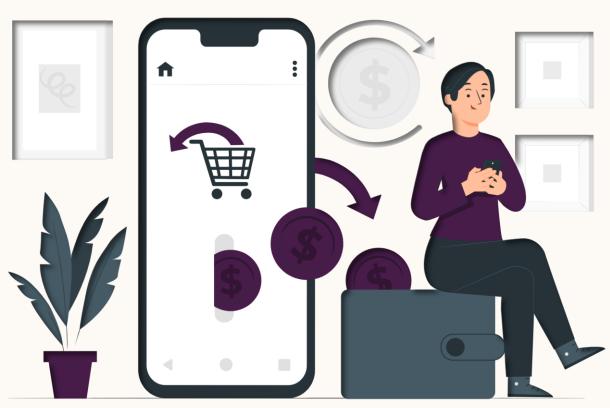
# TMHBKOODO KSMISSK

Хлебникова Мария
 Прохорова Нелли
 Непорада Анастасия
 Заиц Рената





Данные приведены к удобному виду

Проверены типы данных и пропуски Исследовано наличие дубликатов



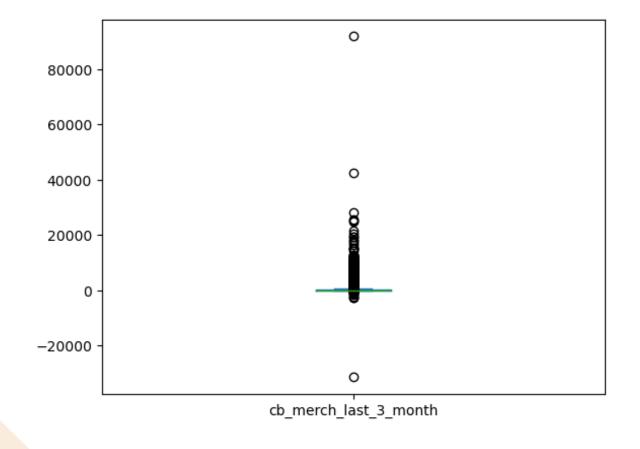


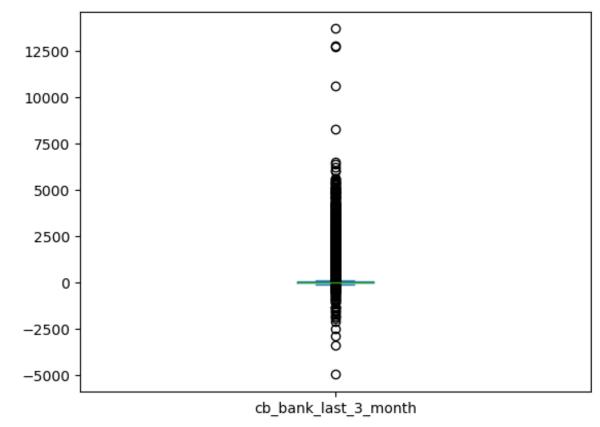




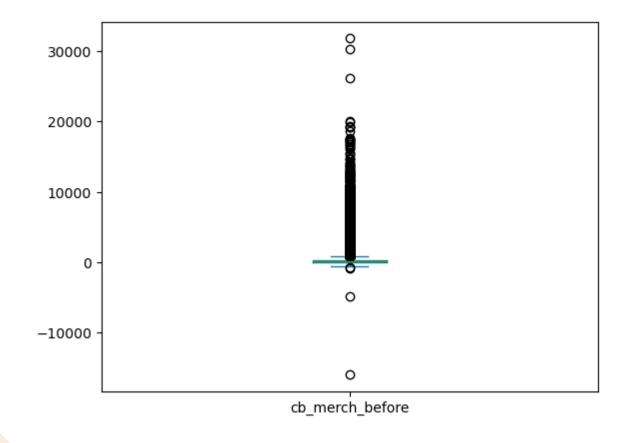


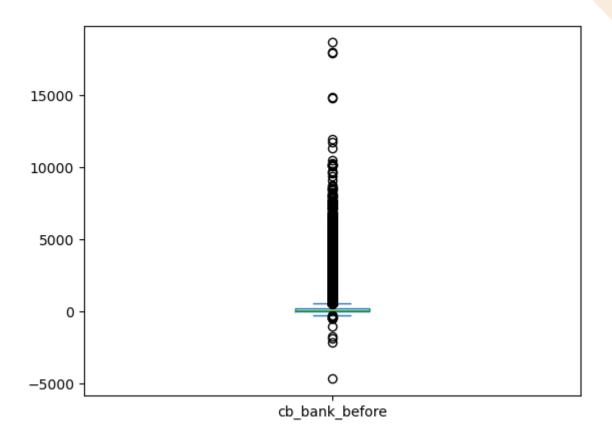




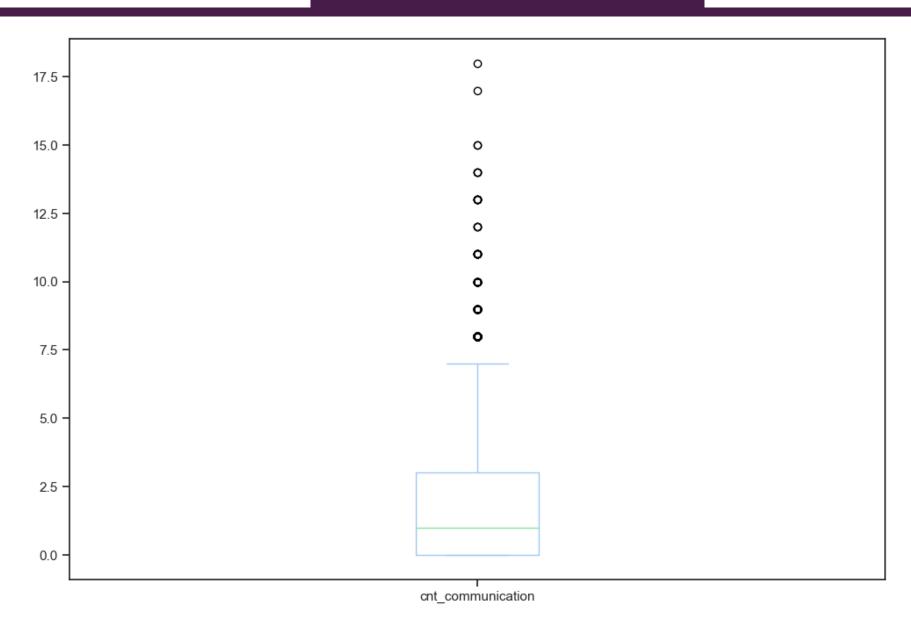






