

Департамент образования и науки города Москвы

**Государственное автономное образовательное учреждение высшего  
образования города Москвы**

**«Московский городской педагогический университет»**

Институт цифрового образования

Департамент информатики, управления и технологий

Башкатова Анна Денисовна

## **ОТЧЕТ**

по дисциплине «Инструменты для хранения и обработки больших данных»

**Тема: «01-1 Визуализация данных из CSV-файла. Индивидуальное  
задание Вариант 9»**

Направление подготовки (специальность) 38.03.05 – бизнес-информатика

Направленность (профиль) образовательной программы «Аналитика данных  
и эффективное управление»

Курс обучения: 3

Форма обучения: очная

Руководитель: Босенко Т. М.

Москва  
2023

**Ссылка на ДАШБОРД:** <https://datalens.yandex/785vhjfo18qx>

1. Просмотр данных Internet Prices around 200+ countries in 2022 (<https://www.kaggle.com/datasets/ramjasmaurya/1-gb-internet-price>)
2. Загрузка данных в colab.research.

[https://colab.research.google.com/drive/1sI0e\\_h5Gtn854nf2nrWikkHGcytb7ndz?usp=sharing](https://colab.research.google.com/drive/1sI0e_h5Gtn854nf2nrWikkHGcytb7ndz?usp=sharing)

PR1\_ИХОБД\_Башкатова.ipynb

Файл Изменить Вид Вставка Среда выполнения Инструменты Справка

+ Код + Текст

Подключиться повторно

```
[ ] import numpy as np
import pandas as pd

[ ] from google.colab import files
unuploaded = files.upload() #загрузка файла worldwide internet.csv

Выбор файла: worldwide internet.csv
• worldwide internet.csv(text/csv) - 20982 bytes, last modified: 15.02.2023 - 100% done
Saving worldwide internet.csv to worldwide internet (4).csv

[ ] df = pd.read_csv('worldwide internet (4).csv', index_col=0) #преобразование файла в датафрейм

[ ] df.head(3)
```

S.NO	Country code	Country	Continental region	NO. OF Internet Plans	Average price of 1GB (USD)	Cheapest 1GB for 30 days (USD)	Most expensive 1GB (USD)	Average price of 1GB (USD at the start of 2021)	Average price of 1GB (USD - at start of 2020)	Internet users	Population	Avg \n(Mbit/s)Ookla
0	IL	Israel	NEAR EAST	27.0	0.05	0.02	20.95	0.11	0.9	6,788,737	8,381,516	28.01
1	KG	Kyrgyzstan	CIS (FORMER USSR)	20.0	0.15	0.10	7.08	0.21	0.27	2,309,235	6,304,030	16.30
2	FJ	Fiji	OCEANIA	18.0	0.19	0.05	0.85	0.59	3.57	452,479	883,483	25.99

PR1\_ИХОБД\_Башкатова.ipynb

Файл Изменить Вид Вставка Среда выполнения Инструменты Справка

+ Код + Текст

Подключиться повторно

```
[ ] df.rename(columns = {'Avg \n(Mbit/s)Ookla':'Avg (Mbit/s)Ookla'}, inplace = True)

df = df.replace(' ','', regex=True)

df.head(3)

[ ] df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 242 entries, 0 to 243
Data columns (total 12 columns):
 #   column                                Non-Null Count  Dtype
---  -
 0   Country code                        242 non-null   object
 1   Country                            242 non-null   object
 2   Continental region                  242 non-null   object
 3   NO. OF Internet Plans               231 non-null   float64
 4   Average price of 1GB (USD)          242 non-null   object
 5   Cheapest 1GB for 30 days (USD)      231 non-null   float64
 6   Most expensive 1GB (USD)            231 non-null   float64
 7   Average price of 1GB (USD at the start of 2021)  231 non-null   object
 8   Average price of 1GB (USD - at start of 2020)  231 non-null   object
 9   Internet users                      212 non-null   object
10   Population                          210 non-null   object
11   Avg (Mbit/s)Ookla                   141 non-null   float64
dtypes: float64(4), object(8)
memory usage: 24.6+ KB
```

PR1 ИХОБД\_Башкатова.ipynb

Файл Изменить Вид Вставка Среда выполнения Инструменты Справка Сохранение...

