

```
import pandas as pd
%matplotlib inline

data = pd.read_csv('modified_smartphone_data.csv')
data

{"type": "dataframe", "variable_name": "data"}

data.isnull().sum()

brand_name      0
model           0
price           0
rating          99
has_5g          0
has_nfc         0
has_ir_blaster  0
num_cores       6
processor_speed 41
battery_capacity 11
fast_charging_available 0
fast_charging   205
ram_capacity    0
internal_memory 0
screen_size     0
refresh_rate    0
resolution      0
num_rear_cameras 0
primary_camera_rear 0
extended_memory_available 0
extended_upto   472
dtype: int64
```

Пропущенные значения содержатся в столбцах rating, processor_brand, num_cores, processor_speed, battery_capacity, fast_charging, num_front_cameras, os, primary_camera_front и extended_upto. Колонка extended_upto содержит 480 пропущенных значений, а ее информация может быть скоррелирована с другими признаками. Таким образом, колонка extended_upto не является важной для решения нашей задачи.

```
data.drop('extended_upto', axis=1, inplace=True)
```

Для более точного анализа, пропущенные значения в колонке rating заполним на основе данных о бренде. Необходимо сгруппировать телефоны по бренду и вычислить медианный рейтинг для каждой группы с помощью метода groupby() и функции median(). Затем мы использовать цикл for для заполнения пропущенных значений

```
data.brand_name.unique()

array(['oneplus', 'samsung', 'motorola', 'realme', 'apple', 'xiaomi',
       'nothing', 'oppo', 'vivo', 'poco', 'iqoo', 'jio', 'gionee',
```

```

        'tecno', 'google', 'infinix', 'letv', 'ikall', 'nokia', 'lava',
        'honor', 'nubia', 'redmi', 'asus', 'itel', 'royole', 'sony',
        'oukitel', 'lyf', 'huawei', 'zte', 'lenovo', 'lg', 'micromax',
        'doogee'], dtype=object)

median_rating_by_brand = data.groupby('brand_name')['rating'].median()

# заполнить пропущенные значения медианными для каждой группы
for brand in data.brand_name.unique():
    data.loc[data['rating'].isnull() & (data['brand_name'] == brand),
             'rating'] = median_rating_by_brand[brand]

```

Пропуски в остальных колонках заполним модой.

```

most_frequent_num_cores = data['num_cores'].mode()[0]
data['num_cores'].fillna(most_frequent_num_cores, inplace=True)

most_frequent_processor_speed = data['processor_speed'].mode()[0]
data['processor_speed'].fillna(most_frequent_processor_speed,
                               inplace=True)

most_frequent_battery_capacity = data['battery_capacity'].mode()[0]
data['battery_capacity'].fillna(most_frequent_battery_capacity,
                                inplace=True)

```

Проверим теперь датасет на пропуски:

```

data.isnull().sum()

```

brand_name	0
model	0
price	0
rating	0
has_5g	0
has_nfc	0
has_ir_blaster	0
num_cores	0
processor_speed	0
battery_capacity	0
fast_charging_available	0
ram_capacity	0
internal_memory	0
screen_size	0
refresh_rate	0
resolution	0
num_rear_cameras	0
primary_camera_rear	0
extended_memory_available	0

```
color                                0
dtype: int64
```

