

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Московский институт электроники и математики им. А.Н. Тихонова

Шембель Даниил Альбертович, группа БИТ213

Модульная работа №1

Семестр 2

по дисциплине «Прикладной Статистический
анализ данных»

Дата сдачи отчета: 5.04.23

Москва 2023 г.

Оглавление

Тема.....	3
Актуальность	3
Исследуемые показатели	3
Вступление	4
Предварительный анализ	4
Корреляционный анализ	5
Кластерный анализ	6
Выводы	9
Прикрепленные файлы.....	9

Тема

В качестве темы было выбрано нахождение зависимости между темпом роста ВВП стран, предоставивших статистические данные в Европейскую экономическую комиссию ООН, от ряда макроэкономических переменных.

Актуальность

Изучение состояния экономики мировых держав является важной задачей в современном мире. Это позволяет получить представление о тенденциях развития мировой экономики, определить наиболее перспективные направления для инвестиций, разработки экономической политики, а также прогнозировать возможные экономические кризисы и риски. Исследование состояний экономики мировых держав может включать в себя анализ таких показателей, как ВВП, уровень безработицы, индексы потребительских цен, внешний баланс, торговый оборот и другие. Полученные данные могут быть использованы не только экономистами и финансистами, но и политиками, научными исследователями и общественностью в целом для принятия важных решений в области экономики, социальной политики и развития стран.

Исследуемые показатели

1. Индекс потребительских цен, 2010 г.=100: это показатель, который отражает уровень инфляции в стране, где базовый год - 2010 год.
2. Расходы на конечное потребление на душу населения в долл. США: это показатель, который отражает средние ежегодные расходы на каждого человека в стране.
3. Внешний баланс товаров и услуг, % от ВВП: это показатель, который отражает разницу между экспортом и импортом товаров и услуг в процентном отношении к ВВП страны.
4. Экспорт товаров и услуг, доля в процентах от ВВП: это показатель, который отражает долю экспорта товаров и услуг в ВВП страны.
5. Импорт товаров и услуг, доля в процентах от ВВП: это показатель, который отражает долю импорта товаров и услуг в ВВП страны.
6. Обменный курс, единиц национальной валюты за долл. США: это показатель, который отражает стоимость единицы национальной валюты относительно доллара США.
7. Уровень безработицы, %: это показатель, который отражает долю безработных в экономике страны.
8. ВВП на душу населения в ценах и ППС(США) 2010 г., темп роста: это показатель, который отражает средний ВВП на каждого человека в стране, выраженный в ценах 2010 года, с учетом покупательной способности. Темп роста указывает на изменение ВВП за определенный период времени.

Вступление

Валовой Внутренний Продукт (ВВП) является ключевым показателем экономической производительности страны, отражающим стоимость всех товаров и услуг, произведенных в ее границах. ВВП широко используется как мера экономического благополучия нации и часто служит ориентиром для принимающих решения и инвесторов.

Однако, рост ВВП сам по себе не может дать полную картину здоровья экономики. Важно понимать взаимосвязь между ростом ВВП и другими ключевыми экономическими показателями, такими как индекс потребительских цен (ИПЦ), уровень безработицы, конечные расходы на потребление, внешний баланс, доля экспорта в ВВП и доля импорта в ВВП, а также курс национальной валюты к доллару США.

Эти показатели предоставляют информацию о различных аспектах экономики и могут помочь понять, как изменения в одной переменной могут повлиять на другие. Например, высокий рост ВВП может сопутствовать снижению уровня безработицы, но также может привести к инфляции, если спрос на товары и услуги превышает предложение.

Поэтому важно изучать корреляцию между ростом ВВП и этими показателями, чтобы лучше понимать динамику экономики и принимать решения в области политики, которые могут обеспечить устойчивый долгосрочный экономический рост и стабильность.