Комментировать Поделиться Настройки Профиль

Код Текст

Подключиться повторно

df.isnull().sum()

Country code 0  
Country 0  
Continental region 0  
NO. OF Internet Plans 11  
Average price of 1GB (USD) 0  
Cheapest 1GB for 30 days (USD) 11  
Most expensive 1GB (USD) 11  
Average price of 1GB (USD at the start of 2021) 11  
Average price of 1GB (USD - at start of 2020) 11  
Internet users 30  
Population 32  
Avg (Mbit/s)Ookla 101  
dtype: int64

Из-за высокого числа NaN в переменной Avg (Мбит/с)Ookla мы исключим ее из анализа.

df.drop(["Avg (Mbit/s)Ookla"], axis = 1, inplace=True)

df.isnull().sum()

Country code 0  
Country 0  
Continental region 0  
NO. OF Internet Plans 11  
Average price of 1GB (USD) 0  
Cheapest 1GB for 30 days (USD) 11  
Most expensive 1GB (USD) 11  
Average price of 1GB (USD at the start of 2021) 11  
Average price of 1GB (USD - at start of 2020) 11  
Internet users 30

PR1 ИХОБД\_Башкатова.ipynb

Файл Изменить Вид Вставка Среда выполнения Инструменты Справка Изменения сохранены

Комментировать Поделиться Настройки Профиль

Код Текст

Подключиться повторно

df.dropna(inplace=True)  
df.info()

<class 'pandas.core.frame.DataFrame'>  
Int64Index: 203 entries, 0 to 232  
Data columns (total 11 columns):  
# Column Non-Null Count Dtype  
---  
0 Country code 203 non-null object  
1 Country 203 non-null object  
2 Continental region 203 non-null object  
3 NO. OF Internet Plans 203 non-null float64  
4 Average price of 1GB (USD) 203 non-null object  
5 Cheapest 1GB for 30 days (USD) 203 non-null float64  
6 Most expensive 1GB (USD) 203 non-null float64  
7 Average price of 1GB (USD at the start of 2021) 203 non-null object  
8 Average price of 1GB (USD - at start of 2020) 203 non-null object  
9 Internet users 203 non-null object  
10 Population 203 non-null object  
dtypes: float64(3), object(8)  
memory usage: 19.0+ KB

df[['Internet users']] = df[['Internet users']].astype(int)  
df[['Population']] = df[['Population']].astype(int)

df['Average price of 1GB (USD at the start of 2021)'].replace("NO PACKAGES", '0', inplace=True)  
df['Average price of 1GB (USD - at start of 2020)'].replace("NO PACKAGES", '0', inplace=True)

df[['Average price of 1GB (USD at the start of 2021)']] = df[['Average price of 1GB (USD at the start of 2021)']].astype(float)  
df[['Average price of 1GB (USD - at start of 2020)']] = df[['Average price of 1GB (USD - at start of 2020)']].astype(float)

PR1 ИХОБД\_Башкатова.ipynb

Файл Изменить Вид Вставка Среда выполнения Инструменты Справка Изменения сохранены

Комментировать Поделиться Настройки Профиль

Код Текст

Подключиться повторно

df.drop(df[df['Average price of 1GB (USD)'] == 'NO PROVIDERS'].index,inplace = True)  
df.drop(df[df['Average price of 1GB (USD)'] == 'HYPERINFLATION'].index,inplace = True)  
df.drop(df[df['Average price of 1GB (USD)'] == "Prices listed in non-convertible 'units'"].index,inplace = True)  
df.drop(df[df['Average price of 1GB (USD - at start of 2020)'] == 'NO PACKAGES'].index,inplace = True)  
df.drop(df[df['Average price of 1GB (USD at the start of 2021)'] == 'NO PACKAGES'].index,inplace = True)

df.info()

<class 'pandas.core.frame.DataFrame'>  
Int64Index: 203 entries, 0 to 232  
Data columns (total 11 columns):  
# Column Non-Null Count Dtype  
---  
0 Country code 203 non-null object  
1 Country 203 non-null object  
2 Continental region 203 non-null object  
3 NO. OF Internet Plans 203 non-null float64  
4 Average price of 1GB (USD) 203 non-null object  
5 Cheapest 1GB for 30 days (USD) 203 non-null float64  
6 Most expensive 1GB (USD) 203 non-null float64  
7 Average price of 1GB (USD at the start of 2021) 203 non-null float64  
8 Average price of 1GB (USD - at start of 2020) 203 non-null float64  
9 Internet users 203 non-null int64  
10 Population 203 non-null int64  
dtypes: float64(5), int64(2), object(4)  
memory usage: 27.1+ KB

