



# КОПИТЬ ИЛИ КУПИТЬ?

The Mind: Иванова Ульяна, Дидык Ирина

БД: САМОЛЁТ





# Описание данных

Всего 23332 строк, 202 колонки. Каждая строка - отчет по клиенту, интересовавшемуся недвижимостью. Всего 8302 уникальных клиента.

Дата получения информации: ежемесячно с февраля 2021 по июль 2023.

Отчеты содержат данные:

- по клиентам (id клиента, проводилась ли коммуникация, вероятность наличия машины),
- по звонкам,
- по интересам клиентов,
- по договорам,
- по ипотечным кредитам.

A detailed architectural floor plan of a house is shown, featuring a living room, two bedrooms, a bathroom, and a porch. A pencil and an eraser are resting on the plan. The drawing includes various dimensions and room labels. A semi-transparent grey box with a blue and black border is overlaid on the top right of the image, containing the title.

# План исследования

1. Подготовка данных и предварительный анализ
2. Постановка исследовательского вопроса и гипотезы
3. Математический анализ
4. Выводы
5. Ограничения и перспективы
6. Policy implication

# Чистка от выбросов и подготовка данных

Осталось 114 колонок, 8 302 строки.

## Чистка

- Отсутствующие значения заменены на ноль.
- Удалили клиентов с суммой договора более 100 млн.
- Убраны колонки, в которых только ноль.
- Убраны этажи меньше 1.

## Подготовка

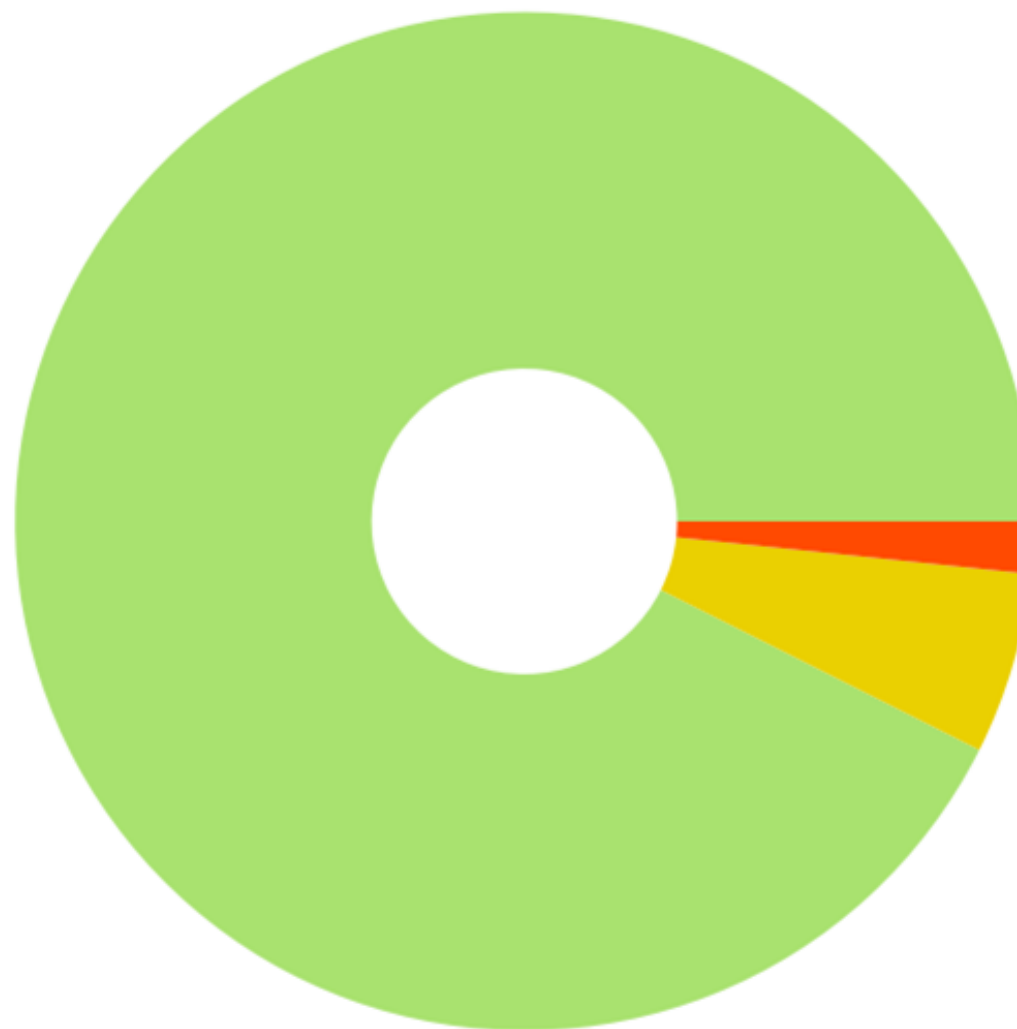
- Даты преобразованы из строкового типа в тип дат и времени
- Значения в колонках с количественными значениями преобразованы в целочисленный тип данных.
- По каждому клиенту взят последний отчет.

