Завдання до файлу loan\_applic.csv:

1. В якості навчальної вибірки обрати дані до 1 вересня 2018 року, тестова вибірка починається з 1 вересня 2018 року
2. усе подальше виконується на навчальній вибірці, відповідні перетворення даних просто переносяться на навчальну вибірку
3. провести розвідувальний аналіз даних:

- визначити типи даних,

- кількість пропусків,

- визначення числових та категорійних вхідних характеристик,

- подивитись розподіли числових характеристик, подумати які з них варто нормувати

- визначити кореляції між числовими характеристиками, провести дослідження на наявність мультикореляції в даних, прийняти міри у випадку її присутності

- визначити можливий вплив категорійних характеристик на таргетну змінну і обміркувати його можливе використання для формування rule-based approach навіть до побудови моделі.

- дослідити кожну числову характеристику на предмет наявності викидів у даних – обдумати способи боротьби з ними

1. Провести Feature Engineering – можливо, в тому числі, застосувавши кластеризацію до певних груп характеристик - спробувати розбити елементи навчальної вхідної інформації на групи за схожістю.
2. Провести аналіз інформативності та важливості фіч, підготувати їх на вхід до моделі
3. Обґрунтувати вид обраної моделі, провести порівняльний аналіз технік, що залучаються при роботі з незбалансованими даними – порівняти результати їх застосування на основі оцінок роботи обраної моделі машинного навчання