

Анализ данных клиентов банка по оформлению депозитов в Google Colaboratory

Анишкова Анастасия Сергеевна

Приамурский государственный университет имени Шолом-Алейхема

Студент

Аннотация

Целью исследования является анализ данных клиентов банка для выявления процента оформленных депозитов. Для реализации использовался облачный сервис Google Colaboratory. Результат анализа можно использовать для улучшения системы банка привлечения клиентов для оформления депозита.

Ключевые слова: Google Colaboratory, депозит, банк.

Analysis of the bank's customer data on the purchase of deposits in Google Colaboratory

Anishkova Anastasia Sergeevna

Sholom-Aleichem Priamurskiy State University

Student

Abstract

The purpose of the study is to analyze the bank's customer data to identify the percentage of deposits made. The Google Colaboratory cloud service was used for implementation. The result of the analysis can be used to improve the bank's system of attracting customers for making a deposit.

Key words: Google Colaboratory, deposit, bank.

1 Введение

1.1 Актуальность

Анализ поможет банку понять, почему одни клиенты оформляют депозиты, а другие нет. Это может быть связано с различными факторами, такими как уровень дохода, финансовые цели, уровень образования, степень доверия к банку и т.д. Понимание этих причин поможет банку разработать более эффективные стратегии привлечения новых клиентов и удержания существующих.

1.2 Обзор исследований

Э.Бисонг, Э.Бисонг описал как работать с библиотеками Matplotlib и Seaborn [1], визуализировал статистических данных при помощи библиотеки Seaborn М. Л. Васком [2]. С. В. Шайтура провел интеллектуальный анализ данных В частности, описан процесс интеллектуального анализа данных в трёх стадиях: выявление закономерностей; использование выявленных

закономерностей для предсказания неизвестных значений; анализ исключений, предназначенный для выявления и толкования аномалий в найденных закономерностях [3]. Качественный анализ данных сделали А. Е. Войскунский, С. В. Скрипкин [4]. Е. А. Григорьев, Н. С. Климов привели пример разведочного анализа данных с помощью python [5].

1.3 Цель исследования

Целью исследования является анализ данных клиентов, оформивших депозит в Google Colaboratory.

2 Материалы и методы

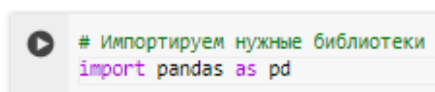
В данном исследовании используется облачный сервис Google Colaboratory.

3 Результаты

Ссылка на ноутбук
<https://colab.research.google.com/drive/1qKpFdJUQTT3fQjrThMPnaj9DE3YQaF43?usp=sharing>

Подключим библиотеки для осуществления анализа данных (см. рис.1).

Ссылка на датасет <https://cloud.mail.ru/public/CtPj/maw1ta3cR>.



```
# Импортируем нужные библиотеки
import pandas as pd
```

Рисунок 1 – Подключение библиотеки

Подгрузим и считаем датасет с данными по банку и вывести первые 10 строк с таблицы (см. рис.2).



```
# Загружаем датасет
df = pd.read_csv('bank-full.csv', sep=';')
df.sample(n=10)
```

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	plays	previous	postcode	y
42907	34	management	married	tertiary	no	1776	no	no	cellular	5	feb	89	1	33	3	success	no
16203	58	housemaid	married	primary	no	1399	yes	yes	cellular	22	jul	1226	2	-1	0	unknown	no
13755	47	management	married	secondary	no	894	yes	no	cellular	10	jul	185	1	-1	0	unknown	no
40696	37	admin	single	secondary	no	115	no	no	cellular	4	aug	236	1	82	4	failure	no
42652	28	unemployed	married	secondary	no	687	no	no	cellular	13	jan	113	-3	33	2	other	no
15704	50	technician	married	secondary	no	-78	yes	yes	cellular	21	jul	357	1	-1	0	unknown	no
16228	48	retired	married	tertiary	no	4	no	yes	cellular	25	jul	79	5	-1	0	unknown	no
13437	45	technician	married	secondary	no	2020	yes	no	cellular	9	jul	85	1	-1	0	unknown	no
14903	36	technician	married	secondary	no	217	no	yes	cellular	16	jul	93	3	-1	0	unknown	no
37134	33	technician	single	tertiary	no	2585	yes	no	cellular	13	may	128	2	-1	0	unknown	no

Рисунок 2 – Считывание данных

Далее построим графики. Первый график будет о том, сколько клиентов в разные месяцы. Для этого необходимо перевести дату в требуемый формат (см. рис. 3).

```
[176] df['month']
0      may
1      may
2      may
3      may
4      may
...
45206  nov
45207  nov
45208  nov
45209  nov
45210  nov
Name: month, Length: 45211, dtype: object
```

```
# Меняем тип переменной
df['month'] = pd.to_datetime(df['month'], format='%b').dt.month
df['month']
```

```
0      5
1      5
2      5
3      5
4      5
..
45206 11
45207 11
45208 11
45209 11
45210 11
Name: month, Length: 45211, dtype: int64
```

Рисунок 3 – Преобразование даты в нужный формат

Сгруппируем данные по месяцу и выявим сколько было клиентов в это время (см. рис. 4).

```
# Готовим данные для графика в виде сводной таблицы
data = df.groupby('age')['convert'].sum().reset_index()
data.head()
```

	age	convert
0	18	7
1	19	11
2	20	15
3	21	22
4	22	40

Рисунок 4 – группировка данных

В результате получили сводную таблицу, которая подсчитала количество клиентов. После проведения данной манипуляции изобразим данные в графике (см. рис.5).



Рисунок 5 – график охвата клиентов

Так же клиенты банка можно отсортировать по семейному положению (см. рис.6).

