

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАДИМА ГЕТЬМАНА**

Інститут інформаційних технологій в економіці

Кафедра математичного моделювання і статистики

Освітньо-професійна програма

Економічна кібернетика і Дата Сайнс

Галузь знань

05 Соціальні та поведінкові науки

Спеціальність

051 Економіка

Форма навчання: дистанційна

КВАЛІФІКАЦІЙНА МАГІСТЕРСЬКА РОБОТА

на тему «Аналіз глобального індексу штучного інтелекту»

здобувача Ведути Павла Івановича

(ПІБ, підпис)

Науковий керівник: кандидат економічних наук, доцент, Ткач Олег

Валентинович

(науковий ступінь, учене звання, ПІБ)

(підпис)

**Робота допущена до захисту перед екзаменаційною комісією
з атестації здобувачів вищої освіти (ЕК)**

В.о. завідувача кафедри: кандидат фізико-математичних наук, професор,

Великоіваненко Галина Іванівна

(науковий ступінь, учене звання, ПІБ)

(підпис)

Київ 2024

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ

ІМЕНІ ВАДИМА ГЕТЬМАНА

Навчально-науковий інститут

«Інститут інформаційних технологій в економіці»

Кафедра математичного моделювання та статистики

Освітньо-професійна програма

Економічна кібернетика та Датасайнс

Галузь знань

05 Соціальні та поведінкові науки

Спеціальність

051 Економіка

ПОГОДЖЕНО

Керівник проектної групи (гарант)
освітньо-професійної програми

О.В.Піскунова

(підпис)

20__ р.

ЗАТВЕРДЖУЮ

В.о. завідувача кафедри
математичного моделювання та
статистики

Г.І.Великоіваненко

(підпис)

20__ р.

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ

здобувачу вищої освіти Ведуті Павлу Івановичу

дистанційної форми навчання

на підготовку кваліфікаційної магістерської роботи

на тему «Аналіз глобального індексу штучного інтелекту»

Тему затверджено наказом ректора Університету від « » _____ 2023 рр. №

Кваліфікаційна магістерська робота виконується на матеріалах

_____ даних про індекс штучного інтелекту з відкритих джерел _____

План кваліфікаційної магістерської роботи

Розділ 1	Сутність та обстеження індексу штучного інтелекту (назва розділу)
Розділ 2	Дослідження чинників індексу штучного інтелекту (назва розділу)
Об'єкт дослідження:	Моделі і методи аналізу індексу ШІ
Предмет дослідження:	Процеси розвитку штучного інтелекту у світі
Мета кваліфікаційної магістерської роботи:	Аналіз даних щодо глобального індексу ШІ, провести кластеризацію країн за різними ознаками, розробити багатофакторну модель для аналізу індексу ШІ

Конкретні завдання, які здобувач повинен виконати для досягнення поставленої мети:

У розділі 1	Розглянути: сутність штучного інтелекту та його індексу; основи дисперсійно-кореляційного, регресійного аналізу; засади кластеризації та побудови багатофакторної моделі.
У розділі 2	Здійснити: кластеризацію країн за різними ознаками; виконати дисперсійно-кореляційний аналіз чинників та розробити модель аналізу індексу ШІ.

Завдання підготував

науковий керівник

О.В. Ткач

(підпис)

(ініціали, прізвище)

« ____ » _____ 20__ р.

Завдання одержав здобувач

П.І. Ведута

(підпис)

(ініціали, прізвище)

« ____ » _____ 20__ р.

Реферат

Кваліфікаційна магістерська робота містить 73 сторінки, 1 таблицю, 27 рисунків, список використаних джерел з 55 найменувань.

«Аналіз індексу штучного інтелекту»

Об'єктом дослідження кваліфікаційної магістерської роботи є моделі і методи аналізу індексу ШІ.

Предметом дослідження є процеси розвитку штучного інтелекту у світі.

Мета і завдання дослідження: розглянути дані, виконати кластеризацію країн, побудувати багатофакторну модель для аналізу індексу ШІ.

Відповідно до поставленої мети визначені такі завдання:

- Розкрити сутність штучного інтелекту та його індекса;
- Здійснити дисперсійно-кореляційний та регресійний аналіз даних індекса ШІ;
- Виконати кластеризацію країн за різними ознаками;
- Побудувати багатофакторну модель аналізу індексу ШІ;
- Виконати аналіз чинників індексу ШІ.

Методична значущість отриманих результатів полягає у створенні підходів до вивчення глобального індексу штучного інтелекту, що сприяють розвитку методологій аналізу, стандартизації та визначенню кращих практик в галузі Штучного інтелекту

Практична значущість полягає у здатності країн виявляти та розуміти ключові аспекти розвитку штучного інтелекту, розробляти ефективні стратегії та впроваджувати інноваційні рішення, сприяючи подальшому розвитку та використанню ШІ в різних сферах, включаючи економіку, освіту та медицину.

Ключові слова: штучний інтелект, глобальний індекс ШІ, кластеризація, багатофакторна модель, чинники ШІ, аналіз індексу ШІ.

Рік виконання кваліфікаційної магістерської роботи – 2024.

Рік захисту роботи – 2024.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів	3
ВСТУП	4
 РОЗДІЛ 1	
ТЕОРЕТИЧНІ ЗАСАДИ ДОСЛІДЖЕННЯ ІНДЕКСА ШТУЧНОГО ІНТЕЛЕКТУ	8
1.1 Сутність штучного інтелекту. Обчислення індексу ШІ	8
1.2 Засади дисперсійно-кореляційного та регресійного аналізу	9
1.3 Основи побудови багатofакторної моделі	10
 Висновки до розділу 1	22
 РОЗДІЛ 2	
ДОСЛІДЖЕННЯ ЧИННИКІВ ІНДЕКСУ ШІ	24
2.1 Кластеризація країн за різними ознаками	31
2.2 Побудова та аналіз багатofакторної моделі індексу ШІ	34
2.3 Результати аналізу глобального індексу ШІ	59
 Висновок до розділу 2	65
 ВИСНОВКИ	66
 ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ	72
Додаток А	73
Додаток Б	95

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ІІІ або AI(Artificial intelligence) – штучний інтелект.

EDA (Exploratory data analysis) – Дослідницький аналіз даних.

Індекс – букви, число або інша комбінація символів, що вказує на місце елемента в сукупності або характеризує стан деякої системи [15]. Зокрема, кожен індекс ІІІ має свою оцінку за 100-ою шкалою [17].

ВСТУП

Штучний інтелект (ШІ) стоїть на передньому краї трансформаційних технологій, змінюючи галузі, економіку та саму структуру суспільства. Оскільки країни змагаються за лідерство в цій технологічній революції, то глобальний індекс ШІ стає компасом, що пропонує розуміння заплутаної мережі факторів, які сприяють або перешкоджають розвитку можливостей штучного інтелекту в усьому світі.

В епоху, коли штучний інтелект не тільки лежить в основі технологічного прогресу, але й має етичні, соціальні й економічні наслідки, дослідження повинно розкрити складність екосистеми ШІ. Вивчаючи дані Глобального індексу, ми прагнемо виявити моделі, кореляції та відмінності, які формують глобальний наратив ШІ [1].

До суті, це дослідження слугує маяком, висвітлюючи контури розвитку ШІ-ландшафту та надаючи зацікавленим сторонам, політикам та ентузіастам тонке розуміння глобальних сил, які керують революцією ШІ. Ознайомлюючись із багатим розмаїттям Глобального індексу штучного інтелекту, ми прагнемо зробити свій внесок у поточний діалог щодо формування відповідального, інклюзивного та ефективного майбутнього ШІ. Таким чином збільшення значення штучного інтелекту у всіх аспектах глобального розвитку суспільства та економіки робить його аналіз необхідним.

Актуальність теми полягає в тому, що штучний інтелект визначає сучасну технологічну революцію, і розуміння його стану та тенденцій за допомогою глобального індексу є важливим для розвитку стратегій, ефективних рішень щодо сталого розвитку будь-якої країни завдяки пришвидшенню інновацій в результаті розширення сфер застосування ШІ.

Аналіз дослідження і публікацій: Прелік тих науковців, що займались проблемою заданої теми: *Geoffrey Hinton* [26]: Визнаний канадський

комп'ютерний вчений, який вніс вагомий внесок у глибоке навчання і нейронні мережі. *Yann LeCun* [25]: Французький вчений, що займається машинним навчанням, відомий своїми роботами в області сверточних нейронних мереж. *Andrew Ng* [24]: Китайсько-американський вчений, спеціаліст у сфері машинного навчання, співзасновник Google Brain та Coursera. *Fei-Fei Li* [23]: Комп'ютерний вчений і професор, відомий своєю роботою в області комп'ютерного зору та нейроінформатики. *Stuart Russell*: Британський вчений, автор спільно з Пітером Норвігом відомого підручника "Штучний інтелект: Сучасний підхід.". Вітчизняні вчені, що займаються такою проблематикою: Олександр Білецький [54]: Український вчений, який активно працює у сфері машинного навчання та обробки природної мови; Любомир Турсунов [55]: Дослідник в області комп'ютерного зору та глибокого навчання, засновник компанії LookInna, яка спеціалізується на аналізі відео за допомогою штучного інтелекту, Андрій Матвійчук: дослідник в області машинного навчання [53].

Об`єкт: моделі і методи аналізу індексу ШІ;

Предмет: процеси розвитку штучного інтелекту у світі;

Методи дослідження, які були використані в процесі дослідження та обробки матеріалів: порівняльний, графічний, розрахунково-аналітичний.

Метою кваліфікаційної магістерської роботи є аналіз даних щодо глобального індексу ШІ, провести кластеризацію країн за різними ознаками, розробити багатофакторну модель для аналізу індексу ШІ;

Теоретична значущість роботи у поглибленні розуміння тенденцій та стану у сфері штучного інтелекту та створює можливість пришвидшення інноваційного розвитку будь-якої країни.

Практична значущість роботи полягає у поліпшенні здатності країн виявляти та розуміти ключові аспекти розвитку штучного інтелекту у світі, розробляти ефективні стратегії та впроваджувати інноваційні рішення, сприяючи подальшому розвитку та використанню ШІ в різних сферах, включаючи економіку, освіту та медицину.

Методична значущість полягає у створенні підходів до вивчення глобального індексу штучного інтелекту, що сприяють розвитку методологій аналізу, стандартизації та визначенню кращих практик в галузі Штучного інтелекту.

Інформаційна база: праці відчизняних та зарубіжних науковців і практиків, статистичні дані з вебсайта kaggle

Структура роботи: у першому розділі розглянуто сутність штучного інтелекту та його індексу, описано засади дисперсійно-кореляційного, регресійного аналізу, та основ побудови багатофакторної моделі; у другому розділі виконано кластеризацію країн за різними ознаками що стосуються штучного інтелекту, розроблено багатофакторну модель, здійснено всебічний аналіз глобального індексу ШІ.

РОЗДІЛ 1

ТЕОРЕТИЧНІ ЗАСАДИ ДОСЛІДЖЕННЯ ІНДЕКСА ШТУЧНОГО ІНТЕЛЕКТУ

1.1. Сутність штучного інтелекту. Обчислення індексу ШІ

Суперечності між видатними науковцями та технологічними лідерами щодо впливу штучного інтелекту (ШІ) на майбутнє людства стають об'єктом уваги. Ілон Маск, Стівен Гокінг та Білл Гейтс висловлюють побоювання стосовно можливих негативних наслідків розвитку ШІ, у той час як інші, такі як Марк Цукерберг, вбачають позитивні перспективи для людства. Створений на основі ШІ робот Софія, виявляючи свої вміння та наміри, піднімає питання, як штучний інтелект може змінити світ – на краще або гірше.

Народження штучного інтелекту можна віднести до Дартмутської конференції 1956 року, де Джон Маккарті, Марвін Мінські, Натаніель Рочестер і Клод Шеннон ввели термін «штучний інтелект». Конференція заклала основу для галузі та створила ШІ як окрему дисципліну [8]. У наступні роки перші піонери ШІ зосередилися [6] на символічному ШІ, створюючи програми, здатні до символічного міркування та вирішення проблем. Аллен Ньюелл і Герберт А. Саймон розробили Logic Theorist, першу програму штучного інтелекту, у 1955 році, а Артур Самуель створив першу самонавчальну програму для ігор у 1959 році. Незважаючи на ранній ентузіазм, галузь зіткнулася з невдачами під час «зими ШІ» у 1970-х і 1980-х роках [35].

Протягом 2023 року розгорнулися дискусії стосовно впливу розвитку інструментів штучного інтелекту на сферу мистецтва і його цінність. Деякі

українські художники висловлювали обурення, стверджуючи, що інструменти, наприклад, Midjourney, можуть привести до деперсоналізації та дискредитації самого мистецтва. Amazon, з свого боку, розпочав видаляти книги зі свого сайту, які, ймовірно, були створені за допомогою систем штучного інтелекту. Однак у 2023 році виникли серйозні виклики для розвитку інструментів штучного інтелекту [7], таких як нереалістичні очікування, зменшення фінансування та труднощі в розробці передових технологій. Це призвело до зменшення інтересу та ресурсів для досліджень в галузі ШІ. Замість цього, увага переключилася на експертні системи - програми, що імітують досвід людини на основі правил. Хоча компанії вкладали значні кошти в ці системи, обмеження в роботі з невизначеністю та розумінням контексту призвели до нового періоду скептицизму. Все ж таки, відродження штучного інтелекту прийшло разом із прогресом у галузі машинного навчання. Перемога 1997 року компанії IBM Deep Blue над чемпіоном із шахів Гаррі Каспаровим стала важливою віхою. У 2000-х роках набули популярності методи машинного навчання, зокрема нейронні мережі. Великі дані та глибоке навчання. Нинішня епоха (2010-ті роки – теперішній час) є свідком розвитку штучного інтелекту завдяки великим даним і глибокому навчанню. Покращена обчислювальна потужність і величезні набори даних дозволяють глибоким нейронним мережам досягати безпрецедентних результатів у таких завданнях, як розпізнавання зображень, обробка природної мови та ігри [9].

Штучний інтелект з його потенціалом стимулювати інновації, підвищувати ефективність і вирішувати складні проблеми стає невід'ємним аспектом глобальної конкурентоспроможності. Глобальний індекс служить барометром, який вимірює готовність і прогрес націй у використанні трансформаційної сили ШІ. Це дослідження заглиблюється в незліченну кількість показників, інкапсульованих у Глобальному індексі штучного інтелекту, розбираючи пули талантів, надійність інфраструктури, операційне

середовище, зусилля з досліджень і розробок, державні стратегії та комерційний ландшафт.

У проміжку 2023 року виникли дискусії щодо впливу розвитку інструментів штучного інтелекту на мистецтво та його цінність. Деякі українські художники висловлювали негодження, стверджуючи, що інструменти, такі як Midjourney, можуть призвести до втрати індивідуальності та дискредитації суті мистецтва. З іншого боку, Amazon розпочав видалення книг зі свого сайту, припускаючи, що вони можуть бути створені за допомогою систем штучного інтелекту. Тим не менш, протягом 2023 року виникли значущі виклики для розвитку інструментів ШІ, такі як нереалістичні очікування, скорочення фінансування та труднощі в розробці передових технологій [49]. Це викликало зниження інтересу та ресурсів для досліджень в галузі ШІ. Натомість, акцент був зміщений на експертні системи - програми, що імітують досвід людини на основі правил. Хоча компанії вкладали значні кошти у ці системи, обмеження в роботі з невизначеністю та розумінням контексту викликали новий період скептицизму. Проте відродження штучного інтелекту прийшло із прогресом у галузі машинного навчання [36].

Вищий індекс штучного інтелекту (ШІ) вказує на більший рівень розвитку сфери штучного інтелекту в даній країні. Індекс може відображати різні аспекти розвитку ШІ, такі як наукові дослідження, інноваційність, інфраструктура, освіта, промисловість та інші фактори. Зазвичай вищий індекс свідчить про те, що країна має ефективну і високотехнологічну систему ШІ, розвинену інфраструктуру для досліджень та розробок, високий рівень освіти в галузі штучного інтелекту, а також активність у сфері стартапів та інновацій.

1.2 Засади дисперсійно-кореляційного та регресійного аналізу

Дисперсійно-кореляційний метод – це потужний аналітичний підхід, який використовується в аналізі даних для виявлення закономірностей, зв'язків і варіацій у наборах даних. Це особливо цінно під час вивчення багатовимірних наборів даних, таких як ті, що знаходяться в Глобальному індексі ШІ. Цей метод об'єднує два ключові статистичні методи: дисперсійний аналіз і кореляційний аналіз [21].

Кореляційний аналіз – це статистичне дослідження (стохастичної) залежності між випадковими величинами (англ. correlation – взаємозв'язок). У найпростішому випадку досліджують дві вибірки (набори даних), у Кореляційний аналіз – це статистичне дослідження взаємозв'язку між випадковими величинами, що базується на стохастичних залежностях. У простіших випадках вивчаються взаємозв'язки між двома вибірками, а в загальному випадку – між багатовимірними комплексами геологічних параметрів чи об'єктів. Кореляційний аналіз відзначається широким спектром застосувань у різних галузях науки та бізнесу. Застосування кореляційного аналізу тісно пов'язане з регресійним аналізом, іноді його називають кореляційно-регресійним [48]. Головні завдання кореляційного аналізу включають визначення форми зв'язку між величинами, вимірювання щільності (сили) цього зв'язку та виявлення впливу факторів на результативну ознаку.

Дисперсійний аналіз представляє собою систему статистичних методів, які призначені для перевірки гіпотез про взаємозв'язок між конкретною ознакою та досліджуваними факторами, що не мають кількісного опису [52]. Цей метод також використовується для визначення ступеня впливу факторів та їх взаємодії. У спеціалізованій літературі дисперсійний аналіз часто позначається як ANOVA (англійський аббревіативний вираз Analysis of Variations). Розроблений у 1925 році Р. Фішером, цей метод став одним із

ключових інструментів статистичної оцінки результатів експерименту. Він також широко застосовується в аналізі економічної інформації, дозволяючи встановити, наскільки вибіркові показники взаємозв'язку між результативними та факторними ознаками є достатніми для узагальнення отриманих даних на генеральну сукупність. Цей метод є особливо корисним при аналізі вибірок невеликої чисельності, надаючи надійні висновки [37].

Дисперсійний аналіз ґрунтується на припущенні, що середні значення залежної кількісної змінної відрізняються між групами фактора. Важливість цієї різниці визначається розкладанням загальної дисперсії залежної змінної на складові: міжгрупову дисперсію (за рахунок груп фактора) та внутрішньогрупову (залишкову) дисперсію [47]. Визначаючи співвідношення міжгрупової дисперсії та загальної, можна оцінити частку впливу фактора на залежну змінну (R^2). Значущість цього впливу оцінюється за допомогою F-критерію Фішера: якщо фактичне значення перевищує табличне при рівні значущості 0,95, робиться висновок про значущий вплив фактора на вивчений показник якості. Проте F-критерій не вказує, між якими саме групами фактора спостерігається значуща різниця у середніх значеннях залежної змінної, тому завершальною метою дисперсійного аналізу є виявлення конкретних груп фактора, що спричиняють відмінність середніх значень.

Дисперсійний аналіз, також відомий як аналіз мінливості, зосереджується на розумінні поширення чи мінливості в наборі даних. Характеристики розсіювання, такі як дисперсія та стандартне відхилення, кількісно визначають ступінь відхилення окремих точок даних від центральної тенденції (рис. 1.1) [19]. Формула для розрахунку дисперсії (C^2) виглядає так:

$$C^2 = \frac{\sum (X_i - X_s)^2}{N}$$

де: X_i - кожна окрема точка даних,

X_s є середнім значенням набору даних, i

N - це загальна кількість точок даних.

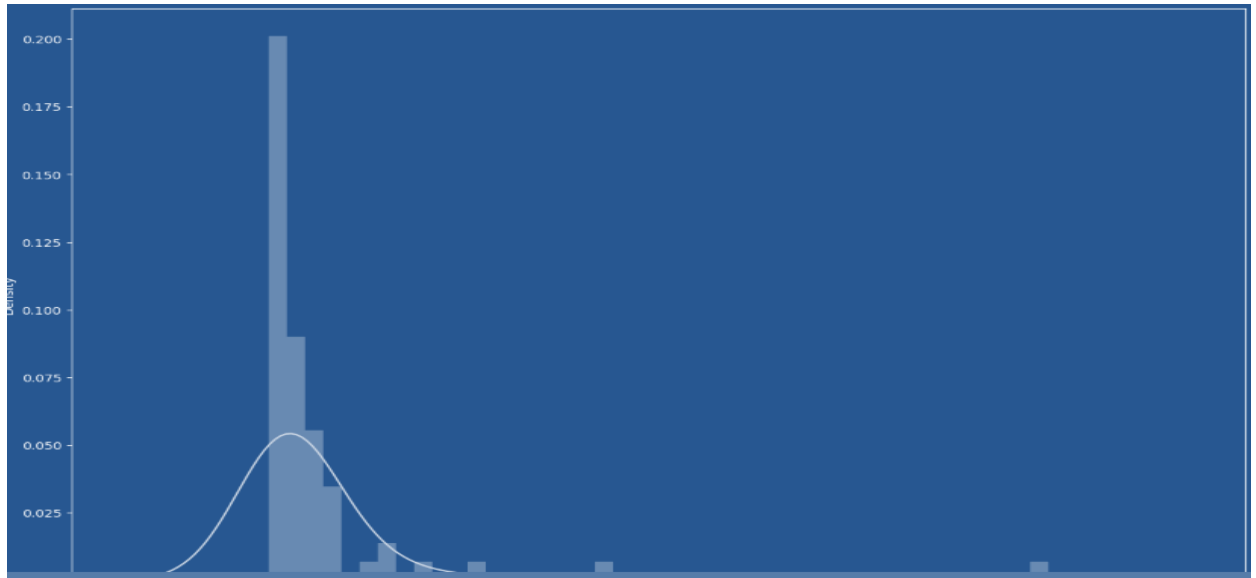


Рисунок 1.1 - Ступінь відхилення окремих точок даних від центральної тенденції

Джерело:[4]

У контексті Глобального індексу штучного інтелекту дисперсійний аналіз допомагає визначити ступінь мінливості ключових показників, таких як талант, інфраструктура, робоче середовище, дослідження та розробки, державна стратегія та комерційні аспекти [4]. Усе це за регіонами, кластерами, групами доходів і політичним режимом. Кореляційний аналіз досліджує силу та спрямованість зв'язків між змінними. Коефіцієнт кореляції Пірсона, кількісно визначає лінійний зв'язок між двома змінними [36]. Якщо ми малюємо це співвідношення у двовимірному просторі (між двома змінними), ми отримуємо пряму лінію (рис. 1.2).

Регресійний аналіз – це статистичний процес вивчення взаємозв'язків між змінними, який включає різноманітні методи моделювання та аналізу декількох змінних. Основна увага при цьому зосереджена на зв'язку між залежною змінною та однією або кількома незалежними змінними.

Незалежна змінна – це фактор, який діє самостійно і не залежить від інших змінних. При коригуванні незалежної змінної рівні залежної змінної можуть змінюватися, а залежна змінна, яка вивчається, є тим, що регресійна модель намагається передбачити. У випадках лінійної регресії кожне спостереження включає значення як залежної, так і незалежної змінної [51]. У загальному вигляді лінійна модель має вигляд об'єкта $x = (x_1, x_2, \dots, x_d)$

Лінійна регресія виконує завдання прогнозування значення залежної змінної (y) на основі заданої незалежної змінної (x). Таким чином, цей метод регресії виявляє лінійну залежність між x (вхід) та y (вихід). Отже, ім'я є лінійною регресією [38]. Якщо ми побудуємо незалежну змінну (x) на осі x та залежну змінну (y) на осі y , лінійна регресія дасть нам пряму лінію, яка найкраще відповідає точкам даних, як показано на рис.1.2.

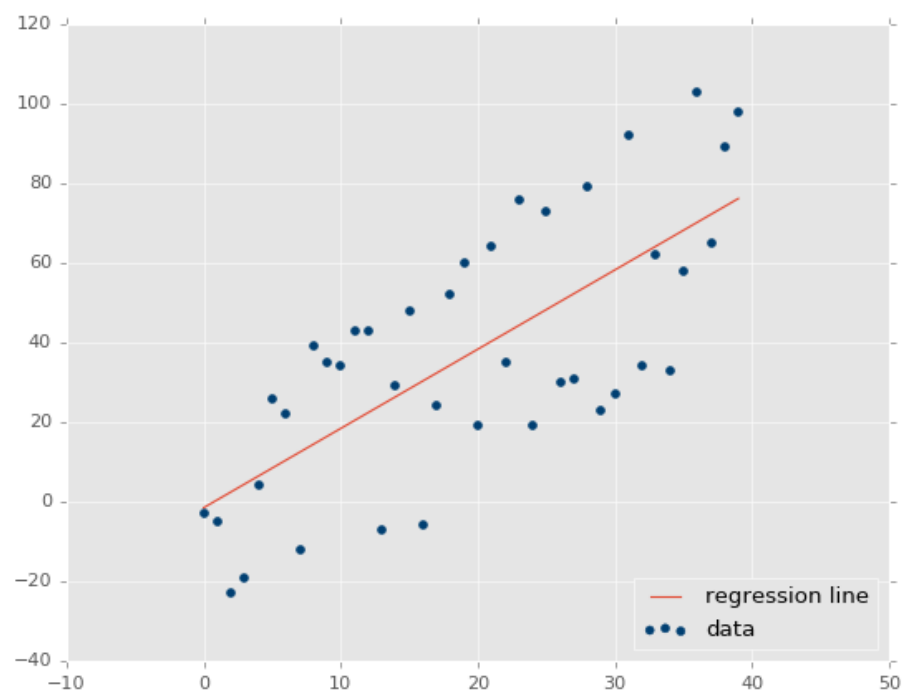


Рисунок 1.2 – Лінія регресії на кореляційному полі даних

Джерело:[4]

Рівняння парної регресії:

$$Y = mx + b,$$

де b – точка перетину, а m – нахил лінії.