PR1\_IХOБД Башкатова.ipynb

КомментироватьПоделитьсяНастроить

ФайлИзменитьВидВставкаСреда выполненияИнструментыСправкаИзменения сохранены

+ Код+ Текст

Подключиться повторно

```
[ ] df.reset_index(drop=True, inplace=True)
```

```
[ ] df
```

	Country code	Country	Continental region	NO. OF Internet Plans	Average price of 1GB (USD)	Cheapest 1GB for 30 days (USD)	Most expensive 1GB (USD)	Average price of 1GB (USD at the start of 2021)	Average price of 1GB (USD - at start of 2020)	Internet users	Population
0	IL	Israel	NEAR EAST	27.0	0.05	0.02	20.95	0.11	0.90	6788737	8381516
1	KG	Kyrgyzstan	CIS (FORMER USSR)	20.0	0.15	0.10	7.08	0.21	0.27	2309235	6304030
2	FJ	Fiji	OCEANIA	18.0	0.19	0.05	0.85	0.59	3.57	452479	883483
3	IT	Italy	WESTERN EUROPE	29.0	0.27	0.09	3.54	0.43	1.73	50540000	60627291
4	SD	Sudan	SUB-SAHARAN AFRICA	33.0	0.27	0.03	0.92	0.63	0.68	12512639	41801533
...	...	...	...	...	...	...	...	...	...	...	...

PR1\_IХOБД Башкатова.ipynb

КомментироватьПоделитьсяНастроить

ФайлИзменитьВидВставкаСреда выполненияИнструментыСправкаИзменения сохранены

+ Код+ Текст

Подключиться повторно

```
[ ] df[['Average price of 1GB (USD)']] = df[['Average price of 1GB (USD)']].astype(float)
```

```
[ ] df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 203 entries, 0 to 202
Data columns (total 11 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Country code                          203 non-null    object
1   Country                              203 non-null    object
2   Continental region                    203 non-null    object
3   NO. OF Internet Plans                  203 non-null    float64
4   Average price of 1GB (USD)             203 non-null    float64
5   Cheapest 1GB for 30 days (USD)         203 non-null    float64
6   Most expensive 1GB (USD)               203 non-null    float64
7   Average price of 1GB (USD at the start of 2021)  203 non-null    float64
8   Average price of 1GB (USD - at start of 2020)  203 non-null    float64
9   Internet users                         203 non-null    int64
10  Population                             203 non-null    int64
dtypes: float64(6), int64(2), object(3)
memory usage: 17.6+ KB
```

```
[ ] df[df['Country code'] == 'AW']
```

	Country code	Country	Continental region	NO. OF Internet Plans	Average price of 1GB (USD)	Cheapest 1GB for 30 days (USD)	Most expensive 1GB (USD)	Average price of 1GB (USD at the start of 2021)	Average price of 1GB (USD - at start of 2020)	Internet users	Population
154	AW	Aruba (Netherlands)	CARIBBEAN	17.0	4.44	0.74	8.96	9.11	5.56	15877494	17059560
...	...	Aruba	CARIBBEAN	17.0	4.44	0.74	8.96	9.11	5.56	15877494	17059560

PR1\_IХOБД Башкатова.ipynb

КомментироватьПоделитьсяНастроить

ФайлИзменитьВидВставкаСреда выполненияИнструментыСправкаИзменения сохранены

+ Код+ Текст

Подключиться повторно

```
[ ] df.drop(index=154, inplace=True)
```

```
[ ] df[df['Country code']!= 'AW']
```

	Country code	Country	Continental region	NO. OF Internet Plans	Average price of 1GB (USD)	Cheapest 1GB for 30 days (USD)	Most expensive 1GB (USD)	Average price of 1GB (USD at the start of 2021)	Average price of 1GB (USD - at start of 2020)	Internet users	Population
155	AW	Aruba (Netherlands)	CARIBBEAN	17.0	4.44	0.74	8.96	9.11	5.56	102285	105845

🔗

```
[ ] df.reset_index(drop=True, inplace=True)
```

```
[ ] df.rename(
    columns={
        "Continental region": "Region",
        "NO. OF Internet Plans": "Internet Plans",
        "Average price of 1GB (USD)": "Average Price",
        "Cheapest 1GB for 30 days (USD)": "Cheapest Price",
        "Most expensive 1GB (USD)": "Most expensive Price",
        "Average price of 1GB (USD at the start of 2021)": "Avg price 2021",
        "Average price of 1GB (USD - at start of 2020)": "Avg price 2020",
        "Internet users": "Internet Users",
    }, inplace=True)
```