В данном исследовании мы подробно рассмотрим взаимосвязь между ростом ВВП и вышеуказанными показателями. Мы проанализируем их исторические тенденции и изучим, как изменения в одном показателе могут повлиять на другие. Мы также рассмотрим роль внешних факторов, таких как глобальные экономические условия и международная торговля, в формировании этих взаимоотношений.

В целом, данное исследование направлено на предоставление комплексного анализа взаимосвязи между ростом ВВП и приведенными показателями.

Предварительный анализ

Среднее значение (mean) для всех показателей, кроме Внешнего баланса товаров и услуг, % от ВВП (-0,964150943), находится выше медианы (median), что говорит о наличии положительной скошенности в распределении данных. Для Внешнего баланса товаров и услуг, % от ВВП значение среднего и медианы равны, что указывает на симметричность его распределения. У большинства показателей отсутствует мода (mode), что свидетельствует о том, что данные распределены равномерно или имеют множество пиков с похожими значениями. Размах для большинства показателей достаточно большой, что указывает на разнообразие данных. Коэффициент вариации (коэф. Вар.) показывает относительную вариабельность показателей, и в данном случае значение колеблется от 0,34 (для показателя Индекса потребительских цен, 2010 г.=100) до 4,47 (для Обменного курса, единиц национальной валюты за долл. США"). Показатели с высоким коэффициентом вариации имеют большую степень изменчивости, а показатели с низким коэффициентом вариации - меньшую степень. Дисперсия (Дисп.) для большинства показателей также достаточно большая, что свидетельствует о значительной разнице между значениями показателей. Стандартное отклонение (станд.отк.) для всех показателей, кроме Внешнего баланса товаров и услуг, % от ВВП (-0,964150943), находится

выше IQR, что указывает на наличие выбросов в данных. В целом, первичный анализ показал, что данные имеют высокую вариативность, разнообразны и содержат выбросы. Также данные имеют смещенное распределение в большинстве случаев, что может указывать на неравномерность их распределения. Точечные и ящичковые диаграммы, а также расчеты верхних и нижних границ по правилам 1,5 IQR и 3 сигма, также указывают на наличие выбросов в данных.

Все расчеты предварительного анализа приведены в Semestr2.xlsx (см. прикрепленные файлы)

Корреляционный анализ

После проведения корреляционного анализа были получены коэффициенты корреляции между каждой парой переменных. Из таблицы коэффициентов корреляции можно сделать следующие выводы:

- Индекс потребительских цен слабо отрицательно связан с расходами на конечное потребление на душу населения и с ВВП на душу населения в ценах и ППС. Это означает, что при увеличении индекса потребительских цен, эти показатели снижаются.
- Расходы на конечное потребление на душу населения сильно положительно связаны с внешним балансом товаров и услуг и умеренно положительно связаны с экспортом товаров и услуг. Это означает, что при увеличении расходов на конечное потребление на душу населения, эти показатели также увеличиваются.
- Внешний баланс товаров и услуг умеренно положительно связан с экспортом товаров и услуг и слабо положительно связан с ВВП на душу населения в ценах и ППС.
- Экспорт товаров и услуг сильно положительно связан с импортом товаров и услуг и умеренно положительно связан с ВВП на душу населения в ценах и ППС.
- Обменный курс имеет слабую положительную связь с индексом потребительских цен и умеренную отрицательную связь с уровнем безработицы.
- Уровень безработицы слабо отрицательно связан с внешним балансом товаров и услуг и умеренно положительно связан с ВВП на душу населения в ценах и ППС.
- ВВП на душу населения в ценах и ППС имеет умеренную отрицательную связь с индексом потребительских цен и слабую положительную связь с внешним балансом товаров и услуг, экспортом товаров и услуг и импортом товаров и услуг.

Корреляционный анализ после удаления выбросов из первичных данных может дать более точные и надежные результаты, поскольку выбросы могут исказить корреляционные связи между переменными.

