Кейс:

кэшбэк

Состав команды - студенты группы Э-2209
Василенко Егор
Казённых Максим
Лёвкина Екатерина
Серкибаев Аким
Салихова Самира

01 ф Коммуникаци ф

Как взаимосвязаны коммуникации и покупки с кэшбэком?



Цели

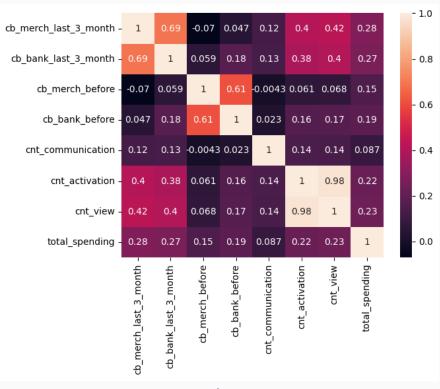
- выделить группы клиентов, схожих по активности
- установить взаимосвязь между коммуникациями и выплаченным кэшбэком, коммуникациями и тратами клиентов, количеством высланных коммуникаций о предложениях с кэшбэком и просмотров этих предложений, количеством просмотров и активаций предложений с кэшбэком

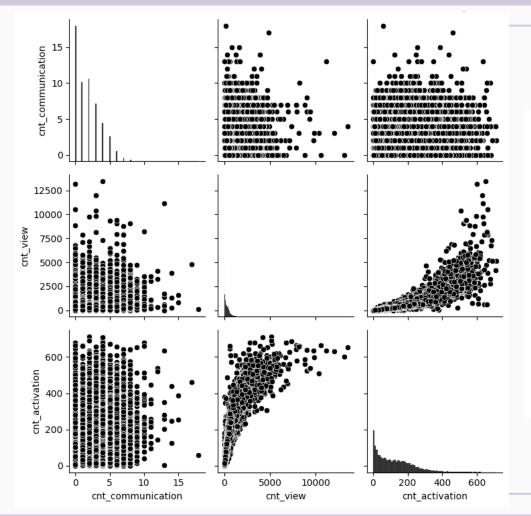
Столбцы в датасете:

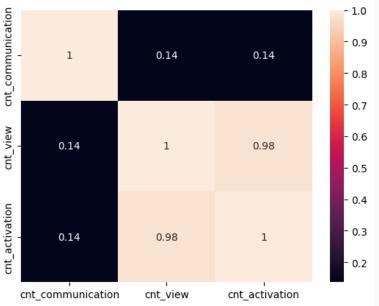
- cnt_communication число отправленных коммуникаций о кэшбэках
- cnt_activation число активированных предложений за последние 3 месяца
- cnt_view число просмотров предложений с кэшбэком за последние 3 месяца



Корреляция

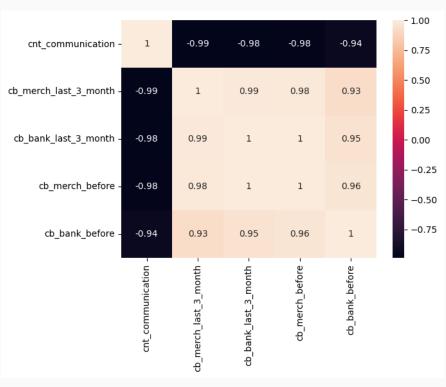








Коммуникации и кэшбэк

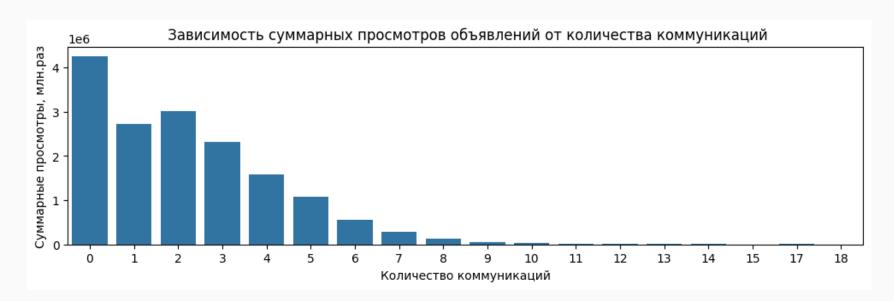




Медианные траты клиентов на покупки с кэшбэком за последние 3 месяца при определенном количестве коммуникаций 20000 -Траты клиентов, руб

