Raport walidacyjny

Etap EDA, raport numer 1

Korekty wymagają następujące elementy:

1. Brak wskazanych kolumn o znaczącej korelacji.
2. W przypadku analizy cech pod kątem korelacji zabrakło wniosków, jak silnie jedna zmienna zależy od drugiej. (np. “*Age Distribution by Marital Status”, “Income Distribution by Education Level”).*
3. W podrozdziale ‘Learning data correlation and column engineering’ brak wniosków co do zależności kolumn w macierzy korelacji (np. pomiędzy zmiennymi finalized\_loan i lenght\_relationship\_with\_client czy debit\_card i salary\_account).
4. Ramka danych “x\_train.csv” ma nazwy kolumn zawierające spacje.
5. Brak informacji o niezrównoważeniu zmiennej do przewidywania i brak informacji o korekcie tego w ramce danych do trenowania.
6. (Detale) *„Count of Categorical Variables”*: podwykresy są zbyt małe, przez co są nieczytelne. *„Count of Categorical Variables”*: “Most have finishes a university.” – błąd koniugacji. *“Income Distribution by Education Level”:* “We may be able to see that Uniwersity graduates[…]”, “# and map it as half of possible numbers or 2000 ti indicate highest one” literówki.