# Table of Contents

# Роялти за 1 квартал 2022 г

## 1 План работы

- чтение данных и предобработка
- обработка неявных дубликатов
-
-
- Исследовательский анализ данных
- карта (WorldMap)

In [115]:
```
# импорт бибилеотеки pandas
import pandas as pd
# import seaborn as sns
import numpy as np
# import matplotlib.pyplot as plt
import plotly.express as px
import plotly
```

In [116]:
```
# настройки отображения
# показывать до 40ка колонок
pd.set_option('display.max.columns', 40)
# установка формата вывода на дисплей численных значений
pd.options.display.float_format = '{:,.2f}'.format
```

## 2 Чтение данных и предобработка

### 2.1 Чтение данных

In [117]:
```
# чтение файлы данных
df = pd.read_csv("https://eddydewrussia.ru/download/muzik_q1_22/?wpdmdl=5243&masterkey=Fs8MQutYt22fdFY9tSjg-Bz2LUUqkiv_GaWLSjGS7IGZE9Iu2O8iq9dY62oSuUQXZvasAu10egrZD7eCXp34
df.head()
```

Out[117]:

| | Labelname | ISRC | EAN/UPC | Artist | Producttitle | Tracktitle | ArtNo | Outletname | Format | Territory | Sales Period | Units | Royalty Amount Customer | Unnamed: 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Sila navsegda | RUA3R2128055 | 7,423,703,037,037.00 | Eduard Dementyev | Luchina (it 's Not the Wind That 's Driving th... | Luchina (it 's Not the Wind That 's Driving th... | NaN | Amazon | Track | IT | 202,202.00 | 1.00 | 0,006605 | NaN |
| **1** | Sila navsegda | RUA3R2123705 | 7,423,739,417,452.00 | Эдуард Дементьев feat. Roma Skeptik | Ой, мороз, мороз. | Ой, мороз, мороз. | NaN | Tidal | Track | AR | 202,201.00 | 1.00 | 0,000875 | NaN |
| **2** | Sila navsegda | RUA3R2120520 | 7,423,714,522,560.00 | Eduard Dementyev | Black raven. Song of a Russian warriorю | Chjornyj Voron. Psaltery | NaN | Amazon | Track | US | 202,202.00 | 1.00 | 0,014177 | NaN |
| **3** | Sila navsegda | RUA3R2128055 | 7,423,703,037,037.00 | Eduard Dementyev | Luchina (it 's Not the Wind That 's Driving th... | Luchina (it 's Not the Wind That 's Driving th... | NaN | Amazon | Track | IN | 202,202.00 | 1.00 | 0,001168 | NaN |
| **4** | Sila navsegda | RUA3R2123705 | 7,423,739,417,452.00 | Эдуард Дементьев feat. Roma Skeptik | Ой, мороз, мороз. | Ой, мороз, мороз. | NaN | Youtube | Track | RU | 202,201.00 | 0.00 | 0,003039 | NaN |

In [118]:
```
# получаю информацию о таблице
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 127 entries, 0 to 126
Data columns (total 14 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Labelname               127 non-null    object
 1   ISRC                    126 non-null    object
 2   EAN/UPC                 126 non-null    float64
 3   Artist                  126 non-null    object
 4   Producttitle            106 non-null    object
 5   Tracktitle              126 non-null    object
 6   ArtNo                   31 non-null     object
 7   Outletname              126 non-null    object
 8   Format                  126 non-null    object
 9   Territory               126 non-null    object
 10  Sales Period            126 non-null    float64
 11  Units                   126 non-null    float64
 12  Royalty Amount Customer  126 non-null   object
 13  Unnamed: 13             0 non-null      float64
dtypes: float64(4), object(10)
memory usage: 14.0+ KB
```

## 2.2 EAN/UPC, Units - переведём в целочисленные

In [119]:
```
df = df.astype({"EAN/UPC": "Int64", "Units": "Int64"})
```

In [120]:
```
1 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 127 entries, 0 to 126
Data columns (total 14 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Labelname               127 non-null    object
 1   ISRC                    126 non-null    object
 2   EAN/UPC                 126 non-null    Int64
 3   Artist                  126 non-null    object
 4   Producttitle            106 non-null    object
 5   Tracktitle              126 non-null    object
 6   ArtNo                   31 non-null     object
 7   Outletname              126 non-null    object
 8   Format                  126 non-null    object
 9   Territory               126 non-null    object
 10  Sales Period            126 non-null    float64
 11  Units                   126 non-null    Int64
 12  Royalty Amount Customer 126 non-null    object
 13  Unnamed: 13             0 non-null      float64
dtypes: Int64(2), float64(2), object(10)
memory usage: 14.3+ KB
```

In [121]:
```
1 df.head()
```

Out[121]:

| | Labelname | ISRC | EAN/UPC | Artist | Producttitle | Tracktitle | ArtNo | Outletname | Format | Territory | Sales Period | Units | Royalty Amount Customer | Unnamed: 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Sila navsegda | RUA3R2128055 | 7423703037037 | Eduard Dementyev | Luchina (it 's Not the Wind That 's Driving th... | Luchina (it 's Not the Wind That 's Driving th... | NaN | Amazon | Track | IT | 202,202.00 | 1 | 0,006605 | NaN |
| **1** | Sila navsegda | RUA3R2123705 | 7423739417452 | Эдуард Дементьев feat. Roma Skeptik | Ой, мороз, мороз. | Ой, мороз, мороз. | NaN | Tidal | Track | AR | 202,201.00 | 1 | 0,000875 | NaN |
| **2** | Sila navsegda | RUA3R2120520 | 7423714522560 | Eduard Dementyev | Black raven. Song of a Russian warriorю | Chjornyj Voron. Psaltery | NaN | Amazon | Track | US | 202,202.00 | 1 | 0,014177 | NaN |
| **3** | Sila navsegda | RUA3R2128055 | 7423703037037 | Eduard Dementyev | Luchina (it 's Not the Wind That 's Driving th... | Luchina (it 's Not the Wind That 's Driving th... | NaN | Amazon | Track | IN | 202,202.00 | 1 | 0,001168 | NaN |
| **4** | Sila navsegda | RUA3R2123705 | 7423739417452 | Эдуард Дементьев feat. Roma Skeptik | Ой, мороз, мороз. | Ой, мороз, мороз. | NaN | Youtube | Track | RU | 202,201.00 | 0 | 0,003039 | NaN |

### 2.3 странное количество строк в первой колонке

- есть колонка с пустыми значениями
- названия колонок не в змеином_регистре

In [122]:
```
1 df.head()
```

Out[122]:

| | Labelname | ISRC | EAN/UPC | Artist | Producttitle | Tracktitle | ArtNo | Outletname | Format | Territory | Sales Period | Units | Royalty Amount Customer | Unnamed: 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Sila navsegda | RUA3R2128055 | 7423703037037 | Eduard Dementyev | Luchina (it 's Not the Wind That 's Driving th... | Luchina (it 's Not the Wind That 's Driving th... | NaN | Amazon | Track | IT | 202,202.00 | 1 | 0,006605 | NaN |
| **1** | Sila navsegda | RUA3R2123705 | 7423739417452 | Эдуард Дементьев feat. Roma Skeptik | Ой, мороз, мороз. | Ой, мороз, мороз. | NaN | Tidal | Track | AR | 202,201.00 | 1 | 0,000875 | NaN |
| **2** | Sila navsegda | RUA3R2120520 | 7423714522560 | Eduard Dementyev | Black raven. Song of a Russian warriorю | Chjornyj Voron. Psaltery | NaN | Amazon | Track | US | 202,202.00 | 1 | 0,014177 | NaN |
| **3** | Sila navsegda | RUA3R2128055 | 7423703037037 | Eduard Dementyev | Luchina (it 's Not the Wind That 's Driving th... | Luchina (it 's Not the Wind That 's Driving th... | NaN | Amazon | Track | IN | 202,202.00 | 1 | 0,001168 | NaN |
| **4** | Sila navsegda | RUA3R2123705 | 7423739417452 | Эдуард Дементьев feat. Roma Skeptik | Ой, мороз, мороз. | Ой, мороз, мороз. | NaN | Youtube | Track | RU | 202,201.00 | 0 | 0,003039 | NaN |

In [123]:
```
1 df.tail()
```

Out[123]:

| | Labelname | ISRC | EAN/UPC | Artist | Producttitle | Tracktitle | ArtNo | Outletname | Format | Territory | Sales Period | Units | Royalty Amount Customer | Unnamed: 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **122** | Sila navsegda | RUA3R2121453 | 7423735364378 | Эдуард Дементьев | Чёрный ворон | Чёрный ворон - Под гусли | NaN | Spotify | Track | RU | 202,201.00 | 1 | 0,000028 | NaN |
| **123** | Sila navsegda | RUA3R2121453 | 7423735364378 | Эдуард Дементьев | Чёрный ворон | Чёрный ворон - Под гусли | NaN | Spotify | Track | RU | 202,202.00 | 1 | 0,000371 | NaN |
| **124** | Sila navsegda | RUA3R2121453 | 7423735364378 | Эдуард Дементьев | Чёрный ворон | Чёрный ворон - Под гусли | NaN | Spotify | Track | RU | 202,203.00 | 2 | 0,000596 | NaN |
| **125** | Sila navsegda | RUA3R2120520 | 7423714522560 | Eduard Dementyev | Black raven. Song of a Russian warriorю | Chjornyj Voron. Psaltery | NaN | Spotify | Track | US | 202,201.00 | 1 | 0,004072 | NaN |
| **126** | MULTIZA_STATEMENT_EXPORT_=Sila navsegda=_=2022... | NaN | <NA> | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | <NA> | NaN | NaN |

