

```
import pandas as pd
import numpy as np
import datetime as dt
import seaborn as sns
import matplotlib.pyplot as plt
from scipy.stats import ttest_ind
from scipy.stats import mannwhitneyu
```

Импорт и анализ таблиц

```
df = pd.read_excel('Новый диплом датасет.xlsx', sheet_name = 'Данные')
df_clients = pd.read_excel('Новый диплом датасет.xlsx', sheet_name =
'Clients')
df_region_dict = pd.read_excel('Новый диплом датасет.xlsx', sheet_name
= 'Region_dict')
```

```
df.head()
```

	id_order	id_client	amt_payment	dtype_pay
0	100238	194569.0	5373.0	2022-10-28 04:05:15.221000
1	100242	191056.0	4151.0	2022-08-01 02:42:54.992000
2	100249	200049.0	5688.0	2022-08-24 12:45:12.744000
3	100258	206161.0	4330.0	2022-12-09 15:13:26.437000
4	100267	190945.0	5488.0	2022-08-03 02:07:31.468000

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 37989 entries, 0 to 37988
Data columns (total 4 columns):
#   Column          Non-Null Count  Dtype
---  -
0   id_order         37989 non-null   int64
1   id_client        37401 non-null   float64
2   amt_payment      35845 non-null   float64
3   dtype_pay        35828 non-null   object
dtypes: float64(2), int64(1), object(1)
memory usage: 1.2+ MB
```

Имеем нулловые значения. При исключении теряем около 5% выборки.

При этом строки, в которых нет данных по ид клиента не несут для нас никакой информации, их необходимо исключить. Строки, в которых нет данных о сумме оплаты также можем убрать сразу, так как при дальнейшей группировке получим тот же результат, что и при проставлении нулей.

```
df = df.dropna(subset=['amt_payment', 'id_client']).reset_index(drop =
True)
df_clients.head()
```

	id_client	datetime_ad	nflag_test	id_trading_point
0	180844	2022-06-08 18:38:41.414	0	212
1	193942	NaT	0	13
2	226069	2022-07-11 16:28:38.511	1	54
3	183981	2022-06-16 12:23:59.289	1	991
4	322530	2022-07-08 08:56:08.714	0	1015

```
df_clients.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 55605 entries, 0 to 55604
Data columns (total 4 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id_client             55605 non-null  int64
1   datetime_ad           55356 non-null  datetime64[ns]
2   nflag_test            55605 non-null  int64
3   id_trading_point      55605 non-null  int64
dtypes: datetime64[ns](1), int64(3)
memory usage: 1.7 MB
```

Имеем нуловые значения по дате показа рекламного объявления. Их необходимо исключить, ведь если клиент не получал уведомление, мы не можем оценивать реакцию на него.

```
df_clients.dropna(inplace = True)
```

```
df_region_dict.head()
```

	id_trading_point	city
0	1	Петрозаводск
1	2	Дмитров
2	4	Чехов
3	7	Ростов Великий
4	11	Владимир

