

Churn customer prediction

Марат Зиятдинов
Джоксан Асокар

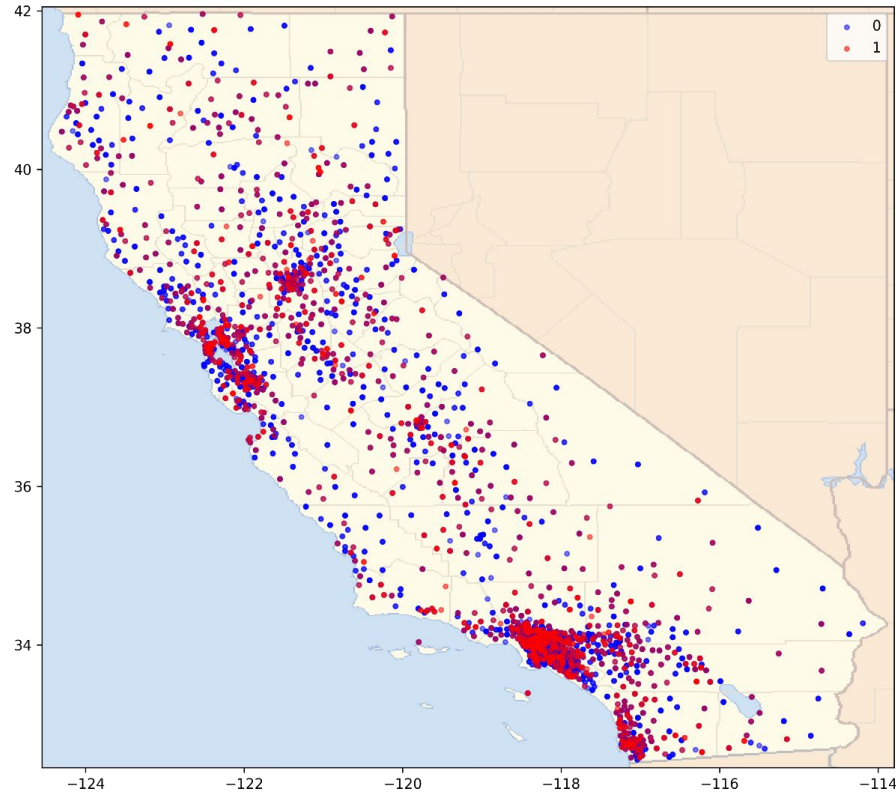
EDA — Preliminary Data Information

- Train: 5,282 customers
 - Label 0: 3,912 customers (74.06%)
 - Label 1: 1,370 customers (25.93%).
 - 31 columns