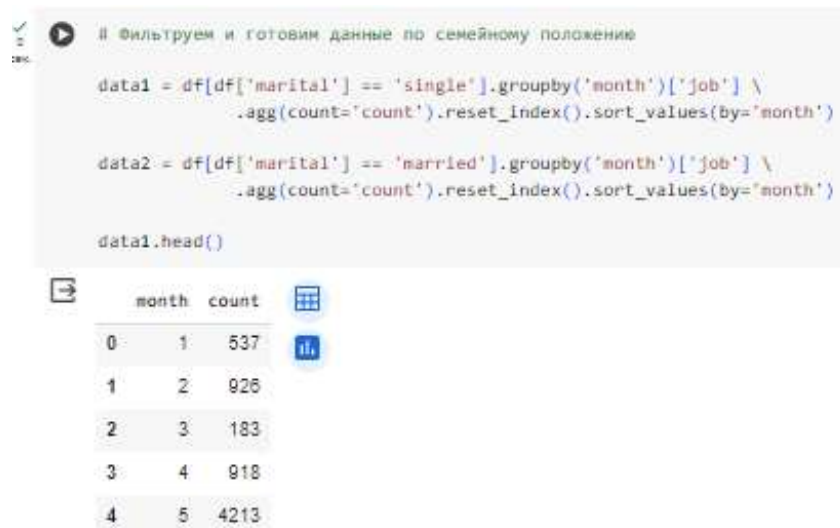


Рисунок 6 – Сводная таблица по семейному положению

На основании сводной таблицы необходимо построить график (см. рис.7).

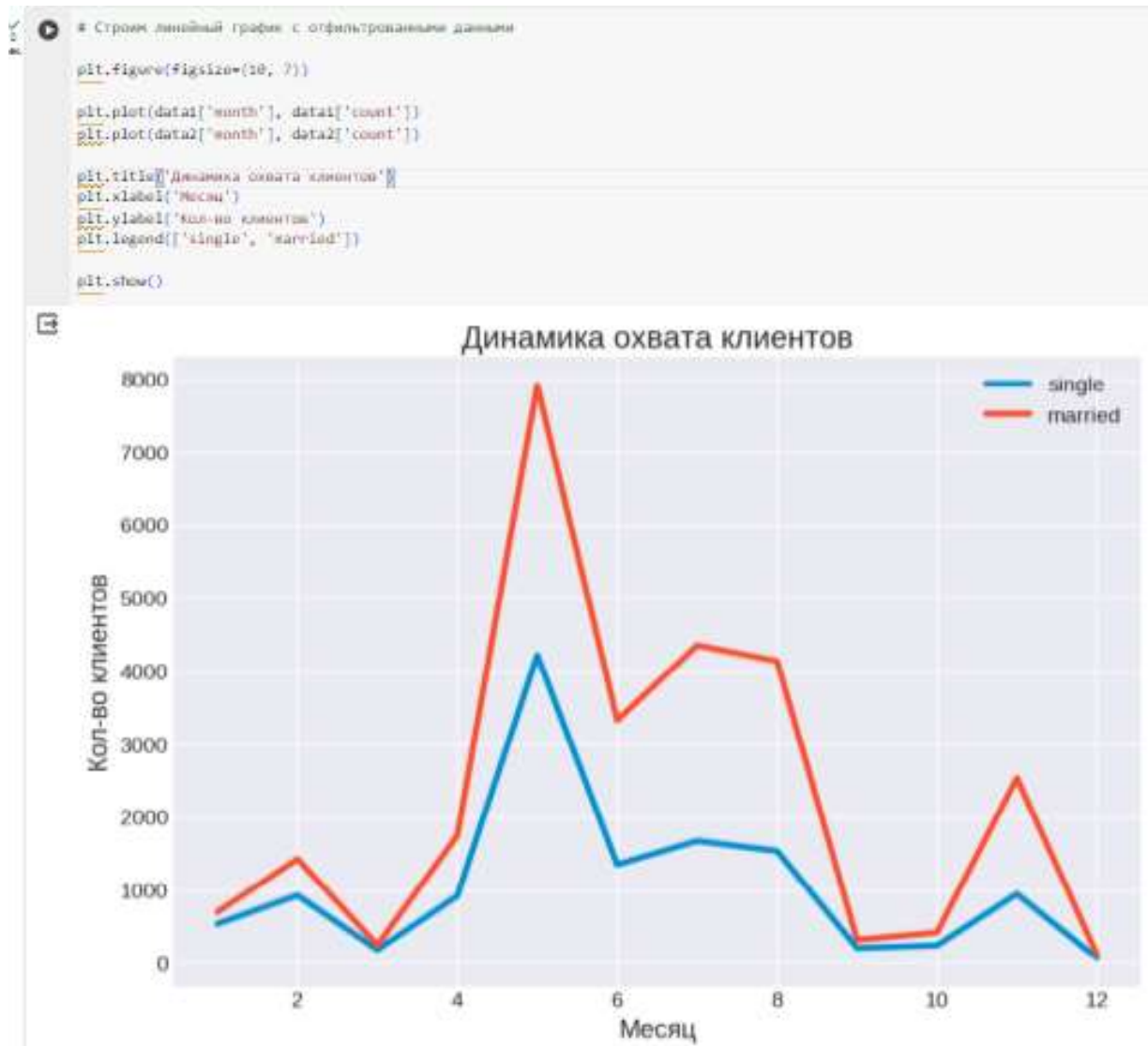


Рисунок 7 – Динамика охвата клиентов семейных и свободных

Далее изобразим возвратной диапазон клиентов (см. рис.8).

