

Исследование клиентов кредитных карт & Credit card customers research

**Люляев Иван
студент МГТУ им. Н.Э. Баумана**

**Почему я
выбрал
эту тему ?**

**Почему я
выбрал
эту тему ?**

Актуальность.

Актуальность

Актуальность

**Рынок банковских карт
стремительно растет**

Актуальность

**Рынок банковских карт
стремительно растет**



**Детская карта
от Сбербанка**



СБЕРБАНК



Актуальность

Рынок банковских карт
стремительно растет



слава богу не
кредитная

Детская карта
от Сбербанка



СБЕРБАНК



Актуальность

Рынок банковских карт
стремительно растет

100 000

новых кредитный карт









**Почему я
выбрал
эту тему ?**

**Почему я
выбрал
эту тему ?**

Хорошие данные. Наконец-то

Хорошие данные. Наконец-то

Trending YouTube Video Statistics

<https://www.kaggle.com/datasnaek/youtube-new>

inspiration:

- Analysing what factors affect how popular a YouTube video will be.

Аналитика клиентов e-commerce

Datasets:

1. <https://www.kaggle.com/prachi13/customer-analytics>
2. <https://www.kaggle.com/aungpyaeap/supermarket-sales> - продажи гипермаркета (1000 записей)

Анализ гипермаркетов

Datasets:

1. <https://www.kaggle.com/itssuru/super-store>

Бизнесов задачи:

1. Найти слабые места бизнеса, над чем нужно поработать чтобы сделать больше прибыли
2. Найти факторы, которые влияют на прибыльность гипермаркетов

Taxi

Chicago Taxi Trips

<https://www.kaggle.com/chicago/chicago-taxi-trips-bq>

Uber Pickups in New York City (*)

<https://www.kaggle.com/fivethirtyeight/uber-pickups-in-new-york-city?select=uber-raw-data-apr14.csv>

Статья на основе (*)

<https://fivethirtyeight.com/features/public-transit-should-be-ubers-new-best-friend/>

Хорошие данные. Наконец-то

Исследование ресторанного бизнеса во время пандемии

Datasets:

1. Restaurant Business Rankings 2020

Top 250, Top 100 Independents and Future 50

<https://www.kaggle.com/michau96/restaurant-business-rankings-2020>

Там 3 таблицы

Исследование с хорошими примерами визуализации данных этого [dataseta](https://www.kaggle.com/akashram/so-what-happened-to-the-restaurants-last-year):
<https://www.kaggle.com/akashram/so-what-happened-to-the-restaurants-last-year>

Исследовательские вопросы

- Каков был 2020 год для ресторанного бизнеса? Рассказать историю в данных

Бизнесовые вопросы

- Что делали те рестораны (что их объединяло), которые преуспели во время пандемии?
- Что делали те рестораны которые разорились/не преуспели во время пандемии
- Что в ресторанном бизнесе будет популярно в будущем?

Задачи по идеям:

- ☐ Проверить позволяют ли имеющиеся данные ответить на эти вопросы
- Из данных по которым можно определить успешность:

- Количество продаж
- Динамика продаж от года к году
- Количество точек
- Динамика количества точек от года к году

Из данных, характеризующих рестораны только:

- Какая кухня представлена
- Тип ресторана (quick service, Fast Casual, Casual dining, Family Dining etc)

Анализ гипермаркетов

Datasets:

1. <https://www.kaggle.com/itssuru/super-store>

Бизнесовые задачи:

- Найти слабые места бизнеса, над чем нужно поработать чтобы сделать больше прибыли
- Найти факторы, которые влияют на прибыльность гипермаркетов

Рынок недвижимости в Мельбурне

<https://www.kaggle.com/ajaypalsinghlo/world-happiness-report-2021> (отсюда ответить на вопрос зачем мне покупать квартиру в Австралии?)

- Зачем мне покупать квартиру в Австралии?
- Потому что она занимает 2 место по уровню счастья
- Одно из лидирующих мест (показать места Австралии в разрезе разных параметров (из которых составляется рейтинг) и сказать почему это оптимальные условия и комбинация факторов)

Taxi Trips

www.kaggle.com/ct

Skips in New York

www.kaggle.com/fjiv

на основе (*)

ivethirtyeight.com/

Twitter Datasets:

1. Tweets at

<https://www.kaggle.com/omermettin/to-2020?select=CompanyValues.csv>

2. Values of

<https://www.kaggle.com/omermettin/to-2020?select=CompanyValues.csv>

Исследовательские вопросы:

- Исследовать взаимосвязь между

нии на рынок
ритель гипотезы
у на акции

Аналитика

Datasets:

- <https://www.kaggle.com/prachi13/customer-analytics>
- <https://www.kaggle.com/aungpyaeap/supermarket-sales> - продажи гипермаркета (1000 записей)

Бизнесовые задачи:

сете предст
эщью самоп
ре и использовать **тональность постов** в анализе.

<https://www.kaggle.com/anthonypino/melbourne-housing-market> (тут с адресом апартаментов)

<https://www.kaggle.com/dansbecker/melbourne-housing-snapshot>

- Где стоит купить квартиру в Австралии?
- Какую квартиру лучше всего покупать в Австралии?
- Предложить конкретные варианты

Анализ какие видео становятся популярными на Ютубе

Datasets:

Trending YouTube Video Statistics

<https://www.kaggle.com/datasnaek/youtube-new>

inspiration:

- Analysing what factors affect how popular a YouTube video will be.

ПОГОВОРИМ

О

ДАННЫХ

Коротко, но понятно.

Коротко о данных

Демография

Лояльность

Коротко о данных

Демография

Пол

Семейное
положение

Образование

Возраст

Дети

Доход

Лояльность

Коротко о данных

Демография

- Пол
- Семейное положение
- Образование
- Возраст
- Дети
- Доход

Лояльность

- Тип карты
- Время сотрудничества
- Утилизация карты
- Кредитный лимит
- Траты по карте

Коротко о данных

Attrition_Flag



Ушедшие



Активные

Key value

Attrition_Flag активный пользователь или ушел

Demographics

Customer_Age

Gender

Dependent_count

количество людей, которые человек обеспечивает

Education_Level

Marital_Status

Income_Category

Loyalty

Card_Category

Months_on_book

сколько человек уже является клиентом банка

Months_Inactive_12_mon

Total_Relationship_Count

сколько продуктов приобрел человек с момента становления клиентом

Avg_Utilization_Ratio

утилизация кредитного лимита клиентом

Credit_Limit

кредитный лимит по кредитной карте

revolving_balance

Не погашенная часть долга по кредитке, перенесенная на следующий месяц

значение усреднено за последние 12 месяцев

High revolving balances may indicate that a borrower is relying too much on credit →