### 2.4 видно, что последняя строка лишняя. её лучше удалить

In [124]:
```
df.drop(df.tail(1).index,inplace=True)
df.tail()
```

Out[124]:

| | Labelname | ISRC | EAN/UPC | Artist | Producttitle | Tracktitle | ArtNo | Outletname | Format | Territory | Sales Period | Units | Royalty Amount Customer | Unnamed: 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **121** | Sila navsegda | RUA3R2121453 | 7423735364378 | Эдуард Дементьев | Чёрный ворон | Чёрный ворон - Под гусли | NaN | Spotify | Track | FI | 202,201.00 | 1 | 0,005603 | NaN |
| **122** | Sila navsegda | RUA3R2121453 | 7423735364378 | Эдуард Дементьев | Чёрный ворон | Чёрный ворон - Под гусли | NaN | Spotify | Track | RU | 202,201.00 | 1 | 0,000028 | NaN |
| **123** | Sila navsegda | RUA3R2121453 | 7423735364378 | Эдуард Дементьев | Чёрный ворон | Чёрный ворон - Под гусли | NaN | Spotify | Track | RU | 202,202.00 | 1 | 0,000371 | NaN |
| **124** | Sila navsegda | RUA3R2121453 | 7423735364378 | Эдуард Дементьев | Чёрный ворон | Чёрный ворон - Под гусли | NaN | Spotify | Track | RU | 202,203.00 | 2 | 0,000596 | NaN |
| **125** | Sila navsegda | RUA3R2120520 | 7423714522560 | Eduard Dementyev | Black raven. Song of a Russian warriorю | Chjornyj Voron. Psaltery | NaN | Spotify | Track | US | 202,201.00 | 1 | 0,004072 | NaN |

In [125]:
```
# удаляю пустую колонку
del df['Unnamed: 13']
df.head()
```

Out[125]:

| | Labelname | ISRC | EAN/UPC | Artist | Producttitle | Tracktitle | ArtNo | Outletname | Format | Territory | Sales Period | Units | Royalty Amount Customer |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Sila navsegda | RUA3R2128055 | 7423703037037 | Eduard Dementyev | Luchina (it 's Not the Wind That 's Driving th... | Luchina (it 's Not the Wind That 's Driving th... | NaN | Amazon | Track | IT | 202,202.00 | 1 | 0,006605 |
| **1** | Sila navsegda | RUA3R2123705 | 7423739417452 | Эдуард Дементьев feat. Roma Skeptik | Ой, мороз, мороз. | Ой, мороз, мороз. | NaN | Tidal | Track | AR | 202,201.00 | 1 | 0,000875 |
| **2** | Sila navsegda | RUA3R2120520 | 7423714522560 | Eduard Dementyev | Black raven. Song of a Russian warriorю | Chjornyj Voron. Psaltery | NaN | Amazon | Track | US | 202,202.00 | 1 | 0,014177 |
| **3** | Sila navsegda | RUA3R2128055 | 7423703037037 | Eduard Dementyev | Luchina (it 's Not the Wind That 's Driving th... | Luchina (it 's Not the Wind That 's Driving th... | NaN | Amazon | Track | IN | 202,202.00 | 1 | 0,001168 |
| **4** | Sila navsegda | RUA3R2123705 | 7423739417452 | Эдуард Дементьев feat. Roma Skeptik | Ой, мороз, мороз. | Ой, мороз, мороз. | NaN | Youtube | Track | RU | 202,201.00 | 0 | 0,003039 |

### 2.5 перевожу названия колонок в змеиный_регистр

In [126]:
```
df = df.rename(columns={'Labelname': 'labelname', 'ISRC': 'isrc', 'EAN/UPC': 'ean_upc',
                        'Artist': 'artist', 'Producttitle': 'producttitle', 'Tracktitle': 'tracktitle',
                        'Outletname': 'outletname', 'Format': 'format', 'Territory': 'territory',
                        'ArtNo':'art_no',
                        'Sales Period': 'sales_period', 'Units': 'units',
                        'Royalty Amount Customer': 'royalty_amount_customer'})
df.head()

```

Out[126]:

| | labelname | isrc | ean_upc | artist | producttitle | tracktitle | art_no | outletname | format | territory | sales_period | units | royalty_amount_customer |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Sila navsegda | RUA3R2128055 | 7423703037037 | Eduard Dementyev | Luchina (it 's Not the Wind That 's Driving th... | Luchina (it 's Not the Wind That 's Driving th... | NaN | Amazon | Track | IT | 202,202.00 | 1 | 0,006605 |
| **1** | Sila navsegda | RUA3R2123705 | 7423739417452 | Эдуард Дементьев feat. Roma Skeptik | Ой, мороз, мороз. | Ой, мороз, мороз. | NaN | Tidal | Track | AR | 202,201.00 | 1 | 0,000875 |
| **2** | Sila navsegda | RUA3R2120520 | 7423714522560 | Eduard Dementyev | Black raven. Song of a Russian warriorю | Chjornyj Voron. Psaltery | NaN | Amazon | Track | US | 202,202.00 | 1 | 0,014177 |
| **3** | Sila navsegda | RUA3R2128055 | 7423703037037 | Eduard Dementyev | Luchina (it 's Not the Wind That 's Driving th... | Luchina (it 's Not the Wind That 's Driving th... | NaN | Amazon | Track | IN | 202,202.00 | 1 | 0,001168 |
| **4** | Sila navsegda | RUA3R2123705 | 7423739417452 | Эдуард Дементьев feat. Roma Skeptik | Ой, мороз, мороз. | Ой, мороз, мороз. | NaN | Youtube | Track | RU | 202,201.00 | 0 | 0,003039 |

## 2.6 проверка на уникальность

## 2.7 Несколько проблем:

- посмотреть что такое art_no (13 строк с данными)
- в artist ??????
- producttitle развалилась - вроде бы разные типы данных

### 2.7.1 art_no

In [127]:
```
df['art_no'].unique()
```

Out[127]:
```
array([nan, 'VKONTAKTE', 'ODNOKLASSNIKI', 'BOOM APP'], dtype=object)
```

### 2.7.2 выведем строки с art_no не NaN

In [128]:
```
# заполним NaN "0"
df['art_no'] = df['art_no'].fillna(0)
df.query('art_no != 0')
```

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **55** | Sila navsegda | RUA3R2129315 | 7423711745757 | Roma Skeptik feat. Эдуард Дементьев | Ой, мороз, мороз | Ой, мороз, мороз | VKONTAKTE | VKONTAKTE | Track | BY | 202,201.00 | 1 | 0,000204 |
| **56** | Sila navsegda | RUA3R2129315 | 7423711745757 | Roma Skeptik feat. Эдуард Дементьев | Ой, мороз, мороз | Ой, мороз, мороз | VKONTAKTE | VKONTAKTE | Track | RU | 202,201.00 | 5 | 0,001022 |
| **57** | Sila navsegda | RUA3R2128418 | 7423701579591 | Эдуард Дементьев | Лучина (То не ветер ветку клонит) | Лучина (То не ветер ветку клонит) | VKONTAKTE | VKONTAKTE | Track | BY | 202,201.00 | 1 | 0,000204 |
| **58** | Sila navsegda | RUA3R2128418 | 7423701579591 | Эдуард Дементьев | Лучина (То не ветер ветку клонит) | Лучина (То не ветер ветку клонит) | VKONTAKTE | VKONTAKTE | Track | RU | 202,201.00 | 6 | 0,001227 |
| **59** | Sila navsegda | RUA3R2117013 | 7423710418409 | Эдуард Дементьев | Ой, мороз, мороз. | Ой, мороз, мороз. | VKONTAKTE | VKONTAKTE | Track | BY | 202,201.00 | 1 | 0,000204 |
| **60** | Sila navsegda | RUA3R2117013 | 7423710418409 | Эдуард Дементьев | Ой, мороз, мороз. | Ой, мороз, мороз. | VKONTAKTE | VKONTAKTE | Track | RU | 202,201.00 | 18 | 0,003681 |
| **61** | Sila navsegda | RUA3R2121453 | 7423735364378 | Эдуард Дементьев | Чёрный ворон | Чёрный ворон | VKONTAKTE | VKONTAKTE | Track | RU | 202,201.00 | 10 | 0,002045 |
| **62** | Sila navsegda | RUA3R2123705 | 7423739417452 | Эдуард Дементьев feat. Roma Skeptik | Ой, мороз, мороз. | Ой, мороз, мороз. | VKONTAKTE | VKONTAKTE | Track | BY | 202,201.00 | 3 | 0,000613 |

Значение в df['art_no'] полностью дублируется в outletname
**колонку можно art_no удалить**

