



МИНОБРНАУКИ РОССИИ

**Федеральное государственное бюджетное образовательное
учреждение высшего образования
«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ»**

Факультет информатики и прикладной математики
Кафедра прикладной математики и экономико-математических методов

ОТЧЁТ

по дисциплине:

«Модели экономической динамики»

на тему:

«Кластеризация экономик. Динамика экономики Японии.»

Направление: 01.03.02

Обучающийся: Бронников Егор Игоревич

Группа: ПМ-1901

Санкт-Петербург
2022

Задание

Выполнить кластеризацию экономик, используя следующий набор показателей (отдельно за 2019, 2013 и с использованием среднего темпового показателя на временном интервале с 2013 по 2019):

- темпы роста ВВП;
- подушевой ВВП, темп роста ВВП;
- подушевой ВВП, темпы роста ВВП, темп инфляции.

Проанализировать полученное распределение по кластерам, миграцию между кластерами, дать содержательную интерпретацию результатов кластеризации.

Решение

Данные были взяты с сайта Всемирного банка.

Первый этап заключался в выделении объединений стран и исключении их из исходного набора данных. Был создан файл с наименованиями этих объединений. (Рисунок 1)

▼ Список объединений стран

```
[3]: with open("./data/CountryAlliances", "r") as f:
      country_alliances = list(map(lambda x: x[:-1], f.readlines()))
```

Рис. 1: Заполнения списка объединений стран

Далее были загружены датасеты с темпами роста ВВП (%), подушевого ВВП (текущий US\$), темпами инфляции (%). (Рисунок 2)

▼ Темы роста ВВП (%)

```
[4]: df_gdp_growth_data = pd.read_csv("./data/GDP_growth.csv", skiprows=4)
df_gdp_growth_data = df_gdp_growth_data.loc[~df_gdp_growth_data["Country Name"].isin(country_alliances)]
df_gdp_growth_data.head()

# df_gdp_growth_data = wb.data.DataFrame("NY.GDP.MKTP.KD.ZG", labels=True)
# df_gdp_growth_data.head()
```

	Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	1962	1963	1964	1965	...
0	Aruba	ABW	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	NaN	NaN	NaN	NaN	NaN	NaN	...
1	Afghanistan	AFG	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	NaN	NaN	NaN	NaN	NaN	NaN	...
2	Angola	AGO	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	NaN	NaN	NaN	NaN	NaN	NaN	...
3	Albania	ALB	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	NaN	NaN	NaN	NaN	NaN	NaN	...
4	Andorra	AND	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	NaN	NaN	NaN	NaN	NaN	NaN	...

5 rows x 66 columns

Рис. 2: Пример подгруженных данных

На следующем этапе было решено начать с *кластеризации стран на основании темпов роста ВВП за 2019 год*.

Были отобраны соответствующие данные. Также были удалены страны, которые не содержат данных. (Рисунок 3)

