



Home Credit Scorecard

RAKAMIN ACADEMY X HOME CREDIT INDONESIA

Ongky Arif Wibowo

Github :

[https://github.com/mamasarif1003/
HomeCreditIndonesiaProjects](https://github.com/mamasarif1003/HomeCreditIndonesiaProjects)



BACKGROUND RESEARCH

Home Credit Indonesia ingin membuat machine learning untuk membantu tim menentukan apakah pengajuan pinjaman dari pelanggan akan mengalami kendala dalam proses pelunasan kredit atau tidak.

Dari data yang ada, Home Credit Indonesia ingin mengetahui apa saja kriteria pelanggan yang tidak mengalami kendala dalam proses pelunasan kredit sehingga dapat membantu meningkatkan pendapatan.

DATASET OVERVIEW

df.shape
✓ 0.0s
(303239, 153)

TARGET	Customers	Customers_pct
0	0	278843
1	1	24396

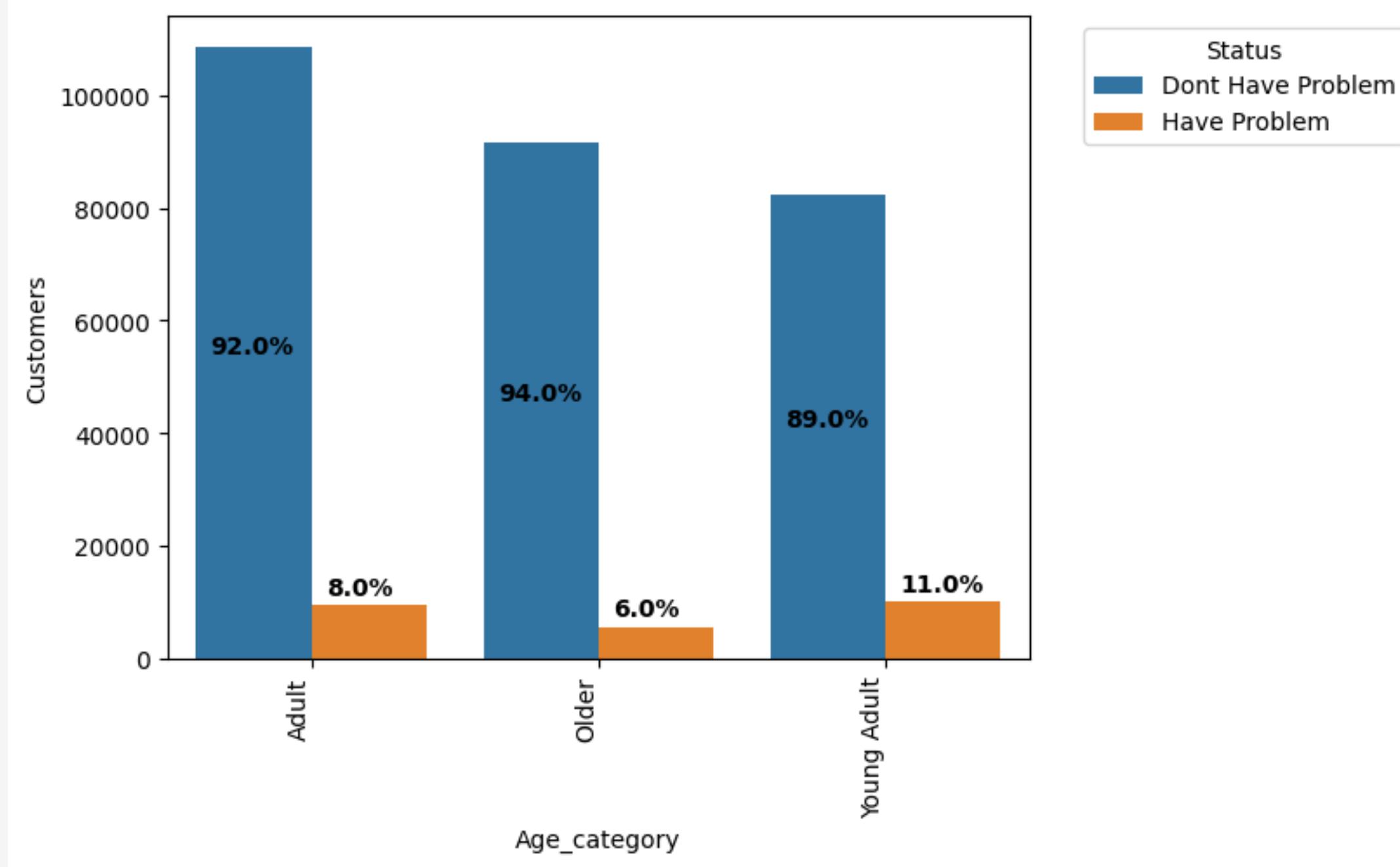
df.shape
✓ 0.1s
(307511, 122)