Загалом, алгоритм лінійної регресії дає найбільш оптимальне значення для перетину і нахилу (у двох вимірах). Змінні y та x залишаються незмінними, оскільки вони є характеристиками даних і не можуть бути змінені. Значення, які ми можемо контролювати, - це точка перетину (b) і нахил (m). Залежно від значень точки перетину та нахилу може бути кілька прямих ліній. По суті, алгоритм лінійної регресії вписується в кілька рядків у точках даних та повертає рядок, який дає найменшу помилку [39].

У контексті Global III Index кореляційний аналіз дозволяє визначити взаємозалежності між різними факторами [22]. Наприклад, це допомагає визначити, чи корелює високий рівень таланту з надійною інфраструктурою, чи державні стратегії корелюють з розробкою комерційного ІІІ.У нашій роботі будемо визначати відношення інвестицій до множини іновацій і впроваджень. Тобто наскільки фінансове забезпечення та прихильність уряду відноситься до всіх інших індексів що описують такі аспекти як кваліфікованих практиків, інфраструктури доступу, громадській думці, досліджень і дослідників, платформ і алгоритмів.

Іновацій до множини інвестицій і впроваджень. Тобто відношення множини досліджень і дослідників, платформ та алгоритмів до інвестицій та кваліфікованих практиків, інфраструктури доступу, громадській думці [50].

Впроваджень до інвестицій та іновацій. Тобто практиків, інфраструктури та громадській думці до фінансового забезпечення, прихильності уряду та дослідників разом з алгоритмами. Дисперсійно-кореляційний метод поєднує ці два аналізи, щоб забезпечити повне розуміння набору даних [23]. Оцінюючи дисперсію окремих показників і досліджуючи їхні кореляції, дослідники можуть виявити складні зв'язки та відмінності в Глобальному індексі штучного інтелекту. Високий розкид певного показника може вказувати на різні рівні розвитку між націями, тоді як сильні кореляції можуть підкреслювати сфери взаємного впливу між різними аспектами готовності до ІІІ [40].

Дослідницький аналіз даних (EDA) - це те, як ми розуміємо дані, перетворюючи їх із сирої форми в більш інформативну. Цей метод аналізу передбачає використання статистичних і графічних методів для глибшого розуміння структури та характеристик набору даних [24]. Ця теорія викладає ключові принципи та методології, що лежать в основі EDA в контексті дослідної роботи.

EDA [7] починається з узагальнення ключових характеристик і описової статистики набору даних. Це передбачає обчислення таких показників, як середнє значення, медіана, дисперсія та стандартне відхилення. У Глобальному індексі штучного інтелекту ці зведені статистичні дані надають огляд основних тенденцій і мінливості таких показників, як талант, інфраструктура, робоче середовище, дослідження та розробки, державна стратегія та комерційні фактори.

Дисперсійний однофакторний аналіз застосовується у випадках дослідження впливу зміни умов або градацій фактора на зміну результативної ознаки. Основна ідея математичних перетворень у дисперсійному методі полягає в порівнянні дисперсій між факторами і загальною дисперсією, отриманою в ході експерименту. Для проведення однофакторного аналізу необхідно мати не менше трьох рівнів фактора та не менше двох випробувань у кожному рівні. Важливо впевнитися в нормальності розподілу випадкової величини та відсутності відмінностей у дисперсіях підгруп, що можна здійснити за допомогою методів перевірки статистичних гіпотез [41].

Однофакторний аналіз передбачає дослідження розподілу та характеристик окремих змінних ізольовано [25]. Гістограми, прямокутні діаграми та графіки ядерної щільності допомагають візуалізувати розподіл кожного показника. Розуміння форми, центральної тенденції та поширення цих змінних закладає основу для подальшого аналізу.

Двофакторний і багатфакторний аналізи досліджують взаємозв'язки між двома чи більше змінними. Діаграми розсіювання, кореляційні матриці

та теплові карти допомагають ідентифікувати закономірності та залежності в наборі даних. У Глобальному індексі штучного інтелекту двовимірний аналіз може виявити, як талант співвідноситься з дослідженнями та розробкою, тоді як багатовимірний аналіз може розкрити складну взаємодію між кількома показниками [26].

EDA включає ідентифікацію викидів — точок даних, які значно відхиляються від загальної моделі. Ящикові діаграми, діаграми розсіювання та z-показники — це інструменти, які використовуються для виявлення та розуміння впливу викидів на набір даних.

EDA може включати методи перетворення даних, такі як логарифмічне масштабування або нормалізація, щоб вирішити такі проблеми, як асиметрія або гетероскедастичність. Трансформації підвищують інтерпретабельність даних і гарантують, що наступні аналізи не будуть зміщені вихідним масштабом [42].

Візуалізація є угловим каміням EDA, пропонуючи потужний засіб для передачі розуміння. Різноманітні графічні представлення, включаючи точкові діаграми, стовпчасті діаграми та географічні карти, забезпечують візуальну розповідь про набір даних. Візуалізація допомагає передати складні зв'язки та шаблони зрозумілим способом[6].

EDA включає ретельний аналіз шаблонів відсутніх даних. Розуміння обсягу та причин відсутності даних має вирішальне значення для отримання надійних висновків. На основі характеру та впливу відсутніх значень у Глобальному індексі штучного інтелекту можна застосовувати такі методи, як імпутація чи виключення.

EDA — це не лише підтвердження гіпотез, а й створення нових. Початкові дослідження можуть призвести до формулювання гіпотез, які можуть бути ретельно перевірені в наступних аналізах. Наприклад, EDA може викликати гіпотези про зв'язок між інвестиціями в інфраструктуру та розробкою комерційного ІІІ.

EDA часто є повторюваним процесом. Початкові висновки можуть викликати додаткові запитання та глибші дослідження. Коли дослідники виявляють ідеї, вони вдосконалюють свій аналіз і повертаються до попередніх етапів EDA. Цей циклічний підхід забезпечує всебічне дослідження набору даних.

Політичні режими представляють головні системи управління в державі, що формує взаємодію між урядом та його громадянами. Курсова робота з політичних режимів передбачає комплексне дослідження цих систем, враховуючи історичні передумови, сучасні прояви та наслідки для суспільства.

Центральним у цій теоретичній структурі є розуміння того, що політичні режими є не статичними утвореннями, а динамічними структурами, які розвиваються з часом. Історичні контексти та соціокультурні фактори впливають на виникнення та розвиток політичних режимів, формуючи їх ідеології, структури та методи правління. Заглиблюючись у ці історичні корені, дослідники отримують цінну інформацію про основи, на яких побудовані сучасні політичні системи.

Динаміка лідерства в політичних режимах відіграє ключову роль у визначенні характеру управління. Концентрація влади, процеси прийняття рішень і роль лідерства у формуванні політичних напрямів є критичними аспектами аналізу [3]. У курсовій роботі має бути розглянуто взаємозв'язок між стилями лідерства та загальною стабільністю та ефективністю політичного режиму.

Дослідження державних інституцій і структур управління забезпечує лінзу, через яке можна зрозуміти розподіл і здійснення політичної влади. Наявність або відсутність системи стримувань і противаг, роль судової системи та функціонування законодавчих органів – усе це сприяє загальному функціонуванню політичного режиму.

Ідеологічні основи є невід'ємною частиною дослідження політичних режимів. Незалежно від того, чи базується вона на принципах демократії,

авторитаризму, соціалізму чи інших принципах, переважаюча ідеологія впливає на політичні рішення, участь громадян і загальний характер управління. Аналіз ідеологічних засад допомагає контекстуалізувати дії та політику режиму.

Відносини між політичними режимами та громадянами є динамічною взаємодією, яка формує соціальну структуру суспільства. Соціальна динаміка, участь громадян і громадянська активність дають змогу зрозуміти рівень політичної обізнаності та інклюзивність процесів прийняття рішень. Вивчення чутливості політичних режимів до суспільних потреб і проблем є вирішальним для розуміння їх легітимності.

Порівняльний аналіз розширює теоретичну базу, розміщуючи певний політичний режим у ширшому глобальному чи регіональному контексті. Проведення паралелей і контрастів з іншими режимами пропонує цінні перспективи щодо сильних і слабких сторін та унікальних характеристик досліджуваного режиму.

Політична стабільність і здатність долати виклики є ключовими факторами. Стабільність часто асоціюється з довголіттям і стійкістю політичних режимів, але розуміння того, як ці режими справляються з внутрішніми незгодами, зовнішнім тиском і суспільними змінами, сприяє детальнішому аналізу.

1.3 Основи побудови багатofакторної моделі

Багатofакторна регресія - це статистичний метод дослідження, спрямований на вивчення взаємозв'язку між залежною змінною та кількома незалежними змінними. Цей метод відрізняється від простої регресії тим, що враховуються впливи не одного, а декількох факторів на змінну відгуку. Багатofакторна регресія дозволяє моделювати більш складні взаємозв'язки

порівняно із простою регресією, де враховується лише один фактор, що впливає на змінну відгуку. У багатофакторній регресії розглядаються одночасно впливи різних факторів, що дозволяє отримати більш повне та реалістичне уявлення про відносини між змінними [44]. Основною метою багатофакторної регресії є створення математичної моделі, яка враховує взаємодії різних факторів і призначена для прогнозування значень залежної змінної на основі введених значень незалежних змінних. Цей метод широко використовується у різних галузях, таких як економіка, фінанси, медицина, соціологія та інші. Наприклад, в економіці його можна використовувати для аналізу впливу різних факторів, таких як доход, ціни на товари, рівень безробіття тощо, на обсяг виробництва.

Мультиколінеарність вказує на наявність тісної лінійної залежності або кореляції між двома чи більше пояснювальними змінними. Це явище може негативно позначитися на кількісних характеристиках економетричної моделі або ускладнити її побудову [27]. Наприклад, мультиколінеарність виникає, коли два предиктори у моделі представлені однаковою інформацією, але в різних метричних шкалах, таких як ріст людини в сантиметрах і дюймах. Це може призвести до коефіцієнта кореляції близько до 1. Для уникнення мультиколінеарності рекомендується включати в модель лише одну з двох змінних.

Гетероскедастичність - це властивість послідовності випадкових величин, яку вивчають у курсі економетрики. Якщо дисперсія залишків у економетричному моделюванні змінюється для кожного спостереження або для груп спостережень, це явище називається гетероскедастичністю.

Побудова моделей множинної регресії включає в себе кілька етапів:

- вибір форми зв'язку (рівняння регресії);
- відбір факторних ознак;
- забезпечення достатнього обсягу сукупності. Визначення типу рівняння може тим, що для будь-якої форми залежності можна

вибрати цілий ряд рівнянь, які певною мірою будуть описувати ці зв'язки. Особливе значення мають лінійні моделі в силу простоти і логічності їх економічної інтерпретації.

Багатофакторна регресія – це складний та міцний метод, що передбачає послідовний ряд етапів для аналізу впливу кількох факторів на залежну змінну. Розглянемо детально кожен етап цього процесу:

- **Визначення змінних:** Перший крок полягає в чіткому визначенні змінних. Зазвичай, є одна залежна змінна, яку ми прагнемо пояснити, і декілька незалежних змінних, які можуть впливати на залежну змінну. Наприклад, якщо ми досліджуємо вартість житла, то ціна може залежати від площі, кількості спалень, розташування та інших факторів.
- **Збирання даних:** Для проведення багатофакторної регресії необхідно мати наявності дані, які включають значення всіх змінних для кожного спостереження. Дані можуть бути отримані з опитувань, експериментів, баз даних, тощо.
- **Побудова моделі:** На цьому етапі створюється математична модель, яка виражає взаємозв'язок між залежною та незалежними змінними. Модель може бути лінійною або нелінійною в залежності від природи досліджуваної проблеми.
- **Оцінювання коефіцієнтів:** Коли модель побудована, необхідно визначити коефіцієнти регресії – числа, які вказують на міру впливу кожної незалежної змінної на залежну. Цей процес використовує методи мінімізації помилок, такі як метод найменших квадратів.
- **Аналіз результатів:** Після визначення коефіцієнтів регресії проводиться аналіз їхньої статистичної значущості. Це дозволяє визначити, які фактори суттєво впливають на залежну змінну.

- Прогнозування: Завершальний етап включає в себе використання розробленої моделі для прогнозування значень залежної змінної на основі введених значень незалежних змінних.

У багатофакторних моделях мультиколінеарність вказує на високу взаємну кореляцію між незалежними змінними, що може суттєво вплинути на стабільність та точність оцінок параметрів моделі [45]. Засоби виявлення мультиколінеарності включають:

Обчислення варіаційного коефіцієнта: Цей коефіцієнт вказує на ступінь розподілу змінної між різними незалежними факторами. Значення варіаційного коефіцієнта може вказувати на наявність мультиколінеарності [28].

- Матричний підхід: Обчислення оберненої кореляційної матриці може виявити великі або навіть нескінченні значення, що є показником проблеми мультиколінеарності.
- Аналіз детермінанта матриці: Детермінант матриці служить оцінкою ступеня мультиколінеарності. Малий детермінант може вказувати на наявність проблеми.
- У разі виявлення мультиколінеарності можуть бути застосовані заходи, такі як зменшення кількості змінних, об'єднання корельованих змінних або використання методів регуляризації.

Перевірка гетероскедастичності вказує на нерівномірність дисперсії помилок у моделі, що може призвести до неточних оцінок стандартних помилок та неправильних інтерпретацій. Кілька методів виявлення гетероскедастичності включають:

- Графічні методи: Дослідження графіків залишкових діаграм та гістограм помилок може виявити відхилення в дисперсії.

- Тести на базі моделі: Тести, такі як тест Бройша-Пагана або тест Голдфелда-Квандта, використовують статистичні показники для оцінки гомоскедастичності.
- Тести на основі регресії: Застосування додаткових регресій для аналізу залишкових дисперсій може допомогти виявити гетероскедастичність.

Якщо виявлена гетероскедастичність, можливі підходи включають узагальнення моделі, використання ваг, що зворотно пропорційні дисперсії, або перетворення змінних для зменшення нерівномірності дисперсії.

Перевірка на мультиколінеарність у багатофакторних моделях є важливим кроком для забезпечення надійності та точності отриманих оцінок. Мультиколінеарність виникає, коли незалежні змінні у моделі є високо корельованими, що може призвести до проблем в оцінках параметрів та їхньої інтерпретації. Одним з методів виявлення мультиколінеарності є обчислення варіаційного коефіцієнта для кожної змінної. Високий варіаційний коефіцієнт свідчить про велику змінливість змінної, що може бути індикатором мультиколінеарності. Матричний підхід включає обчислення оберненої кореляційної матриці. Якщо обернена матриця має великі значення чи навіть нескінченні, це може бути ознакою мультиколінеарності [29]. Аналіз детермінанту матриці також може виявити проблему. Малий детермінант свідчить про високу кореляцію між змінними. Якщо мультиколінеарність виявляється, можливі заходи включають редукцію кількості змінних, об'єднання корельованих змінних або використання регуляризаційних методів для зменшення впливу кореляції між незалежними змінними.

Перевірка багатофакторної моделі на гетероскедастичність є важливим етапом аналізу регресії. Гетероскедастичність вказує на нерівномірність дисперсії помилок моделі. Для виявлення цього явища можна застосовувати різні методи. Один із методів - це графічний аналіз. Дослідження графіків

залишкових діаграм та гістограм помилок може розкрити нерівномірність дисперсії. Додатково використовують статистичні тести, наприклад, тест Бройша-Пагана чи тест Голдфельда-Квандта [46]. Ці тести базуються на порівнянні дисперсій помилок для різних підгруп або різних частин вибірки. Для обрання правильного методу корисно також використовувати допоміжні регресійні підходи. Зокрема, можна дослідити залишкові дисперсії в залежності від прогнозованих значень, щоб виявити гетероскедастичність. У випадку виявлення гетероскедастичності, можливі корекційні дії включають в себе застосування ваг, що зворотно пропорційні дисперсії, або використання альтернативних методів оцінювання, які враховують нерівномірність дисперсії в помилках моделі.

Висновки до розділу 1

Дослідження історії штучного інтелекту (ШІ) надає фундаментальне розуміння еволюції цієї трансформаційної сфери, де відстеження історичної траєкторії від концептуалізації до сучасних застосувань свідчить про стійкість і адаптивність технологій ШІ. Розпізнавання віх, проривів і змін парадигми в історії штучного інтелекту служить компасом для оцінки поточного ландшафту та передбачення майбутніх досягнень.

Використання методу дисперсії-кореляції в аналізі даних пропонує надійний підхід до розгадування складних наборів інформації, об'єднуючи дисперсійний та кореляційний аналіз для глибокого дослідження взаємодії між різними параметрами. Цей інтегрований метод дозволяє детально розглядати мінливість окремих показників та їх взаємозв'язки, забезпечуючи точний та комплексний аналіз даних.

У контексті аналізу глобального індексу штучного інтелекту, багатофакторна регресія виступає як важливий метод для вивчення взаємозв'язків між різними факторами, які впливають на розвиток цієї сфери.

РОЗДІЛ 2

ДОСЛІДЖЕННЯ ЧИННИКІВ ІНДЕКСУ ШІ

2.1 Кластеризація країн за різними ознаками

Кластеризація – це процес групування набору фраз у кластери, які об'єднують схожі висловлення. У сучасному машинному навчанні цей метод є необхідною складовою для обробки та аналізу даних у рамках глобального індексу штучного інтелекту. Кластеризація визначається як процес групування подібних об'єктів разом у кластери або групи, що є ключовим елементом оптимізації систем штучного інтелекту. Розподіл схожих об'єктів за категоріями сприяє розпізнаванню патернів та полегшує аналіз даних, покращуючи ефективність прийняття рішень та точність прогнозування. Кластеризація в індексі штучного інтелекту використовується для оптимізації класифікації, сегментації та рекомендаційних систем, адаптуючи штучний інтелект до змін в обсягах даних та виявляючи невідомі структури та зв'язки в наборі даних [30].

Основна мета кластеризації полягає у визначенні "природних" груп в даних, де об'єкти в одній групі схожі між собою, але відрізняються від об'єктів інших груп. У машинному навчанні цей процес дозволяє виявити подібність між об'єктами та створити абстрактні моделі для роботи з складними даними. Різні методи кластеризації, такі як k-середні, ієрархічна кластеризація та методи на основі найближчих сусідів, використовують різні метрики для визначення схожості та розподілу об'єктів. В контексті індексу штучного інтелекту кластеризація виконує ключову роль у аналізі та обробці великих обсягів даних, щоб поліпшити роботу алгоритмів штучного інтелекту. Цей процес допомагає відокремити об'єкти з схожими характеристиками, забезпечуючи більш ефективні алгоритми для розв'язання

завдань, пов'язаних з розпізнаванням патернів та класифікацією. Створені кластери служать основою для подальших аналізів та вивчення складних залежностей між об'єктами у контексті інтелектуальних систем. Кластеризація також може пристосовуватися до динамічних змін у великих потоках даних, роблячи її важливим елементом для систем штучного інтелекту, орієнтованих на реальний час. Загалом, сутність кластеризації в машинному навчанні для індексу штучного інтелекту полягає в ефективному структуруванні та аналізі обширних даних, забезпечуючи більшу точність та ефективність систем штучного інтелекту.

Головна мета кластеризації в індексі штучного інтелекту полягає в створенні оптимальних груп для подальшої класифікації об'єктів. Мета виявлення природних груп об'єктів дозволяє інтелектуальним системам виявляти та аналізувати закономірності в наборі даних. Кластеризація сприяє поліпшенню алгоритмів розпізнавання патернів, що важливо для систем, що працюють із зображеннями, текстами чи звуками. Розподіл об'єктів на групи спрощує взаємодію систем штучного інтелекту з великим обсягом різноманітних даних. Виявлення груп подібних об'єктів сприяє ефективній роботі систем рекомендацій, що важливо для персоналізованих послуг. Метою є створення прогностичних моделей, заснованих на групах схожих об'єктів, що дозволяє системам штучного інтелекту передбачати події та тренди. Кластеризація спроможна динамічно адаптуватися до змін у даних, підтримуючи високу точність у реальному часі. Кластеризація допомагає створювати персоналізовані та адаптивні інтерфейси, поліпшуючи користувацький досвід. Оптимізація обробки даних за допомогою кластеризації дозволяє економити обчислювальні ресурси та підвищує швидкодію систем штучного інтелекту. Загалом, мета кластеризації в індексі штучного інтелекту полягає в розвитку адаптивних, гнучких та ефективних систем, здатних адаптуватися до різноманітних вимог та змін у навколишньому середовищі.

Кластеризація в індексі штучного інтелекту використовується для покращення класифікації об'єктів, забезпечуючи більш точні та раціональні результати. Системи реклами та персоналізованих послуг використовують кластеризацію для ефективної сегментації аудиторії та забезпечення рекламних рішень. Кластеризація допомагає розкривати та аналізувати патерни поведінки користувачів, полегшуючи розуміння їхніх вподобань та потреб. Використання кластеризації в індексі штучного інтелекту дозволяє ефективно виявляти аномалії та непередбачені зміни в системі. Системи рекомендацій використовують результати кластеризації для надання користувачам більш точних та персоналізованих рекомендацій. Кластеризація служить для ефективної диференціації та організації потоків даних в системах обробки інформації. Системи штучного інтелекту використовують кластеризацію для адаптації до змін у потребах користувачів та покращення якості обслуговування.

Країни потужні гравці представляють країни зі значним глобальним впливом у різних сферах, включаючи політичну, економічну та військову сфери. Теоретична основа досліджує історичні корені їхньої могутності, досліджуючи такі фактори, як історична спадщина, військовий потенціал, економічна міць і дипломатична доблесть. Він заглиблюється в динаміку влади та стратегічні міркування, які позиціонують ці країни як головних гравців на світовій арені.

Традиційні чемпіони – це країни, які історично мали значний вплив і керівні ролі в певних регіонах або доменах. Теоретична основа досліджує стійкі фактори, що сприяють їх традиційному чемпіонському статусу, включаючи історичні альянси, регіональне домінування та культурні зв'язки. Вона досліджує, як ці країни орієнтуються в глобальній динаміці, що розвивається, зберігаючи свої усталені ролі та стосунки. Також це можуть бути країни західного світу [18].

Розвинені країни – це країни, які переживають швидке економічне зростання, технологічний прогрес і зростаючу світову популярність.

Теоретична основа розглядає рушійні сили, що стоять за їх сходженням, включаючи економічну політику, інновації та геополітичне позиціонування. Він досліджує виклики та можливості, з якими стикаються ці країни, коли вони перетворюються з нових гравців на впливових акторів на світовій арені.

Країни, що розвиваються, – це ті, що перебувають на ранніх стадіях розвитку або міжнародної участі, з потенціалом стати помітним у майбутньому. Теоретична основа зосереджена на факторах, що впливають на їхній початковий статус, таких як економічні реформи, геополітичні стратегії та суспільні трансформації. Він досліджує, як ці країни орієнтуються в складних глобальних справах і позиціонують себе для майбутнього впливу.

Дослідження, розвиток, таланти та комерційні змінні мають найбільш відчутний внесок у загальну оцінку індексу ШІ. Урядова стратегія та інфраструктура мають менший вплив на загальний індекс ШІ.

Два потужні гравці (США, Китай) значно перевершують решту країн за показниками Індексу штучного інтелекту; водночас ці країни демонструють різні підходи до побудови та розвитку своєї інфраструктури/технологій/екосистем ШІ, а паралельність (або навіть м'яка «війна») між ними сформує глобальний ландшафт впровадження ШІ в усьому світі. США та Китай лідирують за загальним показником ШІ Index, хоча розрив між загальним балом США та Китаю сам по собі є значним. Деякі з країн у кластері «Розвинені країни» (Ізраїль, Сінгапур, Південна Корея, Нідерланди) перевершують деяких «Країн традиційних чемпіонів» (тобто Німеччину та Францію) за загальним балом. Сполучені Штати є першою країною за кількістю талантів ШІ (середній бал за всіма факторами становить 100)

Індія не входить до першої десятки країн за загальним балом, хоча займає друге місце за балом ШІ Talent (Талантів). У топ-5 країн із найвищим рівнем впровадження штучного інтелекту входять США, Нідерланди, Канада, Китай і Велика Британія. Лише 35% проаналізованих країн мають оцінку впровадження ШІ вище 50. Інфраструктура Китаю найбільше

підходить для ШІ. Суспільство Саудівської Аравії найбільш доброзичливо ставиться до штучного інтелекту (найвищі оцінки операційного середовища). До п'ятірки країн із найвищим рівнем інновацій у сфері штучного інтелекту входять США, Китай, Південна Корея, Австралія та Швейцарія. Проте лише перші дві з вищенаведеного мають бал інноваційності вище 50. Це свідчить про низький рівень розвитку ШІ серед країн (порівняно з лідером – США), незважаючи на часткові докази експоненціальності ШІ у всьому світі.