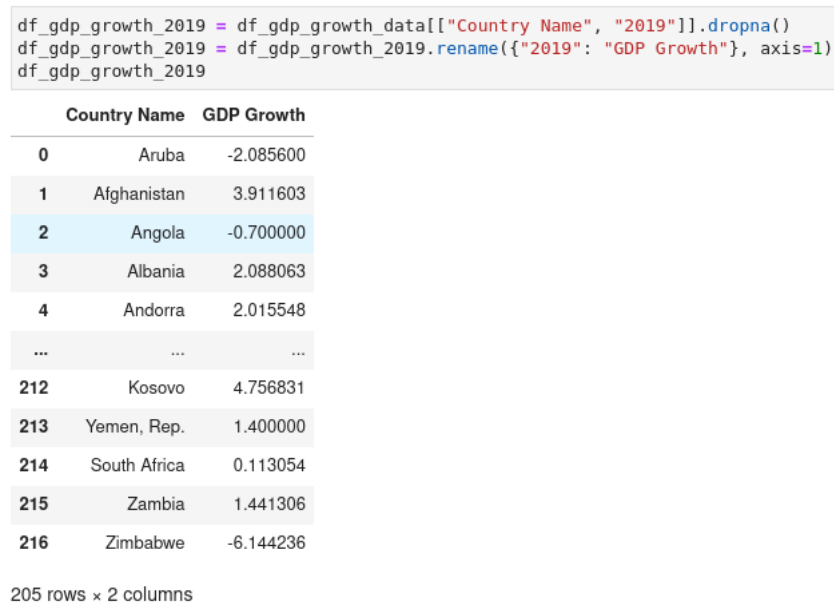


Рис. 3: Пример отбора необходимых данных

Далее была построена дендрограмма и на основании получившегося результата было решено взять для рассмотрения 5 кластеров, для более качественной интерпретации. (Рисунок 4)

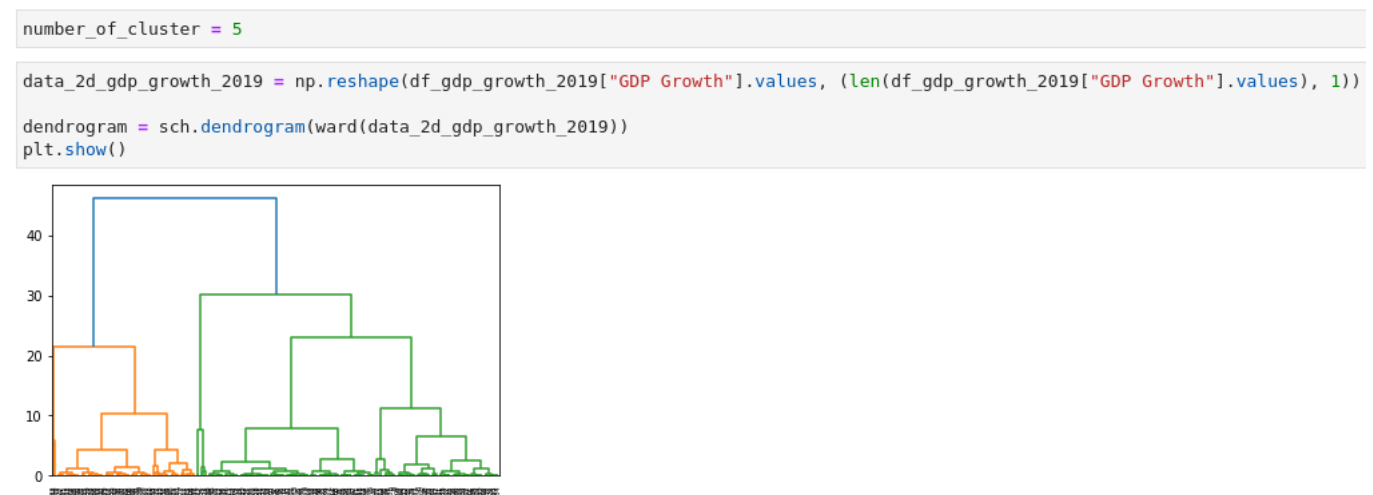


Рис. 4: Построение дерева дендрограммы

Далее происходил процесс кластеризации методом *Agglomerative Clustering*. (Рисунок 5)

```
hc = AgglomerativeClustering(n_clusters=number_of_cluster, affinity="euclidean", linkage="ward")
y_hc = hc.fit_predict(data_2d_gdp_growth_2019)

dict_names_gdp_2019 = {i: [] for i in range(number_of_cluster)}
for i in range(len(y_hc)):
    dict_names_gdp_2019[y_hc[i]].append(df_gdp_growth_2019["Country Name"].values[i])
```

