

Анализ данных оттока клиентов в банке при помощи библиотеки Seaborn в Google Colaboratory

Анишкова Анастасия Сергеевна

Приамурский государственный университет имени Шолом-Алейхема

Студент

Аннотация

Целью исследования является анализ данных оттока клиентов в банке при помощи библиотеки Seaborn. Для реализации использовался облачный сервис Google Colaboratory. Полученный результат использовался для определения количества оттока клиентов.

Ключевые слова: Google Colaboratory, отток клиентов, график, гистограмма, seaborn.

Analysis of customer outflow data in the bank using the Seaborn library in Google Colaboratory

Anishkova Anastasia Sergeevna

Sholom-Aleichem Priamurskiy State University

Student

Abstract

The purpose of the study is to analyze customer churn data in the bank using the Seaborn library. The Google Colaboratory cloud service was used for implementation. The result obtained was used to determine the amount of customer churn.

Key words: Google Colaboratory, customer churn, graph, histogram, seaborn.

1 Введение

1.1 Актуальность

Анализ данных с помощью библиотеки Seaborn является очень актуальным, поскольку Seaborn предоставляет мощные инструменты для визуализации данных и статистического анализа. Google Colaboratory — это не так давно появившийся облачный сервис, направленный на упрощение исследований в области машинного и глубокого обучения.

1.2 Обзор исследований

Э. Бисонг, Э. Бисонг продемонстрировали как осуществлять анализ данных при помощи библиотек Matplotlib и seaborn [1], Анализ болезни по болезни сердца при помощи библиотеки Seaborn проделал А. М. Черкашин [2], А. М. Васильченко пояснил как проводить анализ данных при помощи библиотек Python[3], Разведочный анализ данных в python (GoogleColab)

описала А. С. Матвеева [4], М. С. Федотова, Н. Н. Барышева написал обзор информационных систем на языке python в государственном управлении [5].

1.3 Цель исследования

Целью исследования является анализ данных оттока клиентов в банке при помощи библиотеки Seaborn в Google Colaboratory.

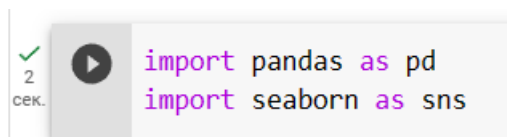
2 Материалы и методы

В данном исследовании используется облачный сервис Google Colaboratory.

3 Результаты

Для анализа данных при помощи библиотеки Seaborn будет использоваться файл с данными о оттоке клиентов с банка. Скачать файл можно по ссылке: [Churn-Modelling-Dataset/Churn_Modelling.csv](#) в мастере · sharmaroshan/Churn-Modelling-Dataset · GitHub

Первый шаг – это подключение библиотеки Pandas и Seaborn (см. рис.1).



```
import pandas as pd
import seaborn as sns
```

Рисунок 1 – Подключение библиотеки

Необходимо подгрузить файл к Google Colab, для этого переходим в «файлы», после этого кликаем на «Загрузить в сессионное хранилище», после совершения этих действий в появившемся окне выбираем файл «Churn_Modelling.csv» и в поле «Файлы» появится файл для дальнейшей работы (см. рис. 2).

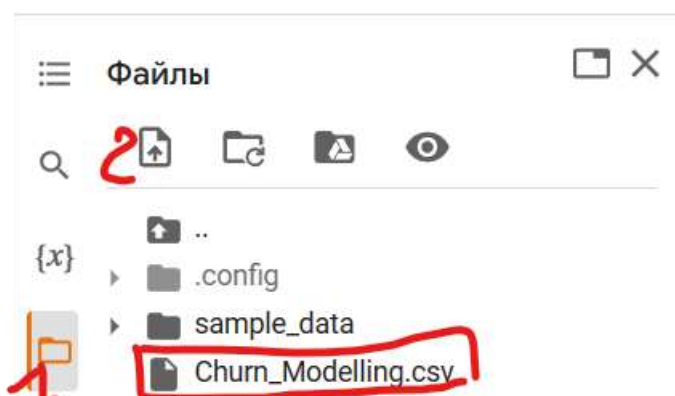


Рисунок 2 – Подгрузка файла

Далее подключаем данные при помощи библиотеки Pandas (см. рис. 3).

```
[8] colors=('Creditscore','Geography','Age','Tenure','Balance','NumOfProduct','IsActiveMember','EstimatedSalary','Exited')
df_churn=pd.read_csv('churn_Modelling.csv')
```

