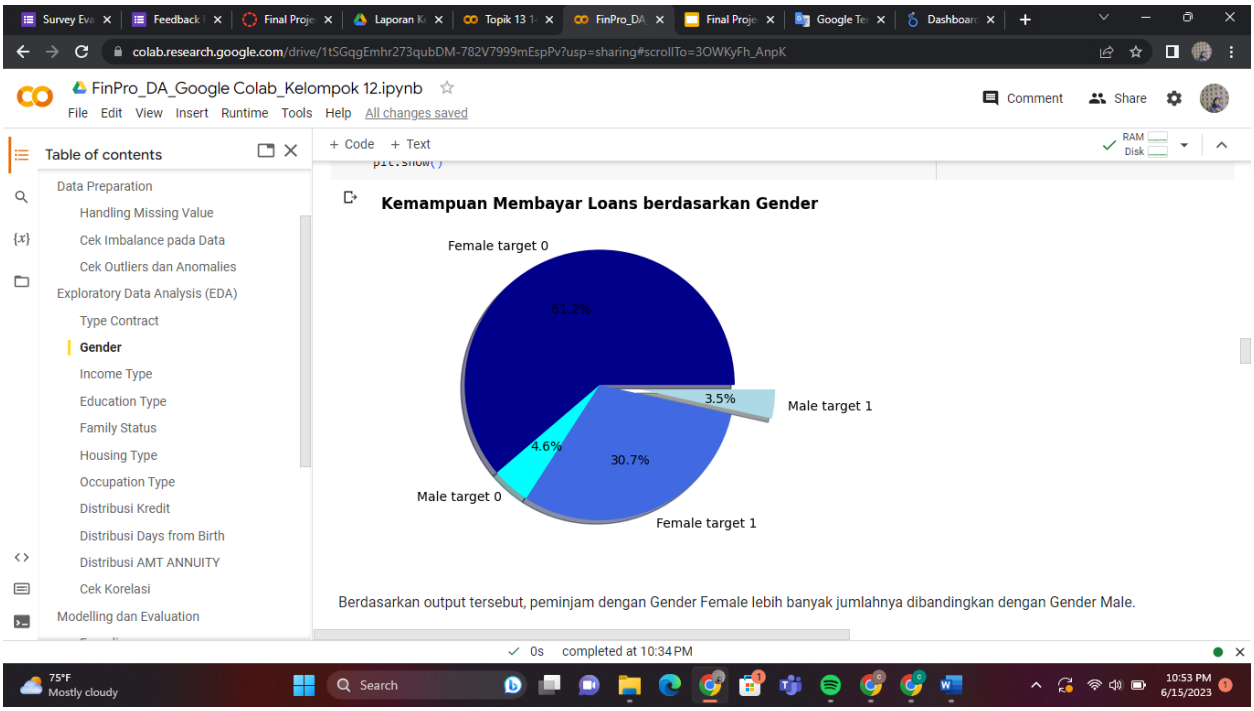
- Membantu dalam Handling Missing Value pada data inrelevan dan menambahkan summary.

