

Анализ данных по показателям ЦУР на национальном уровне

Командный проект при ТГУ

Визуализация и анализ динамики показателей (слайды 6-7), корреляция (слайды 28-30), выводы (слайд 42) – Ткачев И.

Постановка задачи

Цели устойчивого роста являются амбициозной и глобальной задачей на период до 2030 года.


Для оценки прогресса по 17 целям осуществляется мониторинг 230 показателей. Динамика и связь между ними на национальном уровне не исследуется.


Необходимо провести анализ динамики и взаимосвязи данных, построить прогноз развития отдельных показателей.




Гипотезы, цели и задачи



-  **Гипотеза** Между показателями ЦУР существует взаимосвязь. Меры, реализуемые для достижения одной из ЦУР, оказывают влияние на другие ЦУР

-  **Цель** Исследование показателей ЦУР на национальном уровне, анализ динамики и взаимосвязей показателей

-  **Задачи**
 - 1. Сформировать массив данных
 - 2. Проанализировать динамику основных показателей по целям
 - 3. Проанализировать взаимосвязь между показателями
 - 4. Построить прогноз отдельных показателей

Реализация исследования



- Сбор данных и предобработка. Формирование единого массива.
- Анализ динамики отдельных показателей. Визуализация.
- Выявление взаимосвязей с помощью корреляционного и сетевого анализа.
- Построение прогноза

Результат: аналитический отчет о динамике показателей ЦУР, взаимосвязи показателей, прогнозе данных.

Источники информации



Федеральная служба
государственной статистики



United
Nations



SUSTAINABLE
DEVELOPMENT
GOALS
KNOWLEDGE PLATFORM

Методы и подходы



- Для лучшего понимания данных использовались инструменты визуализации. С их помощью выявлялась динамика и возможные взаимосвязи.
- Подтверждение взаимосвязей проводилось путем корреляционного и сетевого анализа.
- Прогноз показателей строился методом скользящей средней. Модель была выбрана исходя из специфики данных (временные ряды) и их ограниченности (с 2010 по 2020 гг.)

Инструменты для реализации

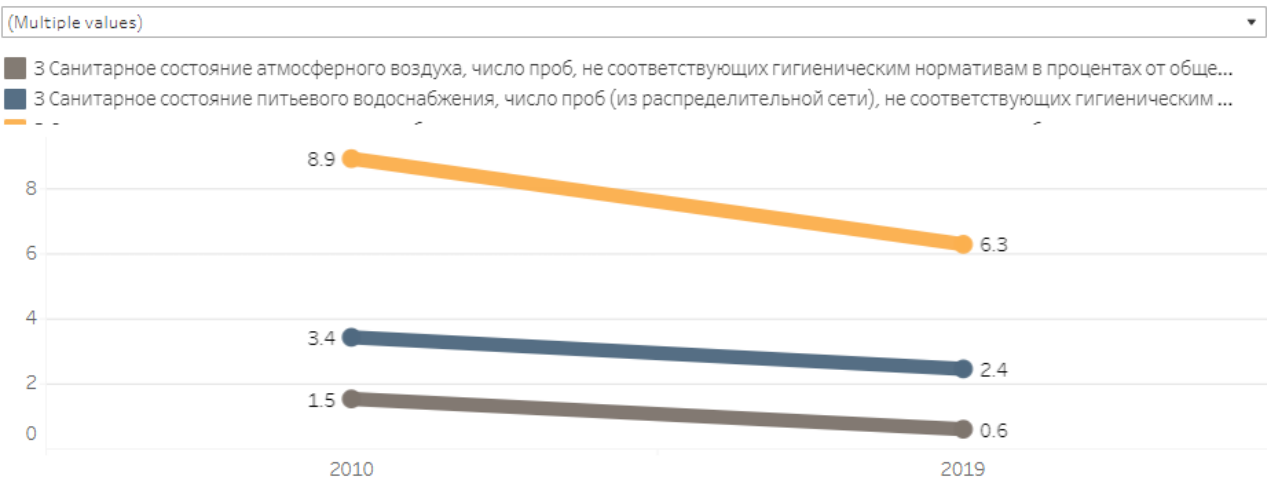
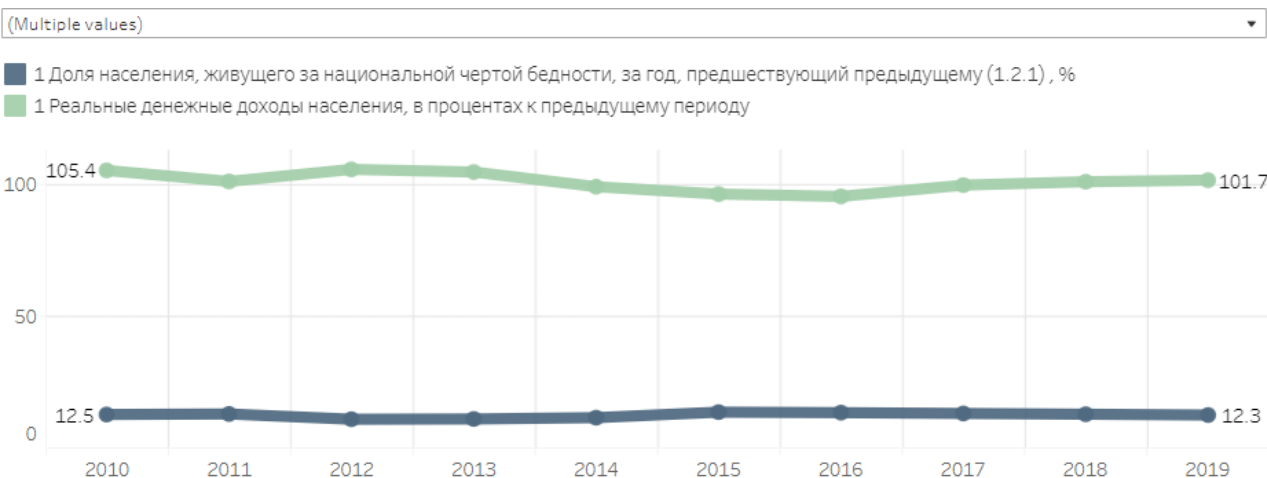


Yandex DataLens



Динамика показателей ЦУР

Данные по всем показателям ЦУР по ссылке: <https://public.tableau.com/app/profile/ivan8718/viz/SDGTotal/Dashboard1?publish=yes>



Анализ динамики. Выводы.



- С учетом того, что показателей ЦУР большое количество, для визуализации и анализа динамики были отобраны наиболее значимые и имеющие данные за 5 и более лет.
- В ходе анализа динамики показателей ЦУР выявлены следующие тенденции:
 1. Положительная тенденция по показателям:
2, 3(частично), 4, 5, 6, 7, 8(частично), 9, 11, 12, 13, 15, 17
 2. Отрицательная тенденция по показателям:
1, 14(частично), 16(частично)
 3. Отсутствие значимых изменений по показателям:
10, 3(частично)
- В целом видна положительная динамика по 13-ти из 17-ти показателям ЦУР - из чего можно сделать вывод о правильных мерах, принимаемых в данных направлениях.
- Однако присутствуют и отрицательные моменты по важным показателям (например, таким как №1. Ликвидация нищеты) на которые необходимо обратить более пристальное внимание.

Взаимосвязь целей и показателей



Проведенный анализ наличия взаимосвязи между ЦУР основан на расчете парных коэффициентов корреляции между показателями



Нулевая гипотеза:

ЦУР, а значит, и сферы жизни общества, не демонстрируют значимых взаимосвязей



Альтернативная гипотеза:

Хотя бы две ЦУР, а значит, и сферы жизни общества, демонстрируют значимую взаимосвязь



Условия и риски:

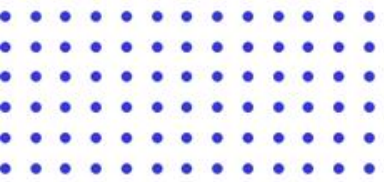
- ограниченность временных рядов
- неоднородность показателей
- несоответствие набора показателей, характеризующих ЦУР, набору показателей, характеризующих состояние соответствующих сфер жизни общества, принятых в национальной статистике



Инструменты:

