

```
In [2]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sbs
import plotly
import plotly.express as px
plt.style.use('default')
```

```
In [ ]: # Загрузка данных

churn_data = pd.read_csv('data/churn.csv')
churn_data.head()
```

```
Out[ ]:   RowNumber  CustomerId  Surname  CreditScore  Geography  Gender  Age  Tenure

0         1    15634602   Hargrave         619      France  Female   42         2
1         2    15647311     Hill         608      Spain  Female   41         1
2         3    15619304     Onio         502      France  Female   42         8
3         4    15701354     Boni         699      France  Female   39         1
4         5    15737888  Mitchell         850      Spain  Female   43         2
```

Столбцы таблицы:

1. RowNumber — номер строки таблицы (это лишняя информация, поэтому можете сразу от неё избавиться)
2. CustomerId — идентификатор клиента
3. Surname — фамилия клиента
4. CreditScore — кредитный рейтинг клиента (чем он выше, тем больше клиент брал кредитов и возвращал их)
5. Geography — страна клиента (банк международный)
6. Gender — пол клиента
7. Age — возраст клиента
8. Tenure — сколько лет клиент пользуется услугами банка
9. Balance — баланс на счетах клиента в банке
10. NumOfProducts — количество услуг банка, которые приобрёл клиент
11. HasCrCard — есть ли у клиента кредитная карта (1 — да, 0 — нет)
12. IsActiveMember — есть ли у клиента статус активного клиента банка (1 — да, 0 — нет)
13. EstimatedSalary — предполагаемая заработная плата клиента
14. Exited — статус лояльности (1 — ушедший клиент, 0 — лояльный клиент)

Итак, банк обращается к вам за помощью: он хочет разработать кампанию лояльности по удержанию клиентов, но для этого ему необходимо, чтобы вы выяснили основные причины оттока клиентов. Иными словами, нужно установить, чем ушедшие клиенты отличаются от лояльных и как между собой связаны различные признаки, определяющие клиентов.

После разведывательного анализа, с целью выявления наиболее важных признаков оттока, банк сможет построить модель машинного обучения, которая будет прогнозировать уход клиента.

## Задания

### 9.2.

**Каково соотношение ушедших и лояльных клиентов? Покажите это на графике и дайте комментарий по соотношению.**

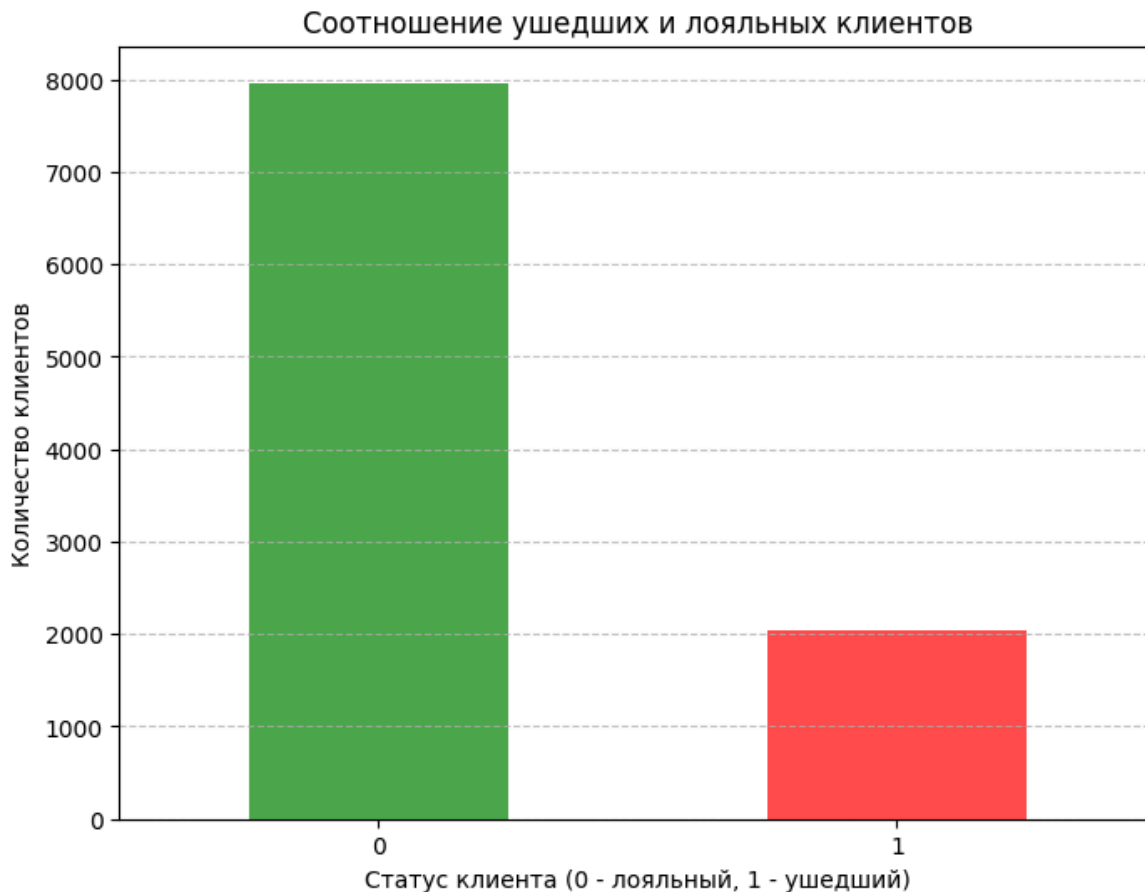
```
In [11]: # Подсчет количества ушедших и лояльных клиентов
counts = churn_data['Exited'].value_counts()

# Построение графика
plt.figure(figsize=(8, 6))
counts.plot(kind='bar', color=['green', 'red'], alpha=0.7)
plt.title('Соотношение ушедших и лояльных клиентов')
plt.xlabel('Статус клиента (0 - лояльный, 1 - ушедший)')
plt.ylabel('Количество клиентов')
plt.xticks(rotation=0)
plt.grid(axis='y', linestyle='--', alpha=0.7)

# Показ графика
plt.show()

# Расчет долей
loyal_ratio = counts[0] / counts.sum()
exited_ratio = counts[1] / counts.sum()

print(f"Лояльные клиенты: {loyal_ratio:.2%}")
print(f"Ушедшие клиенты: {exited_ratio:.2%}")
```

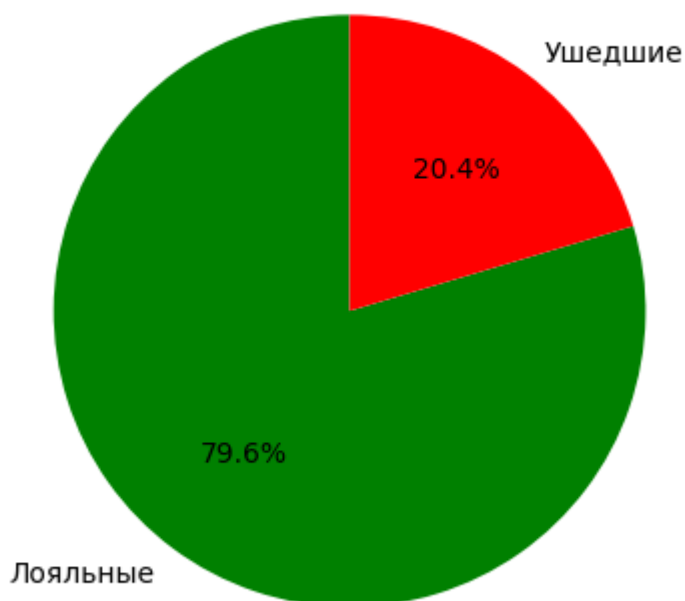


Лояльные клиенты: 79.63%

Ушедшие клиенты: 20.37%

```
In [9]: counts.plot(kind='pie', autopct='%1.1f%%', colors=['green', 'red'], startangle=90)
plt.title('Соотношение ушедших и лояльных клиентов')
plt.ylabel('') # Убираем метку оси Y
plt.show()
```

**Соотношение ушедших и лояльных клиентов**



**Комментарий к диаграммам:**

Диаграмма наглядно подтверждает, что большая часть клиентов лояльна банку (79.63%), в то время как ушедшие клиенты составляют 20.37%.

Хотя процент ушедших клиентов не является катастрофическим, он всё равно может оказать значительное влияние на доходность банка, особенно если это клиенты с высокими балансами или кредитными рейтингами. Поэтому важно разработать стратегии для уменьшения этого процента.