Выполним кодирование категориального признака color, обозначающего цвет телефона. Используем one-hot кодирование.

```
color_dummies = pd.get_dummies(data['color'], prefix='Color')
data_oh = pd.concat([data, color_dummies], axis=1)
data_oh.drop('color', axis=1, inplace=True)

data.color.unique()

array(['Brick Red', 'Copper White', 'Prism Dot Gray', 'Bolivia Blue',
      'Deep Black', 'Clearly White', 'Genuine Leather Black',
      'Garlic',
      'Cloud Pink', 'Diamond Glow', 'Dusk Blue', 'Gunmetal Silver',
      'Electric Black', 'Dashing Blue', 'Cloud Mint', 'Matrix
Purple',
      'Sandstone Black', 'Starry purple', 'Volcano Grey', 'Starry
Night',
      'Crystal Black', 'Infinite Black', 'Orchid Grey', 'Tempered
Blue',
      'Lavender ', 'Vanilla Mint', 'Midnight Black',
      'Prism Crush Black', 'Phantom Violet', 'Shadow Black', 'Fine
Gold',
      'Mercury Silver', 'White ', 'Pacific Pearl', 'Aurora Silver',
      'Nebula Blue', 'Twilight Purple', 'Pearl white', 'Moss Green',
      'White & Copper', 'Black/Tuxedo Black', 'Mirage Blue', 'So
Blue',
      'Celestial Snow', 'Voyager Grey', 'Black & Copper', 'Pheonix
Red',
      'Frosted Pearl', 'Blazing Blue', 'Ruby Red', 'Magic Blue',
      'Brown',
      'Prism Crush Blue', 'Meteor Grey', 'Armoured Edition',
      'Classic Silver', 'Stardust White', 'Neo Black', 'Glacier
Blue',
      'Pink Sand', 'Twilight Orange', 'Slate Blue', 'So White',
      'Bordeaux Red', 'Diamond Sapphire', 'Laser Blue', 'Aqua Green',
      'Copper', 'Crystal Green', 'SeaBlue', 'Rich Cranberry',
      'Daybreak Blue', 'Lunar White', 'Marble Green', 'Midnight
Jazz',
      'Space Gray', 'Enigma Black', 'Out of Blue', 'More Than White',
      'Camo Green', 'Blue & Silver', 'Black and Gold', 'Waterfall
Grey',
      'Bold Red', 'Laser Green', 'That Green', 'COCKTAIL ORANGE',
      'AURORA SILVER', 'Rose Pink', 'Mocha Brown', 'Predator Black',
      'Lavender Purple', 'Mystic Blue', 'Sea Blue', 'Diamond Black',
      'Blue', 'Thunder Blue', 'Night | Dark Blue', 'Pearl Blue',
      'Illuminating Yellow', 'Silky White', 'Fusion Green',
```

'Tornado Black', 'Cosmic Black', 'Cyber Silver', 'Coral Blue',
 'Fusion Blue', 'Ice Blue', 'Crystal Silver', 'Fluid Black',
 'White Knight', 'Galactic Blue', 'Fantastic Rainbow', 'Onyx
 Black',
 'Fantasy White', 'Awesome Violet', 'Rich Grey',