Avg_Open_To_Buy

Средства, доступные к займу в этом месяце

$\text{credit_limit} - \text{revolving_balance}$

значение усреднено за последние 12 месяцев

Total_Trans_Amt

Сумма трат по кредитной карте за последние 12 месяцев

Total_Amt_Chng_Q4_Q1

Отношение трат по кредитной карте в конце года к тратам в начале года

Q4 over Q1

Говорит о динамике трат клиента от начала года к концу

Задачи моего исследования

2 штуки

Задачи моего исследования

1 **Понять, из-за чего
уходят клиенты банка**

2 **Сформировать
гипотезы по
удержанию клиентов**

**Перейдем
к
исследованию**

План

1 Траты по кредитке

2 непогашенная сумма долга

**3 Как связаны возраст и
LTV клиента?**

Самодельный график

2 группы - 2 гистограммы

Самодельный график

2 группы - 2 гистограммы



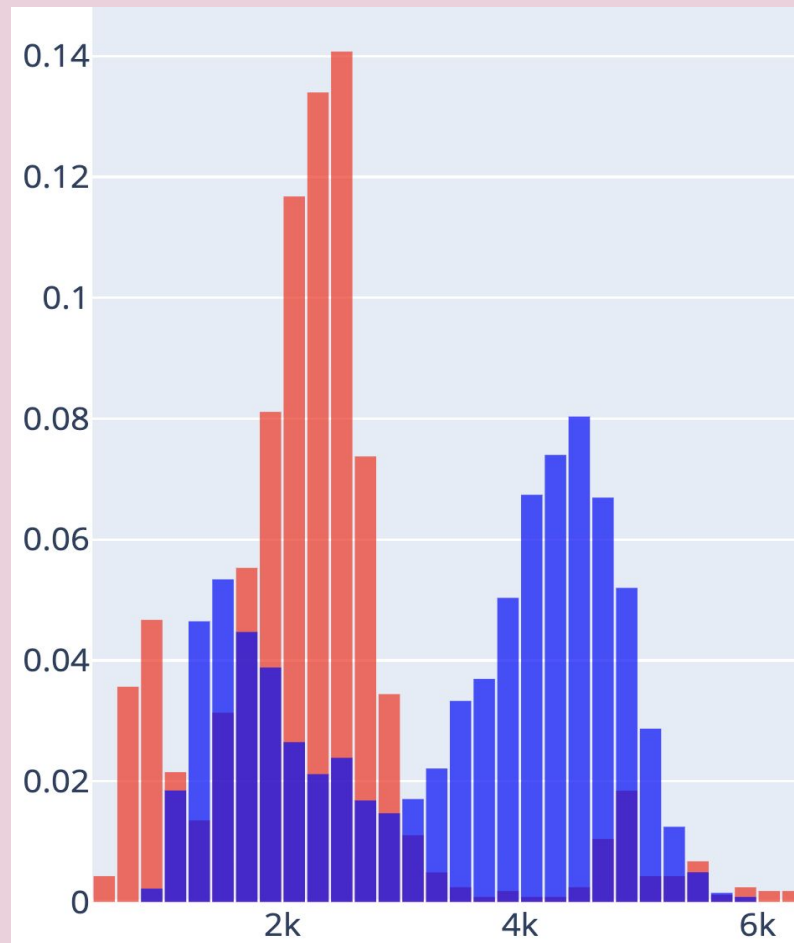
Ушедшие



Активные

Траты по кредитке

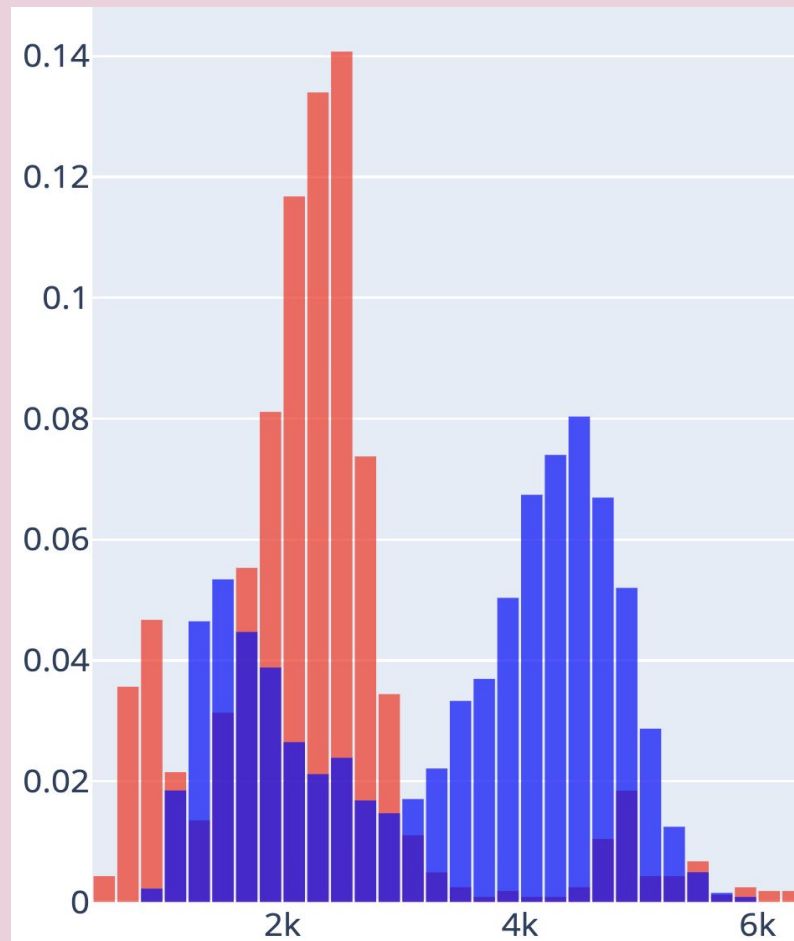