Рекомендации:

- Удержание клиентов: Для дальнейшего снижения оттока стоит провести анализ факторов, которые могут повлиять на уход клиентов, например, недостаток активных продуктов, низкий кредитный рейтинг или отсутствие кредитных карт.
  - Целевая кампания: Возможно, стоит сосредоточиться на группах клиентов, которые наиболее склонны к уходу (например, те, у кого низкий баланс или активность), и предложить им персонализированные условия или дополнительные услуги для повышения лояльности.
1. Столбчатая диаграмма показывает четкое соотношение между ушедшими и лояльными клиентами.
  2. Круговая диаграмма наглядно отображает процентное соотношение ушедших и лояльных клиентов.

---

## 9.2.

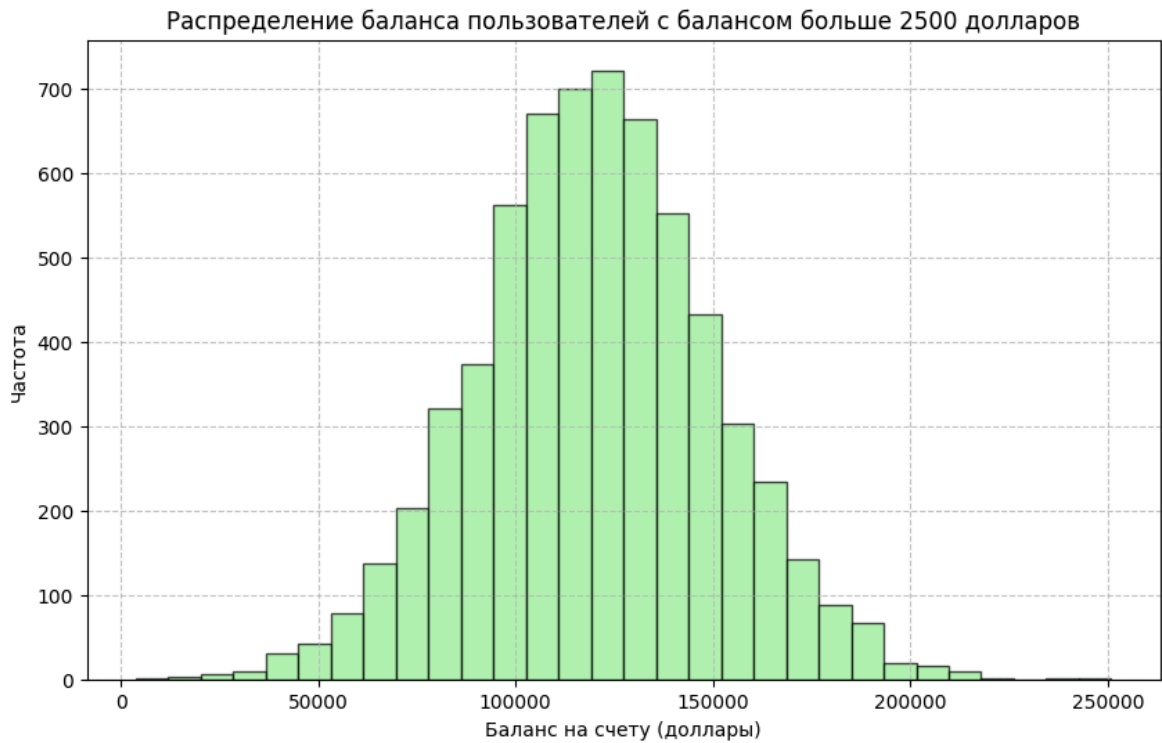
***Постройте график, показывающий распределение баланса пользователей, у которых на счету больше 2 500 долларов. Опишите распределение и сделайте выводы.***

```
In [14]: # Фильтрация данных по балансу больше 2500
df_filtered = churn_data[churn_data['Balance'] > 2500]

# Построение гистограммы
plt.figure(figsize=(10, 6))
plt.hist(df_filtered['Balance'], bins=30, color='lightgreen', edgecolor='black',
plt.title('Распределение баланса пользователей с балансом больше 2500 долларов')
plt.xlabel('Баланс на счету (доллары)')
plt.ylabel('Частота')
plt.grid(True, linestyle='--', alpha=0.7)

# Показать график
plt.show()

# Вывод статистики
print(df_filtered['Balance'].describe())
```



```
count      6383.000000
mean       119827.493793
std        30095.056462
min         3768.690000
25%        100181.975000
50%        119839.690000
75%        139512.290000
max        250898.090000
Name: Balance, dtype: float64
```

#### Статистические данные:

- Количество клиентов: 6383
- Средний баланс: 119,827.49 долларов
- Стандартное отклонение: 30,095.06 долларов
- Минимальный баланс: 3,768.69 долларов
- 25-й процентиль (Q1): 100,181.98 долларов
- Медиана (50-й процентиль): 119,839.69 долларов
- 75-й процентиль (Q3): 139,512.29 долларов
- Максимальный баланс: 250,898.09 долларов

#### Описание распределения:

1. Среднее значение (119,827.49 долларов) достаточно высокое, что указывает на то, что в среднем клиенты имеют значительные суммы на счетах. Это говорит о том, что банк обслуживает клиентов с достаточно высокими финансовыми возможностями.
2. Минимальный баланс (3,768.69 долларов) довольно низкий, что говорит о наличии клиентов с относительно небольшими суммами на счетах. Однако, это значение все равно значительно выше 2 500 долларов, что делает эту выборку релевантной для анализа.

3. Стандартное отклонение (30,095.06 долларов) показывает, что существует значительная вариативность в балансах клиентов. Это означает, что среди клиентов есть как те, кто держит на счетах относительно небольшие суммы, так и те, у кого гораздо большие суммы. Это может указывать на широкое разнообразие в финансовом положении клиентов.
4. Процентили:
  - 25-й процентиль (100,181.98 долларов) показывает, что 25% клиентов имеют баланс ниже этой суммы.
  - Медиана (50-й процентиль) на уровне 119,839.69 долларов говорит о том, что половина клиентов имеет баланс ниже этого значения, а другая половина — выше.
  - 75-й процентиль (139,512.29 долларов) показывает, что 75% клиентов имеют баланс ниже этой суммы, и только 25% имеют баланс выше.
5. Максимальный баланс (250,898.09 долларов) указывает на наличие клиентов с очень высокими суммами на счетах, что может быть связано с "премиум" клиентами.

**Выводы:**

- Склонность к высокому балансу: Средний баланс и медиана (около 120,000 долларов) показывают, что большинство клиентов имеют достаточно высокие суммы на счетах, что свидетельствует о клиентской базе среднего и выше среднего уровня. Это может указывать на стабильность и финансовую устойчивость банка.
- Большая вариативность: Стандартное отклонение в 30,095 долларов подтверждает, что существует значительная вариативность среди клиентов. Хотя большинство клиентов имеет баланс в пределах 100,000–140,000 долларов, некоторые могут иметь гораздо меньшие или большие суммы. Это также подтверждается максимальным значением, которое почти в два раза превышает средний баланс.
- Премиум клиенты: Наличие клиентов с балансом до 250,000 долларов говорит о присутствии в клиентской базе крупных вкладчиков, что может быть важным сегментом для банка.
- Умеренная концентрация: 50% клиентов имеют баланс от 100,000 до 140,000 долларов, что может быть индикатором того, что большинство клиентов находится в этом диапазоне, и банк может сосредоточиться на удержании этого сегмента.

Таким образом, банк имеет разнообразную клиентскую базу с достаточно высокими средними балансами, что важно для стратегии управления активами и маркетинга.

