

ИССЛЕДОВАНИЕ СВЯЗИ УРОВНЯ ЖИЗНИ НАСЕЛЕНИЯ И ОБЪЕМА ВЫБРОСОВ CO₂ С ИСПОЛЬЗОВАНИЕМ БИБЛИОТЕКИ PANDAS

Козлова Т.В., студент гр. 353504

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Жвакина А.В. – канд. техн. наук, доцент

Аннотация. В работе исследуется зависимость объема выбросов CO₂ в городах от социально-экономических показателей с применением информационных технологий для обработки данных и вычисления статистических значений.

Ключевые слова. Анализ данных, pandas, социально-экономические факторы, выбросы CO₂, корреляция, t-тест.

Анализ данных – это процесс сбора, обработки, анализа и интерпретации информации, в результате которого разрозненные данные превращаются в законченную информационную продукцию – аналитический документ.

Социально-экономические статистические показатели характеризуют развитие экономики и социальной сферы.

Углекислый газ (CO₂) – одна из ключевых угроз для экологии и климата Земли. Сжигание ископаемого топлива, промышленность, транспорт и вырубка лесов ежегодно выбрасывают в атмосферу десятки миллиардов тонн CO₂, усиливая парниковый эффект и ускоряя глобальное потепление.

Поэтому для объекта исследования выбран объем выбросов CO₂ (или же загрязнение) в различных городах с учетом таких социально-экономических факторов, как индекс покупательной способности (ИПС), индекс времени (ИВ). Зависимость изучается за последние несколько лет (2021-2025) для определения динамики роста(снижения) показателей. Для исследования выбрано 249 городов по всему миру (большая часть из Европы, Азии и Северной Америки) [1 – 3].

Индекс покупательной способности показывает относительную покупательную способность на основе средней чистой заработной платы. Значение 40 означает, что в среднем жители могут позволить себе на 60% меньше товаров и услуг, чем жители Нью-Йорка со средней зарплатой. Чем выше ИПС, тем лучше.

Индекс времени представляет собой среднее ежедневное время в пути в одну сторону в минутах. Он дает представление о времени, которое требуется для путешествия из одного места в другое в пределах города.

В данной работе применена библиотека pandas для статистических вычислений, импортирования набора данных и экспорта результатов. DataFrame – это двумерная маркированная структура, фактически таблица данных.

Медиана набора чисел – число, которое находится в середине этого набора, если его упорядочить по возрастанию.

Коэффициент корреляции – значение, показывающее прямолинейную связь между переменными. Он принимает значение от -1 до 1. Чем ближе значение к 1, тем выше положительная корреляция между показателями. Если оно, наоборот, ближе к -1 – корреляция отрицательная.

T-тест – общее название для класса методов статистической проверки гипотез (статистических критериев), основанных на распределении Стюдента. T-статистика строится обычно по следующему общему принципу: в числителе – случайная величина с нулевым математическим ожиданием (при выполнении нулевой гипотезы), а в знаменателе – выборочное стандартное отклонение этой случайной величины, получаемое как квадратный корень из несмещенной оценки дисперсии.

P-значение – вероятность получить для данной вероятностной модели распределения значений случайной величины такое же или более экстремальное значение статистики (среднего арифметического, медианы и др.), по сравнению с ранее наблюдаемым, при условии, что нулевая гипотеза верна.

Методология исследования зависимости реализована через следующие шаги: для каждого года последовательно вычисление медиан сравниваемых и сравнивающих столбцов (метод median класса DataFrame), разделение данных на группы по медиане сравнивающего столбца, расчет медиан сравниваемого столбца в этих группах, а также определение коэффициента корреляции (метод corr) и p-значения t-теста (библиотека scipy.stats).

Результаты исследования зависимости выбросов CO₂ от социально-экономических факторов за 2023-2025 представлен в таблице 1.

Таблица 1 – Результаты исследования зависимости преступности

Фактор	СЗ фактора	СЗ группы выше медианы зависимости	СЗ группы ниже медианы зависимости	Разница	Процент разницы от среднего	Корреляция	Р-значение
2025							
ИПС	92.7	38.7	64.3	25.6	48.03	-0.568	0.0
ИВ	35.35	60.0	40.2	19.8	38.335	0.482	0.0
2024							
ИПС	69.5	40.6	67.4	26.8	48.029	-0.543	0.0
ИВ	37.4	58.0	41.1	16.9	32.314	0.477	0.0
2023							
ИПС	69.45	38.6	67.8	29.2	53.043	-0.639	0.0
ИВ	35.7	61.3	38.8	22.5	44.204	0.516	0.0
2022							
ИПС	61.25	39.15	69.2	30.05	52.998	-0.635	0.0
ИВ	35.1	60.0	39.45	20.55	41.348	0.522	0.0
2021							
ИПС	56.15	38.3	67.9	29.6	52.857	-0.672	0.0
ИВ	35.2	62.2	40.9	21.3	42.095	0.508	0.0

Результаты показывают наличие значимых зависимостей. Если смотреть исключительно на процент разницы от среднего значения, то можно только сделать предположение о некоторой зависимости, и неясно, какой процент можно считать достаточно большим, а какой нет. Для уточнения результатов и была исследована корреляция, а после проведен статистический t-тест.