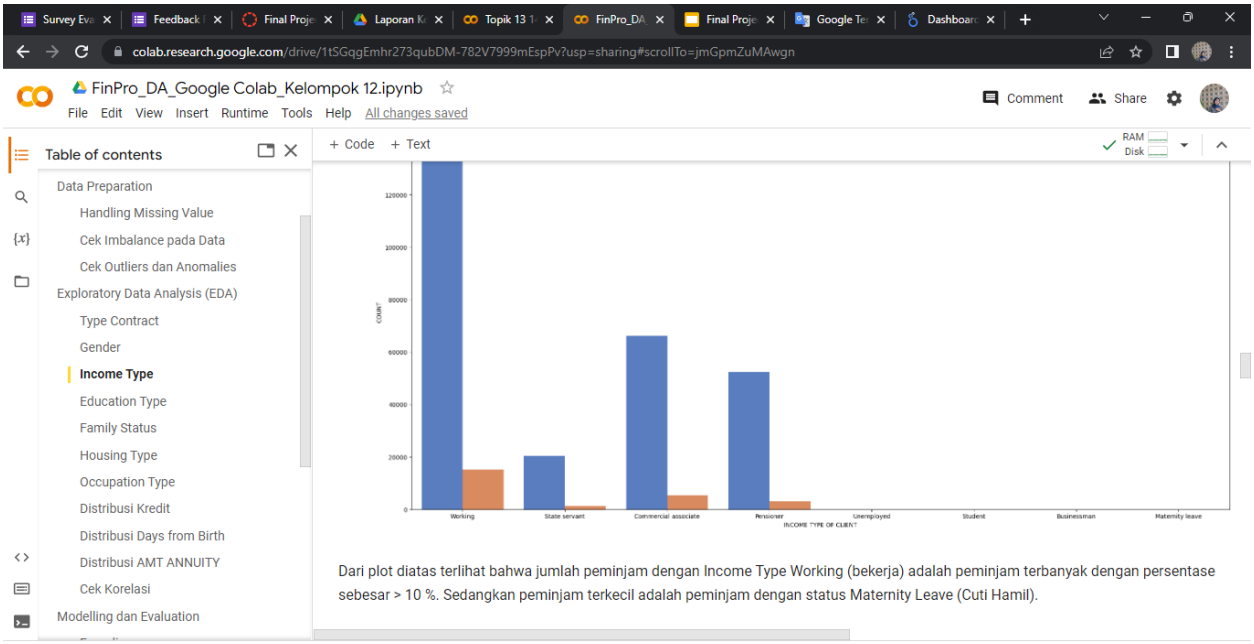


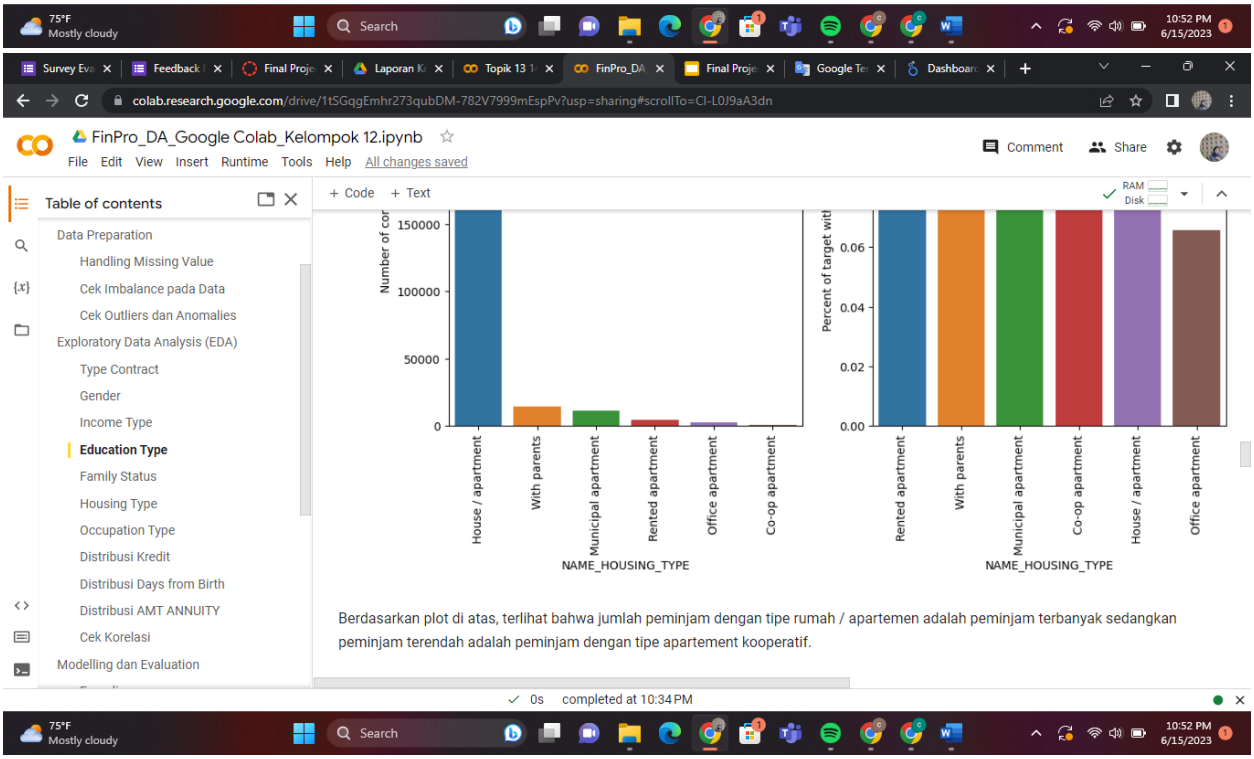
- Membantu dalam Outlier and Anomalies Checking dan menambahkan summary.