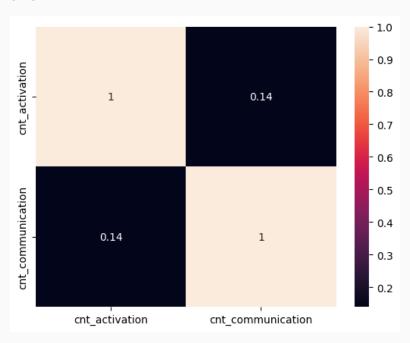
Количество коммуникаций







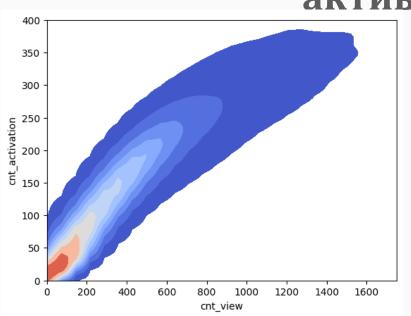
Количество отправленных предложений и активации

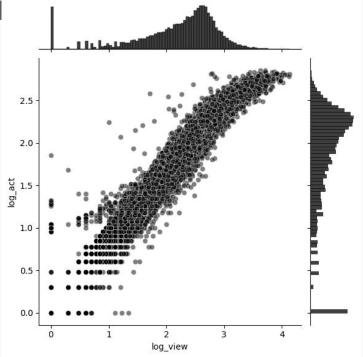




Просмотры предложений и их



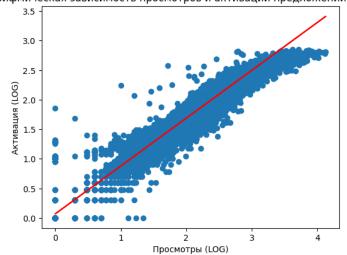


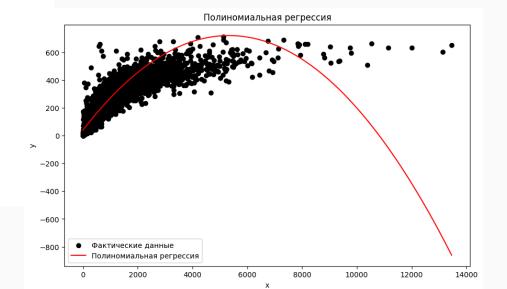


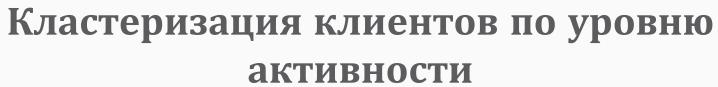


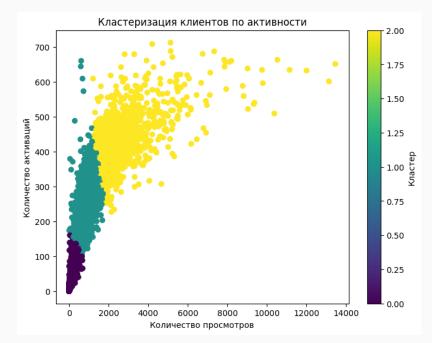
Просмотры предложений и их

Логарифмическая зависимость просмотров и активации предложений с кэшбэком **ация**









- Кластер 0 «наименее активные пользователи»
- Кластер 1 «средне-активные пользователи»
- Кластер 2 «пользователи с высокой активностью»