Data **disediakan oleh tim Home Credit Indonesia**, dengan dimensi 122 kolom dan 307+ ribu baris untuk data train. Berisikan data pelanggan yang bermasalah dan tidak bermasalah dalam melunasi pinjaman yang diberikan oleh Home Credit Indonesia dengan **perbandingan 8% : 92%**.

Pada data test memiliki dimensi **153 kolom** dengan banyak fitur untuk dilakukan prediksi dalam menentukan apakah pelanggan tersebut bermasalah atau tidak.

Business Insight OVERVIEW

Age Category of Customers

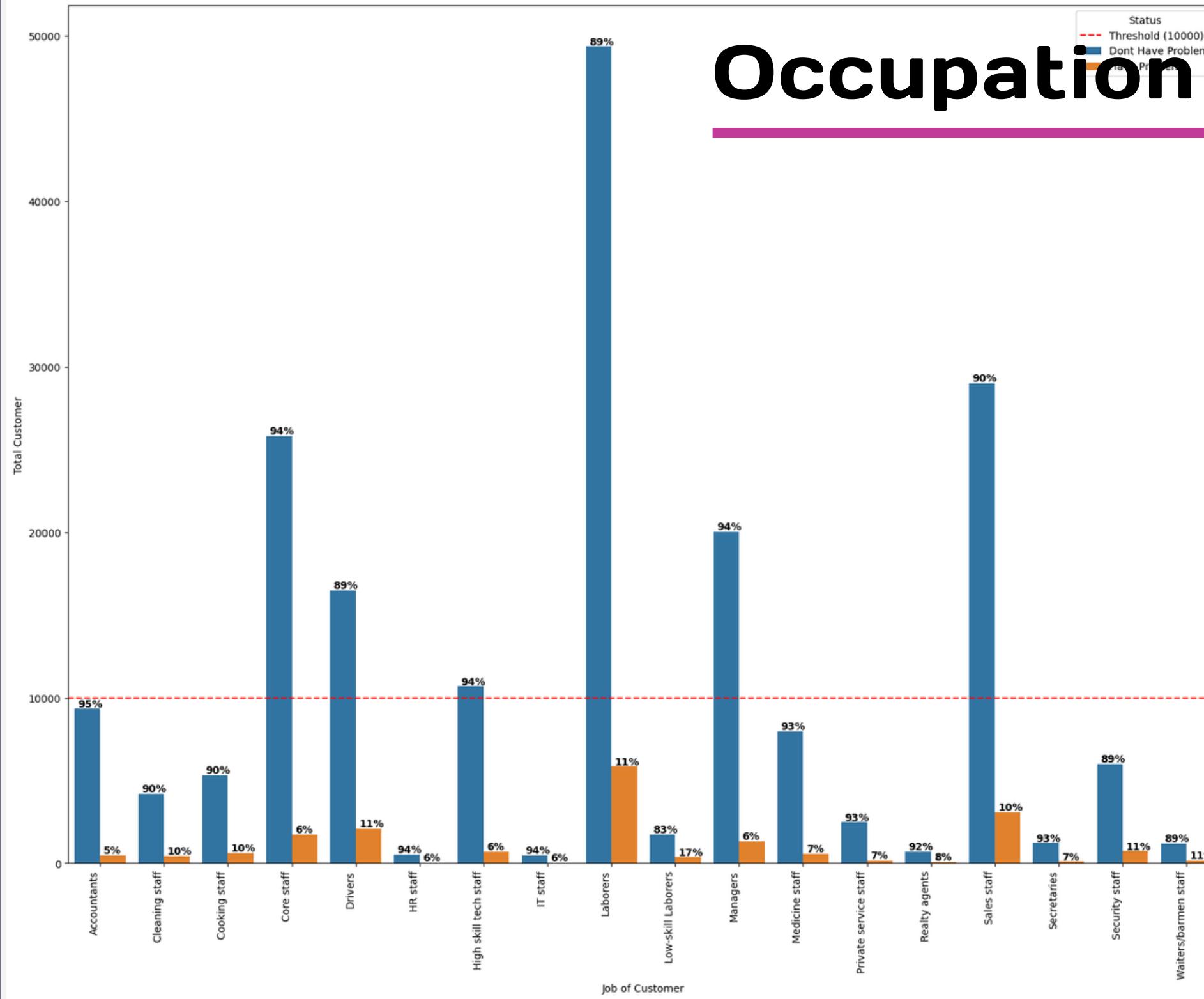


Terlihat dari data bahwa customer pada **kategori adult** dengan range umur 35-50 tahun, **sangat mendominasi** dari keseluruhan jumlah customer.