Траты по кредитке



Распределение трат по кредитке

Траты по кредитке

зона риска: 1500-3000\$

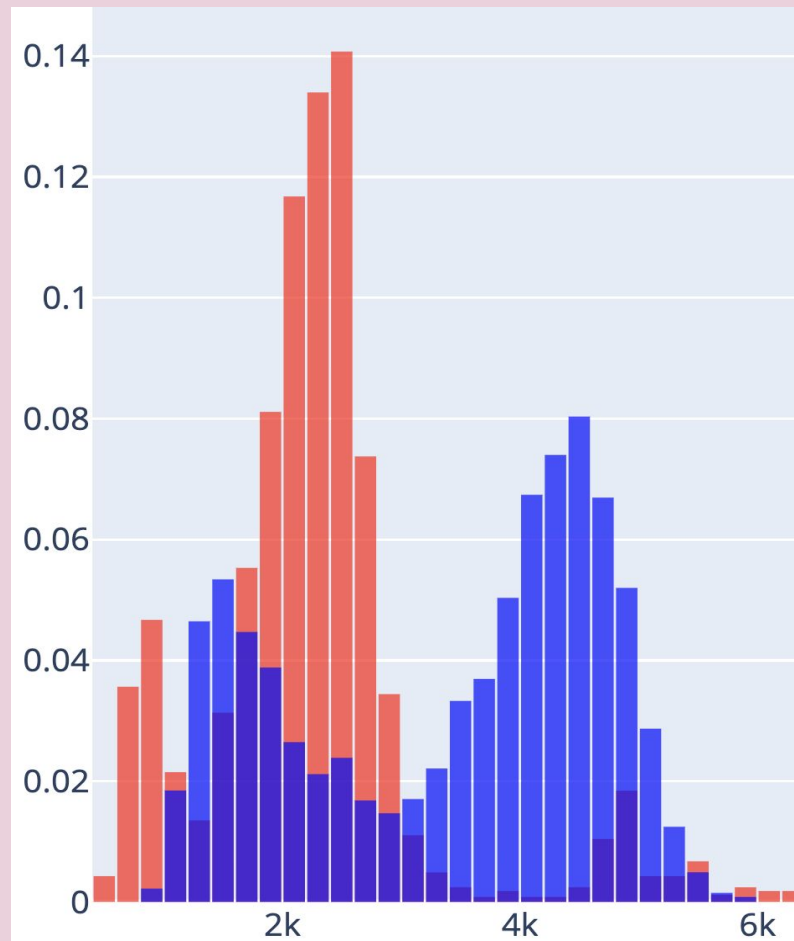


Распределение трат по кредитке

Траты по кредитке

зона риска: 1500-3000\$

**Отдельно работать с этим
сегментом клиентов**



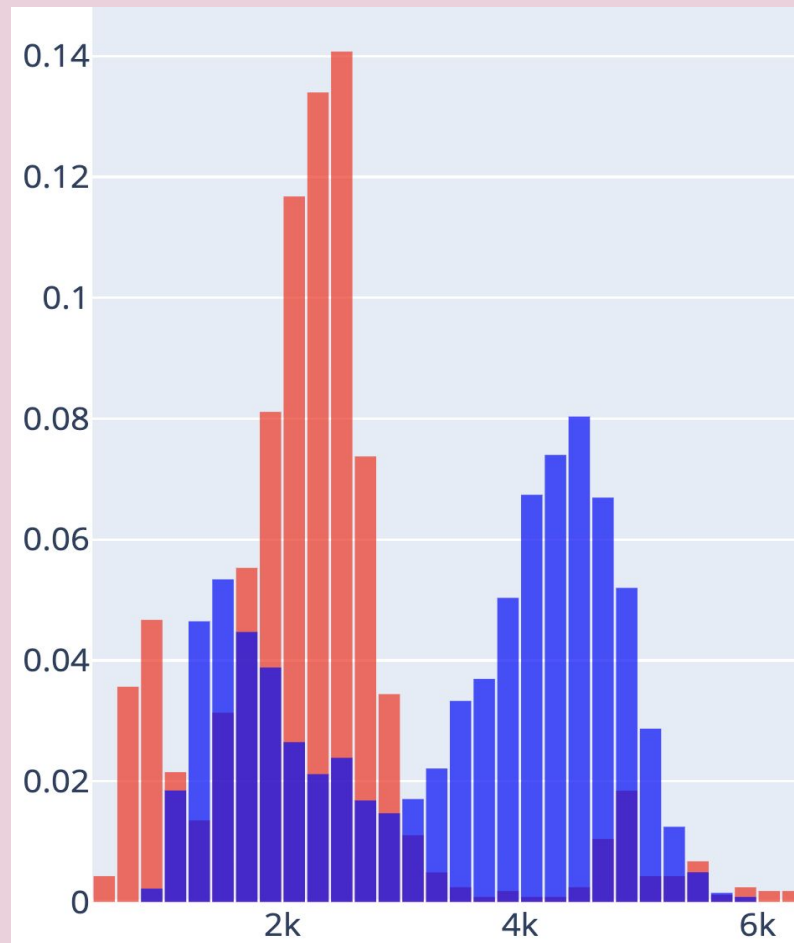
Распределение трат по кредитке

Траты по кредитке

зона риска: 1500-3000\$

**Отдельно работать с этим
сегментом клиентов**

- + Исследовать UX этих клиентов**
- + Интересоваться, что не нравится клиенту**



Распределение трат по кредитке

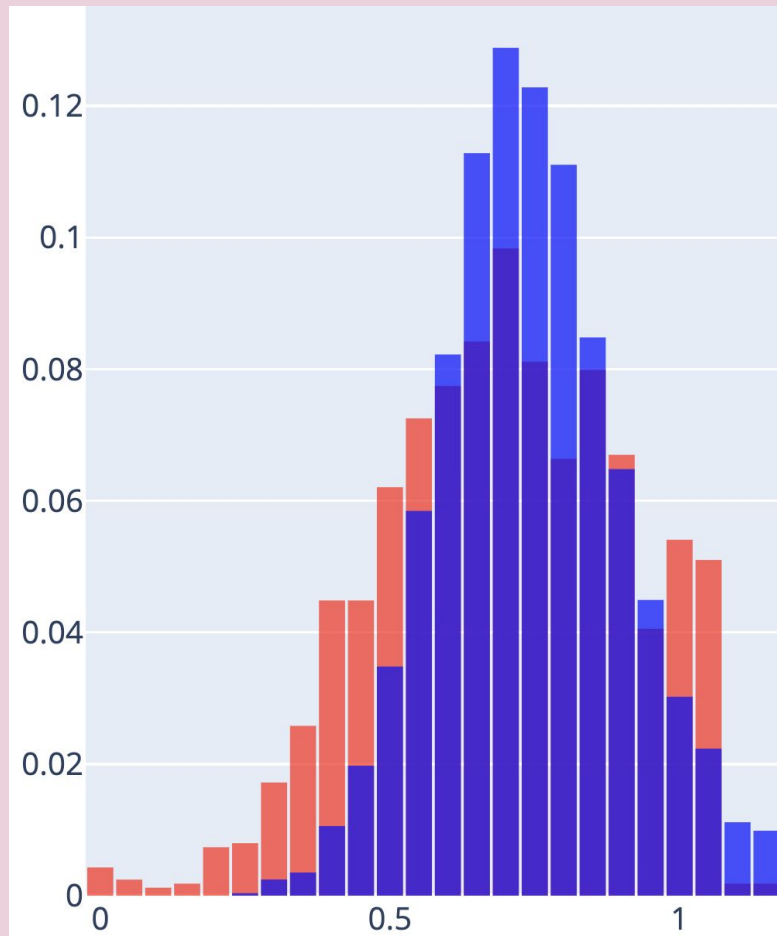
Траты по кредитке

Динамика суммы покупок:

Средняя динамика **активных** = 0.77

Средняя динамика **ушедших** = 0.69

Ушедшие снизили сумму покупок на **8%** больше



Распределение динамики суммы покупок

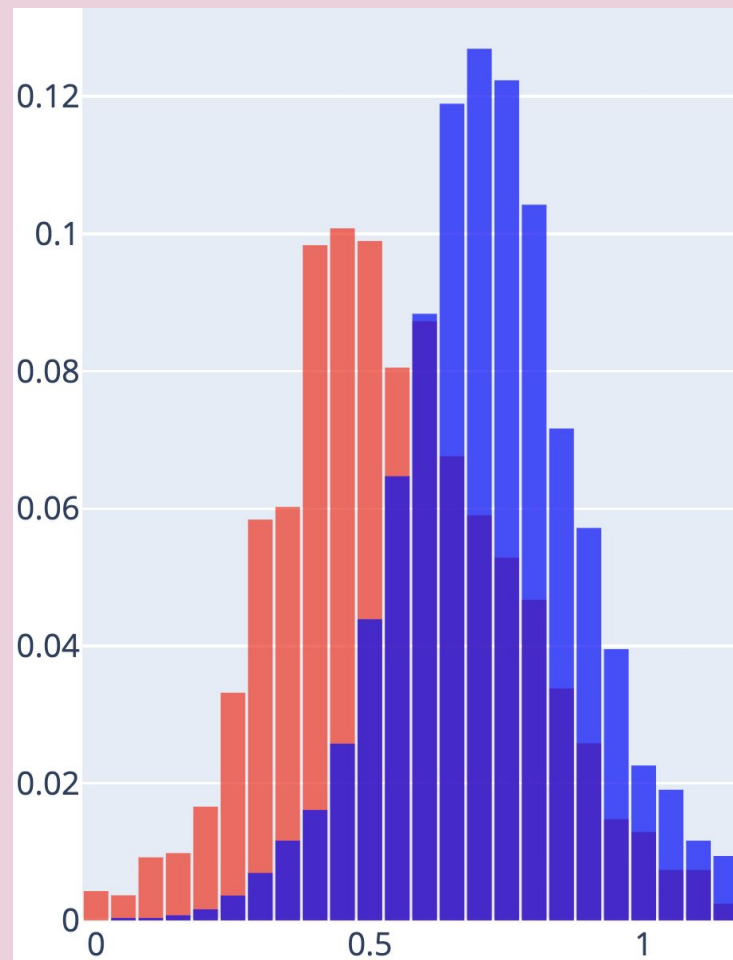
Траты по кредитке

Динамика кол-ва покупок:

Средняя динамика **активных** = 0.74

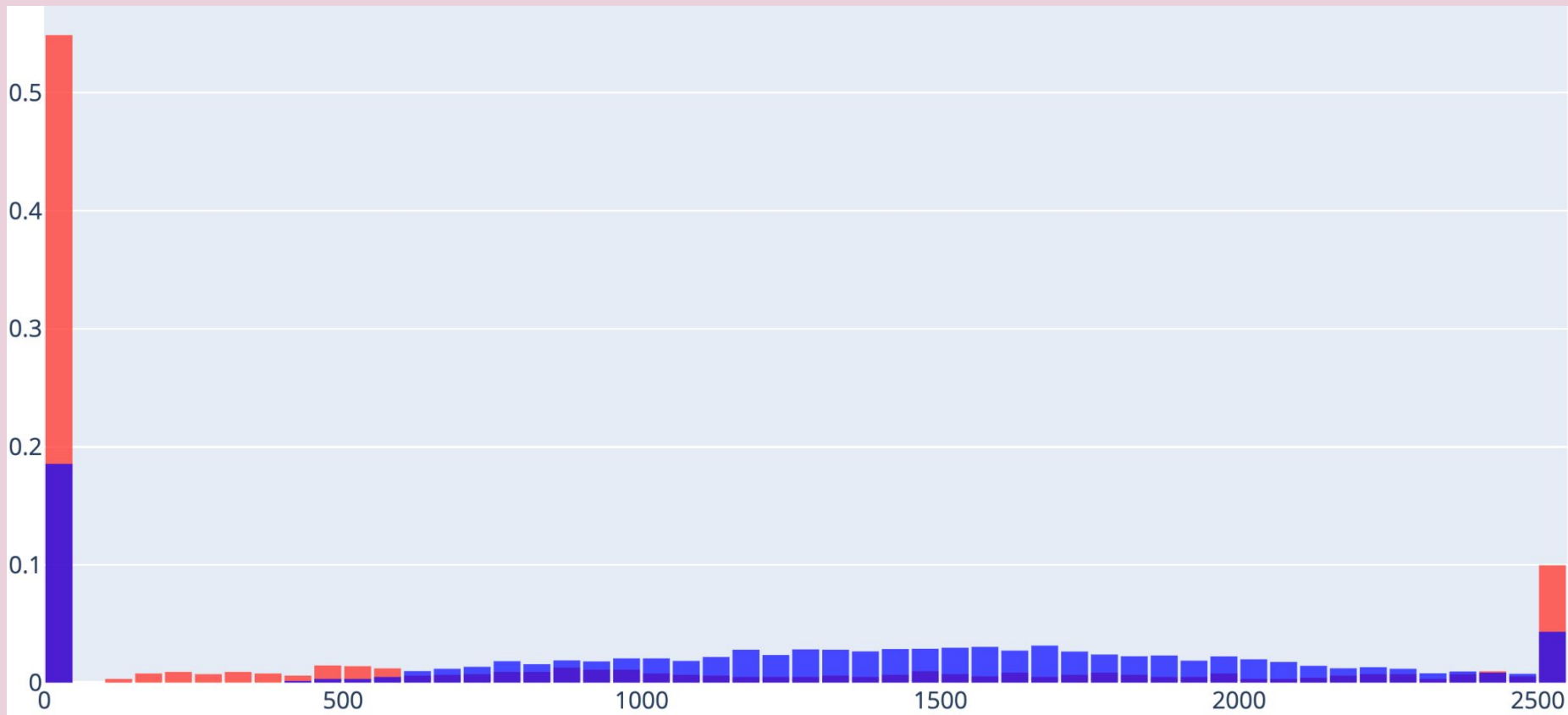
Средняя динамика **ушедших** = 0.55

Ушедшие снизили кол-во покупок на **19%**
больше

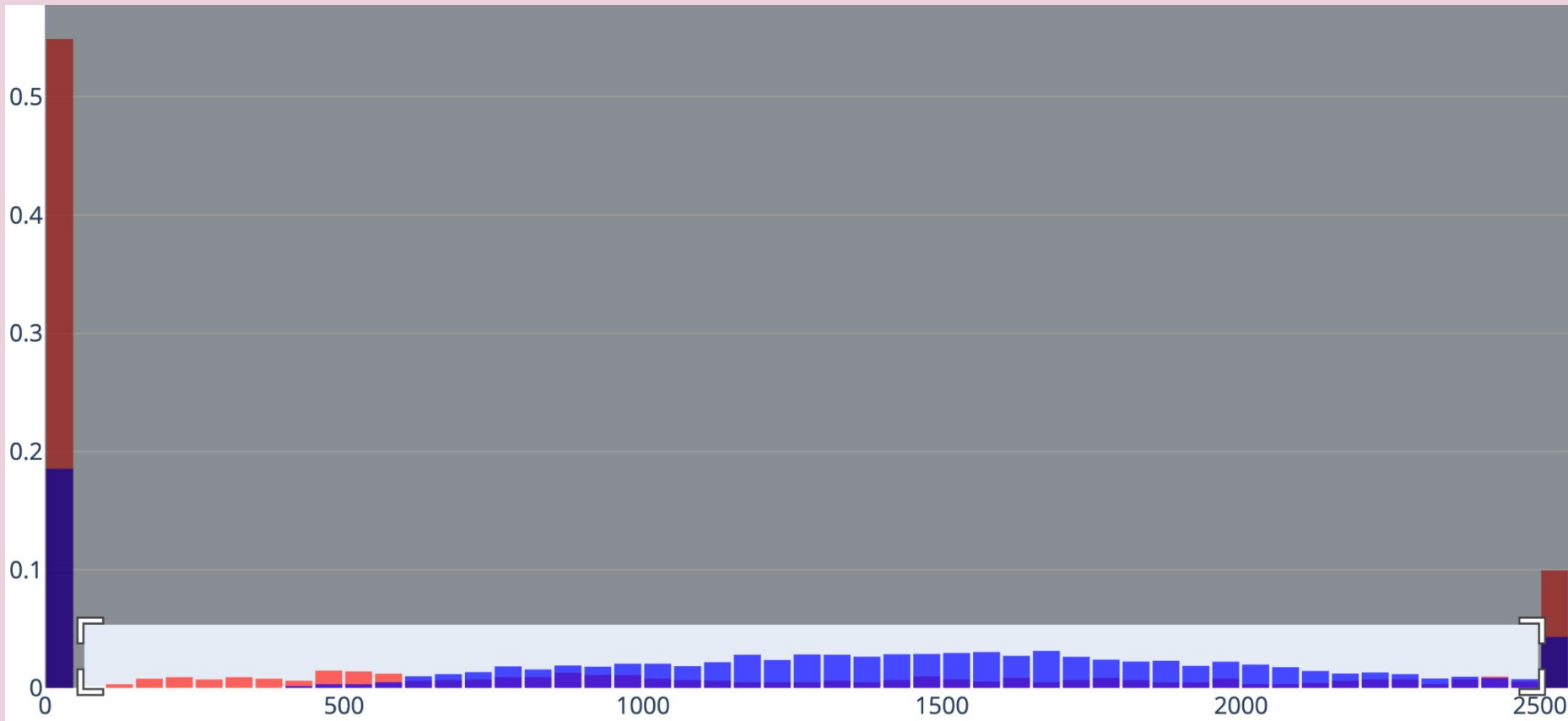


Распределение динамики кол-ва покупок

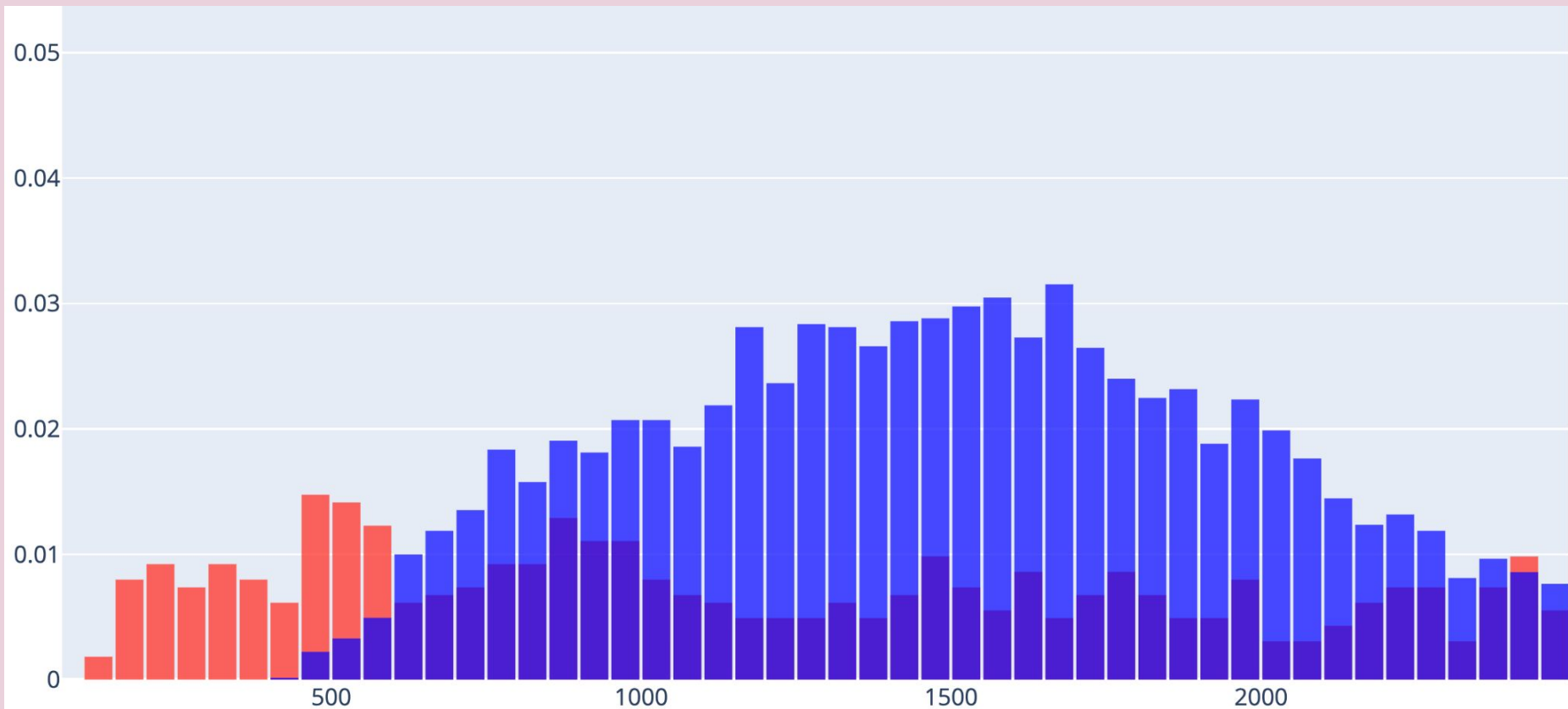
Непогашенная сумма долга



Непогашенная сумма долга



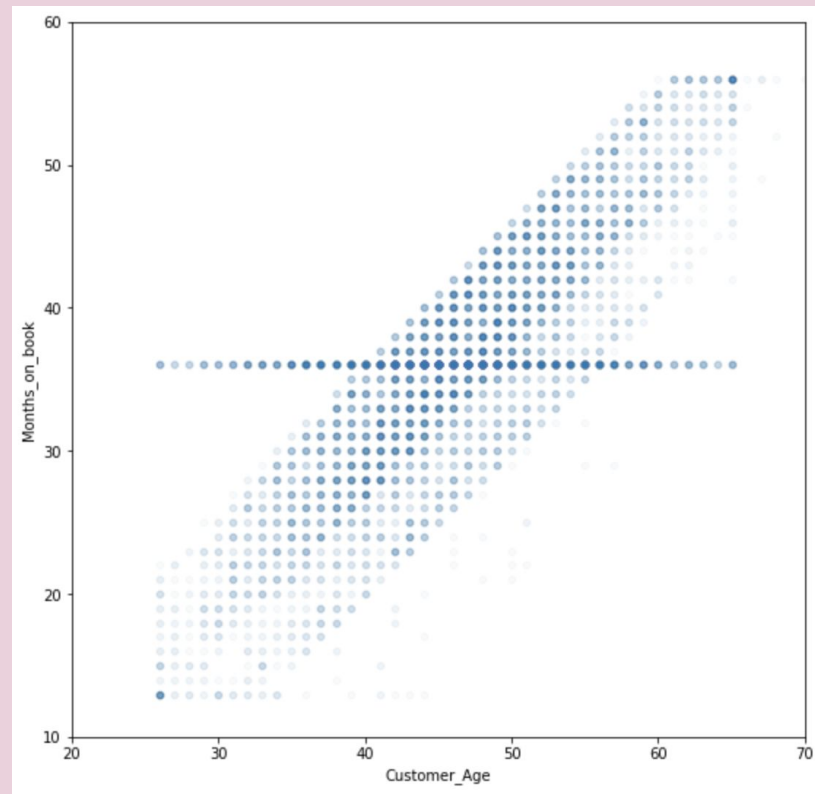