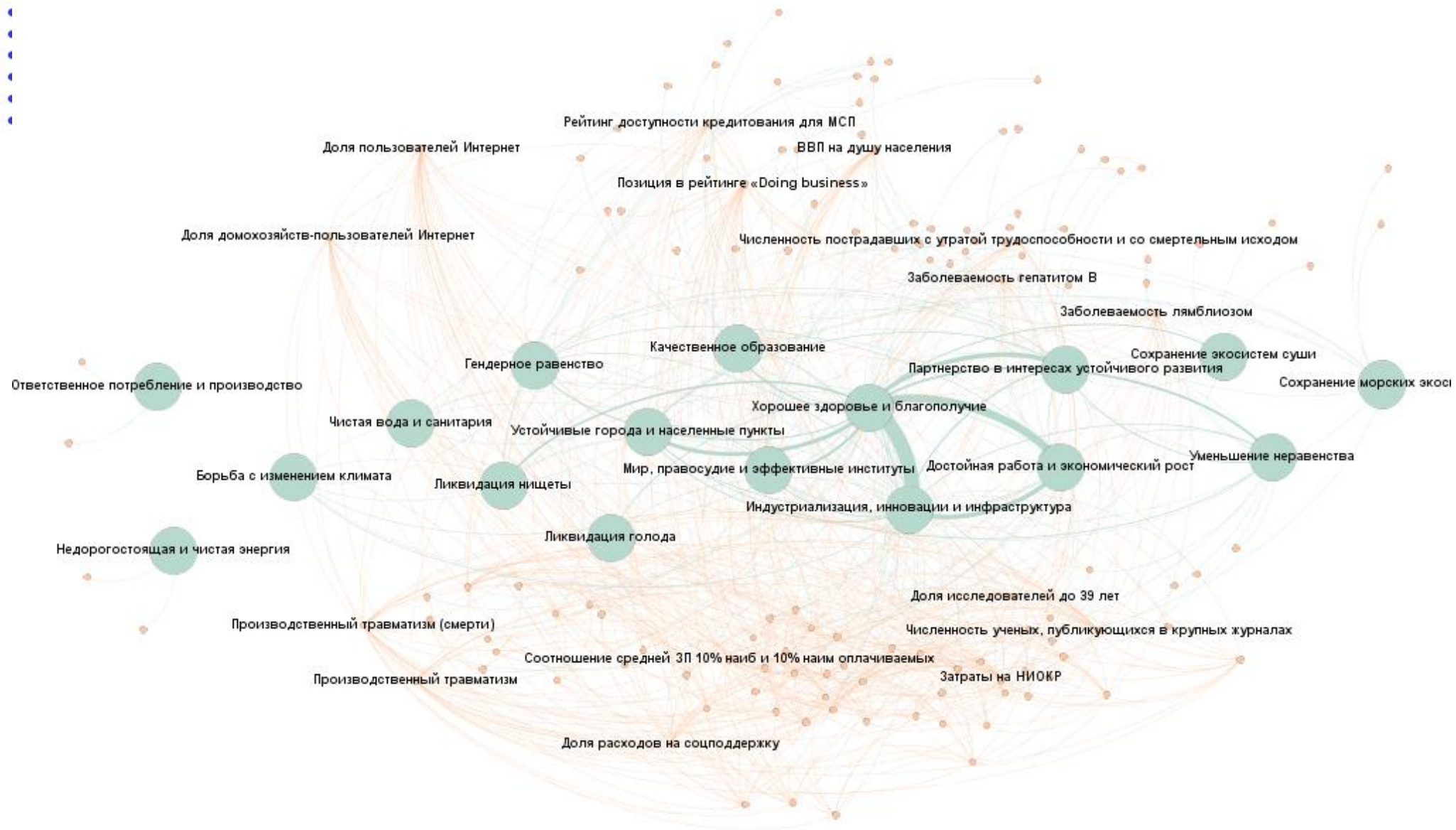
- Для расчета - функции библиотек Pandas, Seaborn и Scipy.stats.
- Для визуализации взаимосвязи - инструменты приложения Gephi для сетевого анализа, Seaborn при анализе корреляции

Взаимосвязь целей и показателей



- Критерии отбора пар показателей:
абсолютное значение коэффициента корреляции больше 0,95
р-уровень значимости менее 0,01
- Основной вывод:
есть основания отклонить нулевую гипотезу, так как по итогам анализа сводного массива (134 показателя) корреляцию, соответствующую критериям, показали 725 пар.
- Параметры графа:
Вершины - ЦУР и соответствующие им показатели (всего 152)
Ребра - выявленные связи между показателями, между ЦУР и соответствующими им показателями, а также между ЦУР, которым соответствуют коррелирующие показатели (всего 1004)

Взаимосвязь целей и показателей

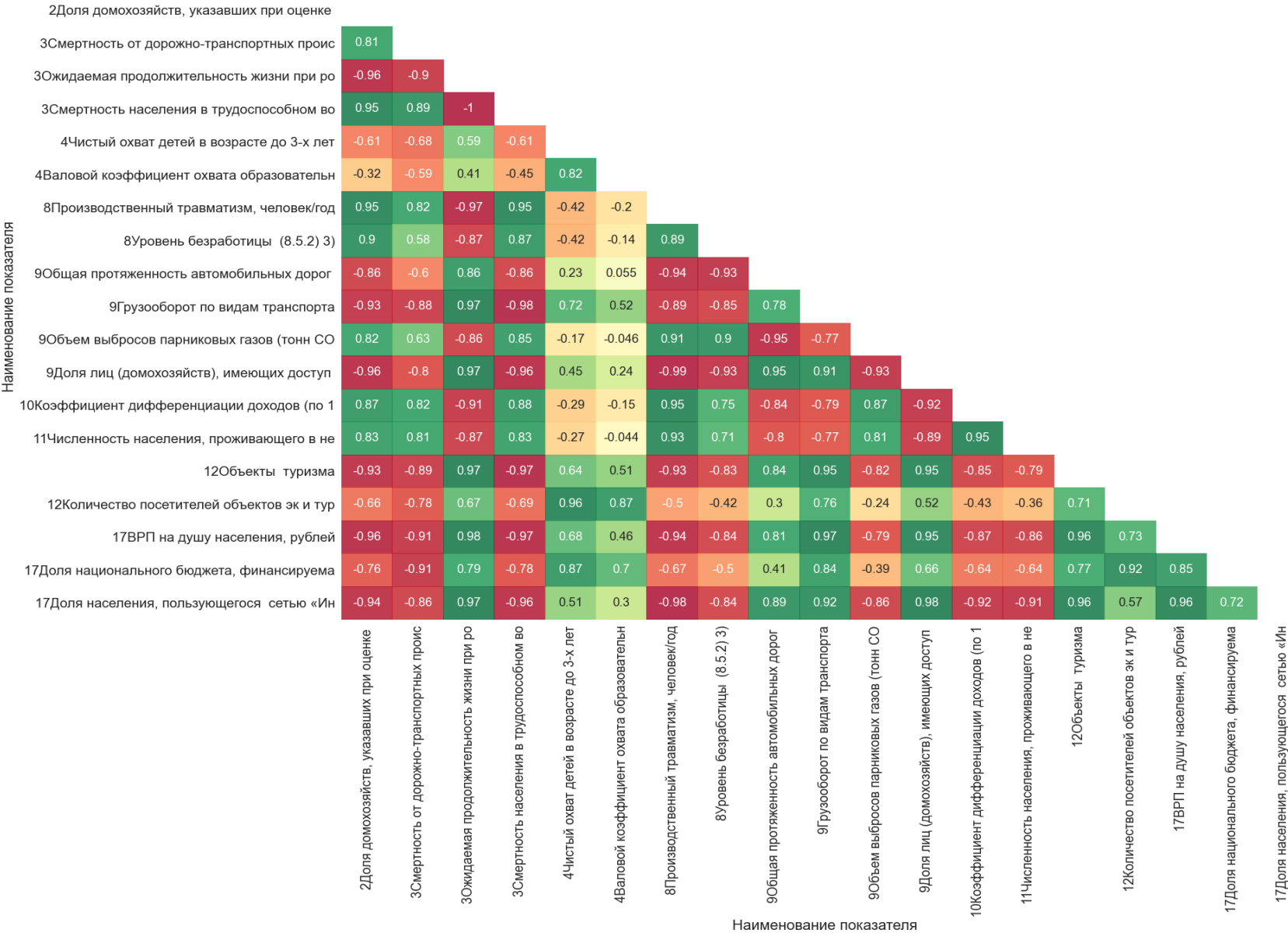
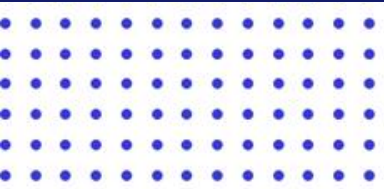




Выводы по итогам сетевого анализа:

- Толщина ребер зависит от количества установленных связей (но не говорит о силе взаимосвязи). Самое большое количество связей между целями:
 - “Хорошее здоровье и благополучие” и “Индустриализация, инновации и инфраструктура”
 - “Хорошее здоровье и благополучие” и “Достойная работа и экономический рост”
 - “Партнерство в интересах устойчивого развития” и “Хорошее здоровье и благополучие”
 - “Индустриализация, инновации и инфраструктура” и “Достойная работа и экономический рост”
- Наибольшая мощность и показатели центральности также у упомянутых целей, а также у ЦУР
- “Мир, правосудие и эффективные институты”, “Устойчивые города и населенные пункты”
- На графе представлены также показатели, демонстрирующие наиболее высокие значения мощности и центральности

Взаимосвязь показателей (“Тепловая карта”)



Взаимосвязь целей и показателей



№ ЦУР - Показатель	№ ЦУР - Показатель2	Кэфф. Корреляции
2Доля домохозяйств, указавших при оценке своего материального положения на нехватку денег на еду	3Ожидаемая продолжительность жизни при рождении	-0,96
2Доля домохозяйств, указавших при оценке своего материального положения на нехватку денег на еду	17ВРП на душу населения, рублей	-0,96
2Доля домохозяйств, указавших при оценке своего материального положения на нехватку денег на еду	17Доля населения, пользующегося сетью «Интернет» (%)	-0,94
2Доля домохозяйств, указавших при оценке своего материального положения на нехватку денег на еду	12Объекты туризма	-0,93
3Ожидаемая продолжительность жизни при рождении	11Численность населения, проживающего в неблагоприятных экологических условиях (в городах с высоким и очень высоким уровнем загрязнения атмосферного воздуха)	-0,87
9Объем выбросов парниковых газов (тонн СО 2-экв.) на единицу валового внутреннего продукта (в текущих ценах, млн. руб.) (9.4.1)	3Ожидаемая продолжительность жизни при рождении	-0,86
8Уровень безработицы (8.5.2) 3)	17ВРП на душу населения, рублей	-0,84
9Объем выбросов парниковых газов (тонн СО 2-экв.) на единицу валового внутреннего продукта (в текущих ценах, млн. руб.) (9.4.1)	12Объекты туризма	-0,82
3Смертность от дорожно-транспортных происшествий (3.6.1)	11Численность населения, проживающего в неблагоприятных экологических условиях (в городах с высоким и очень высоким уровнем загрязнения атмосферного воздуха)	0,81
9Объем выбросов парниковых газов (тонн СО 2-экв.) на единицу валового внутреннего продукта (в текущих ценах, млн. руб.) (9.4.1)	11Численность населения, проживающего в неблагоприятных экологических условиях (в городах с высоким и очень высоким уровнем загрязнения атмосферного воздуха)	0,81
8Производственный травматизм, человек/год (8.8.1)	3Смертность от дорожно-транспортных происшествий (3.6.1)	0,82
2Доля домохозяйств, указавших при оценке своего материального положения на нехватку денег на еду	11Численность населения, проживающего в неблагоприятных экологических условиях (в городах с высоким и очень высоким уровнем загрязнения атмосферного воздуха)	0,83
11Численность населения, проживающего в неблагоприятных экологических условиях (в городах с высоким и очень высоким уровнем загрязнения атмосферного воздуха)	3Смертность населения в трудоспособном возрасте	0,83
9Грузооборот по видам транспорта	17Доля национального бюджета, финансируемая внутр. налогами (%)	0,84
12Объекты туризма	9Общая протяженность автомобильных дорог местного значения	0,84
9Объем выбросов парниковых газов (тонн СО 2-экв.) на единицу валового внутреннего продукта (в текущих ценах, млн. руб.) (9.4.1)	3Смертность населения в трудоспособном возрасте	0,85
4Валовой коэффициент охвата образовательными программами и среднего профессионального образования в процентах к численности населения в возрасте 15-19 лет	12Количество посетителей объектов эк и тур деят-ти	0,87
2Доля домохозяйств, указавших при оценке своего материального положения на нехватку денег на еду	10Коэффициент дифференциации доходов (по 10-процентным группам населения): децильный коэффициент фондов	0,87
4Чистый охват детей в возрасте до 3-х лет дошкольным образованием	17Доля национального бюджета, финансируемая внутр. налогами (%)	0,87
8Производственный травматизм, человек/год (8.8.1)	9Объем выбросов парниковых газов (тонн СО 2-экв.) на единицу валового внутреннего продукта (в текущих ценах, млн. руб.) (9.4.1)	0,91
17Доля национального бюджета, финансируемая внутр. налогами (%)	12Количество посетителей объектов эк и тур деят-ти	0,92
11Численность населения, проживающего в неблагоприятных экологических условиях (в городах с высоким и очень высоким уровнем загрязнения атмосферного воздуха)	8Производственный травматизм, человек/год (8.8.1)	0,93
8Производственный травматизм, человек/год (8.8.1)	3Смертность населения в трудоспособном возрасте	0,95
17ВРП на душу населения, рублей	12Объекты туризма	0,96
9Доля лиц (домохозяйств), имеющих доступ к сети Интернет	17Доля населения, пользующегося сетью «Интернет» (%)	0,98



Выводы по корреляции между показателями:

- С целью проведения анализа корреляционной зависимости между показателями, относящимися к различным ЦУРам, ввиду большего количества показателей (134) в корреляционную матрицу, построенную в виде “тепловой карты”, было включено 19 наиболее важных по мнению группы.
- С учётом отбраковки явно не связанных между собой показателей была составлена таблица(слайд №28) показателей с сильной ($>0,8$) положительной и отрицательной корреляцией.
- На данном основании можно сделать вывод о наличии сильных статистических взаимосвязей между показателями, принадлежащими к разным ЦУР. Т.о. альтернативная гипотеза подтверждается: хотя бы две ЦУР, а значит, и сферы жизни общества, демонстрируют значимую взаимосвязь.
- Есть предложение для специалистов Росстата для введения новых производных показателей на основании данных зависимостей и последующее включение их в показатели ЦУР.

Прогнозы



Выбраны показатели

(в соответствии с приоритетными целями, определенными заказчиком):



Ликвидация нищеты

- Доля населения, живущего за национальной чертой бедности, за год, предшествующий предыдущему (1.2.1)
- Реальные денежные доходы населения, в % к предыдущему периоду



Ликвидация голода

- Индекс производства продукции с/х в сопоставимых ценах к предыдущему году
- Доля домохозяйств, указавших при оценке своего материального положения на нехватку денег на еду



Качество образования

- Валовой коэффициент охвата образовательными программами СПО в % к численности населения в возрасте 15-19 лет
- Доля взрослых, обладающих навыками в области информационно-коммуникационных технологий (4.4.1), в %



Достойная работа и экономический рост

- Индекс физического объема ВВП на душу населения за год, предшествующий предыдущему (8.1.1)
- Индекс производительности труда
- Уровень безработицы (8.5.2)

Прогнозы



Условия и риски:

- ограниченность временных рядов показателей
- высокая изменчивость некоторых признаков
- полученные данные о взаимосвязи показателей между собой недостаточны для построения регрессионной модели



Метод прогнозирования

С учетом условий выбрана экстраполяция методом скользящей средней, так как этот метод:

- сглаживает изменчивость признака
- применим для построения краткосрочных прогнозов в условиях ограниченности временного ряда

Размер окна сглаживания - 3 (с учетом доступности данных в большинстве случаев с 2010-2012 года)

Средняя относительная ошибка - в пределах 10%

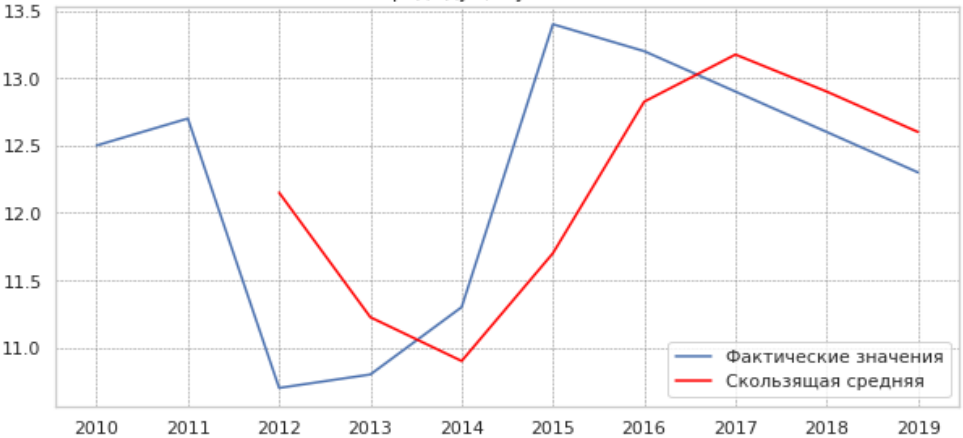
Прогноз доли населения, живущего за национальной чертой бедности, за год, предшествующий предыдущему (1.2.1), %