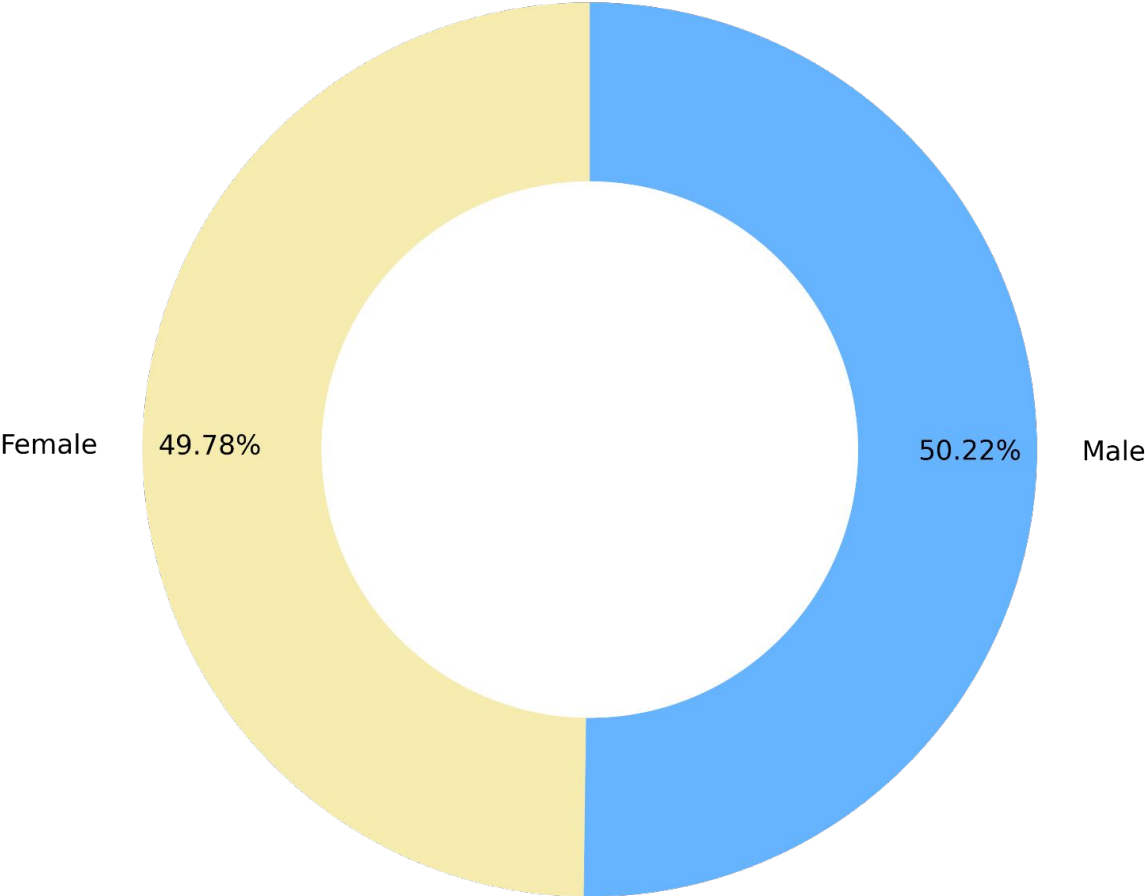
EDA — Preliminary Data Information

- Train: 5,282 customers
 - Label 0: 3,912 customers (74.06%)
 - Label 1: 1,370 customers (25.93%).
 - 31 columns
 - Important features:
 - Churn value (target)
 - Churn score
 - CLTV
 - Churn reason
- Test: 1,761 customers.