- Membantu dalam Exploratory Data Analysis (EDA) dan menambahkan summary

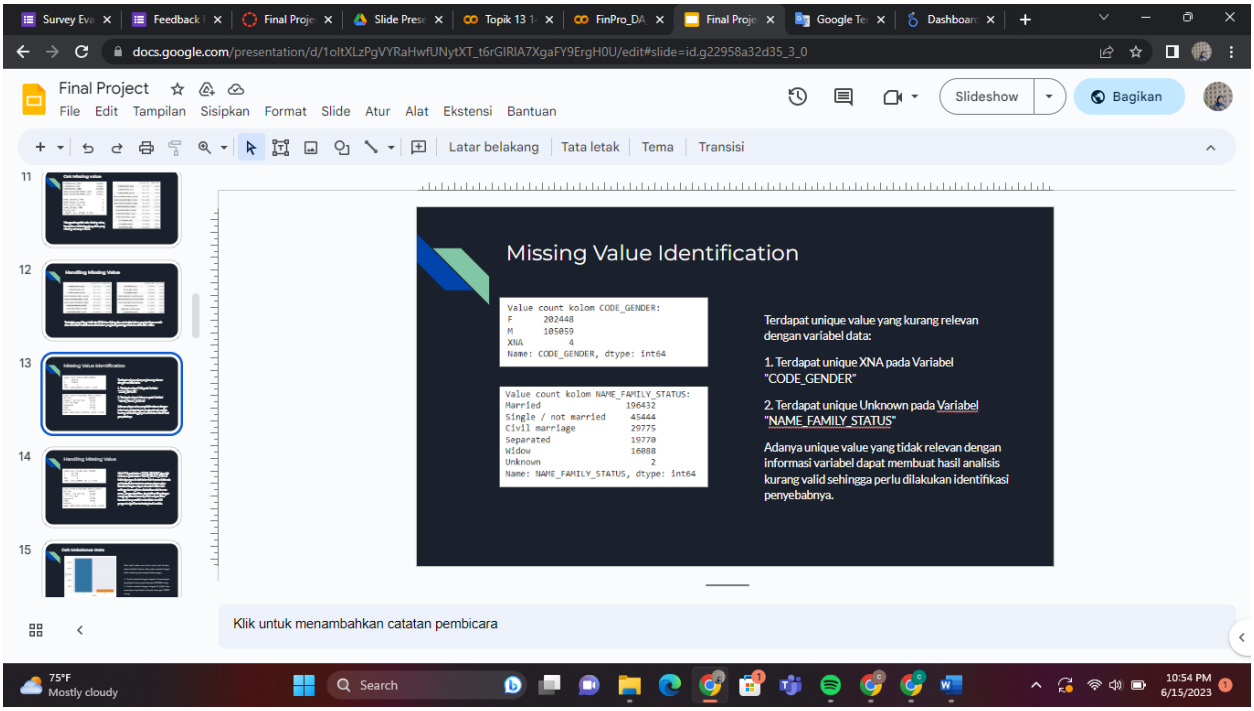






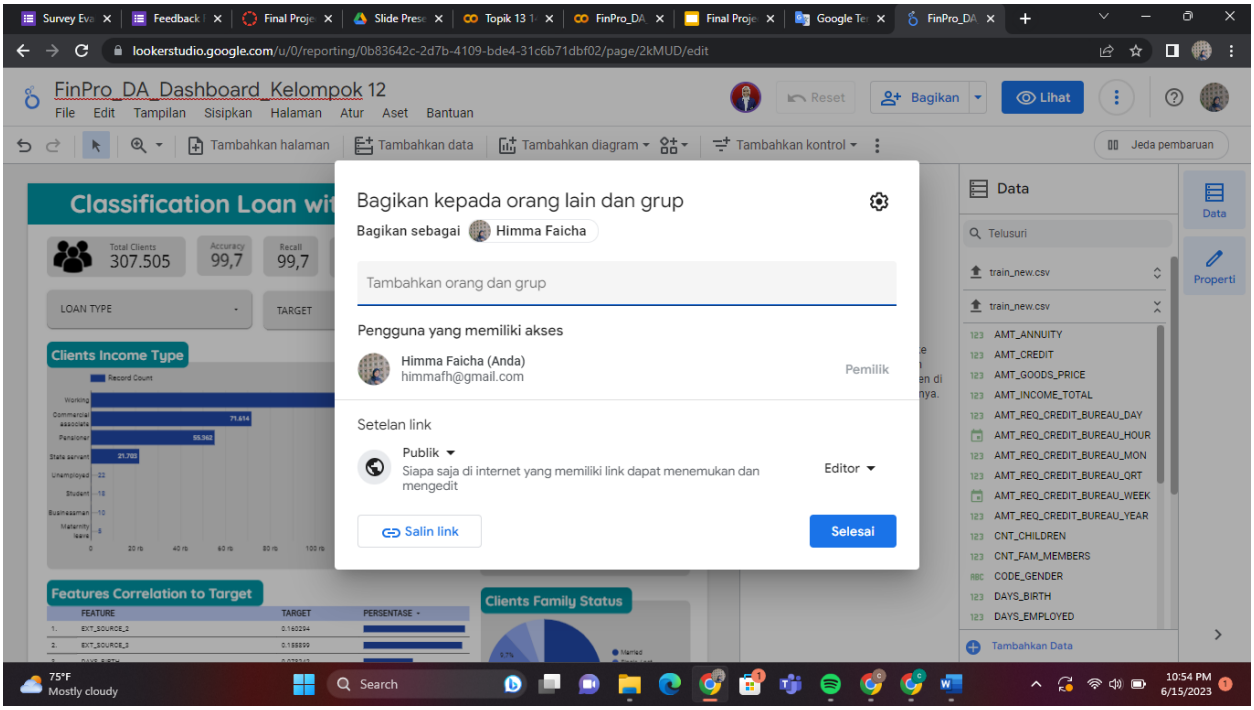
2. Link PPT : https://docs.google.com/presentation/d/1oItXLzPgVYRaHwfUNytXT_t6rGIRIA7XgaFY9ErgH0U/edit?usp=sharing