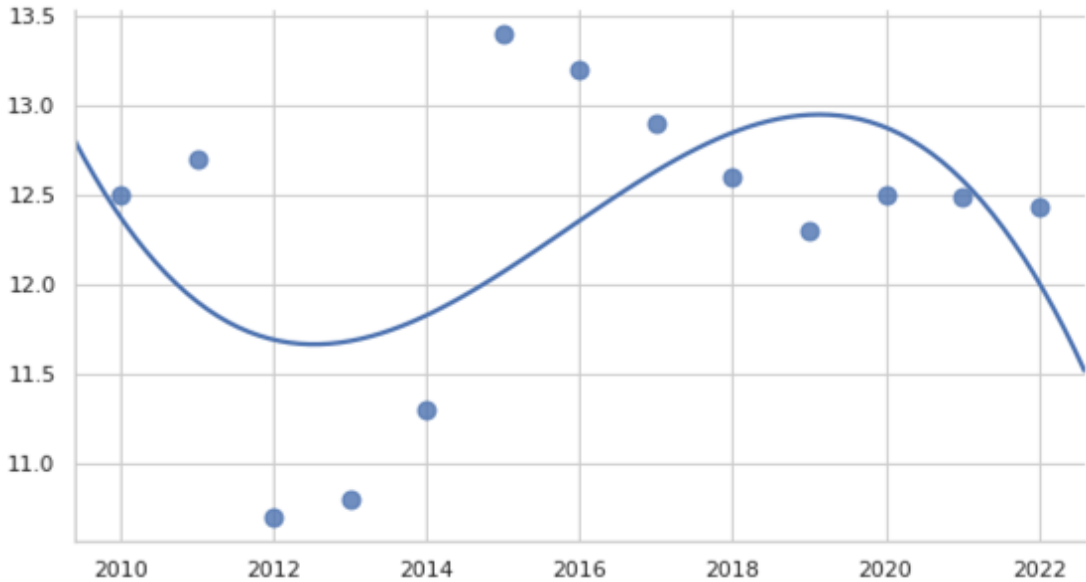
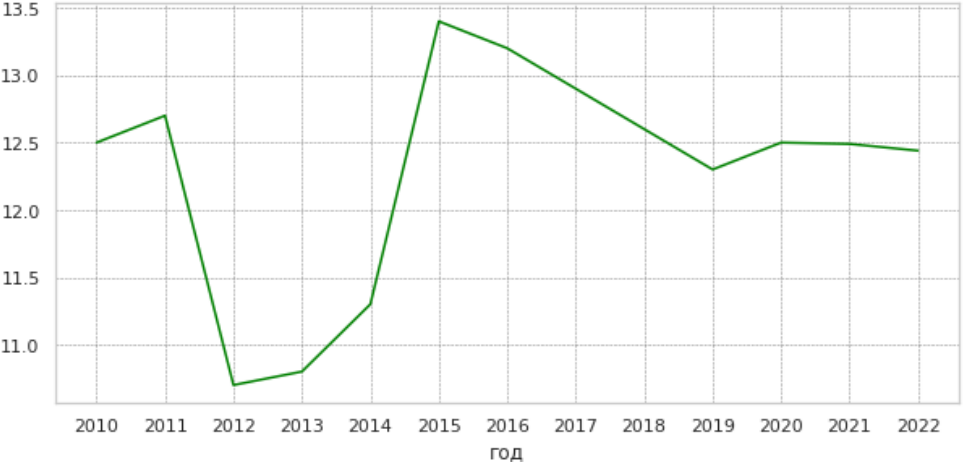
Средняя относительная ошибка при
размере окна 3 0.00413

Год	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Доля населения, живущего за национальной чертой бедности, за год, предшествующий предыдущему (1.2.1) , %	12,5	12,7	10,7	10,8	11,3	13,4	13,2	12,9	12,6	12,3	12,5	12,49	12,44

Доля населения, живущего за национальной чертой бедности, за год, предшествующий предыдущему (1.2.1) , %



Прогноз методом скользящей средней



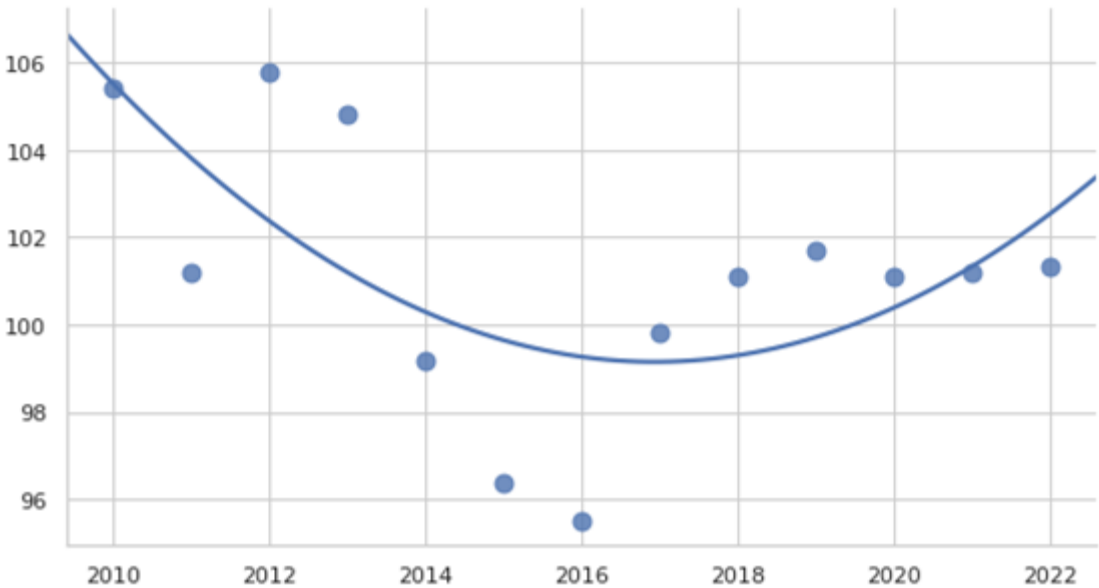
Быстрый рост показателя в кризисном 2014 году
влияет на модель, но основная тенденция - к
снижению

Прогноз темпа роста реальных денежных доходов населения, в % к предыдущему году



Средняя относительная ошибка при
размере окна 3 0.00085

Год	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Реальные денежные доходы населения, в процентах к предыдущему периоду	105,4	101,2	105,8	104,8	99,2	96,4	95,5	99,8	101,1	101,7	101,12	101,21	101,32



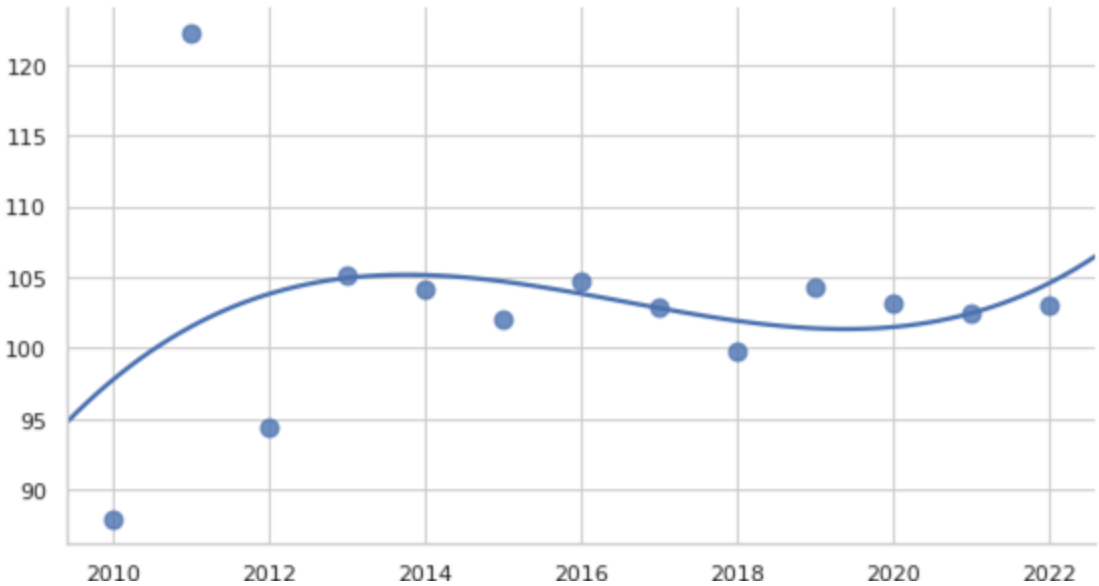
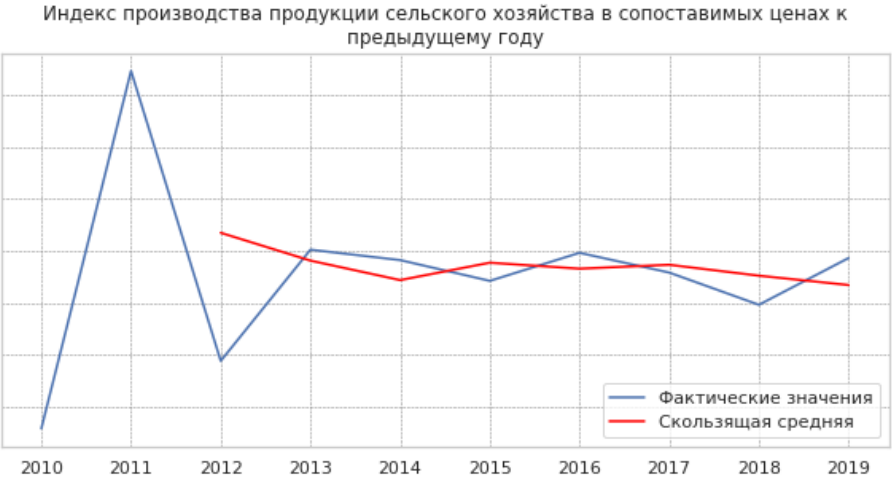
Модель сглаживает колебания показателя, но в целом отражает тенденцию к восстановлению медленного роста