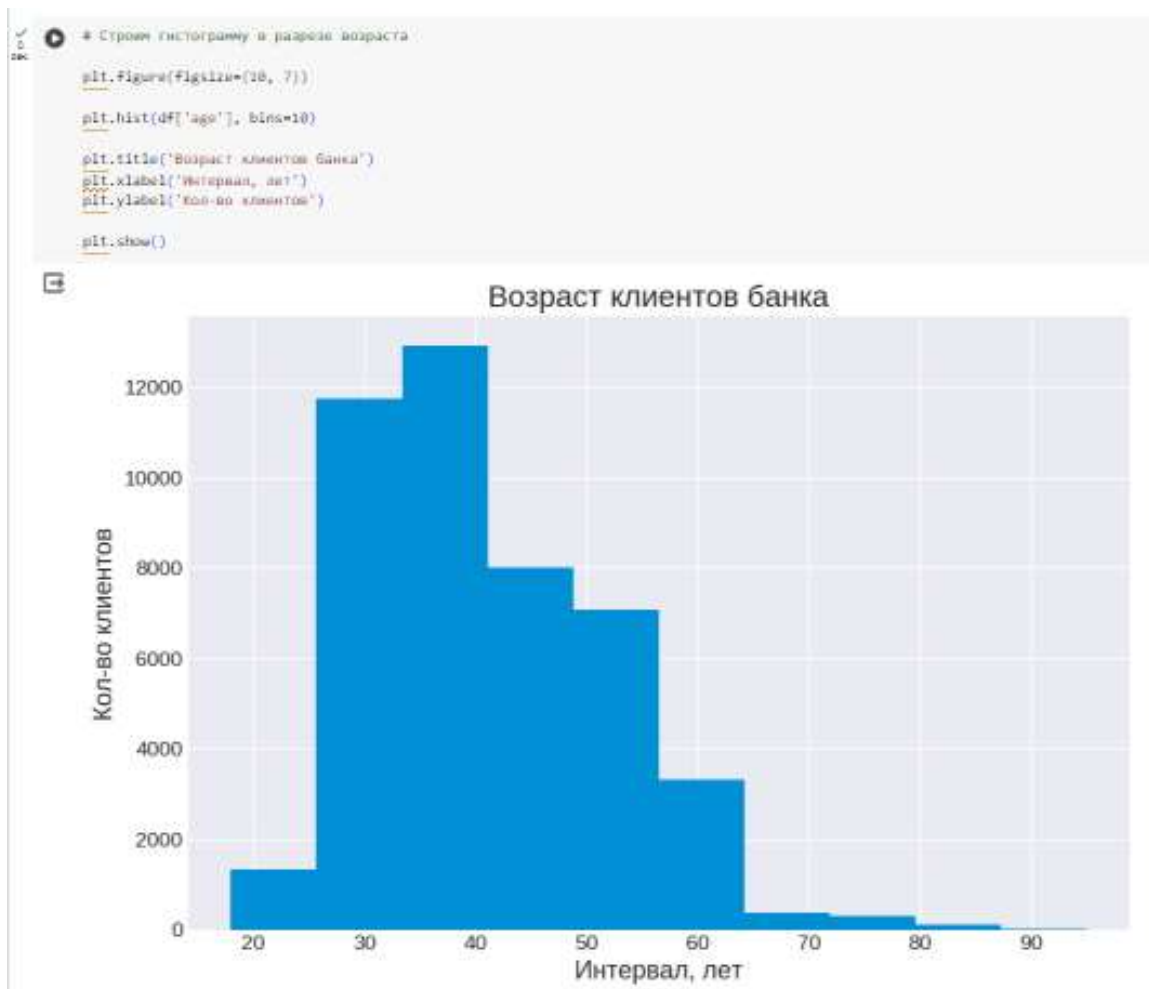


Рисунок 8 – Возвратной диапазон клиентов

Далее изобразим клиентов по целевому признаку, то есть стал клиент владельцем вклада или нет (см. рис.9).