---

## 9.3.

**Посмотрите на распределение баланса клиента в разрезе признака оттока. Как различаются суммы на накопительном счёте ушедших и лояльных клиентов? Подумайте и напишите, с чем это может быть связано, что может не устраивать ушедших клиентов в банке.**

```
In [17]: # Разделение данных на лояльных и ушедших клиентов
loyal_customers = churn_data[churn_data['Exited'] == 0]
exited_customers = churn_data[churn_data['Exited'] == 1]

# Построение гистограммы для лояльных клиентов
plt.figure(figsize=(12, 6))

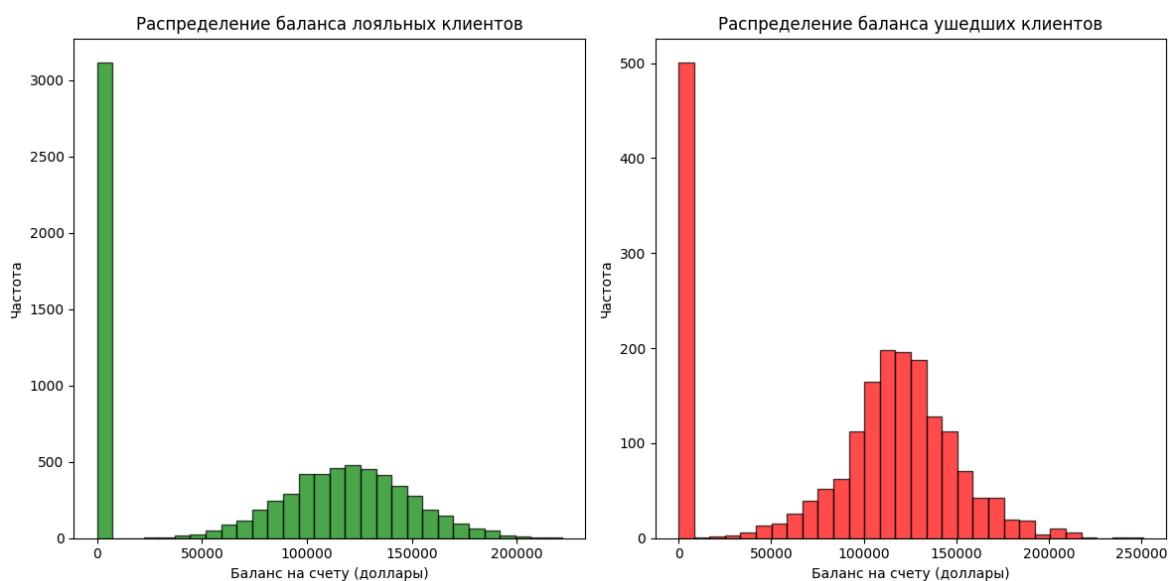
plt.subplot(1, 2, 1)
plt.hist(loyal_customers['Balance'], bins=30, color='green', edgecolor='black',
plt.title('Распределение баланса лояльных клиентов')
plt.xlabel('Баланс на счету (доллары)')
plt.ylabel('Частота')

# Построение гистограммы для ушедших клиентов
plt.subplot(1, 2, 2)
plt.hist(exited_customers['Balance'], bins=30, color='red', edgecolor='black', a
plt.title('Распределение баланса ушедших клиентов')
plt.xlabel('Баланс на счету (доллары)')
plt.ylabel('Частота')

# Показать графики
plt.tight_layout()
plt.show()

# Статистические данные для сравнения
print("Статистика для лояльных клиентов:")
print(loyal_customers['Balance'].describe())

print("\nСтатистика для ушедших клиентов:")
print(exited_customers['Balance'].describe())
```



Статистика для лояльных клиентов:

```
count      7963.000000
mean       72745.296779
std        62848.040701
min         0.000000
25%         0.000000
50%        92072.680000
75%       126410.280000
max       221532.800000
Name: Balance, dtype: float64
```

Статистика для ушедших клиентов:

```
count      2037.000000
mean       91108.539337
std        58360.794816
min         0.000000
25%        38340.020000
50%       109349.290000
75%       131433.330000
max       250898.090000
Name: Balance, dtype: float64
```

### Статистика:

#### 1. Лояльные клиенты:

- Средний баланс: 72,745.30 долларов
- Минимальный баланс: 0 долларов (возможно, это клиенты с нулевыми или очень низкими суммами на счетах)
- 25-й перцентиль (Q1): 0 долларов (половина клиентов имеет баланс менее 92,072.68 долларов)
- Медиана: 92,072.68 долларов (половина лояльных клиентов имеют баланс ниже этой суммы)
- Максимальный баланс: 221,532.80 долларов

#### 2. Ушедшие клиенты:

- Средний баланс: 91,108.54 долларов
- Минимальный баланс: 0 долларов
- 25-й перцентиль (Q1): 38,340.02 долларов
- Медиана: 109,349.29 долларов
- Максимальный баланс: 250,898.09 долларов

### Анализ различий:

#### 1. Средний баланс:

- У ушедших клиентов средний баланс значительно выше (91,108.54 долларов), чем у лояльных клиентов (72,745.30 долларов).
- Это может указывать на то, что клиенты с более высокими суммами на счетах, возможно, ожидают более высоких уровней обслуживания, и если банк не оправдывает их ожидания, они могут покинуть его.

#### 2. Минимальный баланс:



- Как у лояльных, так и у ушедших клиентов есть клиенты с нулевыми балансами, что может указывать на счета, которые не используются активно или на клиентов, которые недавно начали использовать услуги банка. Однако для ушедших клиентов этот факт может быть связан с уходом клиентов, у которых баланс стал равным нулю, возможно, из-за списания средств.

### 3. Процентили:

- Для лояльных клиентов 25% имеют баланс 0 долларов, что может свидетельствовать о наличии клиентов с очень низкими суммами или даже без средств на счетах. Это может быть связано с тем, что они не пользуются активными услугами банка, но при этом остаются в качестве клиентов.
- У ушедших клиентов 25% имеют баланс в районе 38,340.02 долларов, что значительно выше, чем у лояльных. Это может говорить о том, что ушедшие клиенты в среднем имеют более высокие суммы на счетах, но, возможно, не удовлетворены условиями обслуживания, что привело к их уходу.

### 4. Медиана:

- Медиана для лояльных клиентов (92,072.68 долларов) и для ушедших клиентов (109,349.29 долларов) показывает, что большинство ушедших клиентов имеют более высокие суммы на счетах. Это также может свидетельствовать о том, что клиенты с более высокими суммами, скорее всего, ожидали более качественного обслуживания, что может объяснять их уход.

### 5. Максимальный баланс:

- Максимальные значения для обеих групп — 221,532.80 долларов (лояльные) и 250,898.09 долларов (ушедшие) — показывают, что среди ушедших клиентов есть несколько с очень большими суммами на счетах. Это может быть важным сигналом для банка, что высокобалансные клиенты могут быть более чувствительны к качеству обслуживания.

### **Возможные причины ухода клиентов с высокими балансами:**

- Неудовлетворенность условиями обслуживания: Ушедшие клиенты с более высокими балансами могут ожидать более выгодных условий, например, по процентным ставкам, бонусам, привилегиям или более качественного обслуживания. Если банк не удовлетворяет эти ожидания, клиенты могут уйти к конкурентам.
- Высокие комиссии и расходы: Клиенты с высокими балансами могут быть более чувствительны к скрытым комиссиям или невыгодным условиям для крупных вкладчиков. Если банк не предлагает конкурентоспособные ставки или услуги, это может привести к уходу.
- Отсутствие персонализированного обслуживания: Клиенты с высокими суммами на счетах могут ожидать индивидуального подхода и премиум-услуг. Если банк не предоставляет такие услуги или не удовлетворяет ожидания клиентов в этом плане, это может стать причиной оттока.

- Конкуренция: Клиенты с большими суммами могут быть более склонны менять банк, если конкуренты предлагают более привлекательные условия (например, лучшие ставки по депозитам или кредитам, более выгодные условия для премиум-клиентов).

Отток клиентов с более высокими балансами может быть связан с их более высокими ожиданиями от банка. Чтобы уменьшить отток, банк должен сфокусироваться на улучшении условий для крупных клиентов, предложении премиум-услуг и персонализированного обслуживания.