Непогашенная сумма долга



Непогашенная сумма долга

- + Стимулировать клиентов выходить на оптимальный остаток по счету - 1487\$**
- + Отдельно работать с клиентами, которые набирают большой долг**

Как связаны возраст и lifetime клиента?

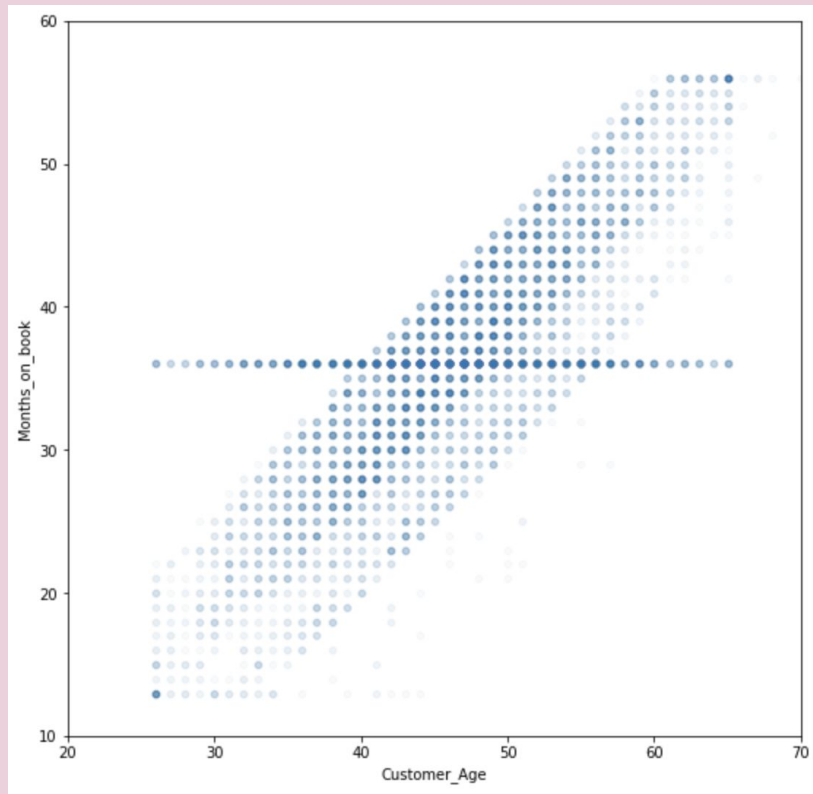


Months_on_Book - время сотрудничества с клиентом, в месяцах

Как связаны возраст и lifetime клиента?