# Распределение целевых интересов клиентов

По данным за 3 месяца

Подавляющее большинство  
клиентов интересуются  
недвижимостью для себя.

Для себя (92.59%)



Для других (1.65%)

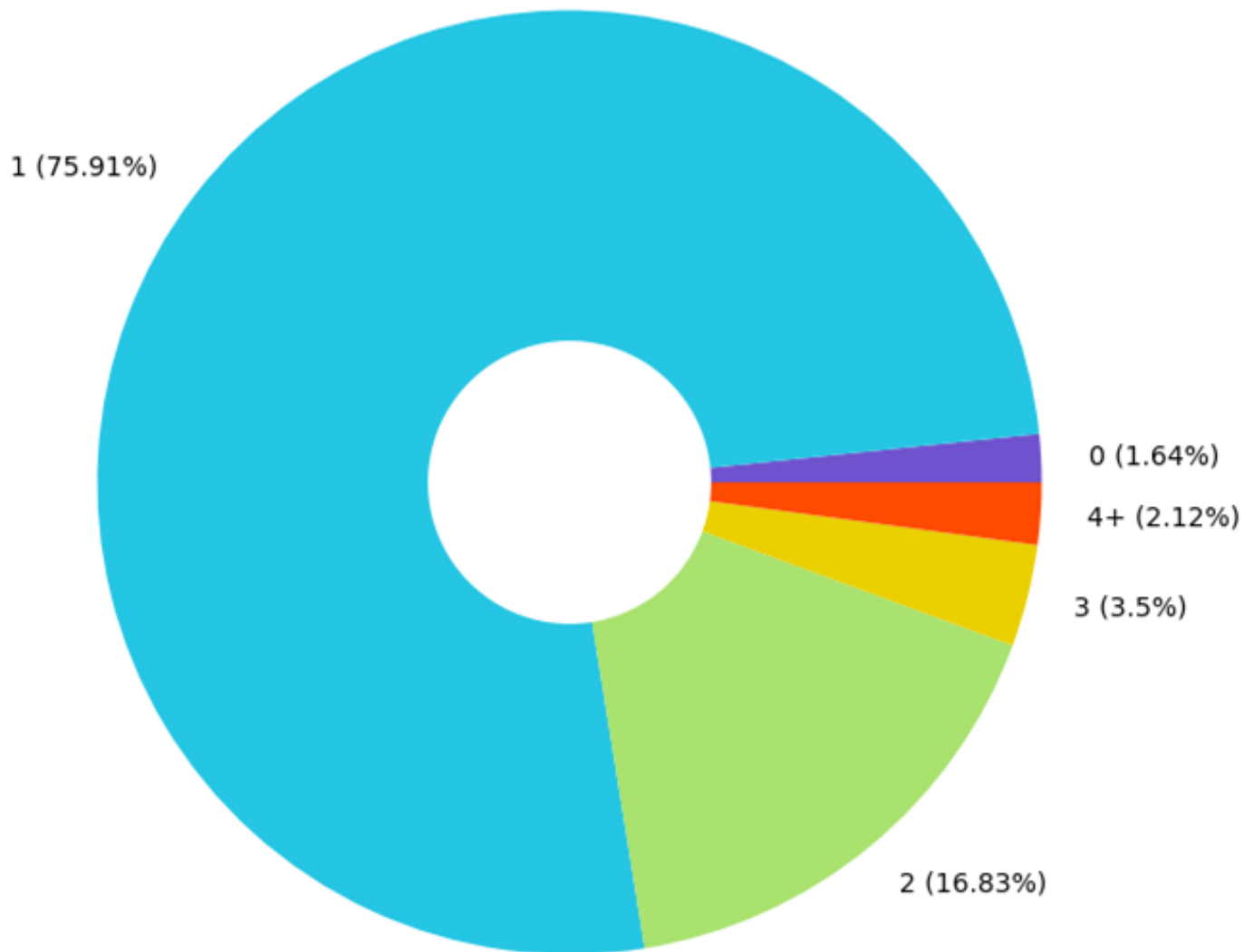
Для инвестиций (5.76%)