## 9.4.

**Посмотрите на распределение возраста в разрезе признака оттока. В какой группе больше потенциальных выбросов? На какую возрастную категорию клиентов стоит обратить внимание банку?**

```
In [20]: # Разделение данных на лояльных и ушедших клиентов
loyal_customers = churn_data[churn_data['Exited'] == 0]
exited_customers = churn_data[churn_data['Exited'] == 1]

# Построение боксплота для возраста лояльных и ушедших клиентов
plt.figure(figsize=(12, 6))
sbs.boxplot(x='Exited', y='Age', data=churn_data, palette="Set2")

plt.title('Распределение возраста по признаку оттока')
plt.xlabel('Отток клиентов (0 - лояльные, 1 - ушедшие)')
plt.ylabel('Возраст')
plt.show()

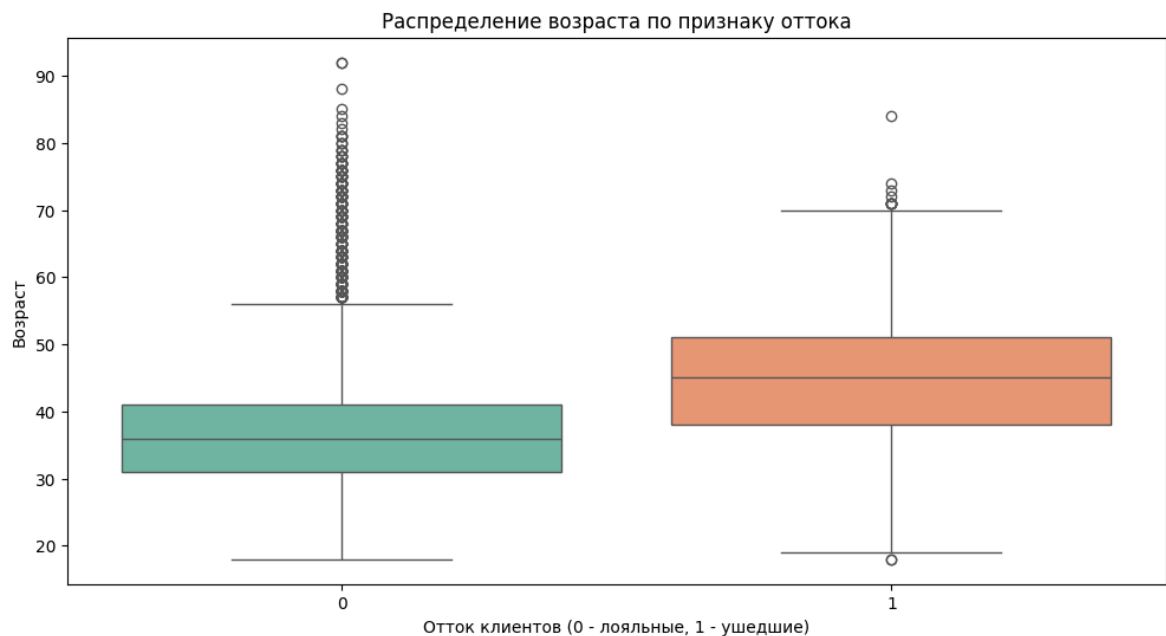
# Статистическая информация по возрасту
print("Статистика для лояльных клиентов:")
print(loyal_customers['Age'].describe())

print("\nСтатистика для ушедших клиентов:")
print(exited_customers['Age'].describe())
```

C:\Users\1\AppData\Local\Temp\ipykernel\_24396\1154953150.py:7: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v 0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

```
sbs.boxplot(x='Exited', y='Age', data=churn_data, palette="Set2")
```



Статистика для лояльных клиентов:

```
count    7963.000000
mean      37.408389
std       10.125363
min       18.000000
25%       31.000000
50%       36.000000
75%       41.000000
max       92.000000
```

Name: Age, dtype: float64

Статистика для ушедших клиентов:

```
count    2037.000000
mean      44.837997
std        9.761562
min       18.000000
25%       38.000000
50%       45.000000
75%       51.000000
max       84.000000
```

Name: Age, dtype: float64

### Статистика:

1. Статистика для лояльных клиентов:

- Средний возраст: 37.41 лет
- Минимальный возраст: 18 лет
- 25-й процентиль (Q1): 31 лет
- Медиана: 36 лет
- 75-й процентиль (Q3): 41 лет
- Максимальный возраст: 92 года

2. Статистика для ушедших клиентов:

- Средний возраст: 44.84 года
- Минимальный возраст: 18 лет
- 25-й процентиль (Q1): 38 лет

- Медиана: 45 лет
- 75-й процентиль (Q3): 51 год
- Максимальный возраст: 84 года

### **Основные выводы:**

#### 1. Возраст ушедших клиентов:

- Средний возраст ушедших клиентов значительно выше (44.84 года) по сравнению с лояльными клиентами (37.41 года).
- Ушедшие клиенты, как правило, старше, и большинство из них находится в возрастной категории от 38 до 51 года.
- Это может свидетельствовать о том, что старшие клиенты могут быть более чувствительны к изменениям в условиях банка, и если банк не удовлетворяет их ожиданиям, они более склонны уйти.

#### 2. Возраст лояльных клиентов:

- Лояльные клиенты в среднем моложе (37.41 года), и их возраст преимущественно варьируется от 18 до 41 года.
- Эта группа более однородна, и, возможно, это клиенты, которые только начинают активно пользоваться банком, что делает их более лояльными в силу привязанности к новым финансовым услугам.

#### 3. Выбросы:

- В обеих группах (лояльных и ушедших) минимальный возраст равен 18 годам, что может свидетельствовать о наличии молодых клиентов, но нет явных возрастных выбросов, таких как очень молодые или очень старые клиенты.
- Однако ушедшие клиенты могут включать больше людей в старших возрастных категориях, что делает эту группу более разнообразной и может указывать на необходимость предложений, ориентированных на старшее поколение.

### **Что это может значить для банка:**

- Старшие клиенты (38-51 год): Для этой группы клиентов банк может разработать специализированные продукты или услуги, такие как пенсионные программы, более выгодные условия по кредитам или депозита, которые будут соответствовать их потребностям и ожиданиям.
- Молодые клиенты (18-37 лет): Молодые клиенты, вероятно, остаются лояльными, так как они только начинают взаимодействовать с банком. Банк может предложить им дополнительные продукты, такие как программы лояльности или бонусы за использование банковских услуг.
- Выбросы и нестандартные возрастные группы: Для ушедших клиентов старшего возраста стоит обратить внимание на возможные изменения в финансовых продуктах или обслуживании, которые могут быть неэффективны для этой возрастной группы, что может объяснить их уход.

Таким образом, для уменьшения оттока клиентов банк может сфокусироваться на создании персонализированных предложений для старших клиентов, улучшении условий для них и более активном вовлечении молодых клиентов через программы лояльности и специальные предложения.

## 9.5.

**Постройте график, который показывает взаимосвязь кредитного рейтинга клиента и его предполагаемой зарплаты. Добавьте расцветку по признаку оттока клиентов. Какова взаимосвязь между признаками? Если не видите явной взаимосвязи, укажите это.**

```
In [30]: # Построение диаграммы рассеяния с расцветкой по признаку оттока
plt.figure(figsize=(10, 6))
sns.scatterplot(x='CreditScore', y='EstimatedSalary', hue='Exited', data=churn_data)

# Настройки графика
plt.title('Взаимосвязь между кредитным рейтингом и предполагаемой зарплатой клиентов')
plt.xlabel('Кредитный рейтинг')
plt.ylabel('Предполагаемая зарплата')
plt.legend(title='Отток клиентов', loc='upper right')
plt.show()

# Статистика по кредитному рейтингу и зарплате для лояльных и ушедших клиентов
loyal_clients_stats = churn_data[churn_data['Exited'] == 0][['CreditScore', 'EstimatedSalary']].describe()
churned_clients_stats = churn_data[churn_data['Exited'] == 1][['CreditScore', 'EstimatedSalary']].describe()

print("Статистика для лояльных клиентов:")
print(loyal_clients_stats)
print("\nСтатистика для ушедших клиентов:")
print(churned_clients_stats)
```



Статистика для лояльных клиентов:

	CreditScore	EstimatedSalary
count	7963.000000	7963.000000
mean	651.853196	99738.391772
std	95.653837	57405.586966
min	405.000000	90.070000
25%	585.000000	50783.490000
50%	653.000000	99645.040000
75%	718.000000	148609.955000
max	850.000000	199992.480000

Статистика для ушедших клиентов:

	CreditScore	EstimatedSalary
count	2037.000000	2037.000000
mean	645.351497	101465.677531
std	100.321503	57912.418071
min	350.000000	11.580000
25%	578.000000	51907.720000
50%	646.000000	102460.840000
75%	716.000000	152422.910000
max	850.000000	199808.100000

```
In [31]: # Корреляция между кредитным рейтингом и зарплатой
correlation = churn_data['CreditScore'].corr(churn_data['EstimatedSalary'])
print(f'Корреляция между кредитным рейтингом и зарплатой: {correlation:.2f}')
```

Корреляция между кредитным рейтингом и зарплатой: -0.00

```
In [32]: # Статистика по кредитному рейтингу и предполагаемой зарплате
credit_salary_stats = churn_data[['CreditScore', 'EstimatedSalary']].describe()
print(credit_salary_stats)
```

	CreditScore	EstimatedSalary
count	10000.000000	10000.000000
mean	650.528800	100090.239881
std	96.653299	57510.492818
min	350.000000	11.580000
25%	584.000000	51002.110000
50%	652.000000	100193.915000
75%	718.000000	149388.247500
max	850.000000	199992.480000

### Выводы по полученным данным:

#### 1. Средний кредитный рейтинг:

- У лояльных клиентов: 651.85.
- У ушедших клиентов: 645.35.
- Разница в среднем рейтинге незначительна (около 6 пунктов), что говорит о том, что кредитный рейтинг сам по себе не является определяющим фактором для оттока.