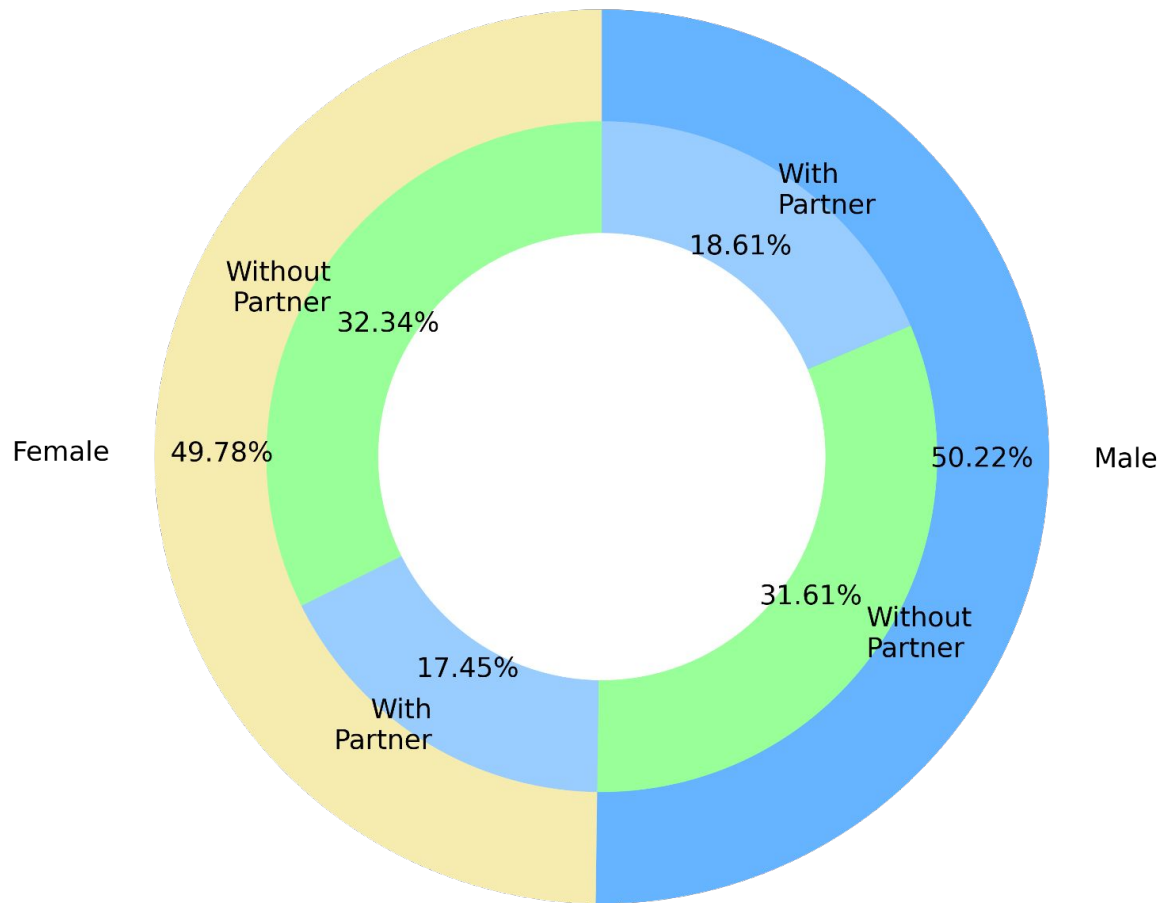
EDA — Longitude and Latitude



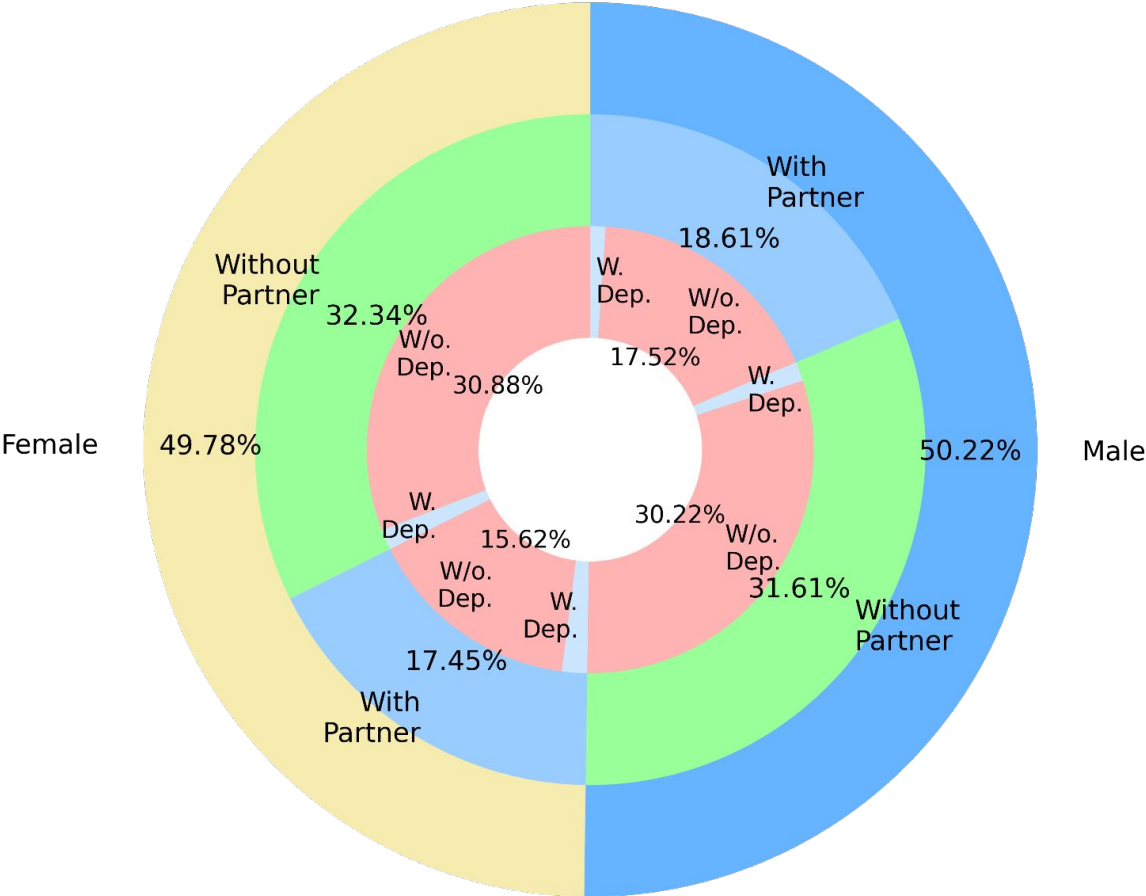
EDA — Gender



EDA — Gender + Partner

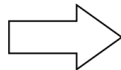


EDA — Gender + Partner + Dependents



EDA - New Features 1.0

internet_service	
10	Fiber optic
11	DSL
12	No
13	Fiber optic
14	DSL

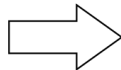


internet_service		internet_type
10	Yes	Fiber optic
11	Yes	DSL
12	No	No
13	Yes	Fiber optic
14	Yes	DSL

Binarize internet_service

EDA - New Features 1.0

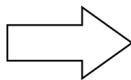
internet_service	
10	Fiber optic
11	DSL
12	No
13	Fiber optic
14	DSL