# Количество договоров на одного клиента

По данным за все время

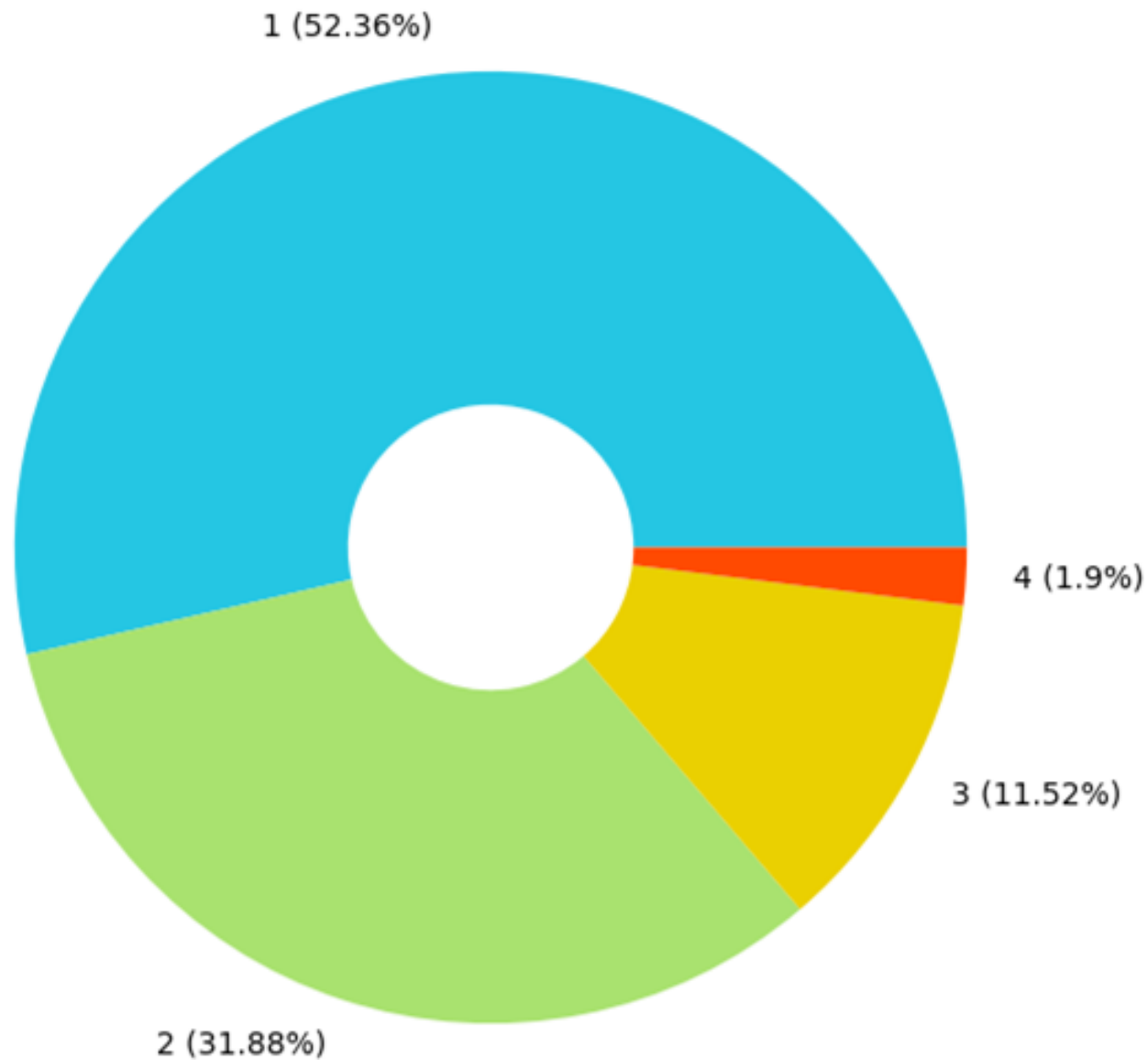
Большинство клиентов заключили только один договор, крайне мало клиентов подписали более 3 договоров или не подписали не одного.



# Количество покупок по комнатности

По данным за все время; для клиентов, купивших 1 квартиру.

