

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Західноукраїнський національний університет
Факультет комп'ютерних інформаційних технологій
Кафедра інформаційно-обчислювальних систем і управління

КОТКЕВИЧ Юрій Володимирович

**Модуль визначення сприятливого середовища для життя
та розвитку людей / Module for determining a favorable
environment for life and development of people**

Спеціальність 122 – Комп'ютерні науки
Освітньо-професійна програма – Комп'ютерні науки

Дипломний проект

Виконав студент групи КН-41
Ю. В. Коткевич

Науковий керівник:
викладач В.І. Дорош

Дипломний проект допущено до захисту
«__» _____ 2023 р.

Завідувач кафедри
_____ М.П. Комар

Тернопіль – 2023

Факультет комп'ютерних інформаційних технологій
Кафедра інформаційно-обчислювальних систем і управління
Освітній ступінь «бакалавр»
Спеціальність 122 – Комп'ютерні науки
Освітньо-професійна програма – Комп'ютерні науки

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ М.П. Комар
« _____ » _____ 2022р.

З А В Д А Н Н Я

НА ДИПЛОМНИЙ ПРОЕКТ СТУДЕНТУ

Коткевичу Юрію Володимировичу

(прізвище, ім'я, по батькові)

1. Тема проекту: Модуль визначення сприятливого середовища для життя та розвитку людей / Module for determining a favorable environment for life and development of people

керівник проекту викладач В.І. Дорош

затверджені наказом по університету від 08 грудня 2022 р. № 491.

2. Строк подання студентом закінченого проекту 01 червня 2023 р.

3. Вихідні дані до проекту: технічне завдання.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)

- оглянути принципи та етапи визначення сприятливого середовища для життя та розвитку людей.
- проаналізувати існуючі підходи до визначення сприятливого середовища для життя та розвитку людей.
- сформулювати задачі для розробки модуля визначення сприятливого середовища для життя та розвитку людей на основі індексу щастя.
- розробити алгоритми на основі машинного навчання для визначення сприятливого середовища для життя та розвитку людей на основі індексу щастя.
- реалізувати програмний модуль для визначення сприятливого середовища для життя та розвитку людей на основі індексу щастя.
- перевірити ефективність та точність розробленого модулю на реальних даних.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)
- алгоритми на основі машинного навчання для визначення сприятливого середовища для життя та розвитку людей на основі індексу щастя.

6. Консультанти розділів проекту

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		Завдання видав	Завдання прийняв
Н. контроль	викладач В.І. Дорош		

7. Дата видачі завдання 08 грудня 2022 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проекту	Строк виконання етапів проекту	Примітка
1	Аналіз методів та засобів визначення сприятливого середовища для життя та розвитку людей	30.12.2022	
2	Розробка алгоритмів на основі машинного навчання для визначення сприятливого середовища для життя та розвитку людей на основі індексу щастя	24.03.2023	
3	Програмна реалізація модуля визначення сприятливого середовища для життя та розвитку людей на основі індексу щастя	12.05.2023	
4	Повне завершення та оформлення дипломного проекту	01.06.2023	

Студент _____ Ю. В. Коткевич
(підпис)

Керівник проекту _____ В.І. Дорош
(підпис)

РЕФЕРАТ

Пояснювальна записка до дипломного проекту: 68 с., 21 рис., 2 додатки, 54 джерел.

Метою дипломного проекту є розробка модуля, який використовує методи машинного навчання для аналізу різних факторів, що впливають на якість життя та рівень щастя людей, і визначення сприятливих умов для створення середовища, яке сприяє розвитку людей в містах.

Об'єкт дослідження є взаємозв'язки між різними факторами, такими як економічні, соціальні, екологічні та інфраструктурні, що впливають на якість життя та рівень щастя населення, а також методи машинного навчання, які дозволяють аналізувати та прогнозувати ці взаємозв'язки.

Предмет дослідження є вивчення методів машинного навчання, таких як кластеризація K-means та множинна регресійна модель, для аналізу та прогнозування різних факторів, що впливають на якість життя та рівень щастя людей, і визначення сприятливого середовища для розвитку людей в містах.

Розроблено та досліджено програмне забезпечення для визначенні факторів, які сприяють покращенню якості життя та розвитку людей в містах.

Застосування методів машинного навчання, таких як K-means та множинна регресійна модель, дозволяє автоматично аналізувати великі обсяги даних та визначати залежності між різними показниками якості життя та рівнем щастя населення. Отримані результати можуть бути корисні для прийняття рішень щодо політики та інфраструктури, що сприятимуть покращенню якості життя та збільшенню рівня щастя населення. Крім того, дослідження може відкрити нові можливості для подальшого дослідження в галузі соціальної політики та психології.

ПРОГНОЗУВАННЯ, МАШИННЕ НАВЧАННЯ, ЖИТТЯ ТА РОЗВИТОК ЛЮДЕЙ, ІНДЕКС ЩАСТЯ

ABSTRACT

The bachelor's thesis report: 68 pages, 21 figures, 2 appendices, 54 references.

The aim of the diploma project is to develop a module that utilizes machine learning methods to analyze various factors that influence the quality of life and happiness levels of people, and determine favorable conditions for creating an environment that promotes human development in cities.

The research object is the interconnections between various factors, such as economic, social, environmental, and infrastructural, that affect the quality of life and happiness levels of the population, as well as machine learning methods that enable the analysis and prediction of these interrelationships.

The research subject is the study of machine learning methods, such as K-means clustering and multiple regression model, for the analysis and prediction of different factors that influence the quality of life and happiness levels of people, and determining a favorable environment for human development in cities.

Software has been developed and investigated to determine the factors that contribute to improving the quality of life and human development in cities.

The application of machine learning methods, such as K-means clustering and multiple regression model, allows for automatic analysis of large volumes of data and identifying dependencies between different indicators of the quality of life and happiness levels of the population. The obtained results can be useful for decision-making regarding policies and infrastructure that will enhance the quality of life and increase the happiness levels of the population. Moreover, the research may open up new opportunities for further exploration in the field of social policy and psychology.

FORECASTING, MACHINE LEARNING, LIFE AND HUMAN DEVELOPMENT, HAPPINESS INDEX.

ТЕХНІЧНЕ ЗАВДАННЯ

1. НАЙМЕНУВАННЯ ТА ОБЛАСТЬ ЗАСТОСУВАННЯ

1.1 Модуль визначення сприятливого середовища для життя та розвитку людей.

1.2 Область застосування – аналіз міського середовища для поліпшення якості життя та розвитку населення..

2. ОСНОВА ДЛЯ РОЗРОБЛЕННЯ

Основою для розроблення є завдання на дипломний проект, затверджене кафедрою інформаційно-обчислювальних систем і управління факультету комп'ютерних інформаційних технологій Західноукраїнського національного університету.

3. ПРИЗНАЧЕННЯ РОЗРОБЛЕНОГО КОМПЛЕКСУ

Метою дипломного проекту є розробка модуля, який використовує методи машинного навчання для аналізу різних факторів, що впливають на якість життя та рівень щастя людей, і визначення сприятливих умов для створення середовища, яке сприяє розвитку людей в містах..

4. ДЖЕРЕЛА РОЗРОБЛЕННЯ

Джерелами даної розробки є матеріали навчальної і реферативної літератури, технічна документація, науково-дослідні статті, журнали, Інтернет.

5. ТЕХНІЧНІ ВИМОГИ

5.1 Основні функціональні вимоги до програмної системи:

– вибір даних (пакетних даних / даних у реальному часі) з різних архітектур великих даних;

– провести попередню обробку даних для інтелектуального аналізу

тексту.

- кластеризація країн за світовим індексом щастя
- передбачення значення індексу щастя

5.2 Вимоги до апаратних засобів:

– кластер запускається на фізичній машині Macbook Pro з 16 гігабайт оперативної пам'яті та процесором 2.3 GHz Intel Core i5.

5.3 Вимоги до програмних засобів:

- для розробки програмне забезпечення - Python 3.7;
- для створення графічного інтерфейсу користувача використано – tkinter, Adobe XD;
- для реалізації моделей навчання – фреймворк sklearn.

6. ПОРЯДОК КОНТРОЛЮ

6.1 Представлення дипломного проекту на попередній захист.

6.2 Представлення дипломного проекту на захист.

Завдання прийняв до виконання _____ Ю. В. Коткевич
(підпис) (прізвище та ініціали)

Керівник дипломного проекту _____ В.І. Дорош
(підпис) (прізвище та ініціали)

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ І ТЕРМІНІВ	2
ВСТУП	3
1 Аналіз методів та засобів визначення сприятливого середовища для життя та розвитку людей	6
1.1 Принципи та етапи визначення сприятливого середовища для життя та розвитку людей.....	6
1.2 Огляд і аналіз існуючих підходів визначення сприятливого середовища для життя та розвитку людей.....	10
1.3 Постановка задачі для розробки модуля визначення сприятливого середовища для життя та розвитку людей	15
2 Розробка алгоритмів для визначення сприятливого середовища для життя на основі машинного навчання.....	18
2.1 Алгоритм категоризації країн за індексом щастя.....	18
2.2 Алгоритм кластеризації даних за допомогою KMeans	21
2.3 Алгоритм прогнозування індексу щастя в Україні за допомогою лінійної множинної регресії.....	25
3 Програмна реалізація модуля визначення сприятливого середовища для життя та розвитку людей на основі індексу щастя.....	30
3.1 Опис та аналіз набору даних.....	30
3.2 Кластеризація країн за світовим індексом щастя	41
3.3 Передбачення значення індексу щастя.....	50
ВИСНОВКИ.....	57
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	60
ДОДАТОК А.....	65
ДОДАТОК Б	66

					<i>ДП.КН.8091494.076.ПЗ</i>			
Зм.	Арк.	№ докум.	Підпис	Дата				
<i>Розроб.</i>		Коткевич Ю.			Модуль визначення сприятливого середовища для життя та розвитку людей в місті	Літ.	Арк.	Аркушів
<i>Перевір.</i>		Дорош В.І.					8	76
<i>Н. Контр.</i>		Дорош В.І.				ЗУНУ.ФКІТ.КН-41		
<i>Затверд.</i>		Комар М.П.						

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

- EDA: Exploratory Data Analysis - метод аналізу даних, що використовується для розуміння особливостей даних, виявлення патернів, візуалізації та виявлення аномалій.
- ML: Machine Learning - машинне навчання, галузь штучного інтелекту, що вивчає розробку алгоритмів та моделей, що дозволяють комп'ютерам самостійно вчитися та покращувати свою продуктивність на основі даних.
- KMeans: K-середніх - алгоритм кластеризації, що використовується для групування об'єктів на основі їхньої подібності до кількох заданих кластерів.
- Множинна лінійна регресія (MLR): Multiple Linear Regression - метод регресійного аналізу, що використовується для моделювання залежності між незалежними змінними та залежною змінною, коли вплив на залежну змінну враховується декількома незалежними змінними.
- RMSE: Root Mean Squared Error - квадратний корінь з середньоквадратичної помилки, метрика, що використовується для вимірювання розбіжності між прогнозованими значеннями і спостережуваними значеннями в задачах прогнозування.
- Коефіцієнт детермінації (R²): Coefficient of Determination - статистична метрика, що вимірює відповідність моделі до вихідних даних шляхом визначення відсотка варіації залежної змінної, який пояснюється моделлю.

									Арк.
									2
Зм.	Арк.	№ докум.	Підпис	Дата	ДП.КН. 8091494.076.ПЗ				

ВСТУП

Актуальність теми. Розробка модуля визначення сприятливого середовища для життя та розвитку людей стає все більш актуальною через зростаючу урбанізацію та складність проблем, пов'язаних з міським середовищем. У сучасному світі індекс щастя вважається важливим показником, який допомагає визначити рівень комфорту та благополуччя населення в різних містах та країнах.

Цей індекс допомагає усвідомити, які фактори впливають на якість життя людей, а також може надати інформацію для прийняття рішень щодо політики, інфраструктури, освіти, медицини та інших аспектів міського розвитку. Це може сприяти покращенню умов життя та створенню сприятливого середовища для розвитку людей в містах.

