

МГТУ им. Н. Э. Баумана, кафедра ИУ5
курс “Технология машинного обучения”

Рубежный контроль №1

«Технологии разведочного анализа и обработки
данных»

ВЫПОЛНИЛ:

Ерохин И.А.

Группа: ИУ5-61Б

Вариант: 8

ПРОВЕРИЛ:

Гапанюк Ю.Е.

Задание:

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.

Набор данных - 8

Задание группы:

Построить диаграмму рассеивания.

Текст программы и экранные формы с примерами выполнения программы:

Yandex Browser (beta) C6, 08:23

PK - Jupyter Notebook

PK - Jupyter Notebook

Last Checkpoint: несколько секунд назад (autosaved)

File Edit View Insert Cell Kernel Help Trusted Python 3

```
In [171]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

%matplotlib inline
sns.set(style="ticks")

In [172]: data = pd.read_csv('RK/googleplaystore.csv', sep=',')

In [173]: data.dtypes
Out[173]: App                object
Category                object
Rating                 float64
Reviews                object
Size                   object
Installs               object
Price                  object
Content Rating         object
Genres                 object
Last Updated           object
Current Ver            object
Android Ver            object
dtype: object

In [174]: data.head()
Out[174]:
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Cam & 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;Puzzle	January 15, 2018	2.0.0	4.0.3 and up
2	U Launcher Lite - FREE Live Cool Themes, HD ...	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

```
In [175]: # нахождение пустых значений
data.isnull().sum()
Out[175]: App                0
Category                0
Rating                 1474
Reviews                0
Size                   0
Installs               1
Price                  0
Content Rating         1
Genres                 0
Last Updated           0
Current Ver            0
Android Ver            3
dtype: int64
```

Yandex Browser (beta) C6, 08:23

PK - Jupyter Notebook

PK - Jupyter Notebook

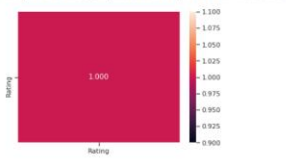
Last Checkpoint: несколько секунд назад (autosaved)

File Edit View Insert Cell Kernel Help Trusted Python 3

```
In [175]: # нахождение пустых значений
data.isnull().sum()
Out[175]: App                0
Category                0
Rating                 1474
Reviews                0
Size                   0
Installs               1
Price                  0
Content Rating         1
Genres                 0
Last Updated           0
Current Ver            0
Android Ver            3
dtype: int64

In [176]: # Удаление строк, содержащих пустые значения
data_new_2 = data.dropna(axis=0, how='any')
(data.shape, data_new_2.shape)
Out[176]: ((10841, 13), (9360, 13))

In [177]: sns.heatmap(data.corr(), annot=True, fct='f')
Out[177]: <matplotlib.axes._subplots.AxesSubplot at 0x7f23c9f0d8d0>
```



```
In [178]: # как мы можем видеть корреляция по 2 признаку ничего не дает,
# один из вариантов привести числовые признаки к типам

In [179]: from sklearn.preprocessing import OneHotEncoder

In [180]: buff = pd.DataFrame(columns=['Rating', 'Price', 'Reviews'])
buff[Reviews] = pd.to_numeric(data_new_2.Reviews, 1)
```