В таблице коэффициентов корреляции можно заметить, что после удаления выбросов некоторые коэффициенты корреляции изменились. Например,

- коэффициент корреляции между Индексом потребительских цен, 2010 г. = 100 и Расходами на конечное потребление на душу населения в долл. США изменился с -0,2718 до -0,2549, что является небольшим изменением в отрицательную сторону.
- коэффициент корреляции между ВВП на душу населения в ценах и ППС(США) 2010 г., темп роста и Индексом потребительских цен, 2010 г. = 100 изменился с -0,3116 до -0,4144, что является значительным изменением в отрицательную сторону.

Общая тенденция изменения коэффициентов корреляции после удаления выбросов заключается в уменьшении их значений, что может свидетельствовать о том, что связь между переменными стала более слабой.

Анализируя данную матрицу частных коэффициентов корреляции, можно сделать следующие выводы:

1. Индекс потребительских цен имеет слабую положительную корреляцию с расходами на конечное потребление на душу населения в долларах США и обменным курсом.
2. Расходы на конечное потребление на душу населения в долларах США имеют умеренную положительную корреляцию с внешним балансом товаров и услуг и слабую положительную корреляцию с индексом потребительских цен и экспортом товаров и услуг.
3. Внешний баланс товаров и услуг имеет умеренную положительную корреляцию с расходами на конечное потребление на душу населения в долларах США и сильную отрицательную корреляцию с импортом товаров и услуг.
4. Экспорт и импорт товаров и услуг имеют высокую положительную корреляцию между собой, что может свидетельствовать о наличии мультиколлинеарности в данных.
5. Обменный курс имеет слабую положительную корреляцию с индексом потребительских цен и умеренную положительную корреляцию с ВВП на душу населения в ценах и ППС.
6. Уровень безработицы имеет умеренную положительную корреляцию с внешним балансом товаров и услуг и слабую отрицательную корреляцию с ВВП на душу населения в ценах и ППС.
7. ВВП на душу населения в ценах и ППС имеет сильную отрицательную корреляцию с индексом потребительских цен и слабую отрицательную корреляцию с уровнем безработицы.

Все расчеты по корреляционному анализу можно найти на соответствующем листе Semestr2.xlsx а также в вспомогательном расчетном файле Al_dor.py (см. прикрепленные файлы)

Кластерный анализ

Было выбрано евклидово расстояние, так как оно является универсальной и удобной мерой расстояния для использования в кластерном анализе, особенно в тех случаях, когда объекты представляются в виде векторов признаков с числовыми значениями. Был построен график для определения оптимального числа кластеров (Clusters.ipynb, Приложение №). По графику

оптимальное число кластеров-6, поэтому далее будет проведена работа именно с этим числом кластеров.

