

## РК1 Коростелёв В. М. ИУ5-21М

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.

```
In [1]: #подключаем библиотеки
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
sns.set(style="ticks")
```

```
In [2]: #подключаем датасет
data = pd.read_csv('googleplaystore.csv', sep=",")
```

```
In [3]: data.dtypes
```

```
Out[3]: App                object
Category                object
Rating                 float64
Reviews                object
Size                   object
Installs               object
Type                   object
Price                  object
Content Rating         object
Genres                 object
Last Updated           object
Current Ver            object
Android Ver            object
dtype: object
```

```
In [4]: #проверяем количество пустых ячеек
data.isnull().sum()
```

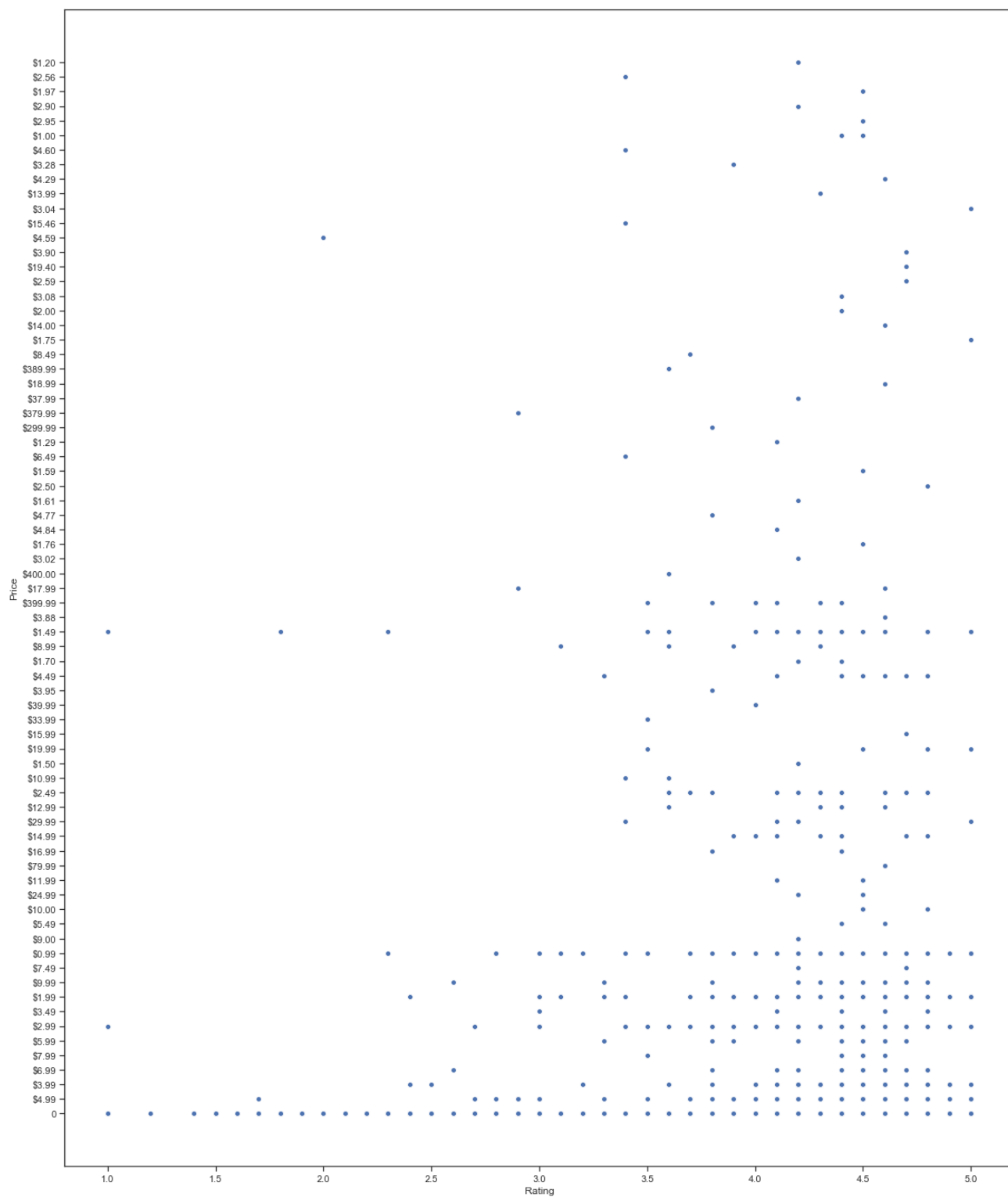
```
Out[4]: App                0
Category                0
Rating                 1474
Reviews                0
Size                   0
Installs               0
Type                   1
Price                  0
Content Rating         1
Genres                 0
Last Updated           0
Current Ver            8
Android Ver            3
dtype: int64
```

```
In [5]: data3 = data.dropna(axis=0, how='any')
(data.shape, data3.shape)
```

```
Out[5]: ((10841, 13), (9360, 13))
```

```
In [6]: fig, ax = plt.subplots(figsize=(20,25))
sns.scatterplot(ax=ax, x='Rating', y='Price', data=data3)
```

```
Out[6]: <matplotlib.axes._subplots.AxesSubplot at 0x297aaf20f28>
```



```
In [7]: fig, ax = plt.subplots(figsize=(20,25))
sns.scatterplot(ax=ax, x='Rating', y='Price', data=data3, hue='Installs')
```

Out[7]: <matplotlib.axes.\_subplots.AxesSubplot at 0x297adab0358>