#### 2. Средняя зарплата:

- У лояльных клиентов: 99,738.39.
- У ушедших клиентов: 101,465.68.
- У ушедших клиентов зарплата в среднем чуть выше (~1,727.29), что может свидетельствовать о том, что клиенты с более высокой зарплатой могут искать

лучшие условия в других банках.

### 3. Диапазон значений:

- Минимальные значения CreditScore и EstimatedSalary схожи, однако у ушедших клиентов встречаются очень низкие значения зарплат (11.58). Это может быть связано с отсутствием данных или некорректными записями.
- Максимальные значения для обоих признаков почти идентичны, что указывает на схожий уровень высокооплачиваемых клиентов в обеих группах.

### 4. Распределение:

- Межквартильный размах (IQR):
  - У лояльных клиентов зарплата более равномерно распределена (25% клиентов имеют зарплату менее 50,783, а 75% — менее 148,610).
  - У ушедших клиентов диапазон зарплат чуть выше (25% — менее 51,907, 75% — менее 152,423).
- Кредитный рейтинг в обеих группах имеет схожий межквартильный размах, но медиана у ушедших клиентов немного ниже (646 против 653).

### 5. Корреляция:

- Корреляция между кредитным рейтингом и предполагаемой зарплатой: -0.00.
- Это говорит об отсутствии линейной взаимосвязи между этими двумя признаками. Высокий кредитный рейтинг не зависит от предполагаемой зарплаты и наоборот.

### Заключение:

Кредитный рейтинг и зарплата в обеих группах не сильно отличаются. Это означает, что для удержания клиентов эти признаки не являются ключевыми индикаторами. Лояльные клиенты имеют немного выше медианный кредитный рейтинг, что может говорить о том, что банк удерживает клиентов с более стабильной финансовой историей.

У ушедших клиентов зарплата в среднем выше, что может свидетельствовать о том, что они ищут лучшие условия обслуживания или продукты, которые больше соответствуют их уровню дохода. Банку стоит обратить внимание на клиентов с высоким доходом и средней кредитной историей, так как они могут быть более склонны к уходу.

---

## 9.6.

***Кто чаще уходит, мужчины или женщины? Постройте график, который иллюстрирует это.***

### Подсказка

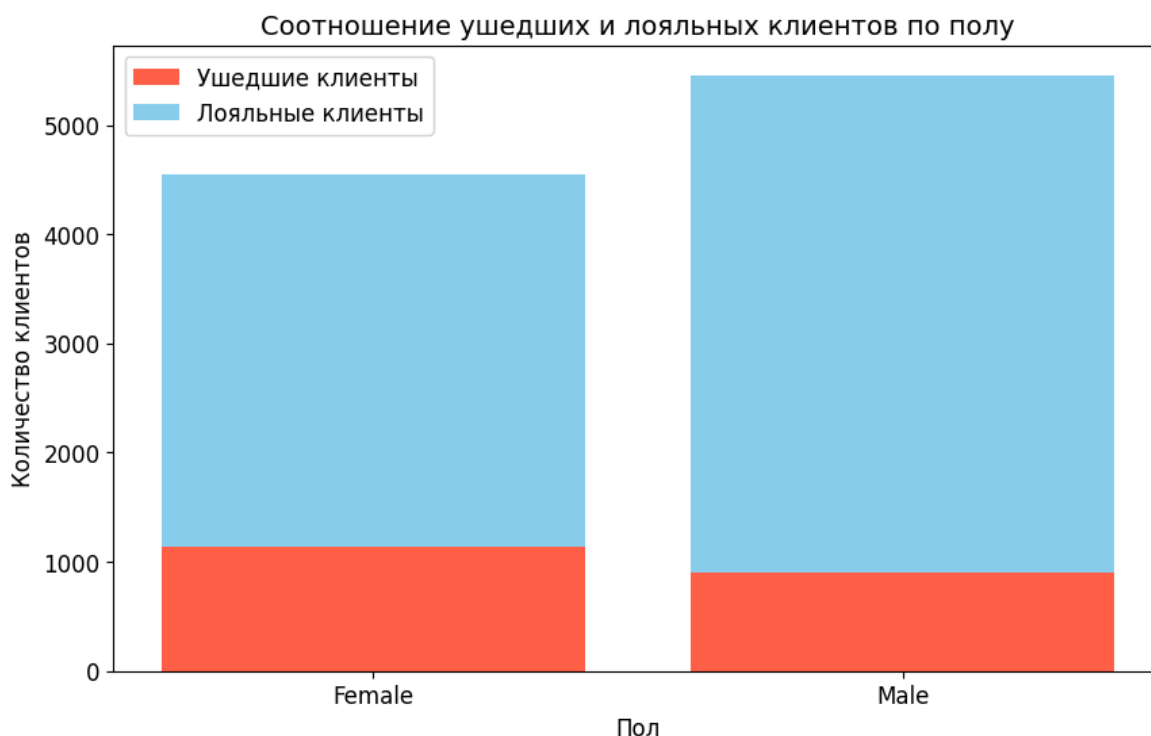
Процент ушедших клиентов в каждой группе можно рассчитать как среднее по

столбцу Exited (так как 1 — это ушедшие клиенты, а 0 — лояльные, среднее арифметическое по столбцу обозначает долю ушедших клиентов).

```
In [41]: # Группировка данных
gender_stats = churn_data.groupby('Gender').agg(
    TotalClients=('Exited', 'count'),
    ExitedClients=('Exited', 'sum'),
    LoyalClients=('Exited', lambda x: (x == 0).sum())
).reset_index()

# Построение составной диаграммы
plt.figure(figsize=(10, 6))
plt.bar(gender_stats['Gender'], gender_stats['ExitedClients'], label='Ушедшие кл
plt.bar(gender_stats['Gender'], gender_stats['LoyalClients'], bottom=gender_stat

# Настройка графика
plt.title('Соотношение ушедших и лояльных клиентов по полу', fontsize=14)
plt.xlabel('Пол', fontsize=12)
plt.ylabel('Количество клиентов', fontsize=12)
plt.legend(fontsize=12)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.show()
```



```
In [42]: # Группировка по полу и расчет статистики
gender_stats = churn_data.groupby('Gender').agg(
    TotalClients=('Exited', 'count'),
    ExitedClients=('Exited', 'sum'),
    ExitRate=('Exited', 'mean')
).reset_index()

# Переименуем столбцы для удобства
gender_stats.columns = ['Gender', 'TotalClients', 'ExitedClients', 'ExitRate']

# Вывод статистики
print(gender_stats)
```



	Gender	TotalClients	ExitedClients	ExitRate
0	Female	4543	1139	0.250715
1	Male	5457	898	0.164559

### Подробный анализ данных

Исходные данные:

1. Количество клиентов по полу:

- Женщины: 4,543 клиента.
- Мужчины: 5,457 клиентов.

2. Количество ушедших клиентов:

- Женщины: 1,139 ушедших клиентов.
- Мужчины: 898 ушедших клиентов.

3. Доля ушедших клиентов (ExitRate):

- Женщины: 25.07% (0.2507).
- Мужчины: 16.46% (0.1646).

### Интерпретация:

1. Общее количество клиентов:

- Мужчин в выборке больше, чем женщин (54.57% против 45.43%). Это говорит о том, что мужчины составляют большую часть клиентской базы банка.

2. Доля ушедших клиентов:

- Женщины чаще уходят из банка, чем мужчины. Доля ушедших среди женщин составляет 25.07%, что значительно выше, чем у мужчин (16.46%). Это может свидетельствовать о том, что женщины менее удовлетворены услугами банка или более склонны переходить к конкурентам.

3. Абсолютное количество ушедших:

- Хотя женщин в клиентской базе меньше, их абсолютное количество среди ушедших (1,139) больше, чем у мужчин (898). Это подчеркивает, что проблема удержания женщин-клиентов для банка более остра.

### Возможные причины более высокого оттока среди женщин:

1. Недостаточная персонализация услуг:

- Женщины могут иметь другие финансовые приоритеты (например, накопления, семейные расходы), которые банк не удовлетворяет в полной мере.

2. Меньшее вовлечение:

- Возможно, женщины менее вовлечены в долгосрочные финансовые продукты, которые связывают клиента с банком.

### 3. Конкуренция:

- Женщины могут быть более чувствительны к условиям обслуживания, предложениям и удобству других банков.

### Рекомендации банку:

#### 1. Персонализация продуктов:

- Разработать специальные предложения для женщин, учитывающие их финансовые потребности (например, продукты для накоплений или семейного бюджета).