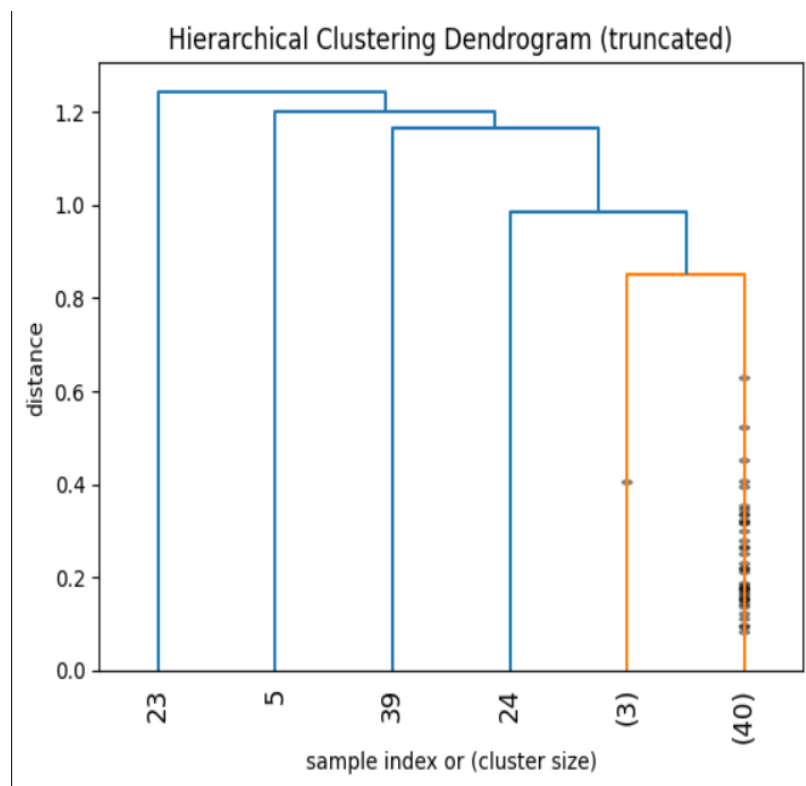


Рис.1 Дендрограмма для определения числа кластеров

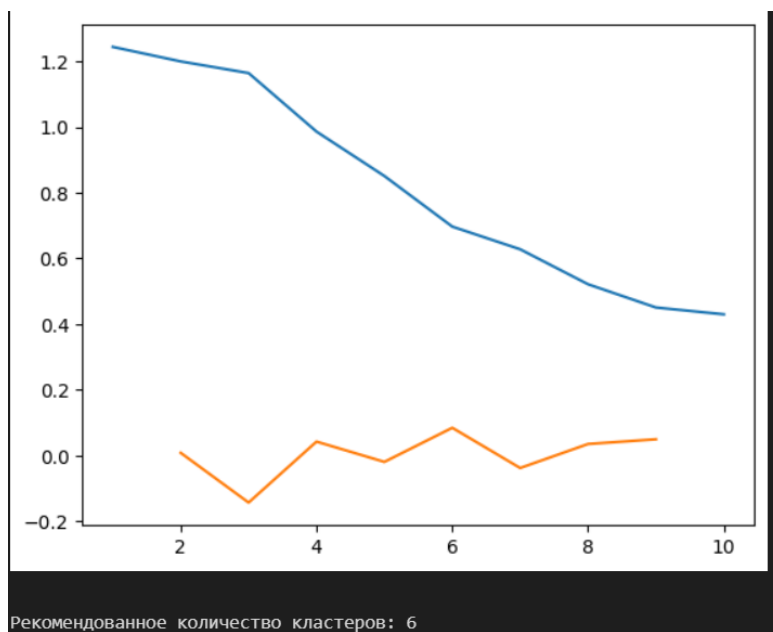


Рис.2 График для определения числа кластеров

Была проведена кластеризация методами (ближнего и дальнего соседа, среднего значения, ward (MoreClusterOptions.ipynb), иерархическая (Clusters.ipynb) , а также kmeans (Kmeans.ipynb)

Исходя из таблицы со средними значениями для каждого кластера, можно сделать следующие выводы:

По итогам kmeans кластеризации были получены следующие точечные диаграммы:

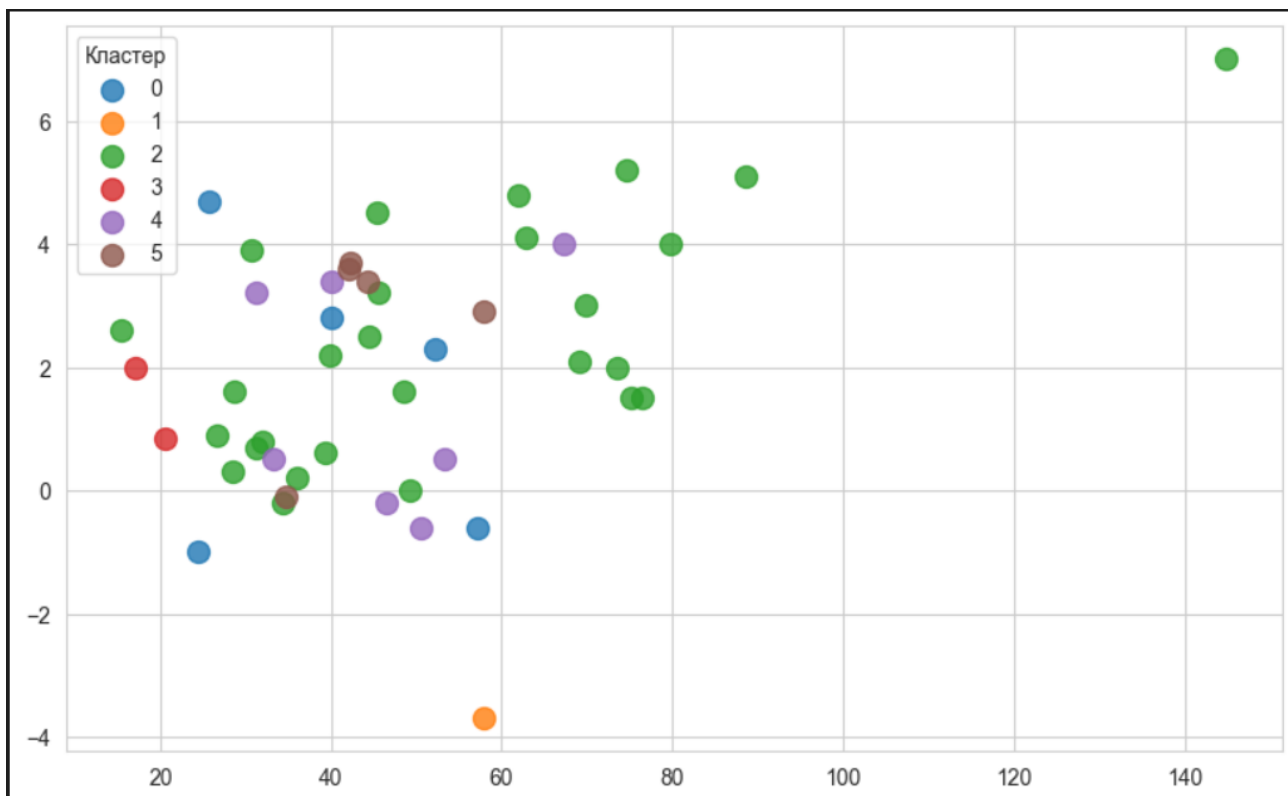


Рис. 3 точечная диаграмма для ненормализованных данных с проведенной kmeans кластеризацией