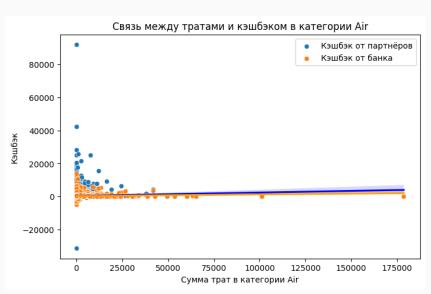
02 †

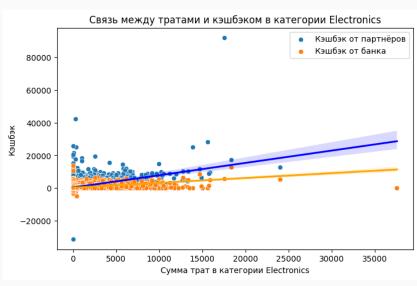
Обороты клиентов

Как взаимосвязаны обороты клиентов и выплаченный кэшбэк?

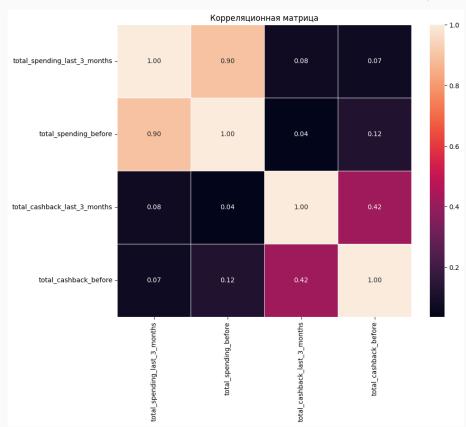


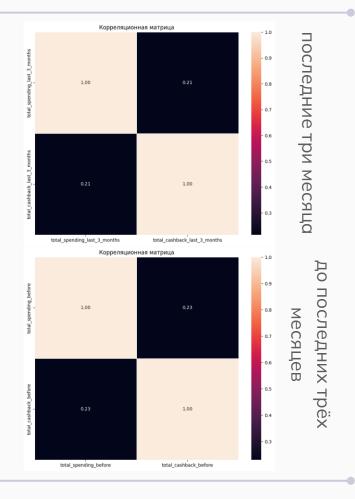
Для изначального датасета





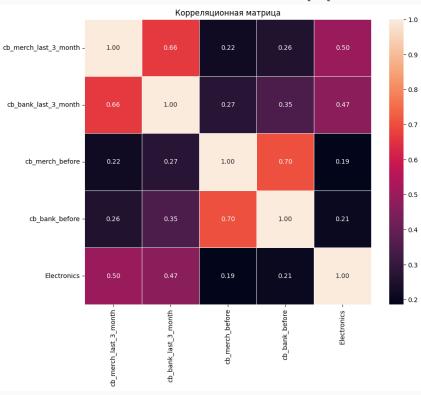






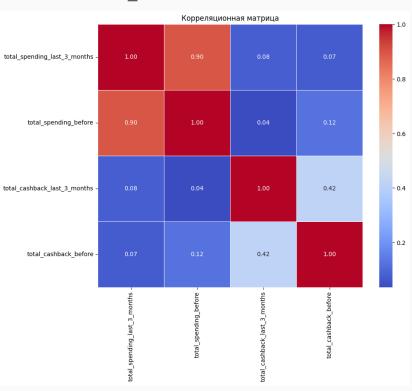


Для изначального датасета