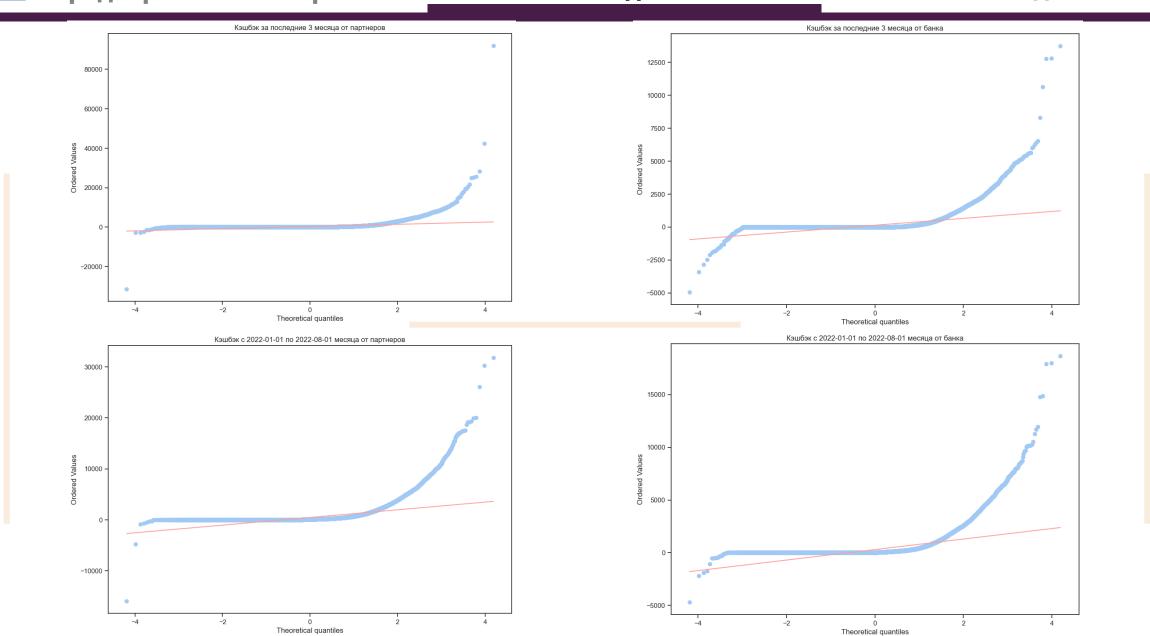


### **Анализ данных**<sup>1</sup>





### **Анализ данных**<sup>1</sup>





### КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ КЕНДАЛЛА

# СОЗДАЛИ НОВЫЕ СТОЛБЦЫ ПО ВРЕМЕННОМУ ПРИЗНАКУ И ПО СУБЪЕКТУ ПОЛУЧЕНИЯ КЭШБЭКА И ПОСТРОИЛИ МАТРИЦУ КОРРЕЛЯЦИЙ

```
df_copy = df.copy()
#Coздадим два столбца: первые по временному признаку

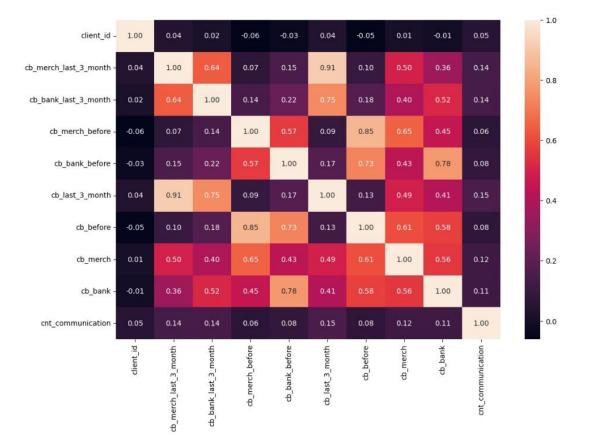
df_copy['cb_last_3_month'] = df['cb_merch_last_3_month'] + df['cb_bank_last_3_month'] #Кэшбэк за последние три месяца

df_copy['cb_before'] = df['cb_merch_before'] + df['cb_bank_before'] #Кэшбэк за первые месяцы

#Третий и четвертый по субъекту получения кэшбэка

df_copy['cb_merch'] = df['cb_merch_last_3_month'] + df['cb_merch_before'] #Кэшбэк от партнеров за год

df_copy['cb_bank'] = df['cb_bank_last_3_month'] + df['cb_bank_before'] #Кэшбэк от банка за год
```