Рисунок 3 – Подключение данных

Выведем первых пять записей на экран (см. рис. 4).

```
df_churn.head()
```

	RowNumber	CustomerId	Surname	Creditscore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary
0	1	15634602	Hargrave	619	France	Female	42	2	0.00	1	1	1	
1	2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	
2	3	15619304	Onio	502	France	Female	42	8	150660.80	3	1	0	
3	4	15701354	Boni	699	France	Female	39	1	0.00	2	0	0	
4	5	15737888	Mitchell	850	Spain	Female	43	2	125510.82	1	1	1	

Рисунок 4 – Данные таблицы

На основе этих данных необходимо построим график частотности признака «Geography», с условием их подсчета (см. рис. 5).

```
ax=sns.catplot(x='Geography', kind="count", data=df_churn)
```

Рисунок 5 – Код для построения графика

В результате написанного кода появится график со странами и их количеством (см. рис.6).

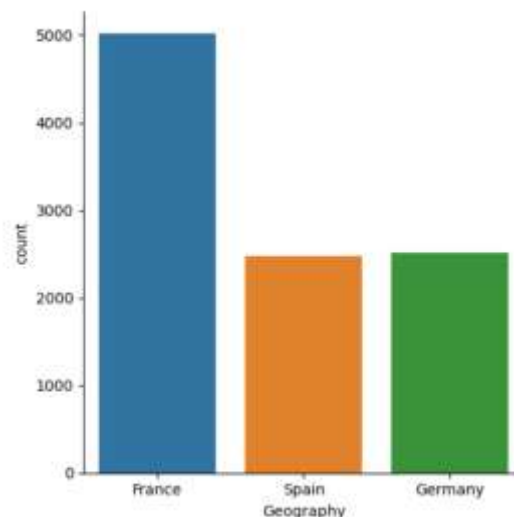


Рисунок 6 – График «Количество стран»

Следующий шаг нам позволит узнать к каким странам больше относится отток клиентов (см. рис. 7)

```
ax=sns.catplot(x='Geography', kind='count', hue='Exited', data=df_churn)
```

Рисунок 7 – Код для построения графика

Теперь можно наблюдать, что для каждой страны появилось два столбца, в одном отображены постоянные клиенты, в другом отток клиентов. Анализируя график видно, что самый большой отток клиентов в банке у Германии (см. рис. 8)

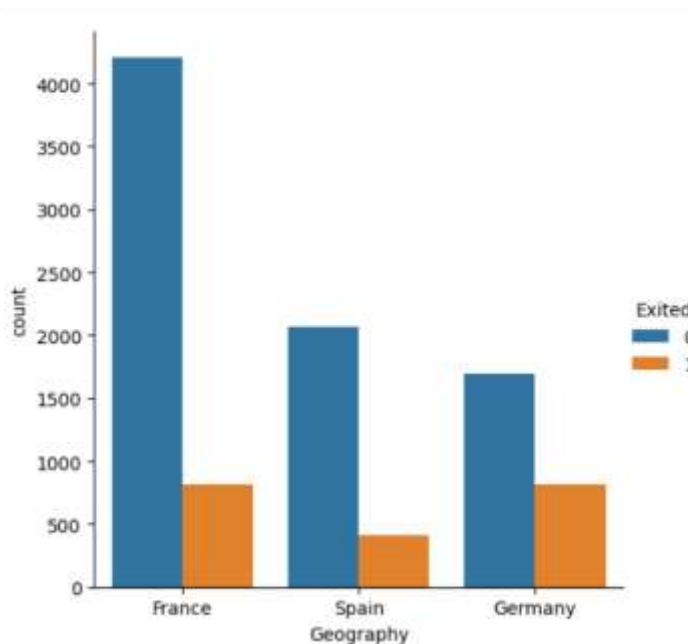


Рисунок 8 – График «Отток клиентов»

Далее проанализируем активных и пассивных клиентов в этих странах (см. рис. 9).

```
ax=sns.catplot(x='Geography', kind="count", hue='IsActiveMember', data=df_churn)
```

Рисунок 9 – Код для построения графика

В графике можем наблюдать количество активных и неактивных клиентов, первый столбец каждой страны – активные клиенты, второй столбец каждой страны – неактивные (см. рис. 11)