```
df_region_dict.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 66 entries, 0 to 65
Data columns (total 2 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id_trading_point      66 non-null    int64
1   city                  66 non-null    object
dtypes: int64(1), object(1)
memory usage: 1.2+ KB
```

Исследуем количество торговых точек в каждом городе

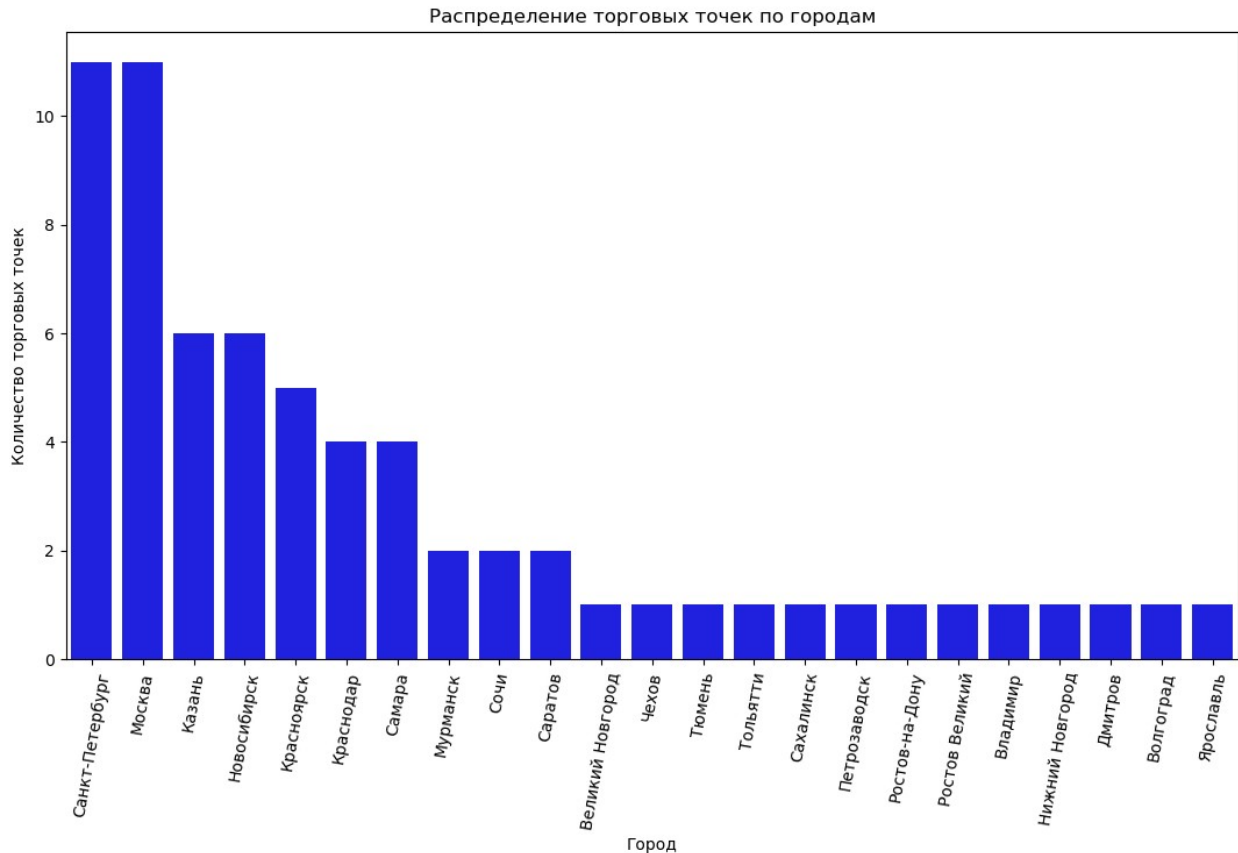
```
df_region_group = df_region_dict.groupby('city')
['id_trading_point'].count().\
reset_index().sort_values(by = 'id_trading_point', ascending = False)
df_region_group
```

	city	id_trading_point
15	Санкт-Петербург	11
7	Москва	11
4	Казань	6
10	Новосибирск	6
6	Красноярск	5
5	Краснодар	4
14	Самара	4
8	Мурманск	2
18	Сочи	2
16	Саратов	2
0	Великий Новгород	1
21	Чехов	1
20	Тюмень	1
19	Тольятти	1
17	Сахалинск	1
11	Петрозаводск	1
13	Ростов-на-Дону	1
12	Ростов Великий	1
1	Владимир	1
9	Нижний Новгород	1
3	Дмитров	1
2	Волгоград	1
22	Ярославль	1

```
plt.figure(figsize=(13,7))
sns.barplot(x = df_region_group['city'], y =
df_region_group['id_trading_point'], color = 'blue')

plt.xlabel('Город')
plt.ylabel('Количество торговых точек')
plt.title('Распределение торговых точек по городам')
plt.xticks(rotation=80)

plt.show()
```



Объединение таблиц

Вычислим сумму платежей для каждого клиента и создадим флаг оплаты

```
df_payment = df.groupby('id_client')
['amt_payment'].sum().reset_index()
df_payment.head()
```

	id_client	amt_payment
0	178561.0	3052.0
1	178562.0	2439.0
2	178563.0	768.0
3	178565.0	958.0
4	178566.0	5474.0

```
df_payment = df_clients.merge(df_payment, how = 'left', on =
'id_client')\
.merge(df_region_dict, how = 'left', on = 'id_trading_point')
```

```
df_payment['amt_payment'].fillna(0.0, inplace=True)
```

```
df_payment['flag_payment'] = np.where(df_payment['amt_payment'] <= 0,
0, 1)
```

