

Анализ данных обнаружения людей в помещении в Google Colaboratory

Анишкова Анастасия Сергеевна

Приамурский государственный университет имени Шолом-Алейхема

Студент

Аннотация

Целью исследования является анализ данных обнаружения людей в помещении при помощи библиотеки seaborn и matplotlib.pyplot. Для реализации использовался облачный сервис Google Colaboratory. Полученный результат использовался для определения человека в помещении.

Ключевые слова: Google Colaboratory, seaborn, matplotlib.pyplot.

Analyzing the data of detecting people in a room in the Google Coollaboratory

Anishkova Anastasia Sergeevna

Sholom-Aleichem Priamurskiy State University

Student

Abstract

The purpose of the study is to analyze the data of detecting people in a room using the seaborn library and matplotlib.pyplot. The Google Coollaboratory cloud service was used for implementation. The result obtained was used to identify the person in the room.

Keywords: Google Colaboratory, seaborn, matplotlib.pyplot.

1 Введение

1.1 Актуальность

Анализ данных обнаружения людей в помещении в Google Colaboratory может быть актуальным для обеспечения безопасности, управления персоналом, оптимизации использования пространства, маркетинга и исследований.

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 [3], качественный анализ данных сделал 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 оттока клиентов в банке при помощи библиотеки Seaborn и matplotlib.pyplot в Google Colaboratory.

2 Материалы и методы

В данном исследовании используется облачный сервис Google Colaboratory.

3 Результаты

Для анализа используем файл с данными по обнаружению людей в здании. Файл состоит из таких столбцов:

date – дата и время.

Temperature - температура.

Humidity - относительная влажность в %.

Light - освещенность в Люксах.

CO2 - концентрация углекислого газа в миллионных долях.

HumidityRatio - величина, производная от температуры и относительной влажности.

Occupancy - целевой признак. Если в помещении находятся люди, то 1, иначе 0.

Скачать данные можно по ссылке <https://cloud.mail.ru/public/5JN1/stqDfoqyV>

Для анализа необходимо в Google Colab подключить необходимые библиотеки (см. рис.1).

```
[31] import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

Рисунок 1 – Подключение библиотек

Далее необходимо загрузить данные, которых будем проводить анализ (см. рис.2).



Рисунок 2 – Загрузка данных

Подключаем датасет с помощью библиотеки Pandas (см. рис.3).

```
[32] data=pd.read_csv('datatraining.txt', sep=',')
```

Рисунок 3 – Подключение датасета

Визуализируем первые 5 записей на экран.

```
data.head()
```

Теперь посмотрим размер данных в датасете (см. рис.4)

```
line_count=data.shape[0]
column_count=data.shape[1]
print('Всего строк: {}'.format(line_count))
print('Всего колонок: {}'.format(column_count))
```

Всего строк: 8143
Всего колонок: 7

Рисунок 4 – размер данных

При помощи команды `columns` выведем название столбцов (см. рис.5).

```
[20] data.columns
```

```
Index(['date', 'Temperature', 'Humidity', 'Light', 'CO2', 'HumidityRatio',  
       'Occupancy'],  
      dtype='object')
```

Рисунок 5 – Название столбцов

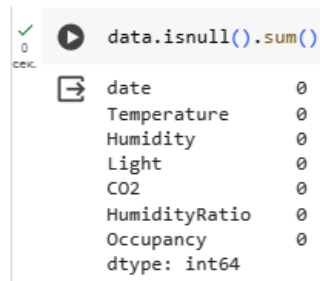
Так же выведем типы данных столбцов (см. рис.6).