In [129]:
```
del df['art_no']
df.head()
```

Out[129]:

| | labelname | isrc | ean_upc | artist | producttitle | tracktitle | outletname | format | territory | sales_period | units | royalty_amount_customer |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Sila navsegda | RUA3R2128055 | 7423703037037 | Eduard Dementyev | Luchina (it 's Not the Wind That 's Driving th... | Luchina (it 's Not the Wind That 's Driving th... | Amazon | Track | IT | 202,202.00 | 1 | 0,006605 |
| **1** | Sila navsegda | RUA3R2123705 | 7423739417452 | Эдуард Дементьев feat. Roma Skeptik | Ой, мороз, мороз. | Ой, мороз, мороз. | Tidal | Track | AR | 202,201.00 | 1 | 0,000875 |
| **2** | Sila navsegda | RUA3R2120520 | 7423714522560 | Eduard Dementyev | Black raven. Song of a Russian warriorю | Chjornyj Voron. Psaltery | Amazon | Track | US | 202,202.00 | 1 | 0,014177 |
| **3** | Sila navsegda | RUA3R2128055 | 7423703037037 | Eduard Dementyev | Luchina (it 's Not the Wind That 's Driving th... | Luchina (it 's Not the Wind That 's Driving th... | Amazon | Track | IN | 202,202.00 | 1 | 0,001168 |
| **4** | Sila navsegda | RUA3R2123705 | 7423739417452 | Эдуард Дементьев feat. Roma Skeptik | Ой, мороз, мороз. | Ой, мороз, мороз. | Youtube | Track | RU | 202,201.00 | 0 | 0,003039 |

### 2.7.3 producttitle - уникальные значения

In [130]:
```
1 df['producttitle'].unique()
```

Out[130]:
```
array(["Luchina (it 's Not the Wind That 's Driving the Branch)",
       'Ой, мороз, мороз.', 'Black raven. Song of a Russian warrior',
       'Чёрный ворон', 'Russian fairy tale', 'Russian Fairy Tale',
       'Russian Fairy Tale.', '?????? (?? ?? ????? ????? ??????)',
       '?????? ?????', '??, ?????, ?????.', '??, ?????, ?????',
       'Ой, мороз, мороз. (Клубная версия)',
       'Black raven. Song of a Russian warrior.',
       'Лучина (То не ветер ветку клонит) (под гусли)',
       "Russian Fairy Tale. (the Skeptik's Version)",
       'Russian fairy tale (Acoustics Version)',
       'Russian Fairy Tale (Club house version)',
       'Ой, мороз, мороз. (Под гусли)', 'Чёрный ворон (Под гусли)',
       'Ой, мороз, мороз (под гусли)', 'Ой, мороз, мороз',
       'Лучина (То не ветер ветку клонит)',
       'Luchina (it s Not the Wind That s Driving the Branch)',
       'Лучина (То не ветер ветку клонит) ( под гусли)', nan],
      dtype=object)
```

### 2.7.4 Заполняем NaN ы

In [131]:
```
1 df['producttitle'] = df['producttitle'].fillna("faf")
```

In [132]:
```
1 df['producttitle'].unique()
```

Out[132]:
```
array(["Luchina (it 's Not the Wind That 's Driving the Branch)",
       'Ой, мороз, мороз.', 'Black raven. Song of a Russian warrior',
       'Чёрный ворон', 'Russian fairy tale', 'Russian Fairy Tale',
       'Russian Fairy Tale.', '?????? (?? ?? ????? ????? ??????)',
       '?????? ?????', '??, ?????, ?????.', '??, ?????, ?????',
       'Ой, мороз, мороз. (Клубная версия)',
       'Black raven. Song of a Russian warrior.',
       'Лучина (То не ветер ветку клонит) (под гусли)',
       "Russian Fairy Tale. (the Skeptik's Version)",
       'Russian fairy tale (Acoustics Version)',
       'Russian Fairy Tale (Club house version)',
       'Ой, мороз, мороз. (Под гусли)', 'Чёрный ворон (Под гусли)',
       'Ой, мороз, мороз (под гусли)', 'Ой, мороз, мороз',
       'Лучина (То не ветер ветку клонит)',
       'Luchina (it s Not the Wind That s Driving the Branch)',
       'Лучина (То не ветер ветку клонит) ( под гусли)', 'faf'],
      dtype=object)
```

In [133]:
```
1 df['tracktitle'].unique()
```

Out[133]:
```
array(["Luchina (it 's Not the Wind That 's Driving the Branch)",
       'Ой, мороз, мороз.', 'Chjornyj Voron. Psaltery', 'Чёрный ворон',
       'Russian fairy tale', 'Russian Fairy Tale', 'Russian Fairy Tale.',
       '?????? (?? ?? ????? ????? ??????)', '?????? ?????',
       '??, ?????, ?????.', '??, ?????, ?????',
       'Ой, мороз, мороз. (Клубная версия)',
       'Black raven. Song of a Russian warrior.',
       'Лучина (То не ветер ветку клонит) (под гусли)',
       "Russian Fairy Tale. (the Skeptik's Version)",
       'Russian fairy tale (Acoustics Version)',
       'Russian Fairy Tale (Club house version)',
       'Ой, мороз, мороз. (Под гусли)', 'Чёрный ворон (Под гусли)',
       'Ой, мороз, мороз (под гусли)', 'Ой, мороз, мороз',
       'Лучина (То не ветер ветку клонит)',
       'Russian Fairy Tale. (feat. Eduard Dementyev)',
       "Russian Fairy Tale. - the Skeptik's Version",
       'Лучина (То не ветер ветку клонит) - под гусли',
       'Ой, мороз, мороз. - Под гусли', 'Ой, мороз, мороз - под гусли',
       'Чёрный ворон - Под гусли'], dtype=object)
```

In [134]:
```
1 df['tracktitle'] = df['tracktitle'].fillna("faf")
```

In [135]:
```
1 df['tracktitle'].unique()
```

Out[135]:
```
array(["Luchina (it 's Not the Wind That 's Driving the Branch)",
       'Ой, мороз, мороз.', 'Chjornyj Voron. Psaltery', 'Чёрный ворон',
       'Russian fairy tale', 'Russian Fairy Tale', 'Russian Fairy Tale.',
       '?????? (?? ?? ????? ????? ??????)', '?????? ?????',
       '??, ?????, ?????.', '??, ?????, ?????',
       'Ой, мороз, мороз. (Клубная версия)',
       'Black raven. Song of a Russian warrior.',
       'Лучина (То не ветер ветку клонит) (под гусли)',
       "Russian Fairy Tale. (the Skeptik's Version)",
       'Russian fairy tale (Acoustics Version)',
       'Russian Fairy Tale (Club house version)',
       'Ой, мороз, мороз. (Под гусли)', 'Чёрный ворон (Под гусли)',
       'Ой, мороз, мороз (под гусли)', 'Ой, мороз, мороз',
       'Лучина (То не ветер ветку клонит)',
       'Russian Fairy Tale. (feat. Eduard Dementyev)',
       "Russian Fairy Tale. - the Skeptik's Version",
       'Лучина (То не ветер ветку клонит) - под гусли',
       'Ой, мороз, мороз. - Под гусли', 'Ой, мороз, мороз - под гусли',
       'Чёрный ворон - Под гусли'], dtype=object)
```