```
df_payment.head()
```

	id_client	dttime_ad	nflag_test	id_trading_point	\
0	180844	2022-06-08 18:38:41.414	0	212	
1	226069	2022-07-11 16:28:38.511	1	54	
2	183981	2022-06-16 12:23:59.289	1	991	
3	322530	2022-07-08 08:56:08.714	0	1015	
4	254313	2022-06-19 22:18:01.770	1	453	

	amt_payment	city	flag_payment
0	5349.0	Красноярск	1
1	0.0	Мурманск	0
2	6322.0	Казань	1
3	0.0	Краснодар	0
4	0.0	Санкт-Петербург	0

Автоматизация статистических вычислений

Создадим функции test_calc и mann_whitney_func

```
def test_calc(r1, r2, alpha = 0.05):

    s, p = ttest_ind(r1, r2)
    print('Значение t критерия:', s)
    print('Значение p value:', p)
    if p<alpha:
        print('Разница между средними статистически значима.')
    else:
        print('Разница между средними случайна')
    # return s, p

def mann_whitney_func(r1, r2, alpha = 0.05):

    u, p = mannwhitneyu(x = r1, y = r2, alternative = 'two-sided')
    print('Значение u критерия:', u)
    print('Значение p value:', p)
    if p<alpha:
        print('Разница между распределениями статистически значима.')
    else:
        print('Разница между распределениями случайна')
    # return u, p
```

Чистка неверно заполненных точек

Соберем торговые точки, в которых не было заплачено ни одного рубля ни одним клиентом.

```
null_payment_point = []

df_sum_payment = df_payment.groupby(['id_trading_point', 'nflag_test'])
['amt_payment'].sum().reset_index()

for i in df_sum_payment.index:
```

```

    if df_sum_payment['amt_payment'][i]==0:
        null_payment_point.append(df_sum_payment['id_trading_point']
[i])

null_payment_point = list(set(null_payment_point))
print(null_payment_point)

[1, 4, 7, 13, 46, 23, 26]

```

Соберем торговые точки, в которых пустует либо контрольная, либо тестовая группа.

```

null_count_point = []

df_count_client =
df_payment.groupby(['id_trading_point', 'nflag_test'])
['id_client'].count().reset_index()
df_count_group =
df_count_client.groupby('id_trading_point').count().reset_index()

for i in df_count_group.index:
    if df_count_group['nflag_test'][i]<2:
        null_count_point.append(df_count_group['id_trading_point'][i])
print(null_count_point)

[228, 603, 739, 800, 810, 866, 1015, 1099]

```

Отбросим определенные ранее торговые точки

```

df_payment =
df_payment[~df_payment['id_trading_point'].isin(null_payment_point)]
df_payment =
df_payment[~df_payment['id_trading_point'].isin(null_count_point)]

```

Расчет общих результатов АБ Теста

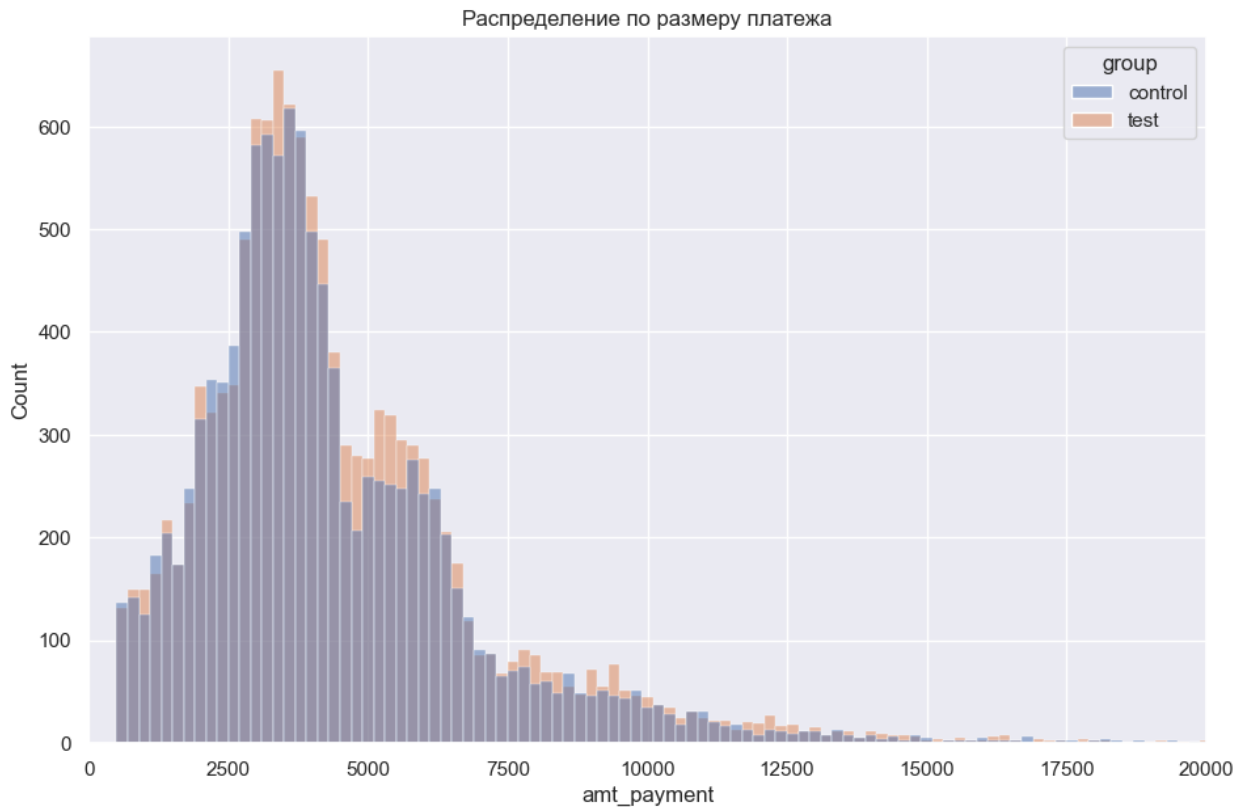
Рассмотрим распределение платежей