При розробці модуля можуть бути застосовані методи машинного навчання, такі як лінійна регресія, кластеризація, аналіз кореляції та інші, які дозволяють отримати передбачення та категорії на основі існуючих даних. Це дозволяє виявити закономірності та тенденції, які впливають на індекс щастя, а також допомагає у розумінні особливостей різних місць, що може бути корисним для планування міського розвитку.

Окрім того, цей модуль може сприяти залученню громадськості та стейкхолдерів до процесу прийняття рішень, а також підвищенню обізнаності населення про важливість сприятливого середовища для життя та розвитку в містах.

В цілому, розробка модуля визначення сприятливого середовища а для життя та розвитку людей в місті має велике значення для суспільства. Це може допомогти місцевим та національним урядам, дослідникам, організаціям та іншим стейкхолдерам виявляти проблеми та розробляти стратегії щодо міського планування, інфраструктури, соціальних послуг та інших важливих аспектів розвитку міст.

Зокрема, з використанням такого модуля можна проводити аналіз не тільки за країнами, але й за містами, якщо доступні відповідні дані. Це дозволить

					<i>ДП.КН. 8091494.076.ПЗ</i>	Арк.
						3
Зм.	Арк.	№ докум.	Підпис	Дата		

отримати більш детальну картину щодо рівня життя та благополуччя населення в конкретних регіонах та містах, що допоможе краще розуміти специфіку місцевих умов та відповідно реагувати на виявлені проблеми.

Для країн, таких як Україна, розробка та застосування такого модуля може мати особливу важливість, оскільки це сприятиме покращенню якості життя та забезпеченню сталого розвитку міських середовищ. Застосування методів машинного навчання та аналізу даних у цій сфері може відкрити нові можливості для інформаційної підтримки прийняття рішень на всіх рівнях управління містами та державними структурами.

Метою дипломного проекту є розробка модуля, який використовує методи машинного навчання для аналізу різних факторів, що впливають на якість життя та рівень щастя людей, і визначення сприятливих умов для створення середовища, яке сприяє розвитку людей в містах. Тому для досягнення мети потрібно вирішити такі завдання:

1. оглянути принципи та етапи визначення сприятливого середовища для життя та розвитку людей.
2. проаналізувати існуючі підходи до визначення сприятливого середовища для життя та розвитку людей.
3. сформулювати задачі для розробки модуля визначення сприятливого середовища для життя та розвитку людей на основі індексу щастя.
4. розробити алгоритми на основі машинного навчання для визначення сприятливого середовища для життя та розвитку людей на основі індексу щастя.
5. реалізувати програмний модуль для визначення сприятливого середовища для життя та розвитку людей на основі індексу щастя.
6. перевірити ефективність та точність розробленого модулю на реальних даних.

Об'єкт дослідження є взаємозв'язки між різними факторами, такими як економічні, соціальні, екологічні та інфраструктурні, що впливають на якість

					<i>ДП.КН. 8091494.076.ПЗ</i>	Арк.
Зм.	Арк.	№ докум.	Підпис	Дата		4

життя та рівень щастя населення, а також методи машинного навчання, які дозволяють аналізувати та прогнозувати ці взаємозв'язки.

Предмет дослідження є вивчення методів машинного навчання, таких як кластеризація K-means та множинна регресійна модель, для аналізу та прогнозування різних факторів, що впливають на якість життя та рівень щастя людей, і визначення сприятливого середовища для розвитку людей в містах.

Методами дослідження в даному проекті є методи машинного навчання, зокрема, методи кластеризації та регресійної аналізу. Для кластеризації даних використовується алгоритм K-means, що дозволяє розділити дані на групи зі схожими характеристиками. Для передбачення індексу щастя за допомогою регресійної моделі використовується множинна лінійна регресія, що дозволяє знаходити залежності між декількома змінними та цільовим значенням. У дослідженні також використовуються методи аналізу кореляції та коваріації для вивчення залежностей між змінними.

Практична цінність дослідження полягає в тому, що воно може допомогти у визначенні факторів, які сприяють покращенню якості життя та розвитку людей в містах. Застосування методів машинного навчання, таких як K-means та множинна регресійна модель, дозволяє автоматично аналізувати великі обсяги даних та визначати залежності між різними показниками якості життя та рівнем щастя населення. Отримані результати можуть бути корисні для прийняття рішень щодо політики та інфраструктури, що сприятимуть покращенню якості життя та збільшенню рівня щастя населення. Крім того, дослідження може відкрити нові можливості для подальшого дослідження в галузі соціальної політики та психології..

Структура та обсяги роботи. Дипломний проект складається з вступу, трьох розділів, висновків, списку використаних джерел і додатків. У проекті міститься 21 рисунок. Список використаних джерел включає 54 пункти і займає 5 сторінок. Загальний обсяг проекту складає 76 сторінок, з яких 68 сторінок припадають на основний текст.

					<i>ДП.КН. 8091494.076.ПЗ</i>	Арк.
						5
Зм.	Арк.	№ докум.	Підпис	Дата		

1 АНАЛІЗ МЕТОДІВ ТА ЗАСОБІВ ВИЗНАЧЕННЯ СПРИЯТЛИВОГО СЕРЕДОВИЩА ДЛЯ ЖИТТЯ ТА РОЗВИТКУ ЛЮДЕЙ

1.1 Принципи та етапи визначення сприятливого середовища для життя та розвитку людей

Всілякі індекси щастя розраховуються різними міжнародними організаціями і науковими центрами, які використовують найрізноманітніші методики. Найвідомішою і найбільш широко прийнятою альтернативою ВВП став Індекс людського розвитку (ІЛР). Вона включає в себе як невід'ємну частину ВВП на душу населення, так і такі показники, як доступ до освіти, тривалість життя і ряд інших факторів.

ІЛР щорічно розраховується аналітиками Програми розвитку ООН (ПРООН) спільно з групою незалежних міжнародних експертів. Для визначення цього індексу поряд з аналітичними розробками використовується статистика національних інститутів і міжнародних організацій. Звіти про дослідження ПРООН публікуються з 1990 року.

При складанні рейтингу враховується багато факторів - зокрема, ситуація у сфері прав людини та громадянських свобод, можливість участі в суспільному житті, соціальний захист, ступінь територіальної та соціальної мобільності населення, показники рівня культурного розвитку, доступу до інформації, здоров'я, злочинності тощо.

При всіх своїх перевагах ІЛР також має ряд істотних недоліків: наприклад, спирається на національні середні показники, які не відображають метрики в розподілі товарів, не враховуються фактори навколишнього середовища і питання духовно-морального розвитку людини. Тому, незважаючи на комплексний підхід, з урахуванням багаточисників показників, що використовуються при розрахунку ІЛР, робляться спроби вдосконалити цю методологію і знайти оптимальний і універсальний метод вимірювання щастя.

					<i>ДП.КН. 8091494.076.ПЗ</i>	Арк.
						6
Зм.	Арк.	№ докум.	Підпис	Дата		

політиків, економістів, соціологів і психологів, також може здійснюватися за рахунок «зростання нещастя». Питання лінійної залежності щастя від добробуту залишається відкритим. Доведено, що суб'єктивне сприйняття задоволеності життям більш об'єктивно впливає на соціальну ситуацію, ніж на реальний стан справ. У зв'язку з цим в науковому співтоваристві все частіше виникає питання недостатності використання суто економічних показників як оцінки ефективності та результативності різних соціально-економічних заходів. Важливість рівня щастя населення, як альтернативи показнику валового внутрішнього продукту і критерію оцінки ефективності державної політики, визнають політики, соціологи, економісти і психологи всього світу. Поступово з'являється розуміння, що «опитування про щастя можуть служити важливим допоміжним інструментом формування державної політики», які проводять міждержавні порівняння, а також численні анкетні опитування населення, що дозволяють глибше вивчити цю проблему.

Сфери, які вивчаються в рамках різних методологій індексу, зазвичай включають: економічний розвиток, охорону навколишнього середовища, просування національної культури, ефективність державного управління, стійкість розвитку, безпеку, політичні права людей, стан екології, доступність послуг соціальних установ, витрати на дослідження, освіту, культуру і спорт, здатність людей брати участь у суспільному житті, ступінь територіального і соціального життя. мобільності населення і т.д.

Аналіз досвіду практичного застосування індексних методів оцінки рівня щастя дозволяє прийти до висновку про істотні недоліки такого підходу. Перш за все, до них відносяться наступні:

- Інтегральний показник не відображає проблем в напрямках приватних індексів, тому жоден з індексів не може бути одноразово використаний для оцінки позиції країни в світі.
- При цьому не враховуються особливості країни, культурні та етнічні відмінності, історичний досвід і поточна ситуація в країні.

					<i>ДП.КН. 8091494.076.ПЗ</i>	Арк.
						9
Зм.	Арк.	№ докум.	Підпис	Дата		

прогнозування національного щастя, важливо застосовувати методику з умінням справлятися зі складними особливостями.

Таким чином, використання опорної векторної машини (SVM) у дослідженні сприятливого середовища для життя та розвитку людей в місті може допомогти виявити релевантні фактори та залежності між ними. Це дасть можливість краще зрозуміти, які аспекти суспільства та інфраструктури впливають на щастя громадян та можуть бути використані для підвищення рівня життя.

Метод SVM може бути особливо корисним при аналізі даних з обмеженою кількістю зразків та високою розмірністю ознак, як це часто буває у випадку досліджень щастя. Завдяки своїй здатності робити гнучкі рішення у високовимірному просторі ознак, SVM може допомогти виявити складні закономірності та взаємозв'язки між факторами, які впливають на щастя людей в містах.

Методи кластеризації також є важливими інструментами машинного навчання для виявлення структури в наборах даних та можуть бути застосовані для визначення сприятливого середовища для життя та розвитку людей. Ці методи допомагають групувати спостереження на основі їхньої подібності, що може виявити закономірності та відносини між різними аспектами життя та щастя.

1. K-means: K-means є одним з найвідоміших методів кластеризації та широко застосовується в різних наукових дослідженнях [14]. У роботі [15] автори застосували метод k-means для кластеризації країн на основі індексу щастя та різних соціально-економічних показників, що дозволило виявити спільні характеристики та відмінності між ними.
2. Ієрархічна кластеризація: Ієрархічні методи кластеризації побудовують деревоподібну структуру кластерів та можуть допомогти у виявленні глибинних відносин між спостереженнями. У статті [16] автори використали ієрархічну кластеризацію для аналізу показників якості життя

										ДП.КН. 8091494.076.ПЗ	Арк.
											12
Зм.	Арк.	№ докум.	Підпис	Дата							

на різних рівнях суспільства, що допомогло зрозуміти різні аспекти, які впливають на щастя людей.

3. DBSCAN: Метод DBSCAN (Density-Based Spatial Clustering of Applications with Noise) є потужним алгоритмом кластеризації, який може виявляти кластери різної форми та розміру, а також виявляти шум у даних. У роботі [17] автори застосували DBSCAN для аналізу соціально-економічних даних, що впливають на якість життя та щастя людей у різних регіонах. Це дослідження допомогло виявити групи регіонів із схожими характеристиками та відмінностями, що можуть бути використані для розробки ефективних стратегій розвитку та політики благополуччя.
4. Агломеративна кластеризація: Цей метод кластеризації об'єднує найближчі кластери в ієрархічному порядку. В статті [18] автори використали агломеративну кластеризацію для аналізу даних про житлові умови та екологічні фактори, що впливають на якість життя в різних містах. Результати допомогли розробити стратегії покращення умов життя для мешканців.
5. Спектральна кластеризація: Спектральна кластеризація використовує графи та спектральні властивості для розподілу даних на кластери. У дослідженні [19] автори використали спектральну кластеризацію для виявлення схожих місць проживання на основі показників щастя та соціально-економічних характеристик, що допомогло виявити потреби місцевого населення та сприяло розвитку сприятливого середовища для життя.