 'Prism Crush White', 'Symphony Cyan', 'Frost Blue',
 'Metallic Grey', 'Serene Gold', 'black sapphire', 'Black&Blue',
 'Mist Black', 'Purple Mist', 'White', 'Sporty Orange',
 'Diamond Blue', 'Pastel Sky', 'Black Ninja', 'Blossom Pink',
 'Watery Grey', 'Racing Yellow', 'Luna White ', 'Pacific
 Sunrise',
 'Cosmic Blue', 'Cool Blue', 'Forest Green', 'Aurora Green',
 'Stardust Black', 'CHAMPAGNE GOLD', 'Matte Black', 'Gold Sand',
 'Violet', 'Luminous Black', 'Cloud Red', 'Champagne Gold',
 'Midnight Grey', 'Brilliant Black', 'Glowing Green', 'Blush
 Gold',
 'Brave Blue', 'Black Blue', 'Carbon', 'Shadow Grey', 'Lunar
 Grey',
 'Phantom Black', 'Obsidian Black', 'Copper Gold', 'Ocean Wave',
 'Frost White', 'CELESTIAL SILVER', 'Metallic Blue', 'Diamond
 Ruby',
 'Caviar Black', 'Blade Silver', 'GREY/BLACK', 'Space Purple',
 'Vinyl Black', 'Polished Copper', 'Mint', 'Caribbean Blue',
 'Jazz Blue', 'Mint Cream', 'Prism Magic', 'Space Silver',
 'Orange',
 'Kind of Grey', 'Dark Pearl', 'Cosmos Black', 'Awesome White',
 'Azure Glow', 'Concrete', 'Energetic Blue', 'Startrails Blue',
 'Sapphire Cyan', 'Charcoal Gray', 'Crystal Blue', 'Night',
 'Sierra Blue', 'Black Titan', 'Red Brick', 'Cosmic Gray',
 'Heart of Ocean', 'Aura White', 'Stream White',
 'Titanium Sapphire', 'Nordic Blue', 'Bordeaux Red ', 'Cyber
 Teal',
 'Twilight Black', 'Cross Black', 'Two shades of black',
 'Asteroid Black', 'Carbon Black', 'Gradation Blue', 'Aura Red',
 'Gold', 'Rainbow Blue', 'Metallic Copper', 'CRYSTAL BLACK',
 'Dynamic Orange', 'Mystic Green', 'Maple Gold', 'Platinum',
 'Fairy White', 'Iceberg blue', 'Nature Green', 'Quartz Black',
 'Mystic Bronze', 'Ink Black', 'Grey / Silver', 'Sunset Red',
 'Glossy Black', 'Angel Gold', 'Blazing Red', 'Green and
 Greener',
 'Lilac Purple', 'Prism Blue', 'Marine Green', 'Cloud Navy',
 'Pebble Grey', 'Starry Night Black', 'Ambitious Black',
 'Coral Red', 'Black Pearl', 'Ceramic Black', 'Sapphire Blue',
 'Prism Crush Red', 'Steel Blue', 'Dawn White', 'Milkyway Grey',
 'Silver Titan', 'Frosted Emerald', 'Pitch Black', 'Slate Gray',
 'Stellar Black', 'Alpine Green', 'Fiery Gold', 'Neptune Blue',
 'Noble Black', 'Celestial Magic', 'Mighty Black', 'Pink',
 'Agate Red', 'Aquamarine Green', 'Watery Blue', 'Midnight',
 'Nordic Secret', 'Glory Silver', 'Magic Gold', 'Lunar Silver',