In [136]:

```
for col in list(df):
    print(col)
    print(np.sort(df[col].unique()))
```

```
labelname
['Sila navsegda']
isrc
['RUA3R2116970' 'RUA3R2117013' 'RUA3R2120520' 'RUA3R2121453'
 'RUA3R2123155' 'RUA3R2123705' 'RUA3R2128055' 'RUA3R2128418'
 'RUA3R2128612' 'RUA3R2129315']
ean_upc
[7423700472473 7423701579591 7423701946959 7423703037037 7423710418409
 7423711745757 7423714522560 7423735364378 7423738235231 7423739417452]
artist
['?????? ?????????' 'Eduard Dementyev'
 'Eduard Dementyev feat. Roma Skeptik' 'Eduard Dementyev;Roma Skeptik'
 'Roma Skeptik' 'Roma Skeptik feat. Eduard Dementyev'
 'Roma Skeptik feat. Эдуард Дементьев' 'Roma Skeptik, Eduard Dementyev'
 'Roma Skeptik, Эдуард Дементьев' 'Roma Skeptik;Эдуард Дементьев'
 'Эдуард Дементьев' 'Эдуард Дементьев feat. Roma Skeptik']
producttitle
['??, ?????, ?????' '??, ?????, ?????.'
 '?????? (?? ?? ????? ????? ??????)' '?????? ?????'
 'Black raven. Song of a Russian warrior.'
 'Black raven. Song of a Russian warriorю'
 "Luchina (it 's Not the Wind That 's Driving the Branch)"
 'Luchina (it s Not the Wind That s Driving the Branch)'
 'Russian Fairy Tale' 'Russian Fairy Tale (Club house version)'
 'Russian Fairy Tale.' "Russian Fairy Tale. (the Skeptik's Version)"
 'Russian fairy tale' 'Russian fairy tale (Acoustics Version)' 'faf'
 'Лучина (То не ветер ветку клонит)'
 'Лучина (То не ветер ветку клонит) ( под гусли)'
 'Лучина (То не ветер ветку клонит) (под гусли)' 'Ой, мороз, мороз'
 'Ой, мороз, мороз (под гусли)' 'Ой, мороз, мороз.'
 'Ой, мороз, мороз. (Клубная версия)' 'Ой, мороз, мороз. (Под гусли)'
 'Чёрный ворон' 'Чёрный ворон (Под гусли)']
tracktitle
['??, ?????, ?????' '??, ?????, ?????.'
 '?????? (?? ?? ????? ????? ??????)' '?????? ?????'
 'Black raven. Song of a Russian warrior.' 'Chjornyj Voron. Psaltery'
 "Luchina (it 's Not the Wind That 's Driving the Branch)"
 'Russian Fairy Tale' 'Russian Fairy Tale (Club house version)'
 'Russian Fairy Tale.' 'Russian Fairy Tale. (feat. Eduard Dementyev)'
 "Russian Fairy Tale. (the Skeptik's Version)"
 "Russian Fairy Tale. - the Skeptik's Version" 'Russian fairy tale'
 'Russian fairy tale (Acoustics Version)'
 'Лучина (То не ветер ветку клонит)'
 'Лучина (То не ветер ветку клонит) (под гусли)'
 'Лучина (То не ветер ветку клонит) - под гусли' 'Ой, мороз, мороз'
 'Ой, мороз, мороз (под гусли)' 'Ой, мороз, мороз - под гусли'
 'Ой, мороз, мороз.' 'Ой, мороз, мороз. (Клубная версия)'
 'Ой, мороз, мороз. (Под гусли)' 'Ой, мороз, мороз. - Под гусли'
 'Чёрный ворон' 'Чёрный ворон (Под гусли)' 'Чёрный ворон - Под гусли']
outletname
['Amazon' 'Apple Music' 'BOOM APP' 'Deezer' 'ODNOKLASSNIKI' 'Spotify'
 'Tidal' 'VKONTAKTE' 'Yandex Music' 'YouTube' 'YouTube Music' 'Youtube']
format
['Track' 'Video']
territory
['AR' 'AT' 'BR' 'BY' 'CA' 'DE' 'FI' 'FR' 'IN' 'IRL' 'IT' 'KG' 'KZ' 'MD'
 'NO' 'RU' 'TW' 'UA' 'US']
```

```
sales_period
[202201. 202202. 202203.]
units
[0 1 2 3 4 5 6 7 8 9 10 12 13 14 15 16 18 20 21 22 23 24 26 27 32 35 37
 149]
royalty_amount_customer
['0' '0,000028' '0,000077' '0,000204' '0,000214' '0,000259' '0,00029'
 '0,000371' '0,000409' '0,000445' '0,000518' '0,000596' '0,000613'
 '0,000651' '0,000676' '0,000777' '0,000778' '0,000818' '0,000875'
 '0,001022' '0,001029' '0,001063' '0,001072' '0,001158' '0,001168'
 '0,001186' '0,001227' '0,001272' '0,001292' '0,00134' '0,001386'
 '0,001554' '0,001636' '0,001691' '0,001711' '0,001712' '0,001863'
 '0,002017' '0,002031' '0,002045' '0,002133' '0,002331' '0,002539'
 '0,00266' '0,002799' '0,002806' '0,002874' '0,002927' '0,003039'
 '0,003087' '0,003091' '0,003317' '0,003681' '0,003793' '0,003875'
 '0,003876' '0,004072' '0,004184' '0,004242' '0,004289' '0,005169'
 '0,005296' '0,005362' '0,005521' '0,005603' '0,005828' '0,006605'
 '0,007477' '0,007684' '0,009196' '0,010723' '0,011041' '0,012173'
 '0,014177' '0,016561' '0,016562' '0,0253' '0,030469' '0,032932'
 '0,035131' '0,038057' '0,040986' '0,043913' '0,056972' '0,05855'
 '0,064402' '0,066467' '0,06733' '0,067332' '0,075963' '0,082289'
 '0,085456' '0,110773' '0,110776' '0,117105']
```

#### 2.7.5 Теперь всё работает. Продолжаем движение

#### 2.7.6 попробуем сделать группировку для определения, что же скрывается за ??? и NaN

In [137]:

```
agg_func_count = {'units': ['sum']}
display(df.groupby(['ean_upc', 'producttitle',  'isrc', 'tracktitle', 'artist']).agg(agg_func_count))
```

| | | | | | units |
|---|---|---|---|---|---|
| | | | | | sum |
| ean_upc | producttitle | isrc | tracktitle | artist | |
| 7423700472473 | Russian Fairy Tale. | RUA3R2128612 | Russian Fairy Tale. | Roma Skeptik | 2 |
| | | | Russian Fairy Tale. - the Skeptik's Version | Roma Skeptik, Eduard Dementyev | 1 |
| | | | Ой, мороз, мороз | Roma Skeptik feat. Eduard Dementyev | 4 |
| | Russian Fairy Tale. (the Skeptik's Version) | RUA3R2128612 | Russian Fairy Tale. (the Skeptik's Version) | Roma Skeptik | 54 |
| | faf | RUA3R2128612 | Russian Fairy Tale. (feat. Eduard Dementyev) | Roma Skeptik | 1 |
| 7423701579591 | ?????? (?? ?? ????? ????? ??????) | RUA3R2128418 | ?????? (?? ?? ????? ????? ??????) | ?????? ????????? | 10 |
| | faf | RUA3R2128418 | Лучина (То не ветер ветку клонит) | Эдуард Дементьев | 19 |
| | Лучина (То не ветер ветку клонит) | RUA3R2128418 | Лучина (То не ветер ветку клонит) | Эдуард Дементьев | 9 |
| | | | Лучина (То не ветер ветку клонит) - под гусли | Эдуард Дементьев | 2 |
| | Лучина (То не ветер ветку клонит) ( под гусли) | RUA3R2128418 | Лучина (То не ветер ветку клонит) | Эдуард Дементьев | 2 |
| | Лучина (То не ветер ветку клонит) (под гусли) | RUA3R2128418 | Лучина (То не ветер ветку клонит) (под гусли) | Эдуард Дементьев | 64 |
| 7423701946959 | Russian fairy tale | RUA3R2116970 | Russian fairy tale | Eduard Dementyev | 2 |
| | | | Ой, мороз, мороз. | Eduard Dementyev | 3 |
| | Russian fairy tale (Acoustics Version) | RUA3R2116970 | Russian fairy tale (Acoustics Version) | Eduard Dementyev | 47 |
| | faf | RUA3R2116970 | Russian Fairy Tale | Eduard Dementyev | 2 |
| 7423703037037 | Luchina (it 's Not the Wind That 's Driving the Branch) | RUA3R2128055 | Luchina (it 's Not the Wind That 's Driving the Branch) | Eduard Dementyev | 68 |
| | Luchina (it s Not the Wind That s Driving the Branch) | RUA3R2128055 | Лучина (То не ветер ветку клонит) | Eduard Dementyev | 8 |
| 7423710418409 | faf | RUA3R2117013 | Ой, мороз, мороз. | Эдуард Дементьев | 4 |
| | Ой, мороз, мороз. | RUA3R2117013 | Ой, мороз, мороз. | Эдуард Дементьев | 22 |
| | | | Ой, мороз, мороз. - Под гусли | Эдуард Дементьев | 1 |
| | Ой, мороз, мороз. (Под гусли) | RUA3R2117013 | Ой, мороз, мороз. | Эдуард Дементьев | 5 |
| | | | Ой, мороз, мороз. (Под гусли) | Эдуард Дементьев | 44 |
| 7423711745757 | ??, ?????, ????? | RUA3R2129315 | ??, ?????, ????? | Roma Skeptik | 3 |
| | Ой, мороз, мороз | RUA3R2129315 | Ой, мороз, мороз | Roma Skeptik feat. Эдуард Дементьев | 6 |
| | | | Ой, мороз, мороз - под гусли | Roma Skeptik, Эдуард Дементьев | 1 |
| | Ой, мороз, мороз (под гусли) | RUA3R2129315 | Ой, мороз, мороз (под гусли) | Roma Skeptik | 24 |
| 7423714522560 | Black raven. Song of a Russian warrior. | RUA3R2120520 | Black raven. Song of a Russian warrior. | Eduard Dementyev | 55 |
| | Black raven. Song of a Russian warriorю | RUA3R2120520 | Chjornyj Voron. Psaltery | Eduard Dementyev | 15 |
| | | | Чёрный ворон | Eduard Dementyev | 2 |
| | faf | RUA3R2120520 | Black raven. Song of a Russian warrior. | Eduard Dementyev | 1 |
| 7423735364378 | ?????? ????? | RUA3R2121453 | ?????? ????? | ?????? ????????? | 9 |
| | faf | RUA3R2121453 | Чёрный ворон | Эдуард Дементьев | 1 |
| | Чёрный ворон | RUA3R2121453 | Чёрный ворон | Эдуард Дементьев | 15 |
| | | | Чёрный ворон - Под гусли | Эдуард Дементьев | 5 |

| ean_upc | producttitle | isrc | tracktitle | artist | units sum |
|---|---|---|---|---|---|
| | Чёрный ворон (Под гусли) | RUA3R2121453 | Чёрный ворон | Эдуард Дементьев | 4 |
| | | | Чёрный ворон (Под гусли) | Эдуард Дементьев | 34 |
| 7423738235231 | Russian Fairy Tale | RUA3R2123155 | Russian Fairy Tale | Eduard Dementyev feat. Roma Skeptik | 3 |
| | | | | Eduard Dementyev;Roma Skeptik | 5 |
| | | | Ой, мороз, мороз. | Eduard Dementyev feat. Roma Skeptik | 11 |
| | Russian Fairy Tale (Club house version) | RUA3R2123155 | Russian Fairy Tale (Club house version) | Eduard Dementyev | 41 |
| 7423739417452 | ??, ?????, ?????. | RUA3R2123705 | ??, ?????, ?????. | ?????? ????????? | 6 |
| | Ой, мороз, мороз. | RUA3R2123705 | Ой, мороз, мороз. | Roma Skeptik;Эдуард Дементьев | 10 |
| | | | | Эдуард Дементьев feat. Roma Skeptik | 204 |
| | Ой, мороз, мороз. (Клубная версия) | RUA3R2123705 | Ой, мороз, мороз. (Клубная версия) | Эдуард Дементьев | 44 |