Застосування різних методів кластеризації в наукових дослідженнях підтверджує важливість цих методів для аналізу сприятливого середовища для життя та розвитку людей. Вони допомагають виявити відносини та закономірності між різними аспектами щастя та можуть бути використані для покращення умов життя та розвитку людей у містах.

У статті [20] дослідники проводять просторово-економетричний аналіз щастя канадських міст, використовуючи дані з опитувань про якість життя.

					<i>ДП.КН. 8091494.076.ПЗ</i>	Арк.
						13
Зм.	Арк.	№ докум.	Підпис	Дата		

Автори застосовують просторові авторегресійні моделі (SAR) для визначення впливу різних факторів на щастя мешканців міст. Результати показують, що доходи, безробіття, національність та інші демографічні характеристики мають важливе значення для рівня щастя міських мешканців.

Стаття [21] вивчає ключові фактори успіху в онлайн-покупках, що можуть стосуватися якості життя в містах, оскільки люди все частіше звертаються до онлайн-покупок для задоволення своїх потреб. Дослідники використовують гібридний підхід, що поєднує машинне навчання та статистичні методи, включаючи регресійний аналіз, для визначення ключових факторів, що впливають на успішність онлайн-покупок. Вони виявляють, що зручність, безпека та якість обслуговування є важливими чинниками для успішного онлайн-покупця.

Обидві статті використовують регресійні моделі для дослідження різних аспектів сприятливого середовища для життя та розвитку людей. З одного боку, Ráez і Scott [20] фокусуються на просторовому аналізі щастя в містах, досліджуючи вплив економічних та демографічних факторів на рівень задоволення життям. Застосування просторових авторегресійних моделей дозволяє враховувати взаємозв'язок між містами і сприяє кращому розумінню факторів, що впливають на щастя на міському рівні.

З іншого боку, Huang і співавтори [21] вивчають питання, пов'язані з якістю життя в містах, з точки зору онлайн-покупок. Вони використовують гібридний підхід, який поєднує регресійний аналіз і машинне навчання, для визначення ключових факторів успішності в онлайн-покупках, що можуть відображати зручність, безпеку та якість обслуговування, які сприяють покращенню якості життя в містах.

Ці дві статті показують, як регресійні моделі можуть бути використані для дослідження різних аспектів сприятливого середовища для життя та розвитку людей, включаючи щастя та зручність, які є важливими складовими якості життя. Використання регресійних моделей і машинного навчання може допомогти дослідникам краще розуміти та прогнозувати фактори, що впливають

					<i>ДП.КН. 8091494.076.ПЗ</i>	Арк.
Зм.	Арк.	№ докум.	Підпис	Дата		14

на сприятливе середовище для життя та розвитку людей, та надавати рекомендації для політиків та розробників урбаністичних проектів.

Підсумовуючи, дослідження в області визначення сприятливого середовища для життя та розвитку людей активно використовують методи машинного навчання, зокрема алгоритми кластеризації, такі як K-means, та множинні регресійні моделі. Застосування K-means дозволяє виявити приховані структури в даних і групувати об'єкти зі схожими характеристиками, що може сприяти покращенню планування міського середовища. Множинні регресійні моделі допомагають вивчати взаємозв'язки між різними факторами та прогнозувати їх вплив на якість життя та рівень щастя. Застосування цих методів дозволяє отримати цінні висновки щодо факторів, які сприяють створенню сприятливого середовища для життя та розвитку людей, та розробляти ефективні стратегії в урбаністичному плануванні та соціальній політиці.

1.3 Постановка задачі для розробки модуля визначення сприятливого середовища для життя та розвитку людей

Розробка модуля визначення сприятливого середовища для життя та розвитку людей на основі індексу щастя є дуже актуальною у зв'язку зі складністю проблем міського середовища. Індекс щастя вважається важливим показником, який допомагає визначити рівень комфорту та благополуччя населення в різних містах та країнах. Модуль може використовувати методи машинного навчання, такі як лінійна регресія, кластеризація та інші, які дозволяють отримати передбачення та категорії на основі існуючих даних. Це допомагає виявити закономірності та тенденції, які впливають на індекс щастя та можуть бути корисними для планування міського розвитку. Модуль може залучати громадськість та стейкхолдерів до процесу прийняття рішень та підвищувати обізнаність населення про важливість сприятливого середовища для життя та розвитку в містах. Для України розробка такого модуля є особливо

					<i>ДП.КН. 8091494.076.ПЗ</i>	Арк.
Зм.	Арк.	№ докум.	Підпис	Дата		15

важливою, оскільки це сприятиме покращенню якості життя та забезпеченню сталого розвитку міських середовищ.

Розробка модуля визначення сприятливого середовища для життя та розвитку людей є дуже актуальною задачею в сучасному світі. Застосування методів машинного навчання та аналізу даних у цій сфері може відкрити нові можливості для інформаційної підтримки прийняття рішень на всіх рівнях управління містами та державними структурами. Індекс щастя допомагає визначити рівень комфорту та благополуччя населення в різних містах та країнах, а це може сприяти покращенню умов життя та створенню сприятливого середовища для розвитку людей в містах. Розробка модуля може також сприяти залученню громадськості та стейкхолдерів до процесу прийняття рішень, а також підвищенню обізнаності населення про важливість сприятливого середовища для життя та розвитку в містах. Отже, цей проект може мати велике значення для розвитку міст та забезпечення сталого розвитку в цілому.

Метою дослідження є розробка модуля, який використовує методи машинного навчання для аналізу різних факторів, що впливають на якість життя та рівень щастя людей, і визначення сприятливих умов для створення середовища, яке сприяє розвитку людей в містах. Тому для досягнення мети потрібно вирішити такі завдання:

1. оглянути принципи та етапи визначення сприятливого середовища для життя та розвитку людей.
2. проаналізувати існуючі підходи до визначення сприятливого середовища для життя та розвитку людей.
3. сформулювати задачі для розробки модуля визначення сприятливого середовища для життя та розвитку людей на основі індексу щастя.
4. розробити алгоритми на основі машинного навчання для визначення сприятливого середовища для життя та розвитку людей на основі індексу щастя.
5. реалізувати програмний модуль для визначення сприятливого середовища для життя та розвитку людей на основі індексу щастя.

					<i>ДП.КН. 8091494.076.ПЗ</i>	Арк.
						16
Зм.	Арк.	№ докум.	Підпис	Дата		

6. перевірити ефективність та точність розробленого модулю на реальних даних.

Завдання 1-3 розглянуто в параграфах 1.1 - 1.3 відповідно, а виконання завдань 5-6 представлені у параграфах 2.1-3.3.

					<i>ДП.КН. 8091494.076.ПЗ</i>	Арк.
Зм.	Арк.	№ докум.	Підпис	Дата		17

2 РОЗРОБКА АЛГОРИТМІВ ДЛЯ ВИЗНАЧЕННЯ СПРИЯТЛИВОГО СЕРЕДОВИЩА ДЛЯ ЖИТТЯ НА ОСНОВІ МАШИННОГО НАВЧАННЯ

2.1 Алгоритм категоризації країн за індексом щастя

У візуалізації даних немає зайвих декоративних елементів, вона спрямована на подання великих обсягів інформації з урахуванням можливих взаємозв'язків. У Data Science візуалізація даних використовується не тільки для наочного подання результатів у вигляді зрозумілих графіків, але й як метод швидкого прототипування. В рамках розвідувального аналізу даних аналітики та Data Scientist використовують безліч візуальних уявлень для виявлення прихованих взаємозв'язків і залежностей, максимального занурення в дані, вибору найбільш важливих змінних, виявлення відхилень та аномалій, перевірки основних гіпотез та розробки початкових моделей.

EDA [22, 23] є складовою підготовки даних для ML-моделювання, яка включає генерацію ознак після вибірки та очищення датасета. Водночас, EDA дозволяє Data Scientist'у (Рис.2.1) переконатися в правильності інтерпретації результатів та їх застосовності до бізнес-контексту. Для бізнес-користувачів, EDA може служити засобом перевірки правильності припущень та питань. EDA також є важливим інструментом валідації даних, оцінки їх відповідності бізнес-цілям. При роботі з великими обсягами даних, EDA є особливо корисною, оскільки датасет може містити інформацію з різних джерел з різними рівнями точності та деталізації. На практиці, EDA може призвести до цікавих бізнес-інсайтів, наприклад, виявлення залежності між сумою чека та часом доби або кореляції між числом відвідувачів та погодними умовами.

EDA в Data Science базується на статистиці та теорії ймовірностей, що включає імовірнісні розподіли змінних, кореляційні матриці, факторний та дискримінантний аналіз, та багатовимірне шкалювання. Для проведення EDA використовують спеціалізовані математичні програми, такі як SAS, Matlab, KNIME, Weka, Orange, системи типу RStudio, оригінальні скрипти на Python, або вбудовані формули табличних редакторів, таких як Excel і Google Sheets. Для

					ДП.КН. 8091494.076.ПЗ	Арк.
Зм.	Арк.	№ докум.	Підпис	Дата		18

- бульбашкову діаграму, яка відображає взаємозв'язки між змінними за допомогою їх розташування і пропорцій.

Правильно обраний вид діаграми для візуалізації даних відповідає таким критеріям [1]:

- стислість - можливість одночасно відобразити багато різнотипних даних;
- відносність і близькість - здатність демонструвати кластери, відносні розміри груп, їх схожість і відмінність, що випадають значення;
- концентрацію і контекст - можливість легко і оперативно взаємодіяти з обраним об'єктом шляхом його інтерактивного перегляду (відображення структури і зв'язків);
- масштабованість - можливість легко і швидко змінювати розміри уявлення;
- зручність користувача за рахунок максимальної наочності надання та підтримка інтуїтивних дій по виявленню закономірностей.

Отже, візуалізація потрібна не тільки для наочного подання результатів, але і для розробки попередніх гіпотез, а також валідації вихідних даних. EDA або розвідувальний аналіз даних - важливий етап підготовки датасета до ML-моделювання та іншим технікам Data Mining. Вибір графіка для візуалізації залежить від мети (порівняння змінних, виявлення взаємозв'язків, уявлення складу і структури або демонстрація статистичного розподілу) і аналізованих категорій (багатовимірний аналіз, тимчасові ряди або кореляція декількох показників).

В процесі EDA (англ. Exploratory Data Analysis) важливо розуміти, які інструменти та методи можна використовувати для аналізу даних. Одним з найпоширеніших методів є візуалізація, яка дозволяє представити дані у зручному та легкозрозумілому вигляді. Одним із видів візуалізації є категоризація, що дозволяє поділити дані на певні групи за заданим критерієм. В даному випадку, розглянемо алгоритм категоризації країн за індексом щастя. Даний алгоритм використовує квартилі для визначення категорій щастя країн.

					<i>ДП.КН. 8091494.076.ПЗ</i>	Арк.
						20
Зм.	Арк.	№ докум.	Підпис	Дата		

Це дозволяє більш детально проаналізувати дані та зробити висновки щодо рівня щастя у різних країнах.

Алгоритм категоризації країн за індексом щастя (див. дод. А):

1. Визначається перший кватиль (Q1) Індексу щастя.
2. Розраховується середнє значення (Q2) Індексу щастя.
3. Визначається третій кватиль (Q3) Індексу щастя.
4. За допомогою циклу **for**, програма перебирає всі рядки в DataFrame **data**.

Для кожної країни визначається категорія щастя на основі значення Індексу щастя:

- Якщо Індекс щастя менше або дорівнює Q1, країні присвоюється категорія 'Sad'.
- Якщо Індекс щастя більше Q1 та менше Q2, країні присвоюється категорія 'Moderately Sad'.
- Якщо Індекс щастя більше Q2 та менше Q3, країні присвоюється категорія 'Happy'.
- Якщо Індекс щастя більше або дорівнює Q3, країні присвоюється категорія 'Very Happy'.

Алгоритм категоризації країн за індексом щастя дозволяє більш детально проаналізувати дані та зробити висновки про рівень щастя країн в залежності від значення їх індексу щастя. В результаті застосування алгоритму, кожній країні присвоюється категорія щастя, що може бути використана для подальшого дослідження і аналізу. Наприклад, зіставлення категорій щастя з іншими показниками, такими як ВВП, освіта, медицина та інші, може допомогти визначити фактори, що впливають на рівень щастя в країнах.