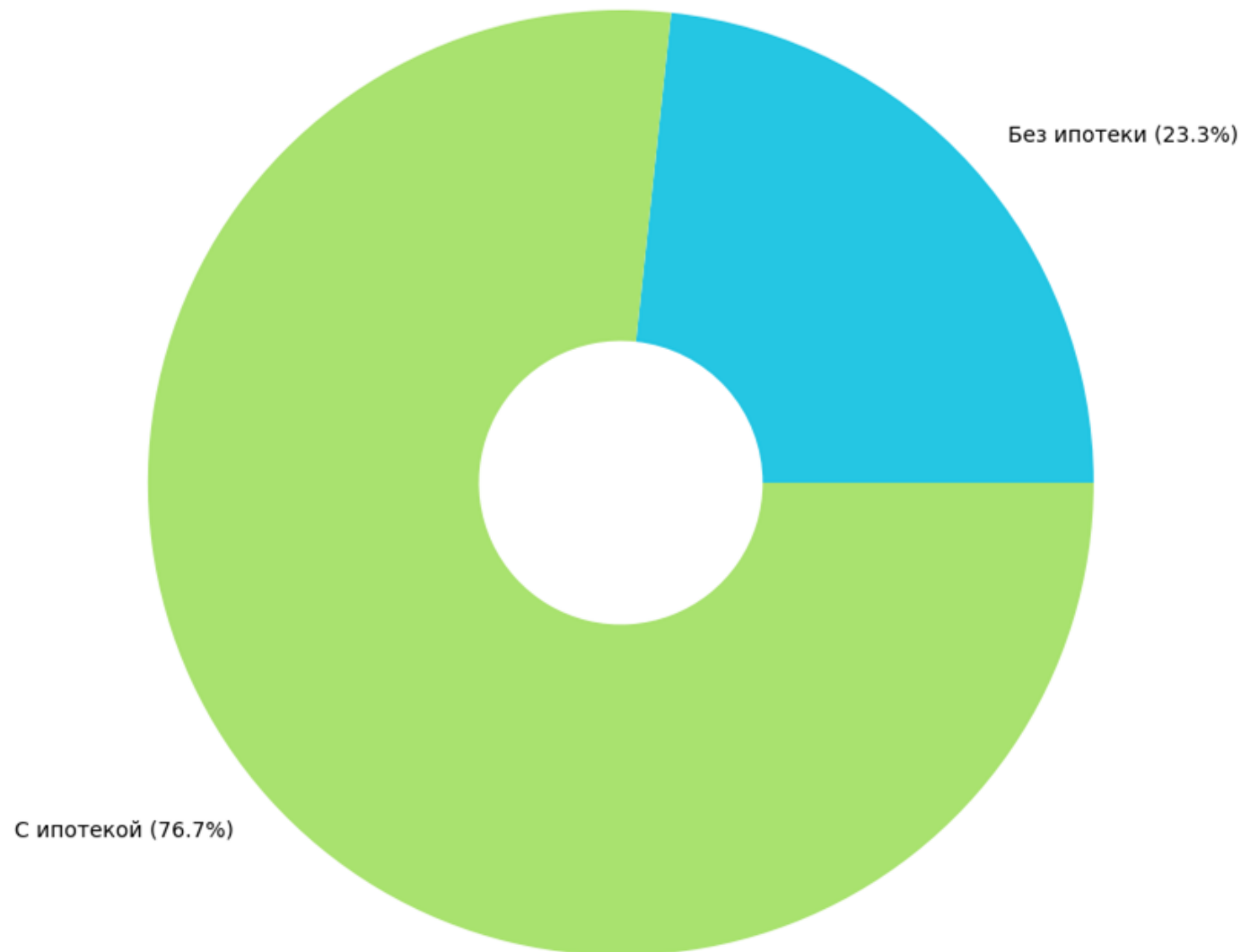
Более половины клиентов купили однокомнатную квартиру.



# Количество договоров с ипотекой и без нее

Для клиентов, купивших одну квартиру,  
по данным за все время

Большинство клиентов  
приобретают квартиру в ипотеку.