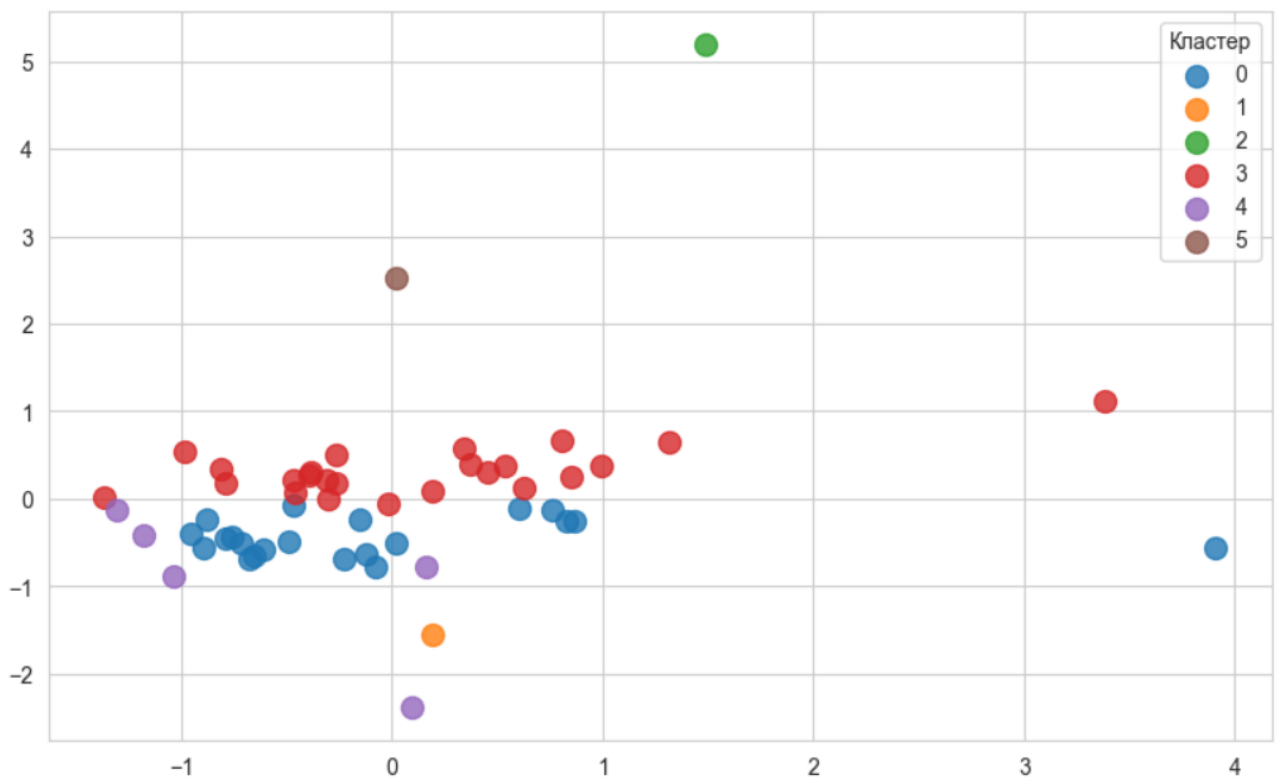


Рис. 4 точечная диаграмма для нормализованных данных с проведенной kmeans кластеризацией

Выводы

Как видно из диаграмм, при нормализации данных различия между кластерами видны более четко, а также сама кластеризация явно выражена на диаграмме и легка к восприятию. Однако финальные выводы будут сделаны по итогам таблицы средних значений для ненормализованных данных, так как они лучше отражают суть самих показателей:

kmeans_6	Расходы на конечное потребление на душу населения в долл. США	Внешний баланс товаров и услуг, % от ВВП	Экспорт товаров и услуг, доля в процентах от ВВП	Импорт товаров и услуг, доля в процентах от ВВП	Обменный курс, единиц национальной валюты за долл. США	Уровень безработицы, %	ВВП на душу населения в ценах и ППС(США) 2010 г., темп роста
0	12363.420000	-12.260000	27.700000	39.960000	71.649300	8.060000	1.640000
1	12856.600000	0.100000	58.000000	57.900000	1.592600	5.200000	-3.700000
2	29620.492593	2.111111	55.911111	53.796296	17.468444	8.966667	2.433333
3	5298.100000	2.450000	21.300000	18.800000	1315.938750	5.400000	1.420000
4	57008.928571	-2.300000	43.671429	46.000000	2.190429	8.628571	1.542857
5	18875.160000	-3.900000	40.280000	44.220000	124.317800	11.560000	2.700000

Рис. 5 итоговая таблица средних значений по кластерам

- Кластер 0 характеризуется относительно низким уровнем расходов на конечное потребление на душу населения и высоким уровнем импорта товаров и услуг относительно экспорта. Обменный курс национальной валюты относительно доллара США в этом кластере также выше, чем в других кластерах.
- Кластер 1 отличается относительно низкими расходами на конечное потребление на душу населения, низким уровнем внешнего баланса товаров и услуг и высоким уровнем

безработицы. Кроме того, данный кластер имеет отрицательный темп роста ВВП на душу населения.