2.2 Алгоритм кластеризації даних за допомогою KMeans

Кластеризація є процесом, під час якого елементи набору даних розділяються на групи на основі їх подібності. Ці групи називають кластерами.

					<i>ДП.КН. 8091494.076.ПЗ</i>	Арк.
Зм.	Арк.	№ докум.	Підпис	Дата		21

Один з елементів, що підлягає кластеризації, називається об'єктом. Для кожного об'єкта створюється вектор характеристик, який описує його.

$$X = \{x_1, x_2, \dots, x_m\}.$$

Об'єкти в алгоритмах кластеризації описуються векторами характеристик, які складаються з окремих компонент x_i . Ці компоненти можуть бути кількісними (наприклад, координати точки) або якісними (наприклад, кольори). Розмірність простору характеристик визначається кількістю компонент, а безліч усіх можливих векторів характеристик позначається як простір характеристик.

$$A = \{X_1, X_2, \dots, X_n\}.$$

Для правильної роботи алгоритмів кластеризації необхідно нормалізувати характеристики, тобто привести їх до одного масштабу. Кластер - це група близьких один до одного об'єктів з множини A . Відстань $\rho(x_i, x_j)$ між об'єктами x_i і x_j визначається за допомогою обраної метрики в просторі характеристик. Для прискорення процесу кластеризації можна зменшити розмірність простору характеристичних векторів, виділивши найбільш важливі властивості об'єктів (наприклад, методом головних компонент). Вибір метрики залежить від простору, в якому знаходяться об'єкти, і неявних характеристик кластерів.

Евклідова відстань

$$\rho(x_i, y_i) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}. \quad (1)$$

Квадрат евклідової відстані - це метрика, яка використовується для вимірювання відстані між двома точками в n -вимірному просторі. Вона є однією з найбільш відомих і часто застосовуваних метрик. Квадрат евклідової відстані

					ДП.КН. 8091494.076.ПЗ	Арк.
						22
Зм.	Арк.	№ докум.	Підпис	Дата		

між двома точками x_i і x_j обчислюється як сума квадратів різниць між відповідними координатами цих точок:

$$\rho(x_i, y_i) = \sum_{i=1}^n (x_i - y_i)^2. \quad (2)$$

де $x_{i,k}$ і $x_{j,k}$ - k -та координата векторів x_i і x_j .

При використанні квадрата евклідової відстані в якості метрики кластеризації більша вага надається ближнім об'єктам, тобто тим, які розташовані ближче до даної точки. Це може бути корисним у випадках, коли ближні об'єкти більш схожі між собою, ніж віддалені. Однак, якщо відстані між об'єктами великі, то квадрат евклідової відстані може призвести до недооцінки віддалених об'єктів і перевищення ваги ближніх.

Манхетенська відстань або відстань міських кварталів є однією з метрик в n -вимірному просторі, яка визначається як сума абсолютних різниць між відповідними координатами двох точок. Назва метрики виникла через те, що в міських кварталах відстані між точками вимірюються вздовж прямих ліній, які паралельні вулицям і перпендикулярні до них. Відстань між двома точками (x_1, y_1) і (x_2, y_2) в n -вимірному просторі буде визначатися формулою:

$$\rho(x_i, y_i) = \sum_{i=1}^n |x_i - y_i|. \quad (3)$$

де $|x_1 - x_2|$ та $|y_1 - y_2|$ - абсолютні різниці між відповідними координатами точок. Ця метрика широко використовується в комп'ютерному зорі, маршрутизації міського транспорту, плануванні містобудівних проектів та інших областях.

Метод k-середніх (KMeans) [24, 25] - це ітеративний алгоритм кластеризації даних, який визначає k кластерів шляхом групування об'єктів в k груп або кластерів на основі подібності між ними. Підхід працює наступним чином:

1. Кількість кластерів k визначається користувачем.
2. Випадковим чином вибираються k об'єктів, які стануть центрами кластерів.
3. Кожен об'єкт присвоюється до найближчого кластеру на основі відстані між об'єктом та центром кластеру.
4. Обчислюється середнє арифметичне всіх об'єктів у кожному кластері, і вибирається новий центр кластеру.
5. Кожен об'єкт знову присвоюється до найближчого кластеру на основі нового центру кластеру.
6. Процес продовжується до тих пір, поки змінність між попереднім та поточним розташуванням центрів кластерів буде менше певного порогового значення або до досягнення максимальної кількості ітерацій.

Кінцевим результатом роботи методу k-середніх є k кластерів, які містять об'єкти з подібними характеристиками. Цей алгоритм є швидким і ефективним в застосуванні, але він також має деякі недоліки, зокрема залежність від початкового вибору центрів кластерів та вразливість до викидів.

Далі наведено псевдокод для кластеризації індексу щастя за допомогою методу k-середніх (див.дод.Б.). Де детально пояснюється кожен крок алгоритму, від відображення даних та їх кодування до виведення інформації про датафрейм та візуалізації матриць кореляції та коваріації. Крім того, докладно розглядається використання методу ліктя для визначення оптимальної кількості кластерів та навчання моделі k-середніх з визначеною кількістю кластерів. Це буде корисно для тих, хто бажає кластеризувати дані та досліджувати залежності між їхніми атрибутами за допомогою кореляційної та коваріаційної матриць.

Псевдокод для кластеризації індексу щастя (див.дод.Б)

1. Відобразити перші рядки датафрейму "data" та його індекс.

					<i>ДП.КН. 8091494.076.ПЗ</i>	Арк.
						24
Зм.	Арк.	№ докум.	Підпис	Дата		

які слід виключити. Значення R2 може коливатися від 0 до 1, де 0 вказує на те, що результат не може бути передбачений жодною з незалежних змінних, а 1 вказує на те, що результат можна передбачити без помилок за допомогою незалежних змінних.

RMSE (Root Mean Squared Error) - це математична метрика для вимірювання середньої квадратичної помилки моделі прогнозування, що використовується для порівняння якості різних моделей прогнозування. Вона обчислюється як квадратний корінь з середнього квадрата різниці між прогнозованими значеннями та фактичними значеннями.

Формула для RMSE:

$$RMSE = \sqrt{\left(\frac{1}{n} * \sum ((y_{pred} - y_{true})^2)\right)}$$

де:

- y_{pred} - вектор прогнозованих значень;
- y_{true} - вектор фактичних значень;
- n - кількість спостережень.

RMSE дозволяє визначити, наскільки далеко відхилення прогнозованих значень від фактичних значень. Чим менше значення RMSE, тим краще прогнозування моделі. Однак, RMSE має одиницю виміру, яка залежить від одиниці виміру вхідних даних, тому порівнювати RMSE між різними моделями можна лише в рамках одних і тих же вхідних даних.

Далі представлено псевдокод, що описує алгоритм для прогнозування індексу щастя в Україні за допомогою лінійної регресії. Ось словесний опис алгоритму, згідно з псевдокодом (Рис.2.3):

1. Імпортувати необхідні бібліотеки (`pandas`, `LinearRegression`, `mean_squared_error`).

									Арк.
									27
Зм.	Арк.	№ докум.	Підпис	Дата					

3 ПРОГРАМНА РЕАЛІЗАЦІЯ МОДУЛЯ ВИЗНАЧЕННЯ СПРИЯТЛИВОГО СЕРЕДОВИЩА ДЛЯ ЖИТТЯ ТА РОЗВИТКУ ЛЮДЕЙ НА ОСНОВІ ІНДЕКСУ ЩАСТЯ

3.1 Опис та аналіз набору даних

Для реалізації запропонованого підходу використано всесвітній звіт [28] про щастя є знаковим оглядом стану глобального щастя. Звіт продовжує здобувати глобальне визнання, оскільки уряди, організації та громадянське суспільство все частіше використовують індикатори щастя для обґрунтування своїх політичних рішень. Провідні експерти в різних галузях – економіка, психологія, аналіз опитувань, національна статистика, охорона здоров'я, державна політика тощо – описують, як вимірювання добробуту можуть бути ефективно використані для оцінки прогресу країн. Звіти розглядають стан щастя в сучасному світі і показують, як нова наука про щастя пояснює особисті та національні відмінності в щасті.

Показники щастя та рейтинги використовують дані Всесвітнього опитування Gallup . Колонки, що слідує за показником щастя, оцінюють, наскільки кожен із шести факторів – економічне виробництво, соціальна підтримка, очікувана тривалість життя, свобода, відсутність корупції та щедрість – сприяють підвищенню оцінок життя в кожній країні, ніж у Антиутопії, гіпотетичній країні, значення якої дорівнюють найнижчим середнім національним показникам у світі для кожного з шести факторів. Вони не впливають на загальний бал для кожної країни, але пояснюють, чому деякі країни займають вищі позиції, ніж інші.

За даними DataFrame є 149 записів (рядків) та 20 стовпців з різними характеристиками. Ось короткий огляд датасету:

1. Country_Name: назва країни (object)
2. Regional_Indicator: регіональний індикатор (object)
3. Ladder_Score: показник сходінки (float64)

					ДП.КН.07122/17.075.ПЗ	Арк.
						30
Зм.	Арк.	№ докум.	Підпис	Дата		

регіонах, а також важливість соціальної підтримки для загального добробуту країни. Використання логарифмічної шкали для осі X може допомогти підкреслити зв'язок між двома змінними для країн з низькою очікуваною тривалістю здорового життя.

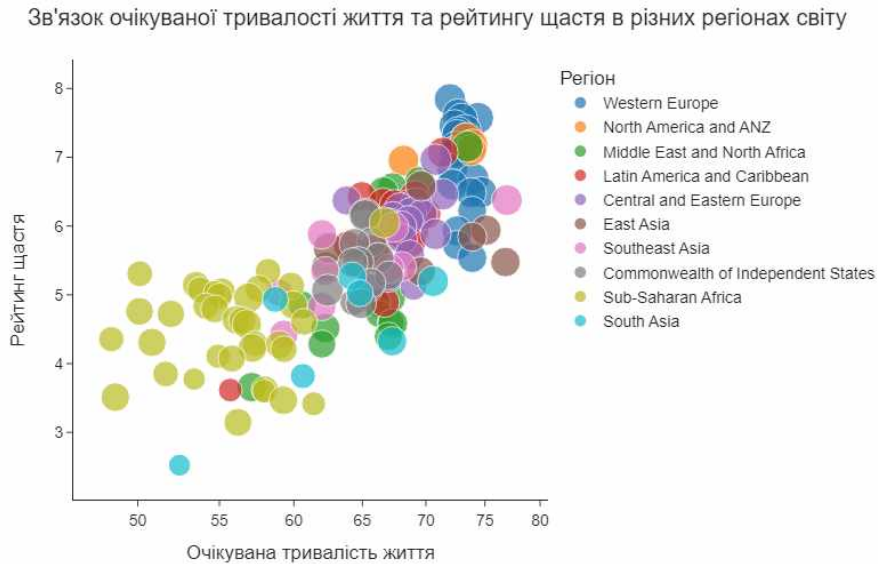


Рисунок 3.4 - Зв'язок очікуваної тривалості життя та рейтингу щастя в різних регіонах світу

Аналіз матриці кореляції (Рис. 3.5) показав високу пряму залежність між індексом щастя та ВВП на душу населення (0.79) та соціальною підтримкою (0.76). Помірна залежність існує між індексом щастя та свободою вибору життя (0.61). Слабкіші зв'язки спостерігаються між іншими параметрами.

Існує позитивна (Рис.3.6) залежність між Логарифмованим ВВП на душу населення та Індексом щастя. Це означає, що країни з вищим ВВП на душу населення зазвичай мають вищий індекс щастя. Різні кольори точок на діаграмі представляють різні регіональні показники. За допомогою кольорів можна відслідкувати тенденції в межах різних регіонів світу. Деякі регіони мають вищий рівень щастя та ВВП, у той час як інші - нижчий. Загалом, діаграма підтверджує результати аналізу кореляції, які вказують на сильну позитивну залежність між Індексом щастя та Логарифмованим ВВП на душу населення.

