

Repo Github :

1. Descriptive Statistics

Df.info()

- Tidak terdapat nilai kosong
- Married/Single diganti Marital_Status
- Income bisa digantikan float

Df.describe()

- Income dan Experience perlu diteliti lebih lanjut untuk potensi outlier karena rentang nilai yang besar.
- CURRENT_HOUSE_YRS memiliki rentang yang sempit, yang mungkin perlu diperiksa lebih lanjut apakah ada bias pada data atau kesalahan pengumpulan data.
- Risk_Flag memiliki distribusi yang condong ke arah nilai 0

2. Univariate Analysis

Skewness KDA

- Distribusi Income hampir seragam, tetapi fluktuasi kecil terlihat. Tidak ada nilai yang terlalu dominan, namun rentang yang lebar mungkin mengindikasikan adanya outlier.
- Distribusi Age cukup merata dari usia 21 hingga 79 tahun. Tidak ada kelompok usia yang sangat dominan, tetapi terdapat sedikit variasi pada frekuensi di beberapa titik usia tertentu.
- Experience menunjukkan distribusi yang cukup merata, tanpa nilai pengalaman kerja yang mendominasi. Ini mengindikasikan variasi pengalaman kerja yang luas di antara individu.
- Distribusi CURRENT_JOB_YRS miring ke kanan, dengan sebagian besar individu memiliki masa kerja yang lebih singkat. Pola ini umum dalam data lama kerja, di mana lebih sedikit orang yang bertahan dalam pekerjaan tertentu selama bertahun-tahun. CURRENT_HOUSE_YRS:
- Distribusi CURRENT_HOUSE_YRS sangat terbatas pada rentang 10 hingga 14 tahun, dengan puncak yang berulang di setiap nilai. Hal ini menunjukkan bahwa hampir semua individu telah tinggal di tempat tinggal mereka selama periode waktu yang hampir sama, yang perlu ditindaklanjuti untuk memastikan relevansi kolom ini.
- Distribusi Risk_Flag menunjukkan ketidakseimbangan yang signifikan, dengan mayoritas nilai 0 (tidak berisiko) dan sebagian kecil nilai 1 (berisiko). Ketidakseimbangan ini perlu ditangani saat pre-processing.

- Outlier Detection pada Income untuk mengidentifikasi dan mengelola nilai ekstrem.
- Class Balancing pada Risk_Flag untuk menangani ketidakseimbangan kelas.
- Validasi Data pada CURRENT_HOUSE_YRS untuk memastikan bahwa nilai-nilai yang mendominasi adalah wajar.
- Normalisasi atau Transformasi pada kolom dengan distribusi yang miring untuk meningkatkan performa analisis atau model.

Categorical

- Sebagian besar data menunjukkan status single, dengan jumlah yang lebih sedikit pada status married.
- Mayoritas data memiliki status rented, dengan sebagian kecil owned dan no_rent_no_own.
- Mayoritas individu tidak memiliki mobil (no), sedangkan sebagian kecil memiliki (yes).
- Kolom Profession memiliki berbagai kategori pekerjaan, namun tidak ada yang benar-benar mendominasi.
Kedua kolom ini memiliki banyak kategori, dengan CITY memiliki lebih banyak variasi dibandingkan STATE. Tindak Lanjut: Jika terlalu banyak kategori, bisa dilakukan pengelompokan atau pengkodean untuk menyederhanakan. Ini akan membantu menghindari masalah sparsity dalam analisis atau pemodelan.

Outlier

- Tidak ada outlier

3. Multivariate Analysis

Korelasi antara Fitur dan Label (Risk_Flag)

- Berdasarkan heatmap korelasi, kita dapat melihat bahwa tidak ada fitur numerik yang memiliki korelasi yang kuat dengan Risk_Flag. Korelasi yang ada berada di kisaran rendah, dengan nilai korelasi terbesar di sekitar -0.035 antara Experience dan Risk_Flag.
- Ini menunjukkan bahwa tidak ada satu fitur numerik yang secara langsung sangat terkait dengan Risk_Flag. Meski begitu, fitur seperti Experience, Age, dan CURRENT_JOB_YRS tetap bisa dipertahankan untuk memperkaya model, karena mereka mungkin memberikan kontribusi minor saat digabungkan dengan variabel lain.

Korelasi Antar-Fitur

- Feature Selection: Pertahankan fitur Experience, Age, dan CURRENT_JOB_YRS, meskipun korelasinya rendah dengan Risk_Flag. Mereka masih mungkin memberikan informasi berguna saat digunakan dalam model machine learning.
- Handling Redundancy: Lakukan uji tambahan (seperti feature importance dalam model) untuk menentukan apakah salah satu dari Experience atau CURRENT_JOB_YRS dapat dihapus untuk mengurangi redundansi tanpa mengorbankan akurasi model.

