

## Задача 1: клиенты авиакомпании

Ссылка на данные:

[https://raw.githubusercontent.com/Murcha1990/from\\_idea\\_to\\_mvp/main/projects/clients.csv](https://raw.githubusercontent.com/Murcha1990/from_idea_to_mvp/main/projects/clients.csv)

Датасет содержит информацию о клиентах некоторой авиакомпании.

*Целевая переменная* (таргет) – **satisfaction** (удовлетворенность клиента полетом), бинарная ('satisfied' или 'neutral or dissatisfied')

*Признаки:*

- **Gender** (categorical: 'Male' или 'Female'): пол клиента
- **Age** (numeric, int): количество полных лет
- **Customer Type** (categorical: 'Loyal Customer' или 'disloyal Customer'): лоялен ли клиент авиакомпании?
- **Type of Travel** (categorical: 'Business travel' или 'Personal Travel'): тип поездки
- **Class** (categorical: 'Business' или 'Eco', или 'Eco Plus'): класс обслуживания в самолете
- **Flight Distance** (numeric, int): дальность перелета (в милях)
- **Departure Delay in Minutes** (numeric, int): задержка отправления (неотрицательная)
- **Arrival Delay in Minutes** (numeric, int): задержка прибытия (неотрицательная)

Признаки, перечисленные ниже, являются числовыми. По смыслу они категориальные: клиент ставил оценку от 1-го до 5-ти включительно. Есть выбросы

- **Inflight wifi service** (categorical, int): оценка клиентом интернета на борту
- **Departure/Arrival time convenient** (categorical, int): оценка клиентом удобство времени прилета и вылета
- **Ease of Online booking** (categorical, int): оценка клиентом удобства онлайн-бронирования
- **Gate location** (categorical, int): оценка клиентом расположения выхода на посадку в аэропорту
- **Food and drink** (categorical, int): оценка клиентом еды и напитков на борту
- **Online boarding** (categorical, int): оценка клиентом выбора места в самолете
- **Seat comfort** (categorical, int): оценка клиентом удобства сиденья
- **Inflight entertainment** (categorical, int): оценка клиентом развлечений на борту
- **On-board service** (categorical, int): оценка клиентом обслуживания на борту
- **Leg room service** (categorical, int): оценка клиентом места в ногах на борту
- **Baggage handling** (categorical, int): оценка клиентом обращения с багажом
- **Checkin service** (categorical, int): оценка клиентом регистрации на рейс
- **Inflight service** (categorical, int): оценка клиентом обслуживания на борту
- **Cleanliness** (categorical, int): оценка клиентом чистоты на борту

## Задача 2: подержанные автомобили

Ссылка на данные:

[https://raw.githubusercontent.com/Murcha1990/from\\_idea\\_to\\_mvp/main/projects/cars\\_train.csv](https://raw.githubusercontent.com/Murcha1990/from_idea_to_mvp/main/projects/cars_train.csv)

Датасет содержит информацию о характеристиках и ценах подержанных автомобилей в некоторой стране

*Целевая переменная* (таргет) – **selling\_price** (цена продажи), числовая

*Признаки*:

- **name** (string): модель автомобиля
- **year** (numeric, int): год выпуска с завода-изготовителя
- **km\_driven** (numeric, int): пробег на дату продажи
- **fuel** (categorical: 'Diesel' или 'Petrol', или 'CNG', или 'LPG', или 'electric'): тип топлива
- **seller\_type** (categorical: 'Individual' или 'Dealer', или 'Trustmark Dealer'): продавец
- **transmission** (categorical: 'Manual' или 'Automatic'): тип трансмиссии
- **owner** (categorical: 'First Owner' или 'Second Owner', или 'Third Owner', или 'Fourth & Above Owner'): какой по счёту хозяин?
- **mileage** (string, по смыслу числовой): пробег, требуется предобработка
- **engine** (string, по смыслу числовой): рабочий объем двигателя, требуется предобработка
- **max\_power** (string, по смыслу числовой): пиковая мощность двигателя, требуется предобработка
- **torque** (string, по смыслу числовой, а то и 2): крутящий момент, требуется предобработка
- **seats** (numeric, float; по смыслу categorical, int)

## Задача 3: кредитный скоринг

Ссылка на данные:

[https://raw.githubusercontent.com/Murcha1990/from\\_idea\\_to\\_mvp/main/projects/credit\\_scoring.csv](https://raw.githubusercontent.com/Murcha1990/from_idea_to_mvp/main/projects/credit_scoring.csv)

Ваша задача состоит в том, чтобы по различным характеристикам клиентов спрогнозировать целевую переменную - имел клиент просрочку 90 и более дней или нет (и если имел, то банк не будет выдавать кредит этому клиенту, а иначе будет).

Ниже находится описание признаков клиентов.

*Целевая переменная (target)*

**SeriousDlqin2yrs:** Клиент имел просрочку 90 и более дней

- **RevolvingUtilizationOfUnsecuredLines:** Общий баланс средств (total balance on credit cards and personal lines of credit except real estate and no installment debt like car loans divided by the sum of credit limits).
- **age:** Возраст заемщика
- **NumberOfTime30-59DaysPastDueNotWorse:** Сколько раз за последние 2 года наблюдалась просрочка 30-59 дней.
- **DebtRatio:** Ежемесячные расходы (платеж по долгам, алиментам, расходы на проживание) деленные на месячный доход.
- **MonthlyIncome:** Ежемесячный доход.
- **NumberOfOpenCreditLinesAndLoans:** Количество открытых кредитов (например, автокредит или ипотека) и кредитных карт.
- **NumberOfTimes90DaysLate:** Сколько раз наблюдалась просрочка (90 и более дней).
- **NumberRealEstateLoansOrLines:** Количество кредитов (в том числе под залог жилья)
- **RealEstateLoansOrLines:** Закодированное количество кредитов (в том числе под залог жилья) - чем больше код буквы, тем больше кредитов
- **NumberOfTime60-89DaysPastDueNotWorse:** Сколько раз за последние 2 года заемщик задержал платеж на 60-89 дней.
- **NumberOfDependents:** Количество иждивенцев на попечении (супруги, дети и др).
- **GroupAge:** закодированная возрастная группа - чем больше код, тем больше возраст.