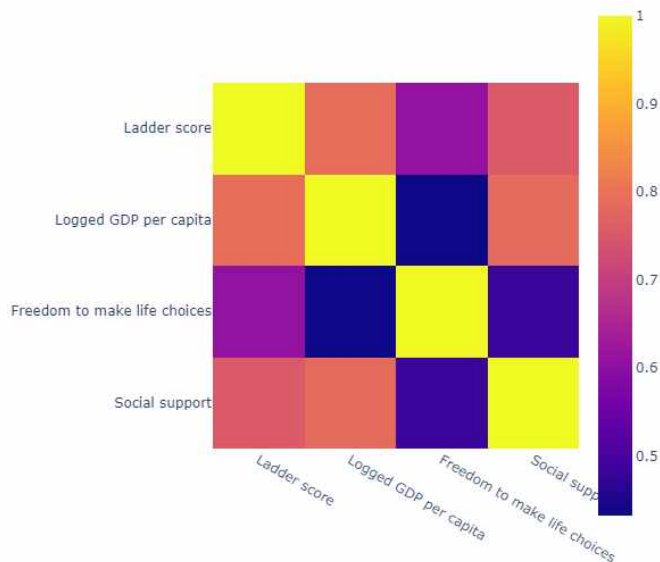


Рисунок 3.5 - Кореляційна матриця

Залежність індексу щастя від ВВП на душу населення

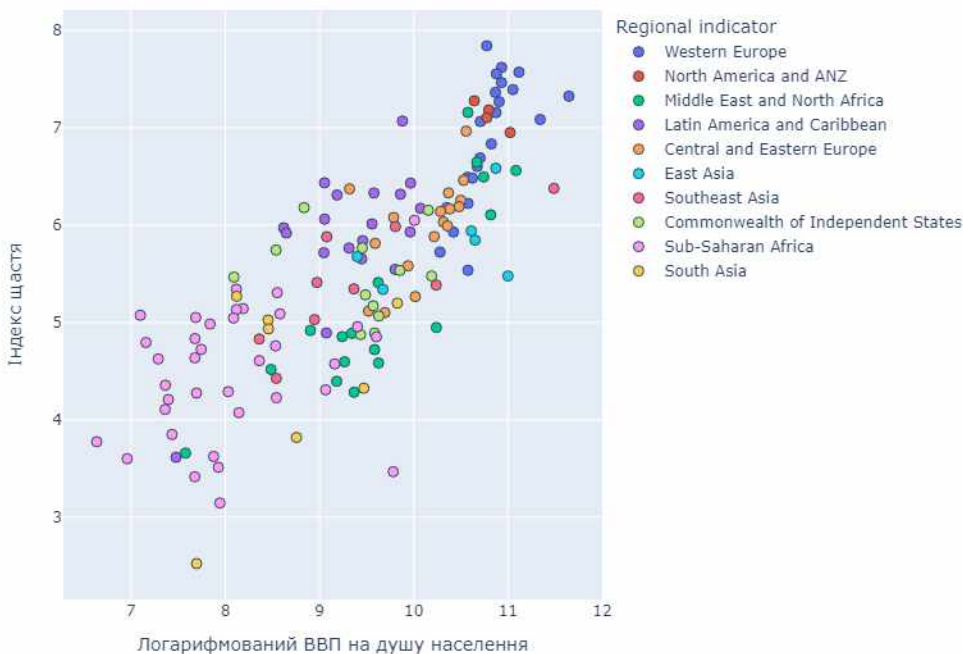


Рисунок 3.6 - Залежність індексу щастя від ВВП на душу населення

Далі розрахуємо перший (Q1), другий (Q2) та третій (Q3) квантили Індексу щастя для набору даних `data_`. На основі цих значень, кожній країні присвоюється певна категорія щастя - 'Sad', 'Moderately Sad', 'Happy' та 'Very Happy' - у новому стовпці 'label' в DataFrame `data`.

```

Q1 = np.percentile(data_, 25, interpolation = 'midpoint')
Q2 = np.mean(data_)
Q3 = np.percentile(data_, 75, interpolation = 'midpoint')
for i in range(len(data_)):
    if data['Ladder score'][i]<=Q1:
        data['label'][i] = 'Sad'
    if data['Ladder score'][i]>Q1 and data['Ladder score'][i]<Q2:
        data['label'][i] = 'Moderately Sad'
    if data['Ladder score'][i]>Q2 and data['Ladder score'][i]<Q3:
        data['label'][i] = 'Happy'
    if data['Ladder score'][i]>=Q3:
        data['label'][i] = 'Very Happy'

```

Результатом виконання коду є DataFrame data, який тепер містить новий стовпець 'label' з категорією щастя для кожної країни. Це допомагає краще розуміти та аналізувати розподіл щастя серед різних країн на основі Індексу щастя.

Далі виведемо стовпчасту діаграму (Рис. 3.7.), що відображає розподіл країн за категоріями щастя ('Sad', 'Moderately Sad', 'Happy' та 'Very Happy'), відсортованих за значеннями Індексу щастя. Кожна категорія щастя має різний колір відповідно до регіонального індикатора. Використовуючи параметр hover_data, діаграма показує назву країни та її Індекс щастя при наведенні курсора на стовпчик. Результати цієї діаграми допомагають проаналізувати розподіл країн за рівнями щастя відповідно до регіональних індикаторів. Це дає змогу порівняти різні регіони та країни за їх категоріями щастя та визначити, які регіони мають вищий або нижчий рівень щастя в цілому.

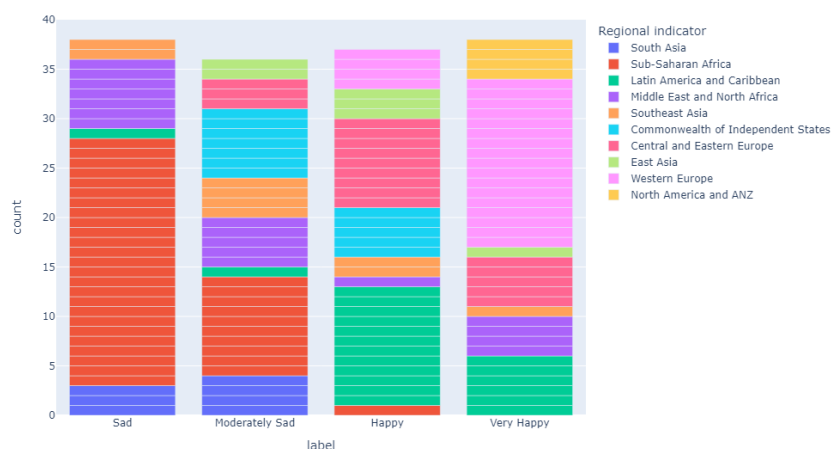


Рисунок 3.7 - Розподіл країн за категоріями щастя з урахуванням регіональних індикаторів

На діаграмі (Рис. 3.8) представлено залежність між Індексом щастя та свободою вибору життя для кожної країни. Діаграма розділена на чотири колонки відповідно до категорій щастя: 'Sad', 'Moderately Sad', 'Happy' та 'Very Happy'.

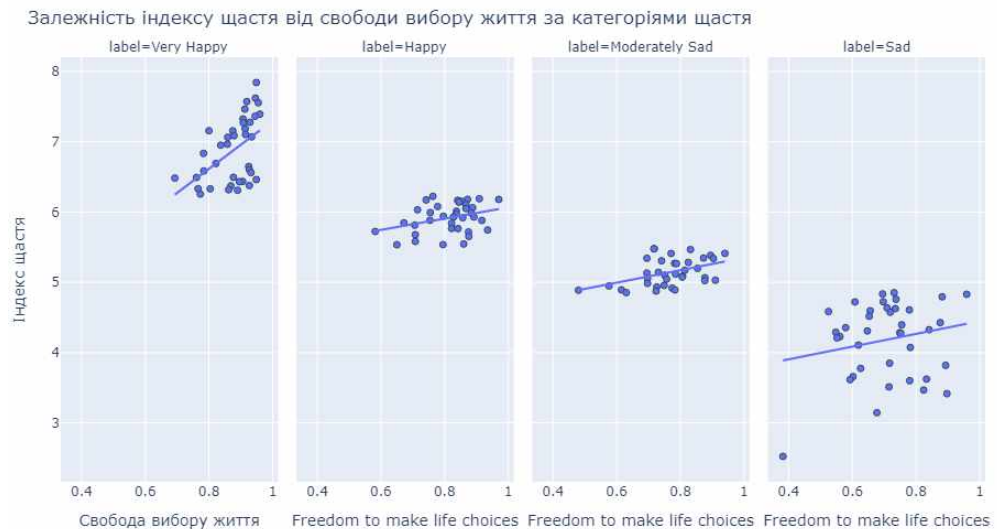


Рисунок 3.8 - Залежність індексу щастя від свободи вибору життя за категоріями щастя

Кожна точка на діаграмі представляє окрему країну, а її розташування відображає значення Індексу щастя та свободи вибору життя. Для кожної категорії щастя також додано лінію тренду за допомогою методу найменших квадратів (OLS). Результати:

1. В категорії 'Sad' можна спостерігати певну позитивну кореляцію між свободою вибору життя та Індексом щастя, що відображається на лінії тренду.
2. Для категорії 'Moderately Sad' кореляція між свободою вибору життя та Індексом щастя також є позитивною, але слабніше вираженою порівняно з категорією 'Sad'.
3. В категорії 'Happy' спостерігається позитивна кореляція між свободою вибору життя та Індексом щастя, хоча розкид точок є ширшим, ніж у попередніх категоріях.

4. Нарешті, для категорії 'Very Happy' кореляція між свободою вибору життя та Індексом щастя є слабкою та позитивною, а точки розташовані відносно близько один до одного.

Загалом, ця діаграма підтверджує наявність позитивної залежності між свободою вибору життя та Індексом щастя для різних категорій щастя, хоча сила залежності може варіюватися.

І на останок виведено хороплет-діаграмі (Рис. 3.9), яка показує розподіл Індексу щастя по всьому світу. Кожна країна забарвлена відповідно до свого значення Індексу щастя.

Індекс щастя по всьому світу

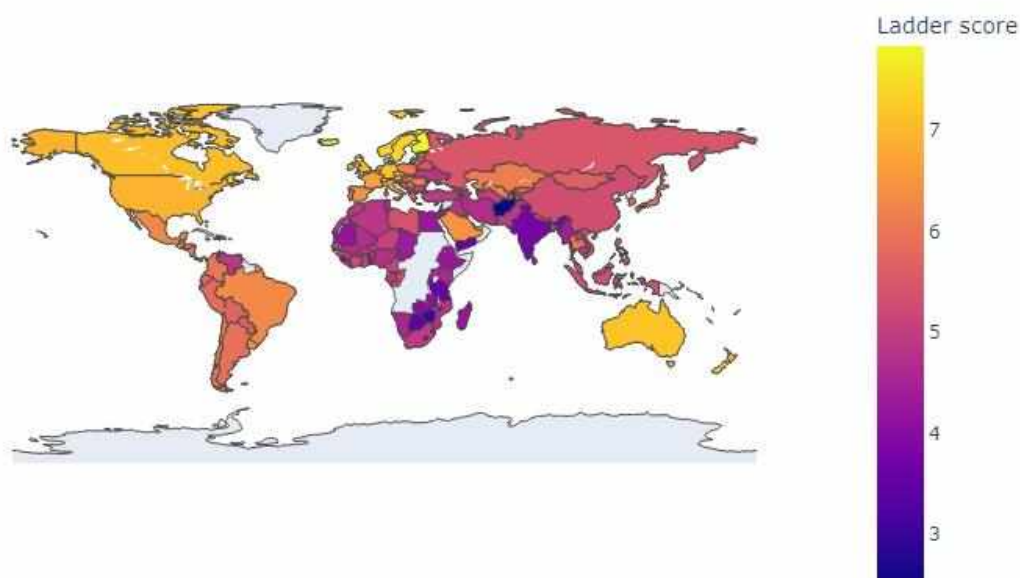


Рисунок 3.9 - Індекс щастя по всьому світу

Чим вище значення індексу, тим світліше кольорове позначення країни на карті. Це дозволяє візуалізувати рівень щастя у різних країнах та сприяє порівнянню між ними. Загальні спостереження з карті включають:

1. Країни Північної Європи (скандинавські країни) мають високі значення Індексу щастя та світлі кольори на карті.
2. Західна Європа, Північна Америка та Австралія також мають відносно високі рівні щастя.