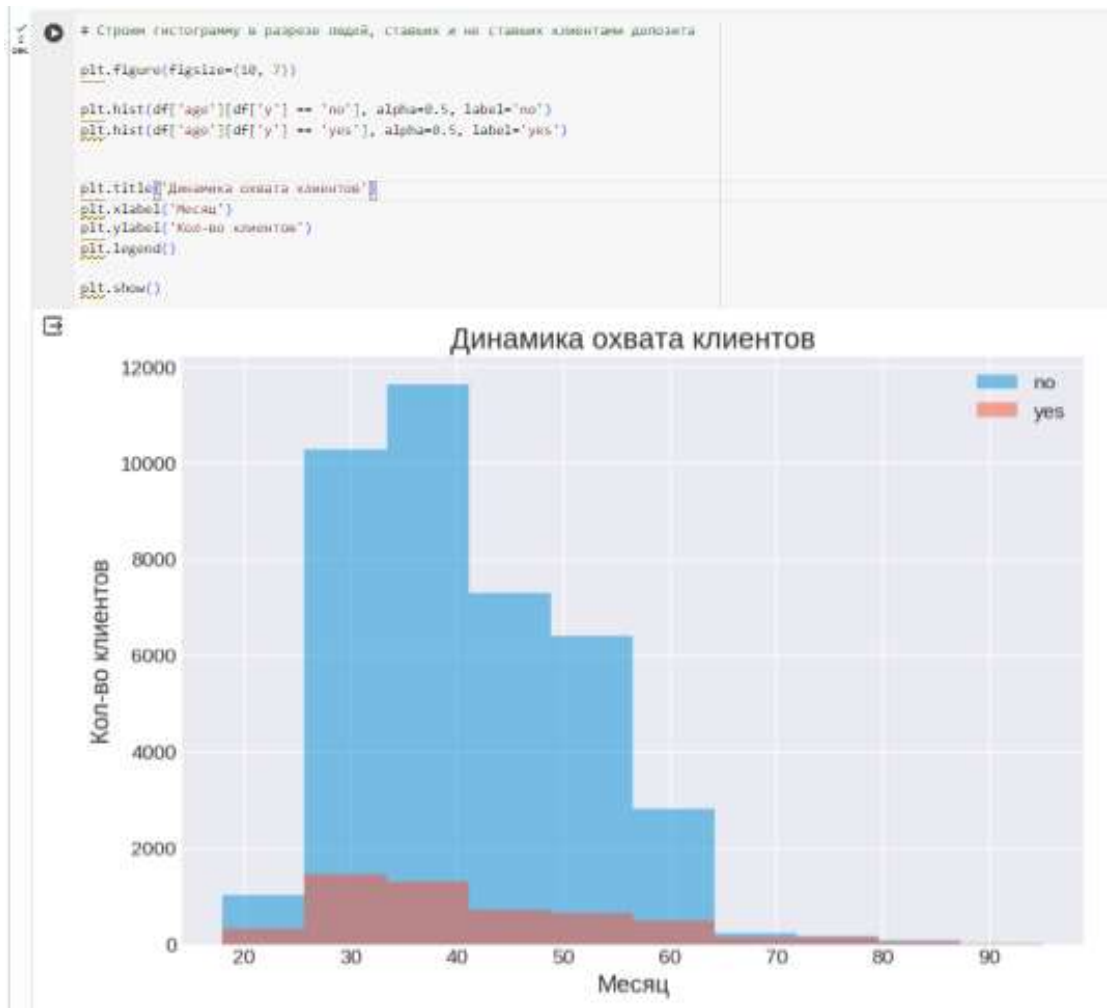


Рисунок 9 – Гистограмма по целевому признаку

Рекомендуется перевести целевой признак в бинарный. Да - 1, нет – 0 (см. рис. 10).



Рисунок 10 – Перевод целевого признака в бинарный и сводная таблица

На основе сводной таблицы посмотрим точечный график (см. рис.11).

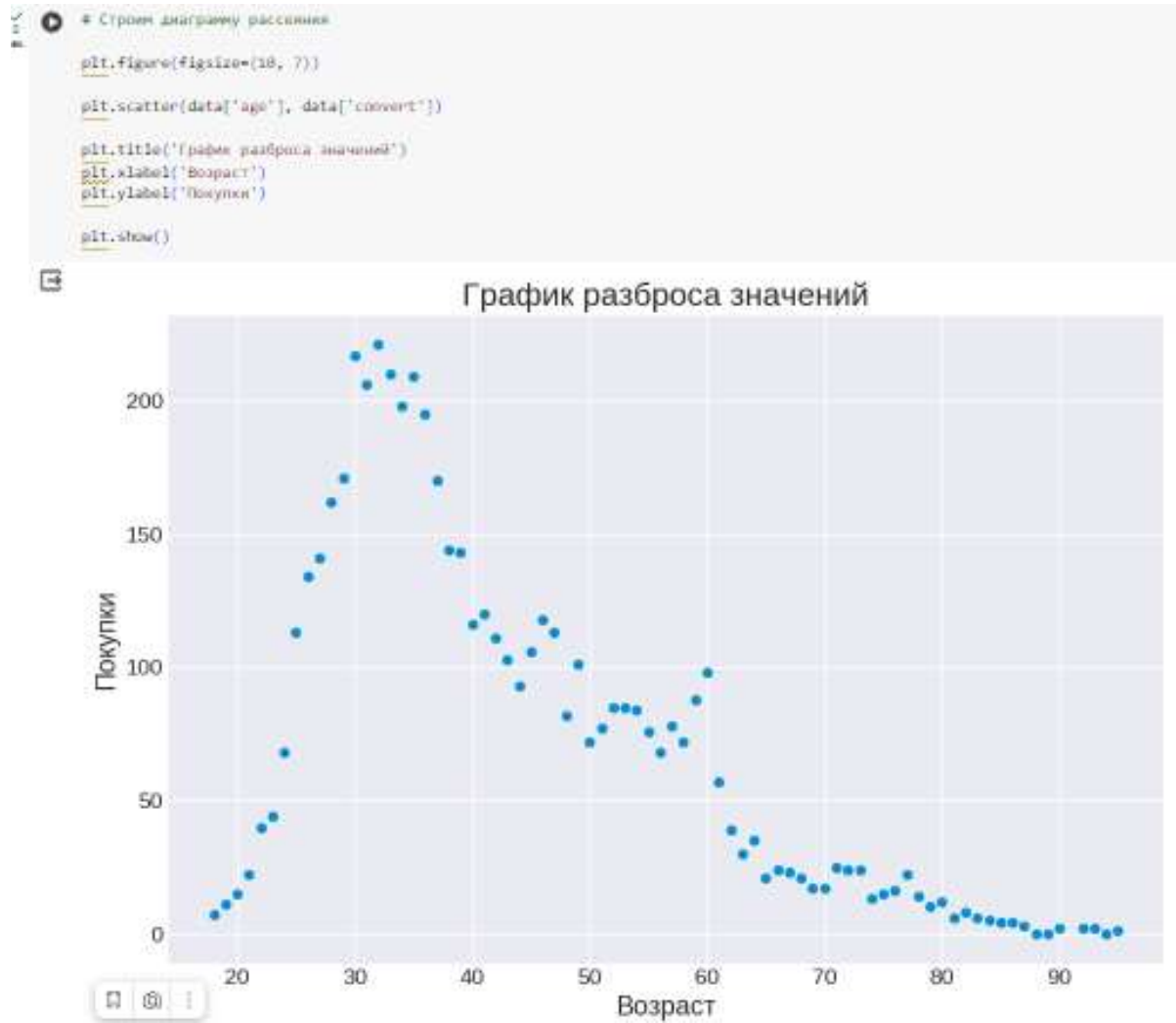


Рисунок 11 – График разброса значений

Изобразим сколько секунд продлился звонок для предложений колл-центра с клиентом (см. рис.12).