internet_service		internet_type
10	Yes	Fiber optic
11	Yes	DSL
12	No	No
13	Yes	Fiber optic
14	Yes	DSL

Binarize internet_service

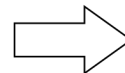
tenure_months	
0	19
1	20
2	72
3	1
4	4



bin_tm	
0	(12, 24]
1	(12, 24]
2	(36, 100]
3	(0, 12]
4	(0, 12]

Bin tenure_months

monthly_charges	
0	24.90
1	19.50
2	111.65
3	25.00
4	89.60



bin_mc	
0	(20.0, 40.0]
1	(0.0, 20.0]
2	(100.0, 120.0]
3	(20.0, 40.0]
4	(80.0, 100.0]

Bin monthly_charges

EDA - New Features 1.0.

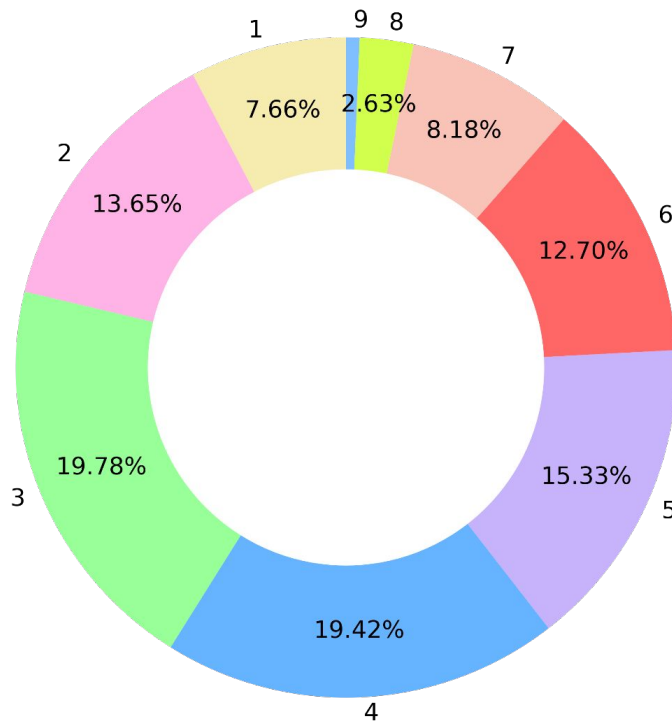
$$\text{service_count} = \sum_{s \in C} [s = 1].$$

$C = \{\text{internet_service}, \text{phone_service}, \text{multiple_lines}, \dots\}$