3. Африканські країни, особливо ті, що розташовані на південь від Сахари, мають низькі значення Індексу щастя та темні кольори на карті.

4. Центральна Азія та деякі країни Близького Сходу також відрізняються низькими рівнями щастя.

Ця діаграма допомагає отримати загальне уявлення про географічний розподіл щастя у світі та може спонукати до подальших досліджень та аналізу причин такого розподілу.

Аналіз результатів світового індексу щастя показує пряму залежність між ВВП на душу населення, соціальною підтримкою та свободою вибору життя. Високі значення кореляції вказують на тісний зв'язок між економічним добробутом та щастям. Візуалізації допомагають спостерігати географічний розподіл щастя та порівняти рівень щастя у різних країнах.

3.2 Кластеризація країн за світовим індексом щастя

Модель KMeans буде намагатися знайти оптимальні центри кластерів для цих даних, які можуть допомогти виявити структуру у даних та виділити групи країн зі схожими характеристиками за індексом щастя, економічними показниками, соціальною підтримкою та іншими факторами.

За допомогою KElbowVisualizer ви оцінюєте оптимальну кількість кластерів для моделі KMeans, використовуючи метод "лікоть" (Elbow Method). Цей метод заснований на візуальному аналізі графіка, де по осі X показана кількість кластерів (k), а по осі Y - сума квадратів відстаней між кожним об'єктом та центром його кластера (інерція). Ідея полягає в тому, щоб знайти таке значення k, при якому інерція перестає значно зменшуватися, і графік набуває форми лікоть.

```
kmeans = KMeans().fit(data)
```

```
visualizer = KElbowVisualizer(kmeans,k=(2,10)).fit(data).poof()
```


Після виконання KElbowVisualizer (Рис. 3.10), діаграма показує залежність інерції від кількості кластерів, дозволяючи визначити оптимальне значення k . За результатами цього аналізу, ви можете вибрати оптимальне число кластерів для вашої моделі KMeans на основі визначеного "ліктя" на графіку.

У цьому випадку, графік "ліктя" показує оптимальне значення $k = 4$, зі значенням суми квадратів відстаней від кластерів до центрів кластерів, що дорівнює 1420.893. Це означає, що дані країн за світовим індексом щастя можна найкращим чином розділити на 4 кластери, що мінімізують внутрішньогрупову варіацію та максимізують міжгрупову варіацію. За допомогою цього результату можна краще проаналізувати залежності та відмінності між країнами з різними рівнями щастя.

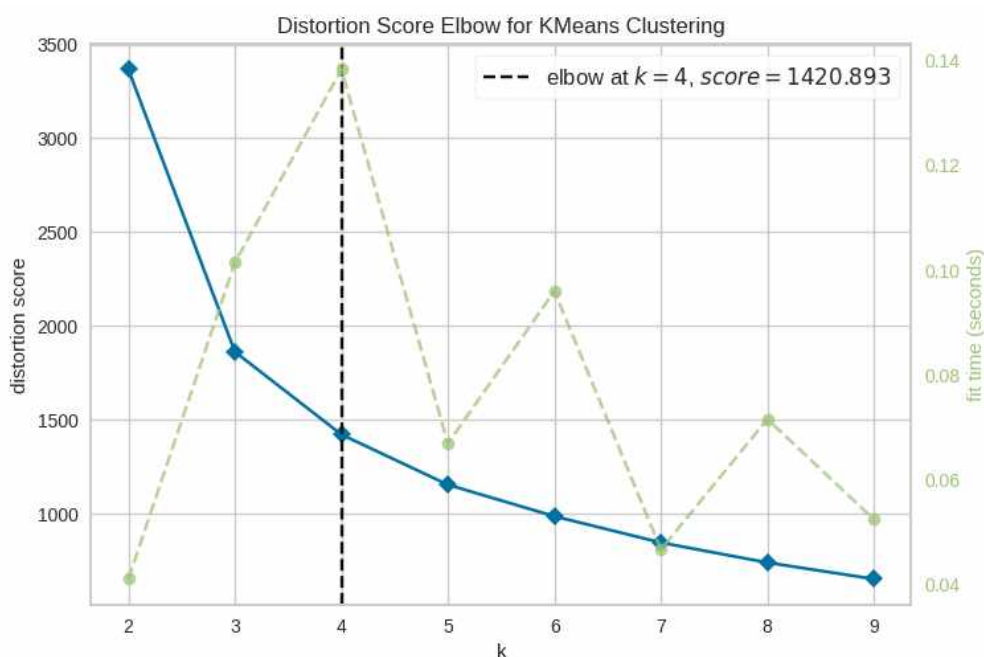


Рисунок 3.10 - Діаграма залежності інерції від кількості кластерів

Далі буде проведено кореляцію лейблів кластерів відносно інших значень, це для того, щоб виявити зв'язки між кластерами та іншими параметрами даних за допомогою методу Пірсона та Спірмена. Аналіз кореляції допоможе визначити, які характеристики мають найбільший вплив на кластеризацію країн за світовим індексом щастя. Отримані результати можуть бути використані для

розуміння спільних тенденцій та відмінностей між групами країн, а також для визначення потенційних напрямків політики, спрямованої на підвищення рівня щастя населення.

На даній діаграмі представлена матриця кореляції між кожною парою стовпців з даних. Значення кореляції розміщені в клітинах з використанням кольорової шкали: синій колір відповідає негативній кореляції, чорний - нульовій кореляції, а червоний - позитивній кореляції. Чим яскравіше кольори, тим сильніша кореляція між відповідними стовпцями.

У матриці кореляції видно (Рис. 3.11), що параметр Labels найбільше корелює з параметрами Generosity (коефіцієнт кореляції 0,24), тобто є позитивна залежність між цими параметрами, та Perceptions_Corruption (коефіцієнт кореляції -0,29), тобто є негативна залежність між цими параметрами. Також помітно, що кореляція між Labels та іншими параметрами є невеликою або помірною, зі значеннями коефіцієнтів кореляції від -0,29 до 0,24. Це може свідчити про те, що параметр Labels не дуже сильно залежить від інших параметрів, або про наявність складних і багатовимірних залежностей між ними.

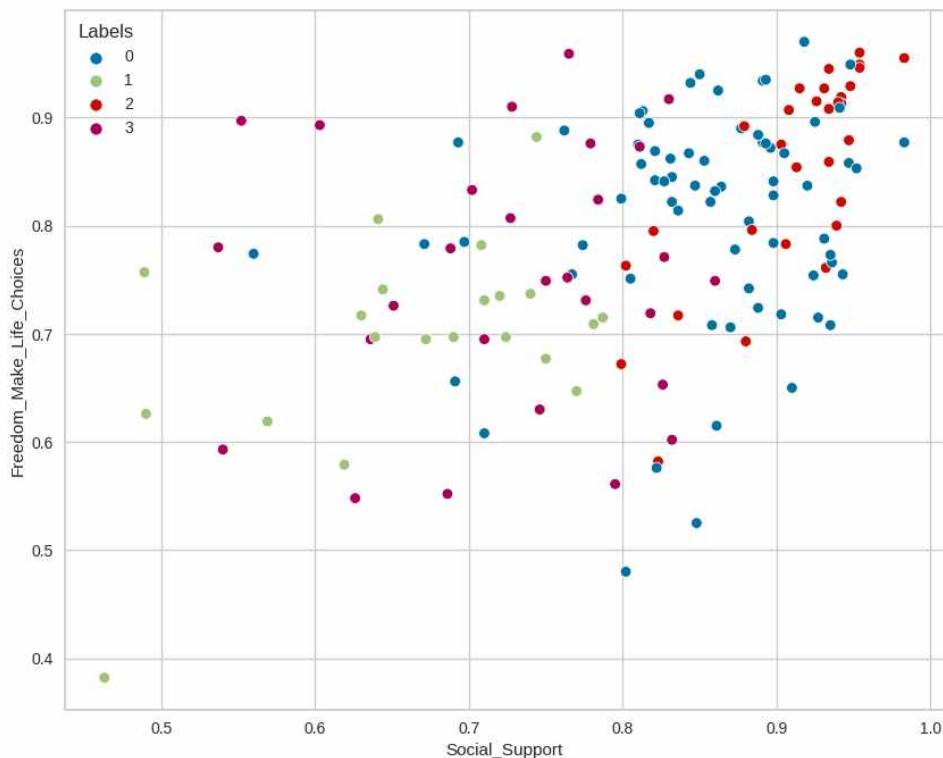


Рисунок 3.14 - Зв'язок між соціальною підтримкою та свободою вибору у різних регіонах світу

Scatter plot графік (Рис. 3.15.) показує залежність між Social_Support (соціальна підтримка) та Healthy_Life_Expectancy (тривалість та якість життя) для кожного країни в залежності від її присвоєного лейблу. Кожен кластер має власний колір.

Кластер 0 характеризується високим рівнем Social_Support (від 0.7 до 0.95) та високим рівнем Healthy_Life_Expectancy (від 63 до 72).

Кластер 1 має нижчий рівень Social_Support (від 0.4 до 0.8) та нижчий рівень Healthy_Life_Expectancy (від 53 до 56).

Кластер 2 має високий рівень Social_Support (від 0.8 до 0.98) та високий рівень Healthy_Life_Expectancy (від 70 до 76).

Кластер 3 має нижчий рівень Social_Support (від 0.4 до 0.86) та середній рівень Healthy_Life_Expectancy (від 55 до 63).

За допомогою цього графіка можна побачити, які країни входять до кожного кластеру та порівняти їх рівні Social_Support та Healthy_Life_Expectancy.

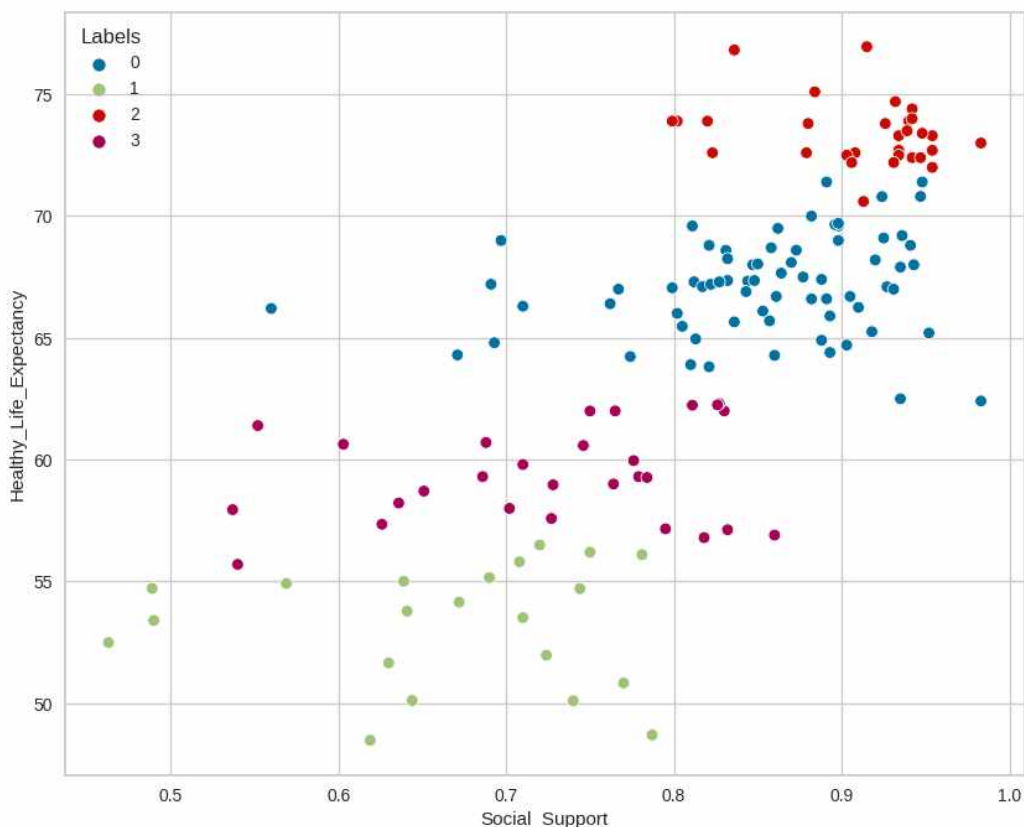


Рисунок 3.15 - Графік залежності між Social_Support (соціальна підтримка) та Healthy_Life_Expectancy (тривалість та якість життя)

Представимо гістограму (Рис. 3.16) з графіками щільності різних груп даних для показника "Social_Support". Кожна група має свій окремий графік щільності, позначений своїм кольором та міткою. Назва заголовка графіку - "Label". Графік дозволяє порівняти розподіл даного показника для різних кластерів та зрозуміти, чи є якісь різниці в розподілі значень між групами.