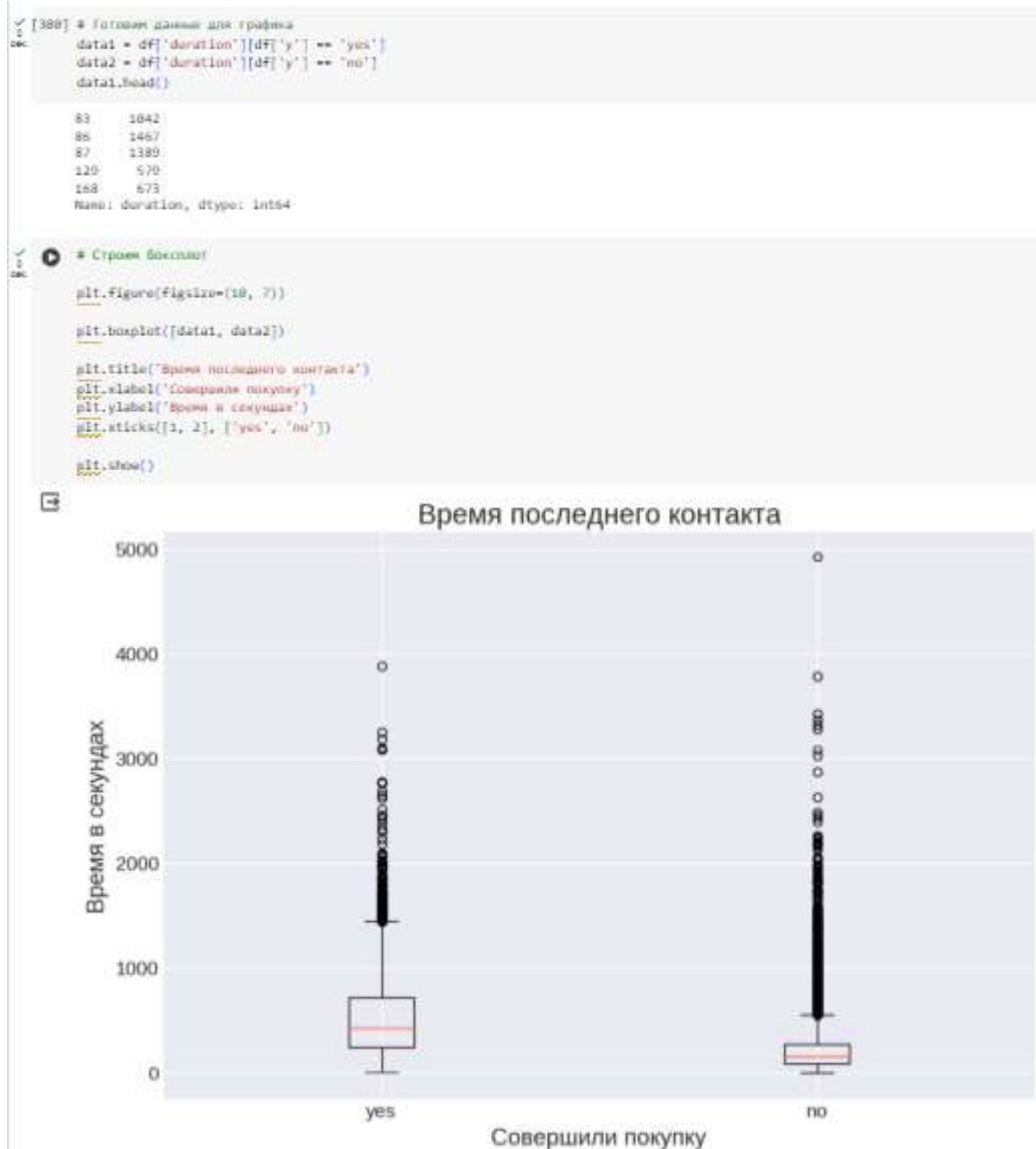


Рисунок 12 – график работы кол-центра

Построим график без выбросов для улучшения визуального восприятия (см. рис.13).