ИПС и ИВ оказались влиятельными показателями: у них относительно высокий процент разницы, корреляция по модулю близка к 0.5. Отрицательное значение корреляции свидетельствует о том, что с увеличением значения индекса покупательной способности уменьшается индекс выбросов CO₂, что и предполагалось.

В свою очередь индекс времени имеет положительную корреляцию, что говорит о том, что он растет вместе с индексом выбросов CO₂.

Во всех случаях р-значение t-теста равно 0, что говорит о существенной зависимости между показателями.

За 5 лет можно наблюдать уменьшение разницы между средним значением загрязнения для стран с лучшими и худшими социально-экономическими факторами. Коэффициент корреляции также уменьшается по модулю с увеличением года, однако в 2025 году есть небольшой прирост.

Графики зависимости индекса выбросов CO₂ от индекса покупательной способности за 2021 и 2025 представлены на рисунке 1.

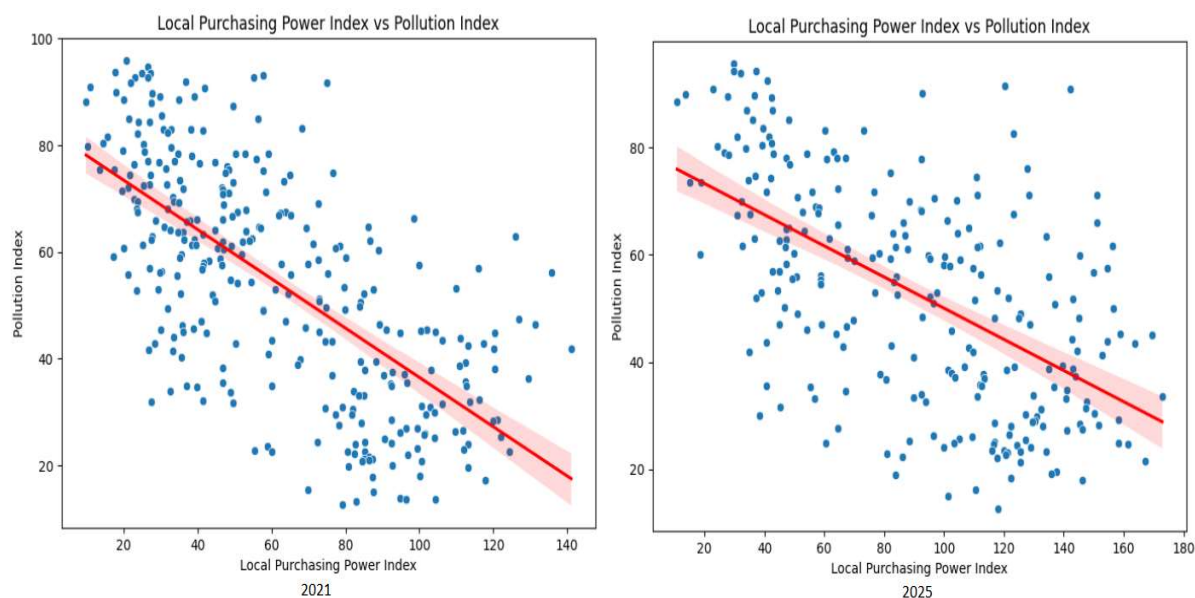


Рисунок 1 – График зависимости индекса загрязнения от ИПС за 2021, 2025

Распределение синих точек (индекс выбросов) показывает, что более высокие значения ИПС в целом связаны с более низкими значениями загрязнения. Однако точки довольно сильно разбросаны, что говорит о том, что связь между этими переменными не является строго линейной, и на индекс выбросов влияют и другие факторы.

Можно наблюдать уменьшение связанности между значением ИПС и индексом выброса CO₂, если сравнивать графики за 2021 и 2025 года.

Графики зависимости индекса выбросов CO₂ от индекса времени за 2021 и 2025 представлены на рисунке 2.