```

df_payment['group'] = np.where(df_payment['nflag_test']==0, 'control',
'test')

sns.set(rc={'figure.figsize':(11,7)})
sns.histplot(data = df_payment[df_payment['flag_payment']==1], x
='amt_payment',
             hue = 'group',
             binwidth=200,
             legend = True
             ).set_title("Распределение по размеру платежа")
plt.xlim(0,20000)
plt.show()

```



```
m0 = round(df_payment[(df_payment['nflag_test']==0) &
(df_payment['flag_payment']==1)][ 'amt_payment' ].mean(),2)
m1 = round(df_payment[(df_payment['nflag_test']==1) &
(df_payment['flag_payment']==1)][ 'amt_payment' ].mean(),2)

print(f'Средний чек в контрольной группе - {m0}')
print(f'Средний чек в тестовой группе - {m1}')
print()
test_calc(df_payment[(df_payment['nflag_test']==0) &
(df_payment['flag_payment']==1)][ 'amt_payment' ],
          df_payment[(df_payment['nflag_test']==1) &
(df_payment['flag_payment']==1)][ 'amt_payment' ])
print()
mann_whitney_func(df_payment[(df_payment['nflag_test']==0) &
(df_payment['flag_payment']==1)][ 'amt_payment' ],
                  df_payment[(df_payment['nflag_test']==1) &
(df_payment['flag_payment']==1)][ 'amt_payment' ])
```

Средний чек в контрольной группе - 4380.48
Средний чек в тестовой группе - 4515.08

Значение t критерия: -3.8083639880071956
Значение p value: 0.00014024278815187957
Разница между средними статистически значима.

Значение u критерия: 67569336.5
Значение p value: 2.719994837555265e-05
Разница между распределениями статистически значима.

Вывод: Изменение типа уведомления положительно сказалось на среднем чеке.

Мы наблюдаем статистически значимое увеличение чека в тестовой группе по сравнению с контрольной на 135 р.

```
conv0 = round(df_payment[df_payment['nflag_test']==0]
['flag_payment'].mean(),2)
conv1 = round(df_payment[df_payment['nflag_test']==1]
['flag_payment'].mean(),2)

print(f'Конверсия из рекламы в покупку в контрольной группе - {conv0}')
print(f'Конверсия из рекламы в покупку в тестовой группе - {conv1}')
print()
test_calc(df_payment[df_payment['nflag_test']==0]['flag_payment'],
df_payment[df_payment['nflag_test']==1]['flag_payment'])
print()
mann_whitney_func(df_payment[df_payment['nflag_test']==0]
['flag_payment'],
df_payment[df_payment['nflag_test']==1]
['flag_payment'])
```

Конверсия из рекламы в покупку в контрольной группе - 0.5
Конверсия из рекламы в покупку в тестовой группе - 0.54

Значение t критерия: -9.406024876422881
Значение p value: 5.384854454774238e-21
Разница между средними статистически значима.

Значение u критерия: 245371284.5
Значение p value: 5.6162497735928165e-21
Разница между распределениями статистически значима.

Вывод: Изменение типа уведомления положительно сказалось на конверсии из рекламы в покупку.

Мы наблюдаем статистически значимое увеличение конверсии в тестовой группе по сравнению с контрольной на 4 процентных пункта.

Сегментация результатов АБ Теста