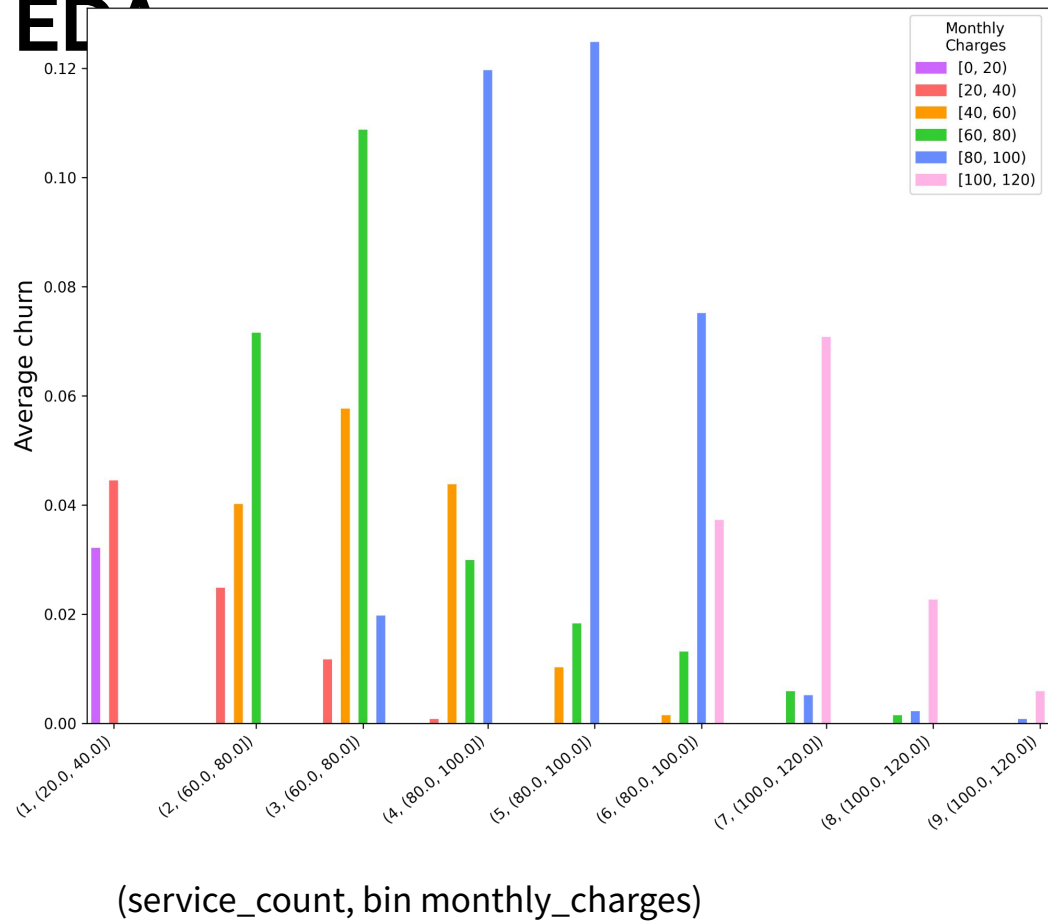
EDA - New Features 1.0.

$$\text{service_count} = \sum_{s \in C} [s = 1].$$

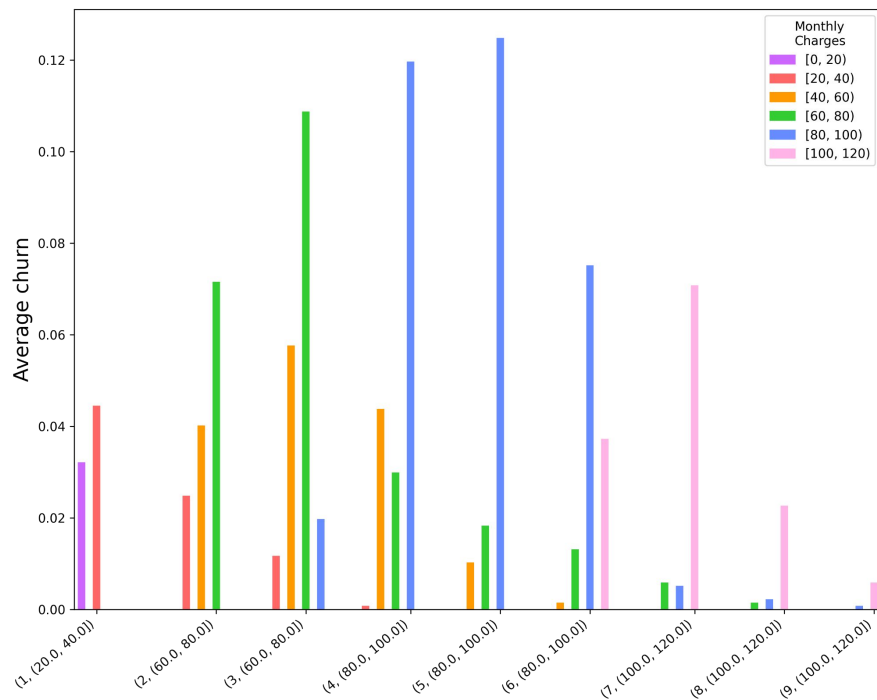
$C = \{\text{internet_service}, \text{phone_service}, \text{multiple_lines}, \dots\}$