Data juga menunjukkan pada kategori **adult dan older memiliki kecenderungan untuk tidak bermasalah dalam pembayaran credit** lebih tinggi.

Business Insight OVERVIEW

Occupation Type of Customers



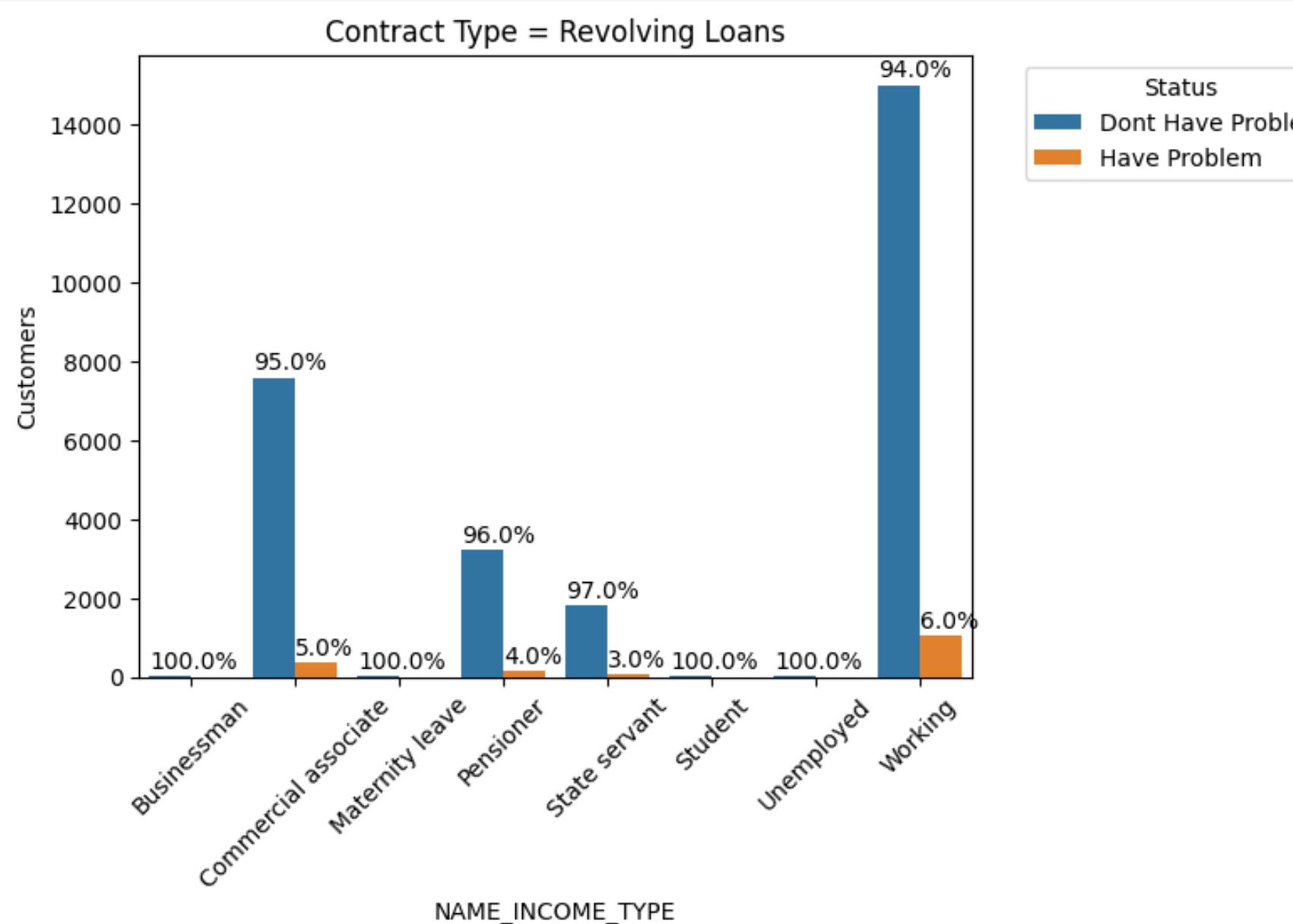
Dari data grafik yang didapat ketika mencari informasi berdasarkan tipe pekerjaan mereka, rata-rata menunjukkan bahwa para kustomer 90% keatas tidak bermasalah.