```
def avg_payment_for_city(name_city):
    print(name_city, '- средний чек:')

    df_0 = df_payment[(df_payment['nflag_test']==0) &
(df_payment['flag_payment']==1) & (df_payment['city'] == name_city)]
    df_1 = df_payment[(df_payment['nflag_test']==1) &
(df_payment['flag_payment']==1) & (df_payment['city'] == name_city)]
```



```

m0 = round(df_0['amt_payment'].mean(),2)
m1 = round(df_1['amt_payment'].mean(),2)
print(f'Средний чек в контрольной группе - {m0}')
print(f'Средний чек в тестовой группе - {m1}')
print()

test_calc(df_0['amt_payment'], df_1['amt_payment'])
print()

mann_whitney_func(df_0['amt_payment'], df_1['amt_payment'])
print('\n')

def conv_for_city(name_city):
    print(name_city, '- конверсия:')

    df_0 = df_payment[(df_payment['nflag_test']==0) &
(df_payment['city'] == name_city)]
    df_1 = df_payment[(df_payment['nflag_test']==1) &
(df_payment['city'] == name_city)]

    conv0 = round(df_0['flag_payment'].mean(),2)
    conv1 = round(df_1['flag_payment'].mean(),2)
    print(f'Конверсия из рекламы в покупку в контрольной группе -
{conv0}')
```

```

    print(f'Конверсия из рекламы в покупку в тестовой группе -
{conv1}')
```

```

    print()

    test_calc(df_0['flag_payment'], df_1['flag_payment'])
    print()

    mann_whitney_func(df_0['flag_payment'], df_1['flag_payment'])

```

Приведем сначала анализ по городам с большим количеством торговых точек - Москве и Санкт-Петербургу.

```

avg_payment_for_city('Москва')
conv_for_city('Москва')
```

Москва - средний чек:

Средний чек в контрольной группе - 4070.46

Средний чек в тестовой группе - 4360.39

Значение t критерия: -4.626738949705741

Значение p value: 3.7880443593827997e-06

Разница между средними статистически значима.

Значение u критерия: 4629615.0

Значение p value: 7.412891199227266e-10

Разница между распределениями статистически значима.

Москва - конверсия:

Конверсия из рекламы в покупку в контрольной группе - 0.49

Конверсия из рекламы в покупку в тестовой группе - 0.55

Значение t критерия: -6.427576036124008

Значение p value: 1.3442208722377196e-10

Разница между средними статистически значима.

Значение u критерия: 17987623.0

Значение p value: 1.3888194541373183e-10

Разница между распределениями статистически значима.

avg_payment_for_city('Санкт-Петербург')

conv_for_city('Санкт-Петербург')

Санкт-Петербург - средний чек:

Средний чек в контрольной группе - 4267.97

Средний чек в тестовой группе - 4436.71

Значение t критерия: -2.2932690256918615

Значение p value: 0.021879269472817743

Разница между средними статистически значима.

Значение u критерия: 2345701.5

Значение p value: 0.08931731500200968

Разница между распределениями случайна

Санкт-Петербург - конверсия:

Конверсия из рекламы в покупку в контрольной группе - 0.45

Конверсия из рекламы в покупку в тестовой группе - 0.55

Значение t критерия: -9.352468686561785

Значение p value: 1.065825659919673e-20

Разница между средними статистически значима.

Значение u критерия: 8881151.0

Значение p value: 1.3118606823466691e-20

Разница между распределениями статистически значима.

Приведем анализ по остальным городам