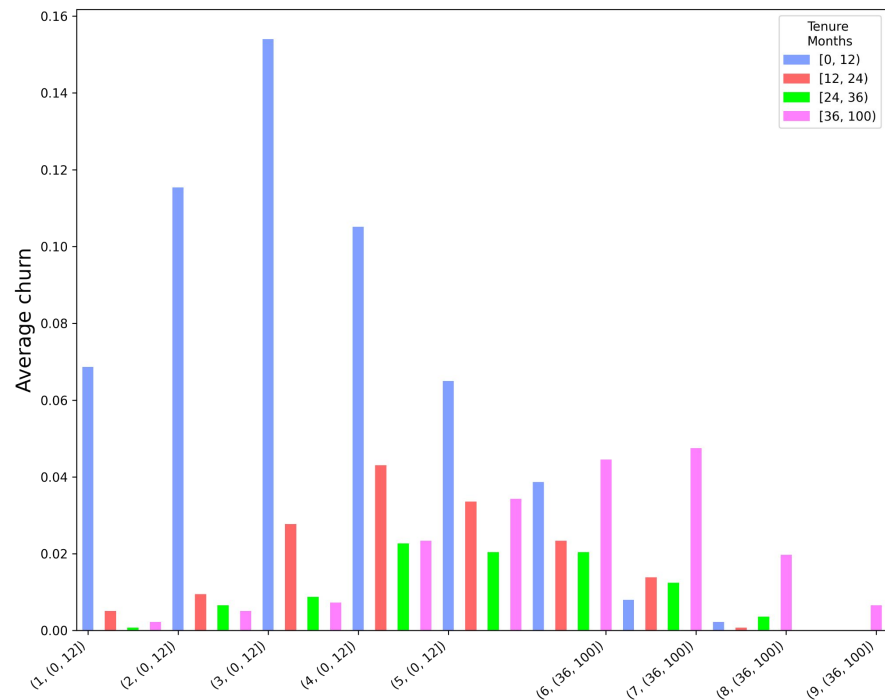
EDA



EDA



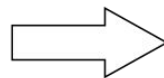
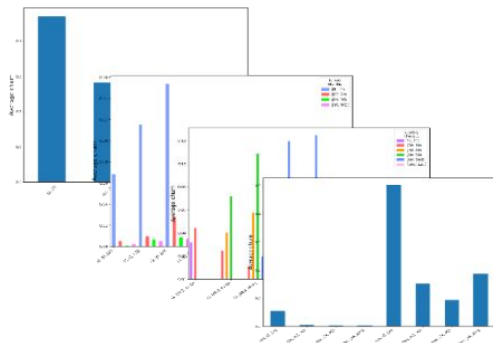
(service_count, bin monthly_charges)



(service_count, bin tenure_months)