'Pearl White', 'Ultra Violet', 'Charcoal Blue', 'Phantom
 Purple',
 'Cosmic Purple', 'Radiant Blue', 'Dusk | Purple',
 'Prism Crush Violet', 'Seawater Blue', 'Midday Dream',
 'Shark Grey', 'Phantom Gray', 'Crystal Symphony', 'Iron',
 'Sunshine Gold', 'Breeze Blue', 'Flame Red', 'Very Silver',
 'Starry Blue', 'Bronze Gold Black', 'That White', 'Tornado
 Black',
 'dark gold', 'Frosted Champagne', 'Graphite Black',
 'Starlight Black', 'Mirror Black', 'Fancy Blue', 'Glaze Blue',
 'Arctic Blue', 'Grey', 'Awesome Mint', 'Quantum Silver', 'Ice',
 'Sunset Jazz', 'Aegean Blue', 'Amazing Silver', 'Sprinkle
 White',
 'Black', 'Metal Grey', 'Denim Black', 'Polaris Blue',
 'Lavender Violet', 'Infinite Blue', 'Nebula', 'Gradation
 Black',
 'Lightning Blue', 'Aura Black', 'Deep Sea Blue', 'Pure White',
 'Fervor Red', 'Auroral Blue', 'Misty Blue', 'Purist Blue',
 'Pewter / White', 'Sonic Blue', 'Dark Nebula', 'Glowing Black',
 'Topaz Blue', 'Midnight Purple', 'Glaring Gold', 'Lunar Gray',
 'Glacier White', 'Sunrise Flare', 'Mystery Black',
 'Frosted Silver', 'Raven Black', 'Sword Black', 'Ceramic
 White',
 'Mirage Black', 'Meteor Black', 'Dark Night', 'Rainbow Black',
 'SAND', 'Twilight Grey', 'Diamond Dazzle', 'Moroccan Blue',
 'Haze Crush Silver', 'Stargaze White', 'Smokey Gray', 'Neo
 Blue',
 'Power Silver', 'Mist White', 'Power Blue', 'Metallic Gray',
 'Grey ', 'Rosso Red', 'CYAN', 'Prism Black', 'Sea Green',
 'Venom Black', 'MAGIC BLUE', 'Wine Red', 'Blue MB', 'Slate
 Black',
 'Polar Gold', 'Space Grey', 'Purple ', 'Black Leather',
 'Lightning Red', 'Astro Moonlight White', 'Stainless Black',
 'Baltic', 'Silver Titanium', 'Moonlight Gold', 'Santorini
 White',
 'Sonic Black', 'Quetzal Cyan', 'Celestial Silver',
 'Sparkling Blue', 'Tradew Grey', 'Space Black', 'Green',
 'Super Polar White', 'Rose Gold', 'Phantom Silver', 'Titan
 Gray',
 'Tahiti Blue', 'Universe Purple', 'Nebula Purple', 'Polished
 Blue',
 'Dark Blue', 'Alpine White', 'Metallic White', 'Diamond White',
 'Metallic Sage', 'Prism White', 'Mocha Gold', 'Classic White',
 'Sunrise Red', 'Twilight Blue', 'Polished Graphite', 'Silver',
 'Eclipse Black', 'Smoky Sangria', 'Cross Blue', 'Astral Blue',
 'Sunset Blue', 'Sunrise Gold', 'Charcoal Grey', 'Cloud Blue',
 'Sprite', 'Sky White', 'Celestial Black', 'Gold & Black',
 'Awesome Blue', 'Platinum Grey', 'Oxford Blue', 'Quartz Green',
 'Meteor Silver', 'Arctic White', 'Blue Lagoon', 'Prism Dot

```

Black',
    'Aurora Grey', 'Fantastic Purple', 'Interstellar Black',
    'Electric Violet', 'Sky Blue', 'Thunder Black', 'Absolute
black',
    'Light Blue', 'Red', 'Sunset Melody', 'Chroma White', 'Fjord
Blue',
    'Dazzling Gold', 'Topaz Gold', 'Onion', 'Mostly Blue',
    'Aquamarine Blue', 'Dusk', 'Copper/Black', 'Silver Blue',
    'Sterling Blue', 'Morandi Green', 'Aurora', 'Mystic Black',
    'Blazing Black', 'That Blue', 'Super Black', 'Lemonade Blue',
    'Iceberg Blue', 'Purple', 'Lake Blue', 'Iris Charcoal',
    'Lime Green', 'Coral Green', 'Prism Crush Silver', 'Latte
Gold',
    'Prime Black', 'Starry Purple', 'Deep Blue', 'midnight blue',
    'COSMIC BLACK', 'Cloud Lavender', 'Black Sky', 'Indigo Black',
    'Solar Red', 'Silver White', 'Black Sapphire', 'Emerald Black',
    'Amber Red', 'IceBerg Blue', 'Graphite', 'Steel', 'Sunrise Blue
',
    'Aurora Dawn', 'Pink Gold', 'Atlantic Blue', 'Pearl Green',
    'Neon Purple', 'Marble White', 'City Blue', 'Jet Black',
    'Mist Blue', 'Jewelry White', 'Bronze Gradient', 'Black
Diamond',
    'Aurora Gray', 'Midnight Green', 'MOONLIGHT WHITE', 'Haze
Green',
    'Laser Black', 'Dynamic Gray', 'Genuine Leather Brown',
    'Lightning Orange', 'Frosted Gold', 'Chic White', 'Stealth
Black',
    'Pacific Blue', 'Bronze Gold'], dtype=object)

data_oh
{"type": "dataframe", "variable_name": "data_oh"}

```

В датасете можно выполнить масштабирование данных с помощью методов MinMaxScaler и StandardScaler из модуля sklearn.preprocessing.

С помощью метода MinMaxScaler приведем значения в колонке Rating к диапазону от 0 до 1

```

from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()

data['rating_scaled'] = scaler.fit_transform(data[['rating']])

data
{"type": "dataframe", "variable_name": "data"}

```

Используем метод StandardScaler (на основе Z-оценки) (удаляет среднее значение и масштабирует данные до единичной дисперсии). Это означает, что среднее значение всех значений признака будет равно 0, а стандартное отклонение будет равно 1. Вычисляет

среднее значение и стандартное отклонение для каждого признака и применяет следующее преобразование для каждого значения признака

```
from sklearn.preprocessing import StandardScaler
standartScaler = StandardScaler()

data['price_scaled'] = scaler.fit_transform(data[['price']])

data

{"type": "dataframe", "variable_name": "data"}
```