```
cities = df_payment['city'].unique()
```

```
other_cities = list(cities)
```

```
other_cities.remove('Москва')
```

```
other_cities.remove('Санкт-Петербург')
```

```
for city in other_cities:  
    avg_payment_for_city(city)  
    conv_for_city(city)  
    print('\n\n')
```

Красноярск - средний чек:
Средний чек в контрольной группе - 4326.45
Средний чек в тестовой группе - 4369.94

Значение t критерия: -0.5266451581598116
Значение p value: 0.5984752086062595
Разница между средними случайна

Значение u критерия: 1372866.0
Значение p value: 0.8267219366517033
Разница между распределениями случайна

Красноярск - конверсия:
Конверсия из рекламы в покупку в контрольной группе - 0.56
Конверсия из рекламы в покупку в тестовой группе - 0.56

Значение t критерия: 0.27863631219717744
Значение p value: 0.7805336501430107
Разница между средними случайна

Значение u критерия: 4419468.0
Значение p value: 0.7805140863923713
Разница между распределениями случайна

Мурманск - средний чек:
Средний чек в контрольной группе - 4434.84
Средний чек в тестовой группе - 4268.68

Значение t критерия: 1.0199434477407143
Значение p value: 0.30802785459679893
Разница между средними случайна

Значение u критерия: 105300.5
Значение p value: 0.5266910456810876
Разница между распределениями случайна

Мурманск - конверсия:
Конверсия из рекламы в покупку в контрольной группе - 0.42
Конверсия из рекламы в покупку в тестовой группе - 0.44

Значение t критерия: -0.8460108060373052
Значение p value: 0.3976422846466432

Разница между средними случайна

Значение u критерия: 551570.0

Значение p value: 0.3975381533962302

Разница между распределениями случайна

Казань - средний чек:

Средний чек в контрольной группе - 5496.52

Средний чек в тестовой группе - 5709.99

Значение t критерия: -1.2363501124573424

Значение p value: 0.21644848079679255

Разница между средними случайна

Значение u критерия: 701750.5

Значение p value: 0.07315822711553638

Разница между распределениями случайна

Казань - конверсия:

Конверсия из рекламы в покупку в контрольной группе - 0.47

Конверсия из рекламы в покупку в тестовой группе - 0.48

Значение t критерия: -0.27304492857426443

Значение p value: 0.784829789025573

Разница между средними случайна

Значение u критерия: 3208915.5

Значение p value: 0.7848080524600568

Разница между распределениями случайна

Саратов - средний чек:

Средний чек в контрольной группе - 4424.94

Средний чек в тестовой группе - 4298.09

Значение t критерия: 0.9379110591392272

Значение p value: 0.34847409618409664

Разница между средними случайна

Значение u критерия: 193935.5

Значение p value: 0.5107809366819125

Разница между распределениями случайна

Саратов - конверсия:

Конверсия из рекламы в покупку в контрольной группе - 0.57

Конверсия из рекламы в покупку в тестовой группе - 0.59

Значение t критерия: -0.6507850002388366

Значение p value: 0.5152558091282406

Разница между средними случайна

Значение u критерия: 553867.0

Значение p value: 0.5151549688905628

Разница между распределениями случайна

Тольятти - средний чек:

Средний чек в контрольной группе - 4199.9

Средний чек в тестовой группе - 4520.15

Значение t критерия: -0.9341936875105548

Значение p value: 0.35124234209754046

Разница между средними случайна

Значение u критерия: 5733.5

Значение p value: 0.6944043312563832

Разница между распределениями случайна

Тольятти - конверсия:

Конверсия из рекламы в покупку в контрольной группе - 0.58

Конверсия из рекламы в покупку в тестовой группе - 0.64

Значение t критерия: -1.185144360467606

Значение p value: 0.2367531179978554

Разница между средними случайна

Значение u критерия: 14893.0

Значение p value: 0.2364676058329549

Разница между распределениями случайна

Тюмень - средний чек:

Средний чек в контрольной группе - 4627.87

Средний чек в тестовой группе - 3907.31

Значение t критерия: 1.5935112159523295

Значение p value: 0.11365448715957271

Разница между средними случайна

Значение u критерия: 2008.0

Значение p value: 0.48427879508500027

Разница между распределениями случайна

Тюмень - конверсия:

Конверсия из рекламы в покупку в контрольной группе - 0.38

Конверсия из рекламы в покупку в тестовой группе - 0.52

Значение t критерия: -2.4598250966325907

Значение p value: 0.01451825932213329

Разница между средними статистически значима.

Значение u критерия: 8101.0

Значение p value: 0.014823398915456424

Разница между распределениями статистически значима.

Волгоград - средний чек:

Средний чек в контрольной группе - 4365.25

Средний чек в тестовой группе - 4142.67

Значение t критерия: 1.3174050880060224

Значение p value: 0.18808971896573595

Разница между средними случайна

Значение u критерия: 80809.0

Значение p value: 0.13079545153843133

Разница между распределениями случайна

Волгоград - конверсия:

Конверсия из рекламы в покупку в контрольной группе - 0.6