Korelasi Setiap Fitur dengan Risk_Flag(Categorical)

- Marital Status: Single lebih berisiko (Risk_Flag 1) dibandingkan married. Relevan.
- House Ownership: Rented memiliki risiko lebih tinggi. Sangat relevan.
- Car Ownership: Pemilik mobil cukup seimbang antara risiko tinggi dan rendah. Relevansi moderat.
- Profession: Risiko bervariasi antar-profesi. Sangat relevan.

Korelasi Antar-Fitur (categorical)

Tidak ada korelasi yang berarti

4. Business Insights

- Risk Flag Distribution Across Age groups

Insight

1. Usia 0-25 Tahun: Di kelompok usia ini, jumlah risiko tinggi relatif rendah. Ini mungkin karena mereka belum memiliki banyak komitmen finansial yang besar.
2. Usia 26-35 Tahun: Risiko tinggi mulai meningkat di kelompok usia ini, yang bisa dikaitkan dengan komitmen finansial awal, seperti cicilan kendaraan atau pembelian rumah.
3. Usia 36-50 Tahun: Kelompok usia ini memiliki jumlah pengguna terbanyak dan menunjukkan risiko yang signifikan. Di usia ini, biasanya ada kewajiban finansial besar, seperti cicilan rumah atau biaya pendidikan anak.

4. Usia 51-65 Tahun dan 66+: Di kelompok usia ini, jumlah risiko tinggi stabil. Mereka mungkin sudah memiliki keuangan yang lebih stabil tetapi mulai memikirkan persiapan pensiun atau risiko kesehatan.

Business Recommendation

1. Usia 0-25 Tahun:
 - Produk Tabungan dan Investasi Awal: Kami menyarankan menyediakan produk tabungan jangka panjang atau investasi kecil untuk membantu mereka mulai menabung.
 - Program Edukasi Finansial: Edukasi tentang manajemen keuangan bisa membangun kebiasaan positif di awal perjalanan finansial mereka.
2. Usia 26-35 Tahun:
 - Pinjaman Konsumtif: Sediakan produk pinjaman seperti untuk kendaraan atau renovasi rumah dengan persyaratan fleksibel, yang relevan dengan gaya hidup mereka.
 - Asuransi Dasar: Mulailah menawarkan asuransi kesehatan atau kecelakaan sederhana yang bisa memberikan perlindungan dasar.
3. Usia 36-50 Tahun:
 - Konsultasi Keuangan: Di usia ini, konsultasi keuangan bisa membantu mereka mengelola aset dan persiapan keuangan masa depan.
 - Asuransi Komprehensif: Produk asuransi kesehatan, jiwa, dan properti komprehensif bisa sesuai dengan kebutuhan mereka yang sudah kompleks.
4. Usia 51-65 Tahun dan 66+:
 - Program Persiapan Pensiun: Produk tabungan pensiun atau investasi stabil yang berisiko rendah bisa menarik bagi mereka.
 - Asuransi Kesehatan Jangka Panjang: Mengingat usia mereka, asuransi kesehatan dan perawatan jangka panjang akan menjadi produk yang sangat relevan.

- Top 5 Lowest and Highest Risk Probability

Insight

- Profesi Risiko Tinggi (Police Officer, Chartered Accountant, Army Officer, Scientist, Civil Engineer):

Tingkat risiko lebih tinggi karena tanggung jawab besar atau lingkungan kerja yang berisiko.

- Profesi Risiko Rendah (contoh: Librarian, Politician, Economist Petroleum Engineer, Technology Specialist):
Posisi stabil dengan penghasilan baik dan tanggung jawab yang lebih terfokus.

Business Recommendation

- Profesi Risiko Tinggi
 - Asuransi Komprehensif: Tawarkan asuransi jiwa dan kecelakaan.
 - Persyaratan Kredit Ketat: Terapkan syarat kredit ketat atau pinjaman dengan jaminan.
- Profesi Risiko Rendah
 - Investasi Premium: Berikan produk investasi dengan imbal hasil tinggi.
 - Kredit Bunga Rendah: Tawarkan pinjaman dengan bunga lebih rendah.