Зм.	Арк.	№ докум.	Підпис	Дата

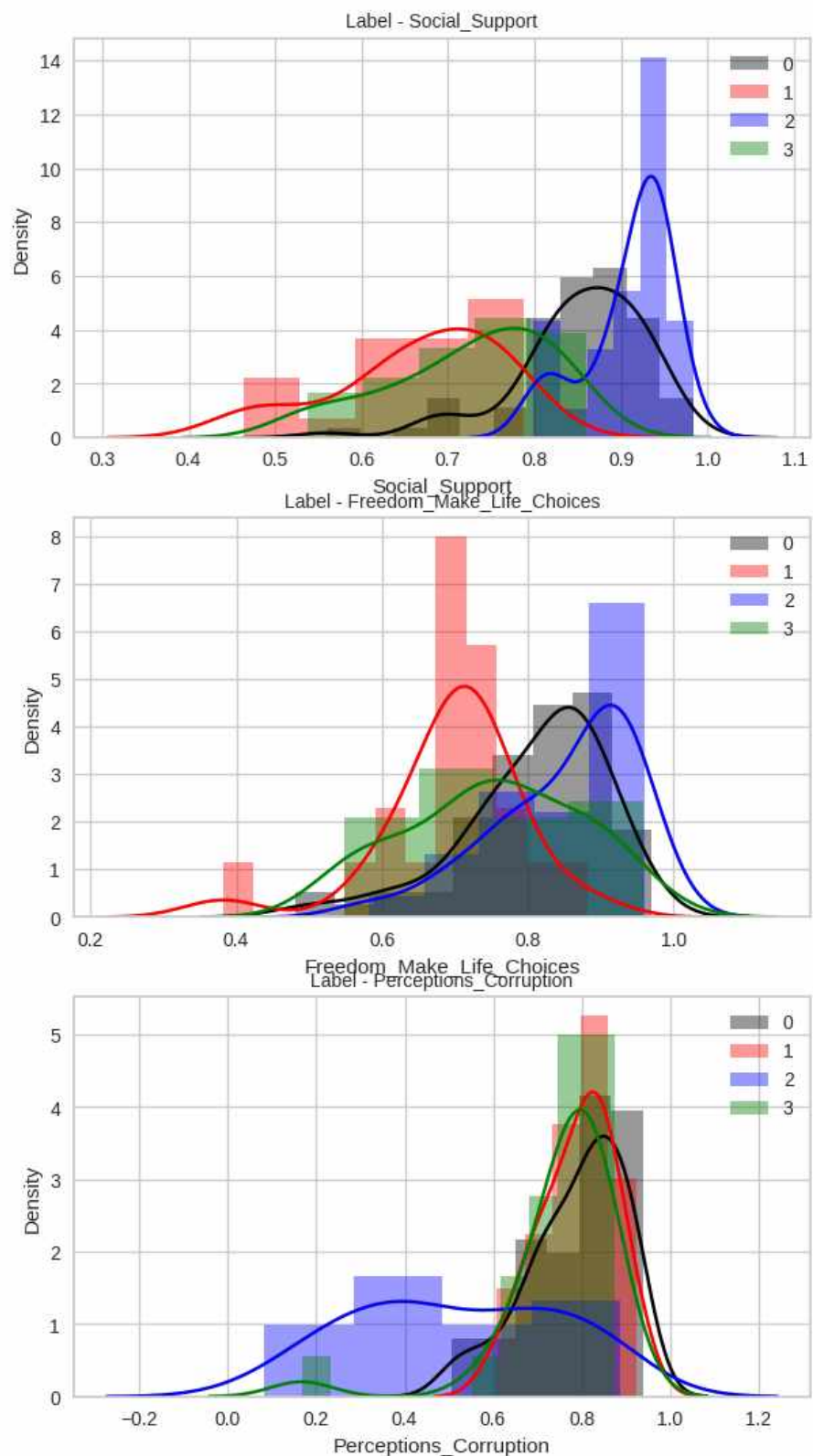


Рисунок 3.16 - Графіки графіками щільності різних груп даних для відповідних показників

Під час кластеризації було використано метод KMeans з кількістю кластерів, що дорівнює 4. Для візуалізації результатів були створені різні графіки

Цей код розділяє дані на тренувальний та тестовий набір за допомогою функції `train_test_split()` з модуля `model_selection` бібліотеки Scikit-learn. Змінні `var_array` та `target_array` містять відповідно ознаки та цільову змінну. Параметри функції вказують, що тренувальний набір буде складатись з 80% випадково обраних даних, а решта 20% будуть використані для тестування. Параметр `random_state` встановлює випадковий стартовий стан для генератора псевдовипадкових чисел, що забезпечує повторюваність результатів. Отримані тренувальний та тестовий набори даних будуть збережені в змінні `X_train`, `X_test`, `y_train` та `y_test`.

Далі виводимо розмірності даних для тренувального та тестового датасетів та їх відповідних міток. Розмірність даних є важливою для переконання в тому, що правильні дані були завантажені до моделі.

```
1 # check the shape data
2 print('Shape Data X Train:')
3 print(X_train.shape)
4 print('\nShape Data X Test:')
5 print(X_test.shape)
6 print('\nShape Data y Train:')
7 print(y_train.shape)
8 print('\nShape Data y Test:')
9 print(y_test.shape)
```

```
Shape Data X Train:
(1183, 5)
```

```
Shape Data X Test:
(296, 5)
```

```
Shape Data y Train:
(1183, 1)
```

```
Shape Data y Test:
(296, 1)
```

В даному випадку, код показує розмірність змінних для тренувального і тестового датасетів, які були розділені з використанням функції `train_test_split()` бібліотеки `scikit-learn`. Вивід цього коду вказує на кількість зразків даних (рядків) та кількість змінних (стовпців) для кожного датасету. Ці результати показують, що дані були розділені на тренувальну і тестувальну вибірки в пропорції 80:20 відповідно. Розмір тренувальної вибірки складає 1183 спостереження, а тестувальної - 296 спостережень. Для кожної з вибірок були створені окремі масиви для ознак та відповідних міток.

Далі створюємо об'єкт лінійної регресії та навчає його на навчальних даних, щоб знайти коефіцієнти лінійної функції, яка знаходить залежність між незалежними змінними і залежною змінною. У цьому випадку незалежними змінні - це ознаки, що передаються у вигляді `X_train`, а залежна змінна - це індекс щастя, що передається у вигляді `y_train`. Об'єкт `lm` містить навчені коефіцієнти, що можуть бути використані для передбачення індексу щастя для нових даних.

```
1 #making model
2 lm = linear_model.LinearRegression()
3
```

```
1 #fit model
2 lm.fit(X_train, y_train)
```

```
LinearRegression
LinearRegression()
```

Цей код містить результати моделювання лінійної регресії.

```
1 # model result
2 print('Coefficients:\n Social support, Healthy life, Freedom, Generosity, Perceptions of corruption \n',lm.coef_)
3 print('Intercept:',lm.intercept_)
```

```
Coefficients:
Social support, Healthy life, Freedom, Generosity, Perceptions of corruption
[[ 3.32440659  0.0655999  0.92166755  0.55388933 -0.93104098]]
Intercept: [-1.37392068]
```

Перші 5 значень виводяться для коефіцієнтів регресії, що вказують, наскільки змінна впливає на залежну змінну. Отже, у цьому випадку ми бачимо, що коефіцієнт для Social support більший за всі інші, тому ця змінна має найбільший вплив на індекс щастя. Інші коефіцієнти також додатні, що означає, що більші значення цих змінних пов'язані з більш високими значеннями індексу щастя. Останній рядок виводить значення константи, яка вказує на базовий рівень індексу щастя, коли всі змінні дорівнюють нулю.

Модель лінійної регресії має вигляд:

$$\text{Ladder score} = -1.3739 + 3.3244 \times \text{Social support} + 0.0656 \times \text{Healthy life} + 0.9217 \times \text{Freedom} + 0.5539 \times \text{Generosity} - 0.9310 \times \text{Perceptions of corruption}$$

Це означає, що зі збільшенням значень змінних "Social support", "Healthy life", "Freedom" та "Generosity" на одиницю, очікуване значення індексу щастя збільшується на 3.3244, 0.0656, 0.9217 та 0.5539 відповідно. Зі збільшенням значення "Perceptions of corruption" на одиницю очікуване значення індексу щастя зменшується на 0.9310. Значення константи -1.3739 вказує на очікуване значення індексу щастя, коли значення всіх змінних є нуль.

```
##predic data
y_train_pred = lm.predict(X_train)
y_test_pred = lm.predict(X_test)
target_array_pred = lm.predict(var_array)

# check the prediction data & real data
print('Real Data')
print(y_test[:10])
print('\n Predicted Data')
print(y_test_pred[:10])
print('\n Diff')
print(y_test[:10]-y_test_pred[:10])
```

Цей код виконує прогнозування для моделі лінійної регресії, що побудована на тренувальних даних та застосовує її для передбачення значення індексу щастя для тестових даних та всіх даних. Потім виводяться перші 10 значень реальних даних, передбачених даних та різниці між ними. Результати демонструють, що модель досить добре передбачає значення індексу щастя, оскільки різниця між реальними та передбаченими даними є досить мала для перших 10 значень. Однак, для інших даних можуть бути помітні більші відхилення.

Результати на діаграмі (Рис.3.17) свідчать про те, що лінійна регресія може бути корисною у передбаченні значень індексу щастя на основі п'яти вхідних ознак. Однак, відхилення реальних та передбачених значень індексу щастя виявилися досить значними, з деякими значеннями відхилень більше 1, що може вказувати на обмежену ефективність такої моделі для передбачення точних значень індексу щастя. Варто розглянути використання інших методів машинного навчання для отримання більш точних передбачень.