Конверсия из рекламы в покупку в тестовой группе - 0.55

Значение t критерия: 1.6222173039489063

Значение p value: 0.10498894497258296

Разница между средними случайна

Значение u критерия: 240982.5

Значение p value: 0.10498253648223195

Разница между распределениями случайна

Сочи - средний чек:

Средний чек в контрольной группе - 4392.23

Средний чек в тестовой группе - 4391.01

Значение t критерия: 0.007565940377089548

Значение p value: 0.993965027319339

Разница между средними случайна

Значение u критерия: 97746.5

Значение p value: 0.7951927092544266

Разница между распределениями случайна

Сочи - конверсия:

Конверсия из рекламы в покупку в контрольной группе - 0.57

Конверсия из рекламы в покупку в тестовой группе - 0.53

Значение t критерия: 1.9849560691617736

Значение p value: 0.047319885252583245

Разница между средними статистически значима.

Значение u критерия: 336608.0

Значение p value: 0.047359239516048784

Разница между распределениями статистически значима.

Владимир - средний чек:

Средний чек в контрольной группе - 4231.14

Средний чек в тестовой группе - 4630.48

Значение t критерия: -1.955214733221153

Значение p value: 0.051036752779746286

Разница между средними случайна

Значение u критерия: 39537.5

Значение p value: 0.15609369806330298

Разница между распределениями случайна

Владимир - конверсия:

Конверсия из рекламы в покупку в контрольной группе - 0.5

Конверсия из рекламы в покупку в тестовой группе - 0.55

Значение t критерия: -1.724937016780416

Значение p value: 0.08481593543159477

Разница между средними случайна

Значение u критерия: 146729.5

Значение p value: 0.08483342425116364

Разница между распределениями случайна

Самара - средний чек:

Средний чек в контрольной группе - 3908.01

Средний чек в тестовой группе - 4550.88

Значение t критерия: -3.975552238878912

Значение p value: 7.556961289731427e-05

Разница между средними статистически значима.

Значение u критерия: 95273.5
Значение p value: 0.004506090651851397
Разница между распределениями статистически значима.

Самара - конверсия:
Конверсия из рекламы в покупку в контрольной группе - 0.44
Конверсия из рекламы в покупку в тестовой группе - 0.72

Значение t критерия: -11.846151689470245
Значение p value: 3.981449742682522e-31
Разница между средними статистически значима.

Значение u критерия: 245193.0
Значение p value: 5.58808973496e-30
Разница между распределениями статистически значима.

Краснодар - средний чек:
Средний чек в контрольной группе - 4458.68
Средний чек в тестовой группе - 4390.35

Значение t критерия: 0.5056745956321305
Значение p value: 0.6131663932420361
Разница между средними случайна

Значение u критерия: 237944.5
Значение p value: 0.5888984729062408
Разница между распределениями случайна

Краснодар - конверсия:
Конверсия из рекламы в покупку в контрольной группе - 0.65
Конверсия из рекламы в покупку в тестовой группе - 0.61

Значение t критерия: 1.8789539353324407
Значение p value: 0.060385275678499745
Разница между средними случайна

Значение u критерия: 605880.5
Значение p value: 0.060406618664924146
Разница между распределениями случайна

Сахалинск - средний чек:
Средний чек в контрольной группе - 4255.64
Средний чек в тестовой группе - 5491.19

Значение t критерия: -1.2424507814524557
Значение p value: 0.225598797984653
Разница между средними случайна

Значение u критерия: 65.0
Значение p value: 0.266872579502037
Разница между распределениями случайна

Сахалинск - конверсия:
Конверсия из рекламы в покупку в контрольной группе - 0.14
Конверсия из рекламы в покупку в тестовой группе - 0.21

Значение t критерия: -1.1704284638604743
Значение p value: 0.24367247406936435
Разница между средними случайна

Значение u критерия: 2713.5
Значение p value: 0.24351609530716278
Разница между распределениями случайна

Дмитров - средний чек:
Средний чек в контрольной группе - 4903.57
Средний чек в тестовой группе - 5020.89

Значение t критерия: -0.08464422217079061
Значение p value: 0.9337427496441195
Разница между средними случайна

Значение u критерия: 31.0
Значение p value: 1.0
Разница между распределениями случайна

Дмитров - конверсия:

Конверсия из рекламы в покупку в контрольной группе - 0.1
Конверсия из рекламы в покупку в тестовой группе - 0.12

Значение t критерия: -0.3285782051097397
Значение p value: 0.7429421277261776
Разница между средними случайна

Значение u критерия: 2722.5
Значение p value: 0.7444023199576546
Разница между распределениями случайна

Статистически значимое изменение и по конверсии, и по среднему чеку наблюдается по крупным городам - Москве и Санкт-Петербургу.

В Тюмени конверсия увеличилась, при этом средний чек не изменился.

В Сочи изменения сказались отрицательно, средний чек не изменился, а конверсия снизилась.

Самара показала увеличение конверсии почти в 2 раза при росте среднего чека.

По остальным городам показатели не изменились.

Отчет по АБ Тесту