## **Анализ дан**ных<sup>2</sup>

#### Выводы

- 1.0

0.8

- 0.6

- 0.4

0.2

# СОЗДАЛИ ДАТАСЕТ С ПОКУПКАМИ ЗА ПОСЛЕДНИЕ ТРИ МЕСЯЦА И ПОСТРОИЛИ МАТРИЦУ КОРРЕЛЯЦИЙ

```
cb last 3 month - 1.00 0.91
                                      0.07 0.11 0.09 0.10 0.08 0.18 0.35 0.19 0.13 0.14 0.08 0.24 0.08 0.19 0.24 0.15
cb merch last 3 month - 0.91 1.00 0.64 0.07 0.10 0.08 0.09 0.06 0.17 0.35 0.18 0.12 0.13 0.08 0.23 0.08 0.18 0.22 0.14
cb bank last 3 month - 0.75 0.64 1.00 0.07 0.09 0.07 0.09 0.08 0.15 0.38 0.17 0.12 0.12 0.06 0.22 0.07 0.17 0.22 0.14
         beauty shop - 0.11 0.10 0.09 0.13 1.00 0.21 0.17 0.12 0.35 0.08 0.28 0.05 0.10 0.19 0.25 0.18 0.28 0.26 0.25
           book store - 0.09 0.08 0.07 0.09 0.21 1.00 0.17 0.12 0.22 0.08 0.23 0.04 0.09 0.13 0.19 0.12 0.19 0.20 0.20
             children - 0.10 0.09 0.09 0.06 0.17 0.17 1.00 0.10 0.27 0.09 0.21 0.20 0.14 0.13 0.23 0.11 0.25 0.28 0.10
              cinema - 0.08 0.06 0.08 0.09 0.12 0.12 0.10 1.00 0.15 0.05 0.19 0.09 0.08 0.07 0.16 0.07 0.12 0.15 0.16
              clothes - 0.18 0.17 0.15 0.16 0.35 0.22 0.27 0.15 1.00 0.14 0.37 0.19 0.19 0.21 0.38 0.18 0.36 0.39 0.28
           electronics - 0.35 0.35 0.38 0.08 0.08 0.08 0.09 0.05 0.14 1.00 0.18 0.15 0.15 0.05 0.23 0.09 0.16 0.21 0.15
             fastfood - 0.19 0.18 0.17 0.21 0.28 0.23 0.21 0.19 0.37 0.18 1.00 0.26 0.24 0.16 0.48 0.20 0.36 0.49 0.48
                  fuel - 0.13 0.12 0.12 0.07 0.05 0.04 0.20 0.09 0.19 0.15 0.26 1.00 0.24 0.08 0.32 0.13 0.25 0.35 0.08
               house - 0.14 0.13 0.12 0.07 0.10 0.09 0.14 0.08 0.19 0.15 0.24 0.24 1.00 0.10 0.26 0.12 0.21 0.29 0.15
              jewelry - 0.08 0.08 0.06 0.09 0.19 0.13 0.13 0.07 0.21 0.05 0.16 0.08 0.10 1.00 0.16 0.09 0.16 0.17 0.13
                other - 0.24 0.23 0.22 0.24 0.25 0.19 0.23 0.16 0.38 0.23 0.48 0.32 0.26 0.16 1.00 0.18 0.38 0.48 0.37
```

## **Анализ данных**<sup>2</sup>

#### Выводы

### СОЗДАЛИ ДАТАСЕТ С ПОКУПКАМИ ЗА ПОСЛЕДНИЕ ТРИ МЕСЯЦА И ПОСТРОИЛИ МАТРИЦУ КОРРЕЛЯЦИЙ

```
#Отдельно посмотрим сумму всех покупок за последние 3 месяца

df_3_all = df_all[['cb_last_3_month', 'cb_merch_last_3_month', 'cb_bank_last_3_month']]

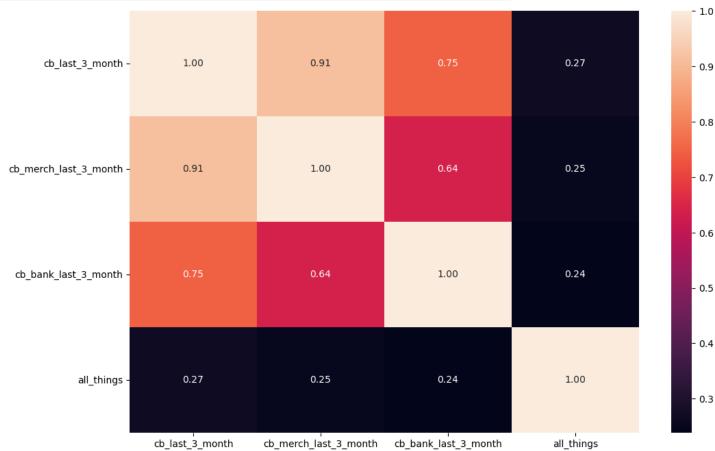
df_3_all['all_things'] = df_3['air'] + df_3['beauty shop'] + df_3['book store'] + df_3['children']

+ df_3['cinema'] + df_3['clothes'] + df_3['electronics'] + df_3['fastfood'] + df_3['fuel'] + df_3['house']

+ df_3['jewelry'] + df_3['other'] + df_3['pet shop'] + df_3['pharmacies'] + df_3['supermarkets'] + df_3['transport']

sns.heatmap(df_3_all.corr(method='kendall'), annot=True, fmt=".2f", linecolor='white', cmap='BuPu')

plt.show()
```