Рис. 5: Кластеризация данных

После чего можно вывести название стран, которые попали в конкретный кластер. (Рисунок 6)

```
df_names_gdp_2019 = pd.DataFrame.from_dict(dict_names_gdp_2019, orient="index").transpose()
df_names_gdp_2019.head()
```

	0	1	2	3	4
0	Aruba	Armenia	Equatorial Guinea	Afghanistan	Timor-Leste
1	Angola	Antigua and Barbuda	Lebanon	Albania	Tuvalu
2	Argentina	Benin	Libya	Andorra	None
3	American Samoa	Burkina Faso	Northern Mariana Islands	United Arab Emirates	None
4	Austria	Bangladesh	Zimbabwe	Australia	None

Рис. 6: Распределение стран по кластерам (названия)

Также можно вывести значение темпов роста ВВП стран в кластерах. (Рисунок 7)

```
dict_vals_gdp_2019 = {i: [df_gdp_growth_2019.loc[df_gdp_growth_2019["Country Name"] == df_
df_vals_gdp_2019 = pd.DataFrame.from_dict(dict_vals_gdp_2019, orient="index").transpose()
df_vals_gdp_2019.head()
```

	0	1	2	3	4
0	-2.085600	7.600000	-5.481823	3.911603	19.535810
1	-0.700000	4.862236	-6.914925	2.088063	13.793103
2	-2.025934	6.865687	-11.195695	2.015548	NaN
3	-0.487805	5.688115	-11.142587	3.411539	NaN
4	1.491211	7.881915	-6.144236	2.113149	NaN

Рис. 7: Распределение стран по кластерам (значения)

Можно выделить численные характеристики каждого получившегося кластера: количество стран в кластере, минимальное значение, максимальное значение и среднее значение. (Рисунок 8)

```
aggregate = {i: ["count", "min", "max", "mean"] for i in range(number_of_cluster)}
df_vals_gdp_2019.agg(aggregate)
```

	0	1	2	3	4
count	57.000000	64.000000	5.000000	77.000000	2.000000
min	-3.779889	4.331735	-11.195695	1.671944	13.793103
max	1.491211	9.460598	-5.481823	4.188260	19.535810
mean	-0.090235	5.739661	-8.175853	2.740411	16.664457

Рис. 8: Численные характеристики кластеров

В данном случае можно выделить, что в 4 кластер попали страны, которые имеют очень высокие темпы роста ВВП, в 1 кластер попали страны с высоким темпом роста ВВП, в 3 кластер попали страны со средними темпами роста ВВП, в 0 кластер попали страны с низким темпом роста ВВП, а во 2 кластер попали страны с очень низким темпом роста ВВП.

Также получившиеся данные сохраняются в *Excel*-файл. На 1 листе будут отображены наименования стран распределённые по кластерам, а на 2 листе будут отражены соответствующие им показатели. (Рисунок 9)

```
save_as_xlsx("./results/gdp_growth_2019.xlsx",
             ["Country Names", "GDP growth"],
             [df_names_gdp_2019, df_vals_gdp_2019])
```

Рис. 9: Сохранение получившихся результатов в Excel-файле

Далее данные шаги повторяются для каждого из годов и соответствующих объединений показателей как указано в условии задачи.

Единственное что стоит отметить, что при рассмотрении факторов, которые имеют разные единицы измерения (% и текущие US\$), перед кластеризацией стоит провести процесс стандартизации. Также стоит отметить, что когда мы рассматриваем средние показатели с 2013 года по 2019 год стоит рассматривать среднее геометрическое показателей если соответствующая величина выражена в % соотношении.

На основании получившихся данных, была рассмотрена миграция стран по кластерам.

В качестве примера рассмотрим миграцию стран по темпам роста ВВП среди стран, которые показали средний темпы роста в 2013 году.