Ясно. надо много переименовывать.
Вернусь к этому попозже

## 2.8 sales_period преобразуем в дату

In [138]:
```
df['sales_period_dt'] = pd.to_datetime(df['sales_period'], format='%Y%m')
df['sales_period_dt'].head()
```

Out[138]:
```
0   2022-02-01
1   2022-01-01
2   2022-02-01
3   2022-02-01
4   2022-01-01
Name: sales_period_dt, dtype: datetime64[ns]
```

### 2.8.1 Колонка Месяц и Год

In [139]:
```
df['year'] = pd.DatetimeIndex(df['sales_period_dt']).year
df['month'] = pd.DatetimeIndex(df['sales_period_dt']).month
df[['sales_period_dt', 'year', 'month']].head()
```

Out[139]:

| | sales_period_dt | year | month |
|---|---|---|---|
| 0 | 2022-02-01 | 2022 | 2 |
| 1 | 2022-01-01 | 2022 | 1 |
| 2 | 2022-02-01 | 2022 | 2 |
| 3 | 2022-02-01 | 2022 | 2 |
| 4 | 2022-01-01 | 2022 | 1 |

## 2.9 Выведем

isrc - producttitle - ean_upc - tracktitle

### 2.9.1 Преобразуем ean_upc в символьный формат

In [140]:

```
df = df.astype({"ean_upc": "object"})
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 126 entries, 0 to 125
Data columns (total 15 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   labelname                126 non-null    object
 1   isrc                     126 non-null    object
 2   ean_upc                  126 non-null    object
 3   artist                   126 non-null    object
 4   producttitle             126 non-null    object
 5   tracktitle               126 non-null    object
 6   outletname               126 non-null    object
 7   format                   126 non-null    object
 8   territory                126 non-null    object
 9   sales_period             126 non-null    float64
 10  units                    126 non-null    Int64
 11  royalty_amount_customer  126 non-null    object
 12  sales_period_dt          126 non-null    datetime64[ns]
 13  year                     126 non-null    int64
 14  month                    126 non-null    int64
dtypes: Int64(1), datetime64[ns](1), float64(1), int64(2), object(10)
memory usage: 15.0+ KB
```

In [141]:

```
agg_func_count = {'units': ['sum']}
display(df.groupby(['ean_upc', 'producttitle',  'isrc', 'tracktitle', 'artist']).agg(agg_func_count))
```

| | | | | | units |
|---|---|---|---|---|---|
| | | | | | sum |
| ean_upc | producttitle | isrc | tracktitle | artist | |
| 7423700472473 | Russian Fairy Tale. | RUA3R2128612 | Russian Fairy Tale. | Roma Skeptik | 2 |
| | | | Russian Fairy Tale. - the Skeptik's Version | Roma Skeptik, Eduard Dementyev | 1 |
| | | | Ой, мороз, мороз | Roma Skeptik feat. Eduard Dementyev | 4 |
| | Russian Fairy Tale. (the Skeptik's Version) | RUA3R2128612 | Russian Fairy Tale. (the Skeptik's Version) | Roma Skeptik | 54 |
| | faf | RUA3R2128612 | Russian Fairy Tale. (feat. Eduard Dementyev) | Roma Skeptik | 1 |
| 7423701579591 | ?????? (?? ?? ????? ????? ??????) | RUA3R2128418 | ?????? (?? ?? ????? ????? ??????) | ?????? ????????? | 10 |
| | faf | RUA3R2128418 | Лучина (То не ветер ветку клонит) | Эдуард Дементьев | 19 |
| | Лучина (То не ветер ветку клонит) | RUA3R2128418 | Лучина (То не ветер ветку клонит) | Эдуард Дементьев | 9 |
| | | | Лучина (То не ветер ветку клонит) - под гусли | Эдуард Дементьев | 2 |
| | Лучина (То не ветер ветку клонит) ( под гусли) | RUA3R2128418 | Лучина (То не ветер ветку клонит) | Эдуард Дементьев | 2 |
| | Лучина (То не ветер ветку клонит) (под гусли) | RUA3R2128418 | Лучина (То не ветер ветку клонит) (под гусли) | Эдуард Дементьев | 64 |
| 7423701946959 | Russian fairy tale | RUA3R2116970 | Russian fairy tale | Eduard Dementyev | 2 |
| | | | Ой, мороз, мороз. | Eduard Dementyev | 3 |
| | Russian fairy tale (Acoustics Version) | RUA3R2116970 | Russian fairy tale (Acoustics Version) | Eduard Dementyev | 47 |
| | faf | RUA3R2116970 | Russian Fairy Tale | Eduard Dementyev | 2 |
| 7423703037037 | Luchina (it 's Not the Wind That 's Driving the Branch) | RUA3R2128055 | Luchina (it 's Not the Wind That 's Driving the Branch) | Eduard Dementyev | 68 |
| | Luchina (it s Not the Wind That s Driving the Branch) | RUA3R2128055 | Лучина (То не ветер ветку клонит) | Eduard Dementyev | 8 |
| 7423710418409 | faf | RUA3R2117013 | Ой, мороз, мороз. | Эдуард Дементьев | 4 |
| | Ой, мороз, мороз. | RUA3R2117013 | Ой, мороз, мороз. | Эдуард Дементьев | 22 |
| | | | Ой, мороз, мороз. - Под гусли | Эдуард Дементьев | 1 |
| | Ой, мороз, мороз. (Под гусли) | RUA3R2117013 | Ой, мороз, мороз. | Эдуард Дементьев | 5 |
| | | | Ой, мороз, мороз. (Под гусли) | Эдуард Дементьев | 44 |
| 7423711745757 | ??, ?????, ????? | RUA3R2129315 | ??, ?????, ????? | Roma Skeptik | 3 |
| | Ой, мороз, мороз | RUA3R2129315 | Ой, мороз, мороз | Roma Skeptik feat. Эдуард Дементьев | 6 |
| | | | Ой, мороз, мороз - под гусли | Roma Skeptik, Эдуард Дементьев | 1 |
| | Ой, мороз, мороз (под гусли) | RUA3R2129315 | Ой, мороз, мороз (под гусли) | Roma Skeptik | 24 |
| 7423714522560 | Black raven. Song of a Russian warrior. | RUA3R2120520 | Black raven. Song of a Russian warrior. | Eduard Dementyev | 55 |
| | Black raven. Song of a Russian warriorю | RUA3R2120520 | Chjornyj Voron. Psaltery | Eduard Dementyev | 15 |
| | | | Чёрный ворон | Eduard Dementyev | 2 |
| | faf | RUA3R2120520 | Black raven. Song of a Russian warrior. | Eduard Dementyev | 1 |
| 7423735364378 | ?????? ????? | RUA3R2121453 | ?????? ????? | ?????? ????????? | 9 |
| | faf | RUA3R2121453 | Чёрный ворон | Эдуард Дементьев | 1 |
| | Чёрный ворон | RUA3R2121453 | Чёрный ворон | Эдуард Дементьев | 15 |
| | | | Чёрный ворон - Под гусли | Эдуард Дементьев | 5 |

| ean_upc | producttitle | isrc | tracktitle | artist | units sum |
|---|---|---|---|---|---|
| | Чёрный ворон (Под гусли) | RUA3R2121453 | Чёрный ворон | Эдуард Дементьев | 4 |
| | | | Чёрный ворон (Под гусли) | Эдуард Дементьев | 34 |
| 7423738235231 | Russian Fairy Tale | RUA3R2123155 | Russian Fairy Tale | Eduard Dementyev feat. Roma Skeptik | 3 |
| | | | | Eduard Dementyev;Roma Skeptik | 5 |
| | | | Ой, мороз, мороз. | Eduard Dementyev feat. Roma Skeptik | 11 |
| | Russian Fairy Tale (Club house version) | RUA3R2123155 | Russian Fairy Tale (Club house version) | Eduard Dementyev | 41 |
| 7423739417452 | ??, ?????, ?????. | RUA3R2123705 | ??, ?????, ?????. | ?????? ????????? | 6 |
| | Ой, мороз, мороз. | RUA3R2123705 | Ой, мороз, мороз. | Roma Skeptik;Эдуард Дементьев | 10 |
| | | | | Эдуард Дементьев feat. Roma Skeptik | 204 |
| | Ой, мороз, мороз. (Клубная версия) | RUA3R2123705 | Ой, мороз, мороз. (Клубная версия) | Эдуард Дементьев | 44 |

## 2.10 Неявные дубликаты producttitle и tracktitle встречаются вперемешку по альбомам.

Исправлять буду по шагам

### 2.10.1 ean_upc 7423700472473

- producttitle **Russian Fairy Tale. (the Skeptik's Version)**
- tracktitle **Russian Fairy Tale. (the Skeptik's Version)**

In [142]:
```
df.loc[(df['ean_upc'] == 7423700472473), ('producttitle', 'tracktitle')] = "Russian Fairy Tale. (the Skeptik's Version)"
```

### 2.10.2 ean_upc 7423701579591

- producttitle **Лучина (То не ветер ветку клонит)**
- tracktitle **Лучина (То не ветер ветку клонит)**

In [143]:
```
df.loc[(df['ean_upc'] == 7423701579591), ('producttitle', 'tracktitle')] = "Лучина (То не ветер ветку клонит)"
```