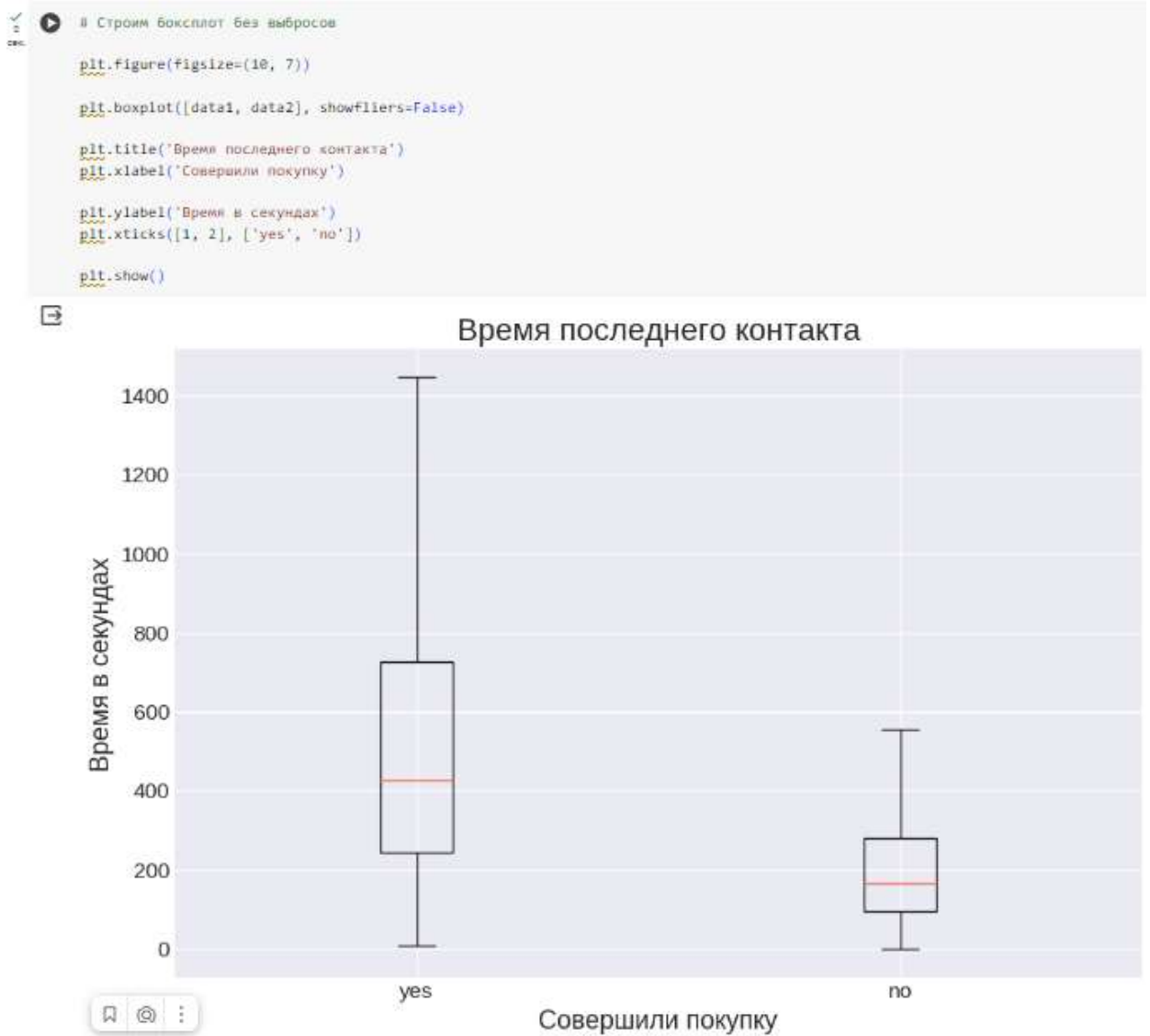


Рисунок 13 – Улучшенный график работы кол-центра

Далее изобразим род деятельности клиентов банка, для этого подготовим сводную таблицу и изобразим результат на графике (см. рис.14).

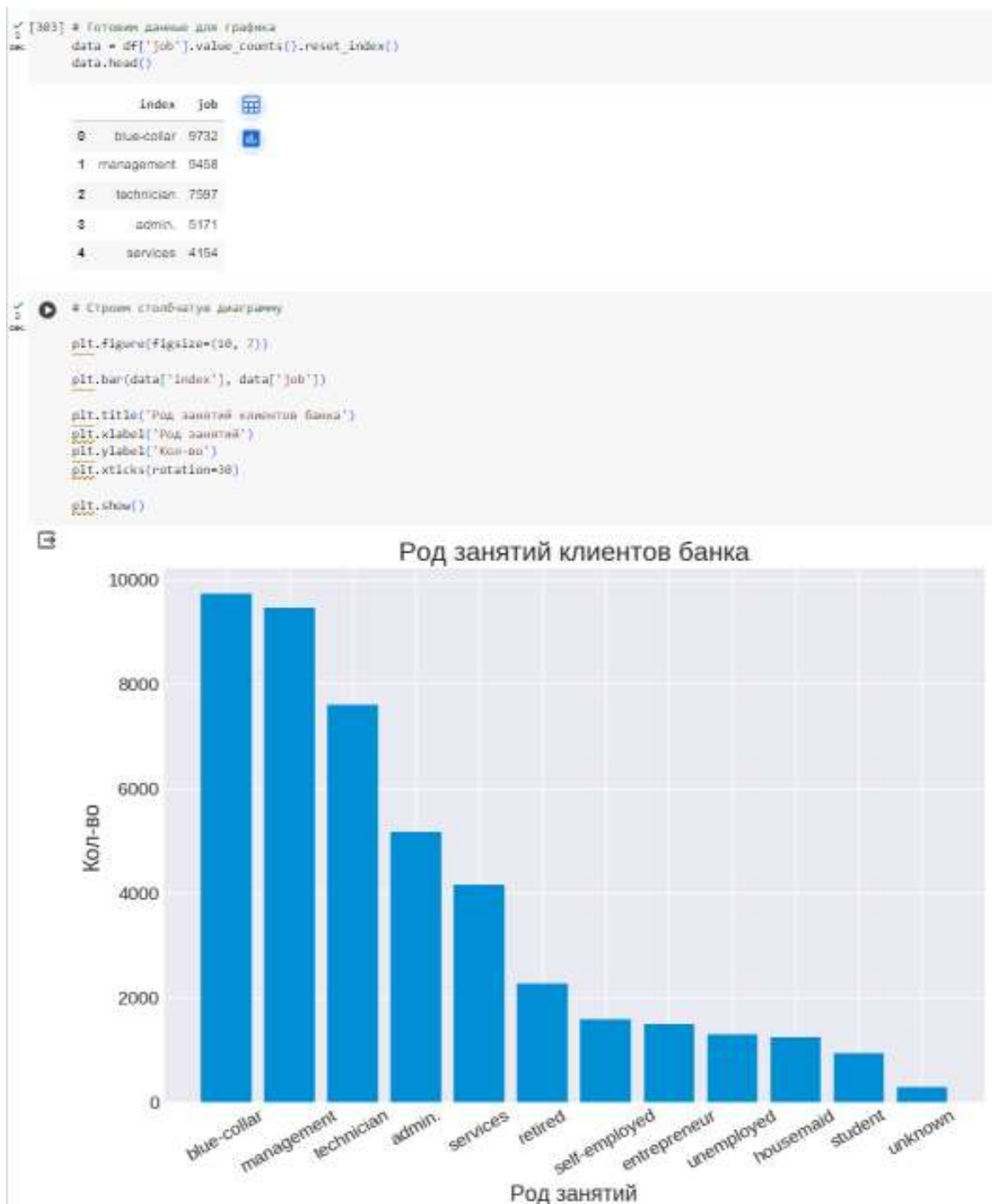


Рисунок 14 – Род деятельности клиентов

Далее отобразим на этом же графике кто приобрел из них вклад (см. рис.15).

```
[ ] # Готовим данные для графика
data = pd.crosstab(df['job'], df['y']).reset_index().sort_values(by='no', ascending=False)
data.head()
```