### **Анализ данных**<sup>2</sup>

#### Выводы

- 1.0

0.8

0.6

0.2

# СОЗДАЛИ ДАТАСЕТ С ПОКУПКАМИ ЗА ПЕРВЫЕ МЕСЯЦЫ И ПОСТРОИЛИ МАТРИЦУ КОРРЕЛЯЦИЙ

```
cb_before - 1.00 0.85 0.
                                 0.12 0.12 0.11 0.11 0.14 0.20 0.19 0.23 0.13 0.16 0.10 0.29 0.10 0.20 0.24 0.18
  cb merch before - 0.85 1.00 0.57 0.11 0.11 0.11 0.10 0.11 0.19 0.18 0.21 0.12 0.15 0.09 0.28 0.09 0.19 0.22 0.17
   cb bank before - 0.73 0.57 1.00 0.10 0.10 0.09 0.10 0.13 0.18 0.17 0.21 0.13 0.16 0.09 0.26 0.10 0.20 0.25 0.17
        air_before - 0.12 0.11 0.10 1.00 0.17 0.14 0.09 0.16 0.21 0.11 0.25 0.09 0.12 0.12 0.28 0.06 0.17 0.18 0.25
beauty shop before - 0.12 0.11 0.10 0.17 1.00 0.24 0.19 0.16 0.38 0.10 0.30 0.05 0.13 0.24 0.27 0.19 0.30 0.27 0.26
 book store_before - 0.11 0.11 0.09 0.14 0.24 1.00 0.16 0.17 0.24 0.10 0.26 0.03 0.12 0.15 0.22 0.14 0.21 0.20 0.23
   children_before - 0.11 0.10 0.10 0.09 0.19 0.16 1.00 0.14 0.27 0.11 0.21 0.20 0.16 0.14 0.24 0.11 0.26 0.29 0.10
    cinema before - 0.14 0.11 0.13 0.16 0.16 0.17 0.14 1.00 0.20 0.13 0.29 0.12 0.14 0.13 0.24 0.10 0.17 0.21 0.23
    clothes before - 0.20 0.19 0.18 0.21 0.38 0.24 0.27 0.20 1.00 0.16 0.39 0.16 0.23 0.25 0.41 0.19 0.37 0.40 0.31
 electronics before - 0.19 0.18 0.17 0.11 0.10 0.10 0.11 0.13 0.16 1.00 0.22 0.16 0.22 0.09 0.25 0.11 0.18 0.23 0.18
   fastfood before - 0.23 0.21 0.21 0.25 0.30 0.26 0.21 0.29 0.39 0.22 1.00 0.23 0.26 0.20 0.47 0.20 0.34 0.46 0.50
       fuel before - 0.13 0.12 0.13 0.09 0.05 0.03 0.20 0.12 0.16 0.16 0.23 1.00 0.24 0.09 0.32 0.14 0.23 0.34 0.06
      other before - 0.29 0.28 0.26 0.28 0.27 0.22 0.24 0.24 0.41 0.25 0.47 0.32 0.33 0.20
  pet shop before - 0.10 0.09 0.10 0.06 0.19 0.14 0.11 0.10 0.19 0.11 0.20 0.14 0.14 0.12 0.19 1.00 0.23 0.25 0.15
pharmacies before - 0.20 0.19 0.20 0.17 0.30 0.21 0.26 0.17 0.37 0.18 0.34 0.23 0.24 0.20 0.40 0.23 1.00 0.46 0.26
```