#### 2. Обратная связь:

- Провести опрос среди ушедших клиенток, чтобы понять их недовольство.

#### 3. Улучшение клиентского опыта:

- Повысить удобство обслуживания (мобильные приложения, скорость обработки запросов).

#### 4. Программы лояльности:

- Запустить акционные программы или бонусы, которые удержат женщин в качестве клиентов.

Банку стоит сосредоточиться на снижении оттока среди женщин, так как это самая уязвимая группа клиентов.

---

## 9.7.

**Как отток клиентов зависит от числа приобретённых у банка услуг? Для ответа на этот вопрос постройте многоуровневую столбчатую диаграмму.**

```
In [52]: # Группировка данных
services_stats = churn_data.groupby(['NumOfProducts', 'Exited']).size().unstack()

# Преобразуем данные в проценты для анализа
services_stats_percent = services_stats.div(services_stats.sum(axis=1), axis=0)

# Построение графика
plt.figure(figsize=(12, 6))
bars = services_stats_percent.plot(
    kind='bar',
    stacked=True,
    figsize=(12, 6),
    color=['skyblue', 'tomato'],
    alpha=0.85,
    edgecolor='black'
)
```

```

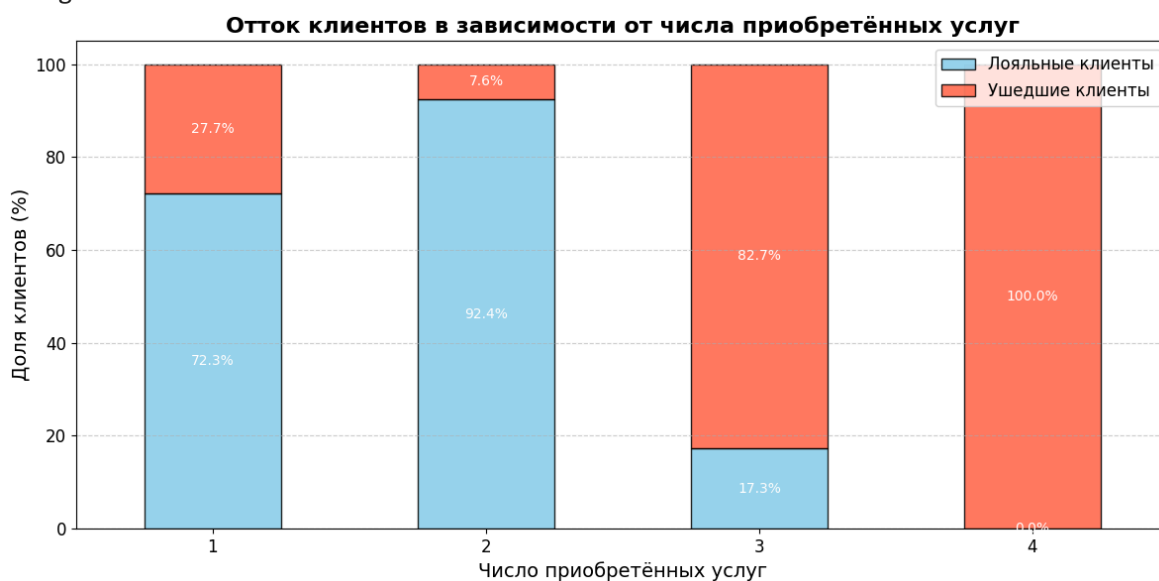
# Настройка аннотаций
for container in bars.containers:
    bars.bar_label(container, fmt='%.1f%%', label_type='center', fontsize=10, co

# Настройка графика
plt.title('Отток клиентов в зависимости от числа приобретённых услуг', fontsize=
plt.xlabel('Число приобретённых услуг', fontsize=14)
plt.ylabel('Доля клиентов (%)', fontsize=14)
plt.xticks(rotation=0, fontsize=12)
plt.yticks(fontsize=12)
plt.legend(['Лояльные клиенты', 'Ушедшие клиенты'], fontsize=12, loc='upper right
plt.tight_layout()
plt.grid(axis='y', linestyle='--', alpha=0.6)

# Отображение графика
plt.show()

```

<Figure size 1200x600 with 0 Axes>



```

In [53]: # Рассчитываем процент ушедших клиентов по числу услуг
exit_rate_by_services = churn_data.groupby('NumOfProducts')['Exited'].mean() * 1

# Выводим статистику для процентного оттока
statistics = exit_rate_by_services.describe()

# Печатаем статистику
print(statistics)

```

```

count      4.000000
mean       54.500716
std        43.911529
min         7.581699
25%        22.681223
50%        55.210583
75%        87.030075
max        100.000000
Name: Exited, dtype: float64

```

### Выводы:

1. Высокий средний отток: Средний процент ушедших клиентов составляет 54.5%, что говорит о достаточно высоком уровне оттока. Банк должен обратить

внимание на это и постараться понять причины, почему клиенты уходят, несмотря на приобретённые услуги.

2. Неоднородность оттока: Стандартное отклонение в 43.91% указывает на значительную вариативность оттока в зависимости от числа услуг. Для некоторых групп клиентов отток крайне низкий, для других — очень высокий.
3. Группы с высоким оттоком: Особенно стоит обратить внимание на группы, где отток превышает 75% (например, для некоторых групп отток составляет 100%). Это может указывать на проблемы с удовлетворённостью клиентов или недостаточной ценностью для них предлагаемых услуг.
4. Группы с низким оттоком: Для групп с меньшим числом услуг (например, 7.58%) отток минимален, что может говорить о том, что клиенты с небольшим количеством услуг более лояльны или довольны текущими условиями.

Таким образом, для эффективной кампании по удержанию клиентов банк должен провести более детальный анализ, чтобы понять, какие именно группы клиентов с определённым числом услуг склонны к уходу, и разработать стратегии для снижения оттока среди них.

## 9.8.

**Как влияет наличие статуса активного клиента на отток клиентов? Постройте диаграмму, иллюстрирующую это. Что бы вы предложили банку, чтобы уменьшить отток клиентов среди неактивных?**

```
In [58]: # Рассчитаем процент ушедших клиентов для активных и неактивных
exit_rate_by_activity = churn_data.groupby('IsActiveMember')['Exited'].mean() *

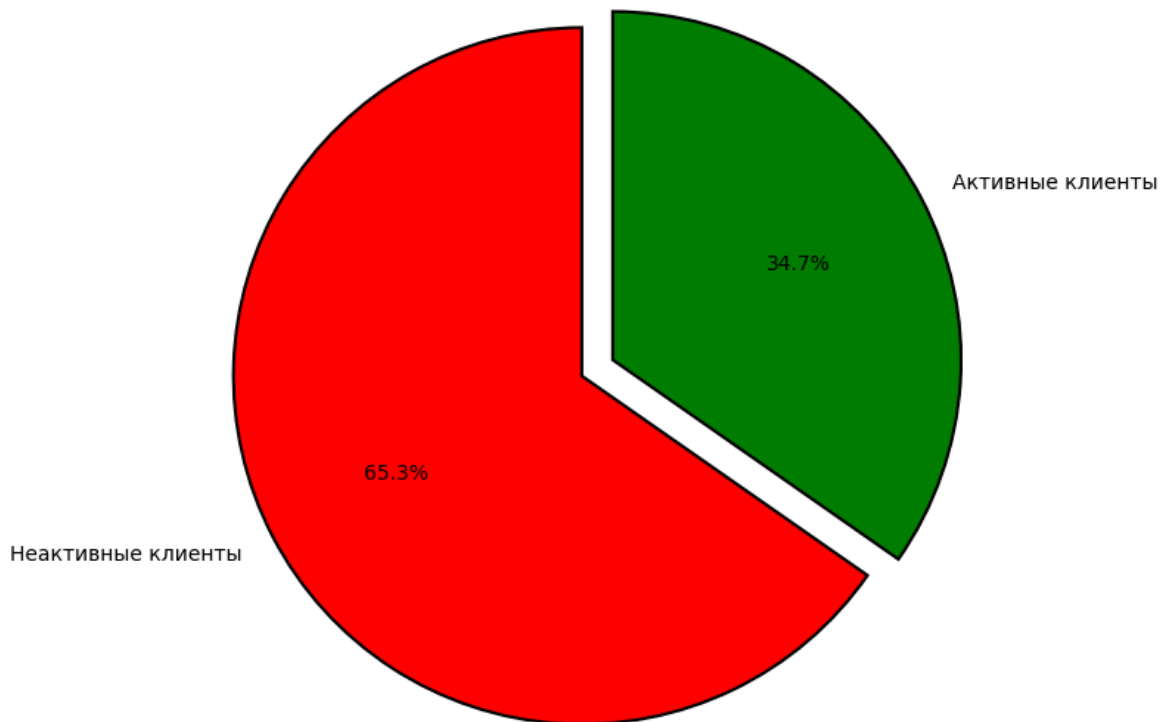
# Данные для круговой диаграммы
labels = ['Неактивные клиенты', 'Активные клиенты']
sizes = exit_rate_by_activity.values
colors = ['red', 'green']
explode = (0.1, 0) # Выделим неактивных клиентов

# Построение круговой диаграммы
plt.figure(figsize=(8, 8))
plt.pie(sizes, explode=explode, labels=labels, colors=colors, autopct='%1.1f%%',
        wedgeprops={'edgecolor': 'black', 'linewidth': 1.5, 'linestyle': 'solid'})

# Заголовок
plt.title('Отток клиентов в зависимости от статуса активности', fontsize=16, font

# Отображение графика
plt.tight_layout()
plt.show()
```