Крім того, ми можемо спостерігати цікавий «випадок», представлений Південною Кореєю. Ця країна демонструє досить низьку цінність досліджень штучного інтелекту, тоді як її індекс розвитку штучного інтелекту входить до топ-5 списку для всього набору даних. Це може вказувати на те, що країна стане потужним центром для створення прототипів різних орієнтованих на штучний інтелект рішень, продуктів і апаратного забезпечення, де ідеї/патенти, запропоновані фахівцями з інших країн, «продукуються» перевіряються практичними застосуваннями.

Майже 70% країн у наборі даних готові до еволюції ШІ на урядовому рівні, їхній показник показника Стратегії уряду перевищує 50.

До 5 кращих країн за рейтингом Урядової стратегії штучного інтелекту входять Канада, Китай, Саудівська Аравія, Іспанія та Франція (це 5 із 6 країн із результатом вище 90). Сполучені Штати лідирують у комерціалізації ШІ і не мають прямої конкуренції в цій сфері. Хоча Стратегія уряду Сполучених Штатів у сфері штучного інтелекту навіть не входить до топ-5 найсильніших, вона має високі бали за показниками інвестицій у штучний інтелект і посідає перше місце. Наступні чотири країни: Китай, Ізраїль, Сполучене Королівство та Канада; їхній рівень інвестицій значно поступається США.

Загалом, на карті є два потужні гравці – Сполучені Штати та Китай; проте Китай майже на 40% слабший за США. Традиційними чемпіонами (за класифікацією Tortoise Media) є наступні дві країни – Велика Британія та Канада. Проте Німеччина та Франція, два інші традиційні чемпіони, стоять на чотирьох позиціях після Ізраїлю, Сінгапуру, Південної Кореї та

Нідерландів (класифікуються як висхідні зірки). Інші розвинені країни – Австралія, Ірландія, Фінляндія, Данія, Японія та Іспанія. Усі інші країни (серед проаналізованих 62 країн), крім Шрі-Ланки, Єгипту, Кенії, Нігерії та Пакистану, є гравцями, що розвиваються. Останні – це країни що зароджуються у ринку ШІ.

Евклідова відстань, також відома як геометрична відстань або L2-відстань, є однією з основних метрик, що вимірює відстань між двома точками в евклідовому просторі. Це поняття виникло на основі геометричних принципів, сформульованих в давньогрецькі часи відомим математиком Евклідом. На математичному рівні евклідова відстань між двома точками A (x_1, y_1) і B (x_2, y_2) в площині визначається формулою:

$$D(A, B) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Загальніше, у тривимірному або n-вимірному евклідовому просторі формула буде містити відповідні складові координати. Евклідова відстань є ключовим поняттям у численних галузях, таких як математика, фізика, комп'ютерні науки та статистика. У векторному просторі евклідова відстань використовується для вимірювання схожості чи відмінностей між векторами. Наприклад, у машинному навчанні вона може використовуватися для класифікації об'єктів за їхнім розташуванням у просторі. Окрім свого основного призначення у вимірюванні відстаней, евклідова метрика також має важливі властивості, які зробили її важливою у різних областях. Наприклад, вона задовольняє нерівності трикутника, що є важливим математичним властивостями для розв'язання різних завдань. Використання евклідової відстані не обмежується лише математикою. Ця метрика має практичне застосування в географії, біології, медицині та інших галузях. Її різноманітність і важливість роблять її фундаментальним інструментом у багатьох аспектах наукового дослідження та практичних застосувань.

На основі евклідової відстані розвинули ряд теорій та підходів, що розширюють її застосування та надають додаткові можливості для вимірювання відстаней у різних контекстах. Декілька з них включають:

Евклідовий простір: Евклідовий простір – це абстрактний математичний об'єкт, в якому можна визначити відстань між будь-якими двома його елементами. Це загальний термін, який включає в себе не лише простір звичайної тривимірної геометрії, але і аналогічні простори вищих розмірностей, які визначаються евклідовою відстанню.

1. **Евклідові групи:** У теорії груп евклідова метрика використовується для вивчення груп, що зберігають цю відстань. Такі групи вивчаються з точки зору трансляцій, обертань та інших геометричних трансформацій, які не порушують евклідові властивості простору.
2. **Евклідові алгоритми:** Евклідова відстань знаходить широке застосування у різних алгоритмах, таких як алгоритми машинного навчання та обробки зображень. Зокрема, вона може служити метрикою для кластеризації даних, розпізнавання образів та вирішення задач регресії.
3. **Евклідова геометрія:** Евклідова відстань є основним поняттям в евклідовій геометрії, яка досліджує властивості геометричних об'єктів у відповідності до евклідових аксіом. Ця геометрія важлива для розуміння простору та відстаней в математиці та її застосуваннях.
4. **Неевклідова геометрія:** Вивчення геометрії, яка відрізняється від евклідової, вело до розвитку неевклідових геометрій. Наприклад, геометрія на сфері чи гіперболої може використовувати свої власні метричні поняття відстаней.
5. **Топологічні властивості:** Евклідова відстань допомагає визначити топологічні властивості простору, такі як відкриті та замкнені множини, що важливо для розуміння форми простору та його конфігурацій.

Кластерний аналіз з'явився порівняно недавно – у 1939 р. Його запропонував вчений К. Тріон. Дослівно термін "кластер" в перекладі з англійської "cluster" означає гроно, згусток, пучок, група.

Ефективним способом виявлення закономірностей у даних є кластерний аналіз. Кластеризація - це процес категоризації схожих об'єктів або спостережень на основі їхніх особливостей або характеристик. Виявлення прихованих взаємозв'язків у даних можна здійснити, визначивши кластери в даних і отримавши уявлення про їхню глибинну структуру. Кластерний аналіз має широкий спектр застосувань - від маркетингу до біології та соціальних наук. Клієнтів можна сегментувати відповідно до їхніх купівельних звичок, гени можна згрупувати відповідно до їхньої експресії, а людей можна класифікувати відповідно до їхніх особистісних рис.

Кластерний аналіз – це статистичний метод, спрямований на групування схожих об'єктів у кластери для виявлення внутрішньої структури даних. Цей підхід може бути ієрархічним або непараметричним, де перший розглядає кластери як дерево подібності, а другий розділяє дані без врахування їхньої структури [31]. Головна мета кластерного аналізу полягає в виявленні природних груп об'єктів у даних, надаючи можливість ефективно визначати патерни та характеристики. Застосування кластерного аналізу розповсюджене: від наукових досліджень у біології та медицині до бізнес-аналітики та маркетингу. У використанні методів, таких як к-середніх та ієрархічна кластеризація, відображається різноманіття підходів. Важливість візуалізації результатів через графіки та діаграми робить аналіз більш доступним. Кластерний аналіз також знаходить застосування у глибокому навчанні та нейромережах для покращення точності та роботи з великими обсягами даних. Переваги цього методу включають можливість класифікації, сегментації та аналізу великих обсягів інформації.

Важливо враховувати, що кластерний аналіз може бути чутливим до початкових умов та обчислювальних методів, і вимагає об'єктивного вибору кількості кластерів. Незважаючи на обмеження, він залишається потужним

інструментом для розуміння структури та взаємозв'язків у складних наборах даних. Додатково, кластерний аналіз знаходить застосування в різних сферах, включаючи геологію для картирования рудних родовищ, або соціологію для класифікації соціокультурних груп. Триває розвиток нових методів, таких як кластерний аналіз графів, що дозволяє аналізувати складні мережеві структури та виявляти ключові взаємозв'язки в даних. Ці інновації посилюють потенціал кластерного аналізу для вирішення сучасних завдань у багатьох галузях [32].

Кластерний аналіз стосується процесу дослідження та інтерпретації даних за допомогою систематичного та ретельного підходу, заснованого на наукових принципах [14]. Це ключовий компонент наукового методу, який є систематичним способом дослідження та розуміння світу природи. Науковий аналіз включає наступні основні етапи:

Спостереження: Першим кроком є спостереження за явищем або збір відповідних даних. Це може включати використання інструментів, проведення експериментів або збір інформації різними способами [13].

Формулювання гіпотези: на основі спостережень вчені формулюють гіпотезу – перевірене та фальсифіковане пояснення спостережуваного явища. Гіпотеза – це попередня ідея, яку можна перевірити шляхом подальших експериментів або спостережень.

Експеримент: експерименти розроблені та проводяться для перевірки гіпотези. Це передбачає маніпулювання змінними та ретельний контроль умов для збору даних у контрольований та повторюваний спосіб.

Збирання даних: під час експериментів дані збираються систематично. Ці дані можуть мати різні форми, такі як числові вимірювання, спостереження або інша відповідна інформація.

Аналіз даних: потім зібрані дані піддаються різноманітним статистичним і аналітичним методам. Цей крок передбачає

впорядкування, інтерпретацію та осмислення даних, щоб зробити значущі висновки [20].

Набір даних[2] включає сам Глобальний індекс ШІ та сім індикаторів, що впливають на Індекс по 62 країнах, а також загальну інформацію про країни (регіон, кластер, група доходу та політичний режим). Талант, інфраструктура та операційне середовище є факторами групи «індикаторів впровадження штучного інтелекту», яка відображає застосування штучного інтелекту професіоналами в різних секторах, таких як підприємства, уряди та громади. У свою чергу дослідження та розробки ШІ є чинниками групи «показників інновацій», які відображають прогрес, досягнутий у технологіях і методології, що означає потенціал штучного інтелекту для розвитку та вдосконалення. Державна стратегія та комерційна діяльність є факторами «інвестиційної групи індикаторів», яка відображає фінансові та процедурні зобов'язання щодо штучного інтелекту.

Індикатор таланту зосереджується на наявності кваліфікованих практиків для надання рішень штучного інтелекту;

Мета: проаналізувати доступність і якість талантів ШІ в усьому світі.

Гіпотеза: регіони, де приділяється сильна увага освіті, навчанню та утриманню талантів у сфері штучного інтелекту, матимуть потужнішу робочу силу для штучного інтелекту.

Індикатор інфраструктури

Індикатор інфраструктури зосереджується на надійності та масштабі інфраструктури доступу, від електрики та Інтернету до суперкомп'ютерних можливостей.

Мета: оцінити технологічну інфраструктуру, що підтримує розробку та розгортання ШІ.

Гіпотеза: країни з розвиненими телекомунікаціями, високошвидкісним Інтернетом і доступом до обчислювальних ресурсів будуть лідерами в інфраструктурі ШІ.

Індикатор операційного середовища

Індикатор операційного середовища зосереджується на нормативному контексті та громадській думці щодо штучного інтелекту.

Мета: Оцінити наявність та ефективність нормативних актів і політик, пов'язаних із ШІ та наявність і дотримання етичних вказівок і принципів у розробці та розгортанні ШІ.

Гіпотеза: регіони з чіткими етичними принципами та суспільним визнанням додатків штучного інтелекту матимуть більш стабільне зростання штучного інтелекту.

Дослідження та розробки ШІ

Індикатор розробки фокусується на розробці фундаментальних платформ і алгоритмів, на яких спираються інноваційні проекти штучного інтелекту.

Індикатор досліджень фокусується на масштабі спеціалізованих досліджень і дослідників; дослідження кількості публікацій і цитувань у надійних наукових журналах.

Мета: вивчити рівень інвестицій і діяльність у дослідженнях і розробках, пов'язаних зі штучним інтелектом.

Гіпотеза: Країни зі значними інвестиціями в науково-дослідні розробки та співпрацю між академічними колами, промисловістю та урядом сприятимуть розвитку ШІ.

Державна стратегія та комерційна діяльність

Індикатор урядової стратегії зосереджується на глибині прихильності національного уряду штучному інтелекту; дослідження зобов'язань щодо витрат і національних стратегій.

Комерційна діяльність фокусується на рівні активності стартапів, інвестицій та бізнес-ініціатив на основі штучного інтелекту.

Мета: вивчити рівень інвестицій, стартапів і національних стратегій на основі штучного інтелекту та у прихильності штучному інтелекту національного уряду країн.

Гіпотеза: Країни зі значними інвестиціями у науково-дослідні розробки та стартапів ШІ, прихильним урядом до штучного інтелекту ,сприятимуть розвитку ШІ.

2.2 Дисперсійно-кореляційний метод аналізу

Дисперсійно-кореляційний аналіз (ДКА) – це статистичний метод, спрямований на вивчення впливу одного або декількох факторів на змінні вимірювання. Цей метод дозволяє визначити, чи є статистично значимим вплив факторів на результати дослідження та взаємозв'язок між змінними. ДКА включає аналіз дисперсії, який оцінює розподіл варіації в групах, та кореляційний аналіз, який визначає ступінь взаємозв'язку між різними змінними. Основна ідея - визначити, чи є різниці в середніх значеннях груп статистично значущими, та яка частина варіації може бути пояснена факторами. ДКА застосовується у різних галузях, включаючи психологію, соціологію, медицину та економіку. Основною метою є виявлення взаємозв'язку між змінними та визначення вагомості внеску кожного фактора у варіацію [33].

Одним із ключових етапів ДКА є визначення гіпотези та вибір репрезентативної вибірки. Для проведення аналізу використовуються різні статистичні показники, такі як F-статистика та кореляційні коефіцієнти. Важливим аспектом ДКА є інтерпретація результатів, яка визначає, чи існують статистично значущі відмінності між групами та які фактори найбільше впливають на змінні. У світлі сучасних тенденцій в аналізі даних, ДКА залишається актуальним методом, особливо при використанні комп'ютерних програм для більш точних та складних обчислень. Такий підхід дозволяє дослідникам отримувати більше інформації зі своїх досліджень та вдосконалювати методологію наукових вивчень. При

існуванні погодженого зміни двох або більше ознак говорять про існування між ними кореляційної зв'язку.

Для того щоб дослідити базу даних, буде використовуватись мова програмування Python зі бібліотекою Pandas [11].

Pandas - це пакет для мови програмування Python, який надає швидкі, гнучкі та виразні структури даних, спеціально розроблені для ефективної роботи з "реляційними" або "міткими" даними. Він створений з метою забезпечити легкість використання та інтуїтивно зрозуміле програмування при роботі з такого роду даними. Основною метою Pandas є стати ключовим інструментом високого рівня для ефективного аналізу реальних даних в середовищі Python. Крім того, він проголошує ширший замір - стати найпотужнішим і найгнучкішим інструментом для аналізу та обробки даних з відкритим кодом, що буде доступний для використання будь-якою мовою програмування. Здебільшого, Pandas успішно виконує ці завдання та займає важливе місце в інструментах для роботи з даними в екосистемі Python (рис.2.1).

	Country	Talent	Infrastructure	Operating_Environment	Research	Development	Government_Strategy	Commercial	Total_score	Region	Cluster	Income group	Political_regime
0	United States of America	100.00	94.02	64.56	100.00	100.00	77.39	100.00	100.00	Americas	Power players	High	Liberal democracy
1	China	16.51	100.00	91.57	71.42	79.97	94.87	44.02	62.92	Asia-Pacific	Power players	Upper middle	Closed autocracy
2	United Kingdom	39.65	71.43	74.65	36.50	25.03	82.82	18.91	40.93	Europe	Traditional champions	High	Liberal democracy
3	Canada	31.28	77.05	93.94	30.67	25.78	100.00	14.88	40.19	Americas	Traditional champions	High	Liberal democracy
4	Israel	35.76	67.58	82.44	32.63	27.96	43.91	27.33	39.89	Middle East	Rising stars	High	Liberal democracy
...	...	...	...	...	...	...	...	...	...	...	...	...	...
57	Sri Lanka	6.27	34.64	35.79	0.12	0.95	35.57	0.09	6.62	Asia-Pacific	Nascent	Lower middle	Electoral democracy
58	Egypt	1.11	38.84	0.00	2.08	1.54	68.72	0.31	4.83	Middle East	Nascent	Lower middle	Electoral autocracy
59	Kenya	0.75	14.11	29.84	0.07	12.15	7.75	0.31	2.30	Africa	Nascent	Lower middle	Electoral autocracy
60	Nigeria	2.74	0.00	50.10	0.45	2.06	7.75	0.33	1.38	Africa	Nascent	Lower middle	Electoral autocracy
61	Pakistan	8.00	2.43	12.48	2.17	1.09	13.92	0.27	0.00	Asia-Pacific	Nascent	Lower middle	Electoral autocracy

Рисунок 2.1 – Візуалізація даних

Джерело:[2]

De Talent, Infrastructure, Operating Environment, Research, Development,

Government, Strategy, Commercial, Total score – індекси (рис. 2.1). Country, Region, Cluster, Income group, Political regime – Країна, регіон, кластер розвитку, група доходу, політичний режим відповідно. Проаналізуємо неперервні змінні описовою статистикою за допомогою функції «describe()» (рис. 2.2).

	Talent	Infrastructure	Operating Environment	Research	Development	Government Strategy	Commercial	Total score
count	62.000000	62.000000	62.000000	62.000000	62.000000	62.000000	62.000000	62.000000
mean	16.803065	63.503710	66.925484	16.610000	14.824677	57.865645	6.171935	23.914677
std	15.214963	20.217525	20.000424	17.413996	19.419279	26.252448	14.029632	15.123586
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	7.365000	55.857500	58.107500	3.032500	1.202500	41.030000	0.697500	14.805000
50%	13.445000	65.230000	69.505000	12.930000	9.005000	63.930000	2.585000	23.220000
75%	24.567500	75.947500	80.500000	25.412500	19.980000	77.952500	5.307500	30.487500
max	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000

Рисунок 2.2 - Візуалізація даних

Джерело:[2]

Рядок «count» вказує кількість рядків кожного індексу, mean – середнє значення, «std» – стандартне відхилення, min – мінімальне значення. Значення «25%» - нижній центиль, «50%» - центиль що дорівнює медіані і «75%» - верхній центиль. Кожен центиль вказує кількість оцінок що менше певного значення. «max» – максимальне значення. Описова статистика включає статистику, яка підсумовує центральну тенденцію, дисперсію та форму розподілу набору даних, за винятком пустих значень [10].

Ми бачимо, що у наборі даних немає відсутніх значень, там представлено 62 країни (записи), щоб описати їхній потенціал ІІІ станом на 2023 рік. У наборі є 8 числових і 4 категоріальні змінні. Усі числові змінні мають значення в діапазоні від 0 до 100 [5]. Змінну загальної оцінки можна вважати цільовою змінною, тоді як інші - можуть описувати різні впливи на неї.

Робимо діаграми [12] і маємо такий опис їх (рис. 2.3):

- Інфраструктура та операційне середовище трохи зміщені вправо
- Інші числові змінні значно зміщені вліво
- Значення «Талант», «Дослідження», «Розробка», «Комерція» та «Загальний бал» мають чіткі викиди в правій частині розподілу (воно пов'язане з двома країнами в кластері потужних країн)



Рисунок 2.3 – Розподіли щільностей основних чинників індексу ІІІ

Джерело:[2]

Спостерігаються одномодальні розподіли 1, 2, 4, 5, 7 (Талант, Інфраструктура, Дослідження, Розвиток, Комерція) та двомодальні 3, 6, 8 (Операційна інфраструктура, Державна стратегія, загальний рахунок).

Для виявлення зв'язків між чинниками побудовано кореляційну матрицю яка відображена на рисунку 2.4

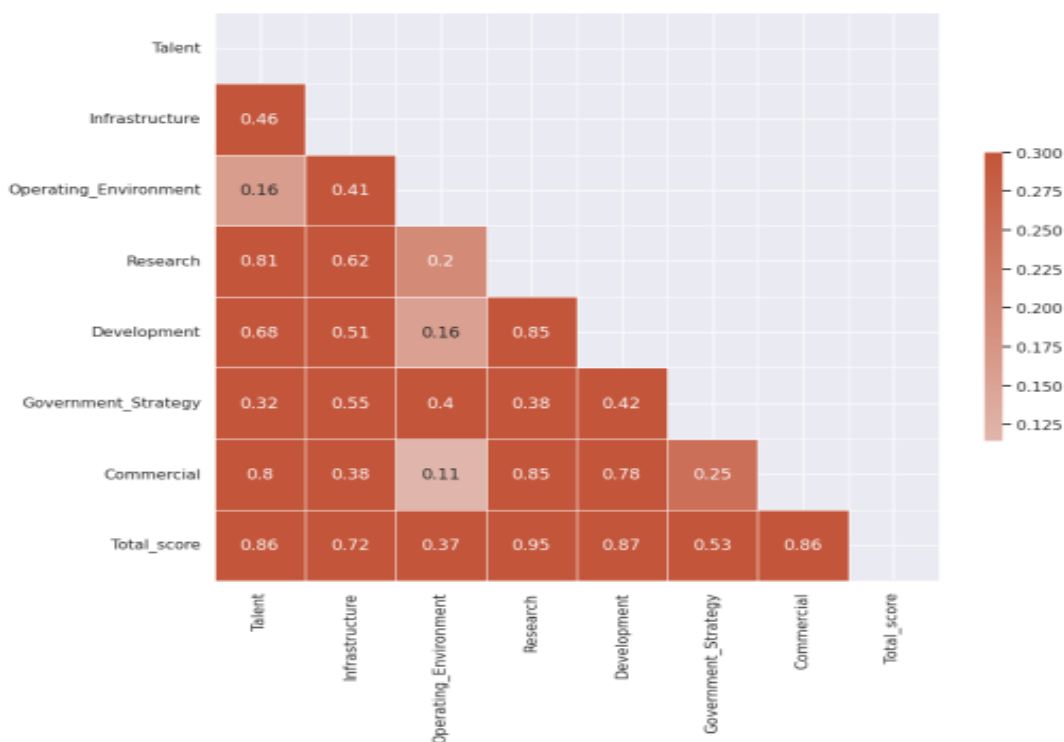


Рисунок 2.4 – Матриця кореляцій між змінними

Джерело:[2]

Спостерігається досить тісний зв'язок між окремими змінними, що свідчить про наявність мультіколінеарності, тому в багатофакторну модель включені лише ті змінні, що мають найтісніший зв'язок з y (індекс ШІ). Найбільші вагові коефіцієнти у commercial, development та research. У загальному модель виглядає наступним чином:

$$y_i = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_3 + e_i, \quad i = \overline{1, n}$$

де

x_1 - commercial

x_2 - development

x_3 – research

Замінімо $x_1 x_2 x_3$ назвами чинників, тоді модель матиме вигляд:

$$y_i = a_0 + a_1 \cdot com + a_2 \cdot dvlp + a_3 \cdot rsch + e_i, \quad i = (\overline{1, n})$$

Коефіцієнти мають наступні числові значення:

$$a_0 = 11.286, a_1 = 0.561, a_2 = 0.157, a_3 = 0.156, R^2 = 0.856$$

Тоді отримуємо модель в аналітичному вигляді:

$$y_i = 11,286 + 0,561 \cdot com + 0,157 \cdot dvlp + 0,156 \cdot rsch + e_i, \quad i = (\overline{1, n})$$

У таблиці 1.1 наведено масив даних, що характеризує параметри та інші показники моделі.

Таблиця 2.1 - Результати обчислення характеристик багатофакторної моделі

	a3	a2	a1	a0
	0.1561305	0.1565141	0.5612878	11.286494
	0.1364431	0.0574118	0.0793583	0.8909312
R2	0.8562793	4.5103578	#Н/Д	#Н/Д
	113.20086	57	#Н/Д	#Н/Д
	6908.6467	1159.5697	#Н/Д	#Н/Д

Джерело:[53]

Загальний бал сильно корелює зі змінними Дослідження (0,95), Розвиток (0,87), Талант (0,86) і Комерція (0,86)

Загальна оцінка відображає середні кореляції зі змінними урядової стратегії та операційного середовища (тобто внесок таких змінних у загальну оцінку є меншим, ніж сильно корельовані змінні).

У роботі здійснено кластерний аналіз за відстанями найближчого та найдальшого сусідів. Об'єднання виконувалось за кожною окремою ознакою.

Загальна оцінка проекту, що є ключовим показником його успішності, демонструє високу ступінь кореляції із такими важливими компонентами, як Дослідження, Розвиток, Талант і Комерція. Наявність цих сильних кореляцій вказує на те, що ефективність проекту тісно пов'язана із його науковим, технічним, кадровим та комерційним аспектами. Проведений детальний аналіз показав, що загальна оцінка проекту визначається переважно впливом урядової стратегії та операційного середовища. Оцінка відображає важливість правильної стратегії управління та адаптації до змін в операційному середовищі для досягнення успіху проекту. Відзначається, що вплив цих факторів становить основу для формування загального балу проекту, але вони менш вагомі, ніж інші ключові аспекти. Окрім того, в рамках проведеного дослідження використовувався кластерний аналіз, в якому враховувалися відстані найближчих та найдальших сусідів. Цей підхід дозволяє розглядати різні аспекти проекту, адаптуючи підхід до кластеризації відповідно до конкретних особливостей кожної ознаки [34].

Кластеризація країн за кожною окремою ознакою представлена на рис. 2.5-2.8.

В результаті аналізу набору даних, було виявлено, що він включає інформацію про 29 країн Європи, що представляє значний обсяг даних з цього регіону. Додатково, наявна інформація стосовно 14 країн Азіатсько-Тихоокеанського регіону, що свідчить про широкий географічний охоплення аналізу. Регіон Америки представлений в наборі даних 8 країнами, в той час як Близький Схід та Африка мають відповідно 6 та 5 країн, вносячи важливий елемент різноманітності та глобальної репрезентації у дослідження. Цей розподіл дозволяє отримати глибший уявлення про політичні режими в різних частинах світу. Зазначена географічна розмаїтість важлива для адекватного розуміння різних культурних, історичних та соціальних контекстів, що впливають на формування політичних систем.