Решение

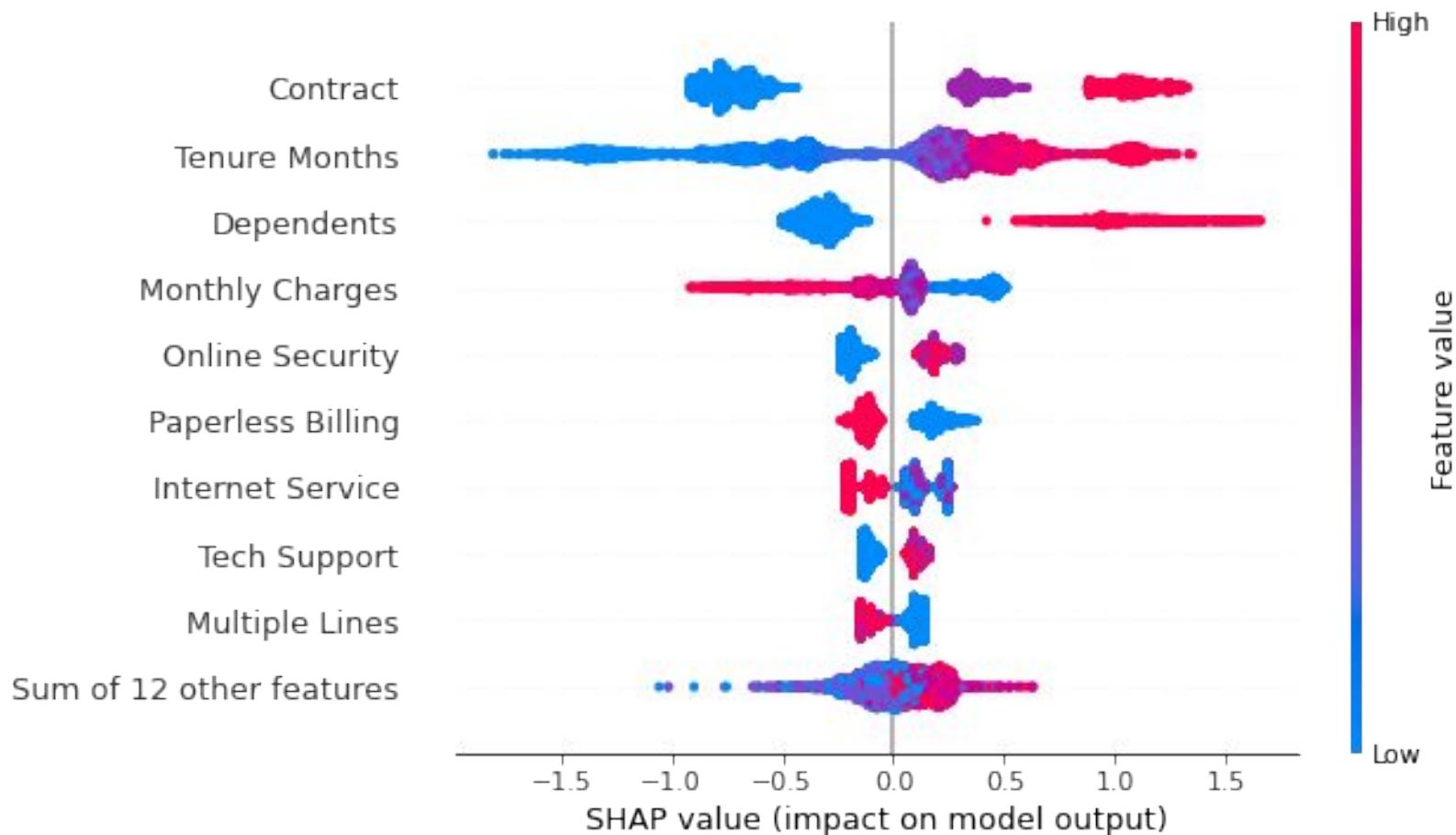
- Удалили колонку `customer_id`
- Не включили в модель колонки: Churn Score, CLTV, Churn Reason и другие, не несущие важную информацию.
- Закодировали категориальные фичи
- Добавили созданные нами фичи (label encoding, target encoding).