y	job	no	yes
1	blue-collar	8024	708
4	management	8157	1301
8	technician	6757	840
0	admin.	4540	531
7	services	3785	359

```
# Строим столбчатую многорядную диаграмму

plt.figure(figsize=(10, 7))

# определяем кол-во делений
n_ticks = np.arange(len(data['job']))

# определяем сдвиг
offset = 0.2

# определяем ширину столбцов
w = 0.4

# добавляем сдвиг к кол-ву делений
plt.bar(n_ticks - offset, data['no'], width=w)
plt.bar(n_ticks + offset, data['yes'], width=w)

plt.title('Род занятий клиентов банка, открывших депозит')
plt.xlabel('Род занятий')
plt.ylabel('кол-во')
plt.legend(['no', 'yes'])
plt.xticks(n_ticks, data['job'], rotation = 30) # добавляем метки делений

plt.show()
```

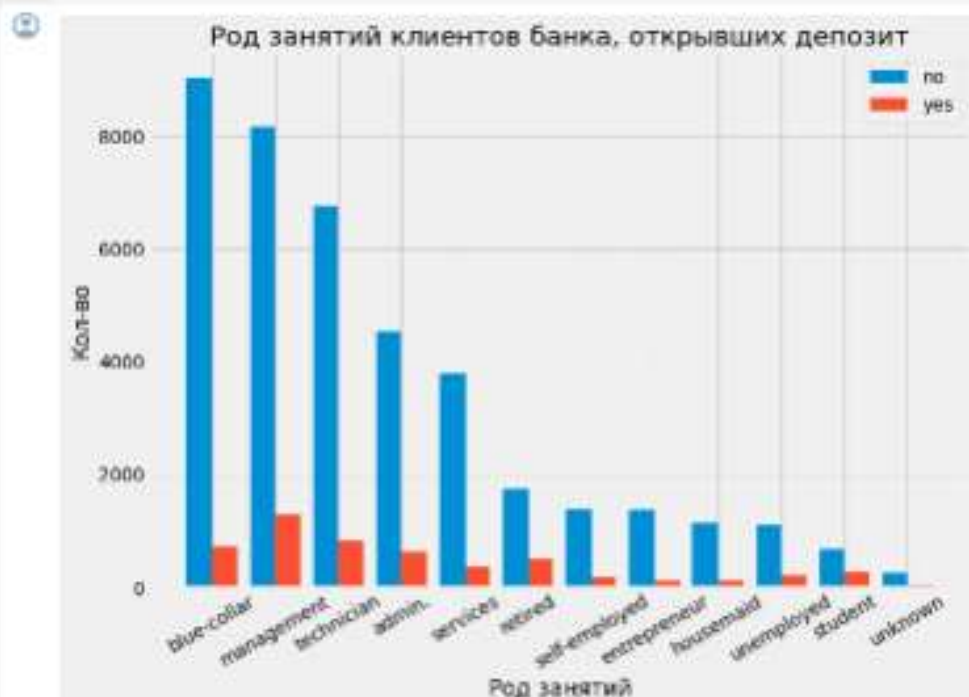


Рисунок 15 – Род деятельности клиентов и наличие вклада

Для наглядности изобразим график с помощью многорядной сложной диаграммой (см. рис. 16). На графике видно, что большую часть клиентов

банка составляет голубые-воротнички и количество депозитов оформлено в их категории больше остальных.

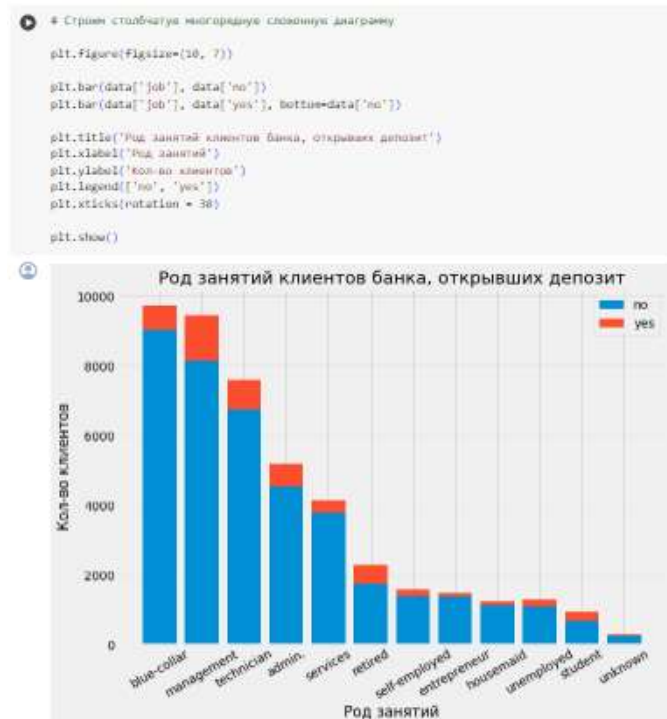


Рисунок 16 – Род деятельности клиентов банка, открывших вклад

Но так как количество клиентов не равное, то необходимо перевести в процентное соотношение количество клиентов как часто брали вклад и не брали (см. рис.17).