Такий підхід розширює обсяг аналізу та робить його більш всебічним, що є ключовим для отримання об'єктивних та репрезентативних висновків.

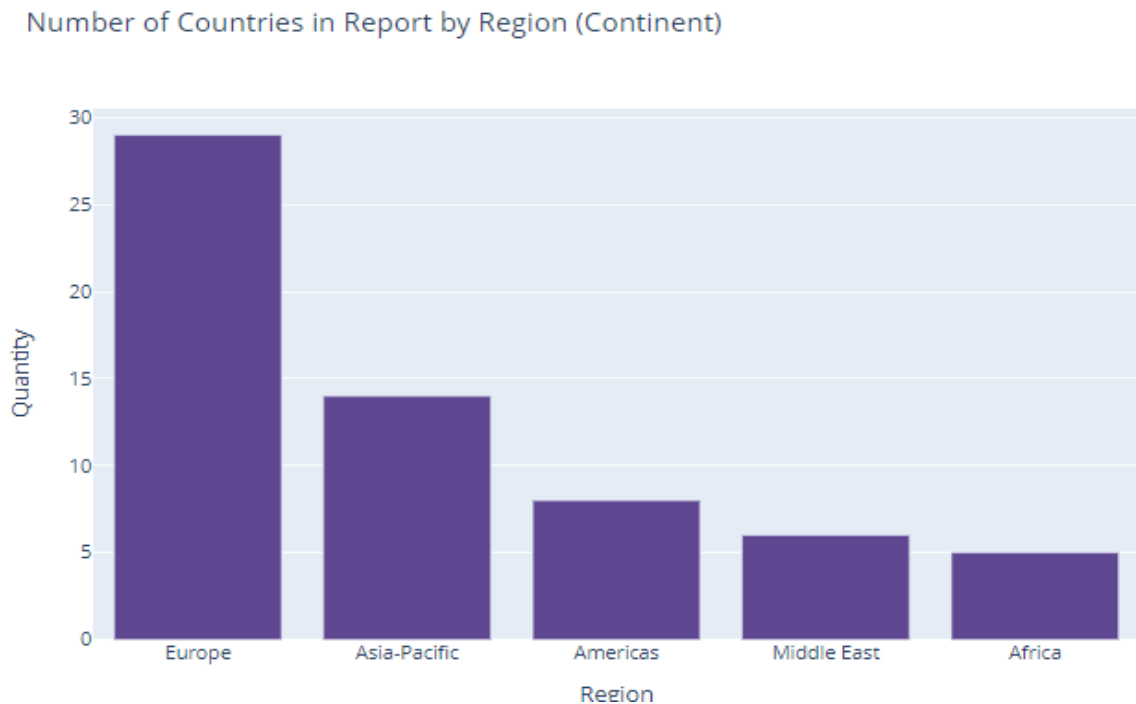


Рисунок 2.5 - Кількість країн у звіті за регіонами

Джерело:[2]

Кластер потужних гравців (рис. 2.6) містить дві провідні країни з точки зору впровадження ІІІ в серйозних масштабах (це США та Китай)

Традиційний кластер найкращих об'єднує 4 країни (Велику Британію, Канаду, Францію та Німеччину), які слідують за потужними гравцями

Кластер Rising Stars об'єднує 11 країн, які перебувають на серйозному підйомі з точки зору розвитку ІІІ/комерціалізації

Кластер Waking up складається з 40 країн, які щойно почали плавний прогрес у ІІІ

Зароджувальний кластер(*Nascent cluster*) – це об'єднання 5 країн, які знаходяться на самому початку шляху до адаптації ІІІ на національному рівні.

Number of Countries in Report by Cluster

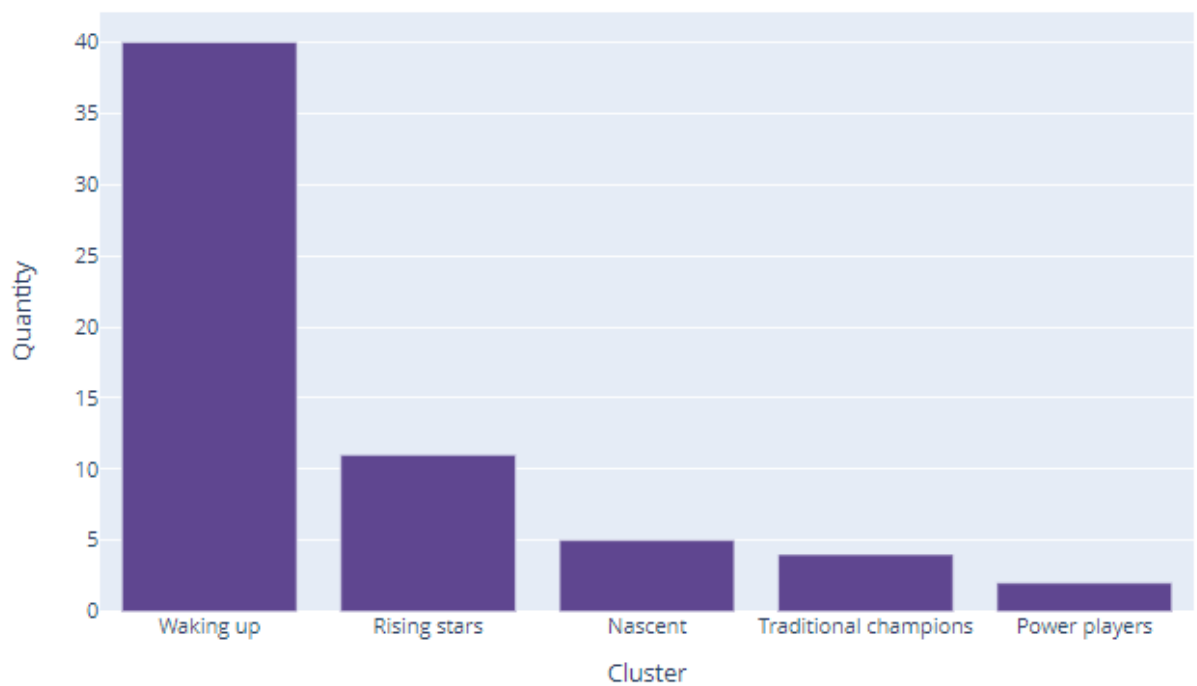


Рисунок 2.6 - Кількість країн у звіті за кластерами

Джерело:[2]

Кількість країн за групами доходів зображено на рис. 2.7

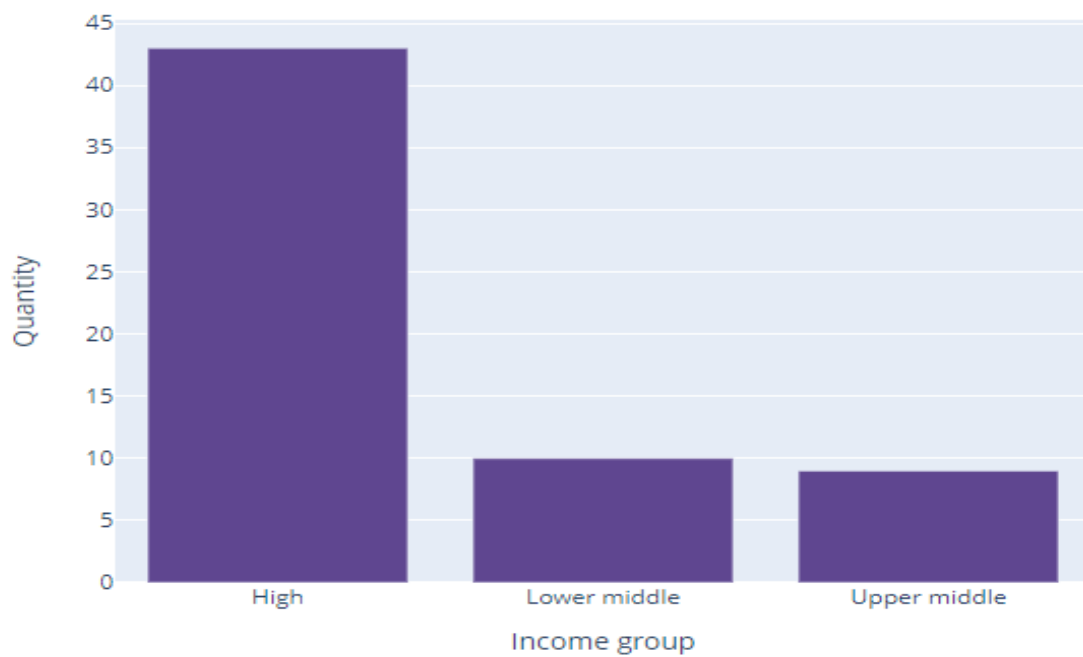


Рисунок 2.7 - Кількість країн за групами доходів

Джерело:[2]

Згідно з рис. 2.3, маємо 43 країни з високим рівнем доходу із 62, представлених у наборі даних), 9 країн з рівнем доходу вище середнього, 8 країн з рівнем доходу нижче середнього

Кількість країн за політичним режимом зображена на рис. 2.8

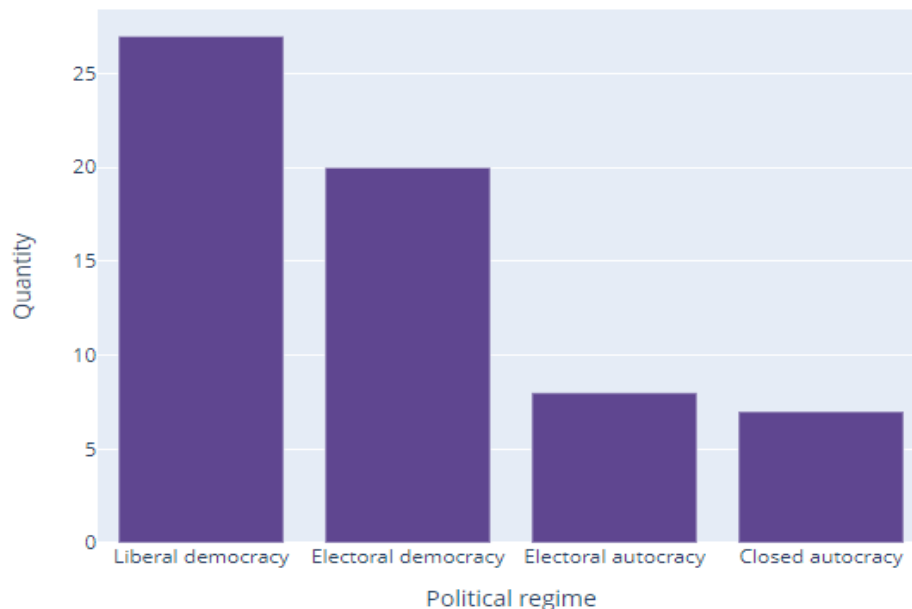


Рисунок 2.8 - Кількість країн за політичним режимом

Джерело:[2]

Стовбчаста діаграма відображає розподіл країн за політичними режимами, де виділено наступні категорії: ліберальна демократія, електоральна демократія, виборча автократія та закрита автократія. За результатами аналізу можна визначити, що 27 країн працюють в умовах ліберальної демократії, що відзначається високим рівнем громадянських свобод і політичних прав. Ще 20 країн входять у категорію електоральної демократії, де проводяться регулярні вибори, але можливо виникають обмеження щодо громадянських свобод. З іншого боку, 8 країн, згідно набору даних, працюють у виборчій автократії, що вказує на владу обраного лідера з обмеженими політичними свободами громадян. Додатково, 7 країн у наборі даних функціонують у режимі закритої автократії, де політична система характеризується відсутністю конкуренції та обмеженим доступом

до політичної участі. Результати цього аналізу вказують на різноманітність політичних режимів у світі та дають можливість глибше розуміти динаміку політичних систем у визначених країнах. Важливо враховувати ці різноманітність та розподіл, оскільки політичний контекст може впливати на розвиток суспільства, громадянські свободи та міжнародні відносини.

Аналіз талантів показано на рисунку 2.9.

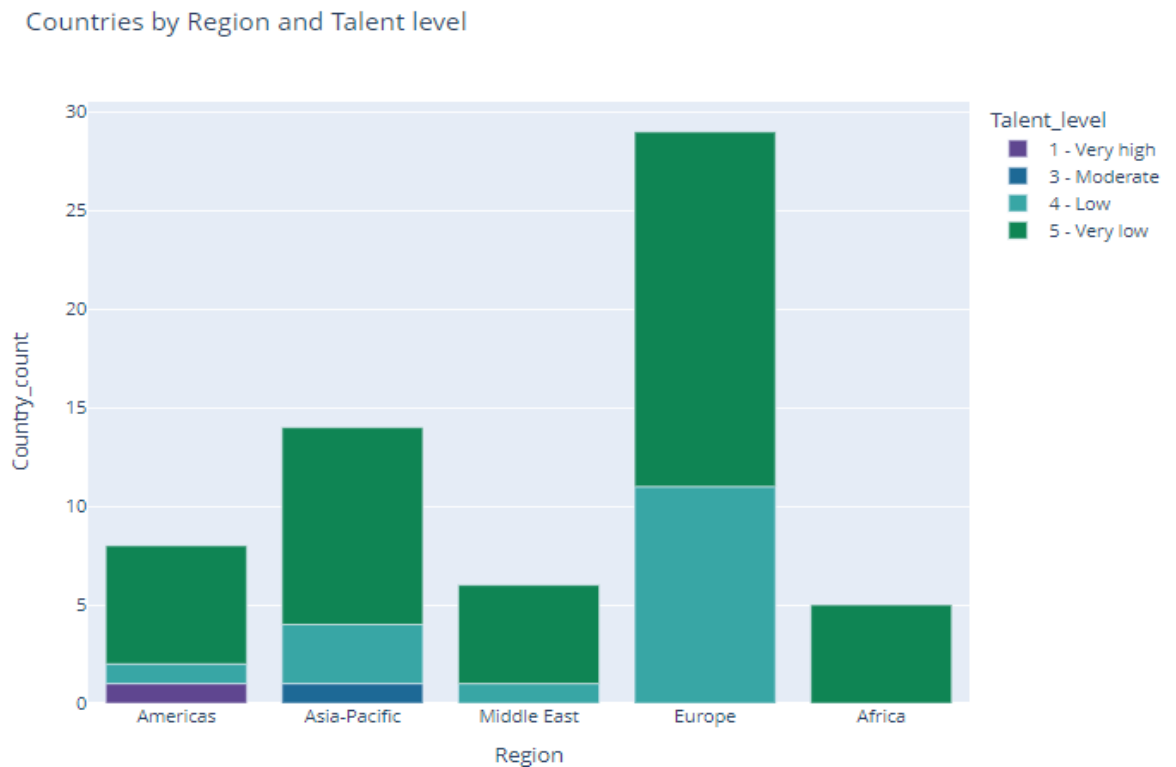


Рисунок 2.9 - Статистика Таланту по регіонах

Джерело:[2]

У висновках з аналізу діаграми, згідно з рисунком 2.9, можна визначити, що Американський континент виділяється як основний центр талановитих фахівців, з високим рівнем експертності. Це свідчить про впливність та великий резервар талантів у даному регіоні. Азіатсько-Тихоокеанський регіон, в свою чергу, відзначається середніми показниками кваліфікації, вказуючи на те, що тут також присутній значущий пул талантів, хоча на рівні, що трошки нижче. Навпаки, Європа та Африка представлені меншими та найменшими показниками талантів відповідно. Це може бути

результатом різниці в рівні освіти, економічного розвитку чи ступеня доступу до технологічних ресурсів у цих регіонах. Дана діаграма вказує на нерівномірний розподіл талановитих фахівців у світі та визначає ключові області, де варто акцентувати увагу для подальшого розвитку та сприяння обміну знаннями та експертами між регіонами. Такий аналіз може служити основою для стратегій глобального управління талантами та розвитку інновацій.

Статистику Таланту по кластерам зображено на рис. 2.11

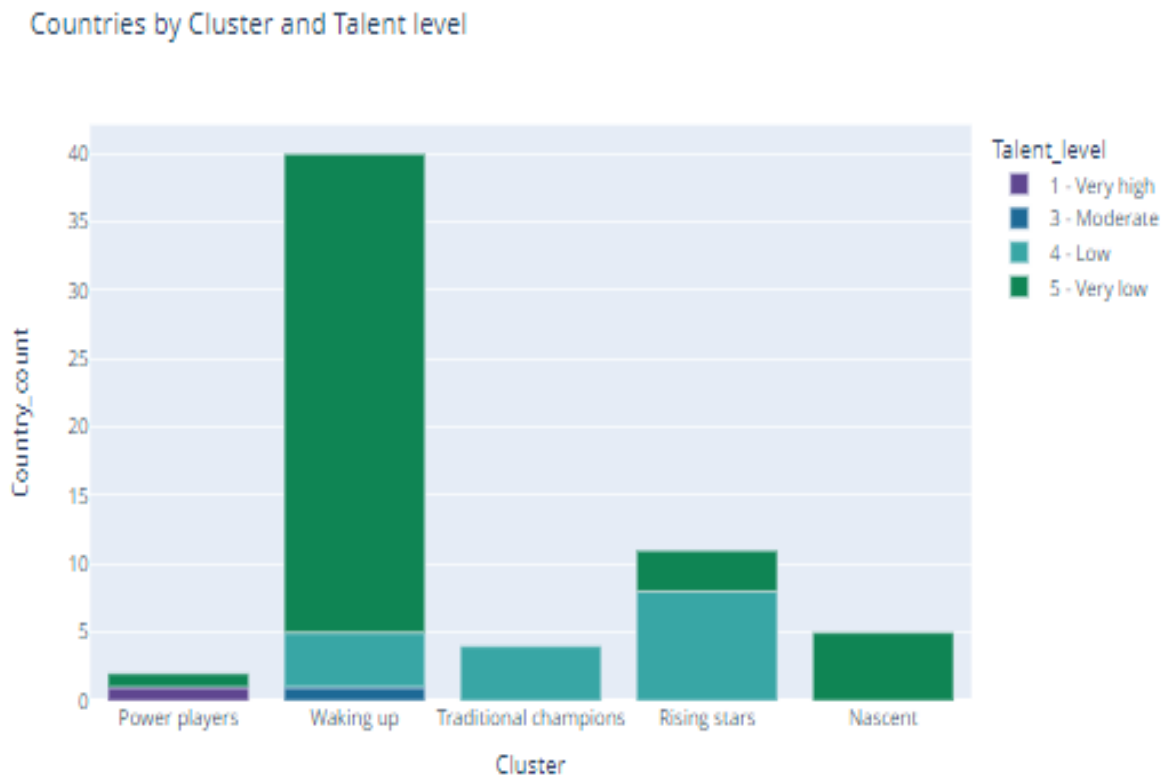


Рисунок 2.10 - Статистика Таланту по кластерам

Джерело:[2]

У рамках аналізу діаграми, що зображена на рисунку 2.10, виділяється наступний аспект: кластер "Power players" привертає увагу як найбільший за кількістю висококваліфікованих фахівців серед представлених країн. Це свідчить про високий рівень експертизи та потенціалу цього кластеру в порівнянні з іншими групами. Зазначений феномен говорить про те, що

країни, які утворюють кластер "Power players", мають значущий внесок у відповідну сферу діяльності, що підкреслює їхню вагу і вплив. Детальне дослідження структури цього кластеру може розкрити ключові аспекти, які призводять до його домінування, такі як високий рівень освіти, наукових досягнень чи інноваційних практик. Такий розподіл спеціалістів свідчить не лише про кількісне переважання, але і про важливі якісні характеристики, що роблять цей кластер ключовим гравцем у глобальному контексті. Діаграма відкриває можливості для подальших досліджень та стратегічного планування, спрямованого на розвиток і вдосконалення спеціалізованих знань та навичок у цьому визначальному кластері.

Статистику Талантів за групами доходу зображено на рис. 2.11

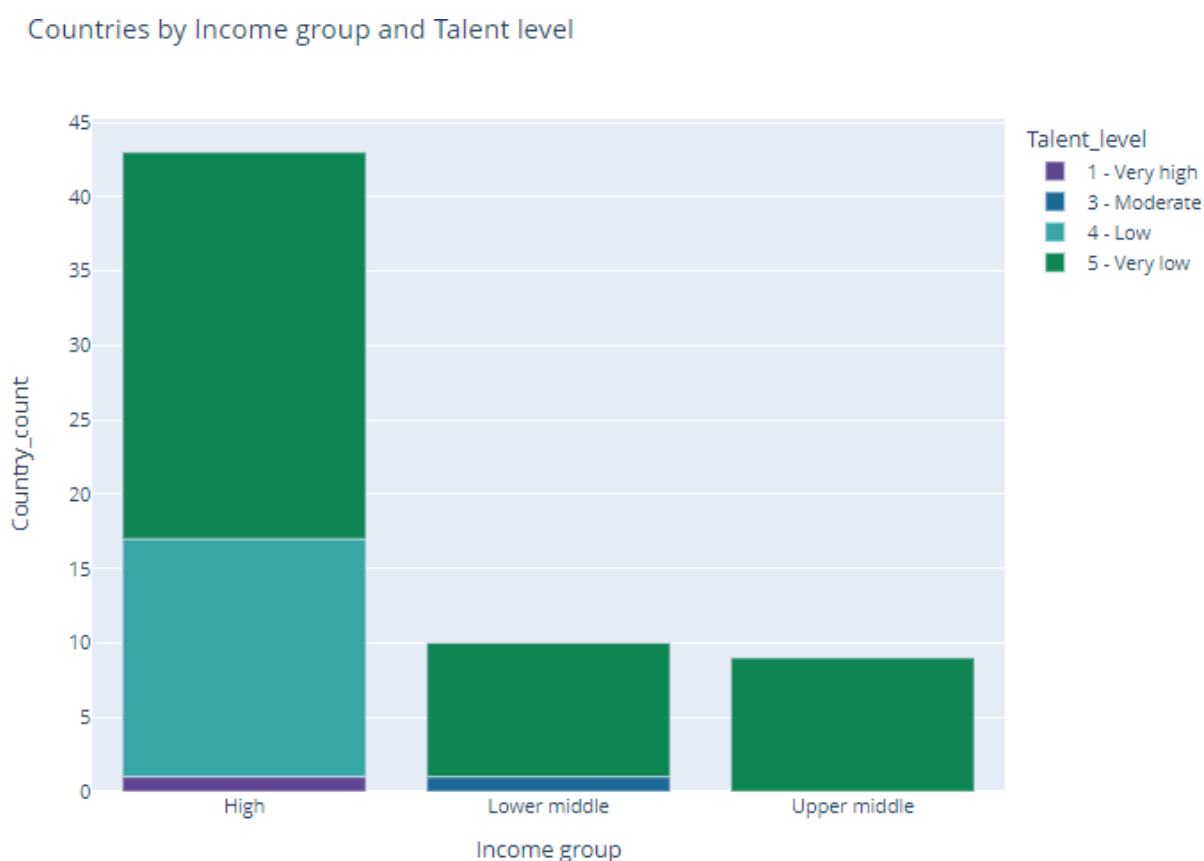


Рисунок 2.11 - Статистика Талантів за групами доходу

Джерело:[2]

Аналізуючи діаграму зі статистикою талантів за групами доходу, що зображена на рисунку 2.11, відзначається, що країни з найвищою групою доходу є лідерами за рівнем талантів. Це свідчить не лише про високий рівень експертності, але і про потужний інноваційний потенціал цих країн. Однак, не менш важливою є констатація, що в цих країнах також існують низькі та дуже низькі рівні талантів, що може вказувати на наявність соціально-економічних викликів та різноманіття в області освіти. У групі країн з "lower middle" доходами виділяється одна країна, яка має помірний рівень талантів. Проте, залишаючись в рамках даної групи, інші країни демонструють дуже низькі рівні талантів. Це може бути пов'язано з різницею в розвитку економік та соціальних секторів серед цих країн. У групі "upper middle" країн виділяється особливий аспект – тут переважають країни із дуже низьким рівнем талантів. Це може свідчити про виклики, що стосуються розвитку інноваційного потенціалу в цих регіонах, або про особливості в системі освіти та доступі до ресурсів. Цей аналіз груп доходу надає контекст для стратегічного планування та розробки політик, спрямованих на підтримку розвитку талановитості та інновацій у різних соціально-економічних контекстах.

Статистику Талантів за політичним режимом можна побачити на рис. 2.12.

Розглядаючи діаграму зі статистикою талантів за політичним режимом, що зображена на рисунку 2.10, можна зробити наступні висновки: ліберальна демократія привертає увагу як категорія, що включає країни з найвищим рівнем талантів. Це може свідчити про те, що такий політичний режим сприяє розвитку та стимулює талановитість в суспільстві. З іншого боку, присутність багатьох країн з низьким рівнем талантів у цій категорії може вказувати на відмінності у рівні розвитку в межах ліберальної демократії. У електоральних автократіях спостерігається помірний рівень таланту, і ця особливість може виникати з обмеженої конкуренції та контролю над політичною системою. Інші політичні режими, такі як електоральна

демократія та закрита автократія, відзначаються в основному наявністю країн зі слабо розвинутим рівнем таланту. Це може свідчити про особливості в управлінні, економічному розвитку чи освітніх політиках цих режимів.