- Membantu menyusun PPT

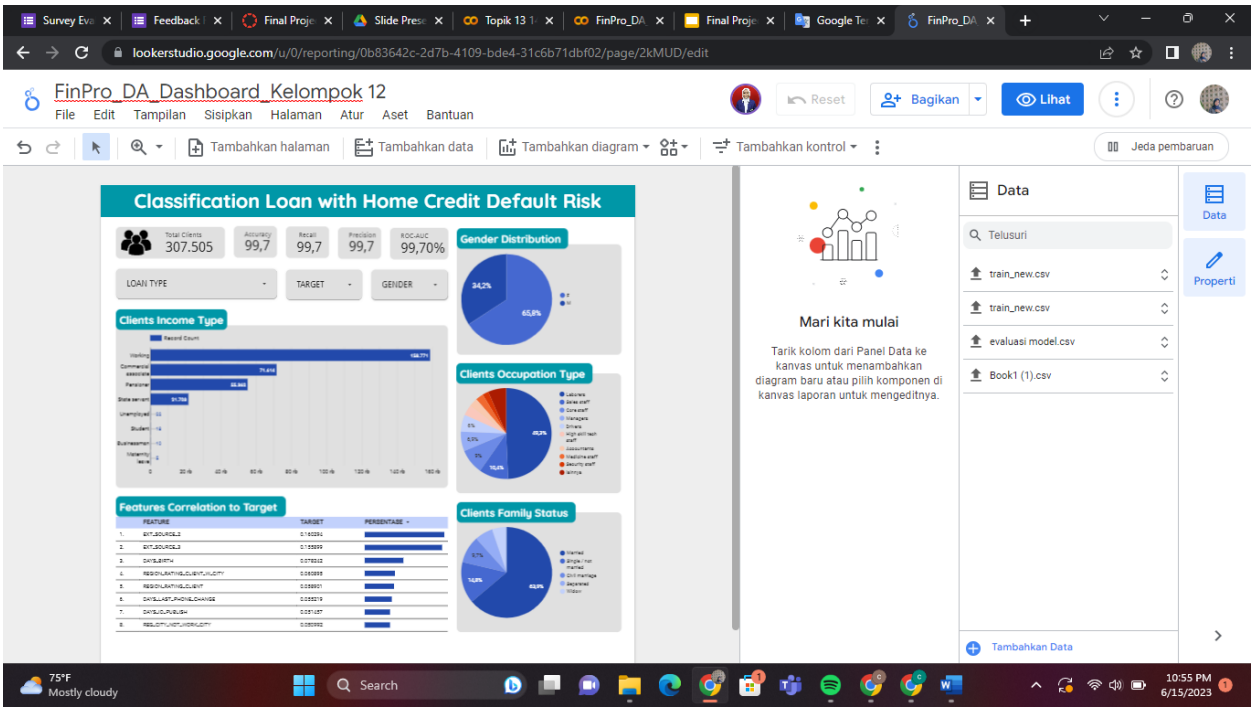


3. Link Dashboard : https://lookerstudio.google.com/u/0/reporting/0b83642c-2d7b-4109-bde4-31c6b71dbf02/page/2kMUD

- Membuat link Dashboard Looker Studio 2



- Membuat dan mengatur susunan data pada dashboard.