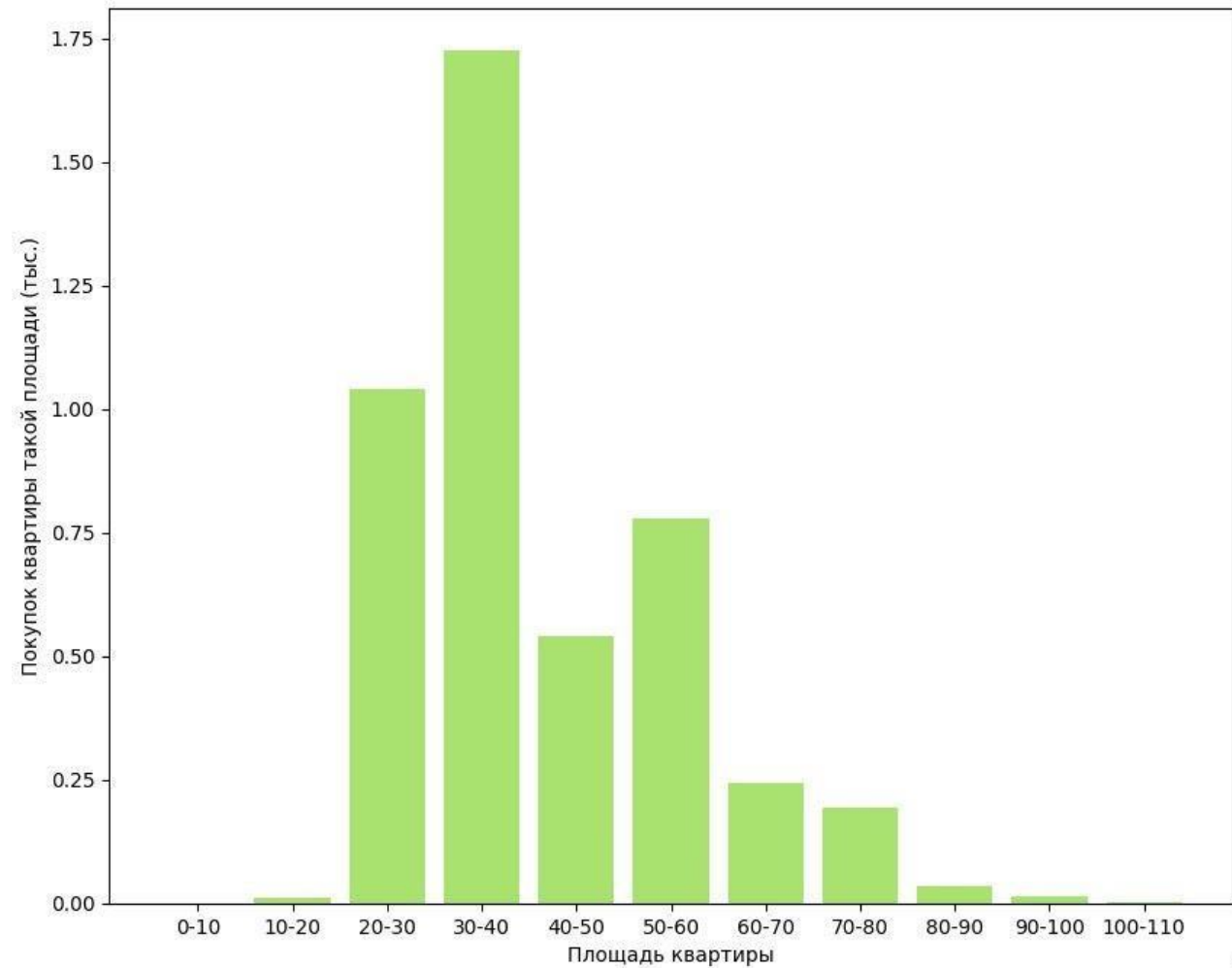
# Квартиры какой площади покупают клиенты

Для клиентов, купивших одну квартиру, по данным за все время

Квартиры небольшой площади - 20-40 кв.м. - самые популярные. Можно предположить, что это студии и однокомнатные квартиры.

Стандартный минимальный метраж — 18 кв. м на 1 человека, однако этот показатель применяют, если в квартире живут 3 и более человек. Человеку, живущему отдельно, полагается 33 кв. м, для 2 жильцов — 42 кв. М.

[Домклик](#)



# Исследовательский вопрос и гипотеза



Есть ли существенная  
разница в площади квартир  
купленных с  
использованием/без  
использования ипотеки?



В ипотеку клиенты покупают  
квартиры большей площади,  
чем без ипотеки.



# Показатели в математическом анализе

Для клиентов, купивших 1 квартиру, за все время.

## Исследуемый показатель

- Средняя общая площадь по договорам покупки квартиры:

`agreements__agreement_object_projected_total_area_mean_interval_ALL_TIME`

## Признак

- Наличие/отсутствие ипотеки (если максимально одобренная сумма ипотеки не нулевая, то считаем, что ипотека есть; если нулевая - нет).

`mortgages__mort_app_approved_amount_max_interval_ALL_TIME`

# Математический анализ

- $p$  - уровень значимости 40%
- средняя площадь квартир, которые берут в ипотеку 40.487 кв.м.
- средняя площадь квартир без ипотеки 40.834 кв.м.

Большой  $p$  - уровень значимости означает, что разница в выборках не статистически значимая. Средние площади квартир купленных в ипотеку и без привлечения ипотеки практически не отличаются.



# Показатели в математическом анализе

Для клиентов, купивших 1 квартиру, за все время.

## Исследуемый показатель

- Средняя общая площадь по договорам покупки квартиры:

`agreements__agreement_object_projected_total_area_mean_interval_ALL_TIME`

## Признаки

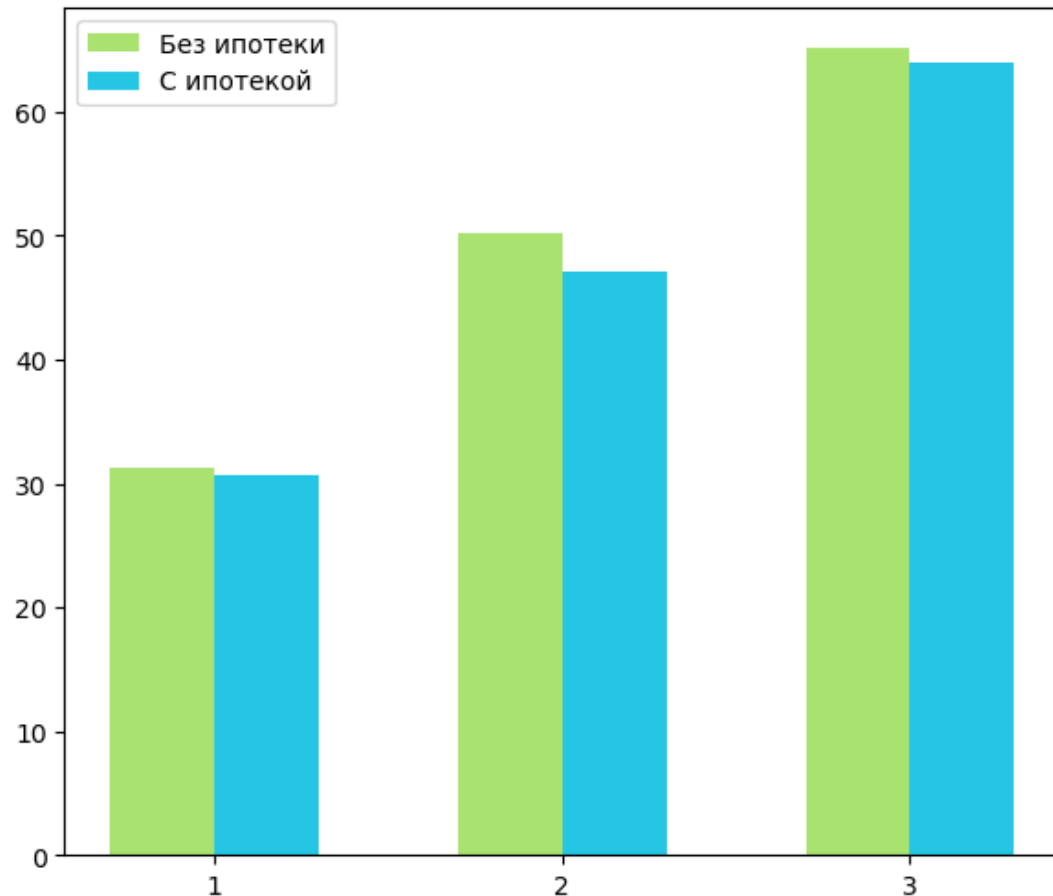
- Наличие/отсутствие ипотеки (если максимально одобренная сумма ипотеки не нулевая, то считаем, что ипотека есть; если нулевая - нет).

`mortgages__mort_app_approved_amount_max_interval_ALL_TIME`

- Среднее количество комнат в договорах (т.к. договор только один, то это количество комнат в купленной квартире)

`agreements__agreement_object_room_quantity_mean_interval_ALL_TIME`

# Средняя площадь в разбивке по типам квартир



## Размеры выборок

	1 КОМН.	2 КОМН.	3 КОМН.	4 КОМН.
С ипотекой	2428	1457	502	82
Без ипотеки	1044	530	142	21

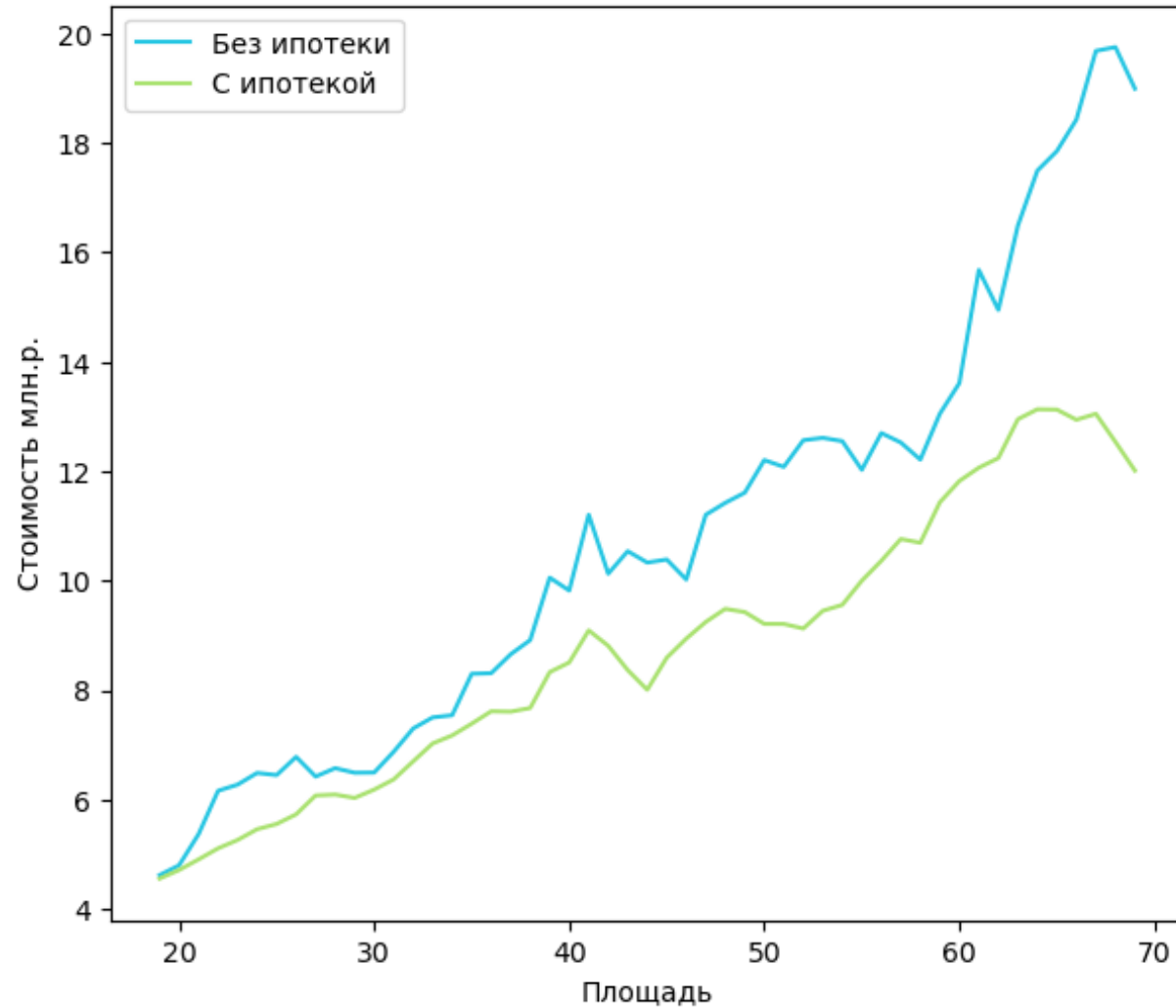
	1 КОМН.	2 КОМН.	3 КОМН.
р - уровень значимости	0.014	≈ 0.0*	0.27
На сколько % больше средняя общая площадь квартир без ипотеки, чем с ипотекой	2 %	6 %	1.5 %