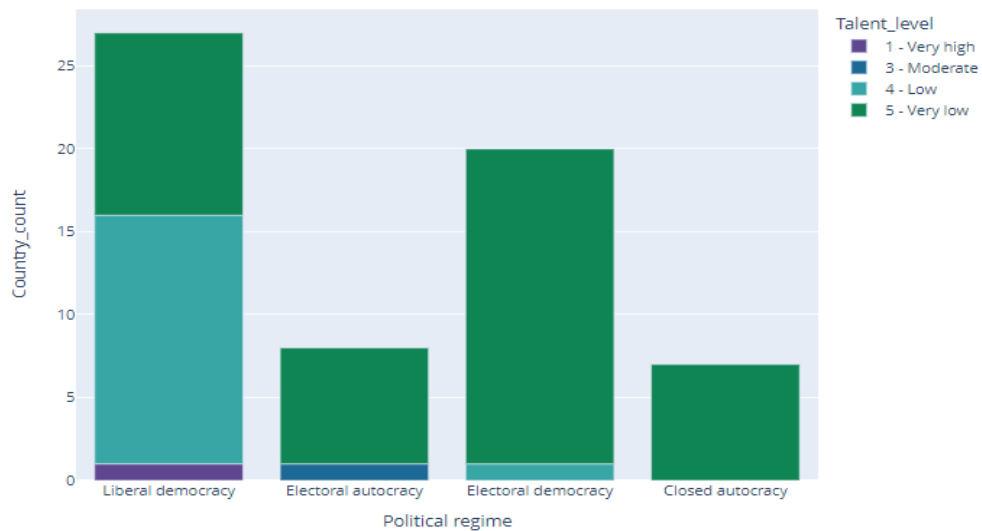


Рисунок 2.12 - Статистика Талантів за політичним режимом

Джерело:[2]

Даний аналіз створює можливість для глибшого розуміння взаємозв'язку між політичним устроєм та рівнем талантів у країнах. Він може послужити основою для розробки стратегій політичного та освітнього розвитку з метою сприяння талановитості та інноваціям у різних частинах світу.

Аналіз даних інфраструктури можна побачити на рис. 2.13.

Аналізуючи статистику інфраструктури за регіонами, що зображена на рис. 2.13, можна виділити ключові тенденції та відмінності в розвитку технічної бази. У американському регіоні спостерігається різноманіття рівнів інфраструктури: 1 країна відзначається дуже високим рівнем, 3 – високим, та 4 – помірним. Це може свідчити про високий рівень технічної готовності країн цього регіону, але також про різницю у розвитку між ними.

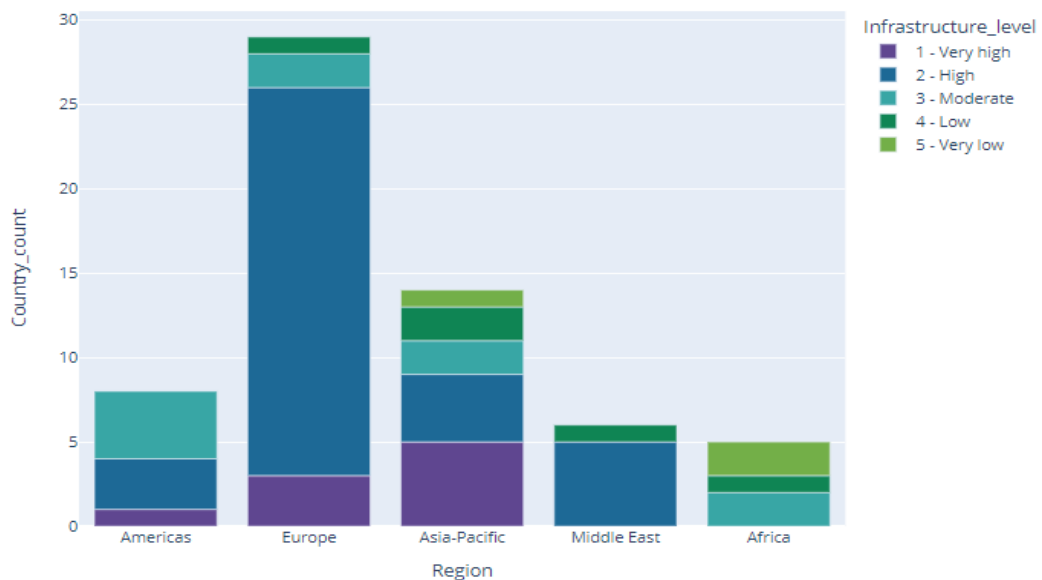


Рисунок 2.13 - Статистика інфраструктури по регіонах

Джерело:[2]

У європейському регіоні зафіксовано значний рівень інфраструктури: 3 країни – з дуже високим, 20 – з високим та 3 – з помірним рівнем. Це свідчить про технічну стабільність та розвиток різноманітних технічних секторів у більшості країн європейського континенту. В азійському регіоні присутній великий розмах рівнів інфраструктури: 5 країн – з дуже високим, 4 – з високим та 5 – з помірним та низьким рівнем.

Це може бути результатом швидкого технологічного розвитку у деяких країнах, в той час як інші можуть стикатися з викликами у сфері інфраструктурних інвестицій. На Середньому Сході виявлено дуже різний рівень інфраструктури: 5 країн – з високим та 1 – з низьким рівнем. Це може бути пов'язано з великими відмінностями в розвитку технічних систем у різних країнах регіону. У африканському регіоні також виявляється різноманітність рівнів інфраструктури: 2 країни – з помірним, 1 – з низьким та 2 – з дуже низьким рівнем. Це свідчить про необхідність удосконалення технічних засобів у деяких частинах Африки.

Статистика інфраструктури по кластерам зображено на рисунку 2.14

Даний аналіз статистики інфраструктури в регіонах, що зображена на рисунку 2.14, може бути корисним для розробки стратегій інфраструктурного розвитку та сприяння технологічному прогресу у різних частинах світу.

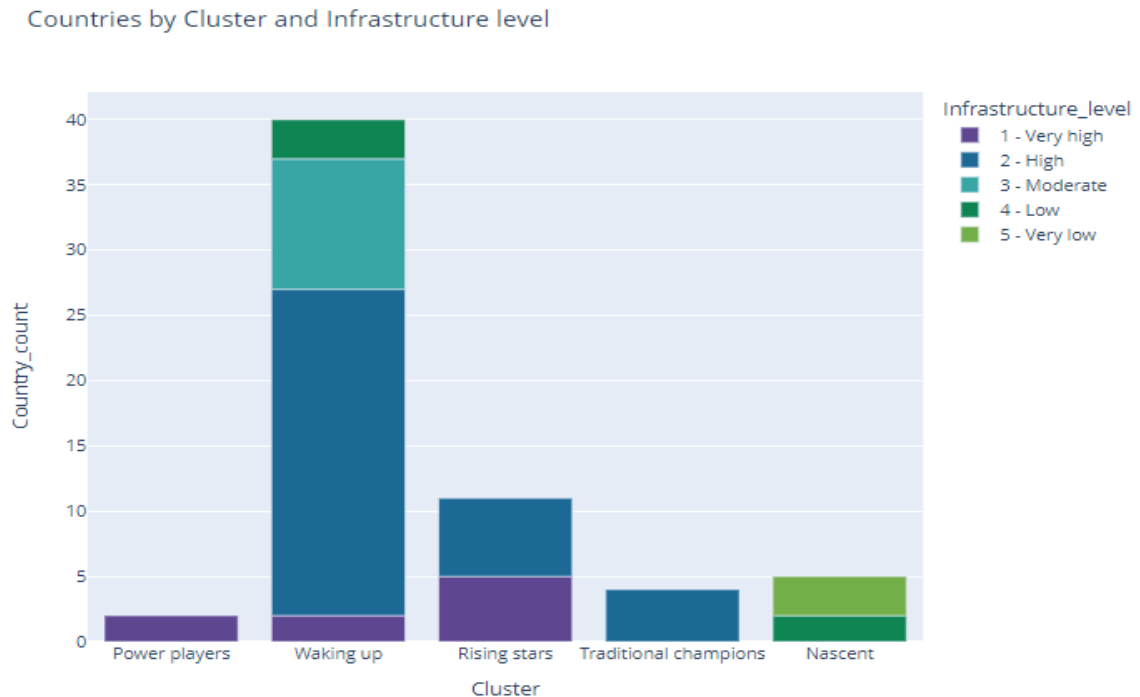


Рисунок 2.14 - Статистика інфраструктури по кластерам

Джерело:[2]

Аналізуючи статистику інфраструктури за класифікацією кластерів, можна зазначити ключові особливості та різноманіття в рівнях технічної готовності країн у різних групах. У кластері "Сильні гравці" виділяються 2 країни з дуже високим рівнем інфраструктури. Це може свідчити про їхню передовість та високий рівень розвиненості технічних систем. Група "Ті, хто тільки починає впроваджувати ШІ" представлена 2 країнами з дуже високим рівнем, 20 – з високим, 20 – з помірним та 3 – з низьким рівнем інфраструктури. Це вказує на широкий розмах розвитку в групі, але й на наявність країн із значущими викликами у розвитку технічних систем. Ті, хто піднімаються, характеризуються 5 країнами з дуже високим та 6 з високим рівнем інфраструктури. Це свідчить про активний розвиток технічних

галузей у цих країнах. У кластері "Традиційні чемпіони" можна відзначити 4 країни з високим рівнем інфраструктури, що свідчить про стабільність та високий рівень технічної готовності у цих країнах. Група "Ті, що зароджуються" представлена 2 країнами з низьким та 3 – з дуже низьким рівнем інфраструктури. Це може вказувати на необхідність значних інвестицій та розвитку технічних секторів у цих регіонах. Даний аналіз надає контекст для формування стратегій розвитку технічної інфраструктури в різних класифікаційних кластерах, а також може бути корисним для визначення пріоритетних напрямків інновацій та розвитку в різних групах країн.

Аналізуючи статистику інфраструктури за політичними режимами, можна визначити основні тенденції та різноманіття в розвитку технічної бази в країнах з різними формами уряду.

Інфраструктуру за групами доходу зображено на рисунку 2.15

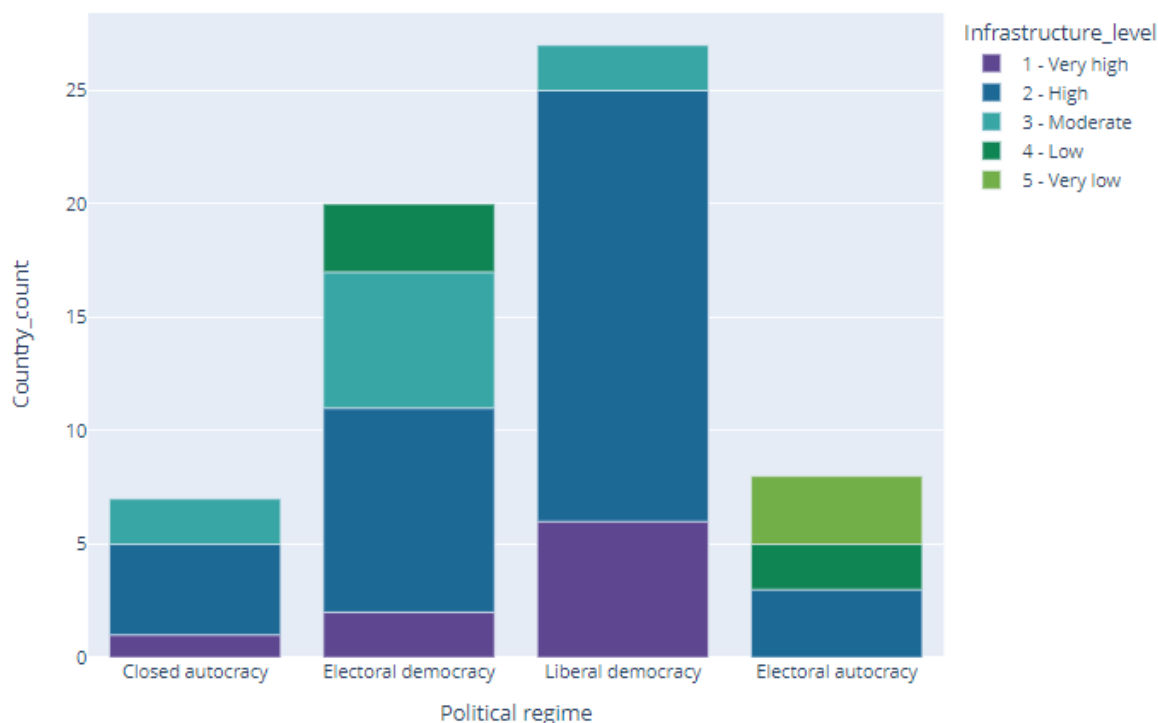


Рисунок 2.15 - Інфраструктура за групами доходу

Джерело:[2]

Згідно з рисунком 2.15, у країнах з закритою автократією виявляється значна різноманітність рівнів інфраструктури: 1 країна – з дуже високим, 4 – з високим та 2 – з помірним рівнем. Це може бути наслідком спрямованості уряду на впровадження технічних ініціатив або обмежень, що впливають на інфраструктурний розвиток. В електоральних демократіях спостерігається різноманіття рівнів інфраструктури: 2 країни – з дуже високим, 10 – з високим, 5 – з помірним та 3 – з низьким рівнем. Це може вказувати на вплив політичного процесу та конкуренції на інфраструктурний розвиток у цих країнах. Ліберальні демократії характеризуються значною кількістю країн з високим та дуже високим рівнем інфраструктури: 6 – з дуже високим, 19 – з високим і 2 – з помірним. Це може свідчити про важливість гармонізації політичного та технічного розвитку у цих країнах. Електоральні автократії відзначаються 3 країнами з високим рівнем інфраструктури, 2 – з низьким і 3 – з дуже низьким. Це може бути пов'язано з особливостями політичного управління та спрямованістю урядів цих країн на розвиток технічної бази. Даний аналіз дозволяє зрозуміти вплив політичного контексту на розвиток інфраструктури та може бути використаний для формування стратегій інновацій та технічного розвитку у різних типах політичних систем.

Статистику інфраструктури у геопросторовому вигляді зображено на рис. 2.16.

З геопросторової точки зору, як це зображено на рисунку 2.16, можемо бачити, що є дуже багато осередків розвитку у країнах Європи, і серед них є приклади з дуже високим рівнем розвитку інфраструктури. Також високі рівні розвитку можемо бачити в США, Китаї, Японії та Канаді. Натомість країни Африки найбільше відстають за цим показником, а наші недоброзичливі сусіди начеб-то розвиваються, але насправді дуже слабкі у даній сфері.

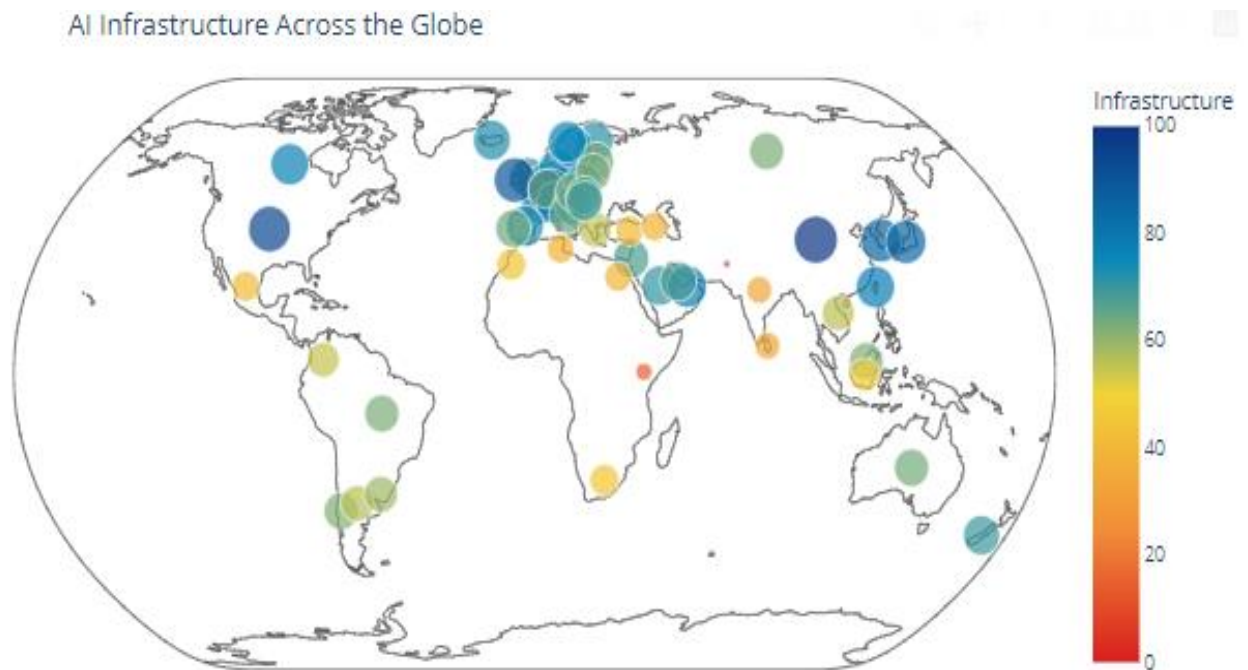


Рисунок 2.16 - Статистика інфраструктури геопросторовий вигляд

Джерело:[2]

Статистику операційного середовища за регіонами зображено на рис. 2.17

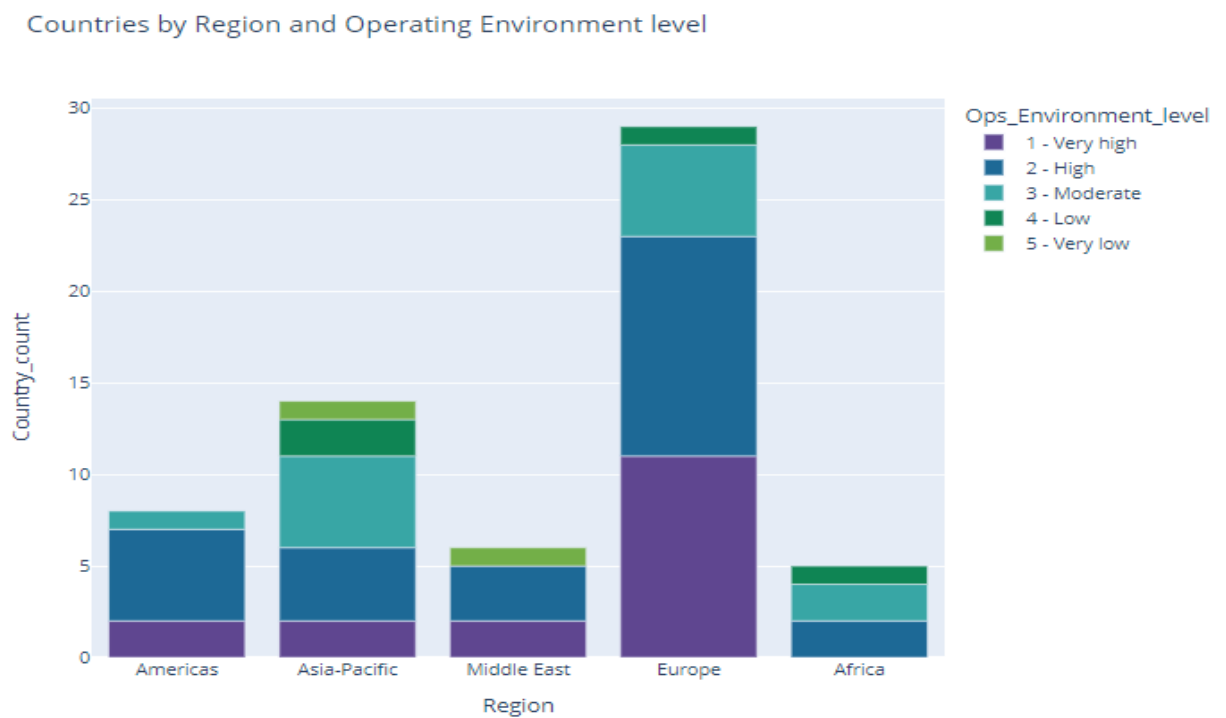


Рисунок 2.17 - Статистика операційного середовища за регіонами

Джерело:[2]

Аналізуючи статистику операційного середовища за регіонами, що зображена на рисунку 2.17, можна визначити основні тенденції та відмінності в рівнях функціонування технічних систем у різних частинах світу. У Європі спостерігається високий рівень операційного середовища, в основному представлений країнами з дуже високим та високим рівнем. Це свідчить про ефективне функціонування технічних систем та інфраструктури в більшості країн європейського континенту. У регіонах Америки, Азії та Середнього Сходу виявляється відносно однаковий рівень країн з дуже високим та високим рівнем операційного середовища. Це свідчить про технічну стабільність і схожість у розвитку технічних систем у цих регіонах. У азійському регіоні відзначається ще багато країн зі слабким і помірним рівнем операційного середовища. Це може бути результатом різноманітності у технічних стратегіях та інвестицій у цих країнах. В африканському регіоні зафіксовано 2 країни з високим рівнем, а також пару країн із зазначеними низьким і помірним рівнем операційного середовища. Це може вказувати на варіативність розвитку технічних систем у різних частинах Африки.

Даний аналіз статистики операційного середовища за регіонами дозволяє здійснити порівняльний огляд рівнів функціонування технічних систем та інфраструктури в різних частинах світу, що може бути корисним для розробки стратегій технічного розвитку та інновацій.

Статистику операційного середовища за кластерами зображено на рис. 2.18.

Аналізуючи дану стовбчасту діаграму, що зображена на рисунку 2.18, можна зробити висновок, що відсотково всі члени кластеру мають однакову кількість членів сформованих по статистиці операційного середовища, окрім тих хто починає свій шлях в сфері штучного інтелекту

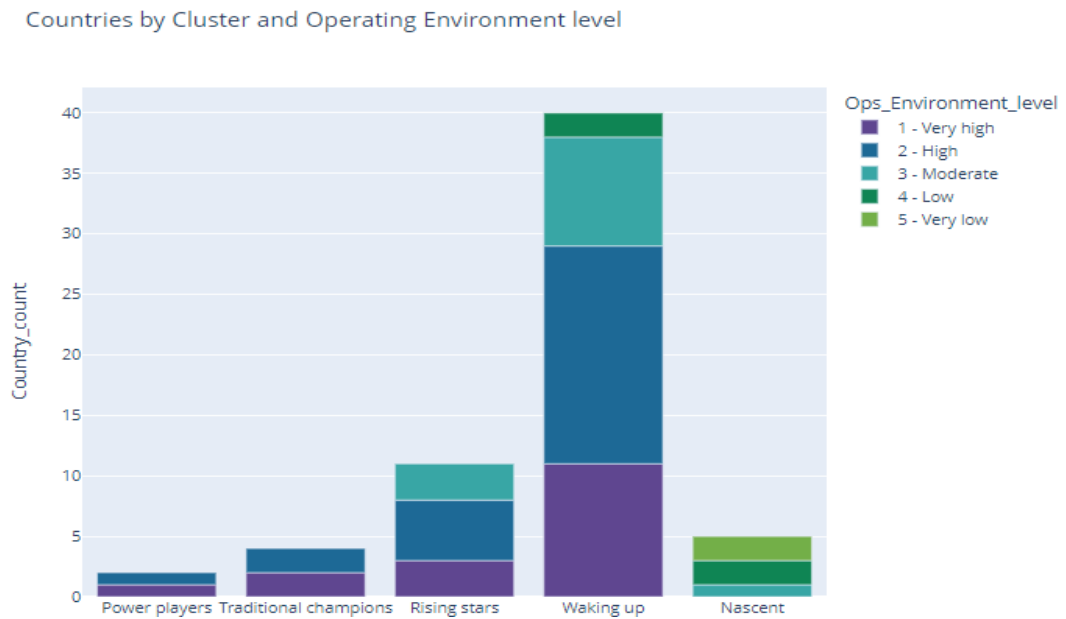


Рисунок 2.18 - Статистика операційного середовища за кластерами

Джерело:[2]

Статистика операційного середовища за групами доходів зображена на рис. 2.19.