- Кластер 2 характеризуется высокими расходами на конечное потребление на душу населения и высоким уровнем экспорта товаров и услуг относительно импорта. Внешний баланс товаров и услуг также находится на относительно высоком уровне.
- Кластер 3 имеет относительно низкие расходы на конечное потребление на душу населения и низкий уровень внешнего баланса товаров и услуг, однако уровень безработицы ниже, чем в кластере 1.
- Кластер 4 характеризуется высокими расходами на конечное потребление на душу населения и высоким уровнем импорта товаров и услуг относительно экспорта. Кроме того, данный кластер имеет наиболее высокий уровень обменного курса национальной валюты относительно доллара США.
- Кластер 5 характеризуется высокими расходами на конечное потребление на душу населения и отрицательным уровнем внешнего баланса товаров и услуг. Уровень безработицы в этом кластере также относительно высокий, а темп роста ВВП на душу населения находится на относительно высоком уровне.

Названия для кластеров

- Кластер 0, который имеет низкие расходы на конечное потребление на душу населения, высокий уровень импорта товаров и услуг относительно экспорта и высокий уровень обменного курса национальной валюты относительно доллара США, можно назвать "Импортный кластер". Страны этого кластера, вероятно, зависят от импорта товаров и услуг, больше чем от экспорта, и тратят большую часть своих доходов на импортированные товары и услуги.
- Кластер 1, который характеризуется низкими расходами на конечное потребление на душу населения, низким уровнем внешнего баланса товаров и услуг и высоким уровнем безработицы, можно назвать "Кризисный кластер". Эти страны находятся, вероятно, в трудном экономическом положении, что может объясняться низкими уровнями расходов на конечное потребление и высоким уровнем безработицы. Кроме того, отрицательный темп роста ВВП на душу населения может свидетельствовать о том, что экономика находится в рецессии.
- Кластер 2, который характеризуется высокими расходами на конечное потребление на душу населения, высоким уровнем экспорта товаров и услуг относительно импорта и высоким уровнем внешнего баланса товаров и услуг, можно назвать "Экспортный кластер". Страны этого кластера могут быть зависимы от экспорта товаров и услуг, а высокий уровень внешнего баланса может свидетельствовать о том, что экспортные доходы составляют основную часть ВВП стран данного кластера.
- Кластер 3 содержит страны среднего уровня дохода, у которых относительно низкие расходы на конечное потребление на душу населения, но высокий уровень экспорта и импорта товаров и услуг. Эти страны также имеют высокий

уровень ВВП на душу населения и низкий уровень безработицы, что указывает на довольно здоровую экономическую ситуацию. Обменный курс в этом кластере является средним. Я бы назвал этот кластер "Страны среднего уровня дохода с высоким уровнем внешней торговли".

- Кластер 4 состоит из стран с низким уровнем дохода и относительно низким уровнем экспорта товаров и услуг. Однако у этих стран высокий процент ВВП, который их экономика получает от внешнего баланса товаров и услуг. Уровень безработицы в этом кластере средний. Эти страны могут иметь трудности в экономическом развитии, но их экономики все же ориентированы на импорт. Я бы назвал этот кластер "Страны с высокой зависимостью от импорта и достаточно низким уровнем дохода".
- Кластер 5 состоит из стран с высокими расходами на конечное потребление на душу населения, но с низким уровнем экспорта и импорта товаров и услуг. Эти страны также имеют высокий уровень безработицы. Обменный курс в этом кластере выше, чем в других кластерах. Я бы назвал этот кластер "Страны с высокими расходами и низким экспортом". Эти страны могут иметь проблемы с внешними инвестициями и экономическим развитием.

Прикрепленные файлы

- *модульная работа семестр 2\Clusters.ipynb*
- *модульная работа семестр 2\Kmean z-cluster.ipynb*
- *модульная работа семестр 2\KMeans.ipynb*
- *модульная работа семестр 2\MoreClusterOptions.ipynb*
- *модульная работа семестр 2\Z-clusters.ipynb*
- *модульная работа семестр 2\Semestr2.xlsx*
- *модульная работа семестр 2\AI_dop.py*