Прогноз индекса производства продукции с/х в сопоставимых ценах к предыдущему году



Средняя относительная ошибка при
размере окна 3 0.00885

Год	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Индекс производства продукции сельского хозяйства в сопоставимых ценах к предыдущему году	87,9	122,3	94,4	105,1	104,1	102,1	104,8	102,9	99,8	104,3	103,2	102,53	103,08



Метод позволил элиминировать колебания
показателя в 2010-2012 годах

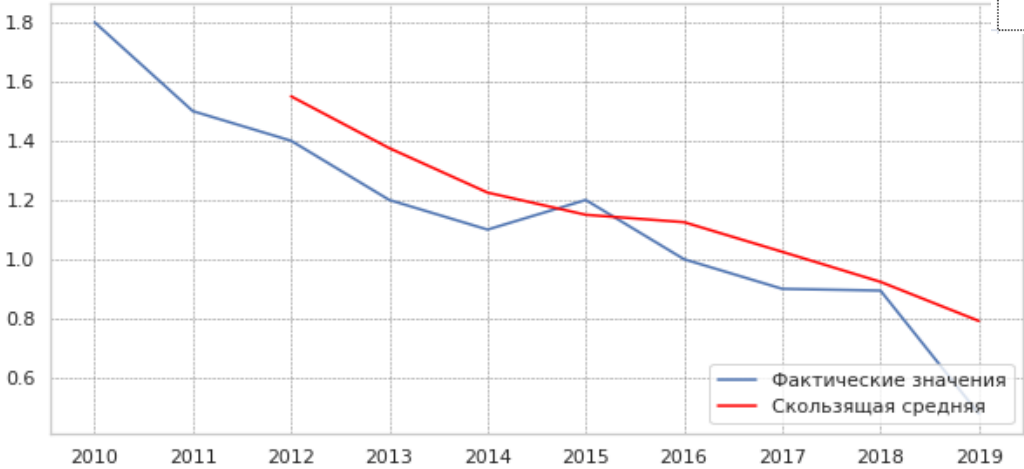
Прогноз доли домохозяйств, указавших при оценке материального положения на нехватку денег на еду, %



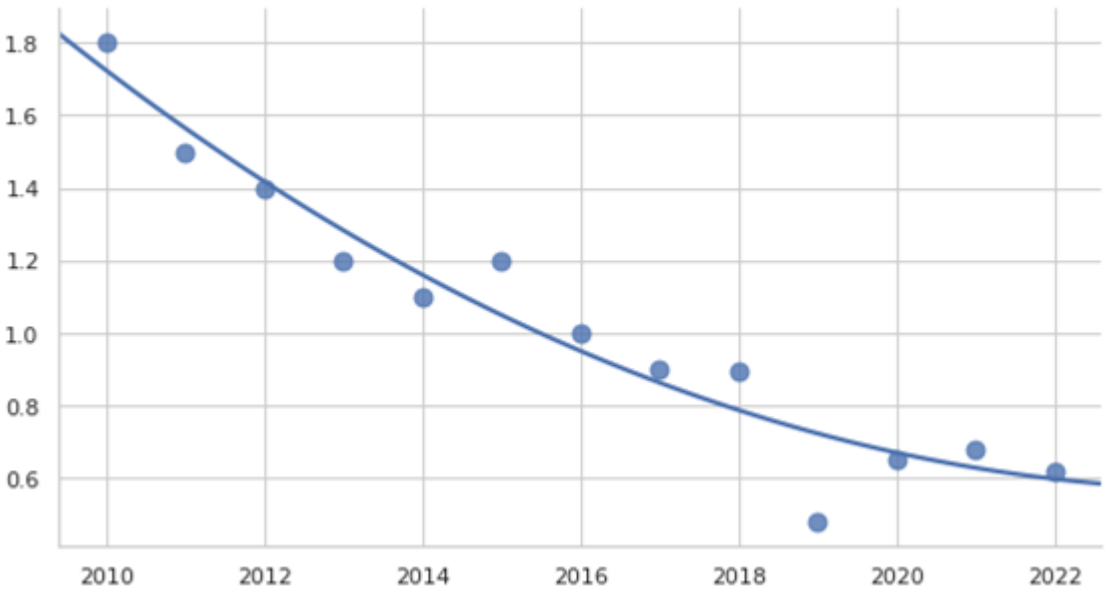
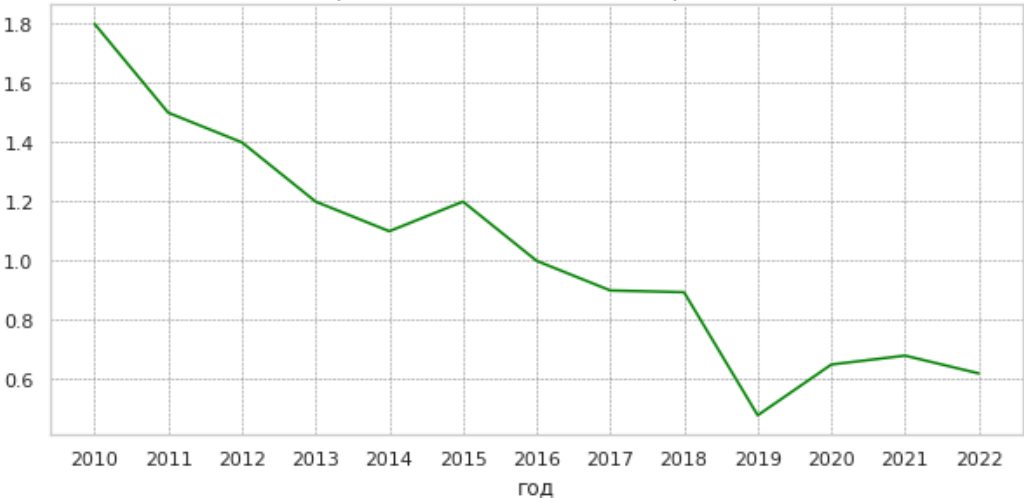
Средняя относительная ошибка при
размере окна 3 0.09815

Год	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Доля домохозяйств, указавших при оценке своего материального положения на нехватку денег на еду	1,8	1,5	1,4	1,2	1,1	1,2	1	0,9	0,894129	0,4786	0,65	0,68	0,62

Доля домохозяйств, указавших при оценке своего материального положения на нехватку денег на еду



Прогноз методом скользящей средней



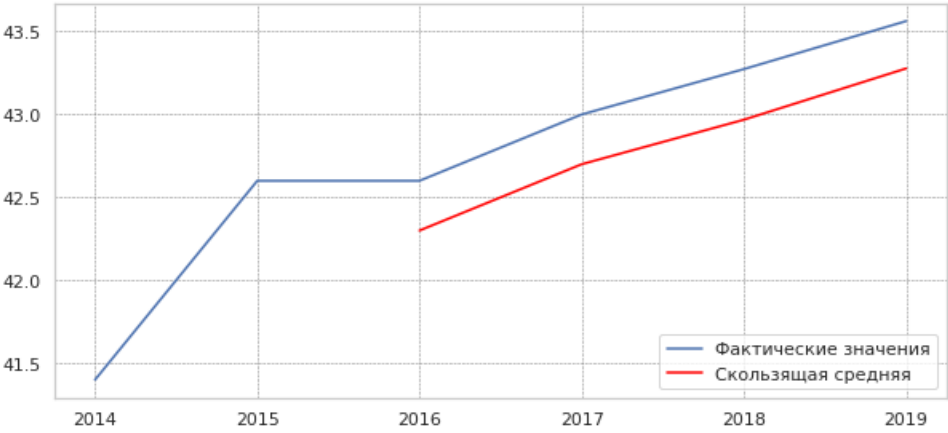
Показатель стабильно снижается, но модель сглаживает резкое снижение в 2019 году