## **Анализ данных**<sup>2</sup>

#### Выводы

# СОЗДАЛИ ДАТАСЕТ С ПОКУПКАМИ ЗА ПЕРВЫЕ МЕСЯЦЫ И ПОСТРОИЛИ МАТРИЦУ КОРРЕЛЯЦИЙ

```
#Οπ∂ε∩ЬΗΟ ΠΟCYUMTAEM CYMMY βCEX ΠΟΚΥΠΟΚ 3α ΠΕΡβЫЙ ΠΕΡΙΟΟ

df_before_all = df_all[['cb_before', 'cb_merch_before', 'cb_bank_before']]

df_before_all['all_things'] = df_before['air_before'] + df_before['beauty shop_before'] + df_before['book store_before']

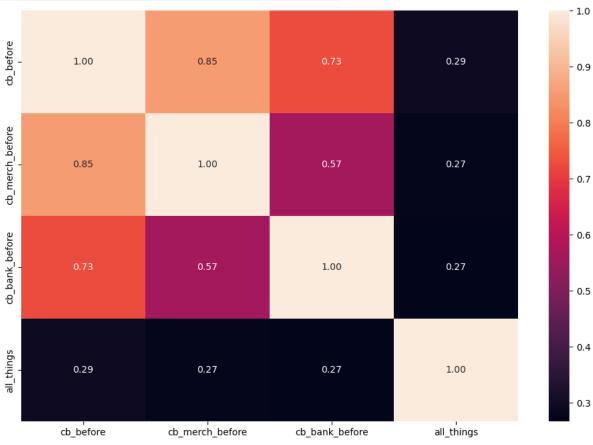
+ df_before['children_before'] + df_before['cinema_before'] + df_before['clothes_before'] + df_before['electronics_before']

+ df_before['fastfood_before'] + df_before['fuel_before'] + df_before['house_before'] + df_before['jewelry_before']

+ df_before['other_before'] + df_before['pet_shop_before'] + df_before['pharmacies_before'] + df_before['supermarkets_before']

sns.heatmap(df_before_all.corr(method='kendall'), annot=True, fmt=".2f", linecolor='white', cmap='BuPu')

plt.show()
```





# БАКЕТНЫЙ ТЕСТ

Разделим пользователей по бакетам:

- 50000 измерений
- 100 бакетов
- по 500 пользователей в каждом бакете



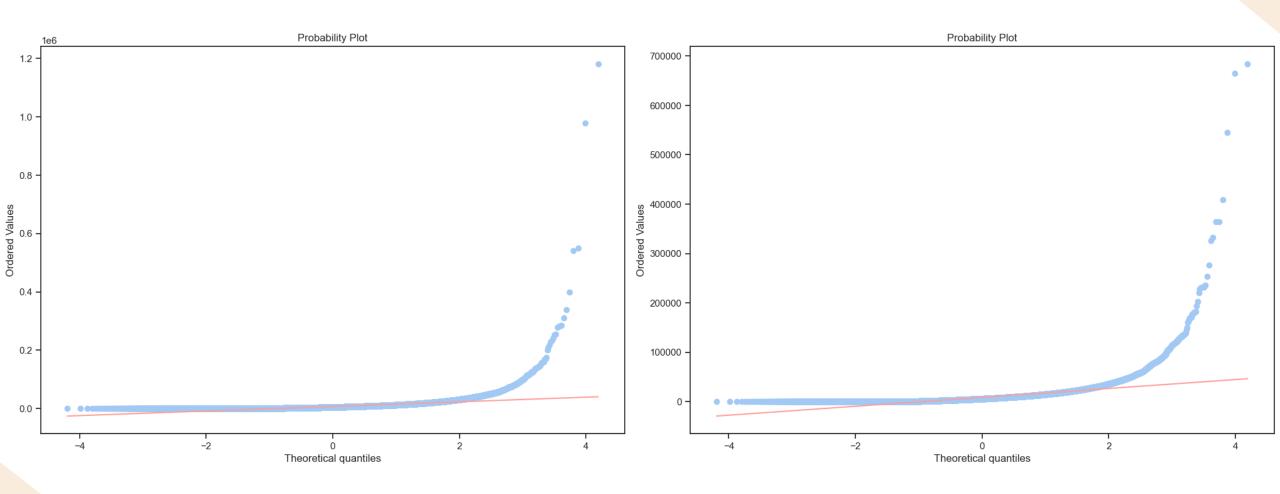
Н0: различий между выборками нет, обновление не играет роли