Коэффициент корреляция Пирсона = 0.79

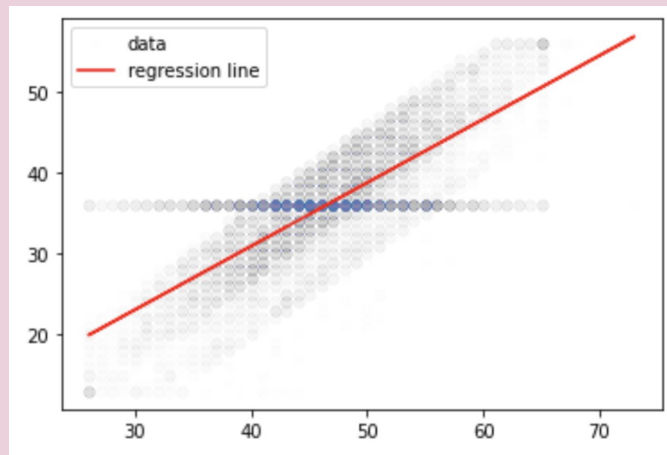
Расчет p-value излишен, т. к. Работаем с базой всех клиентов, т.е. с генеральной совокупностью



Months_on_Book - время сотрудничества с клиентом, в месяцах

Построим регрессионную модель

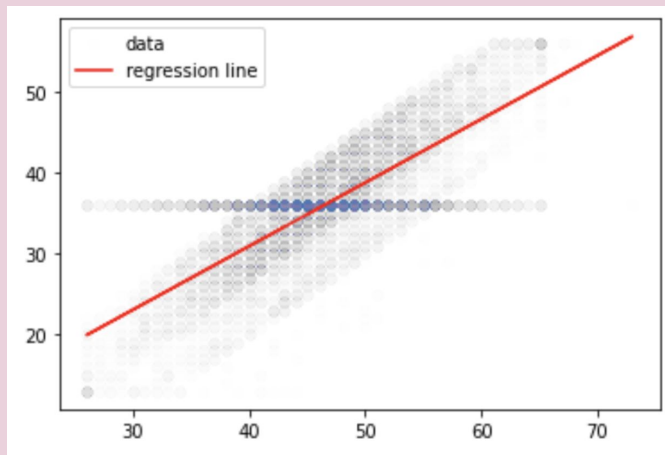
Будем использовать библиотеку statsmodels



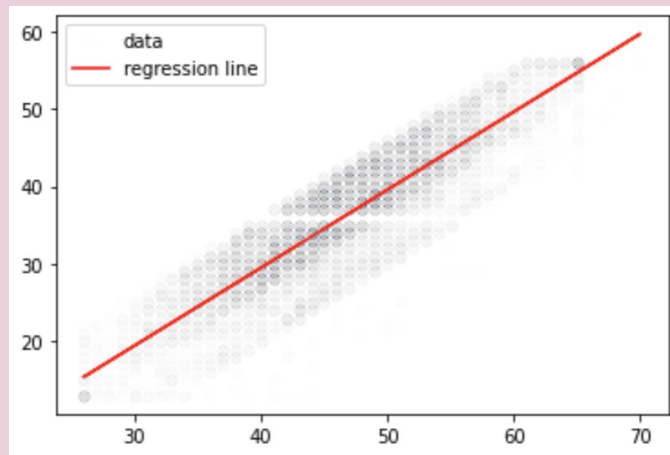
R-squared = 0.622

Построим регрессионную модель

Будем использовать библиотеку statsmodels



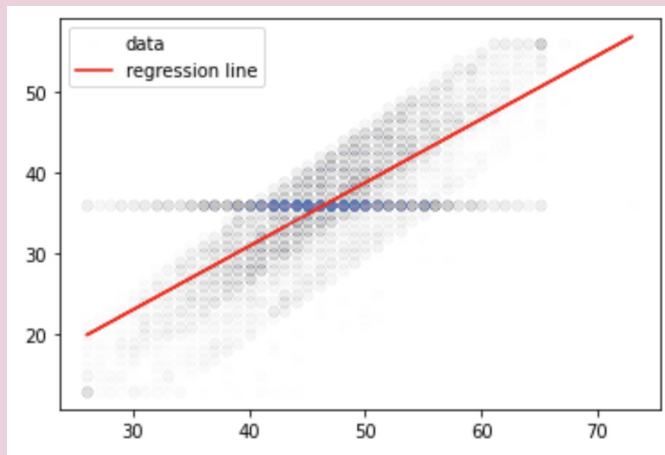
R-squared = 0.622



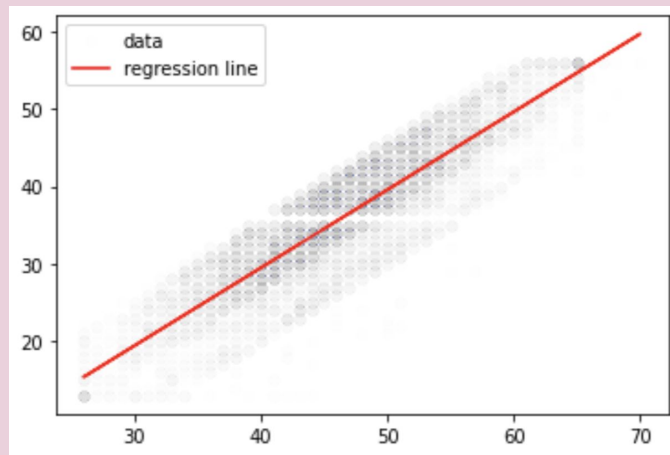
R-squared = 0.797

Построим регрессионную модель

Будем использовать библиотеку statsmodels



R-squared = 0.622



R-squared = 0.797

Модель объясняет 80% изменчивости зависимой переменной

Построим регрессионную модель

$$\text{время_сотрудничества} = -10.77 + \text{возраст} * 1.0064$$

OLS Regression Results

Dep. Variable:	Months_on_book	R-squared:	0.797			
Model:	OLS	Adj. R-squared:	0.797			
Method:	Least Squares	F-statistic:	3.010e+04			
Date:	Sat, 22 May 2021	Prob (F-statistic):	0.00			
Time:	19:45:15	Log-Likelihood:	-21754.			
No. Observations:	7664	AIC:	4.351e+04			
Df Residuals:	7662	BIC:	4.353e+04			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	-10.7739	0.273	-39.440	0.000	-11.309	-10.238
Customer_Age	1.0064	0.006	173.493	0.000	0.995	1.018
Omnibus:	513.257	Durbin-Watson:	2.022			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	619.771			
Skew:	-0.684	Prob(JB):	2.62e-135			
Kurtosis:	3.265	Cond. No.	272.			