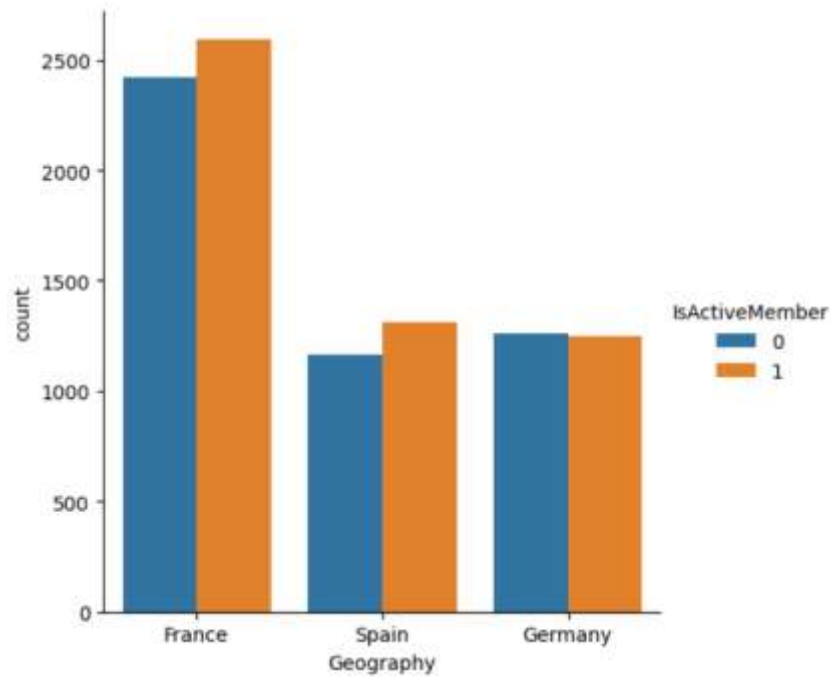


Рисунок 11- График активности клиентов

Так же есть возможность проанализировать использование продуктов клиентами (см. рис. 12).

```
ax=sns.catplot(x='Geography', kind="count", hue='Tenure', data=df_churn)
```

Рисунок 12 – Код для построения графика

Анализируя график можем наблюдать, что во Франции клиенты пользуются самым большим количеством различных услуг (см. рис. 13)

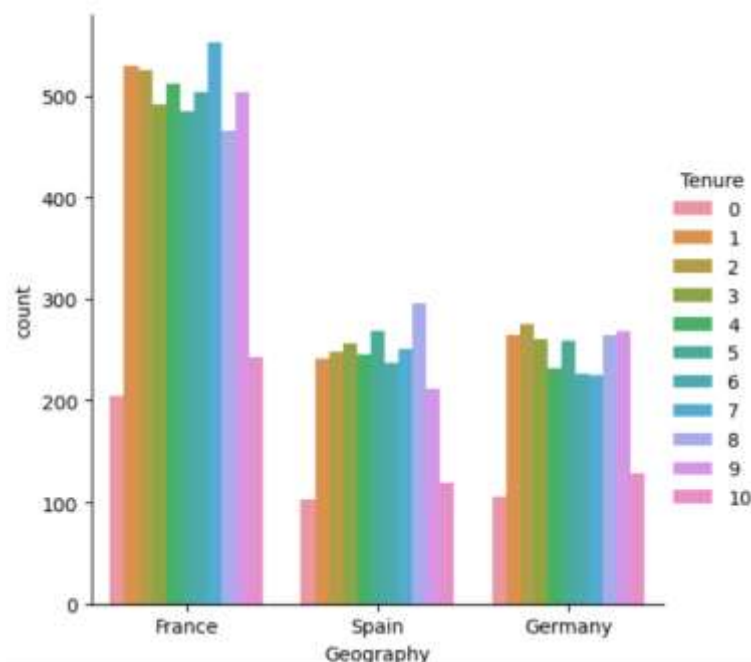


Рисунок 13 – График «Количество используемых услуг»

Следующий анализ – это анализ гистограммы выбывшие и не выбывшие клиенты (см. рис. 14).

```
✓ 1 сек. ax=sns.histplot(x='CreditScore', hue='Exited', data=df_churn)
```

Рисунок 14 – Код для построения гистограммы

Посмотрев на гистограмму можно увидеть количество выбывших – серый цвет колонок, количество оставшихся – синий цвет колонки (см. рис. 15).