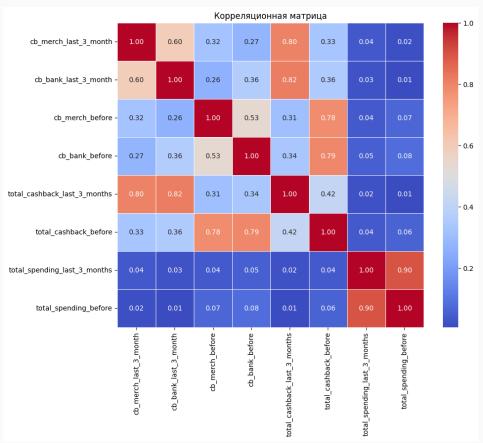




Для нового обработанного датасета

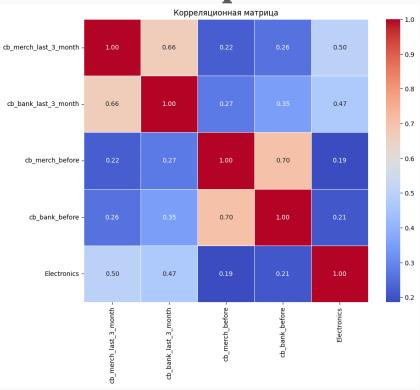








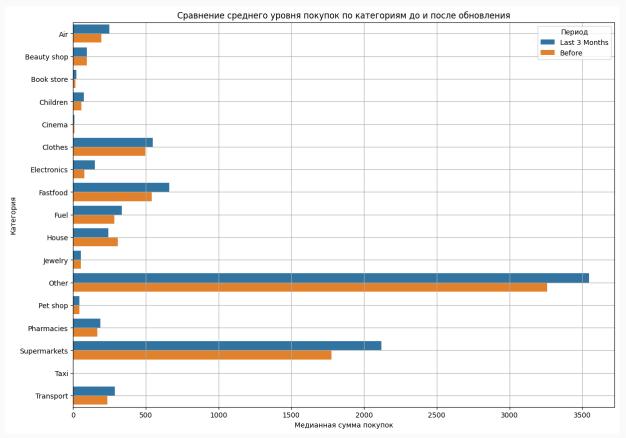
Уточнение по категории «Электроника»



оз ф Обновлени ф

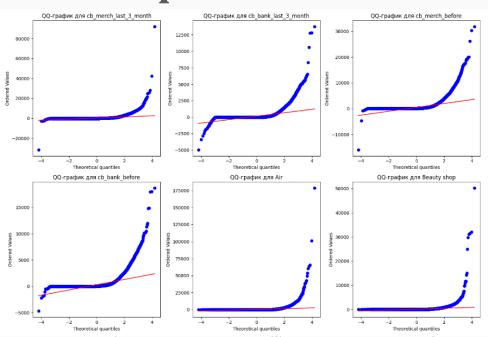
3 месяца назад произошло глобальное обновление приложения, в т.ч. методики взаимодействия с пользователем. Необходимо оценить эффект обновления.