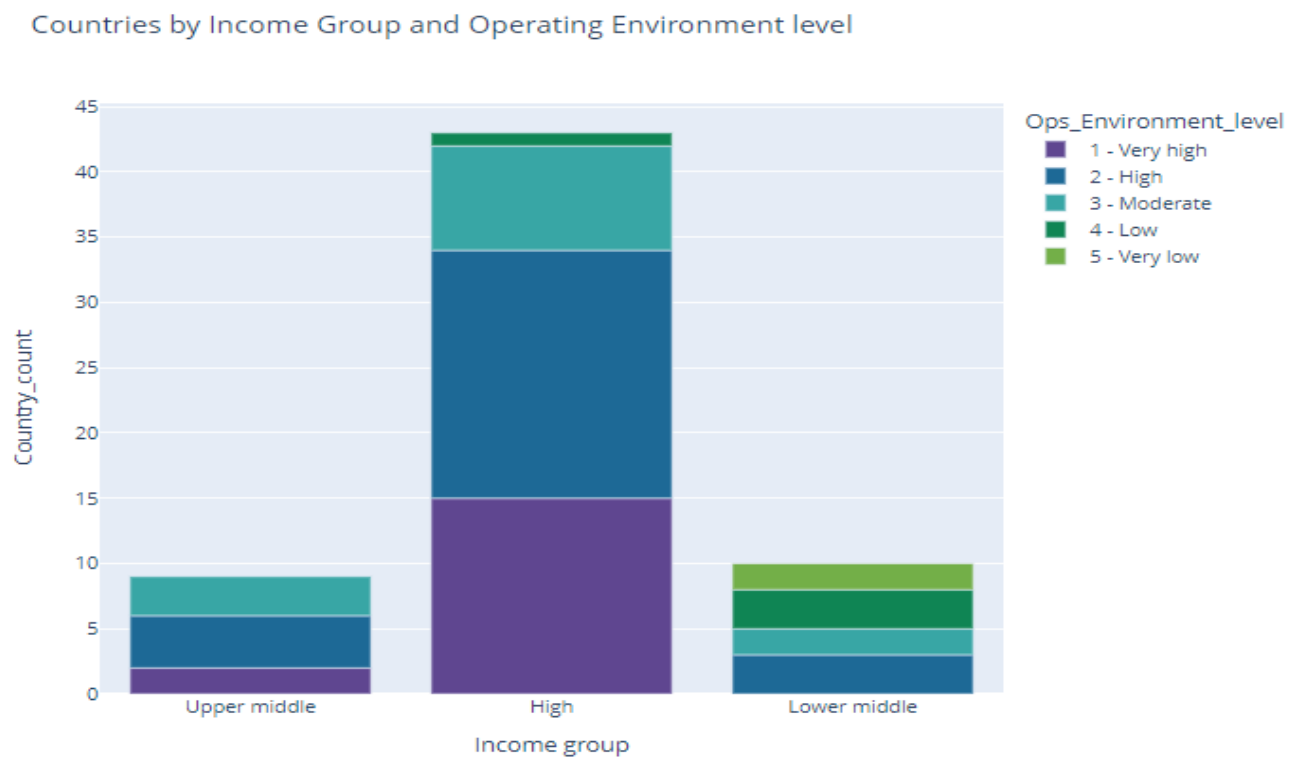


Рисунок 2.19 - Статистика операційного середовища за групами доходів

Джерело:[2]

Розглядаючи дані стосовно операційного середовища в контексті економічного рівня країн, що зображена на рисунку 2.19, можна виділити ключові тенденції та особливості, що вказують на взаємозв'язок між розвитком технічних систем і рівнем доходів. У групі країн із високим рівнем доходів, спостерігається висока ефективність операційного середовища. Велика кількість країн із дуже високим рівнем свідчить про передовий технічний статус цих економік. Зокрема, є певна різноманітність у рівні операційного середовища, що може бути наслідком різних стратегій технічного розвитку. Група країн із рівнем доходів вище середнього вирізняється двома країнами із дуже високим рівнем операційного середовища, вказуючи на високий рівень технічного розвитку. Ця група також містить країни із високим і помірним рівнями, вказуючи на певну різноманітність у розвитку технічних систем. Серед країн із нижчим рівнем доходів виділяється певна динаміка операційного середовища. Країни із високим рівнем вказують на здатність деяких економік ефективно використовувати технічні ресурси при обмеженому фінансуванні. Однак наявність країн із помірним і низьким рівнями може вказувати на технічні виклики та нерівномірності у розвитку. Цей аналіз дозволяє краще зрозуміти вплив економічного фактора на технічний прогрес, визначаючи спільні риси та різноманітність у розвитку операційного середовища в різних країнах.

Статистика операційного середовища за політичним режимом зображено на рис. 2.20.

При аналізі статистики операційного середовища, що відображена на рисунку 2.20, варто відзначити, що закриті автократичні системи включають дві країни із дуже високим рівнем операційного середовища, п'ять країн із високим та одну із помірним. Це може свідчити про різницю у стратегіях управління та ефективності технічних систем у таких автократичних режимах.

Countries by Political Regime and Operating Environment level

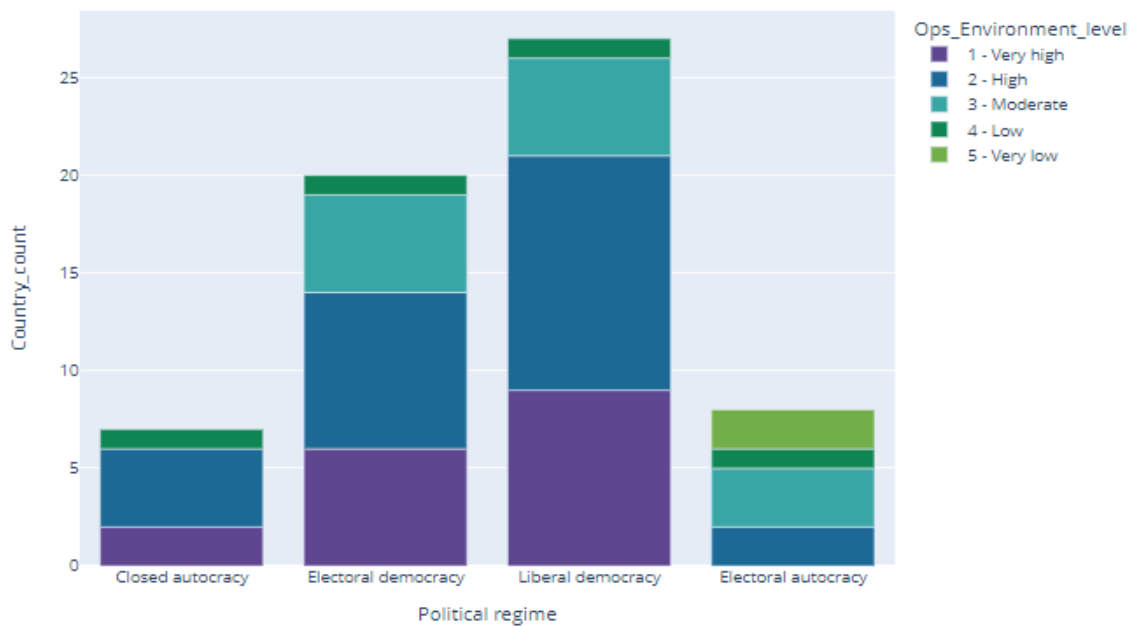


Рисунок 2.20 - Статистика операційного середовища за політичним режимом

Джерело:[2]

Серед систем електоральної демократії переважає високий рівень операційного середовища: шість країн із дуже високим, десять із високим, п'ять із помірним та одна із низьким. Це може свідчити про те, що в демократичних системах, де існує конкуренція та відкритість, сприяється технічному прогресу та оптимальному функціонуванню інфраструктури. Ліберальні демократії відрізняються високим рівнем операційного середовища: дев'ять країн із дуже високим, десять із високим та п'ять із помірним. Це може свідчити про те, що в системах, де враховуються як права та свободи громадян, так і технічні інновації, досягається високий стандарт операційного середовища. Електоральні автократії мають три країни із високим рівнем операційного середовища, три із помірним та три із низьким. Це може вказувати на те, що в деяких автократичних системах, де може бути обмежений політичний плюралізм, технічна ефективність може варіюватися. Розглядаючи ці дані, можна зробити висновок, що взаємодія політичних

режимів та технічного розвитку є складною та має свої власні особливості в кожному конкретному випадку.

Статистику операційного середовища з геопросторовим виглядом зображено на рисунку 2.21

З геопросторової точки зору, представлені на рисунку 2.21, виявляється, що центр операційного середовища визначається головним чином в Європі, включаючи сильно розвинені області в США та Китаї, що стало вже традиційним.

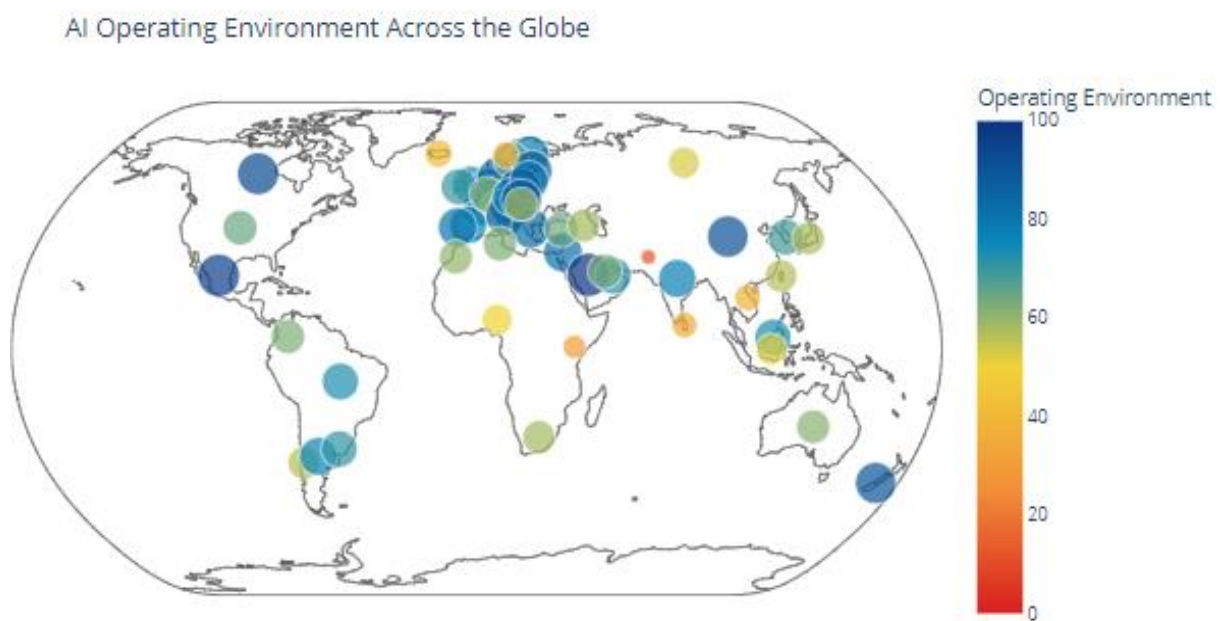


Рисунок 2.21 - Статистика операційного середовища геопросторовий вигляд

Джерело:[2]

Однак значний інтерес та бажання до розвитку в цій сфері проявляються в інших країнах, що також видно на графіку. Існуюча тенденція до підняття країн та їхня загальна нахиленасть до розвитку свідчать про динаміку розподілу операційних середовищ на глобальному рівні. При детальному аналізі можна зауважити, що закриті авторитарні системи виявляються здатними до високого рівня операційного середовища в двох країнах, високого рівня в п'яти країнах і помірного рівня в одній. Це може вказувати на особливості управлінських стратегій та технічної ефективності в цих

автократичних режимах. З іншого боку, системи електоральної демократії характеризуються вищим рівнем операційного середовища, включаючи дуже високий рівень в шести країнах, високий в десяти країнах, помірний в п'яти та низький в одній. Це свідчить, можливо, про те, що демократичні системи, з їхньою конкурентоспроможністю та відкритістю, сприяють технічному прогресу та оптимальному функціонуванню інфраструктури. Ліберальні демократії відрізняються високим рівнем операційного середовища, включаючи дев'ять країн з дуже високим рівнем, десять з високим та п'ять з помірним. Це може свідчити про те, що в системах, де акцентується на правах та свободах громадян, а також технічних інноваціях, досягається високий стандарт операційного середовища. Електоральні автократії мають три країни з високим рівнем операційного середовища, три з помірним та три з низьким, що може вказувати на те, що в деяких автократичних системах, де може бути обмежений політичний плюралізм, технічна ефективність може варіюватися.

Цілісний аналіз впровадження ШІ зображено на рис. 2.22.

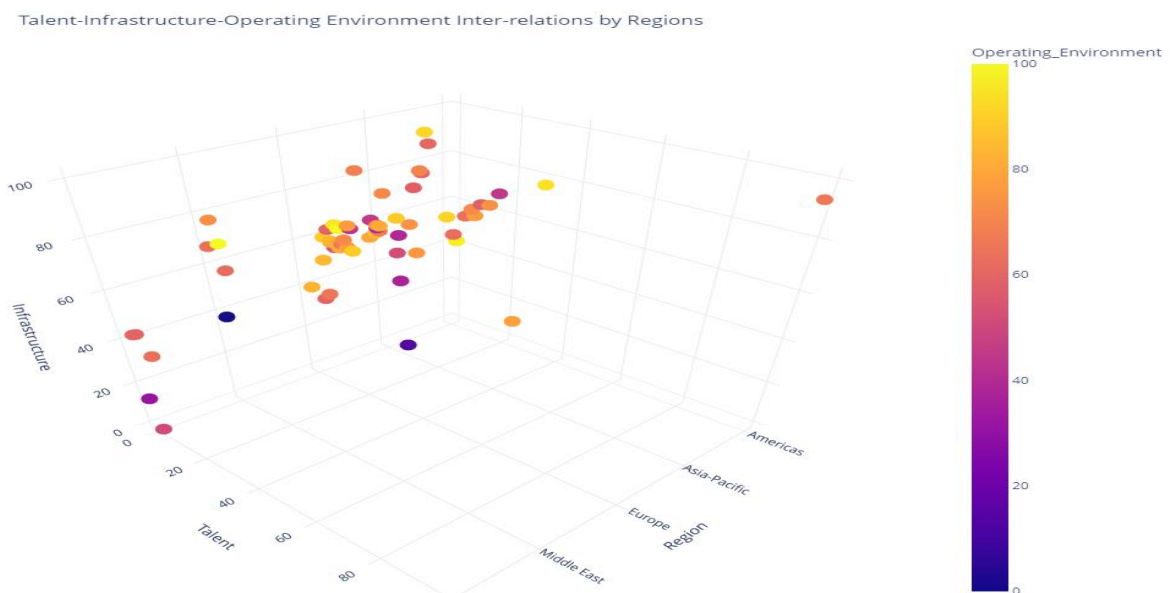


Рисунок 2.22 - Впровадження штучного інтелекту за регіонами та кількістю спеціалістів

Джерело:[2]

На глибокій тривимірній візуалізації, що зображена на рисунку 2.22, маємо можливість більш детально розглядати індекс Штучного Інтелекту за географічними кластерами, що дозволяє отримати вичерпний образ розподілу талантів у цьому стратегічно важливому секторі. Звертає на себе увагу той факт, що Європа представлена значущим кластером, що свідчить про високий рівень розвитку та інноваційної активності в цьому регіоні. Необхідно відзначити, що, не дивлячись на суттєвий вплив Європи, Сполучені Штати Америки утримують лідерські позиції за кількістю кваліфікованих спеціалістів в галузі Штучного Інтелекту. Це може свідчити про глибокий внутрішній потенціал та інноваційну екосистему, що дозволяє США залишатися ключовим гравцем у цій сфері. Такий висновок підкреслює важливість вивчення географічного розподілу талантів в контексті індексу Штучного Інтелекту, щоб зрозуміти не лише поточний стан справ, але і визначити тенденції та можливості для подальшого розвитку цієї сфери в майбутньому.

Цілісний аналіз інвестицій ШІ зображено на рисунку 2.23.

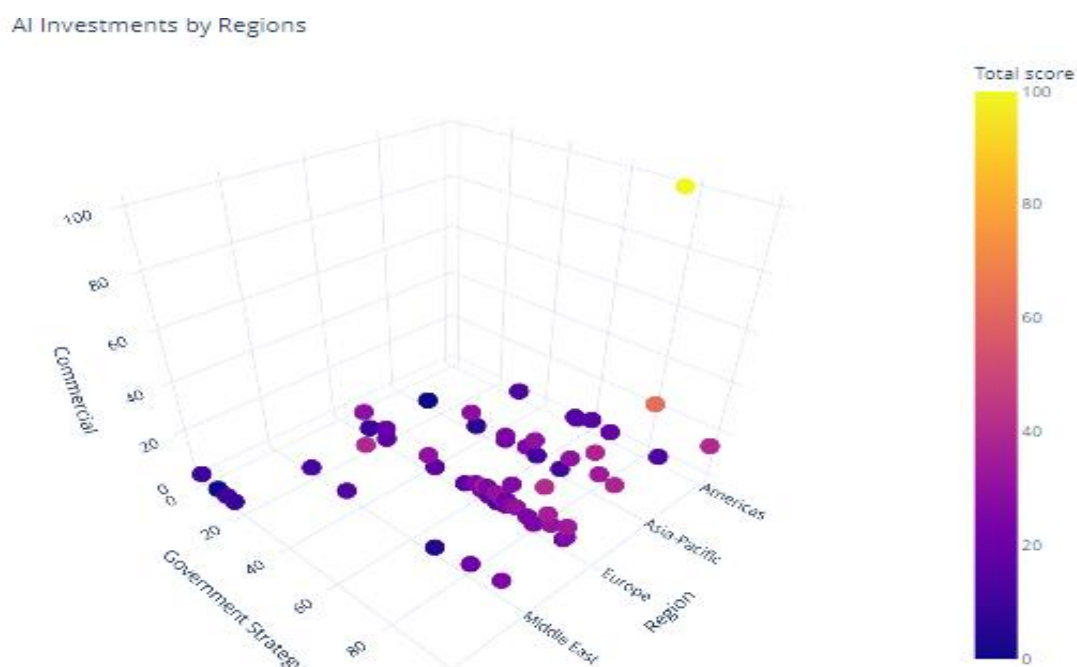


Рисунок 2.23 - Аналіз інвестицій за регіонами

Джерело:[2]

Цей аналіз розподілу інвестицій у сферу Штучного Інтелекту, що зображена на рисунку 2.23, дозволяє спростувати стереотипи і виявити різноманітність стратегій країн Європи. Многовидовий характер інвестицій свідчить про те, що розвиток ШІ в Європі є складним та високодиференційованим процесом, де кожна країна приділяє увагу конкретним аспектам та напрямкам цієї інноваційної галузі. Зокрема, можна відзначити, що окремі країни Європи акцентують увагу на дослідженнях та розвитку новітніх технологій в ШІ, в той час як інші фокусуються на впровадженні виробничих та комерційних аспектів цих технологій.

Такий підхід дозволяє кожній країні максимізувати свій внесок у галузь Штучного Інтелекту з урахуванням власних сильних сторін та особливостей. З іншого боку, вплив Сполучених Штатів Америки виявляється як домінуючий фактор у глобальному розвитку ШІ. Великий обсяг інвестицій та відмінна науково-дослідницька база в США створюють умови для лідерства у впровадженні інновацій та визначенні технологічних трендів в цій галузі. Враховуючи цей аналіз, можна зробити висновок, що розвиток Штучного Інтелекту є глобальним явищем, яке формується відповідно до конкретних стратегій та пріоритетів кожної країни, але при цьому не можна недооцінювати ключову роль Сполучених Штатів у цій екосистемі.

2.3 Результати аналізу глобального індексу ШІ

Регресійний аналіз - це статистичний метод, призначений для виявлення та кількісної оцінки взаємозв'язку між залежною змінною та однією або кількома незалежними змінними. У найпростішому випадку, коли є лише одна незалежна змінна, це відомо як простий регресійний аналіз. А якщо наявність двох чи більше незалежних змінних (понад одну) - це вже

багатофакторний регресійний аналіз. Регресійний аналіз допомагає вивчати, як зміни в значеннях цих незалежних змінних пов'язані зі змінами в залежній змінній, що сприяє розумінню взаємозв'язків у наборі даних.

Процес регресійного аналізу можна поділити на наступні етапи:

1. Каталогізація вхідних даних: Формування систематичного каталогу вхідних даних, які будуть використовуватися для подальших аналізів.
2. Визначення моделі: Обирання типу регресії - простої (з однією незалежною змінною) або множинної (з багатьма незалежними змінними). Побудова математичної моделі, яка відображає взаємозв'язок між залежною та незалежними змінними.
3. Підготовка даних: Проведення аналізу для виявлення аномалій, обробка пропущених значень та стандартизація даних для полегшення аналізу.
4. Оцінка коефіцієнтів: Визначення коефіцієнтів моделі за допомогою методу найменших квадратів або інших статистичних технік.
5. Тестування гіпотез: Визначення статистичної значущості моделі та її компонентів для підтвердження або відхилення гіпотез.
6. Прогнозування: Використання розробленої моделі для передбачення значень залежної змінної на основі вхідних даних.

Використання регресійного аналізу:

1. Прогнозування: Регресійний аналіз є ефективним інструментом для прогнозування майбутніх значень залежної змінної. Це дозволяє аналізувати та передбачати, як зміни в незалежних змінних можуть впливати на залежну змінну в майбутньому.
2. Експлікація Зв'язків: Регресійний аналіз допомагає розкривати та детально розібрати природу та силу взаємозв'язків між різними факторами. Він вказує, які незалежні змінні мають значущий вплив на зміну залежної змінної.

3. Управління Ризиками: Регресійний аналіз дозволяє систематично аналізувати вплив різних факторів на залежну змінну. Це допомагає визначати та керувати ризиками, а також розуміти, як різні змінні можуть взаємодіяти для формування результату.
4. Оцінка Впливу: Регресійний аналіз визначає та оцінює вплив різних змінних на результат. Це дозволяє важливим зацікавленим сторонам отримати глибоке розуміння та кількісну оцінку внеску кожного фактора в результат.

У контексті наших даних Глобального індексу штучного інтелекту ви використовуєте регресійний аналіз, щоб зрозуміти, як ключові показники (талант, інфраструктура, робоче середовище тощо) впливають один на одного та на Глобальний індекс штучного інтелекту або прогнозують його [16].

Звичайно що наші усі індекси 'Talent', 'Infrastructure', 'Operating Environment' та інші. Пропорційно відповідають глобальному індексу ШІ.Щоб упевнитися у цьому ми зробимо регресію індексів із глобальним індексом (рис.2.24):

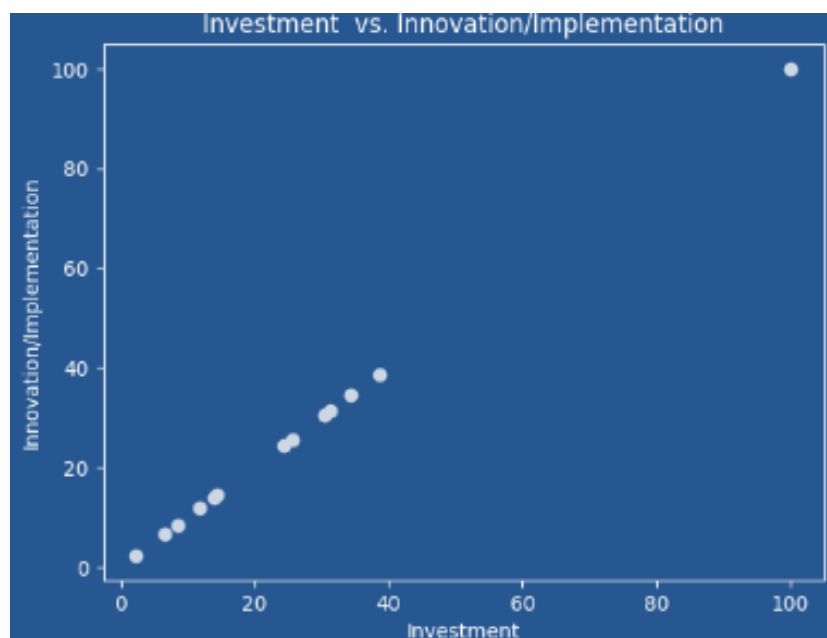


Рисунок 2.24 - Регресійний аналіз Глобального індексу ШІ

Джерело:[2]

Графік з регресійними лініями послужив основою для аналізу взаємодії між незалежними та залежними змінними. Однак при ретельному розгляді можна спостерігати, що значення глобального індексу залишаються невідзначеними, не піддаючись значним коливанням або тенденціям змін. Це важливе спостереження свідчить про стійкість та незмінність глобального індексу в умовах впливу визначених факторів. Наочно відображаючи стабільність цього показника, регресійний аналіз дозволяє визначити, що взаємозв'язок між різними змінними та глобальним індексом є консистентним та сталим у досліджуваному періоді. Цей заключний висновок підкреслює важливість та достовірність знайдених взаємозв'язків, роблячи акцент на тому, що такий сталий характер впливу може бути виявлений не лише для глобального індексу, але й для інших важливих індексів, що визначаються у розглядуваному контексті. Це важливе спостереження підтверджує надійність та сталість зв'язків між визначеними факторами і ключовими показниками, що характеризують величину індексів у вивченій області.

Дослідимо залежність іновацій і впроваджень від фінансових та процедурних зобов'язань тобто інвестицій.

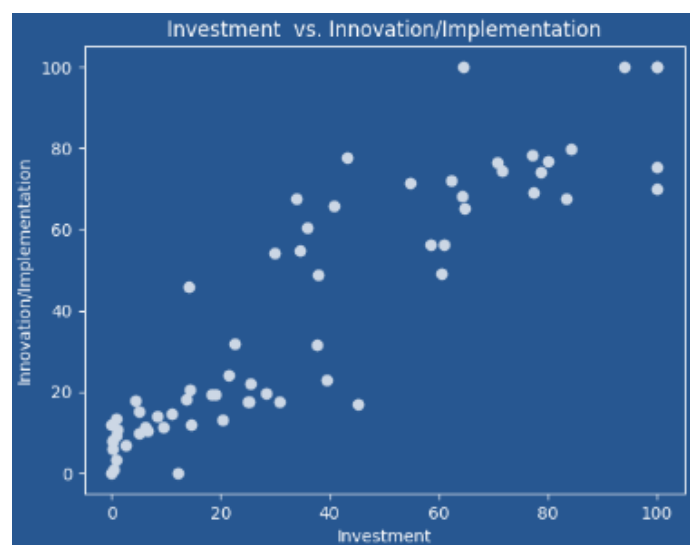


Рисунок 2.25 - Регресійний аналіз інвестицій до множини інновації і впроваджень

Джерело:[2]

У даному регресійному аналізі досліджено взаємозв'язок між рівнем інновацій та впроваджень і фінансовими, а також процедурними зобов'язаннями, що виражаються у величині інвестицій. Важливо відзначити, що отриманий результат виявився додатнім приростом у регресії, що вказує на позитивний вплив фінансових і процедурних зобов'язань на рівень інновацій та їхнє впровадження. Це свідчення про те, що зі збільшенням обсягу інвестицій спостерігається активний розвиток іноваційних і впроваджуваних процесів. Такий позитивний приріст регресії дозволяє зробити висновок про те, що фінансові та процедурні заходи грають ключову роль у стимулюванні та підтримці іноваційних і впроваджуваних ініціатив, що впливає на зростання їхнього рівня у вивченій області.