параметры регрессионной модели

Построим регрессионную модель

$$\text{время_сотрудничества} = -10.77 + \text{возраст} * 1.0064$$

То есть с повышением возраста клиента на 1 год,
время сотрудничества в среднем увеличивается
на 1 месяц !

1 год

1 месяц

OLS Regression Results

Dep. Variable:	Months_on_book	R-squared:	0.797			
Model:	OLS	Adj. R-squared:	0.797			
Method:	Least Squares	F-statistic:	3.010e+04			
Date:	Sat, 22 May 2021	Prob (F-statistic):	0.00			
Time:	19:45:15	Log-Likelihood:	-21754.			
No. Observations:	7664	AIC:	4.351e+04			
Df Residuals:	7662	BIC:	4.353e+04			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	-10.7739	0.273	-39.440	0.000	-11.309	-10.238
Customer_Age	1.0064	0.006	173.493	0.000	0.995	1.018
Omnibus:	513.257	Durbin-Watson:	2.022			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	619.771			
Skew:	-0.684	Prob(JB):	2.62e-135			
Kurtosis:	3.265	Cond. No.	272.			

параметры регрессионной модели

Как связаны возраст и LTV клиента?

Как связаны возраст и LTV клиента?

Среднестатистический клиент **Клиент в возрасте 55-70 лет**

Как связаны возраст и LTV клиента?

**Среднестатистический
клиент**

Прибыль в месяц = 19.44\$

**Клиент в возрасте 55-70
лет**

Прибыль в месяц = 19.96\$

Как связаны возраст и LTV клиента?

**Среднестатистический
клиент**

Прибыль в месяц = 19.44\$

Lifetime, в месяцах = 36

**Клиент в возрасте 55-70
лет**

Прибыль в месяц = 19.96\$

Lifetime, в месяцах = 49

Как связаны возраст и LTV клиента?

**Среднестатистический
клиент**

Прибыль в месяц = 19.44\$

Lifetime, в месяцах = 36

LTV = 697.94\$ USD

**Клиент в возрасте 55-70
лет**

Прибыль в месяц = 19.96\$

Lifetime, в месяцах = 49

LTV = 976.09\$ USD

Как связаны возраст и LTV клиента?

**Среднестатистический
клиент**

Прибыль в месяц = 19.44\$

Lifetime, в месяцах = 36

LTV = 697.94\$ USD

**Клиент в возрасте 55-70
лет**

Прибыль в месяц = 19.96\$

Lifetime, в месяцах = 49

LTV = 976.09\$ USD

Клиенты в возрасте приносят компании на 278\$ больше среднего

Дашборд с метриками-алармами

Churn rate

0.1607

LTV

683.0

Purchase count
change

0.7122

Number of clients with
1500\$-3000\$ expenses

0.2849

Customer Age



Gender

☒ (All)

☐ F

☐ M

Card Category

☒ (All)

☒ Blue

☒ Gold

☒ Platinum

☒ Silver

Marital Status

(All)

Education Level

(All)

Дашборд с метриками-алармами

Churn rate

0.1342

LTV

424.6

Purchase count
change

0.6926

Number of clients with
1500\$-3000\$ expenses

0.5368

Customer Age

26

32



Gender

☒ (All)

☐ F

☐ M

Card Category

☒ (All)

☒ Blue

☒ Gold

☒ Platinum

☒ Silver

Marital Status

(All)

Education Level

(All)

Дашборд с метриками-алармами

Churn rate

0.1432

LTV

963.4

Purchase count
change

0.7126

Number of clients with
1500\$-3000\$ expenses

0.3580

Customer Age

61 73

Gender

☒ (All)

☐ F

☐ M

Card Category

☒ (All)

☒ Blue

☒ Gold

☒ Platinum

☒ Silver

Marital Status

(All)

Education Level

(All)

Гипотезы по улучшению

★ стимулировать клиентов активнее
★ использовать кредитные возможности карты

важные метрики активности

Общая сумма трат

★ выводить клиентов на большие траты по кредитке

Частота покупок

★ Стараться поддерживать частоту покупок клиентов на том же уровне что и была ранее

★ не давать клиентам сильно залезать в долги

на раннем этапе определять клиентов, которые месяц за месяцем накапливают все больший долг по карте, и работать с этой группой пользователей, на то чтобы они не оказались в ситуации максимально накопленного долга

(см. п. 2.4)

★ формировать культуру пользования кредитной картой

стимулировать клиента "выходить" на уровень остатка на счету, соответствующий остатку, характерному для существующих клиентов с его картой

(см. п. 2.4)

★ Точечное улучшение сервиса

(см. п. 2.5)

Исследовать UX клиентов с тратами 1500-3000

Интересоваться, что не нравится клиенту

★ Делать упор на клиентов в возрасте

Стараться привлекать новых клиентов, среди людей зрелого и преклонного возраста 47 - 61 года

65 (максимальный среди представленных в данных возраст) - 4 года (среднее время сотрудничества в таком возрасте) = 61

так мы получим возраст, на момент начала сотрудничества

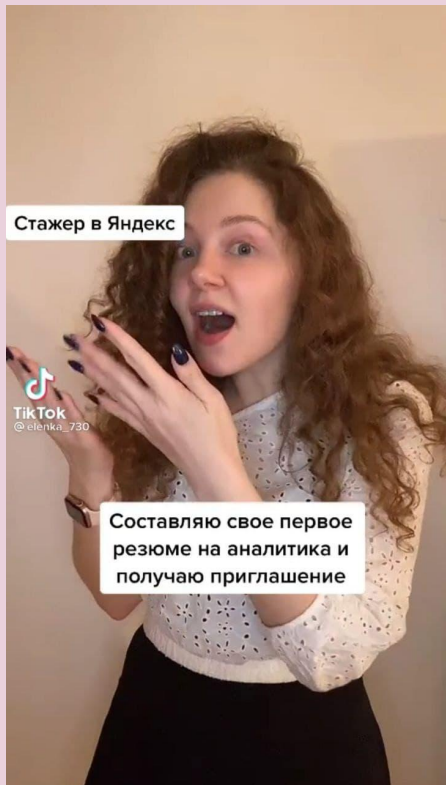
(см. п. 3.1)