PR1\_ИХОБД\_Башкатова.ipynb

Файл Изменить Вид Вставка Среда выполнения Инструменты Справка Изменения сохранены

Комментировать Поделиться

+ Код + Текст

Подключиться повторно

- Country - Название страны.
- Country code - Код страны.
- Region - Регион.
- Internet Plans - количество интернет-услуг, предлагаемых для покупки в стране.
- Average Price - средняя цена за 1 ГБ в US dollars
- Cheapest Price - самая низкая цена за 1 ГБ в US dollars
- Most expensive Price - самая высокая цена за 1 ГБ в US dollars
- Avg price 2021 - средняя цена за 1 ГБ в US dollars в 2021
- Avg price 2020 - средняя цена за 1 ГБ в US dollars в 2020
- Internet Users - кол-во пользователей интернета
- Population - количество людей, проживающих в стране

```
[ ] df.head(10)
```

	Country code	Country	Region	Internet Plans	Average Price	Cheapest Price	Most expensive Price	Avg price 2021	Avg price 2020	Internet Users	Population
0	IL	Israel	NEAR EAST	27.0	0.05	0.02	20.95	0.11	0.90	6788737	8381516
1	KG	Kyrgyzstan	CIS (FORMER USSR)	20.0	0.15	0.10	7.08	0.21	0.27	2309235	6304030
2	FJ	Fiji	OCEANIA	18.0	0.19	0.05	0.85	0.59	3.57	452479	883483
3	IT	Italy	WESTERN EUROPE	29.0	0.27	0.09	3.54	0.43	1.73	50540000	60627291
4	SD	Sudan	SUB-SAHARAN AFRICA	33.0	0.27	0.03	0.92	0.63	0.68	12512639	41801533
5	RU	Russia	CIS (FORMER USSR)	22.0	0.29	0.13	1.86	0.52	0.91	124000000	145734038
6	MD	Moldova	EASTERN EUROPE	18.0	0.32	0.07	2.79	1.12	2.82	3083783	4051944
7	BD	Bangladesh	ASIA (EX. NEAR EAST)	60.0	0.34	0.11	2.22	0.70	0.99	129180000	166303498

PR1\_ИХОБД\_Башкатова.ipynb

Файл Изменить Вид Вставка Среда выполнения Инструменты Справка Изменения сохранены

Комментировать Поделиться

+ Код + Текст

Подключиться повторно

```
[ ] df.groupby('Country').mean()
```

Country	Internet Plans	Average Price	Cheapest Price	Most expensive Price	Avg price 2021	Avg price 2020	Internet Users	Population
Afghanistan	35.0	1.02	0.38	2.12	1.55	1.60	4068194.0	37171921.0
Albania	32.0	2.80	0.72	96.39	2.83	5.28	2105339.0	2882740.0
Algeria	20.0	0.51	0.16	2.24	0.65	5.15	26350000.0	42228408.0
Andorra	7.0	7.49	1.06	65.07	9.54	12.58	76095.0	77006.0
Angola	20.0	1.61	1.03	3.21	5.29	7.95	4271053.0	30809787.0
...	...	...	...	...	...	...	...	...
Uzbekistan	60.0	0.60	0.01	23.75	1.34	3.27	16692456.0	32476244.0
Vanuatu	23.0	6.06	0.75	28.23	4.25	0.00	71050.0	292680.0
Vietnam	46.0	0.49	0.07	5.55	0.57	1.31	68172134.0	95545962.0
Yemen	17.0	15.98	0.40	39.94	15.98	15.73	7548512.0	28498683.0
Zambia	60.0	1.13	0.01	6.80	1.36	2.25	4760715.0	17351708.0

201 rows x 8 columns

```
[ ] from google.colab import files
df.to_csv('worldwide internet (4).csv')
files.download('worldwide internet (4).csv')
```

### 3. Загружаем данные в datalens.yandex.ru

<https://datalens.yandex/785yhjfo18qx>

### 4. Создаем подключение

Users / bashkatova2002 / Новое подключение

Создать подключение

← File

+ Загрузить файлы ?

worldwide internet (4).csv ...