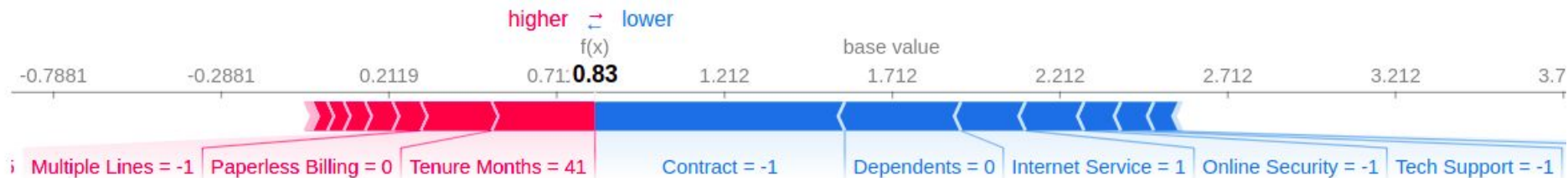
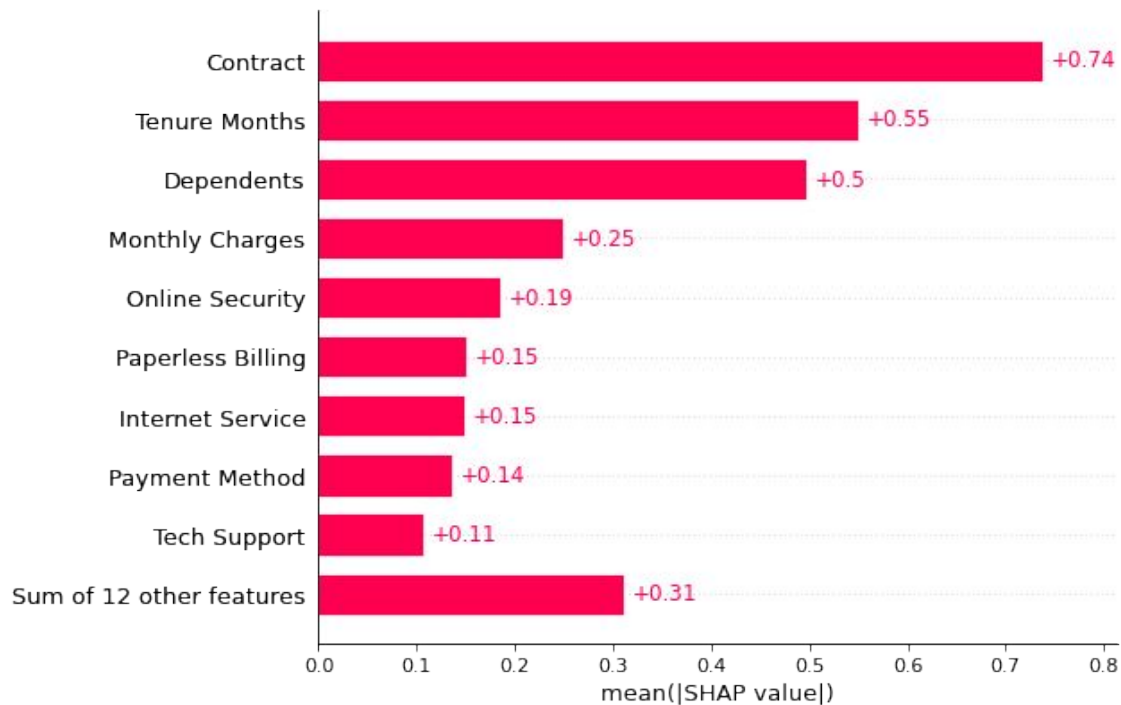


Target Encoding

- Основная модель: градиентный бустинг
- Кросс-валидацию (cv=5) Train, validation: 80% и 20%
- Shap values.

Shap values





Проблемы при внедрении

1. Не все новые признаки полезны.
2. Не анализируются выбросы.
3. Изменение распределении признаков наших клиентов.

Финансовая модель

Траты:

1. закупка оборудования
2. аренда офиса
3. стоимость услуг
4. зарплаты сотрудников
5. возможные скидки

Прибыль:

1. прибыль от стандартного клиента
2. прибыль от клиента со скидками
3. можно попробовать динамически повышать стоимость наших услуг для лояльных клиентов, например с помощью АБ тестов

Стратегия удержания клиента

1. Сделать разные скидки, бонус.
2. Попыаться поменять признаки, которые на `shap values` дают основной негативный вклад, и добавить те, которые вносят основной позитивный вклад.
3. Сделать прозрачную стратегию бонусов с долгосрочной перспективой.