\*6.11 e-11

Для 1 и 2-комнатных квартир разница в площади является статистически значимой, для 3-комнатных нет.

# Средняя стоимость квартиры от её площади

При равной площади люди, покупающие квартиру в ипотеку, выбирают вариант дешевле.



# Интерпретация результатов

---

Россияне с недоверием относятся к ипотеке, их не устраивают высокие проценты. Также люди считают ипотеку долгим обязательством.

Поэтому те, кто покупают квартиру в ипотеку, стараются выбрать вариант подешевле, чтобы минимизировать «потери».



Просто в голове не укладывается: подарить столько денег банку



Когда будет возможность у всей страны брать ипотеку под 2%, тогда это и станет выгодно. А пока даже ставка 10% — это сумасшедшая переплата, особенно на горизонте 20—30 лет.

<https://journal.tinkoff.ru/bad-ipoteka/>



Клиенты, приобретающие недвижимость без ипотеки, могут позволить себе квартиру такой же или чуть большей площади, чем клиенты-ипотечники.

Возможно, что значительная часть клиентов-ипотечников ограничена в средствах, и поэтому чувствительна к цене. Для них не актуально купить большой метраж, и, соответственно, взять большую сумму ипотеки.

# Выводы

Гипотеза не подтвердилась и в общем, и для отдельных типов квартир

## Ограничения

- Нет детальных данных по характеристикам квартир. Нельзя представить полную картину, так как эти данные тоже влияют на клиента.
- Нет данных по географии объектов, в разных регионах могут быть разные тенденции.
- Нет развернутых данных по ипотеке (размер первоначального взноса, наличие льготных условий, процент) для более подробного исследования. Для разной ипотеки могут быть разные соотношения.

## Перспективы

- Для более детального понимания интересов клиентов, как пользующихся ипотекой, так и нет, нужно расширить объем данных для исследования:
  - подробные данные о клиенте:
    - возраст,
    - пол,
    - семейное положение.
  - более подробная характеристика квартир:
    - наличие мебели,
    - высота потолков,
    - вид отделки.



# Policy implication

Для того чтобы скомпенсировать неудобства от небольшой площади можно предложить:

- варианты интересных дизайн-проектов для маленьких квартир,
- дополнительные места для хранения колясок, шин и др. вещей,
- дополнительные зоны для работы и досуга (коворкинги, зоны для отдыха на крыше, общие игровые комнаты для детей).
- услуги по помощи с переездом.