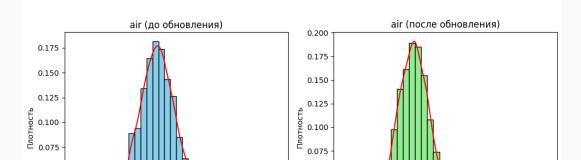




Проверка изначальных данных на нормальность







0.050

0.025

0.000

Гистограммы для air

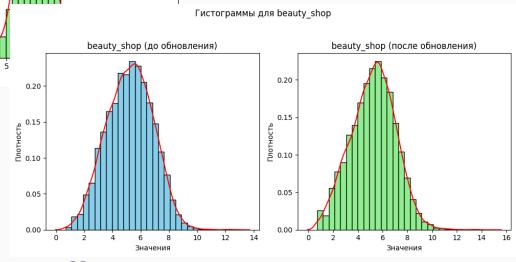
Без обработки выбросов

Значения

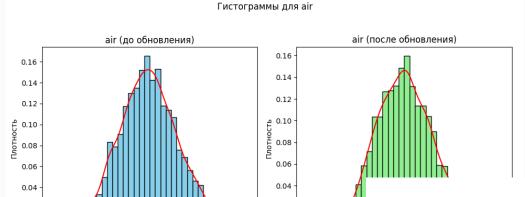
0.050

0.025

0.000







0.02

2.5

С обработкой выбросов

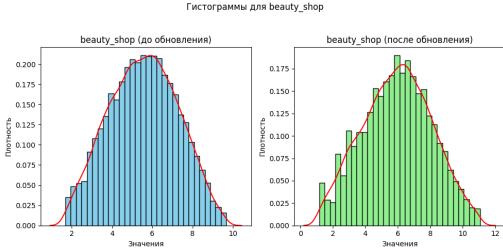
12.5

10.0

Значения

15.0

0.02





Построение гипотез и проведение

- **Нулевая гипотеза (Н0)**: Обновление приложения не оказало значительного влияния на метрику (например, средний кэшбэк, сумму покупок и т.д.) в рассматриваемой категории. То есть средние значения до и после обновления не отличаются.
- **Альтернативная гипотеза (Н1)**: Обновление приложения оказало значительное влияние на метрику в рассматриваемой категории, и средние значения до и после обновления отличаются.

Используем t-тест для зависимых парных выборок (то есть используя метод t test_rel)

Отвергаем нулевую гипотезу о равенстве средних значений до и после обновления.

04 ф Гипотеза ф

Выдвиньте собственную гипотезу на основе имеющихся данных.



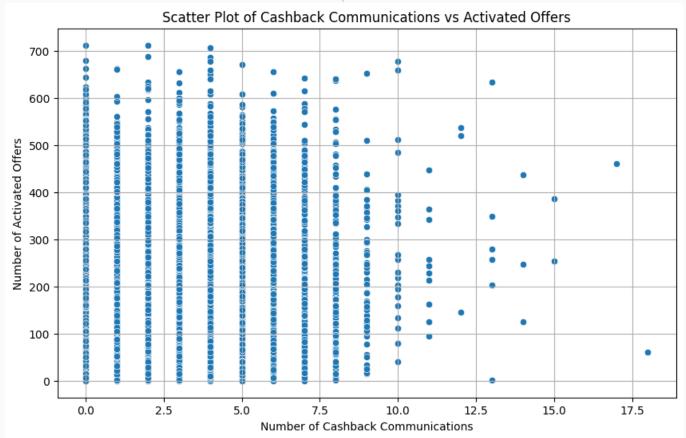
• Н0 (нулевая гипотеза):

Количество сообщений кэшбэка не влияет на количество активированных предложений (т.е. нет значимой связи между количеством отправленных сообщений и количеством активированных предложений).

• Н1 (альтернативная гипотеза):

Количество кэшбэк-рассылок положительно влияет на количество активированных офферов (т.е. существует значимая положительная связь между количеством отправленных рассылок и количеством активированных офферов.







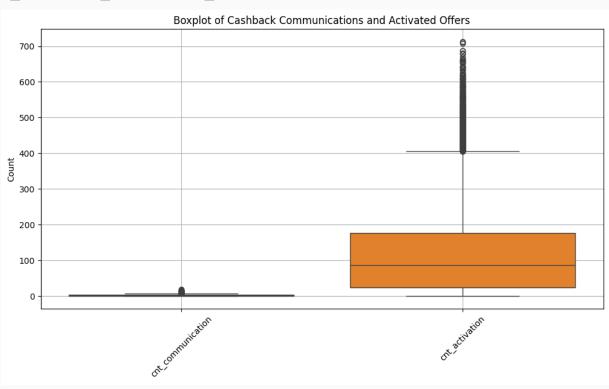
Анализ линейной регрессии

				_	_	
	OL	S Regress	ion Results			
						==
Dep. Variable:	cnt_activation		•		0.034	
Model:			Adj. R-squared:		0.034	
Method:	Least Squares				1780.	
			Prob (F-statistic):			
Time:			Log-Likelihood:		-3.0260e+05	
No. Observations:	50000		AIC:		6.052e+05	
Df Residuals:	49998		BIC:		6.052e+05	
Df Model:		1				
Covariance Type:	no	nrobust				
			t		-	0.975]
const						94.383
cnt_communication	10.9768					
 Omnibus:	8992.310		Durbin-Watson:		1.992	
Prob(Omnibus):	0.000		Jarque-Bera (JB):		15904.471	
skew:			Prob(JB):		0.00	
Kurtosis:	4.505		Cond. No.		3.74	
						==
Notes:						
[1] Standard Errors	assume tha	t the cov	ariance matrix	k of the err	ors is corre	ctly speci-

Учитывая значимое р-значение, мы можем отвергнуть нулевую гипотезу (Н0) и принять альтернативную гипотезу (Н1), заключив, что количество кэшбэк-коммуникаций положительно влияет на количество активированных офферов.