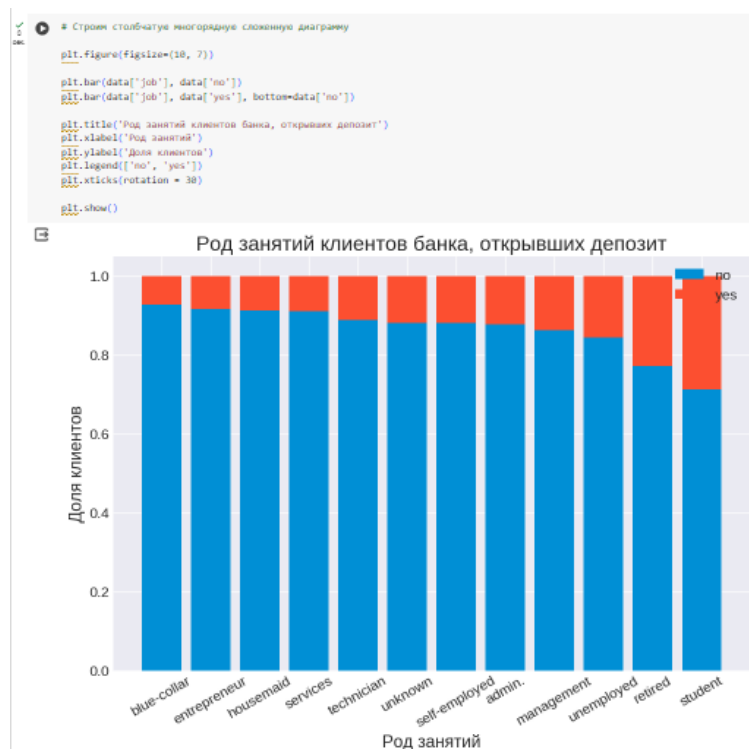


Рисунок 17 – График в процентном соотношении

Последний график содержит в себе информацию о количестве человек клиентов открывший вклад (см. рис. 18).

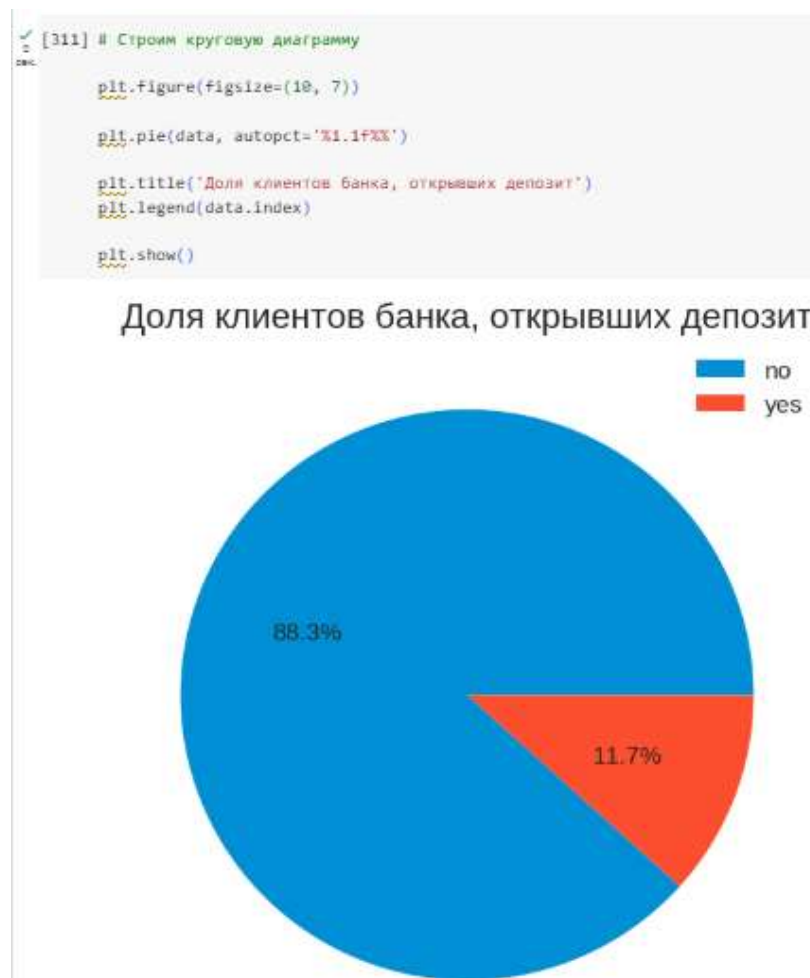


Рисунок 18 – Доля клиентов, открывших вклад

Выводы

На основе анализа в процентном отношении можно делать выводы, что большую часть клиентов банка – это синие-воротнички, но студенты чаще всего оформляют депозит и они являются той аудиторией, на которую нужно делать акцент.

Библиографический список

1. Бисонг Э., Бисонг Э. Matplotlib и Seaborn // Построение моделей машинного обучения и глубокого обучения на облачной платформе Google Cloud Platform: полное руководство для начинающих. 2019. С. 151-165.
2. Васком М. Л. Seaborn: визуализация статистических данных // Журнал программного обеспечения с открытым исходным кодом. 2021. Т. 6. №. 60. С. 3021.
3. Шайтура С. В. Интеллектуальный анализ данных //Славянский форум. 2015. №. 2. С. 341-350.
4. Войсунский А. Е., Скрипкин С. В. Качественный анализ данных //Вестник

Московского университета. Серия. 2001. Т. 14. С. 93-109.

5. Григорьев Е. А., Климов Н. С. Разведочный анализ данных с помощью python //E-Scio. 2020. №. 2 (41). С. 165-176.