**1. Основная информация о данных: количество наблюдений, количество переменных, типы данных, количество пропущенных значений**

```
Количество пропущенных значений:
age          0
job          0
marital      0
education    0
default      0
balance      0
housing      0
loan         0
contact      0
day          0
month        0
duration     0
campaign     0
pdays        0
previous     0
poutcome     0
y            0
```

```
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   age        4521 non-null   int64
 1   job        4521 non-null   object
 2   marital    4521 non-null   object
 3   education  4521 non-null   object
 4   default    4521 non-null   object
 5   balance    4521 non-null   int64
 6   housing    4521 non-null   object
 7   loan       4521 non-null   object
 8   contact    4521 non-null   object
 9   day        4521 non-null   int64
 10  month      4521 non-null   object
 11  duration   4521 non-null   int64
 12  campaign   4521 non-null   int64
 13  pdays      4521 non-null   int64
 14  previous   4521 non-null   int64
 15  poutcome   4521 non-null   object
 16  y          4521 non-null   int64
```

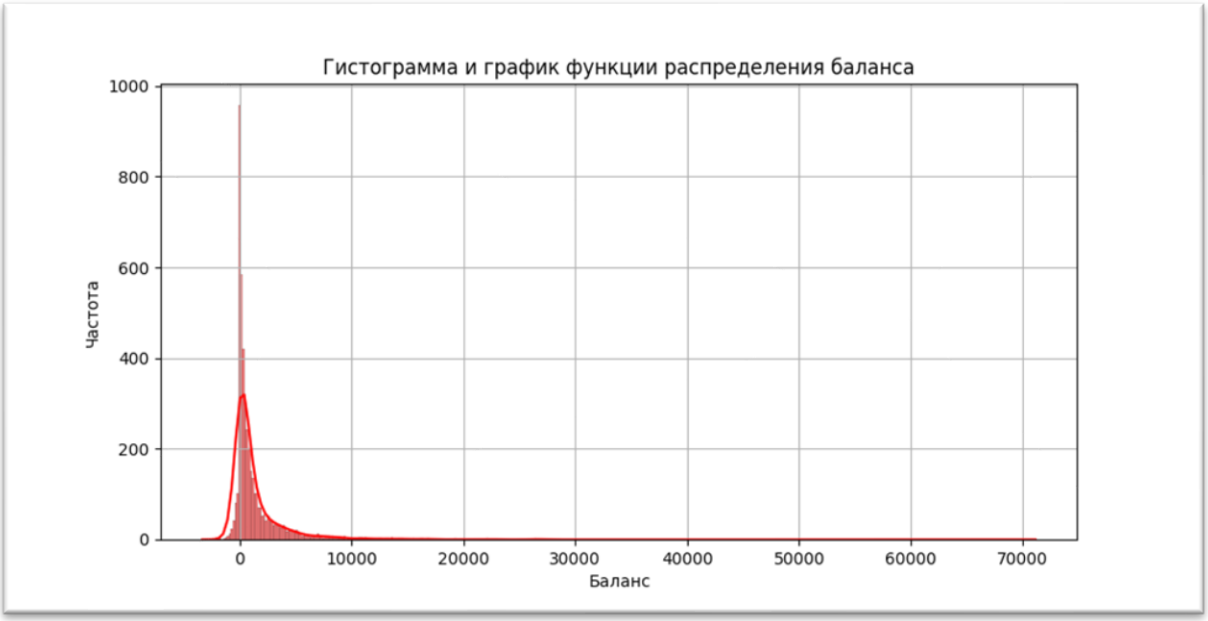
**2. Вычислить основные статистические характеристики данных: средние значения, дисперсии, корреляции, минимумы, максимумы, квартили.**