### 2.10.3 ean_upc 7423701946959

- producttitle **Russian fairy tale (Acoustics Version)**
- tracktitle **Russian fairy tale (Acoustics Version)**

In [144]:
```
df.loc[(df['ean_upc'] == 7423701946959), ('producttitle', 'tracktitle')] = "Russian fairy tale (Acoustics Version)"
```

### 2.10.4 ean_upc 7423703037037

- producttitle **Luchina (it s Not the Wind That s Driving the Branch)**
- tracktitle **Luchina (it s Not the Wind That s Driving the Branch)**

In [145]:
```
df.loc[(df['ean_upc'] == 7423703037037), ('producttitle', 'tracktitle')] = "Luchina (it s Not the Wind That s Driving the Branch)"
```

#### 2.10.5 ean_upc 7423710418409

- producttitle **Ой, мороз, мороз. (Под гусли)**
- tracktitle **Ой, мороз, мороз. (Под гусли)**

In [146]:
```
df.loc[(df['ean_upc'] == 7423710418409), ('producttitle', 'tracktitle')] = "Ой, мороз, мороз. (Под гусли)"
```

#### 2.10.6 ean_upc 7423711745757

- producttitle **Ой, мороз, мороз. (the Skeptik's Version)**
- tracktitle **Ой, мороз, мороз. (the Skeptik's Version)**

In [147]:
```
df.loc[(df['ean_upc'] == 7423711745757), ('producttitle', 'tracktitle')] = "Ой, мороз, мороз. (the Skeptik's Version)"
```

#### 2.10.7 ean_upc 7423714522560

- producttitle **Black raven. Song of a Russian warrior.**
- tracktitle **Black raven. Song of a Russian warrior.**

In [148]:
```
df.loc[(df['ean_upc'] == 7423714522560), ('producttitle', 'tracktitle')] = "Black raven. Song of a Russian warrior."
```

#### 2.10.8 ean_upc 7423735364378

- producttitle **Чёрный ворон**
- tracktitle **Чёрный ворон**

In [149]:
```
df.loc[(df['ean_upc'] == 7423735364378), ('producttitle', 'tracktitle')] = "Чёрный ворон"
```

#### 2.10.9 ean_upc 7423738235231

- producttitle **Russian Fairy Tale (Club house version)**
- tracktitle **Russian Fairy Tale (Club house version)**

In [150]:
```
df.loc[(df['ean_upc'] == 7423738235231), ('producttitle', 'tracktitle')] = "Russian Fairy Tale (Club house version)"
```

#### 2.10.10 ean_upc 7423739417452

- producttitle **Ой, мороз, мороз. (Клубная версия)**
- tracktitle **Ой, мороз, мороз. (Клубная версия)**

In [151]:
```
df.loc[(df['ean_upc'] == 7423739417452), ('producttitle', 'tracktitle')] = "Ой, мороз, мороз. (Клубная версия)"
```

```
#### проверим результат
```

In [152]:
```
agg_func_count = {'units': ['sum']}
display(df.groupby(['ean_upc', 'producttitle',  'isrc', 'tracktitle', 'artist']).agg(agg_func_count))
```

| | | | | | units |
|---|---|---|---|---|---|
| | | | | | sum |
| ean_upc | producttitle | isrc | tracktitle | artist | |
| 7423700472473 | Russian Fairy Tale. (the Skeptik's Version) | RUA3R2128612 | Russian Fairy Tale. (the Skeptik's Version) | Roma Skeptik | 57 |
| | | | | Roma Skeptik feat. Eduard Dementyev | 4 |
| | | | | Roma Skeptik, Eduard Dementyev | 1 |
| 7423701579591 | Лучина (То не ветер ветку клонит) | RUA3R2128418 | Лучина (То не ветер ветку клонит) | ?????? ????????? | 10 |
| | | | | Эдуард Дементьев | 96 |
| 7423701946959 | Russian fairy tale (Acoustics Version) | RUA3R2116970 | Russian fairy tale (Acoustics Version) | Eduard Dementyev | 54 |
| 7423703037037 | Luchina (it s Not the Wind That s Driving the Branch) | RUA3R2128055 | Luchina (it s Not the Wind That s Driving the Branch) | Eduard Dementyev | 76 |
| 7423710418409 | Ой, мороз, мороз. (Под гусли) | RUA3R2117013 | Ой, мороз, мороз. (Под гусли) | Эдуард Дементьев | 76 |
| 7423711745757 | Ой, мороз, мороз. (the Skeptik's Version) | RUA3R2129315 | Ой, мороз, мороз. (the Skeptik's Version) | Roma Skeptik | 27 |
| | | | | Roma Skeptik feat. Эдуард Дементьев | 6 |
| | | | | Roma Skeptik, Эдуард Дементьев | 1 |
| 7423714522560 | Black raven. Song of a Russian warrior. | RUA3R2120520 | Black raven. Song of a Russian warrior. | Eduard Dementyev | 73 |
| 7423735364378 | Чёрный ворон | RUA3R2121453 | Чёрный ворон | ?????? ????????? | 9 |
| | | | | Эдуард Дементьев | 59 |
| 7423738235231 | Russian Fairy Tale (Club house version) | RUA3R2123155 | Russian Fairy Tale (Club house version) | Eduard Dementyev | 41 |
| | | | | Eduard Dementyev feat. Roma Skeptik | 14 |
| | | | | Eduard Dementyev;Roma Skeptik | 5 |
| 7423739417452 | Ой, мороз, мороз. (Клубная версия) | RUA3R2123705 | Ой, мороз, мороз. (Клубная версия) | ?????? ????????? | 6 |
| | | | | Roma Skeptik;Эдуард Дементьев | 10 |
| | | | | Эдуард Дементьев | 44 |
| | | | | Эдуард Дементьев feat. Roma Skeptik | 204 |

## 2.11 Неявные дубликаты artist тоже по альбомам

### 2.11.1 ean_upc 7423700472473

- artist **Roma Skeptik feat. Eduard Dementyev**

In [153]:
```
df.loc[(df['ean_upc'] == 7423700472473), ('artist')] = "Roma Skeptik feat. Eduard Dementyev"
```

### 2.11.2 ean_upc 7423701579591

- artist **Эдуард Дементьев**

In [154]:
```
df.loc[(df['ean_upc'] == 7423701579591), ('artist')] = "Эдуард Дементьев"
```

### 2.11.3 ean_upc 7423739417452

- artist **Эдуард Дементьев feat. Roma Skeptik**

In [155]:
```
df.loc[(df['ean_upc'] == 7423739417452), ('artist')] = "Эдуард Дементьев feat. Roma Skeptik"
```

In [156]:
```
df.loc[(df['ean_upc'] == 7423711745757), ('artist')] = "Эдуард Дементьев feat. Roma Skeptik"
```

In [157]:
```
df.loc[(df['ean_upc'] == 7423735364378), ('artist')] = "Эдуард Дементьев"
```

In [158]:
```
df.loc[(df['ean_upc'] == 7423738235231), ('artist')] = "Eduard Dementyev feat. Roma Skeptik"
```

In [ ]:
```

```

### 2.11.4 Проверим результат

In [159]:
```
agg_func_count = {'units': ['sum']}
display(df.groupby(['ean_upc', 'producttitle',  'isrc', 'tracktitle', 'artist']).agg(agg_func_count))
```

| | | | | | units |
|---|---|---|---|---|---|
| | | | | | sum |
| ean_upc | producttitle | isrc | tracktitle | artist | |
| 7423700472473 | Russian Fairy Tale. (the Skeptik's Version) | RUA3R2128612 | Russian Fairy Tale. (the Skeptik's Version) | Roma Skeptik feat. Eduard Dementyev | 62 |
| 7423701579591 | Лучина (То не ветер ветку клонит) | RUA3R2128418 | Лучина (То не ветер ветку клонит) | Эдуард Дементьев | 106 |
| 7423701946959 | Russian fairy tale (Acoustics Version) | RUA3R2116970 | Russian fairy tale (Acoustics Version) | Eduard Dementyev | 54 |
| 7423703037037 | Luchina (it s Not the Wind That s Driving the Branch) | RUA3R2128055 | Luchina (it s Not the Wind That s Driving the Branch) | Eduard Dementyev | 76 |
| 7423710418409 | Ой, мороз, мороз. (Под гусли) | RUA3R2117013 | Ой, мороз, мороз. (Под гусли) | Эдуард Дементьев | 76 |
| 7423711745757 | Ой, мороз, мороз. (the Skeptik's Version) | RUA3R2129315 | Ой, мороз, мороз. (the Skeptik's Version) | Эдуард Дементьев feat. Roma Skeptik | 34 |
| 7423714522560 | Black raven. Song of a Russian warrior. | RUA3R2120520 | Black raven. Song of a Russian warrior. | Eduard Dementyev | 73 |
| 7423735364378 | Чёрный ворон | RUA3R2121453 | Чёрный ворон | Эдуард Дементьев | 68 |
| 7423738235231 | Russian Fairy Tale (Club house version) | RUA3R2123155 | Russian Fairy Tale (Club house version) | Eduard Dementyev feat. Roma Skeptik | 60 |
| 7423739417452 | Ой, мороз, мороз. (Клубная версия) | RUA3R2123705 | Ой, мороз, мороз. (Клубная версия) | Эдуард Дементьев feat. Roma Skeptik | 264 |