Кодировка utf-8

Разделитель Запятая

Заголовок столбцов Да Нет

Поиск по столбцу

#	field1	Country code	Country	Region	Internet Plans	Average Price	Cheapest Price	Most expensive Price	Avg price 2021
0	IL	Israel	NEAR EAST	27.0	0.05	0.02	20.95	0.11	
1	KG	Kyrgyzstan	CIS (FORMER USSR)	20.0	0.15	0.1	7.08	0.21	
2	FJ	Fiji	OCEANIA	18.0	0.19	0.05	0.85	0.59	
3	IT	Italy	WESTERN EUROPE	29.0	0.27	0.09	3.54	0.43	
4	SD	Sudan	SUB-SAHARAN AFRICA	33.0	0.27	0.03	0.92	0.63	
5	RU	Russia	CIS (FORMER USSR)	22.0	0.29	0.13	1.86	0.52	
6	MD	Moldova	EASTERN EUROPE	18.0	0.32	0.07	2.79	1.12	
7	BD	Bangladesh	ASIA (EX. NEAR EAST)	60.0	0.34	0.11	2.22	0.7	
8	LK	Sri Lanka	ASIA (EX. NEAR EAST)	60.0	0.38	0.0	5.53	0.51	
9	CL	Chile	SOUTH AMERICA	59.0	0.39	0.24	1.83	0.71	
10	FR	France	WESTERN EUROPE	45.0	0.41	0.09	118.2	0.81	
11	ID	Indonesia	ASIA (EX. NEAR EAST)	53.0	0.42	0.17	2.94	0.64	
12	BY	Belarus	CIS (FORMER USSR)	14.0	0.43	0.02	16.62	0.89	

5. Создаем датасет.

Users / bashkatova2002 / Новый датасет

Создать чарт

Сохранить

Источники Поля Параметры Фильтрация

Обновить поля Предпросмотр + Добавить поле Имя поля

#	Имя	Источник поля	Тип	Агрегация	Описание
1	Index	csv.field1	Целое число	Нет	
2	Код страны	csv.country_code	Строка	Нет	
3	Страна	csv.country	Строка	Нет	
4	Регион	csv.region	Строка	Нет	
5	Интернет план	csv.internet_plans	Дробное число	Нет	
6	Средняя цена	csv.average_price	Дробное число	Нет	
7	Мин цена	csv.cheapest_price	Дробное число	Нет	
8	Мак цена	csv.most_expensive_price	Дробное число	Нет	
9	Средняя цена 2021	csv.avg_price_2021	Дробное число	Нет	
10	Средняя цена 2020	csv.avg_price_2020	Дробное число	Нет	
11	Кол-во пользователей интернета	csv.internet_users	Целое число	Нет	

Предпросмотр

Количество строк: 10 не больше 1 000

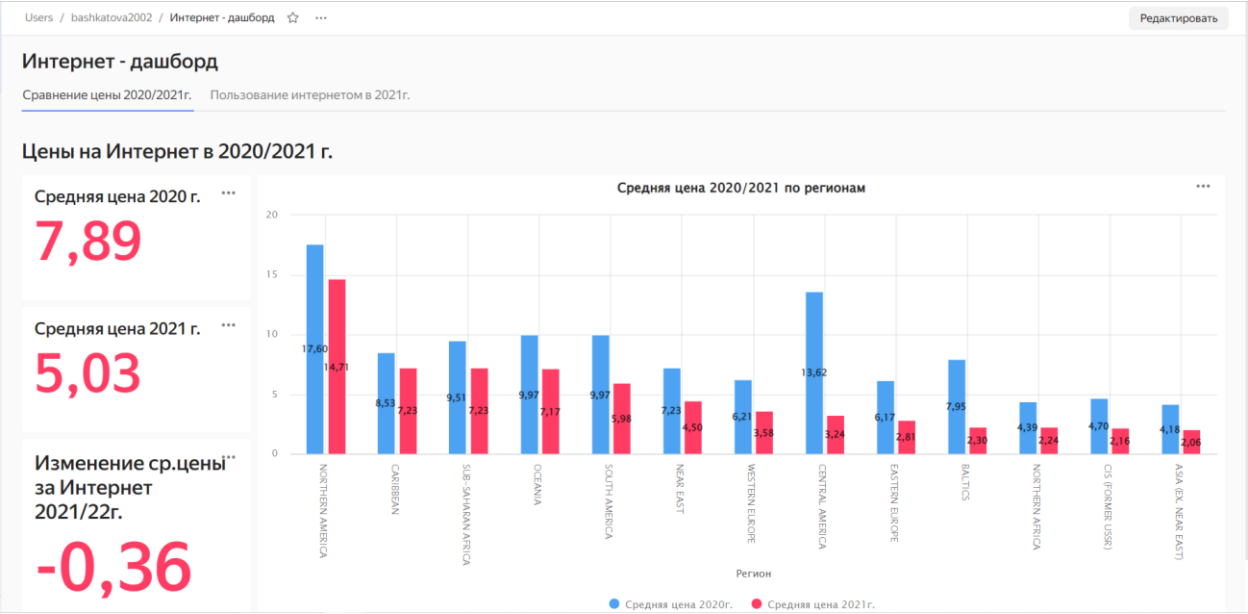
#	field1	Country code	Country	Region	Internet Plans	Average Price	Cheapest Price	Most expensive Price	Avg price 2021	Avg price 2020	Internet Users	Population
7	6	MD	Moldova	EASTERN EUROPE	18.0	0.32	0.07	2.79	1.12	2.82	3083783	4051944
8	7	BD	Bangladesh	ASIA (EX. NEAR EAST)	60.0	0.34	0.11	2.22	0.7	0.99	129180000	166303498
9	8	LK	Sri Lanka	ASIA (EX. NEAR EAST)	60.0	0.38	0.0	5.53	0.51	0.78	7121116	21228763
10	9	CL	Chile	SOUTH AMERICA	59.0	0.39	0.24	1.83	0.71	1.87	14864456	18729160