Прогноз валового охвата программами СПО (15-19 лет)

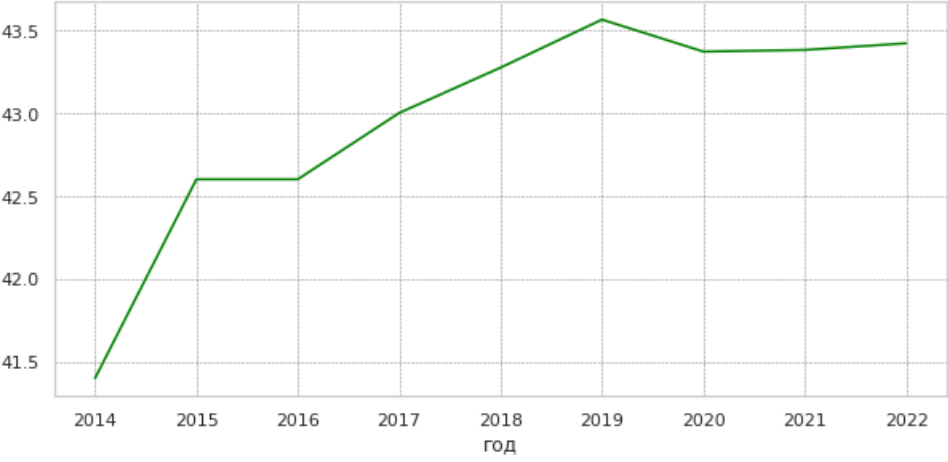


средняя относительная ошибка при
размере окна 3 0.00307

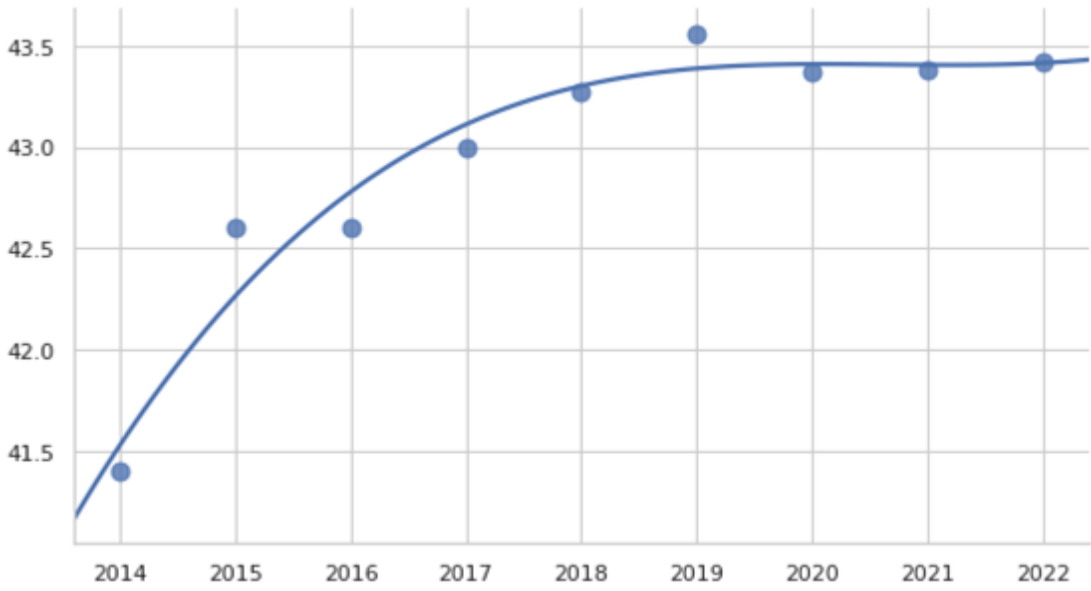
Валовой коэффициент охвата образовательными программами среднего профессионального образования в процентах к численности населения в возрасте 15-19 лет



Прогноз методом скользящей средней



Год	2014	2015	2016	2017	2018	2019	2020	2021	2022
Валовой коэффициент охвата образовательными программами СПО в % к численности населения в возрасте 15-19 лет	41,4	42,6	42,6	43	43,27	43,56	43,37	43,38	43,42



Показатель медленно растет

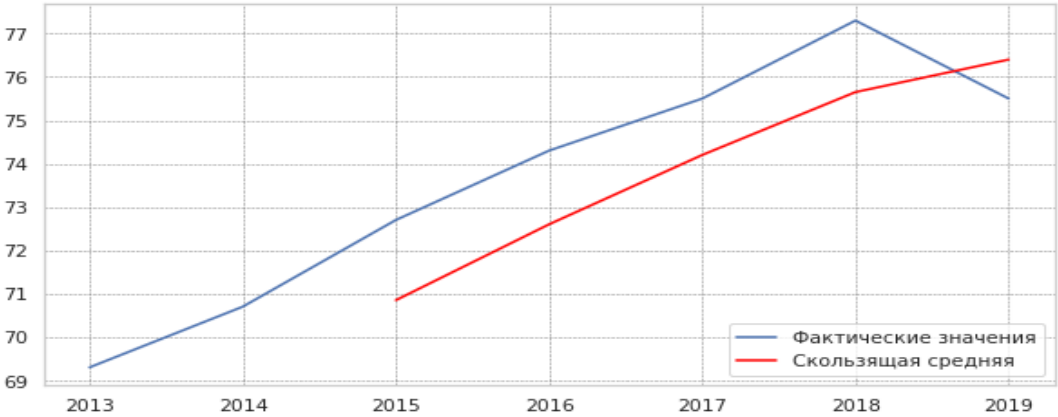
Прогноз доли взрослых, обладающих навыками в области ИКТ (4.4.1)



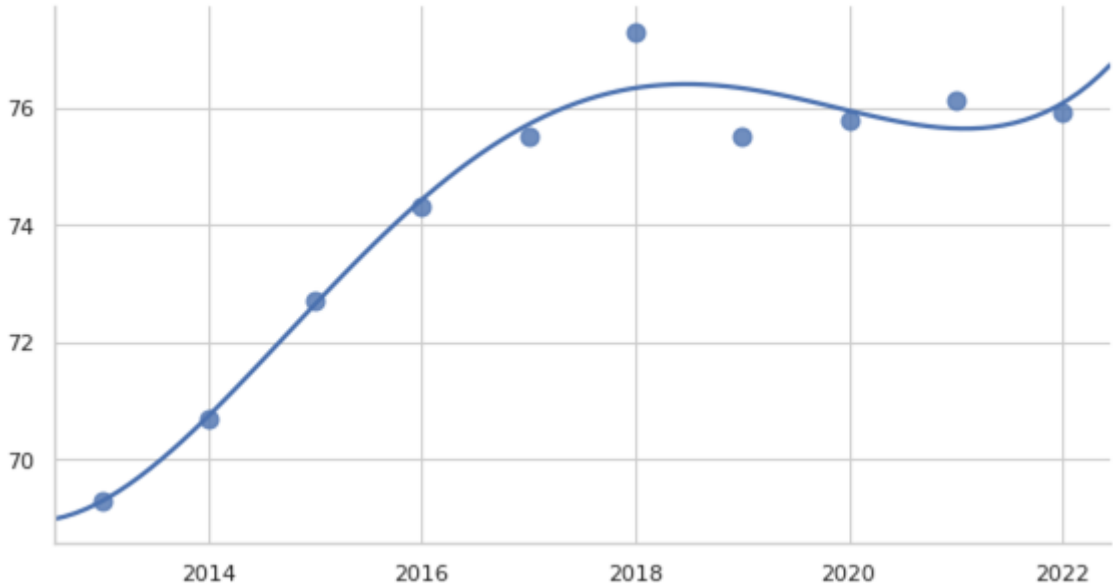
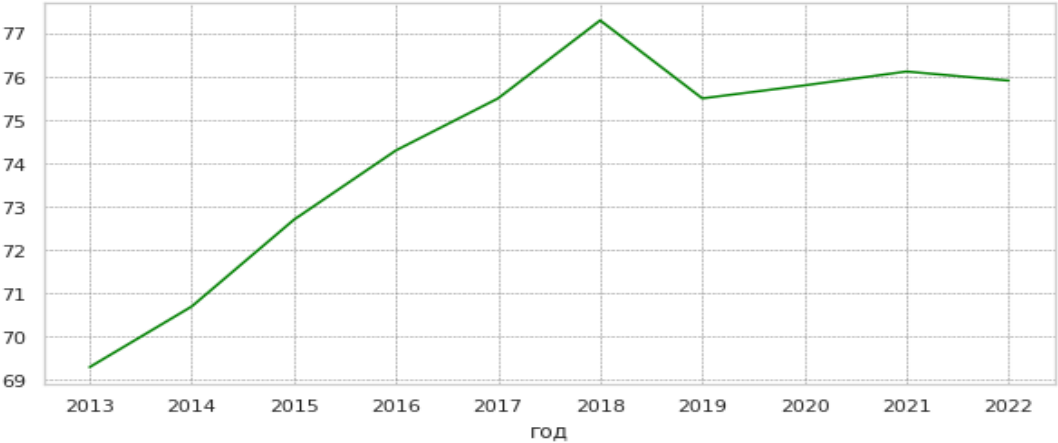
средняя относительная ошибка при
размере окна 3 -0,0075