### 2.11.5 сколько же теперь уникальных значений

In [160]:

```
for col in list(df):
    print(col)
    print(np.sort(df[col].unique()))
```

```
labelname
['Sila navsegda']
isrc
['RUA3R2116970' 'RUA3R2117013' 'RUA3R2120520' 'RUA3R2121453'
 'RUA3R2123155' 'RUA3R2123705' 'RUA3R2128055' 'RUA3R2128418'
 'RUA3R2128612' 'RUA3R2129315']
ean_upc
[7423700472473 7423701579591 7423701946959 7423703037037 7423710418409
 7423711745757 7423714522560 7423735364378 7423738235231 7423739417452]
artist
['Eduard Dementyev' 'Eduard Dementyev feat. Roma Skeptik'
 'Roma Skeptik feat. Eduard Dementyev' 'Эдуард Дементьев'
 'Эдуард Дементьев feat. Roma Skeptik']
producttitle
['Black raven. Song of a Russian warrior.'
 'Luchina (it s Not the Wind That s Driving the Branch)'
 'Russian Fairy Tale (Club house version)'
 "Russian Fairy Tale. (the Skeptik's Version)"
 'Russian fairy tale (Acoustics Version)'
 'Лучина (То не ветер ветку клонит)'
 "Ой, мороз, мороз. (the Skeptik's Version)"
 'Ой, мороз, мороз. (Клубная версия)' 'Ой, мороз, мороз. (Под гусли)'
 'Чёрный ворон']
tracktitle
['Black raven. Song of a Russian warrior.'
 'Luchina (it s Not the Wind That s Driving the Branch)'
 'Russian Fairy Tale (Club house version)'
 "Russian Fairy Tale. (the Skeptik's Version)"
 'Russian fairy tale (Acoustics Version)'
 'Лучина (То не ветер ветку клонит)'
 "Ой, мороз, мороз. (the Skeptik's Version)"
 'Ой, мороз, мороз. (Клубная версия)' 'Ой, мороз, мороз. (Под гусли)'
 'Чёрный ворон']
outletname
['Amazon' 'Apple Music' 'BOOM APP' 'Deezer' 'ODNOKLASSNIKI' 'Spotify'
 'Tidal' 'VKONTAKTE' 'Yandex Music' 'YouTube' 'YouTube Music' 'Youtube']
format
['Track' 'Video']
territory
['AR' 'AT' 'BR' 'BY' 'CA' 'DE' 'FI' 'FR' 'IN' 'IRL' 'IT' 'KG' 'KZ' 'MD'
 'NO' 'RU' 'TW' 'UA' 'US']
sales_period
[202201. 202202. 202203.]
units
[0 1 2 3 4 5 6 7 8 9 10 12 13 14 15 16 18 20 21 22 23 24 26 27 32 35 37
 149]
royalty_amount_customer
['0' '0,000028' '0,000077' '0,000204' '0,000214' '0,000259' '0,00029'
 '0,000371' '0,000409' '0,000445' '0,000518' '0,000596' '0,000613'
 '0,000651' '0,000676' '0,000777' '0,000778' '0,000818' '0,000875'
 '0,001022' '0,001029' '0,001063' '0,001072' '0,001158' '0,001168'
 '0,001186' '0,001227' '0,001272' '0,001292' '0,00134' '0,001386'
 '0,001554' '0,001636' '0,001691' '0,001711' '0,001712' '0,001863'
 '0,002017' '0,002031' '0,002045' '0,002133' '0,002331' '0,002539'
 '0,00266' '0,002799' '0,002806' '0,002874' '0,002927' '0,003039'
 '0,003087' '0,003091' '0,003317' '0,003681' '0,003793' '0,003875'
 '0,003876' '0,004072' '0,004184' '0,004242' '0,004289' '0,005169'
```

```
 '0,005296' '0,005362' '0,005521' '0,005603' '0,005828' '0,006605'
 '0,007477' '0,007684' '0,009196' '0,010723' '0,011041' '0,012173'
 '0,014177' '0,016561' '0,016562' '0,0253' '0,030469' '0,032932'
 '0,035131' '0,038057' '0,040986' '0,043913' '0,056972' '0,05855'
 '0,064402' '0,066467' '0,06733' '0,067332' '0,075963' '0,082289'
 '0,085456' '0,110773' '0,110776' '0,117105']
sales_period_dt
['2022-01-01T00:00:00.000000000' '2022-02-01T00:00:00.000000000'
 '2022-03-01T00:00:00.000000000']
year
[2022]
month
[1 2 3]
```

#### 2.11.6 Успех!

## 2.12 проверяю формат числовых данных

это колонки Units и royalty_amount_customer

In [162]:

```
df.info()
print('units', type(df['units'][5]))
print('royalty_amount_customer', type(df['royalty_amount_customer'][5]))
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 126 entries, 0 to 125
Data columns (total 15 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   labelname               126 non-null    object
 1   isrc                    126 non-null    object
 2   ean_upc                 126 non-null    object
 3   artist                  126 non-null    object
 4   producttitle            126 non-null    object
 5   tracktitle              126 non-null    object
 6   outletname              126 non-null    object
 7   format                  126 non-null    object
 8   territory               126 non-null    object
 9   sales_period            126 non-null    float64
 10  units                   126 non-null    Int64
 11  royalty_amount_customer  126 non-null    object
 12  sales_period_dt         126 non-null    datetime64[ns]
 13  year                    126 non-null    int64
 14  month                   126 non-null    int64
dtypes: Int64(1), datetime64[ns](1), float64(1), int64(2), object(10)
memory usage: 15.0+ KB
units <class 'numpy.int64'>
royalty_amount_customer <class 'str'>
```

## 2.13 тип данных royalty_amount_customer строка

перевожу его в числовой

In [163]:
```
# замена "," на "."
df['royalty_amount_customer'] = df['royalty_amount_customer'].str.replace(',','.')
df.head()
```

Out[163]:

| | labelname | isrc | ean_upc | artist | producttitle | tracktitle | outletname | format | territory | sales_period | units | royalty_amount_customer | sales_period_dt | year | month |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Sila navsegda | RUA3R2128055 | 7423703037037 | Eduard Dementyev | Luchina (it s Not the Wind That s Driving the ... | Luchina (it s Not the Wind That s Driving the ... | Amazon | Track | IT | 202,202.00 | 1 | 0.006605 | 2022-02-01 | 2022 | 2 |
| **1** | Sila navsegda | RUA3R2123705 | 7423739417452 | Эдуард Дементьев feat. Roma Skeptik | Ой, мороз, мороз. (Клубная версия) | Ой, мороз, мороз. (Клубная версия) | Tidal | Track | AR | 202,201.00 | 1 | 0.000875 | 2022-01-01 | 2022 | 1 |
| **2** | Sila navsegda | RUA3R2120520 | 7423714522560 | Eduard Dementyev | Black raven. Song of a Russian warrior. | Black raven. Song of a Russian warrior. | Amazon | Track | US | 202,202.00 | 1 | 0.014177 | 2022-02-01 | 2022 | 2 |
| **3** | Sila navsegda | RUA3R2128055 | 7423703037037 | Eduard Dementyev | Luchina (it s Not the Wind That s Driving the ... | Luchina (it s Not the Wind That s Driving the ... | Amazon | Track | IN | 202,202.00 | 1 | 0.001168 | 2022-02-01 | 2022 | 2 |
| **4** | Sila navsegda | RUA3R2123705 | 7423739417452 | Эдуард Дементьев feat. Roma Skeptik | Ой, мороз, мороз. (Клубная версия) | Ой, мороз, мороз. (Клубная версия) | Youtube | Track | RU | 202,201.00 | 0 | 0.003039 | 2022-01-01 | 2022 | 1 |

In [164]:
```
print('royalty_amount_customer', type(df['royalty_amount_customer'][5]))
```

```
royalty_amount_customer <class 'str'>
```

### 2.14 меняю тип данных со строчных на численные

In [165]:
```
df['royalty_amount_customer'] = df['royalty_amount_customer'].astype('float')

# проверяю тип данных
print('royalty_amount_customer', type(df['royalty_amount_customer'][5]))
```

```
royalty_amount_customer <class 'numpy.float64'>
```

In [166]:
```
df.head()
```

Out[166]:

| | labelname | isrc | ean_upc | artist | producttitle | tracktitle | outletname | format | territory | sales_period | units | royalty_amount_customer | sales_period_dt | year | month |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Sila navsegda | RUA3R2128055 | 7423703037037 | Eduard Dementyev | Luchina (it s Not the Wind That s Driving the ... | Luchina (it s Not the Wind That s Driving the ... | Amazon | Track | IT | 202,202.00 | 1 | 0.01 | 2022-02-01 | 2022 | 2 |
| **1** | Sila navsegda | RUA3R2123705 | 7423739417452 | Эдуард Дементьев feat. Roma Skeptik | Ой, мороз, мороз. (Клубная версия) | Ой, мороз, мороз. (Клубная версия) | Tidal | Track | AR | 202,201.00 | 1 | 0.00 | 2022-01-01 | 2022 | 1 |
| **2** | Sila navsegda | RUA3R2120520 | 7423714522560 | Eduard Dementyev | Black raven. Song of a Russian warrior. | Black raven. Song of a Russian warrior. | Amazon | Track | US | 202,202.00 | 1 | 0.01 | 2022-02-01 | 2022 | 2 |
| **3** | Sila navsegda | RUA3R2128055 | 7423703037037 | Eduard Dementyev | Luchina (it s Not the Wind That s Driving the ... | Luchina (it s Not the Wind That s Driving the ... | Amazon | Track | IN | 202,202.00 | 1 | 0.00 | 2022-02-01 | 2022 | 2 |
| **4** | Sila navsegda | RUA3R2123705 | 7423739417452 | Эдуард Дементьев feat. Roma Skeptik | Ой, мороз, мороз. (Клубная версия) | Ой, мороз, мороз. (Клубная версия) | Youtube | Track | RU | 202,201.00 | 0 | 0.00 | 2022-01-01 | 2022 | 1 |