## Отток клиентов в зависимости от статуса активности



```
In [59]: # Рассчитываем статистику для оттока по статусу активности
exit_stats_by_activity = churn_data.groupby('IsActiveMember')['Exited'].describe

# Выводим статистику
print(exit_stats_by_activity)
```

	count	mean	std	min	25%	50%	75%	max
IsActiveMember								
0	4849.0	0.268509	0.443230	0.0	0.0	0.0	1.0	1.0
1	5151.0	0.142691	0.349791	0.0	0.0	0.0	0.0	1.0

### Статистика:

#### 1. Неактивные клиенты (IsActiveMember = 0):

- Средний процент оттока: 26.85% (0.268509)
- Стандартное отклонение: 44.32% (высокая изменчивость оттока среди неактивных клиентов).
- Минимальное значение оттока: 0%, максимальное значение: 100%.
- 25% клиентов имеют отток 0%, 50% — также 0%, и 75% — имеют отток 100%. Это означает, что среди неактивных клиентов есть значительная группа, которая осталась лояльной, и есть клиенты, которые ушли.

#### 2. Активные клиенты (IsActiveMember = 1):

- Средний процент оттока: 14.27% (0.142691)
- Стандартное отклонение: 34.98% (меньшая изменчивость по сравнению с неактивными клиентами).

- Минимальное значение оттока: 0%, максимальное значение: 100%.
- 25% клиентов имеют отток 0%, 50% — также 0%, и 75% — имеют отток 0%. Это указывает на то, что большинство активных клиентов остаются лояльными, и лишь небольшая доля уходит.

**Выводы:**

1. Активные клиенты имеют значительно меньший процент оттока (14.27%) по сравнению с неактивными клиентами (26.85%). Это свидетельствует о том, что статус активности клиента в банке является важным фактором, влияющим на его лояльность. Неактивные клиенты более склонны уходить, возможно, потому, что они не вовлечены в процесс обслуживания или не получают достаточную ценность от услуг банка.
2. Неактивные клиенты показывают более высокое стандартное отклонение в проценте оттока, что говорит о том, что среди них есть как очень лояльные клиенты, так и те, кто ушел.

**Рекомендации для банка:**

Чтобы уменьшить отток клиентов среди неактивных:

1. Повышение вовлеченности: Банк должен предложить дополнительные стимулы для активации клиентов, такие как персонализированные предложения, бонусы за активность, консультации по использованию дополнительных услуг.
2. Обучение и поддержка: Банку стоит провести образовательные кампании, чтобы показать неактивным клиентам, как можно извлечь максимальную пользу из предлагаемых продуктов и услуг.
3. Регулярная связь с клиентами: Важно поддерживать контакт с клиентами через электронную почту, SMS или мобильное приложение, напоминая о возможностях и акциях, доступных в банке.
4. Анализ причин неактивности: Провести опросы или анализ поведения клиентов, чтобы понять, почему они не используют услуги банка. Возможно, есть проблемы с продуктами или с клиентским обслуживанием.
5. Персонализированные предложения: Для неактивных клиентов можно предложить персонализированные скидки или улучшенные условия на продукты банка, чтобы вернуть их к активному использованию.

Эти шаги помогут банку не только снизить отток среди неактивных клиентов, но и повысить общую лояльность клиентов.

---

## 9.9.

***В какой стране доля ушедших клиентов больше? Постройте тепловую картограмму, которая покажет это соотношение на карте мира. Предположите, с чем это может быть связано.***

```
In [60]: # Группируем данные по странам и рассчитываем долю ушедших клиентов
country_exit_rate = churn_data.groupby('Geography')['Exited'].mean()

# Преобразуем данные в DataFrame для удобной работы
country_exit_rate = country_exit_rate.reset_index()
country_exit_rate.columns = ['Country', 'ExitRate']

# Проверяем данные
print(country_exit_rate.head())
```

```
   Country  ExitRate
0  France    0.161548
1  Germany   0.324432
2   Spain    0.166734
```

```
In [61]: import plotly.express as px

# Построение тепловой карты
fig = px.choropleth(country_exit_rate,
                    locations="Country",
                    locationmode="country names", # Используем название страны
                    color="ExitRate",
                    color_continuous_scale="Viridis", # Выбираем цветовую шкалу
                    labels={"ExitRate": "Доля ушедших клиентов (%)"},
                    title="Доля ушедших клиентов по странам")

# Отображаем карту
fig.show()
```

### Анализ данных:

Из предоставленных данных мы видим следующие значения для доли ушедших клиентов:

- France: 16.15%
- Germany: 32.44%
- Spain: 16.67%

Из этого можно сделать вывод, что Germany (Германия) имеет наибольшую долю ушедших клиентов — 32.44%.

### Возможные причины более высокого оттока в Германии:

1. Конкуренция на рынке: Германия — одна из самых экономически развитых стран Европы с высокоразвитыми финансовыми рынками. В таких странах конкуренция между банками может быть особенно жесткой. Клиенты могут уходить в поисках более выгодных условий, лучших продуктов или более удобных сервисов, таких как мобильные приложения и онлайн-банкинг.
2. Качество обслуживания: Если клиенты в Германии чувствуют, что банк не удовлетворяет их потребности или что качество обслуживания недостаточно высоко, это может привести к более высокому оттоку. Например, долгие ожидания в обслуживании, недостаток персонализированных предложений или проблемы с доступностью продуктов и услуг.

3. Экономические факторы: Германия, несмотря на свою стабильную экономику, может испытывать экономические колебания или изменения в законодательстве, которые могут повлиять на поведение клиентов. Например, изменения в налоговом законодательстве или в банковских правилах могут повлиять на финансовое поведение клиентов.
4. Финансовая грамотность: В странах с высокоразвитыми рынками, такими как Германия, клиенты могут быть более финансово грамотными и осведомленными о возможностях на рынке. Это может означать, что они легче переключаются между банками, если не удовлетворены текущими условиями.
5. Технологические инновации: В странах с высоким уровнем цифровизации, таких как Германия, клиенты могут уходить к более инновационным банкам, которые предлагают более удобные или продвинутые технологии, такие как мобильные платежи, цифровые кошельки или инвестиционные платформы.

### Предложения для банка:

Для уменьшения оттока клиентов в Германии банк может:

- Улучшить качество обслуживания: Важно инвестировать в обучение сотрудников, улучшить поддержку клиентов и предоставить более персонализированные услуги.
- Внедрить инновации: Внедрение новых технологий, таких как мобильные приложения, искусственный интеллект для персонализированных предложений и чат-боты для улучшения взаимодействия с клиентами.
- Анализировать конкурентные предложения: Провести исследование конкурентных предложений на рынке и, возможно, пересмотреть условия для привлечения и удержания клиентов.
- Обратит внимание на лояльность клиентов: Ввести программы лояльности или специальные предложения для клиентов, которые могут стимулировать их оставаться с банком.

Эти шаги помогут снизить отток и повысить лояльность клиентов в Германии.

---

## 9.10.

**Переведите числовой признак `CreditScore` в категориальный. Для этого воспользуйтесь функцией `get_credit_score_cat()`, которая приведена ниже. Примените её к столбцу `CreditScore` и создайте новый признак `CreditScoreCat` — категории кредитного рейтинга.**

Постройте сводную таблицу, строками которой являются категории кредитного рейтинга (`CreditScoreCat`), а столбцами — количество лет, в течение которых клиент пользуется услугами банка (`Tenure`). В ячейках сводной таблицы должно находиться среднее по признаку оттока (`Exited`) — доля ушедших пользователей.