Год	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Доля взрослых, обладающих навыками в области информационно-коммуникационных технологий (4.4.1), в %	69,3	70,7	72,7	74,3	75,5	77,3	75,5	75,8	76,12	75,91

Доля взрослых, обладающих навыками в области информационно-коммуникационных технологий (4.4.1), в %



Прогноз методом скользящей средней



Дальнейшая экстраполяция в целом отражает положительную динамику показателя до 2018 года

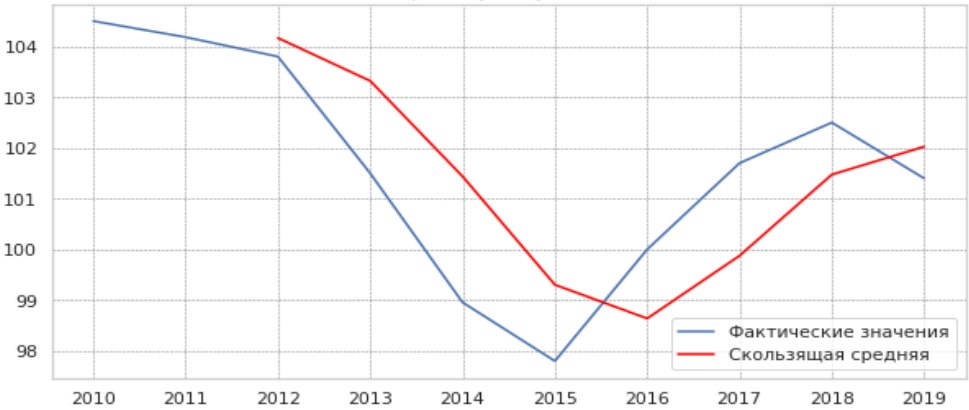
Прогноз индекса физического объема ВВП на душу населения за год, предшествующий предыдущему



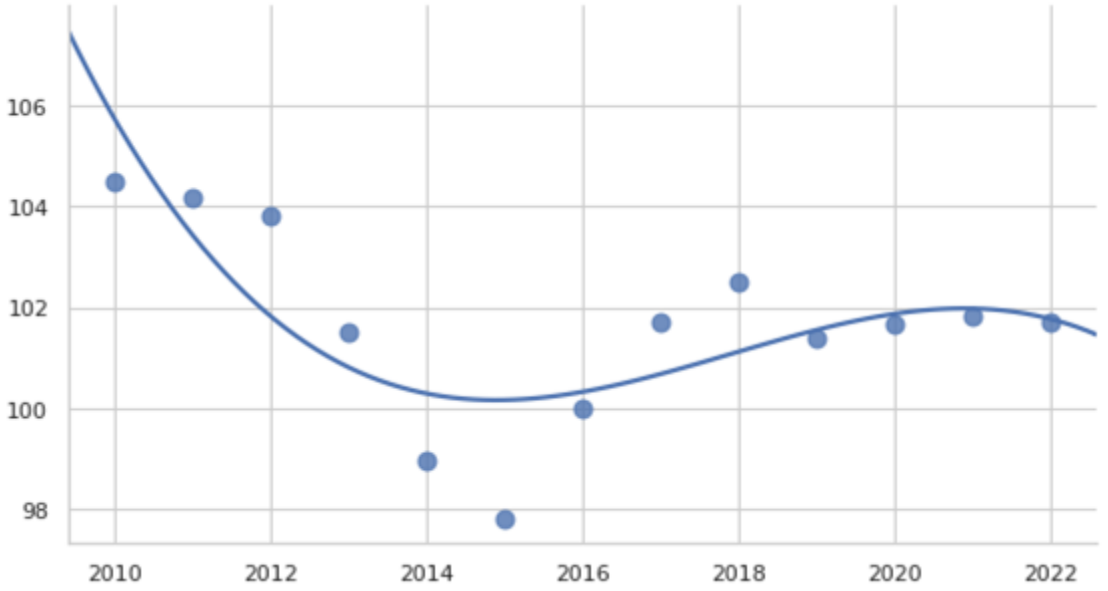
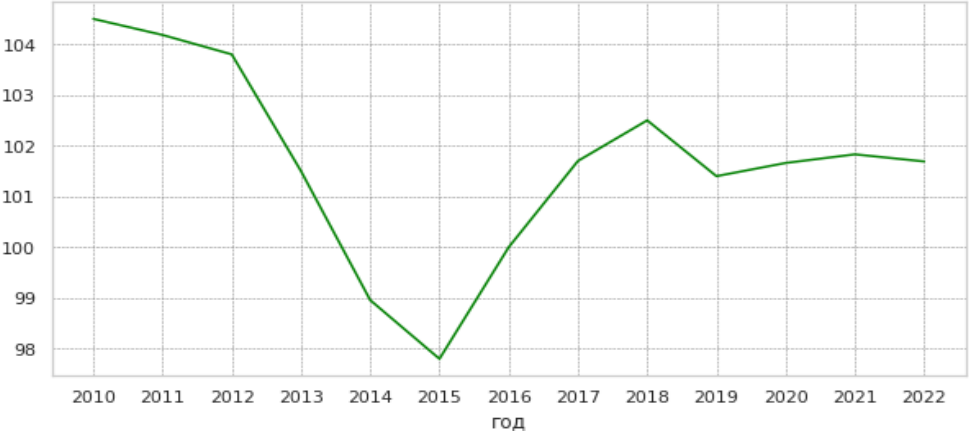
средняя относительная ошибка при
размере окна 3 0.00204

Год	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Индекс физического объема ВВП на душу населения за год, предшествующий предыдущему (8.1.1)	104,5	104,1829	103,8	101,5	98,95703	97,8	100	101,7	102,5	101,4	101,66	101,83	101,69

Индекс физического объема ВВП на душу населения за год, предшествующий предыдущему (8.1.1)