### 2.15 данные подготовлены для анализа

# 3 Исследовательский анализ данных

## 3.1 Выявить самые привлекательные по цене

суммы units по площадкам

In [167]:
```
units_sum_by_outletname = df.groupby('outletname')['units'].sum().sort_values(ascending=False)
print(units_sum_by_outletname)
```

```
outletname
YouTube Music    400
VKONTAKTE        216
YouTube           69
BOOM APP          51
Yandex Music      35
Deezer            30
Apple Music       28
Spotify           18
ODNOKLASSNIKI     11
Youtube           11
Amazon             3
Tidal              1
Name: units, dtype: Int64
```

## 3.2 суммы units по royalti в увеличенном в 1000000 раз для удобства восприятия

In [168]:

```
units_sum_by_royaly = df.groupby('outletname')['royalty_amount_customer'].sum().sort_values(ascending=False) * 100 * 100
print(units_sum_by_royaly)
```

```
outletname
YouTube Music   12,330.06
Youtube            637.65
Apple Music        599.38
VKONTAKTE          441.66
BOOM APP           428.42
YouTube            417.20
Spotify            366.57
Yandex Music       358.13
Amazon             219.50
Deezer              77.70
ODNOKLASSNIKI       22.48
Tidal                8.75
Name: royalty_amount_customer, dtype: float64
```

### 3.3 вычисляю стоимость одного юнита по площадкам

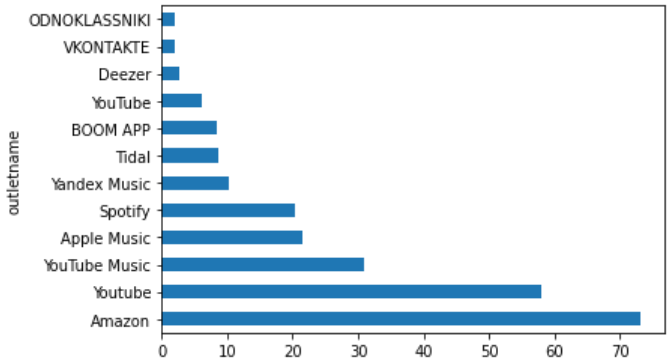
In [169]:

```
unit_price = units_sum_by_royaly / units_sum_by_outletname
print(type(unit_price))
unit_price = unit_price.sort_values(ascending = False)
print(unit_price)
```

```
<class 'pandas.core.series.Series'>
outletname
Amazon          73.17
Youtube         57.97
YouTube Music   30.83
Apple Music     21.41
Spotify         20.36
Yandex Music    10.23
Tidal            8.75
BOOM APP         8.40
YouTube          6.05
Deezer           2.59
VKONTAKTE        2.04
ODNOKLASSNIKI    2.04
dtype: Float64
```

### 3.4 построить график стоимости юнитов

In [170]:
```
1 unit_price.plot.barh();
```

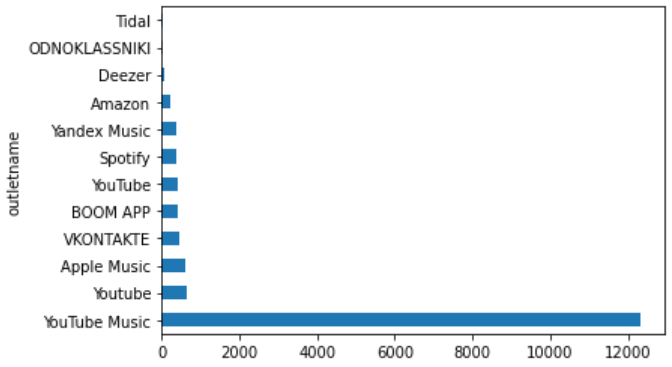



### 3.5 Самый высокооплачиваемый канал музыкального ритейла - **Amazon

Следом за ним следует Youtube

### 3.6 Выявить самые привлекательные по сумме royalty

In [171]:
```
1 units_sum_by_royaly.plot.barh();
```

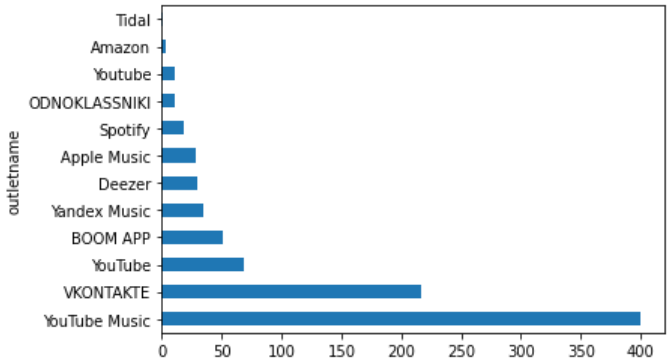



### 3.7 Самый доходный канал музыкального ритейла - YouTube Music

Следом за ним следует **YouTube**

### 3.8 Выявить площадки с наибольшим количеством прослушиваний

In [172]:
```
1 units_sum_by_outletname.plot.barh();
```




### 3.9 Канал с самым большим количеством прослушиваний - Youtube Music

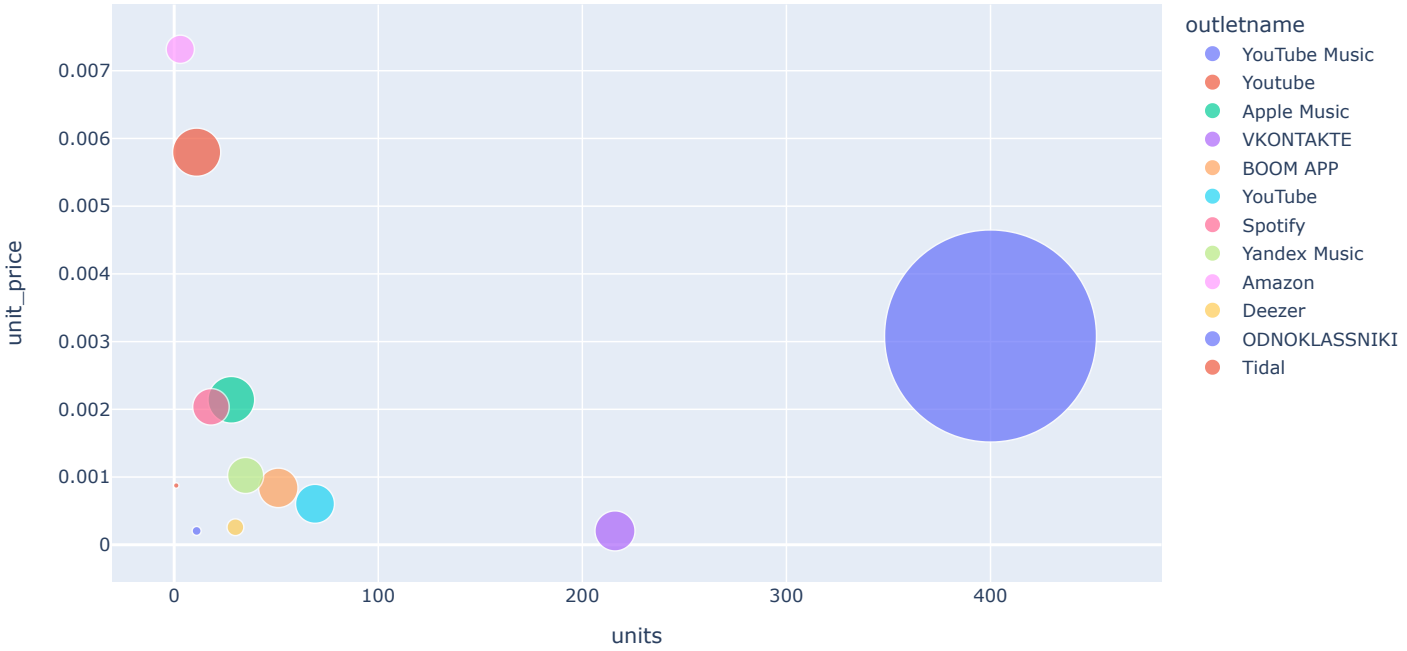
Следом за ним следует **ВКонтакте**

### 3.10 Посмотрим распределение

- цена юнита
- количество прослушиваний
  Размер маркера - **сумма** роялти

In [176]:

```
table_for_scatter = df.groupby(['outletname'])[['units', 'royalty_amount_customer']]\
    .sum().reset_index().sort_values('royalty_amount_customer', ascending=False)
# добавим колонку цены юнита
table_for_scatter['unit_price'] = table_for_scatter['royalty_amount_customer'] / table_for_scatter['units']
fig = px.scatter(table_for_scatter, x="units", y="unit_price", color="outletname",
                 size='royalty_amount_customer', size_max=100)
fig.show()
```




## 4 Самый выгодный канал - Youtube Music

Перспективный канал - **ВКонтаке**
Интересный канал **Apple Music**