Страны, оставшиеся в кластере средних темпов роста ВВП. (Рисунок 10)

```
list(set(df_names_gdp_2013[0]) & set(df_names_gdp_2019[3]))
```

- 'Colombia',
- 'South Asia',
- 'United Kingdom',
- 'Panama',
- 'North Macedonia',
- 'Canada',
- 'Comoros',
- 'Bolivia',
- 'India',
- 'New Zealand',
- 'Jordan',
- 'El Salvador',
- 'Guam',
- 'Thailand',
- 'Eswatini',
- 'United States',
- 'Afghanistan',
- 'Latvia',
- 'Guatemala',
- 'Burundi',
- 'Gabon',
- 'Pakistan',
- 'Lesotho',
- 'Chad',
- 'Korea, Rep.',
- 'North America',
- 'Cameroon',
- 'Honduras',
- 'Mozambique',
- 'Australia',
- 'Luxembourg',
- 'Romania',
- 'Bahrain',
- 'Sao Tome and Principe',
- 'Morocco',
- 'Seychelles',
- 'Iceland',
- 'Azerbaijan',
- 'Nigeria',
- 'Montenegro'

Рис. 10: Оставшиеся в своём кластере

Страны, мигрировавшие в кластер с высокими темпами роста ВВП. (Рисунок 11)

```
list(set(df_names_gdp_2013[0]) & set(df_names_gdp_2019[1]))
```

['Burkina Faso',
'Uganda',
'Serbia',
'Marshall Islands',
'Vietnam',
'Senegal',
'Rwanda',
'Kazakhstan',
'Guinea-Bissau',
'Bangladesh',
'St. Kitts and Nevis',
'Mali',
'Niger',
'Guinea',
'Egypt, Arab Rep.',
'Lithuania',
'Malaysia',
'Indonesia',
'Hungary',
'Bhutan',
'Armenia',
'Togo',
'Malawi',
'Gambia, The',
'Madagascar',
'Tanzania',
'Guyana',
'Malta',
'Georgia',
'Mauritania',
'Nepal',
'Papua New Guinea',
'Kosovo',
'Kenya',
'Dominican Republic',
'Philippines']

Рис. 11: Мигрировавшие в кластер с высокими темпами роста

Страны, мигрировавшие в кластер с очень высокими темпами роста ВВП. (Рисунок 12)

```
list(set(df_names_gdp_2013[0]) & set(df_names_gdp_2019[4]))
```

['Tuvalu', 'Timor-Leste']

Рис. 12: Мигрировавшие в кластер с очень высокими темпами роста

Страны, мигрировавшие в кластер с очень низкими темпами роста ВВП. (Рисунок 13)

```
list(set(df_names_gdp_2013[0]) & set(df_names_gdp_2019[2]))
```

['Northern Mariana Islands', 'Zimbabwe', 'Lebanon']

Рис. 13: Мигрировавшие в кластер с очень низкими темпами роста

Страны, мигрировавшие в кластер с низкими темпами роста ВВП. (Рисунок 14)

```
list(set(df_names_gdp_2013[0]) & set(df_names_gdp_2019[0]))
```

```
['Oman',  
'Brazil',  
'Grenada',  
'Fiji',  
'Hong Kong SAR, China',  
'Qatar',  
'Ecuador',  
'Aruba',  
'Haiti',  
'Kiribati',  
'Chile',  
'Isle of Man',  
'Namibia',  
'St. Vincent and the Grenadines',  
'Saudi Arabia',  
'Singapore',  
'Japan',  
'Solomon Islands',  
'Switzerland',  
'Zambia',  
'Argentina',  
'Trinidad and Tobago',  
'Algeria',  
'Yemen, Rep.',  
'Sudan',  
'Angola',  
'Uruguay',  
'Cuba',  
'Nicaragua',  
'Suriname',  
'South Africa',  
'Tunisia']
```

Рис. 14: Мигрировавшие в кластер с низкими темпами роста

Аналогичным образом были получены результаты для всех остальных миграций по показателям и по их принадлежности к конкретному кластеру.

Задание