Прогноз методом скользящей средней

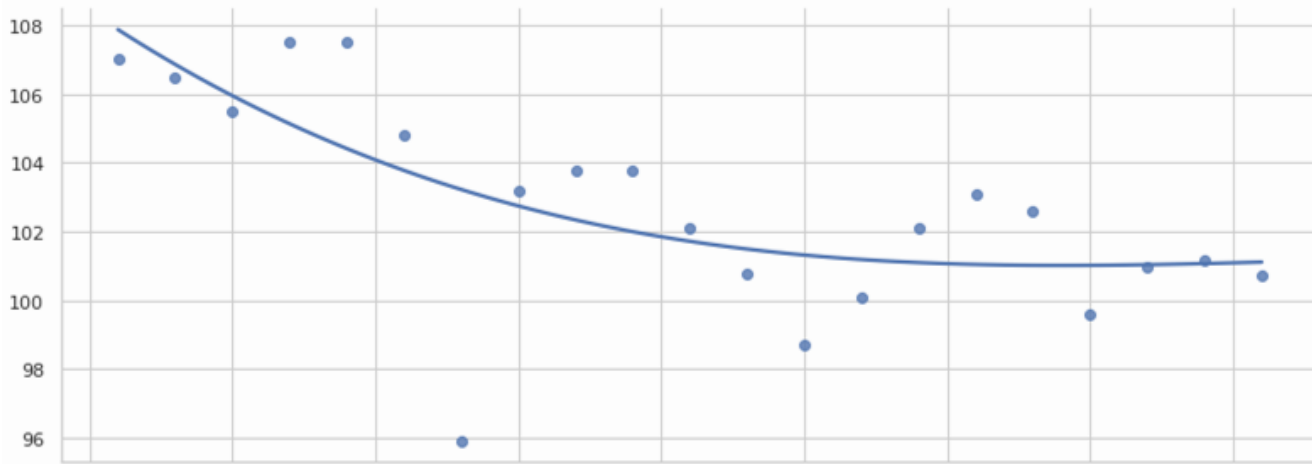
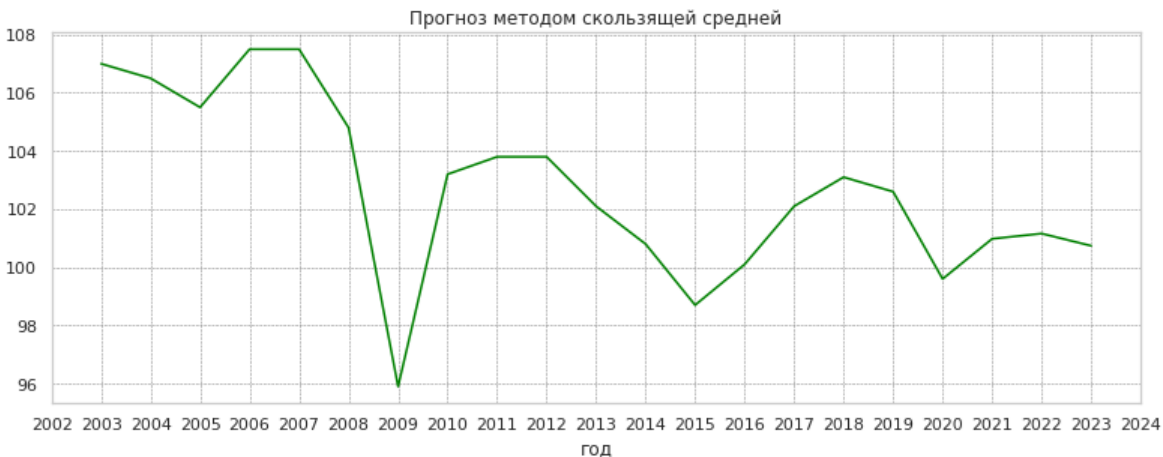


Наблюдается тенденция к замедлению индекса
производительности труда, но прогнозируются
значения выше 100%

Прогноз индекса производительности труда

средняя относительная ошибка при
размере окна 3 0.00325

Год	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Индекс производительности труда в целом по всем видам деятельности, %	107	106,5	105,5	107,5	107,5	104,8	95,9	103,2	103,8	103,8	102,1	100,8	98,7	100,1	102,1	103,1	102,6	99,6	100,98	101,15	100,74



Наблюдается устойчивая тенденция к замедлению индекса производительности труда, но прогнозируются значения выше 100%

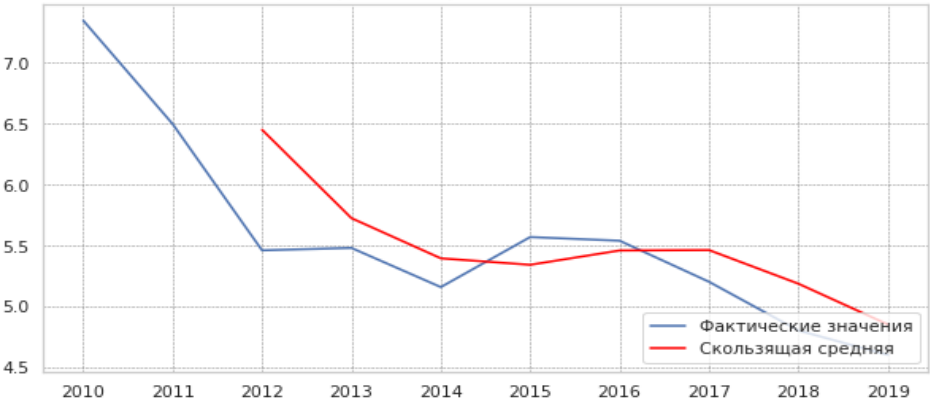
Прогноз уровня безработицы в целом (8.5.2)



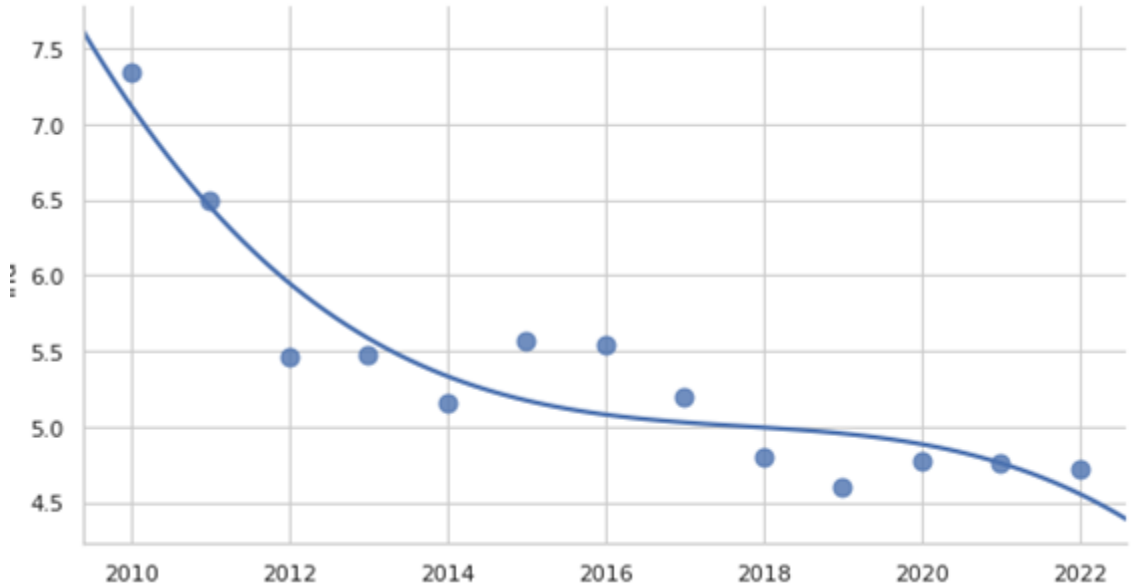
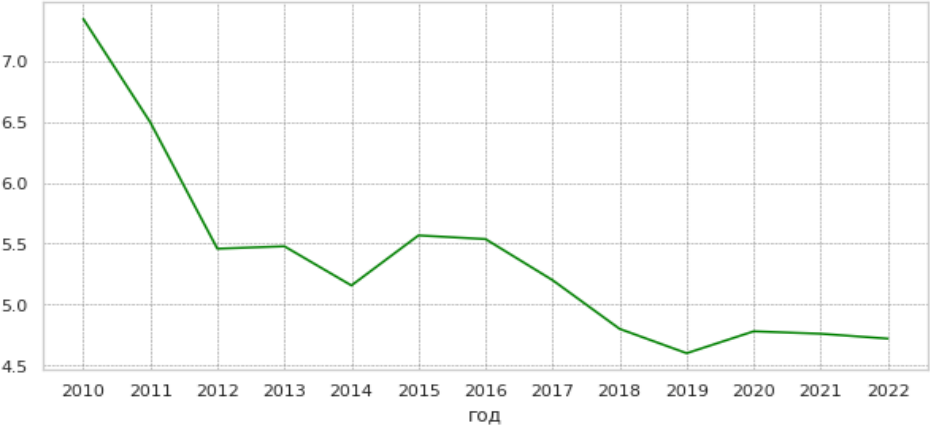
средняя относительная ошибка при
размере окна 3 0.00308

Год	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Уровень безработицы (8.5.2) 3)	7,35	6,50	5,46	5,48	5,16	5,57	5,54	5,20	4,80	4,60	4,78	4,76	4,72

Уровень безработицы (8.5.2) 3)



Прогноз методом скользящей средней



Наблюдается устойчивая тенденция к снижению показателя

[Ссылка на материалы для расчетов](#)

Результаты и выводы



В ходе проведенного исследования показателей ЦУР группой были выполнены все поставленные задачи:

- проведена выгрузка данных по показателям ЦУР
- выполнена визуализация и анализ динамики основных показателей
- подтверждено наличие взаимосвязей между показателями, относящимися к разным ЦУР
- построены прогнозы развития отдельных показателей.

По итогу исследования можно заявить о подтверждении альтернативной гипотезы: между показателями ЦУР существует взаимосвязь, а значит - меры, реализуемые для достижения одной из ЦУР, оказывают влияние на другие ЦУР.

Ссылки на расчеты



■ Ссылка на ноутбук с кодом (исходный датафрейм, корреляция, прогноз) -

Ссылка на файл с итогами прогноза (для удобства - в MSExcel)

■ Ссылка на граф –

■ Датафрейм с данными по показателям



Спасибо за внимание.