**На
ЭТОМ
Все.**

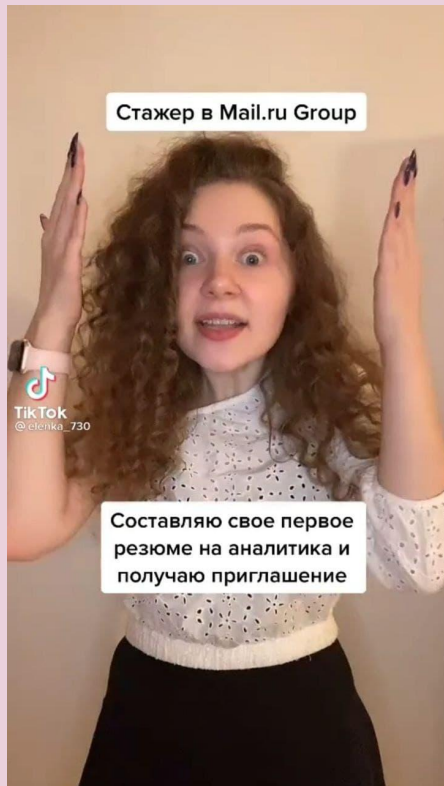
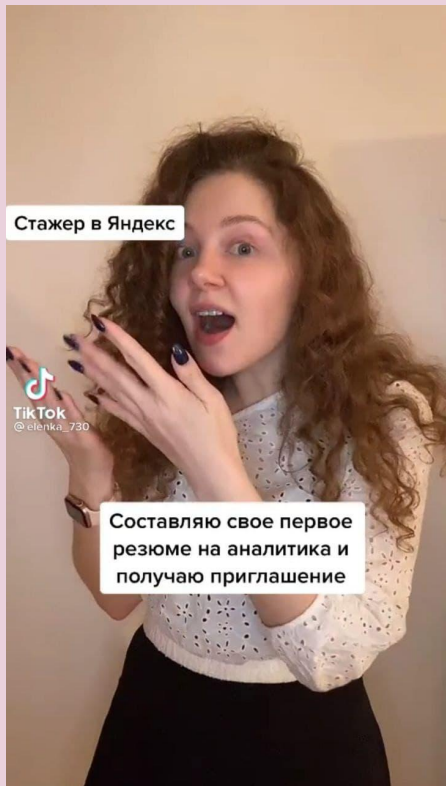
Ну а я пошел рассылать резюме..

Ну а я пошел рассылать резюме..

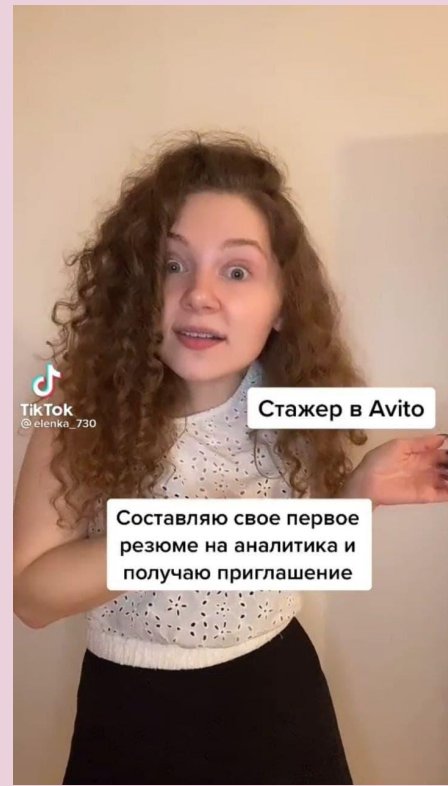
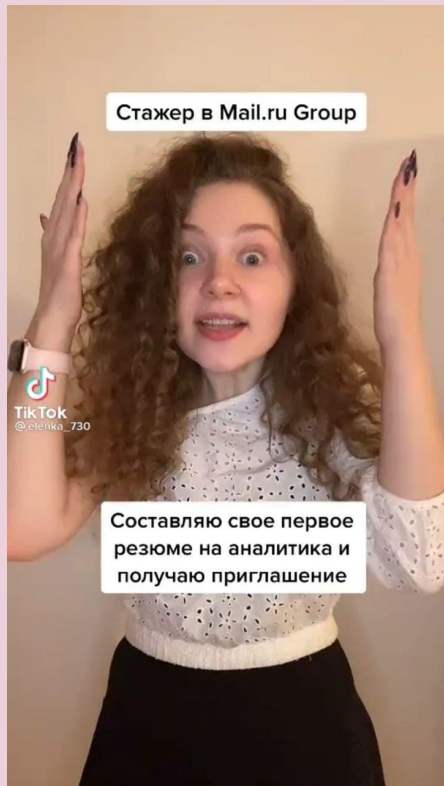
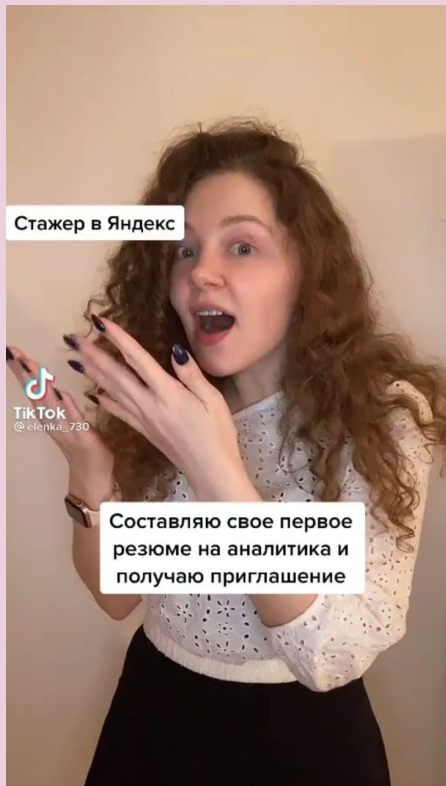
Ну а я пошел рассылать резюме..

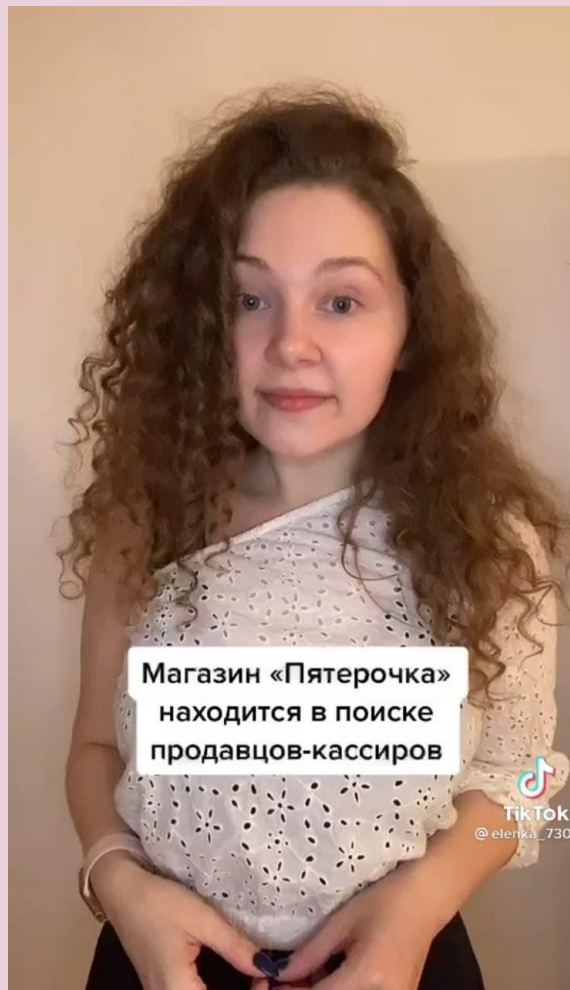


Ну а я пошел рассылать резюме..



Ну а я пошел рассылать резюме..





Магазин «Пятерочка»
находится в поиске
продавцов-кассиров


TikTok
@elenka_730

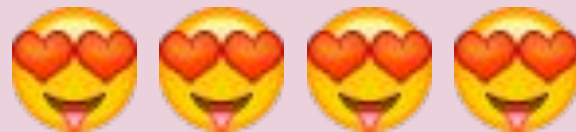
Спасибо

за

ВНИМАНИЕ!

Лена, спасибо за прекрасный курс !

СпИсОк ЛиТеРаТуРы



Само исследование:

<https://github.com/Penguin-Run/TP-Analytics-credit-card-customers>

Данные взяты из Kaggle:

<https://www.kaggle.com/sakshigoyal7/credit-card-customers>