Namun terdapat beberapa data yang menarik dimana pada kriteria yang baik dan diatas rata-rata bahwa mereka tidak bermasalah menunjukkan jumlah kustomer yang sedikit. Data tersebut meliputi 'HR staff', 'IT staff', 'Realty agents', 'Accountants', 'Cleaning Staff', 'Cooking Staff', 'Medicine Staff', dan 'Secretaris'.

Sehingga dari hal ini diperlukan beberapa improvement market agar dapat menjangkau kustomer dengan tipe pekerjaan tersebut.

Business Insight OVERVIEW

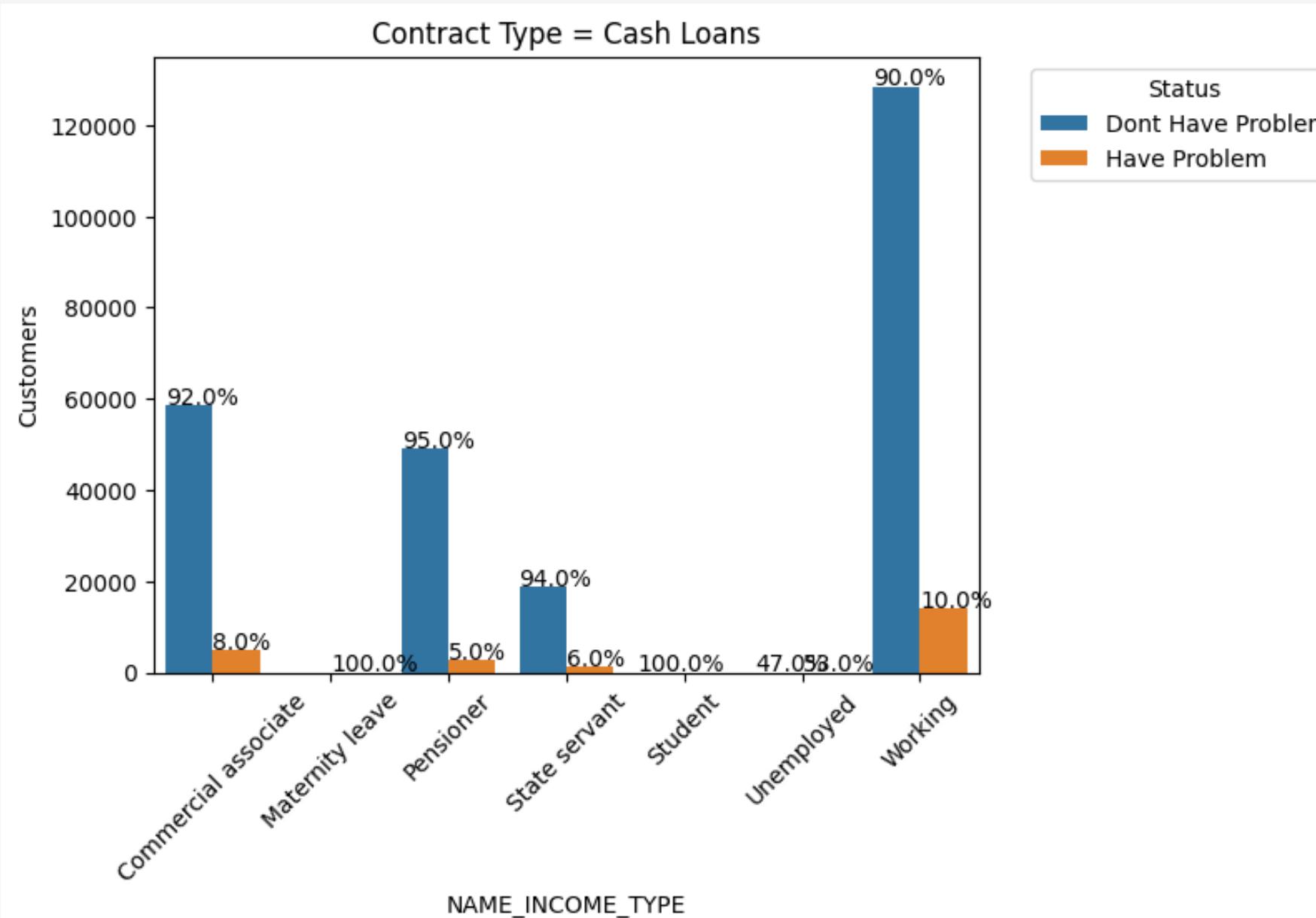
Contract : Revolving Loans



Mengacu pada dataset dan graf dalam kontrak revolving loans, ada empat kategori pendapatan utama, seperti pengusaha, mereka yang sedang cuti hamil, pelajar, dan pengangguran, yang tidak mengalami