```
[21] data.dtypes
```

```
date           object  
Temperature    float64  
Humidity       float64  
Light          float64  
CO2            float64  
HumidityRatio  float64  
Occupancy      int64  
dtype: object
```

Рисунок 6 – типы данных

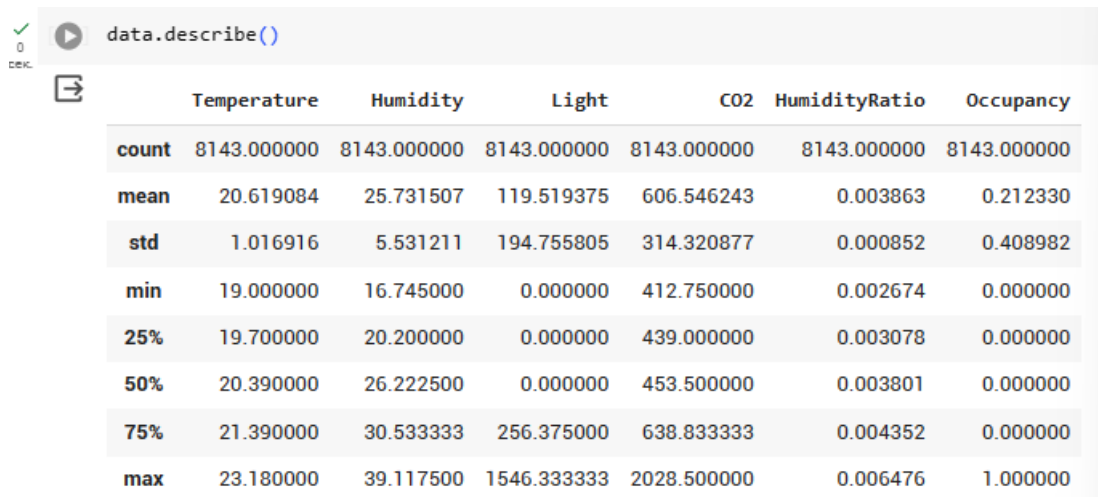
Удостоверимся в отсутствие пропущенных значений (см. рис.7).



```
data.isnull().sum()
date           0
Temperature    0
Humidity       0
Light          0
CO2            0
HumidityRatio  0
Occupancy     0
dtype: int64
```

Рисунок 7 – Пропущенные значения

Пропущенных значений нет, дополнительную обработку данных использовать нет необходимости, поэтому приступим к дальнейшему анализу и посмотрим статистические показатели (см. рис.8). Метод `describe` в Google Colab используется для описания данных, которые были загружены в колонку данных. Этот метод позволяет получить информацию о данных, такие как количество строк и столбцов, типы данных, минимальные и максимальные значения, а также статистические показатели, такие как среднее значение, медиана, стандартное отклонение и квартили.