```
df_res = pd.DataFrame(columns = ['city', 'id_trading_point',  
'count_test', 'count_control', 'count_all', 'percent_count',  
                                'avg_payment_test',  
'avg_payment_control', 'diff', 'sigma_test', 'sigma_control',  
                                'ttest', 'pvalue_ttest', 'flag'])  
  
for city in cities:  
    points = df_payment[df_payment['city'] == city]  
    ['id_trading_point'].unique()  
    for point in points:  
        df_point = df_payment[df_payment['id_trading_point'] == point]  
        count_test = df_point[df_point['nflag_test']==1]  
        ['id_client'].count()  
        count_control = df_point[df_point['nflag_test']==0]  
        ['id_client'].count()  
        count_all = count_test + count_control  
        percent_count = count_all/df_payment['id_client'].count()  
  
        avg_payment_test = df_point[(df_point['nflag_test']==1) &  
(df_point['flag_payment']==1)][ 'amt_payment'].mean()  
        avg_payment_control = df_point[(df_point['nflag_test']==0) &  
(df_point['flag_payment']==1)][ 'amt_payment'].mean()  
        diff = avg_payment_test - avg_payment_control  
  
        sigma_test = df_point[(df_point['nflag_test']==1) &  
(df_point['flag_payment']==1)][ 'amt_payment'].std()  
        sigma_control = df_point[(df_point['nflag_test']==0) &  
(df_point['flag_payment']==1)][ 'amt_payment'].std()  
  
        ttest, pvalue_ttest =  
ttest_ind(df_point[(df_point['nflag_test']==0) &  
(df_point['flag_payment']==1)][ 'amt_payment'],  
df_point[(df_point['nflag_test']==1) & (df_point['flag_payment']==1)]  
[ 'amt_payment'])  
  
        if diff>0 and pvalue_ttest<0.05:  
            flag = 1  
        elif diff<0 and pvalue_ttest<0.05:
```

```

        flag = -1
    else:
        flag = 0

    df_res.loc[len(df_res.index)] = [city, point, count_test,
                                     count_control, count_all, percent_count,
                                     avg_payment_test,
                                     avg_payment_control, diff, sigma_test, sigma_control,
                                     ttest, pvalue_ttest, flag]

df_res.head()

```

	city	id_trading_point	count_test	count_control	count_all
0	Красноярск	212	593	575	1168
1	Красноярск	278	535	489	1024
2	Красноярск	202	551	565	1116
3	Красноярск	444	650	562	1212
4	Красноярск	277	708	709	1417

	percent_count	avg_payment_test	avg_payment_control	diff
0	0.025773	4101.870968	4360.682594	-258.811626
1	0.022596	4517.398649	4141.744186	375.654463
2	0.024626	4634.956897	4203.444142	431.512755
3	0.026744	4316.765193	4402.163077	-85.397884
4	0.031268	4281.246032	4477.501299	-196.255267

	sigma_test	sigma_control	ttest	pvalue_ttest	flag
0	2449.485560	2426.988322	1.302575	0.193219	0
1	2498.933998	2110.603391	-1.896006	0.058481	0
2	2432.827891	2311.357999	-2.432123	0.015256	1
3	2434.025976	2312.912183	0.470048	0.638470	0
4	2287.790254	2446.713156	1.143964	0.252998	0

Выгрузим полученные результаты в Excel.

```

writer = pd.ExcelWriter('Results.xlsx')
df_res[df_res['flag']==1].to_excel(writer, sheet_name = 'Положительные
исходы')
df_res[df_res['flag']==-1].to_excel(writer, sheet_name =
'Отрицательные исходы')
df_res[df_res['flag']==0].to_excel(writer, sheet_name = 'Нейтральные
исходы')
writer.save()

```

```
C:\Users\Daria\AppData\Local\Temp\ipykernel_3420\1243958422.py:5:  
FutureWarning: save is not part of the public API, usage can give  
unexpected results and will be removed in a future version  
writer.save()
```