kesulitan dalam membayar pinjaman mereka. Tetapi, jumlah pelanggan dengan jenis pendapatan ini terbilang sedikit. Oleh karena itu, perlu meningkatkan jumlah pelanggan yang memperoleh pendapatan dari salah satu dari empat jenis tersebut.

Business Insight OVERVIEW

Contract : Cash Loans



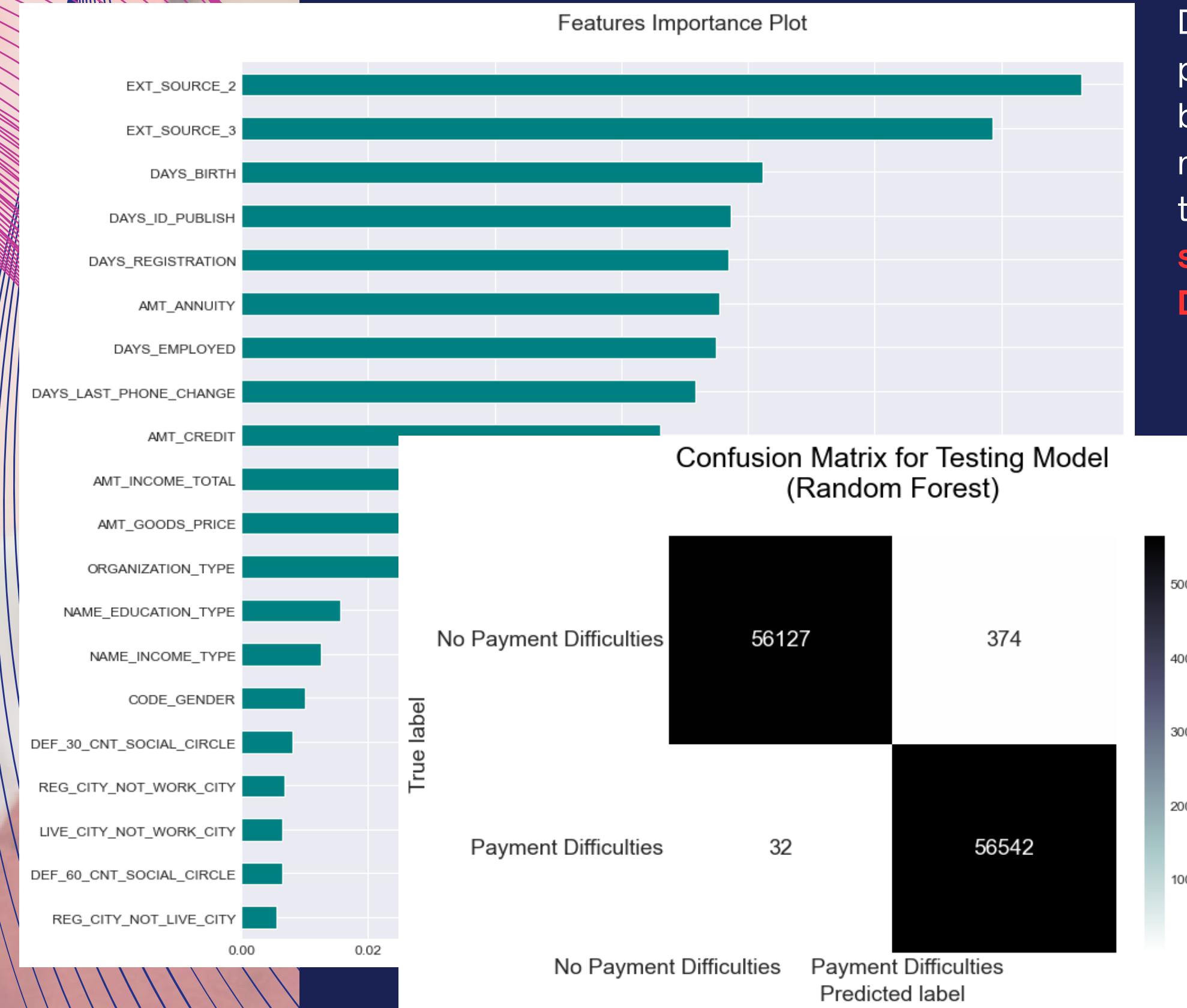
Mengacu pada dataset dan graf dalam kontrak Cash Loans, hanya individu yang sedang cuti hamil yang mengalami kesulitan dalam membayar kembali pinjamannya. Oleh karena itu, dari data tersebut dapat diambil tindakan untuk mempertimbangkan pinjaman dengan kontrak Revolving Loans, karena nasabah yang sedang cuti hamil mengalami keberhasilan yang lebih tinggi dalam pembayaran pinjaman dalam jenis ini.