На основе полученной сводной таблицы постройте тепловую карту с аннотацией. Найдите на тепловой карте категории клиентов, которые уходят чаще всего.

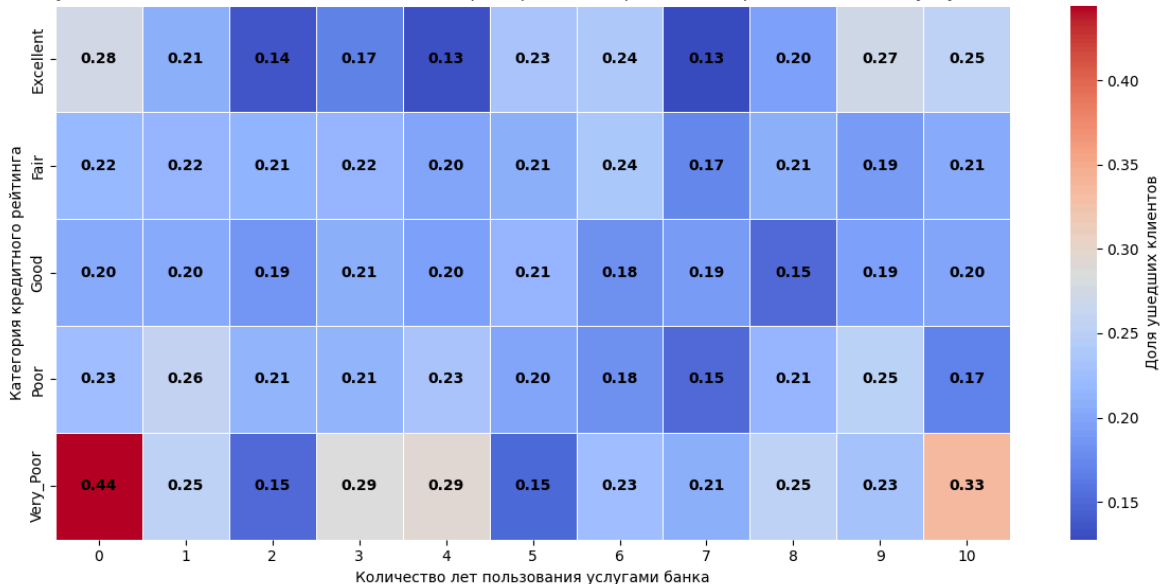
```
In [70]: # Пример функции для преобразования CreditScore в категориальный признак
def get_credit_score_cat(credit_score):
    if credit_score >= 300 and credit_score < 500:
        return "Very_Poor"
    elif credit_score >= 500 and credit_score < 601:
        return "Poor"
    elif credit_score >= 601 and credit_score < 661:
        return "Fair"
    elif credit_score >= 661 and credit_score < 781:
        return "Good"
    elif credit_score >= 781 and credit_score < 851:
        return "Excellent"
    elif credit_score >= 851:
        return "Top"
    elif credit_score < 300:
        return "Deep"

# Применим функцию к столбцу CreditScore для создания нового столбца CreditScore
churn_data['CreditScoreCat'] = churn_data['CreditScore'].apply(get_credit_score_cat)

# Построим сводную таблицу
pivot_table = churn_data.pivot_table(
    values='Exited',
    index='CreditScoreCat',
    columns='Tenure',
    aggfunc='mean'
)

# Построим тепловую карту с настройкой аннотаций
plt.figure(figsize=(12, 6))
sbs.heatmap(
    pivot_table,
    annot=True,
    annot_kws={'size': 10, 'weight': 'bold', 'color': 'black'}, # Настройки для
    cmap='coolwarm',
    fmt='.2f',
    linewidths=0.5,
    cbar_kws={'label': 'Доля ушедших клиентов'}
)
plt.title('Доля ушедших клиентов в зависимости от категории кредитного рейтинга')
plt.xlabel('Количество лет пользования услугами банка')
plt.ylabel('Категория кредитного рейтинга')
plt.tight_layout()
plt.show()
```

Доля ушедших клиентов в зависимости от категории кредитного рейтинга и срока пользования услугами банка



```
In [65]: # Общая статистика для всех числовых столбцов
print(churn_data.describe())
```

	RowNumber	CustomerId	CreditScore	Age	Tenure
count	10000.00000	1.000000e+04	10000.000000	10000.000000	10000.000000
mean	5000.50000	1.569094e+07	650.528800	38.921800	5.012800
std	2886.89568	7.193619e+04	96.653299	10.487806	2.892174
min	1.00000	1.556570e+07	350.000000	18.000000	0.000000
25%	2500.75000	1.562853e+07	584.000000	32.000000	3.000000
50%	5000.50000	1.569074e+07	652.000000	37.000000	5.000000
75%	7500.25000	1.575323e+07	718.000000	44.000000	7.000000
max	10000.00000	1.581569e+07	850.000000	92.000000	10.000000

	Balance	NumOfProducts	HasCrCard	IsActiveMember
count	10000.000000	10000.000000	10000.000000	10000.000000
mean	76485.889288	1.530200	0.70550	0.515100
std	62397.405202	0.581654	0.45584	0.499797
min	0.000000	1.000000	0.00000	0.000000
25%	0.000000	1.000000	0.00000	0.000000
50%	97198.540000	1.000000	1.00000	1.000000
75%	127644.240000	2.000000	1.00000	1.000000
max	250898.090000	4.000000	1.00000	1.000000

	EstimatedSalary	Exited
count	10000.000000	10000.000000
mean	100090.239881	0.203700
std	57510.492818	0.402769
min	11.580000	0.000000
25%	51002.110000	0.000000
50%	100193.915000	0.000000
75%	149388.247500	0.000000
max	199992.480000	1.000000

```
In [66]: # Частотный анализ для категориального столбца
print(churn_data['CreditScoreCat'].value_counts())
print(churn_data['Tenure'].value_counts())
```

```
CreditScoreCat
Good          3703
Poor          2434
Fair          2281
Excellent     950
Very_Poor     632
Name: count, dtype: int64

Tenure
2           1048
1           1035
7           1028
8           1025
5           1012
3           1009
4            989
9            984
6            967
10           490
0            413
Name: count, dtype: int64
```

```
In [67]: # Группировка данных по категории кредитного рейтинга
grouped = churn_data.groupby('CreditScoreCat')['Exited'].mean()
print(grouped)
```

```
CreditScoreCat
Excellent     0.196842
Fair          0.206050
Good          0.193087
Poor          0.211586
Very_Poor     0.237342
Name: Exited, dtype: float64
```

### Доля ушедших клиентов по категориям кредитного рейтинга:

- Very\_Poor (Очень плохой): Доля ушедших клиентов в этой категории составляет 23.73%. Это наибольшая доля среди всех категорий.
- Poor (Плохой): Здесь доля ушедших клиентов составляет 21.16%, что также довольно высоко, но немного ниже, чем в категории "Very\_Poor".
- Fair (Средний): В этой категории доля ушедших клиентов составляет 20.61%, что чуть ниже, чем в категориях "Poor" и "Very\_Poor".
- Good (Хороший): Для клиентов с хорошим кредитным рейтингом доля ушедших клиентов составляет 19.31%. Это также выше среднего, но ниже, чем в предыдущих категориях.
- Excellent (Отличный): Для клиентов с отличным кредитным рейтингом доля ушедших клиентов составляет 19.68%, что является самой низкой среди всех категорий, хотя разница с другими категориями незначительна.

### Анализ тенденций:

- Клиенты с более низким кредитным рейтингом (например, Very\_Poor и Poor) имеют более высокую вероятность покинуть банк. Это может свидетельствовать о том, что клиенты с плохой кредитной историей могут сталкиваться с трудностями в обслуживании или более высокими ставками, что может увеличить их желание уйти.

- Напротив, клиенты с Excellent кредитным рейтингом имеют самую низкую вероятность ухода, что может быть связано с более выгодными условиями для таких клиентов или их общей удовлетворённостью сервисом.

**Выводы для банка:**

- Банку стоит обратить внимание на клиентов с низким кредитным рейтингом (особенно в категориях Very\_Poor и Poor), так как их доля ушедших клиентов значительно выше. Это может указывать на проблемы с удержанием таких клиентов, например, из-за высоких ставок, отказа в кредитах или других факторов.
- Для клиентов с высоким кредитным рейтингом (например, Excellent), несмотря на меньшую долю ухода, важно продолжать поддерживать хорошие условия и повышать лояльность, чтобы предотвратить их уход, особенно в условиях конкурентного рынка.