**Спасибо за внимание!**





# Приложение

Подготовка данных и преданализ

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
df = pd.read_csv('samolet_dataset.csv')
```

```
df['report_date'] = pd.to_datetime(df['report_date'])
```

```
df = df.fillna(0)
```

```
df = df.loc[df.groupby('client_id')['report_date'].idxmax()]
```

```
for i in list(df.columns):
    if list(df[i].unique()) == [0]:
        df = df.drop(columns = [i], axis = 1)
```

```
for i in sorted(list(df.columns)[3:]):
    l = list(df[i].unique())
    print(i, (100 - len(l)) * "-", "> ", min(l), " ", sum(l) / len(l), " ", max(l), "\n")
```

# Приложение

Подготовка данных и построение графика по количеству покупок

```
for i in list(df.columns):  
    if ("qty" in i):  
        df = df.astype({i: "Int64"})
```

```
l = df["agreements__agreements_sales__qty__interval_ALL_TIME"].value_counts()  
l = l.sort_index()  
colors = ["#6F53CE", "#25C6E3", "#A9E26F", "#EAD001", "#FF4A01"]  
  
l2 = list(l)[:4] + [sum(list(l)[4:])]  
  
s = sum(l)  
labels = list(l.index)[:4] + ["4+"]  
  
for i in range(len(labels)):  
    labels[i] = str(labels[i]) + " (" + str(round(l2[i] / s * 100, 2)) + "%)"  
  
plt.figure(figsize = (8, 8))  
plt.pie(l2, labels = labels, colors = colors, wedgeprops = dict(width = 0.7))
```

# Приложение

Построение графика по целевым интересам

```
x = [sum(df["specific_picks__sp_acquisition_myself__qty__interval_0_90"]),
      sum(df["specific_picks__sp_acquisition_investment__qty__interval_0_90"]),
      sum(df["specific_picks__sp_acquisition_others__qty__interval_0_90"])
     ]

s = sum(x)

la = ["Для себя (" + str(round(x[0] / s * 100, 2)) + "%)",
      "Для инвестиций (" + str(round(x[1] / s * 100, 2)) + "%)",
      "Для других (" + str(round(x[2] / s * 100, 2)) + "%)"]

colors = ["#A9E26F", "#EAD001", "#FF4A01"]

plt.figure(figsize = (8, 8))
plt.pie(x, labels = la, colors = colors, wedgeprops = dict(width = 0.7))
```

# Приложение

Построение графика корреляции

```
plt.figure(figsize = (9, 18))

plt.xlabel("Целевых интересов")
plt.ylabel("Кол-во договоров на продажу")

plt.scatter(
    df["specific_picks__specific_picks__qty__interval_0_365"],
    df["agreements__agreements_sales__qty__interval_0_365"],
    color = "#25C6E307",
    s = 300
)
```

# Приложение

Построение графика количества квартир, купленных с ипотекой и без

```
k = [len(df[(df["agreements__agreements_sales__qty__interval_ALL_TIME"] == 1) &
           (df["mortgages__mort_app_approved_amount__max__interval_ALL_TIME"] == 0)]),
      len(df[(df["agreements__agreements_sales__qty__interval_ALL_TIME"] == 1) &
           (df["mortgages__mort_app_approved_amount__max__interval_ALL_TIME"] != 0)])
    ]

colors = ["#25C6E3", "#A9E26F"]

s = sum(k)

plt.figure(figsize = (10, 10))
plt.pie(k, labels = ["Не ипотека (" + str(round(k[0] / s * 100, 2)) + "%)",
                    "Без ипотеки (" + str(round(k[1] / s * 100, 2)) + "%)"],
        colors = colors, wedgeprops = dict(width = 0.7))
```

# Приложение

Построение графика распределения площади купленных квартир

```
df1 = df[(df["agreements__agreements_sales__qty__interval_ALL_TIME"] == 1)]

y = []
la = []

for i in range(0, int(max(df1["agreements__agreement_object_projected_total_area__mean__interval_ALL_TIME"])), 10):
    df2 = df1[(df1["agreements__agreement_object_projected_total_area__mean__interval_ALL_TIME"] > i) &
              (df1["agreements__agreement_object_projected_total_area__mean__interval_ALL_TIME"] <= i + 10)]
    y.append(len(df2))
    la.append(str(i) + "-" + str(i + 10))

plt.figure(figsize = (10, 7))
plt.bar(la, y, color = "#A9E26F")
```



# Приложение

Мат. анализ

```
import scipy.stats as stats

df1 = df[(df["agreements__agreements_sales__qty__interval_ALL_TIME"] == 1) &
         (df["mortgages__mort_app_approved_amount__max__interval_ALL_TIME"] == 0)]

df1 = df1["agreements__agreement_object_projected_total_area__mean__interval_ALL_TIME"]

df2 = df[(df["agreements__agreements_sales__qty__interval_ALL_TIME"] == 1) &
         (df["mortgages__mort_app_approved_amount__max__interval_ALL_TIME"] != 0)]

df2 = df2["agreements__agreement_object_projected_total_area__mean__interval_ALL_TIME"]

print(np.var(df1), np.var(df2))

print(stats.ttest_ind(df1, df2, equal_var = True))

print(sum(df1) / len(df1), sum(df2) / len(df2))
```