```
data.describe()
Temperature  Humidity  Light  CO2  HumidityRatio  Occupancy
count  8143.000000  8143.000000  8143.000000  8143.000000  8143.000000  8143.000000
mean    20.619084   25.731507   119.519375   606.546243    0.003863    0.212330
std     1.016916    5.531211   194.755805   314.320877    0.000852    0.408982
min     19.000000   16.745000    0.000000   412.750000    0.002674    0.000000
25%    19.700000   20.200000    0.000000   439.000000    0.003078    0.000000
50%    20.390000   26.222500    0.000000   453.500000    0.003801    0.000000
75%    21.390000   30.533333   256.375000   638.833333    0.004352    0.000000
max     23.180000   39.117500  1546.333333  2028.500000    0.006476    1.000000
```

Рисунок 8 – Статистические показатели

Далее построим график плотности температуры (см. рис.9).

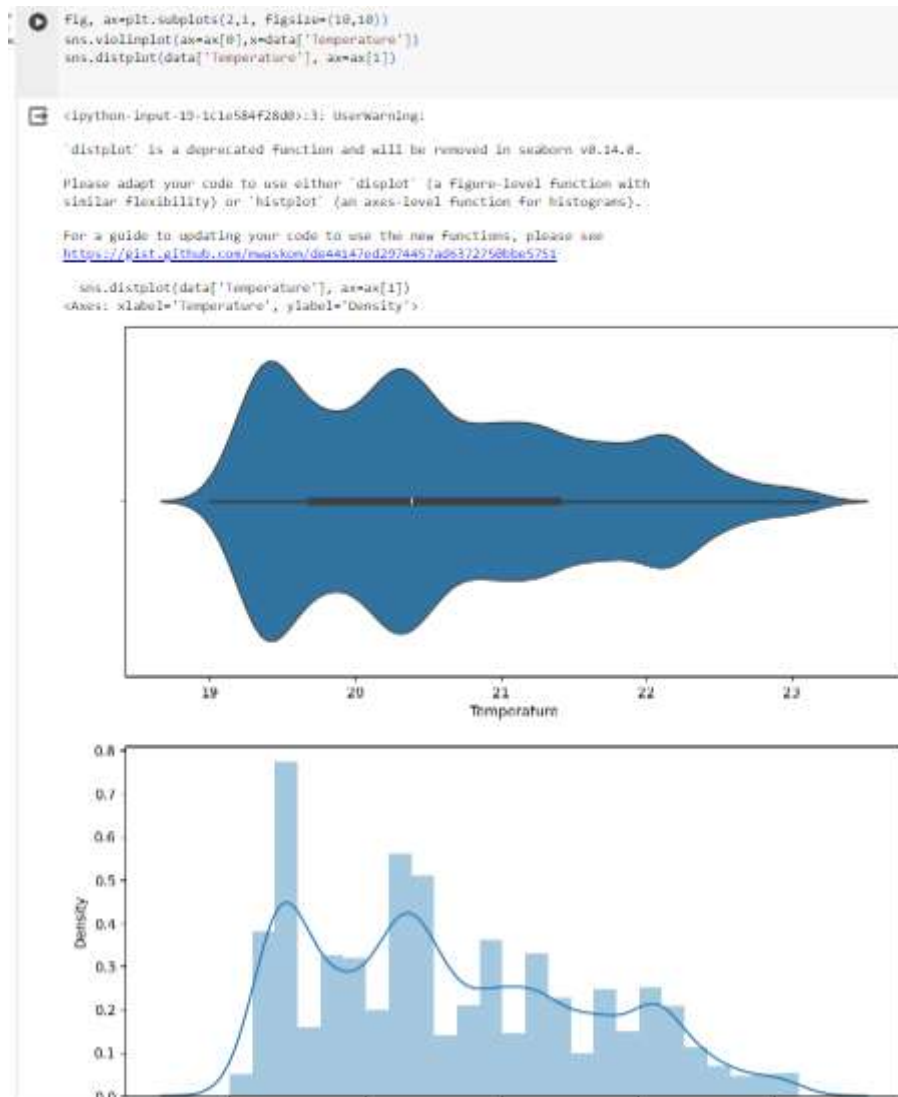


Рисунок 9 – График распределения плотности

Теперь визуализируем распределение вероятностей одномерно (см. рис.10). Это средняя температура на протяжении всего времени: когда в помещении присутствуют люди и отсутствуют.

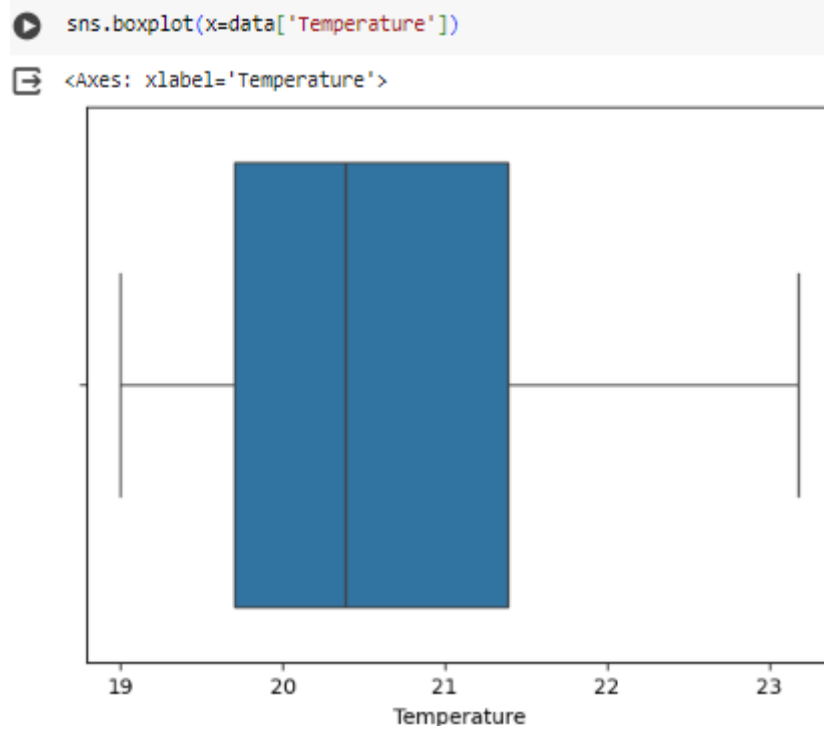


Рисунок 10 – Распределение вероятностей

Необходимо выяснить какая температура должна быть во время присутствия людей. Необходимо отсортировать данные по наличию человека (см. рис. 11). Результат показывает, что температура должна быть около 21,75 градусов.

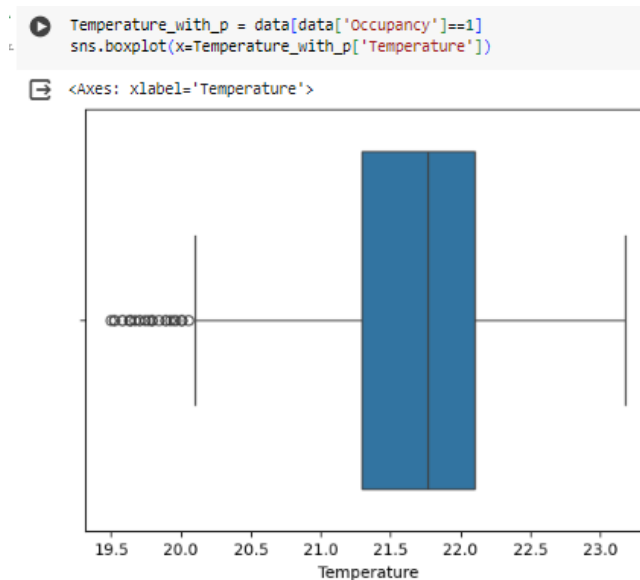