Н1: различия между выборками есть, обновление повлияло на

изменение суммы заказа



## **Анализ данных**<sup>3</sup>





# **НЕНОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ**

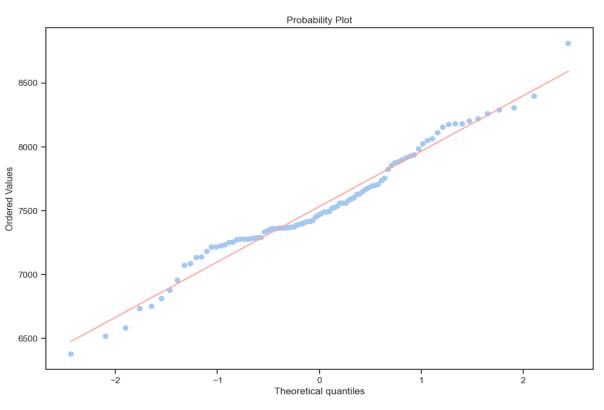
### ТЕСТ МАННА-УИТНИ

### ОТВЕРГАЕМ ГИПОТЕЗУ О НЕВАЖНОСТИ РОЛИ ОБНОВЛЕНИЯ

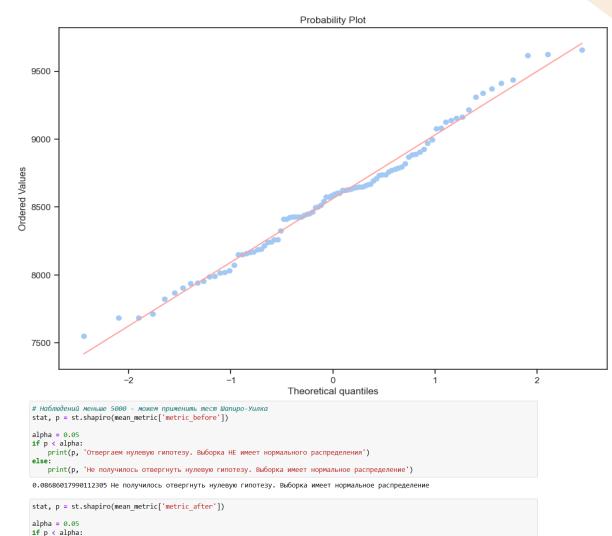
### **Анализ данных**<sup>3</sup>

#### Выводы

### С ПОМОЩЬЮ УДАЛЕНИЯ ВЫБРОСОВ ПРИВЕЛИ РАСПРЕДЕЛЕНИЕ К НОРМАЛЬНОМУ









**НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ** 

t-TECT

ОТВЕРГАЕМ ГИПОТЕЗУ О НЕВАЖНОСТИ РОЛИ ОБНОВЛЕНИЯ



# БАКЕТНЫЙ ТЕСТ

Предварительная обработка

Разделим пользователей по бакетам:

- 50000 измерений
- 100 бакетов
- по 500 пользователей в каждом бакете

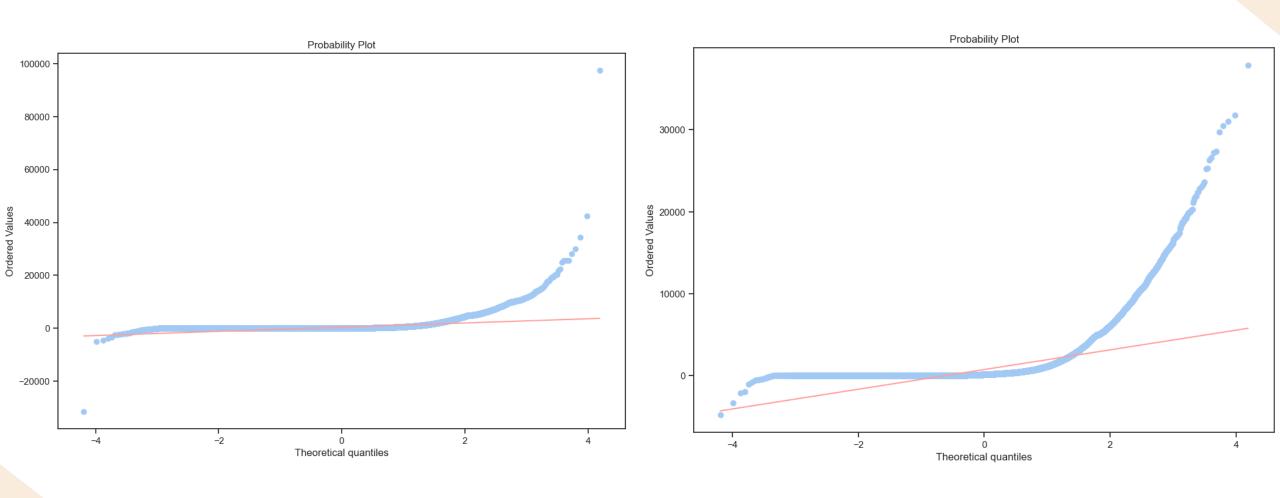


Н0: обновление не повлияло на общий кэшбэк

Н1: обновление повлияло на общий кэшбэк



## **Анализ данных**<sup>4</sup>





**НЕНОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ** 

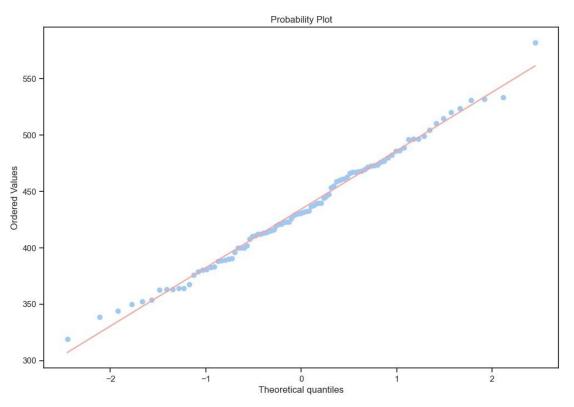
Предварительная обработка

ТЕСТ МАННА-УИТНИ ОТВЕРГАЕМ ГИПОТЕЗУ О НЕВАЖНОСТИ РОЛИ ОБНОВЛЕНИЯ

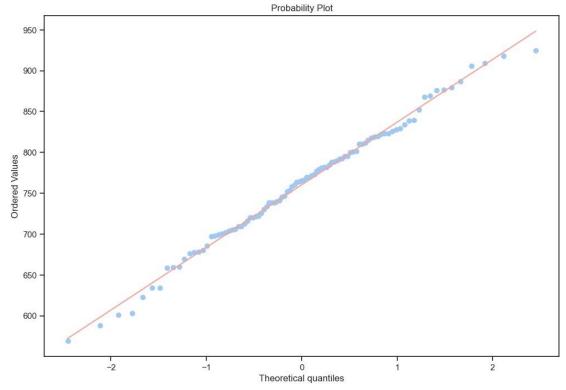
p-value: 0.0

### **Анализ данных**<sup>4</sup>

### С ПОМОЩЬЮ УДАЛЕНИЯ ВЫБРОСОВ ПРИВЕЛИ РАСПРЕДЕЛЕНИЕ К НОРМАЛЬНОМУ







```
# Наблюдений меньше 5000 - можем применить тест Шапиро-Уилка
stat, p = st.shapiro(mean_metric['cb_last_3_month'])

alpha = 0.05
if p < alpha:
    print(p, 'Отвергаем нулевую гипотезу. Выборка НЕ имеет нормального распределения')

else:
    print(p, 'Не получилось отвергнуть нулевую гипотезу. Выборка имеет нормальное распределение')

0.8287481069564819 Не получилось отвергнуть нулевую гипотезу. Выборка имеет нормальное распределение

stat, p = st.shapiro(mean_metric['cb_before'])

alpha = 0.05
if p < alpha:
    print(p, 'Отвергаем нулевую гипотезу. Выборка НЕ имеет нормального распределения')

else:
    print(p, 'Не получилось отвергнуть нулевую гипотезу. Выборка имеет нормальное распределение')

0.824847400188446 Не получилось отвергнуть нулевую гипотезу. Выборка имеет нормальное распределение
```



НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

t-TECT

ОТВЕРГАЕМ ГИПОТЕЗУ О НЕВАЖНОСТИ РОЛИ ОБНОВЛЕНИЯ



- СВЯЗЬ МЕЖДУ ВСЕМИ ПОКАЗАТЕЛЯМИ И КОЛИЧЕСТВОМ КОММУНИКАЦИЙ СЛАБАЯ И НЕЛИНЕЙНАЯ
- 2 СВЯЗЬ МЕЖДУ ОБОРОТОМ КЛИЕНТОВ И ВЫПЛАЧЕННЫМ КЭШБЭКОМ СЛАБАЯ И НЕЛИНЕЙНАЯ
- **3** ОБНОВЛЕНИЕ ПОВЛИЯЛО НА СУММУ ПОКУПОК
- ОБНОВЛЕНИЕ ПОВЛИЯЛО НА СУММУ КЭШБЭКА

