



**Министерство науки и высшего образования
Российской Федерации Федеральное государственное
бюджетное образовательное учреждение высшего
образования «Московский государственный
технический университет имени Н.Э. Баумана**

(национальный исследовательский университет)»

(МГТУ им. Н.Э. Баумана)

Факультет «Информатика и системы управления»

Кафедра ИУ5 «Системы обработки информации и управления»

Отчёт по РК1

«Технологии машинного обучения»

Вариант 7

Выполнил:

студент группы ИУ5-63Б

Некрасов С. А.

Преподаватель:

Гапанюк Ю. Е.

2023 г.

Задание:

Задача №1.

Для заданного набора данных проведите корреляционный анализ. В случае наличия пропусков в данных удалите строки или колонки, содержащие пропуски. Сделайте выводы о возможности построения моделей машинного обучения и о возможном вкладе признаков в модель.

Для студентов групп ИУ5-63Б, ИУ5Ц-83Б - для произвольной колонки данных построить график "Ящик с усами (boxplot)".

<https://www.kaggle.com/lava18/google-play-store-apps>

Решение:

Загружаем датасет и подключаем необходимые библиотеки:

```
In [74]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

```
In [75]: ds = pd.read_csv('googleplaystore.csv')
```

```
In [76]: ds.head()
```

Out[76]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up

Получим информацию о датасете:

```
In [77]: ds.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10841 entries, 0 to 10840
Data columns (total 13 columns):
#   Column              Non-Null Count  Dtype
---  ---
0   App                  10841 non-null  object
1   Category             10841 non-null  object
2   Rating               9367 non-null   float64
3   Reviews              10841 non-null  object
4   Size                 10841 non-null  object
5   Installs              10841 non-null  object
6   Type                 10840 non-null  object
7   Price                10841 non-null  object
8   Content Rating       10840 non-null  object
9   Genres               10841 non-null  object
10  Last Updated         10841 non-null  object
11  Current Ver          10833 non-null  object
12  Android Ver          10838 non-null  object
dtypes: float64(1), object(12)
memory usage: 1.1+ MB
```

Посчитаем количество пустых значений:

```
In [78]: ds.isna().sum()
```

```
Out[78]: App                0
Category                0
Rating                1474
Reviews                0
Size                  0
Installs              0
Type                  1
Price                 0
Content Rating        1
Genres                 0
Last Updated          0
Current Ver           8
Android Ver           3
dtype: int64
```

Удалим строки с пустыми значениями:

```
In [79]: ds.dropna()
```

```
Out[79]:
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design, Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design, Creativity	June 20, 2018	1.1	4.4 and up
...
10834	FR Calculator	FAMILY	4.0	7	2.6M	500+	Free	0	Everyone	Education	June 18, 2017	1.0.0	4.1 and up
10836	Sya9a Maroc - FR	FAMILY	4.5	38	53M	5,000+	Free	0	Everyone	Education	July 25, 2017	1.48	4.1 and up
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3.6M	100+	Free	0	Everyone	Education	July 6, 2018	1.0	4.1 and up

```
In [81]: ds.describe().T
```

```
Out[81]:
```

	count	mean	std	min	25%	50%	75%	max
Rating	9367.0	4.193338	0.537431	1.0	4.0	4.3	4.5	19.0

Оптимизируем память и изменим типы данных на более корректные:

```
In [82]: int_types = ["uint8", "int8", "int16"]
for it in int_types:
    print(np.iinfo(it))
```

```
Machine parameters for uint8
```

```
-----
min = 0
max = 255
-----
```

```
Machine parameters for int8
```

```
-----
min = -128
max = 127
-----
```

```
Machine parameters for int16
```

```
-----
min = -32768
max = 32767
-----
```

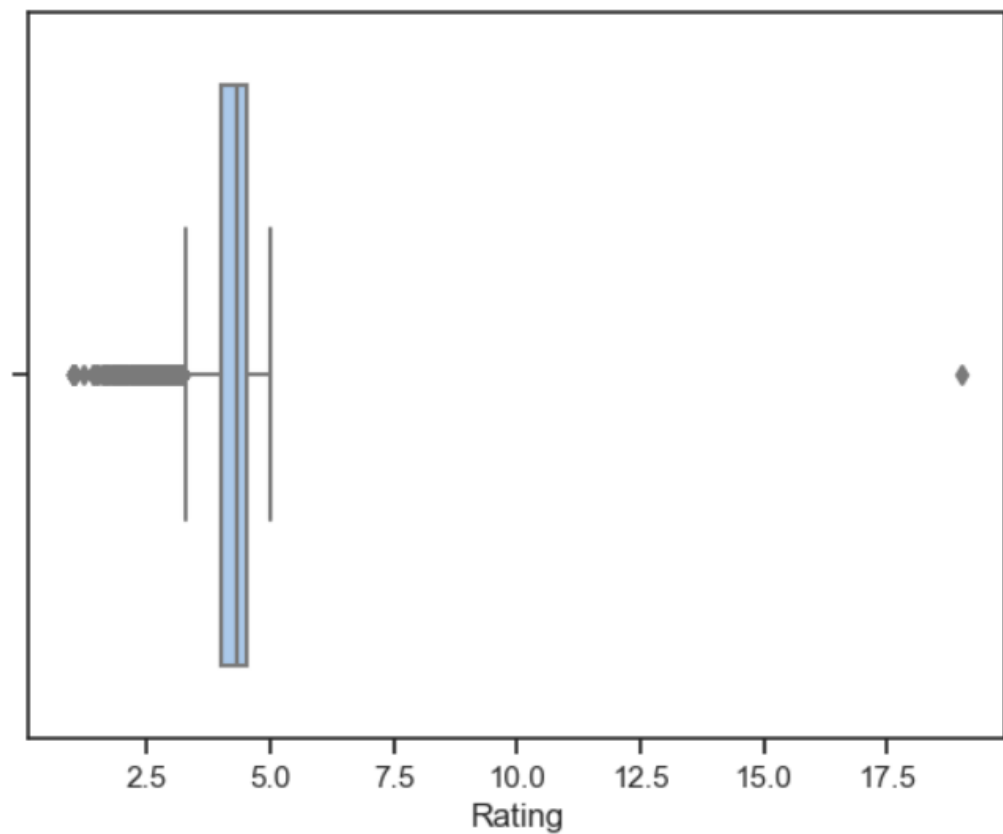
```
In [83]: ds.info(memory_usage='deep')

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10841 entries, 0 to 10840
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  -
0   App                    10841 non-null  object
1   Category               10841 non-null  object
2   Rating                 9367 non-null   float64
3   Reviews                10841 non-null  object
4   Size                   10841 non-null  object
5   Installs               10841 non-null  object
6   Type                   10840 non-null  object
7   Price                  10841 non-null  object
8   Content Rating         10840 non-null  object
9   Genres                 10841 non-null  object
10  Last Updated           10841 non-null  object
11  Current Ver            10833 non-null  object
12  Android Ver            10838 non-null  object
dtypes: float64(1), object(12)
memory usage: 8.3 MB
```

По колонке “Rating” сделаем boxplot (Ящик с усами):

```
In [85]: sns.set_theme(style="ticks", palette="pastel")
sns.boxplot(x=ds["Rating"])
```

```
Out[85]: <Axes: xlabel='Rating'>
```



Корреляционный анализ:

Данный датасет нельзя использовать для корреляционного анализа, так как он почти полностью состоит из неподходящих типов данных. Значит, построение такой модели машинного обучения невозможно.