Рисунок 11 – Наличие человека

Для наглядности построим парные диаграммы (см. рис.10). и корреляционную матрицу (см. рис.12).

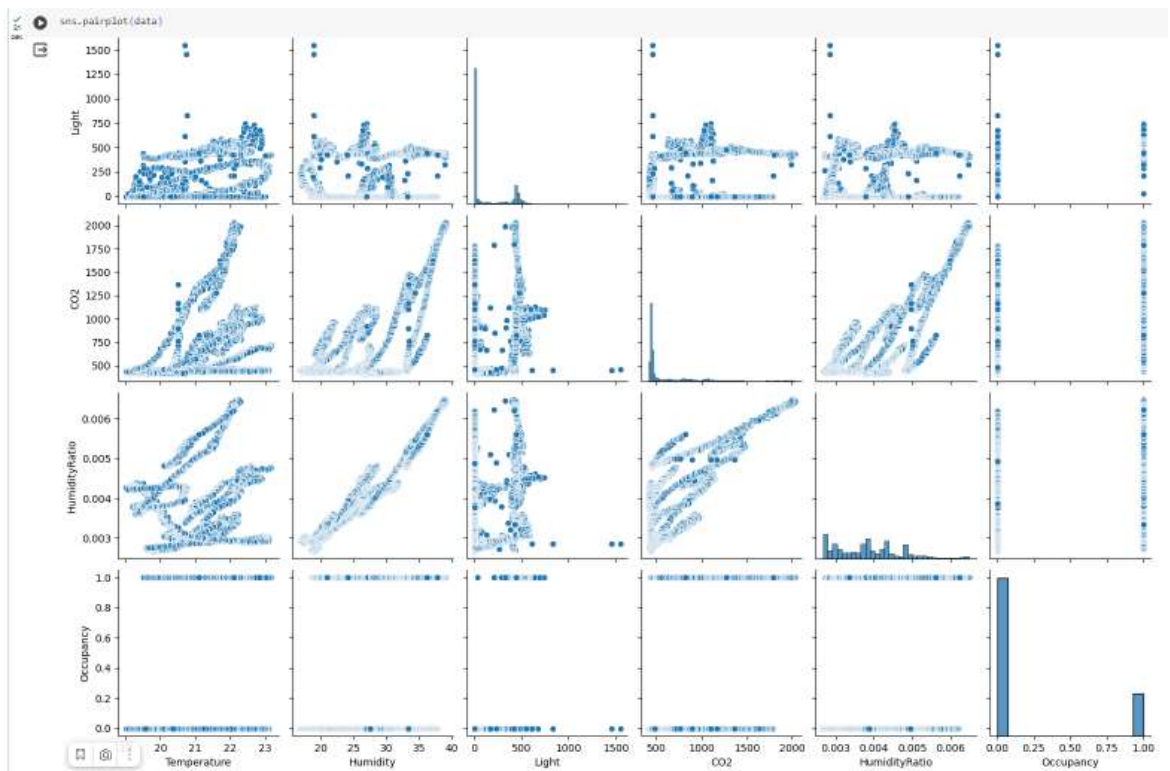


Рисунок 12 – Парные диаграммы

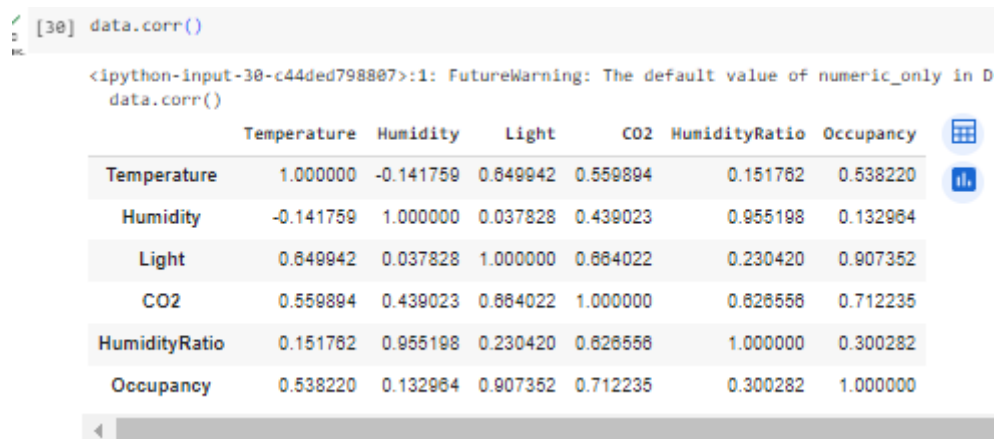


Рисунок 13 - Корреляционная матрица

Распределим значения по убыванию:

1. Light - 0,90
2. CO2 - 0,71
3. Temperature - 0,54
4. HumidityRatio - 0,30
5. Humidity - 0,13.

После анализа можно сделать вывод, что присутствие человека в помещении взаимосвязано с температурой. Умный дом должен сам поддерживать необходимую температуру – это 21,75 градуса

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. Войскунский 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.