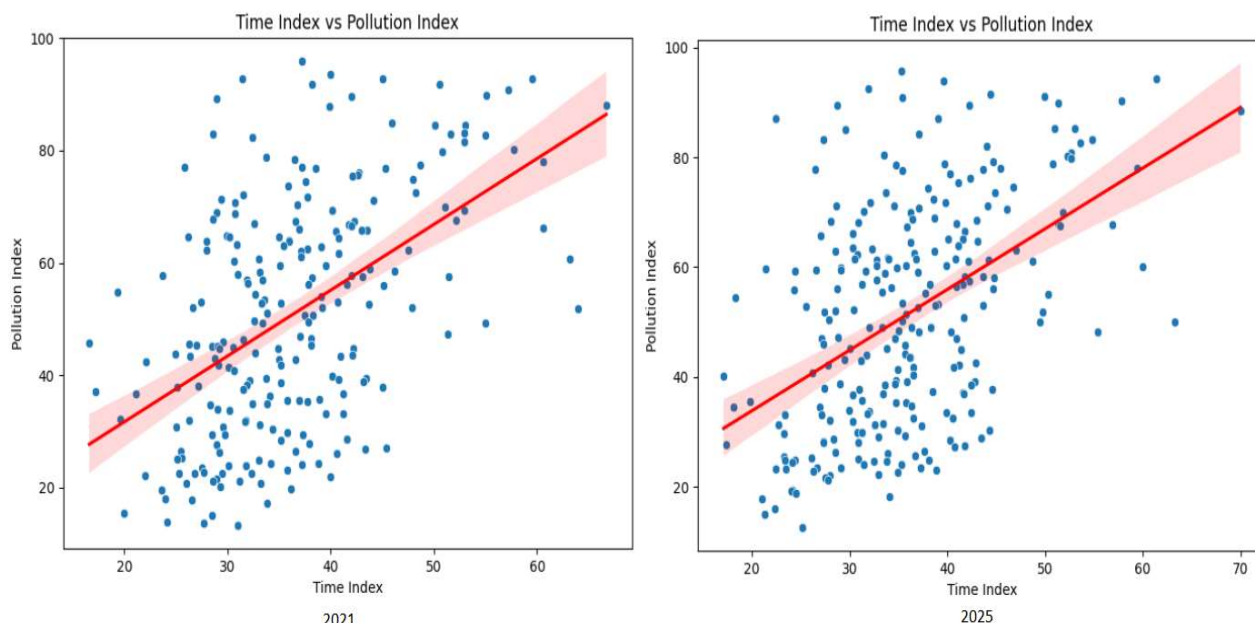


Рисунок 2 – График зависимости индекса загрязнения от ИТ за 2021, 2025

В данном случае видно, что высокие значения индекса выброса CO₂ связаны с более высокими значениями индекса времени. Можно заметить, что за 2021 и 2025 года зависимость практически не изменилась.

Использование pandas определенно упростило анализ данных (выделение нужных данных из множества строк таблиц) и само статистическое исследование.

Изменение зависимости CO₂ от индекса покупательной способности связано с тем, что происходит массовый переход на электромобили – транспорт с нулевыми выбросами, сокращающий эмиссию от сжигания бензина и дизеля. Однако часть этого эффекта нивелируется углеродным следом производства батарей и генерации электроэнергии для зарядки. Также рост числа машин в развивающихся странах увеличивает выбросы.

В том числе изменение зависимости связано с переходом в разных городах к солнечной и ветровой энергетике. Переработка мусора, снижающая объёмы свалок и мусоросжигания, уменьшает выбросы метана и CO₂.

Одной из причин неизменности зависимости выбросов CO₂ от индекса времени, несмотря на распространение электромобилей, остаются пробки, которые влияют на батарею электромобилей (и соответственно увеличивается выброс от производства).

Для снижения CO₂ мы можем предпринять разные меры. К ним относятся распространение экологичного транспорта – использование электромобилей, развитие общественного транспорта (в том числе электробусов) и велодвижения; введение экологических стандартов – налоги на углерод, запрет одноразового пластика, субсидии для "чистых" производств; поддержка научных исследований – инвестиции в инновационные технологии, такие как термоядерная энергетика или искусственный фотосинтез.

Самое важное – переход к чистым источникам энергии, так как выбросы CO₂ напрямую зависят от энергопотребления [4].

Список использованных источников:

1. Numbeo cost of living [Электронный ресурс]. – Режим доступа: <https://www.numbeo.com/cost-of-living/> – Дата доступа: 14.04.2025
2. Numbeo pollution [Электронный ресурс]. – Режим доступа: <https://www.numbeo.com/pollution/> – Дата доступа: 14.04.2025
3. Numbeo traffic [Электронный ресурс]. – Режим доступа: <https://www.numbeo.com/traffic/> – Дата доступа: 14.04.2025
4. Sciencedirect [Электронный ресурс]. – Режим доступа: <https://www.sciencedirect.com/science/article/pii/S026499932300007X#sec4> – Дата доступа: 14.04.2025

STATISTICAL ANALYSIS OF SOCIO-ECONOMIC FACTORS OF CRIME USING THE PANDAS LIBRARY

Kozlova T.V.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Zhvakina A.V. – Associate Professor of the Department of Computer Science

Annotation. The work examines the dependence of CO₂ emissions in cities on socio-economic indicators using information technology for data processing and calculating statistical values.

Keywords. Data analysis, pandas, socio-economic factors, CO₂ emissions, correlation, t-test.