Проверка проведённого анализа



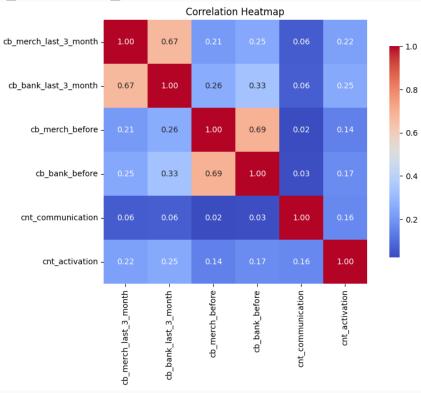


Проверка проведённого анализа

Dep. Variable:	cnt activation		R-squared:		0.034	
Model:	<u>-</u>		Adj. R-squared:		0.034	
Method:	Least Squares		F-statistic:		1780.	
Date:	Thu, 14 Nov 2024		Prob (F-statistic):		0.00	
Time:	00:19:33		Log-Likelihood:		-3.0260e+05	
No. Observations:	50000		AIC:		6.052e+05	
Df Residuals:	49998		BIC:		6.052e+05	
Df Model:		1				
Covariance Type:	no	nrobust				
			t	P> t	[0.025	0.975
const			145.140	0.000	91.868	94.383
cnt_communication	10.9768	0.260	42.186	0.000	10.467	11.487
		=======				==
Omnibus:			Durbin-Watson:		1.992	
Prob(Omnibus):			Jarque-Bera (JB):			
Skew:			Prob(JB):		0.00	
Kurtosis:	4.505		Cond. No.		3.74	



Проверка проведённого анализа





Множественный регрессионный анализ

	OLS Regre	ession	Results			
	- OLS Least Squares Thu, 14 Nov 2024 00:19:35 49028 49022	5 Ad 5 F- 1 Pr 5 Lc 3 AI 2 BI	j. R-squared: statistic: ob (F-statistic): g-Likelihood: C:		0.098 0.098 1070. 0.00 -2.8990e+05 5.798e+05 5.799e+05	
	coef s	td er	r t	P> +	 [0.025	0.9751
const cb_merch_last_3_month cb_bank_last_3_month cb_merch_before cb_bank_before cnt_communication	81.7614 0.0088 0.0339 0.0026 0.0089 8.0968	0.59 0.00 0.00 0.00 0.00	4 137.574 1 15.903 1 25.221 0 5.388 1 11.562 0 33.739	0.000 0.000 0.000 0.000 0.000 0.000	80.597 0.008 0.031 0.002 0.007 7.626	82.926 0.010 0.036 0.004 0.010
Omnibus: Prob(Omnibus): Skew: Kurtosis:	0.746) Ja) Pr	rbin-Watson: rque-Bera (JB): ob(JB): nd. No.		2.001 4582.761 0.00 2.24e+03	

Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 2.24e+03. This might indicate that there are strong multicollinearity or other numerical problems.



Множественный регрессионный анализ

```
import numpy as np

n_bootstrap = 10000

bootstrap_means = []
for _ in range(n_bootstrap):
    sample = np.random.choice(df_cleaned['cnt_activation'], size=len(df_cleaned), replace=True)
    bootstrap_means.append(np.mean(sample))

observed_mean = np.mean(df_cleaned['cnt_activation'])
bootstrap_means = np.array(bootstrap_means)

p_value_bootstrap = np.mean(bootstrap_means >= observed_mean)

(observed_mean, p_value_bootstrap)

(105.88863506567675, 0.493)
```



Спасибо за внимание!