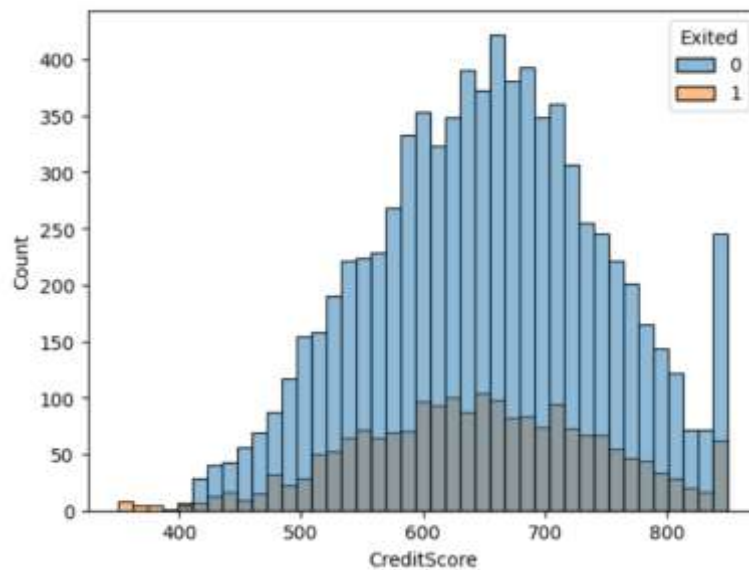


Рисунок 15 – Гистограмма выбывших и не выбывших клиентов в банке

Теперь проанализируем выбывших и оставшихся клиентов по странам (см. рис. 16).

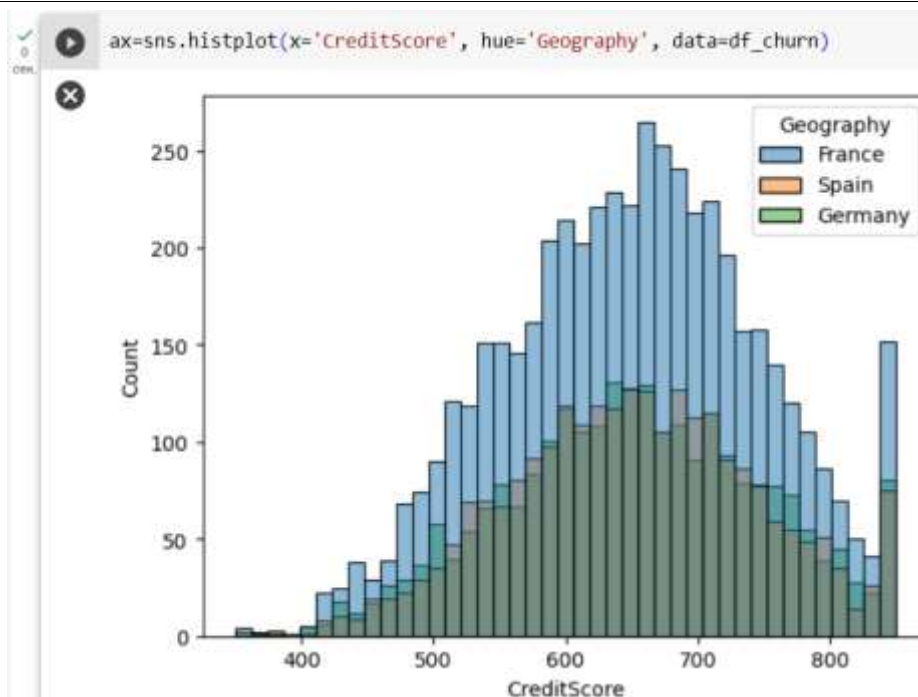


Рисунок 16 – Гистограмма «Выбывшие и не выбывшие клиенты банка по странам»

4. Выводы

Таким образом, в данной статье был рассмотрен анализ данных оттока клиента при помощи библиотеки Seaborn. Полученный результат можно использовать для анализа данных, отслеживание динамики и пособием для обучения. Было отслежено какие страны больше подвержены оттоку клиентов.

Библиографический список

1. Bisong E., Bisong E. Matplotlib and seaborn //Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners. 2019. С. 151-165.
2. Черкашин А. М. Анализ данных по болезни сердца //Постулат. 2022. №. 1.
3. Васильченко А. М. Как проводить анализ данных при помощи python? //Инновации и инвестиции. 2023. №. 5. С. 161-165.
4. Матвеева А. С. Разведочный анализ данных в python (GoogleColab) //Постулат. 2023. №. 1.
5. Федотова М. С., Барышева Н. Н. Обзор информационных систем на языке python в государственном управлении //Наука и образование. 2022. Т. 5. №. 2. С. 374.