MACHINE LEARNING MODEL COMPARISON

Models	Training Accuracy Score	Testing Accuracy Score	ROC Score
3 Random Forest	1.000000	0.996500	0.996500
2 Decision Tree	1.000000	0.882600	0.882600
4 K-Nearest Neighbor	0.915600	0.880700	0.880600
5 Neural Network	0.700100	0.694800	0.694800
0 Logistic Regression	0.671600	0.672900	0.672900
1 Gaussian Naive Bayes	0.602400	0.603900	0.604000

Model **Random Forest** menunjukkan akurasi prediksi yang hampir sama untuk kumpulan data latih dan data uji, yang menunjukkan model yang sangat baik tanpa ada tanda-tanda terjadinya **underfitting atau overfitting**. Sehingga berdasar evaluasi model dengan ROC sendiri yang menunjukan akurasi yang baik, maka model Random Forest dipilih sebagai pilihan optimal untuk melakukan prediksi terhadap kemampuan kustomer dalam menyelesaikan pembayaran.

BEST MODEL FEATURE



Dari model yang kita pilih setelah melakukan perbandingan, yaitu Random Forest. Terlihat beberapa **fitur-fitur penting** yang digunakan oleh model untuk melakukan prediksi seperti lima fitur teratas yang **meliputi Score from external data source 2, Score from external data source 3, Days Birth, Days ID publish, Days registration**

Dilihat dengan menggunakan matrix evaluasi, confussion matrix. Data menunjukan bahwa model melakukan kesalahan yang minim dalam mengklasifikasikan kustomer apakah dia bermasalah atau tidak. **Setidaknya hanya terjadi 406 kesalahan dari 113.075 data.**



HOME CREDIT
Kamu Bisa!

CONCLUSION

Pelanggan dengan latar belakang akuntan menunjukkan kemampuan yang tinggi dalam melunasi pinjaman mereka, dengan persentase persetujuan yang mencapai 95%. Namun, jumlah aplikasi pinjaman dari kelompok ini hanya sebesar 3,19%. Pola yang serupa terlihat pada staf dan manajer teknologi berkompetensi tinggi, meskipun mereka memiliki kemampuan untuk melunasi pinjaman, namun permohonan pinjaman dari kelompok ini tergolong sedikit. Di sisi lain, nasabah yang mengajukan cuti melahirkan dalam kontrak pinjaman tunai cenderung memiliki masalah dalam melunasi pinjaman (100% permohonan ditolak), tetapi ketika beralih ke pinjaman bergulir, semua aplikasi mereka disetujui. Pengangguran memiliki tingkat kesulitan yang signifikan (lebih dari 50%) dalam melunasi pinjaman pada pinjaman tunai, sementara semua pengangguran yang menggunakan pinjaman bergulir mampu membayar kembali pinjaman mereka.

RECOMENDATION

Investigasi lebih lanjut diperlukan, mungkin melalui survei, untuk mengidentifikasi potensi masalah yang timbul ketika klien yang sedang cuti melahirkan atau mereka yang menganggur memilih kontrak pinjaman tunai. Data ini akan sangat penting untuk referensi di masa depan, memungkinkan rekomendasi yang disesuaikan untuk klien dengan latar belakang pendapatan yang sama, memastikan persetujuan permohonan mereka dengan menyarankan jenis kontrak yang sesuai.