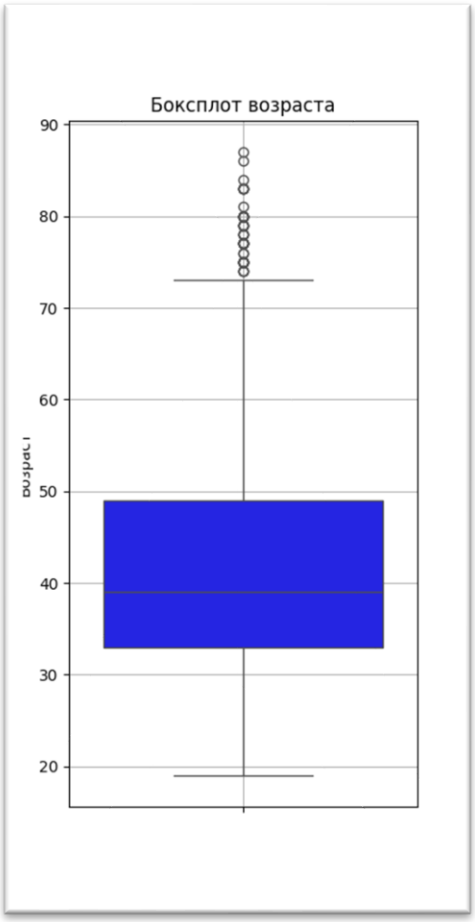
```
Описательные статистики для числовых переменных:
                age        balance
count  4521.000000    4521.000000
mean     41.170095    1422.657819
std      10.576211    3009.638142
min      19.000000   -3313.000000
25%      33.000000      69.000000
50%      39.000000     444.000000
75%      49.000000    1480.000000
max      87.000000   71188.000000

Корреляции между числовыми переменными:
             age  balance
age      1.00000  0.08382
balance  0.08382  1.00000
```

**3. Построить гистограммы с графиком функции распределения, боксплоты и диаграммы рассеивания.**


Гистограмма и график функции распределения возраста
Частота
Возраст
0
50
100
150
200
250
300
350
400
20
30
40
50
60
70
80
90


Гистограмма и график функции распределения баланса
Частота
Баланс
0
200
400
600
800
1000
0
10000
20000
30000
40000
50000
60000
70000


Боксплот возраста
90
80
70
60
50
40
30
20

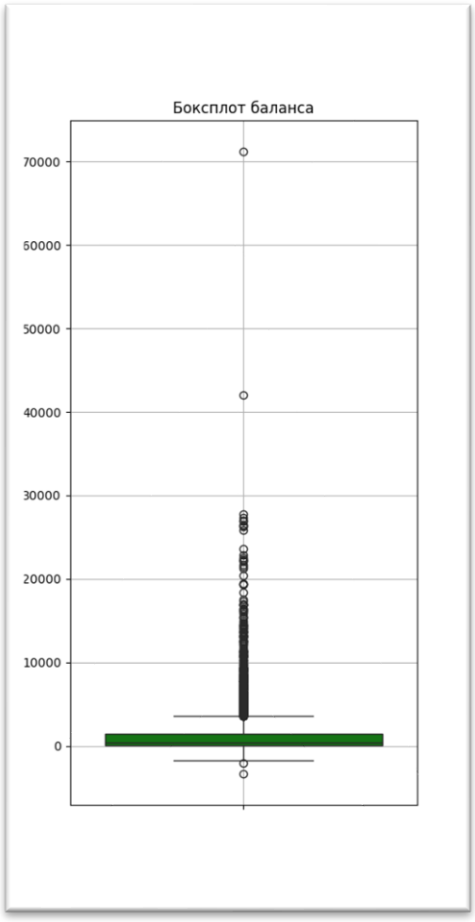

Боксплот баланса
70000
50000
50000
40000
30000
20000
10000
0


Диаграмма рассеивания возраста и баланса
Баланс
70000
60000
50000
40000
30000
20000
10000
0
20
30
40
50
60
70
80
90
Возраст

## 4.Найти выбросы с помощью межквартильного размаха

Выбросы для переменной age:

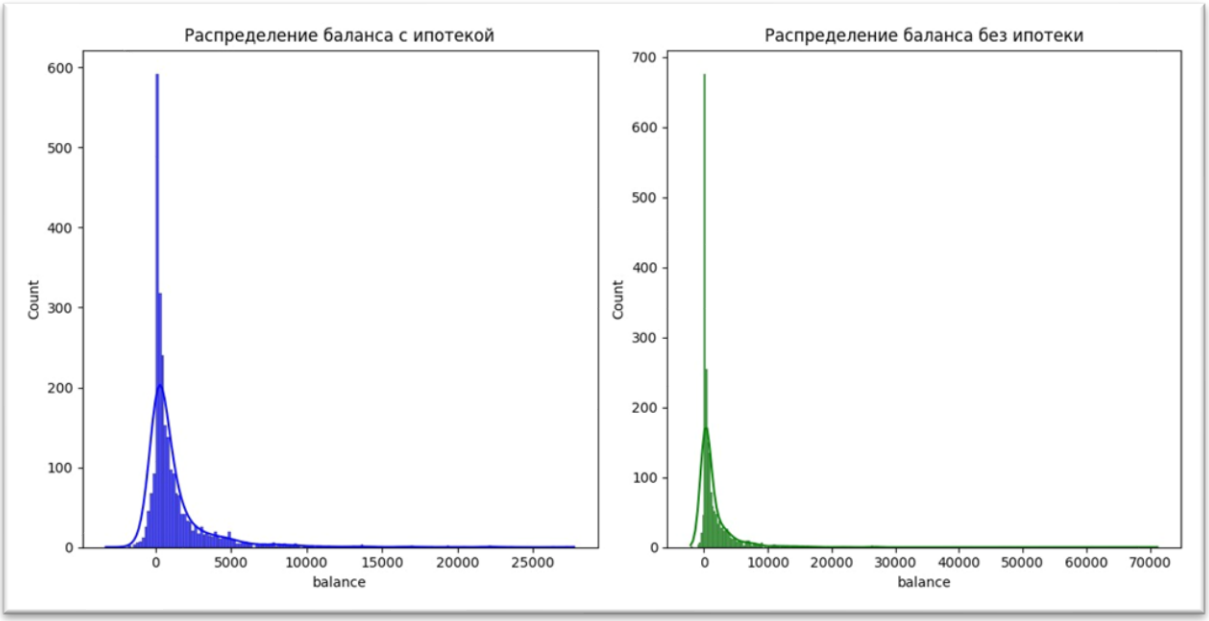
| | age | job | marital | education | ... | pdays | previous | poutcome | y |
|---|---|---|---|---|---|---|---|---|---|
| 36 | 78 | retired | divorced | primary | ... | -1 | 0 | unknown | 1 |
| 129 | 77 | retired | divorced | tertiary | ... | -1 | 0 | unknown | 1 |
| 166 | 78 | housemaid | married | secondary | ... | -1 | 0 | unknown | 0 |
| 199 | 75 | retired | married | secondary | ... | 181 | 2 | success | 1 |
| 412 | 75 | retired | divorced | tertiary | ... | 183 | 1 | failure | 1 |
| 477 | 77 | retired | married | tertiary | ... | 92 | 2 | success | 1 |
| 573 | 81 | retired | married | secondary | ... | -1 | 0 | unknown | 0 |
| 633 | 83 | retired | married | secondary | ... | -1 | 0 | unknown | 0 |
| 688 | 80 | management | married | primary | ... | -1 | 0 | unknown | 0 |
| 1126 | 77 | retired | married | secondary | ... | -1 | 0 | unknown | 0 |
| 1230 | 75 | retired | divorced | secondary | ... | -1 | 0 | unknown | 1 |
| 1312 | 80 | retired | married | secondary | ... | 64 | 12 | failure | 1 |
| 1349 | 83 | retired | married | primary | ... | 140 | 1 | failure | 1 |
| 1415 | 75 | retired | divorced | primary | ... | -1 | 0 | unknown | 0 |
| 1422 | 79 | retired | married | secondary | ... | -1 | 0 | unknown | 0 |
| 1437 | 77 | retired | married | tertiary | ... | 60 | 1 | success | 1 |
| 1866 | 86 | retired | married | secondary | ... | 101 | 1 | other | 0 |
| 1949 | 78 | retired | married | tertiary | ... | -1 | 0 | unknown | 0 |
| 1956 | 77 | retired | married | primary | ... | 89 | 7 | failure | 0 |
| 2014 | 74 | retired | divorced | secondary | ... | -1 | 0 | unknown | 1 |
| 2070 | 80 | retired | married | secondary | ... | 118 | 11 | success | 0 |
| 2678 | 74 | retired | married | secondary | ... | 104 | 1 | other | 0 |
| 2848 | 80 | retired | married | secondary | ... | -1 | 0 | unknown | 0 |
| 2896 | 80 | retired | married | primary | ... | -1 | 0 | unknown | 0 |
| 3157 | 75 | blue-collar | married | secondary | ... | 190 | 1 | failure | 0 |
| 3193 | 76 | retired | divorced | primary | ... | -1 | 0 | unknown | 0 |
| 3202 | 79 | retired | married | primary | ... | 272 | 2 | success | 1 |
| 3205 | 77 | retired | married | primary | ... | 94 | 3 | failure | 0 |
| 3311 | 87 | retired | married | primary | ... | -1 | 0 | unknown | 1 |
| 3360 | 79 | retired | married | primary | ... | -1 | 0 | unknown | 1 |
| 3495 | 76 | retired | married | primary | ... | -1 | 0 | unknown | 0 |
| 3690 | 80 | housemaid | married | primary | ... | 189 | 1 | failure | 1 |
| 3750 | 79 | retired | divorced | unknown | ... | 450 | 2 | failure | 0 |
| 3786 | 74 | retired | married | secondary | ... | -1 | 0 | unknown | 1 |
| 4047 | 75 | retired | married | secondary | ... | -1 | 0 | unknown | 0 |
| 4108 | 84 | retired | divorced | primary | ... | -1 | 0 | unknown | 1 |
| 4323 | 83 | retired | divorced | primary | ... | 77 | 3 | success | 0 |