- Total Counts of Risk Categories by Income Level

Insight

- Pendapatan Menengah: Memiliki jumlah individu berisiko tinggi terbanyak, menunjukkan adanya kerentanan finansial di kelompok ini.
- Pendapatan Tinggi: Memiliki proporsi risiko tinggi yang lebih rendah, menunjukkan stabilitas finansial lebih baik.
- Pendapatan Rendah: Meskipun jumlahnya lebih kecil, terdapat proporsi individu berisiko tinggi yang signifikan.

Business Recommendation

1. Produk Finansial Terarah:
 - Pendapatan Menengah: Tawarkan produk untuk mengelola risiko, seperti asuransi atau investasi moderat.
 - Pendapatan Tinggi: Fokus pada layanan premium seperti manajemen kekayaan.
 - Pendapatan Rendah: Sediakan produk keuangan dasar dan edukasi literasi finansial.
2. Program Mitigasi Risiko:
 - Adakan workshop literasi finansial bagi kelompok menengah dan rendah untuk membantu mereka mengelola risiko finansial.
3. Kampanye Pemasaran Tertarget:

- Sesuaikan kampanye pemasaran dengan kebutuhan spesifik tiap kelompok, khususnya fokus pada mitigasi risiko untuk pendapatan menengah.
- 4. Peluang Cross-Selling:
 - Tawarkan produk investasi untuk individu pendapatan menengah berisiko rendah dan layanan eksklusif untuk pendapatan tinggi.

- Count by Income Quantile and Marital Status

Insight

1. Status Pernikahan dan Kuantile Pendapatan:
 - Mayoritas individu lajang berada di kuantile pendapatan menengah (25-75%), sementara proporsi individu lajang di kuantile atas (>75%) dan bawah (<25%) cukup seimbang.
 - Untuk individu menikah, jumlahnya jauh lebih sedikit di semua kuantile, dengan distribusi yang lebih merata di ketiga kuantile pendapatan.
2. Distribusi yang Dominan pada Individu Lajang:
 - Lajang dengan pendapatan menengah mendominasi secara signifikan, menunjukkan bahwa segmen ini lebih besar dibandingkan dengan yang menikah.

Business Recommendation

1. Strategi Produk untuk Lajang Berpendapatan Menengah:
 - Tawarkan produk investasi atau tabungan jangka panjang yang fleksibel, karena segmen ini kemungkinan besar ingin menabung atau berinvestasi untuk masa depan.
 - Manfaatkan pemasaran digital dengan kampanye yang berfokus pada stabilitas keuangan untuk masa depan, dengan pesan yang relevan bagi individu lajang.
2. Penawaran untuk Pasangan Menikah:
 - Meskipun proporsinya lebih kecil, pasangan menikah mungkin memerlukan produk keuangan bersama, seperti asuransi keluarga atau tabungan pendidikan anak.

- Pertimbangkan kampanye personalisasi yang menyoroti manfaat produk keuangan yang dirancang untuk keluarga kecil.

3. Pendekatan Pemasaran yang Tertarget:

- Untuk individu lajang di kuantile 25-75%, gunakan strategi pemasaran dengan pesan terkait peningkatan dan stabilitas finansial.
- Untuk individu dengan pendapatan >75%, baik lajang maupun menikah, fokus pada layanan premium seperti manajemen kekayaan.

- **Count of Risk Flags by Marital Status and House Ownership**

Insight:

- Mayoritas Lajang Menyewa Rumah: Sebagian besar individu lajang berada dalam kategori "rented."
- Kepemilikan Rumah Rendah: Baik lajang maupun menikah menunjukkan tingkat kepemilikan rumah yang rendah dibandingkan opsi lain.

Business Recommendation:

- Tawarkan Program Kepemilikan Rumah: Sasaran utama adalah lajang yang ingin transisi dari menyewa ke memiliki rumah.
- Produk Asuransi untuk Penyewa: Berikan perlindungan asuransi bagi penyewa, mengingat tingginya jumlah individu yang menyewa.

- **Total Counts of Risk Categories by House and Car Ownership**

Insight:

- Penyewa Tanpa Mobil: Kelompok terbesar adalah individu yang menyewa rumah tanpa memiliki mobil, dengan proporsi risiko rendah dan tinggi yang signifikan.
- Kepemilikan Mobil pada Penyewa: Penyewa yang memiliki mobil menunjukkan jumlah individu berisiko lebih rendah dibandingkan penyewa tanpa mobil.

Business Recommendation :

- Penawaran Asuransi Mobil untuk Penyewa: Berikan insentif asuransi mobil bagi penyewa untuk mengurangi risiko finansial mereka.

- Program Kepemilikan Rumah: Tawarkan program transisi dari menyewa ke memiliki rumah, terutama bagi penyewa tanpa mobil untuk membantu meningkatkan stabilitas finansial mereka.