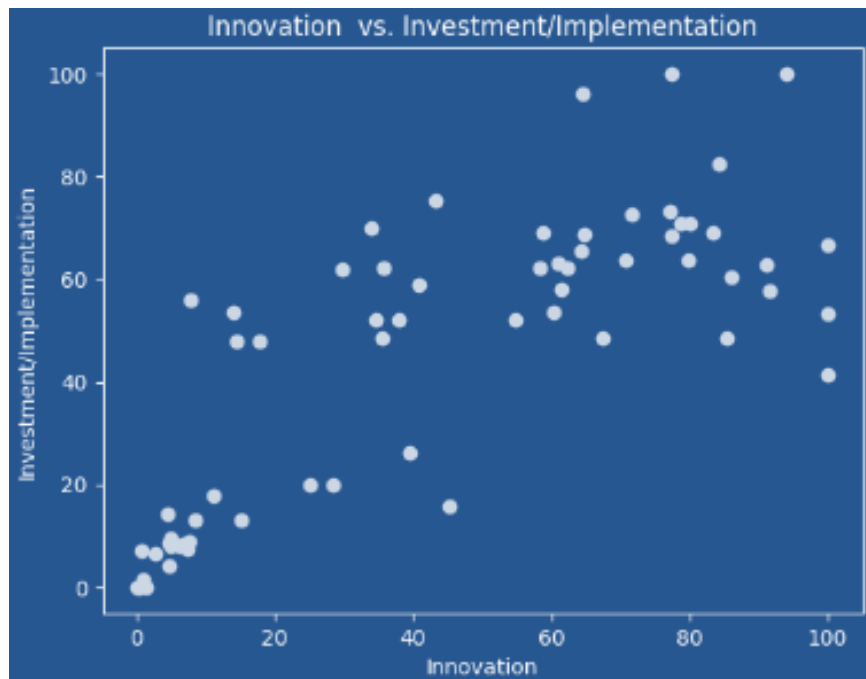


Рисунок 2.26 - Регресійний аналіз інновацій до множини інвестицій і впроваджень

Джерело:[2]

У ході даного регресійного аналізу, спрямованого на вивчення взаємозв'язку між рівнем інновацій та впроваджень і обсягом інвестицій, я спостерігав цікавий феномен. На Рисунку 2.25, що відображає попередні аналізи, помітно більше від'ємних відносин, які вказують на зменшення або

обмеження іновацій при збільшенні інвестицій. Такий контекст може бути сприйнятий як несподіваний та суперечливий. Згодом, на Рисунку 2.26, виявлено, що не дивлячись на більше від'ємних відносин, регресія в цілому має позитивну тенденцію. Це вказує на те, що при узгодженні ряду факторів, інвестиції ще завжди сприяють підвищенню іновацій і впроваджень. Такий результат може свідчити про складні взаємодії та компенсації, які виникають у системі, де різні фактори можуть взаємодіяти та коригувати один одного, надаючи позитивний вплив на іноваційні процеси. Отже, хоча спостереження на Рисунку 2.25 може викликати застереження, загальна тенденція позитивної регресії, відображена на Рисунку 2.26, вказує на те, що інвестиції взагалі сприяють стимулюванню іновацій та їхнього ефективного впровадження у вивченій сфері.

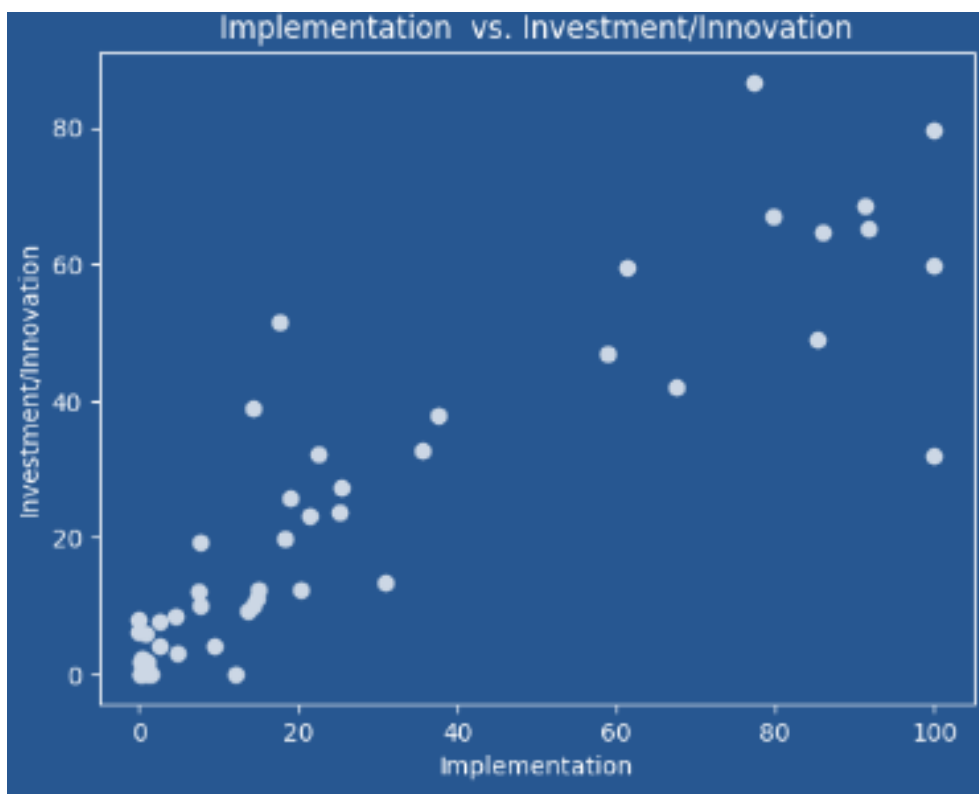


Рисунок 2.27 - Регресійний аналіз впроваджень до множини інвестицій та іновацій

Джерело:[2]

Під час регресійного аналізу, де основним фокусом була вивчення взаємозв'язку між обсягом впроваджень, інвестиціями та рівнем інновацій, виявився дуже цікавий результат. Суттєвий позитивний приріст регресії, свідчить про те, що зі зростанням інвестицій відбувається активне збільшення обсягу впроваджень та, відповідно, розвиток інноваційних процесів. Це важливе спостереження, яке зображено на рисунку 2.27, дозволяє зробити висновок, що фінансова підтримка у формі інвестицій має значущий позитивний вплив на впровадження нововведень у вивченій області. За таких умов навіть невеличкі збільшення обсягів інвестицій можуть генерувати значимий ріст у впровадженні та розвитку інновацій, відзначаючи важливість фінансових ресурсів у стимулюванні інноваційних ініціатив та покращенні їхнього впровадження.

Висновки до розділу 2

Обширне дослідження індексу штучного інтелекту свідчить, що країни "Великої сімки" та інші розвинені держави, такі як США, Японія, Китай та Європейський союз, відзначаються високим рівнем інтеграції штучного інтелекту. Особливо виділяються "потужні гравці" – США та Китай, які займають провідні позиції із найвищим рівнем індексу. Висновок полягає в тому, що країни з розвиненим соціально-економічним становищем та технічною готовністю ефективніше інтегруються в сферу штучного інтелекту.

Далі, розглядаючи аспекти розвитку штучного інтелекту, можна виокремити стратегії впливових країн. Наприклад, США активно інвестують в наукові дослідження та стартапи, Європейський союз встановлює стандарти етики та прозорості, Японія фокусується на робототехніці, а Китай має національний план для автоматизації виробництва. Також варто увагу

приділити регіонам, де індекс штучного інтелекту менш виражений, таким як Африка та Середній Схід, і думати про можливості співпраці для стимулювання їхнього розвитку.

На останньому етапі, індекс штучного інтелекту взаємодіє з соціально-економічними аспектами країн, змінюючи ринок праці, викликаючи етичні дебати та впливаючи на освіту. Також він розкриває важливість врахування культурних та регіональних особливостей для ефективного впровадження та розвитку штучного інтелекту в сучасному світі.

ВИСНОВКИ

Аналіз Глобального індексу штучного інтелекту (ШІ) дає повне розуміння різноманітності та розвитку готовності країн до ШІ. Вивчення ключових показників, зокрема талантів, інфраструктури, операційного середовища, досліджень і розробок, урядової стратегії та комерційної діяльності, проливає світло на особливості динаміки, що формують глобальну екосистему ШІ.

Метрика таланту показує спроможність націй у вихованні кваліфікованих професіоналів у сфері ШІ. Країни з надійною системою освіти, навчальними програмами та культурою інновацій демонструють конкурентну перевагу. Усвідомлюючи вирішальну роль таланту в розвитку ШІ, зусилля з виховання та залучення кваліфікованих людей стають першорядними для сталого зростання.

Інфраструктура є наріжним показником, що відображає технологічну основу нації. Розширені обчислювальні ресурси, інфраструктура даних і підключення є ключовими для розробки та розгортання програм ШІ. Країни з найсучаснішою інфраструктурою мають кращі можливості для використання трансформаційного потенціалу ШІ в різних галузях.

Метрика операційного середовища охоплює нормативні, правові та етичні рамки, які регулюють розгортання ШІ. Країни, які сприяють створенню сприятливого середовища з чіткими правилами, етичними принципами та рамками для відповідального використання ШІ, створюють основу для сталого розвитку ШІ. Встановлення правильного балансу між інноваціями та етичними міркуваннями є обов'язковим для довгострокового успіху.

Метрика досліджень і розробок відображає прихильність країни до просування штучного інтелекту шляхом постійних інновацій. Країни, які інвестують значні кошти в науково-дослідні ініціативи, сприяють співпраці

між академічними та промисловими колами та підтримують передові дослідження, роблять значний внесок у глобальну базу знань ШІ. Надійна екосистема досліджень і розробок відіграє важливу роль у розширенні можливостей ШІ.

Державна стратегія постає як керівна сила, яка формує підхід країни до розвитку ШІ. Чітка політика, стратегічні інвестиції та проактивна позиція щодо нових технологій свідчать про прагнення уряду використовувати переваги ШІ. Скоординовані зусилля державного та приватного секторів лежать в основі успішних стратегій ШІ.

Комерційна діяльність, основний показник, відображає застосування та інтеграцію штучного інтелекту в різних секторах. Країни з жвавими комерційними екосистемами свідчать про широке впровадження технологій штучного інтелекту в галузях, що сприяє економічному зростанню та інноваціям. Надійна комерційна діяльність означає практичне впровадження рішень штучного інтелекту, перетворюючи теоретичні досягнення у відчутні вигоди.

Таким чином успішними будуть ті країни, які швидко адаптуються до нових тенденцій, надають пріоритет етичним міркуванням і сприяють спільному середовищу, яке заохочує відповідальний розвиток і застосування технологій ШІ. Глобальний індекс служить не лише еталоном для порівняльного аналізу, але й компасом, який спрямовує країни до майбутнього, де штучний інтелект робить позитивний внесок у суспільства, економіку та добробут людства.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. Holistic-View-on-AI-Investments URL:
<https://www.kaggle.com/code/gvyshnya/global-ai-index-insights#Holistic-View-on-III-Investments>;
2. AI Global Index <https://www.kaggle.com/datasets/katerynameleshenko/ai-index%E2%80%8D>;
3. Machine learning mastery URL: <https://machinelearningmastery.ua/a-beginners-guide-to-linear-regression-in-python-with-scikit-learn-83a8f7ae2b4f/>;
4. An introduction to seaborn
<https://seaborn.pydata.org/tutorial/introduction.html#a-high-level-api-for-statistical-graphics>;
5. Aiindex <https://aiindex.stanford.edu/report/>;
6. What is exploratory data analysis: <https://www.ibm.com/topics/exploratory-data-analysis>;
7. Exploratory data analysis
https://en.wikipedia.org/wiki/Exploratory_data_analysis;
8. III Narratives: A History of Imaginative Thinking about Intelligent Machines
<https://academic.oup.com/book>;
9. Штучний інтелект. Простий путівник, який допоможе зрозуміти III
<https://www.bbc.com/ukrainian/resources/idt-74697280-e684-43c5-a782-29e9d11fecf3>;
10. Маккинни У. Python та аналіз даних(ДМК Пресс) – 2020. – №2.;
11. Daniel Y. Chen Pandas for everyone– 2023– №2.;
12. Діаграми playground <https://www.yworks.com/products/yfiles/playground>;
13. КІІ ім. Ігоря Сікорського Методологія науково-дослідницької діяльності – 2021;
14. Т.Г. Ярних Методологія та методи наукового аналізу –2018;

15. What Is an Index? Examples, How It's Used
<https://www.investopedia.com/terms/i/index.asp>;
16. Регресійний аналіз - Теорія економічного аналізу
https://pidru4niki.com/17280924/ekonomika/regresiyinyi_analiz;
17. Індекс
<https://uk.wikipedia.org/wiki/%D0%86%D0%BD%D0%B4%D0%B5%D0%BA%D1%81>;
18. Західний світ
https://uk.wikipedia.org/wiki/%D0%97%D0%B0%D1%85%D1%96%D0%B4%D0%BD%D0%B8%D0%B9_%D1%81%D0%B2%D1%96%D1%82;
19. Дисперсійний аналіз <https://habr.com/ru/companies/otus/articles/734258/>;
20. Т. А. Ліхоузова, Ю. О. Олійник Аналіз даних в інформаційних системах – 2022. 328с;
21. Основи статистики та аналізу даних
<https://socialdata.org.ua/manual/manual4/>;
22. What is machine learning. URL: <https://www.ibm.com/topics/machine-learning>;
23. Unsupervised learning for data classification. URL:
<https://developer.ibm.com/articles/cc-unsupervised-learning-data-classification/>
24. Історія Штучного Інтелекту. URL: <https://tyzhden.ua/korotka-istoriia-shtuchnoho-intelektu/>
25. Кластеризація семантичного ядра. URL:
<https://serpstat.com/uk/tutorial/1003-klasterizatciya/>
26. ДИСПЕРСІЙНИЙ АНАЛІЗ. URL:
https://pidru4niki.com/12991010/statistika/dispersiyniy_analiz
27. Аналіз Дисперсійного Аналізу. URL: chrome-extension://efaidnbmnnnibpcajpcgclefindmkaj/http://virtuni.education.zp.ua/edu_cpu/file.php/2753/Analiz_danikh_3.pdf
28. Розвідувальний аналіз даних. URL: <http://surl.li/pyafy>
29. Багатофакторний експеримент і його організація. URL: <http://surl.li/pyagi>

30. Кластерний аналіз. URL:
https://pidru4niki.com/11800912/ekonomika/klasterniy_analiz
31. Розкриття можливостей кластерного аналізу. URL: <http://surl.li/pyaox>
32. Гетероскедастичність. URL: <http://surl.li/pyatx>
33. Гетероскедастичність залишків. URL: <http://surl.li/pyaug>
34. Поняття мультиколінеарності. URL:
<https://studfile.net/preview/9396413/page:2/>
35. Мультиколінеарність. URL: <http://surl.li/pyaun>
36. Регресійний аналіз. URL: <https://epi.cc.ua/282-regressionnyiy-analiz-26451.html>
37. Використання регресійного аналізу для розуміння складних взаємозв'язків. URL: <http://surl.li/pyazv>
38. МЕТОДИ КОРЕЛЯЦІЙНОГО І ДИСПЕРСІЙНОГО АНАЛІЗУ. URL:
https://stud.com.ua/170830/pedagogika/metodi_korelyatsiynogo_dispersiynogo_analizu
39. pandas: powerful Python data analysis toolkit. URL:
<https://pypi.org/project/pandas/>
40. ЩО ТАКЕ ШТУЧНИЙ ІНТЕЛЕКТ: ІСТОРІЯ, ВИДИ ТА СКЛАДОВІ.
URL: <https://gigacloud.ua/blog/navchannja/scho-take-shtuchnij-intelekt-istorija-vidi-ta-skladovi>
41. Як використання ШІ впливає на мистецтво, кіно, музику та літературу.
URL: <https://suspilne.media/culture/645624-vsi-strivozeni-ak-vikoristanna-si-vplivae-na-mistectvo-kino-muziku-ta-literaturu/>
42. Що може зробити зі світом штучний інтелект. URL:
<https://www.radiosvoboda.org/a/details/28891073.html>
43. Кореляційний аналіз, мета і завдання. URL: <http://surl.li/pybfd>
44. Кореляційний аналіз даних: Що це таке і як його застосовувати.
<https://ua5.org/model/1728-korelyacijnyj-analiz-danyh-shho-cze-take-i-yak-jogo-zastosovuvaty.html>

45. Кореляційний аналіз. URL:
https://pidru4niki.com/14990528/ekonomika/korelyatsiyniy_analiz
46. Теоретичні основи і принципова схема дисперсійного аналізу. URL:
https://pidru4niki.com/1929100153035/statistika/dispersiyniy_analiz
47. СТАТИСТИЧНІ МЕТОДИ КОНТРОЛЮ ТА АНАЛІЗУ ЯКОСТІ
ПРОДУКЦІЇ. URL:
https://stud.com.ua/121019/prirodovnavstvo/dispersiyniy_analiz
48. Що таке лінійна регресія. URL: <https://www.unite.ai/uk/what-is-linear-regression/>
49. Застосування лінійної регресії. URL: <https://robotdreams.cc/uk/blog/437-shcho-take-liniyna-regresiya>
50. Лінійна регресія та кореляція. URL: <http://surl.li/pybgu>
51. Диха М.С Мороз В.М. Економетрія. Навчальний посібник. к.: Центр навчальної літератури. 2019. 206с
52. Наконечний С. І., Терещенко Т.О., Романюк Т. П. Економетрія: Підручник. вид. 3-тє, доп. та перероб. к.: КНЕУ, 2004. 520 с.
53. Математичні моделі та методи ринкової економіки : навч. посіб. / В. В. Вітлінський, О. В. Піскунова. К. : КНЕУ, 2010. 531с.
54. Математичні моделі та методи ринкової економіки : практикум / В.В. Вітлінський, О.В. Піскунова, О.В. Ткач, В.І. Скіцько, О.М. Новоселецький. – К.: КНЕУ, 2014.- 362с.
55. МАКРОЕКОНОМІКА: Теорія та політика. Підручник / І. Ф. Радіонова. – К.: Таксон, 2004.- 348с

ДОДАТОК А

```
# -*- coding: utf-8 -*-
"""data_analice.py

"""

from google.colab import drive
drive.mount('/content/drive')

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.cm as cm
import seaborn as sns
import plotly.express as px
import plotly.graph_objects as go
import plotly.offline

from plotly.subplots import make_subplots

plt.style.use('dark_background')

df = pd.read_csv('/content/drive/MyDrive/Python/Pandas/ai_global_index/AI_index_db.csv')

df

fig = make_subplots(rows=2, cols=4, subplot_titles=('<b>Distribution of Talent</b>',
                                                    '<b>Distribution of Infrastructure</b>',
                                                    '<b>Distribution of Operating Environment</b>',
                                                    '<b>Distribution of Research</b>',
                                                    '<b>Distribution of Development</b>',
                                                    '<b>Distribution of Government Strategy</b>'))
```

```

Commercial</b>',
                                '<b>Distribution  of
Total score</b>'
                                ))

fig.add_trace(go.Histogram(x=df['Talent'], nbinsx=30), row=1, col=1)
fig.add_trace(go.Histogram(x=df['Infrastructure']), row=1, col=2)
fig.add_trace(go.Histogram(x=df['Operating_Environment'], nbinsx=30),
row=1, col=3)
fig.add_trace(go.Histogram(x=df['Research'], nbinsx=30), row=1, col=4)

fig.add_trace(go.Histogram(x=df['Development'], nbinsx=30), row=2,
col=1)
fig.add_trace(go.Histogram(x=df['Government_Strategy']), row=2, col=2)
fig.add_trace(go.Histogram(x=df['Commercial'], nbinsx=30), row=2, col=3)
fig.add_trace(go.Histogram(x=df['Total_score'], nbinsx=30), row=2,
col=4)

# Update visual layout
fig.update_layout(
    showlegend=False,
    width=800,
    height=500,
    autosize=False,
    margin=dict(t=15, b=0, l=5, r=5),
    template="plotly_white",
)

# update font size at the axes
fig.update_coloraxes(colorbar_tickfont_size=10)
# Update font in the titles: Apparently subplot titles are annotations
(Subplot font size is hardcoded to 16pt · Issue #985)
fig.update_annotations(font_size=10)
# Reduce opacity
fig.update_traces(opacity=0.75)

fig.show()

# sns.set_theme(style="darkgrid")

```

```

# # make a dataframe with only numeric variables, without 'Outcome'
# d = df.copy()

# # Compute the correlation matrix for numeric features
# corr = d.corr()

# # Generate a mask for the upper triangle
# mask = np.triu(np.ones_like(corr, dtype=bool))

# # Set up the matplotlib figure
# f, ax = plt.subplots(figsize=(11, 9))

# # Generate a custom diverging colormap
# cmap = sns.diverging_palette(230, 20, as_cmap=True)

# # Draw the heatmap with the mask and correct aspect ratio
# sns.heatmap(corr, mask=mask, cmap=cmap, vmax=.3, center=0, annot=True,
#             square=True, linewidths=.5, cbar_kws={"shrink": .5})

# dfg = df['Region'].value_counts().reset_index()
# dfg.columns = ['Region', 'Quantity']
# fig = px.bar(dfg, x='Region', y='Quantity',
#             title='Number of Countries in Report by Region
# (Continent)',
#             color_discrete_sequence=px.colors.qualitative.Prism)
# fig.show()

dfg = df['Cluster'].value_counts().reset_index()
dfg.columns = ['Cluster', 'Quantity']
fig = px.bar(dfg, x='Cluster', y='Quantity',
            title='Number of Countries in Report by Cluster',
            color_discrete_sequence=px.colors.qualitative.Prism)
fig.show()

df.describe()

df['Commercial'].hist()

df.groupby('Country', dropna=False)['Commercial'].mean().plot(kind='bar')

```

```

df.groupby('Region')['Talent'].var()

df.groupby('Region')['Talent'].var().plot(kind='bar')

"""# ***Talent***

---

# Talent by Regions
"""

def set_talent(x):
    res = "N/A"
    if x >= 80:
        res = "1 - Very high"
    elif x >= 60 and x < 80:
        res = "2 - High"
    elif x >= 40 and x < 60:
        res = "3 - Moderate"
    elif x >= 20 and x < 40:
        res = "4 - Low"
    else:
        res = "5 - Very low"
    return res

df["Talent_level"] = df["Talent"].apply(set_talent)

agg_data = df.groupby(['Region',
'Talent_level']).size().reset_index(name='Country_count')

agg_data = agg_data.sort_values(by=["Talent_level", "Country_count"])

fig = px.bar(
    agg_data,
    x='Region',
    y='Country_count',
    color='Talent_level',
    title="Countries by Region and Talent level",
    color_discrete_sequence=px.colors.qualitative.Prism,
    height=600)
fig.show()

```

```
fig = px.violin(df, y="Talent", x="Region", box=True, points="all",
title="Talent by Regions", hover_data=df.columns)
fig.show()
```

```
""""# Total Score by Clusters"""
```

```
agg_data = df.groupby(['Cluster',
'Talent_level']).size().reset_index(name='Country_count')
```

```
agg_data = agg_data.sort_values(by=["Talent_level", "Country_count"])
```

```
fig = px.bar(
    agg_data,
    x='Cluster',
    y='Country_count',
    color='Talent_level',
    title="Countries by Cluster and Talent level",
    color_discrete_sequence=px.colors.qualitative.Prism,
    height=600)
fig.show()
```

```
fig = px.violin(df, y="Talent", x="Cluster", box=True, points="all",
title="Talent by Cluster", hover_data=df.columns)
fig.show()
```

```
""""# Talent by Income groups"""
```

```
fig = px.violin(df, y="Talent", x="Income group", box=True,
points="all", title="Talent by Cluster", hover_data=df.columns)
fig.show()
```

```
agg_data = df.groupby(['Income group',
'Talent_level']).size().reset_index(name='Country_count')
```

```
agg_data = agg_data.sort_values(by=["Talent_level", "Country_count"])
```

```
fig = px.bar(
    agg_data,
    x='Income group',
    y='Country_count',
```

```

        color='Talent_level',
        title="Countries by Income group and Talent level",
        color_discrete_sequence=px.colors.qualitative.Prism,
        height=600)
fig.show()

"""# Talent by Political regimes"""

fig = px.violin(df, y="Talent", x="Political regime", box=True,
points="all", title="Talent by Political regime", hover_data=df.columns)
fig.show()

agg_data = df.groupby(['Political regime',
'Talent_level']).size().reset_index(name='Country_count')

agg_data = agg_data.sort_values(by=["Talent_level", "Country_count"])

fig = px.bar(
    agg_data,
    x='Political regime',
    y='Country_count',
    color='Talent_level',
    title="Countries by Political regime and Talent level",
    color_discrete_sequence=px.colors.qualitative.Prism,
    height=600)
fig.show()

"""# **`Infrastructure`**

---

# ***Infrastructure by Regions***
"""

fig = px.violin(df, y="Infrastructure", x="Region", box=True,
points="all", title="Infrastructure by Region", hover_data=df.columns)
fig.show()

def set_infrastructure_level(x):
    res = "N/A"
    if x >= 80:

```

```

        res = "1 - Very high"
    elif x >= 60 and x < 80:
        res = "2 - High"
    elif x >= 40 and x < 60:
        res = "3 - Moderate"
    elif x >= 20 and x < 40:
        res = "4 - Low"
    else:
        res = "5 - Very low"
    return res

df["Infrastructure_level"] =
df["Infrastructure"].apply(set_infrastructure_level)

agg_data = df.groupby(['Region',
'Infrastructure_level']).size().reset_index(name='Country_count')

agg_data = agg_data.sort_values(by=["Infrastructure_level",
"Country_count"])

fig = px.bar(
    agg_data,
    x='Region',
    y='Country_count',
    color='Infrastructure_level',
    title="Countries by Region and Infrastructure level",
    color_discrete_sequence=px.colors.qualitative.Prism,
    height=600)
fig.show()

"""# Infrastructure by Clusters"""

fig = px.violin(df, y="Infrastructure", x="Cluster", box=True,
points="all", title="Infrastructure by Cluster", hover_data=df.columns)
fig.show()

agg_data = df.groupby(['Cluster',
'Infrastructure_level']).size().reset_index(name='Country_count')

agg_data = agg_data.sort_values(by=["Infrastructure_level",
"Country_count"])

```

```

fig = px.bar(
    agg_data,
    x='Cluster',
    y='Country_count',
    color='Infrastructure_level',
    title="Countries by Cluster and Infrastructure level",
    color_discrete_sequence=px.colors.qualitative.Prism,
    height=600)
fig.show()

"""# Infrastructure by Income groups"""

fig = px.violin(df, y="Infrastructure", x="Income group", box=True,
points="all", title="Infrastructure by Income groups",
hover_data=df.columns)
fig.show()

agg_data = df.groupby(['Income group',
'Infrastructure_level']).size().reset_index(name='Country_count')

agg_data = agg_data.sort_values(by=["Infrastructure_level",
"Country_count"])

fig = px.bar(
    agg_data,
    x='Income group',
    y='Country_count',
    color='Infrastructure_level',
    title="Countries by Income group and Infrastructure level",
    color_discrete_sequence=px.colors.qualitative.Prism,
    height=600)
fig.show()

"""# Infrastructure by Political Regimes"""

fig = px.violin(df, y="Infrastructure", x="Political regime", box=True,
points="all", title="Infrastructure by Political regimes",
hover_data=df.columns)
fig.show()

```

```

agg_data = df.groupby(['Political regime',
'Infrastructure_level']).size().reset_index(name='Country_count')

agg_data = agg_data.sort_values(by=["Infrastructure_level",
"Country_count"])

fig = px.bar(
    agg_data,
    x='Political regime',
    y='Country_count',
    color='Infrastructure_level',
    title="Countries by Political regime and Infrastructure level",
    color_discrete_sequence=px.colors.qualitative.Prism,
    height=600)
fig.show()

"""# Infrastructure: Geospacial view"""

# set the size of the geo bubble
def set_infrastructure_size(value):
    ...

    Takes the numeric value of a parameter to visualize on a map (Plotly
    Geo-Scatter plot)

    Returns a number to indicate the size of a bubble for a country
    which numeric attribute value
    was supplied as an input
    ...

    result = np.log(1+value/1200)

    if result < 0:
        result = 0.001
    return result

df['Size'] = df.apply(lambda x:
set_infrastructure_size(x['Infrastructure']), axis=1)

fig = px.scatter_geo(
    df, locations="Country", locationmode='country names',
    color="Infrastructure",
    size='Size', hover_name="Country",
    range_color= [0, 100],

```

```

        projection="natural earth",
        title='AI Infrastructure Across the Globe',
        color_continuous_scale="portland_r")

# Update visual layout
fig.update_layout(
    showlegend=True,
    width=800,
    height=400,
    autosize=False,
    margin=dict(t=40, b=0, l=5, r=5),
    template="plotly_white",
)

fig.show()

"""# ***Operating Environment***

---

# Operating Environment by Regions
"""

fig = px.violin(df, y="Operating Environment", x="Region", box=True,
points="all", title="Operating Environment by Regions",
hover_data=df.columns)
fig.show()

# implement binning by Operating Environment categories

def set_ops_level(x):
    res = "N/A"
    if x >= 80:
        res = "1 - Very high"
    elif x >= 60 and x < 80:
        res = "2 - High"
    elif x >= 40 and x < 60:
        res = "3 - Moderate"
    elif x >= 20 and x < 40:
        res = "4 - Low"
    else:

```

```

        res = "5 - Very low"
    return res

df["Ops_Environment_level"] = df["Operating
Environment"].apply(set_ops_level)

agg_data = df.groupby(['Region',
'Ops_Environment_level']).size().reset_index(name='Country_count')

agg_data = agg_data.sort_values(by=["Ops_Environment_level",
"Country_count"])

fig = px.bar(
    agg_data,
    x='Region',
    y='Country_count',
    color='Ops_Environment_level',
    title="Countries by Region and Operating Environment level",
    color_discrete_sequence=px.colors.qualitative.Prism,
    height=600)
fig.show()

"""# Operating Environment by Clusters"""

fig = px.violin(df, y="Operating Environment", x="Cluster", box=True,
points="all", title="Operating Environment by Clusters",
hover_data=df.columns)
fig.show()

agg_data = df.groupby(['Cluster',
'Ops_Environment_level']).size().reset_index(name='Country_count')

agg_data = agg_data.sort_values(by=["Ops_Environment_level",
"Country_count"])

fig = px.bar(
    agg_data,
    x='Cluster',
    y='Country_count',
    color='Ops_Environment_level',
    title="Countries by Cluster and Operating Environment level",

```

```

        color_discrete_sequence=px.colors.qualitative.Prism,
        height=600)
fig.show()

"""# Operating Environment by Income Groups"""

fig = px.violin(df, y="Operating Environment", x="Income group",
box=True, points="all", title="Operating Environment by Income Groups",
hover_data=df.columns)
fig.show()

agg_data = df.groupby(['Income group',
'Ops_Environment_level']).size().reset_index(name='Country_count')

agg_data = agg_data.sort_values(by=["Ops_Environment_level",
"Country_count"])

fig = px.bar(
    agg_data,
    x='Income group',
    y='Country_count',
    color='Ops_Environment_level',
    title="Countries by Income Group and Operating Environment level",
    color_discrete_sequence=px.colors.qualitative.Prism,
    height=600)
fig.show()

"""# Operating Environment by Political Regimes"""

fig = px.violin(df, y="Operating Environment", x="Political regime",
box=True, points="all", title="Operating Environment by Political
regimes", hover_data=df.columns)
fig.show()

agg_data = df.groupby(['Political regime',
'Ops_Environment_level']).size().reset_index(name='Country_count')

agg_data = agg_data.sort_values(by=["Ops_Environment_level",
"Country_count"])

fig = px.bar(

```

```

    agg_data,
    x='Political regime',
    y='Country_count',
    color='Ops_Environment_level',
    title="Countries by Political Regime and Operating Environment
level",
    color_discrete_sequence=px.colors.qualitative.Prism,
    height=600)
fig.show()

"""# *Operating Environment: Geospacial view*"""

# set the size of the geo bubble
def set_op_env_size(value):
    ...

    Takes the numeric value of a parameter to visualize on a map (Plotly
Geo-Scatter plot)

    Returns a number to indicate the size of a bubble for a country
which numeric attribute value
was supplied as an input
    ...

    result = np.log(1+value/1200)

    if result < 0:
        result = 0.001
    return result

df['Size'] = df.apply(lambda x: set_op_env_size(x['Operating
Environment']), axis=1)

fig = px.scatter_geo(
    df, locations="Country", locationmode='country names',
    color="Operating Environment",
    size='Size', hover_name="Country",
    range_color= [0, 100],
    projection="natural earth",
    title='AI Operating Environment Across the Globe',
    color_continuous_scale="portland_r")

# Update visual layout
fig.update_layout(

```

```

        showlegend=True,
        width=800,
        height=400,
        autosize=False,
        margin=dict(t=40, b=0, l=5, r=5),
        template="plotly_white",
    )

fig.show()

"""# Holistic View on AI Implementation"""

fig = px.scatter_3d(df, x='Income group', y='Talent',
z='Infrastructure', color='Operating_Environment', hover_name="Country",
                    title='Talent-Infrastructure-Operating Environment
Inter-relations by Income Groups')
# Update visual layout
fig.update_layout(
    showlegend=True,
    width=800,
    height=400,
    autosize=False,
    margin=dict(t=35, b=35, l=5, r=5),
    template="plotly_white",
)
fig.show()

"""# Holistic View on AI Innovations

***R&D Inter-relations by Regions***
"""

fig = px.scatter_3d(df, x='Region', y='Research', z='Development',
color='Total_score', hover_name="Country",
                    title='R&D Inter-relations by Regions')
# Update visual layout
fig.update_layout(
    showlegend=True,
    width=1000,
    height=800,
    autosize=False,

```

```

        margin=dict(t=35, b=35, l=5, r=5),
        template="plotly_white",
    )
    fig.show()

    """***R&D Inter-relations by Clusters***"""

    fig = px.scatter_3d(df, x='Cluster', y='Research', z='Development',
                        color='Total_score', hover_name="Country",
                        title='R&D Inter-relations by Clusters')
    # Update visual layout
    fig.update_layout(
        showlegend=True,
        width=1000,
        height=800,
        autosize=False,
        margin=dict(t=35, b=35, l=5, r=5),
        template="plotly_white",
    )
    fig.show()

    """***R&D Inter-relations by Income groups***"""

    fig = px.scatter_3d(df, x='Income group', y='Research', z='Development',
                        color='Total_score', hover_name="Country",
                        title='R&D Inter-relations by Income groups')
    # Update visual layout
    fig.update_layout(
        showlegend=True,
        width=1000,
        height=800,
        autosize=False,
        margin=dict(t=35, b=35, l=5, r=5),
        template="plotly_white",
    )
    fig.show()

    """***R&D Inter-relations by Political regimes***"""

    fig = px.scatter_3d(df, x='Political_regime', y='Research',
                        z='Development', color='Total_score', hover_name="Country",

```

```

        title='R&D Inter-relations by Political regimes')
# Update visual layout
fig.update_layout(
    showlegend=True,
    width=1000,
    height=800,
    autosize=False,
    margin=dict(t=35, b=35, l=5, r=5),
    template="plotly_white",
)
fig.show()

"""# **Holistic View on AI Investments**

***AI Investments by Regions ***
"""

fig = px.scatter_3d(df, x='Region', y='Government_Strategy',
z='Commercial', color='Total_score', hover_name="Country",
                    title='AI Investments by Regions')
# Update visual layout
fig.update_layout(
    showlegend=True,
    width=800,
    height=600,
    autosize=False,
    margin=dict(t=35, b=35, l=5, r=5),
    template="plotly_white",
)
fig.show()

"""***AI Investments by Clusters***"""

fig = px.scatter_3d(df, x='Cluster', y='Government Strategy',
z='Commercial', color='Total score', hover_name="Country",
                    title='AI Investments by Clusters')
# Update visual layout
fig.update_layout(
    showlegend=True,
    width=800,
    height=700,

```

```

        autosize=False,
        margin=dict(t=35, b=35, l=5, r=5),
        template="plotly_white",
    )
fig.show()

"""AI Investments by Income groups"""

fig = px.scatter_3d(df, x='Income group', y='Government_Strategy',
z='Commercial', color='Total_score', hover_name="Country",
                    title='AI Investments by Income groups')
# Update visual layout
fig.update_layout(
    showlegend=True,
    width=1000,
    height=800,
    autosize=False,
    margin=dict(t=35, b=35, l=5, r=5),
    template="plotly_white",
)
fig.show()

"""***AI Investments by Political regimes***"""

fig = px.scatter_3d(df, x='Political regime', y='Government Strategy',
z='Commercial', color='Total score', hover_name="Country",
                    title='AI Investments by Political regimes')
# Update visual layout
fig.update_layout(
    showlegend=True,
    width=800,
    height=800,
    autosize=False,
    margin=dict(t=35, b=35, l=5, r=5),
    template="plotly_white",
)
fig.show()

```

```

# -*- coding: utf-8 -*-
"""data_anallice2.py

"""

from google.colab import drive
drive.mount('/content/drive')

# Commented out IPython magic to ensure Python compatibility.
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn import metrics

# %matplotlib inline
plt.style.use('dark_background')

df = pd.read_csv('/content/drive/MyDrive/Python/Pandas/ai_global_index/AI_index_db.csv')

df

"""# Investment"""

X = df[['Talent', 'Infrastructure', 'Operating_Environment', 'Research',
'Development', 'Government_Strategy', 'Commercial']]
y = df[['Total_score']]

# Split the data into training and testing sets (optional but recommended)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

model = LinearRegression()
model.fit(X_train, y_train)

```

```

y_pred = model.predict(X_test)

y_pred_clipped = np.clip(y_pred, 0, 100)

plt.scatter(y_test, y_pred_clipped)
plt.xlabel("Investment")
plt.ylabel("Innovation/Implementation")
plt.title("Investment vs. Innovation/Implementation")
plt.show()

plt.plot([min(X_test), max(X_test)], [min(y_test), max(y_test)],
color='red', linewidth=2, label='Regression Line') # Regression line
at the mean

plt.legend()
plt.show()

X = df[['Government_Strategy', 'Commercial']]
y = df[['Talent', 'Infrastructure', 'Operating_Environment', 'Research',
'Development']]

# Split the data into training and testing sets (optional but
recommended)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

model = LinearRegression()
model.fit(X_train, y_train)

y_pred = model.predict(X_test)

y_pred_clipped = np.clip(y_pred, 0, 100)

plt.scatter(y_test, y_pred_clipped)
plt.xlabel("Investment")
plt.ylabel("Innovation/Implementation")
plt.title("Investment vs. Innovation/Implementation")
plt.show()

```

```

plt.plot([min(X_test), max(X_test)], [min(y_test), max(y_test)],
color='red', linewidth=2, label='Regression Line') # Regression line
at the mean

plt.legend()
plt.show()

"""# **Innovation**"""

X = df[['Research', 'Development']]
y = df[['Talent', 'Infrastructure',
'Operating_Environment', 'Government_Strategy', 'Commercial']]

# Split the data into training and testing sets (optional but
recommended)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

model = LinearRegression()
model.fit(X_train, y_train)

y_pred = model.predict(X_test)

y_pred_clipped = np.clip(y_pred, 0, 100)

plt.scatter(y_test, y_pred_clipped)

plt.xlabel("Innovation")
plt.ylabel("Investment/Implementation")
plt.title("Innovation vs. Investment/Implementation")
plt.show()
plt.plot([min(X_test), max(X_test)], [min(y_test), max(y_test)],
color='red', linewidth=2, label='Regression Line') # Regression line
at the mean

plt.legend()
plt.show()

"""# **Implementation**"""

```

```

X = df[['Talent', 'Infrastructure', 'Operating_Environment']]
y = df[['Research', 'Development', 'Government_Strategy', 'Commercial']]

# Split the data into training and testing sets (optional but
recommended)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

model = LinearRegression()
model.fit(X_train, y_train)

y_pred = model.predict(X_test)

y_pred_clipped = np.clip(y_pred, 0, 100)

plt.scatter(y_test, y_pred_clipped)
plt.xlabel("Implementation")
plt.ylabel("Investment/Innovation")
plt.title("Implementation vs. Investment/Innovation")
plt.show()

plt.plot([min(X_test), max(X_test)], [min(y_test), max(y_test)],
color='red', linewidth=2, label='Regression Line') # Regression line
at the mean

plt.legend()
plt.show()

"""# ** _____ **"""

df.describe()

dfm = df.describe()

df.plot(x='Infrastructure', y='Commercial', style='o')
plt.title('Commercial < Infrastructure')

```

```
plt.xlabel('Infrastructure')
plt.ylabel('Commercial')
plt.show()

plt.plot(dfm['Infrastructure']['mean'], dfm['Commercial']['mean'])

plt.figure(figsize=(15,10))
plt.tight_layout()
sns.distplot(x = df['Commercial'], kde=True)
```

ДОДАТОК Б

Country	Talent	Infrastructure	Operating_Environment	Research	Development	Governm ent_Strategy	Commercial	Total_score	Region	Cluster	Income group	Political_regi me
United States of America	100	94.02	64.56	100	100	77.39	100	100	Americas	Power players	High	Liberal democracy
China	16.51	100	91.57	71.42	79.97	94.87	44.02	62.92	Asia-Pacific	Power players	Upper middle	Closed autocracy
United Kingdom	39.65	71.43	74.65	36.5	25.03	82.82	18.91	40.93	Europe	Traditional champions	High	Liberal democracy
Canada	31.28	77.05	93.94	30.67	25.78	100	14.88	40.19	Americas	Traditional champions	High	Liberal democracy
Israel	35.76	67.58	82.44	32.63	27.96	43.91	27.33	39.89	Middle East	Rising stars	High	Liberal democracy
Singapore	39.38	84.3	43.15	37.67	22.55	79.82	15.07	38.67	Asia-Pacific	Rising stars	High	Elector al democracy
South Korea	14.54	85.23	68.86	26.66	77.25	87.5	5.41	38.6	Asia-Pacific	Rising stars	High	Liberal democracy
The Netherlands	33.83	81.99	88.05	25.54	30.17	62.35	4.97	36.35	Europe	Rising stars	High	Liberal democracy
Germany	27.63	77.22	70.22	35.84	24.79	84.65	8.29	36.04	Europe	Traditional champions	High	Liberal democracy
France	28.32	77.15	80.02	25.48	21.44	91.2	7.65	34.42	Europe	Traditional champions	High	Liberal democracy

										mpions		
Australia	25.43	63.43	61.23	32.63	41.15	82.11	6.72	33.86	Asia-Pacific	Rising stars	High	Liberal democracy
Ireland	29.93	89.5	70.15	16.79	30.85	69.44	3.94	33.04	Europe	Rising stars	High	Liberal democracy
Finland	24.99	71.6	78.76	25.21	18.32	85.99	4.64	31.36	Europe	Rising stars	High	Liberal democracy
Denmark	27.07	74.08	85.39	26.01	8.92	74.23	3.46	30.87	Europe	Rising stars	High	Liberal democracy
Luxembourg	21.66	94.88	66.96	19.39	19.95	66.69	4.68	30.73	Europe	Waking up	High	Liberal democracy
Japan	15.18	84.58	57.53	22.51	34.47	71.96	7.31	30.53	Asia-Pacific	Rising stars	High	Liberal democracy
India	45.27	33.91	77.3	18.92	30.86	58.83	7.39	30.36	Asia-Pacific	Waking up	Lower middle	Electoral autocracy
Switzerland	25.63	78.43	44.14	38.24	23.11	12.18	7.76	30.25	Europe	Rising stars	High	Liberal democracy
Sweden	28.21	75.19	66.77	27.61	17.81	40.35	4.51	29.85	Europe	Waking up	High	Liberal democracy
Hong Kong	17.56	96.11	59.5	31.51	8.63	33.29	5.3	29.11	Asia-Pacific	Waking up	High	Electoral democracy
Spain	17.61	73.32	75.36	18.6	10.87	91.28	3.08	26.95	Europe	Rising stars	High	Liberal democracy
Austria	16.97	64.49	76.3	23.56	17.81	72.14	3.08	26.89	Europe	Waking up	High	Electoral democracy
Estonia	18.74	63.65	88.67	11.75	9.31	72.08	12.51	26.6	Europe	Waking up	High	Liberal democracy
Taiwan	12.34	77.86	56.67	25.71	19.99	55.97	2.53	25.79	Asia-	Waking	High	Liberal democ

									Pacific	up		racy
Norway	27.61	76.2	36.65	21.18	13.56	59.05	3.95	25.77	Europe	Waking up	High	Liberal democracy
Saudi Arabia	4.49	70.8	100	13.63	14.38	91.63	4.73	25.6	Middle East	Waking up	High	Closed autocracy
Belgium	15.17	65.1	64.08	22.15	19.81	63.58	5.31	25.52	Europe	Waking up	High	Liberal democracy
Poland	14.21	70.96	99.56	10.6	9.09	78.14	2.25	25.2	Europe	Waking up	High	Electoral democracy
Slovenia	13.02	72.06	94.55	19.1	1.06	80.38	0.61	25.19	Europe	Waking up	High	Electoral democracy
New Zealand	23.3	69.78	90.35	12.23	5.96	47.62	2.49	24.88	Asia-Pacific	Waking up	High	Liberal democracy
Italy	11.09	64.76	83.25	20.3	14.66	61.43	2.64	24.45	Europe	Waking up	High	Liberal democracy
Russia	12.46	62.59	52.85	14.21	19.48	90.4	1.38	21.99	Europe	Waking up	Upper middle	Electoral autocracy
Malta	15.87	67.12	70.96	5.96	11.72	70.49	4.3	21.85	Europe	Waking up	High	Electoral democracy
United Arab Emirates	2.65	79.16	72.12	5.13	15.53	81.38	3.22	21.17	Middle East	Waking up	High	Closed autocracy
Portugal	13.43	64.2	80.66	8.96	3.92	70.69	2.05	20.89	Europe	Waking up	High	Electoral democracy
Czech Republic	11.11	64.26	76.97	11.26	2.7	70.29	1.75	20.31	Europe	Waking up	High	Electoral democracy
Iceland	18.45	72.45	41.19	18.29	0.19	22.15	5.74	19.81	Europe	Waking up	High	Liberal democracy
Lithuania	14.3	63.1	80.67	3.2	6.18	64.28	1.77	19.5	Eur	Wak	High	Elector

		9		2				9	ope	ing up		al democ racy
Brazil	13.46	62.6 1	72.82	4.8 3	5.07	67.72	1.36	18.8 9	Americ as	Waking up	Upper middle	Elector al democ racy
Greece	7.62	55.4 4	83.58	15. 12	2.21	22.15	0.92	17.3 3	Eur ope	Waking up	High	Liberal democ racy
Slovakia	8.55	65.3 6	88.71	2.9 7	0.34	43.07	0.67	17.2 4	Eur ope	Waking up	High	Elector al democ racy
Hungary	10.34	69.1 7	58.01	4.3 1	5.4	55.01	1.08	17	Eur ope	Waking up	High	Elector al autocr acy
Malaysia	10.44	62.0 4	73.24	5.5 7	0.88	47.6	0.63	16.6 6	Asi a- Pac ific	Waking up	Upper middle	Elector al autocr acy
Mexico	1.72	41.8 5	97.03	8.1 1	4.46	54.21	0.78	15.4 8	Americ as	Waking up	Upper middle	Elector al democ racy
Chile	10.56	61.9 7	56.73	1.4 9	0.67	60.5	2.95	15.3 3	Americ as	Waking up	High	Liberal democ racy
Argentina	8.4	56.1 5	75.95	1.2 5	3.19	54.94	0.34	15.2 4	Americ as	Waking up	Upper middle	Elector al democ racy
Qatar	0	67.9 7	62.58	11. 94	0	33.49	0	14.6 6	Midl dle Eas t	Waking up	High	Closed autocr acy
Turkey	8.49	40.7 4	64.34	9.5 3	1.02	67.45	0.95	14.3 9	Eur ope	Waking up	High	Elector al democ racy
Colombia	4.97	54.8	62.37	0	0.89	85.29	0.5	13.8 5	Americ as	Waking up	Upper middle	Elector al democ racy
Uruguay	7.28	58.7 7	70.75	0.7 3	0.26	30.92	0.43	13.2 7	Americ as	Waking up	High	Liberal democ racy
Bahrain	4.99	60.3 9	60.9	2.5 3	0	17.72	0.24	11.7 9	Midl dle Eas	Waking up	High	Closed autocr acy

									t			
Vietnam	6.34	55.76	37.61	2.03	0.3	68.86	0.31	11.62	Asia-Pacific	Waking up	Lower middle	Closed autocracy
Indonesia	5.51	47.52	51.18	0.98	3.52	59.99	0.91	11.47	Asia-Pacific	Waking up	Lower middle	Electoral democracy
Tunisia	8.94	38.59	62.32	3.9	0	12.18	0.15	9.87	Africa	Waking up	Lower middle	Electoral democracy
South Africa	4.61	45.73	58.43	0.83	7.52	0	2.03	9.71	Africa	Waking up	Upper middle	Electoral democracy
Morocco	3.36	44.88	60.17	1.46	0.05	15.9	0.1	8.87	Africa	Waking up	Lower middle	Closed autocracy
Armenia	6.69	37.84	58.4	0.28	0.33	14.4	1.37	8.49	Europe	Waking up	Upper middle	Electoral democracy
Sri Lanka	6.27	34.64	35.79	0.12	0.95	35.57	0.09	6.62	Asia-Pacific	Nascent	Lower middle	Electoral democracy
Egypt	1.11	38.84	0	2.08	1.54	68.72	0.31	4.83	Middle East	Nascent	Lower middle	Electoral autocracy
Kenya	0.75	14.11	29.84	0.07	12.15	7.75	0.31	2.3	Africa	Nascent	Lower middle	Electoral autocracy
Nigeria	2.74	0	50.1	0.45	2.06	7.75	0.33	1.38	Africa	Nascent	Lower middle	Electoral autocracy
Pakistan	8	2.43	12.48	2.17	1.09	13.92	0.27	0	Asia-Pacific	Nascent	Lower middle	Electoral autocracy