Выбросы для переменной balance:

| | age | job | marital | education | ... | pdays | previous | poutcome | y |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 33 | services | married | secondary | ... | 339 | 4 | failure | 0 |
| 10 | 39 | services | married | secondary | ... | -1 | 0 | unknown | 0 |
| 16 | 56 | technician | married | secondary | ... | -1 | 0 | unknown | 0 |
| 25 | 41 | management | married | tertiary | ... | -1 | 0 | unknown | 0 |
| 30 | 68 | retired | divorced | secondary | ... | -1 | 0 | unknown | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | .. |
| 4464 | 53 | services | divorced | secondary | ... | -1 | 0 | unknown | 0 |
| 4473 | 33 | technician | married | secondary | ... | 272 | 2 | failure | 0 |
| 4489 | 45 | management | married | tertiary | ... | 356 | 3 | failure | 0 |
| 4500 | 38 | admin. | married | secondary | ... | -1 | 0 | unknown | 0 |
| 4517 | 57 | self-employed | married | tertiary | ... | -1 | 0 | unknown | 0 |

**5.На основе выбранных данных провести алгоритм проверки гипотезы о параметрах генеральной совокупности и виде распределения.**

Тест Манна-Уитни:




Гипотезы:

- H0 (нулевая гипотеза): Средние значения баланса у клиентов с ипотекой и без ипотеки одинаковы.
- H1 (альтернативная гипотеза): Средние значения баланса у клиентов с ипотекой и без ипотеки различаются

Поскольку p-значение 0.037307663790934094 меньше 0.05, мы отвергаем нулевую гипотезу. Следовательно, существует статистически значимая разница в балансе между клиентами с ипотекой и без ипотеки.

```
Тест Шапиро-Уилка для возраста:
p-значение: 9.427573604575e-34
Отвергаем нулевую гипотезу. Распределение возраста не является нормальным.

Тест Шапиро-Уилка для баланса:
p-значение: 1.124061228956861e-77
Отвергаем нулевую гипотезу. Распределение баланса не является нормальным.
```

**6. Выбрать два признака и построить линейную регрессию.**