6. Создаем чарты (ниже несколько примеров).

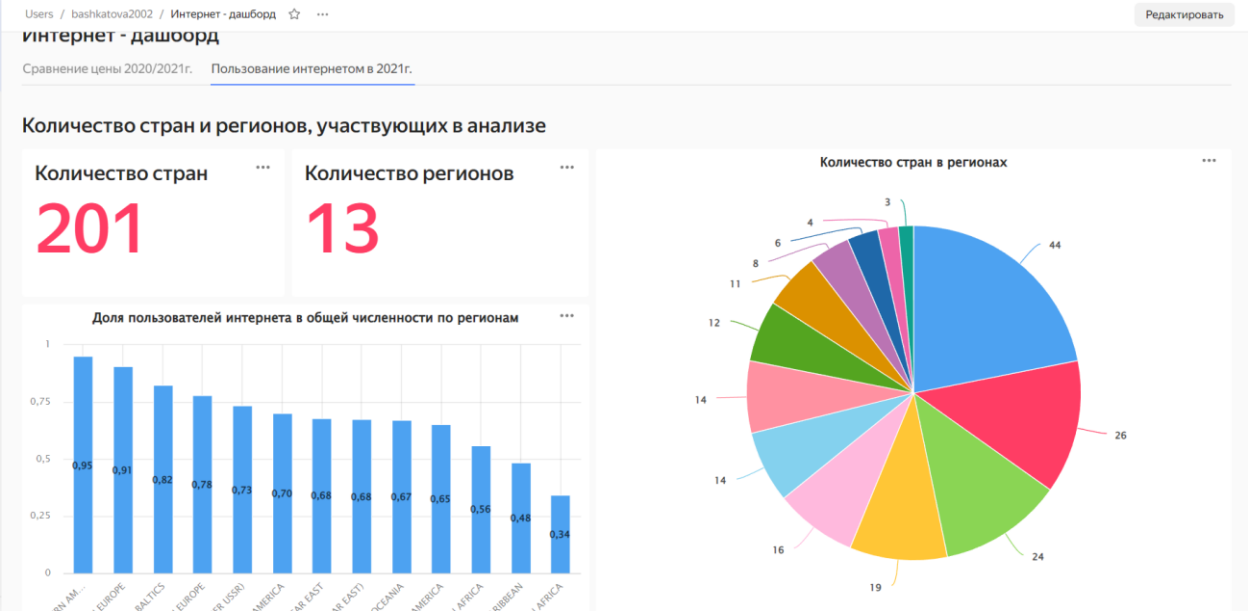


7. Далее размещаем их на дашборде «Интернет-дашборд». Дашборд состоит из двух вкладок – «Сравнение цены 2020/2021г.» и «Пользование интернетом в 2021г.».

Первая вкладка:



Вторая вкладка:



Ниже расположен фильтр, где можно выбрать регион.