Proses Kerja

Tugas Final Project PZIB dikerjakan secara berkelompok. Dalam tugas final project, proses saya dalam bekerja adalah sebagai berikut:

1. Saya turut berperan aktif dalam diskusi kelompok. Dalam diskusi kelompok, pembahasan Final Project dimulai dengan pembahasahan terkait tahapan-tahapan CRISP-DM. Kemudian dilanjutkan dengan pembahasan terkait Bussiness Understanding dengan memahami latar belakang dataset Home Credit Default Risk, tujuan dari analisis data tersebut, dan permasalahan serta solusi bagi permasalahan tersebut. Analisis data dari dataset tersebut dapat menentukan kelayakan kredit dari calon peminjam dengan cepat dan akurat.
2. Selanjutnya dilakukan tahap Data Understanding. Setelah diskusi, kelompok kami sepakat untuk menggunakan dataset application train sebagai dataset utama. Pada Data Understanding, saya melakukan pengecekan terhadap ukuran dataset yang digunakan dan pengecekan duplikat data, serta menambahkan summary pada Data Understanding agar memudahkan untuk membaca output hasil coding. Dari Data Understanding, dapat diketahui bahwa terdapat 307511 baris, 122 kolom, 65 data bertipe float64, 41 data bertipe int64, dan 16 data bertipe object. Kemudian data tersebut saya pisahkan menjadi variabel numerik dan kategorik, serta dilakukan value count pada data kategorik. Setelah dilakukan value count pada data kategorik, ditemukan nilai unique value yang kurang relevan dengan variabel data yaitu nilai XNA pada variabel "CODE_GENDER" dan nilai Unknown

pada variabel "NAME_FAMILY_STATUS" yang dapat membuat hasil analisis kurang valid sehingga perlu dilakukan identifikasi penyebabnya.

3. Setelah Data Understanding, dilakukan Data Preparation untuk membuat data mentah menjadi data yang layak untuk dianalisis. Pada tahap ini, saya membantu dalam pengecekan dan visualisasi Missing Value, serta handling Missing Value berupa menghapus baris yang terdapat nilai XNA pada kolom "CODE_GENDER" dan nilai Unknown pada variabel "NAME_FAMILY_STATUS" karena tidak ada pola keterikatan tertentu antara unique value kolom lain serta jumlahnya yang sedikit. Selain itu, saya juga membantu dalam pengecekan outliers dengan menggunakan metode Boxplot.
4. Selanjutnya tahap Exploratory Data Analysis (EDA) yang bertujuan untuk visualisasi data kategorik dan numerik, serta persebaran data. Dalam tahap EDA, saya membantu membuat pie chart dan barplot untuk melihat persentase dan persebaran data pada variabel berdasarkan target, yaitu berdasarkan tingkat kesulitan pembayaran klien.
5. Setelah EDA, dilakukan tahap Modelling dan Evaluasi. Pada tahap ini, dilakukan encoding dan feature selection dengan menggunakan variabel yang memiliki korelasi tinggi terhadap target. Selain itu, dilakukan handling unbalanced data serta splitting data train menjadi data test dan data train. Pada tahap modelling digunakan 4 percobaan model yaitu, metode Logistic Regression, Random Forest Classifier, Decision Tree, dan KNN. Berdasarkan perbandingan keakuratan, nilai recall, f-score, ROC-AUC, dan *Confussion Matrix* diperoleh hasil bahwa kemungkinan model terbaik adalah model Random Forest Classifier.
6. Setelah proses olah data selesai, selanjutnya dilakukan pembuatan Dashboard pada Looker Studio. Saya membuat link Looker Studio 2 untuk membuat dashboard yang dikerjakan secara bersama. Pada awalnya pembuatan dashboard sudah dilakukan oleh teman saya tetapi karena ada beberapa bagian yang sulit untuk diperbaiki maka saya membuat link Looker Studio yang baru. Pada pembuatan dashboard ini, saya mendownload data yang sudah dibersihkan dan mengunggahnya ke dashboard. Saya menyusun tata letak dan design dari setiap diagram yang ada di Looker Studio, serta menambahkan beberapa feature yang diperlukan.
7. Saya turut berperan dalam proses pembuatan Slide Presentation dengan memasukkan output dari Google Colab dan menambahkan interpretasi.

Persetujuan Anggota

Laporan ini telah disetujui oleh:

- ☑ Adinda Permata Sari
- ☑ Ardia Fatma Sari
- ☑ Raihanah Assa’adah
- ☑ Raihan Tsabita Sabil