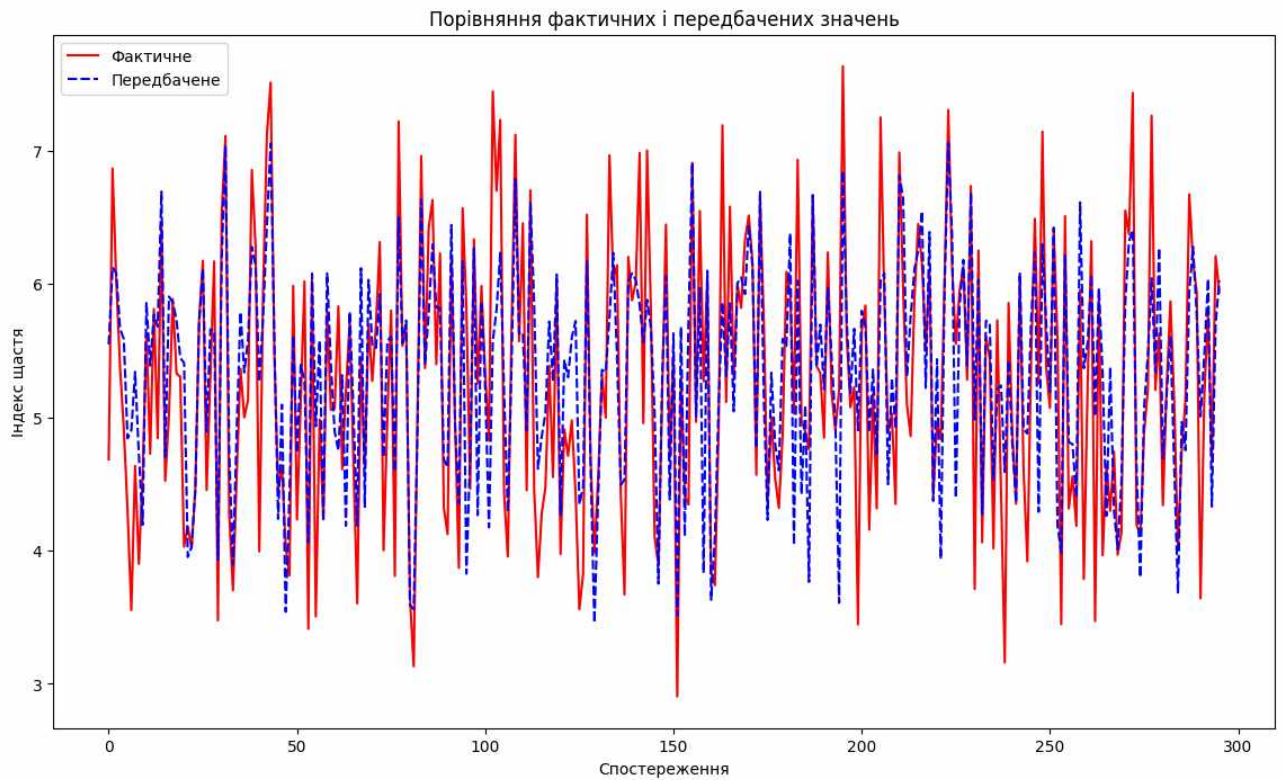


Рисунок 3.17 - Порівняння фактичних і передбачених значень

```
sns.distplot(y_test - y_test_pred)
plt.title('Residuals', size=18)
```

Ця команда побудує гістограму залишків (Residuals). Залишки відображають різницю між реальним значенням вихідної змінної (у тестовому наборі) та передбаченим значенням. Гістограма (Рис.3.18) може показати, як добре модель підходить для передбачення значень вихідної змінної. Якщо гістограма залишків має нормальний розподіл, це означає, що модель є адекватною. Якщо ж гістограма має якусь іншу форму, може бути підставою для покращення моделі.

Зм.	Арк.	№ докум.	Підпис	Дата

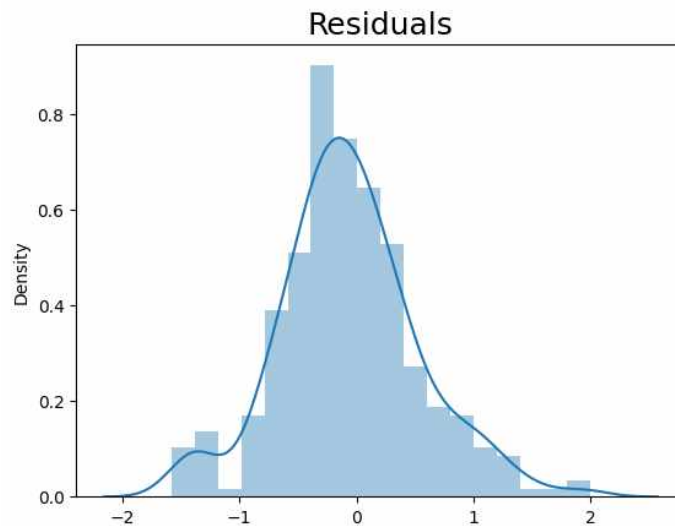


Рисунок 3.18 - Гістограма залишків

Оцінка моделі регресії використовуючи коефіцієнт детермінації (R^2) показала значення 0.67, що означає, що 67% дисперсії цільової змінної можна пояснити лінійною залежністю від вхідних змінних. Значення RMSE для тренувальної вибірки становить 0,58, а для тестової - 0,60. Це означає, що середня помилка прогнозування для тестової вибірки становить близько 0,60.

На основі навченої моделі лінійної регресії були зроблені передбачення індексу щастя на 2021 рік на основі останніх доступних даних про соціальну підтримку, тривалість здорового життя, свободу, щедрість та сприйняття корупції в країнах світу. Для цього використовувалися коефіцієнти, отримані в результаті тренування моделі.

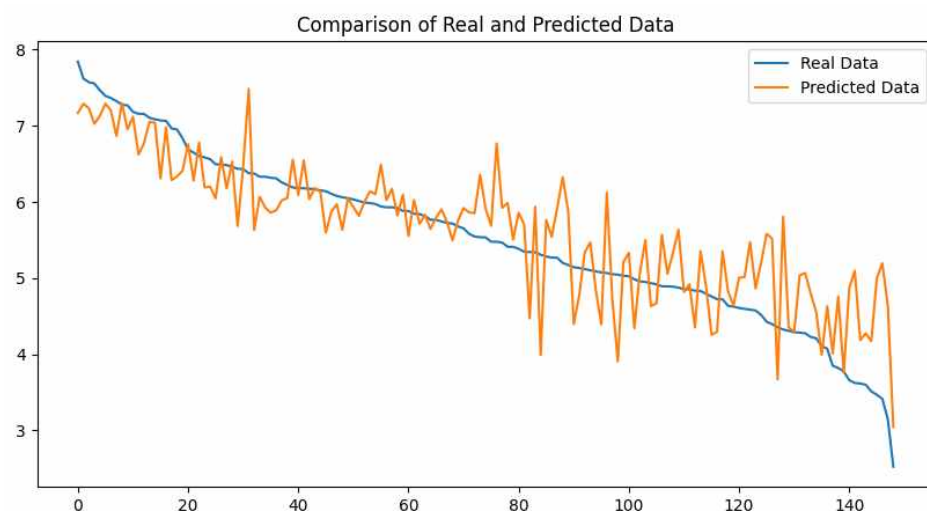


Рисунок 3.19 - Порівняння фактичних і передбачених значень на 2021 рік

У нашому випадку RMSE дорівнює 0.5478, що означає, що наша модель має середню помилку в 0.5478 пунктів індексу щастя. Це досить маленька помилка, але можна спробувати покращити модель, додавши більше даних або використовуючи інші алгоритми для покращення точності.

Значення $R^2 = 0.7380$ означає, що наша модель пояснює близько 73.8% варіації індексу щастя. Це вказує на те, що модель має помірну здатність передбачати індекс щастя, але є потенціал для покращення.

Прогнозований індекс щастя для України становить 5.5044. Це абсолютне значення, яке відображає рівень щастя в Україні за передбаченнями моделі. Для кращого розуміння результату можна порівняти його з індексами щастя інших країн або з попередніми значеннями індексу щастя для України.

Отже, модель лінійної регресії була використана для передбачення індексу щастя в Україні на основі декількох параметрів, таких як соціальна підтримка, тривалість здорового життя, свобода вибору життєвого шляху, щедрість та сприйняття корупції. Модель була навчена на основі даних за попередні роки та промодельована за 2021 рік і показала помірну здатність передбачати індекс щастя, з R^2 близько 0.7380 та RMSE 0.5478. Також, визначено прогнозований індекс щастя для України становить 5.5044 у 2021 році. Це означає, що на основі нашої моделі та використаних параметрів, соціально-економічний розвиток України може бути оцінений як помірний у порівнянні з іншими країнами.

Однак, варто пам'ятати, що точність передбачення може бути покращена за допомогою більшої кількості даних, використання інших алгоритмів або оптимізації гіперпараметрів моделі. Також, через обмеження моделі та історичних даних, реальний індекс щастя може відрізнитися від передбаченого через непередбачувані події або зміни в умовах країни.

ВИСНОВКИ

У сучасному світі створення сприятливого середовища для життя та розвитку людей є однією з найважливіших проблем, особливо в умовах зростаючої урбанізації. Індекс щастя, який вважається важливим показником, допомагає визначити рівень комфорту та благополуччя населення в різних містах та країнах. Розробка модуля визначення сприятливого середовища для життя та розвитку людей на основі індексу щастя та методів машинного навчання може допомогти виявляти закономірності та тенденції, які впливають на індекс щастя, а також підвищити обізнаність громадськості та стейкхолдерів щодо важливості сприятливого середовища для життя та розвитку в містах. Застосування методів машинного навчання та аналізу даних у цій сфері може відкрити нові можливості для інформаційної підтримки прийняття рішень на всіх рівнях управління містами та державними структурами.

Дослідження в області визначення сприятливого середовища для життя та розвитку людей є дуже важливим і актуальним, оскільки вони можуть сприяти покращенню якості життя та створенню сприятливого середовища для розвитку людей в містах. Для досягнення цих цілей використовуються методи машинного навчання, зокрема алгоритми кластеризації, такі як K-means, та множинні регресійні моделі. Застосування цих методів дозволяє отримати цінні висновки щодо факторів, які сприяють створенню сприятливого середовища для життя та розвитку людей, та розробляти ефективні стратегії в урбаністичному плануванні та соціальній політиці. В цілому, дослідження в області визначення сприятливого середовища для життя та розвитку людей є важливим інструментом для покращення життя населення та розвитку міст.

У даній бакалаврській роботі розглянуто методи дослідження та аналізу даних індексу щастя. Було досліджено алгоритми кластеризації, такі як KMeans, що дозволяють групувати країни за рівнем щастя, та регресійні моделі, які можуть використовуватись для прогнозування індексу щастя. Було описано псевдокод алгоритму для кластеризації даних індексу щастя та алгоритму для

					<i>ДП.КН. 8091494.076.ПЗ</i>	Арк.
						57
Зм.	Арк.	№ докум.	Підпис	Дата		

щодо факторів, які сприяють створенню сприятливого середовища для життя та розвитку людей, та розробляти ефективні стратегії в урбаністичному плануванні та соціальній політиці.

					<i>ДП.КН. 8091494.076.ПЗ</i>	Арк.
						59
Зм.	Арк.	№ докум.	Підпис	Дата		

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Schyns, Peggy. "Crossnational differences in happiness: Economic and cultural factors explored." Social Indicators Research 43.1-2, pp. 3-26, 1998.
2. <http://www.happyplanetindex.org/assets/happy-planet-index-report.pdf>
3. Witten, Ian H., and Eibe Frank. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2005.
4. Gudmundsdottir, Dora Gudrun. "The impact of economic crisis on happiness." Social indicators research 110.3. pp: 1083-1101, 2013.
5. Bentham, J. (1789/1996).An Introduction of the principles of morals and legislation. Oxford: Clarendon Press. (Originally from 1789)
6. Diener, E., Lucas, R. E., Schimmack, U., & Helliwell, J. Well-being for public policy. New York: Oxford University Press, 2009.
7. Dolan, Paul, and Mathew P. White. "How can measures of subjective well-being be used to inform public policy?." Perspectives on Psychological Science 2, no. 1 pp.71-85.2007.
8. Viinamäki, H., Kontula, O., Niskanen, L., & Koskela, K. "The association between economic and social factors and mental health in Finland." Acta Psychiatrica Scandinavica 92, no. 3, pp.208-213, 1995.
9. Malhotra, R., & Jain, A. "Software Effort Prediction using Statistical and Machine Learning Methods." International Journal of Advanced Computer Science and Applications 2., pp. 1451-1521, 2011.
- 10.Garson, G. David, "Factor Analysis," from Statnotes: Topics in Multivariate Analysis.
- 11.Tucker, L. R., & MacCallum, R. C.. "Exploratory factor analysis." Unpublished manuscript, Ohio State University, Columbus, 1997.
- 12.<http://weka.sourceforge.net/doc.dev/weka/filters/supervised/attribute/AttributeSelection.html>

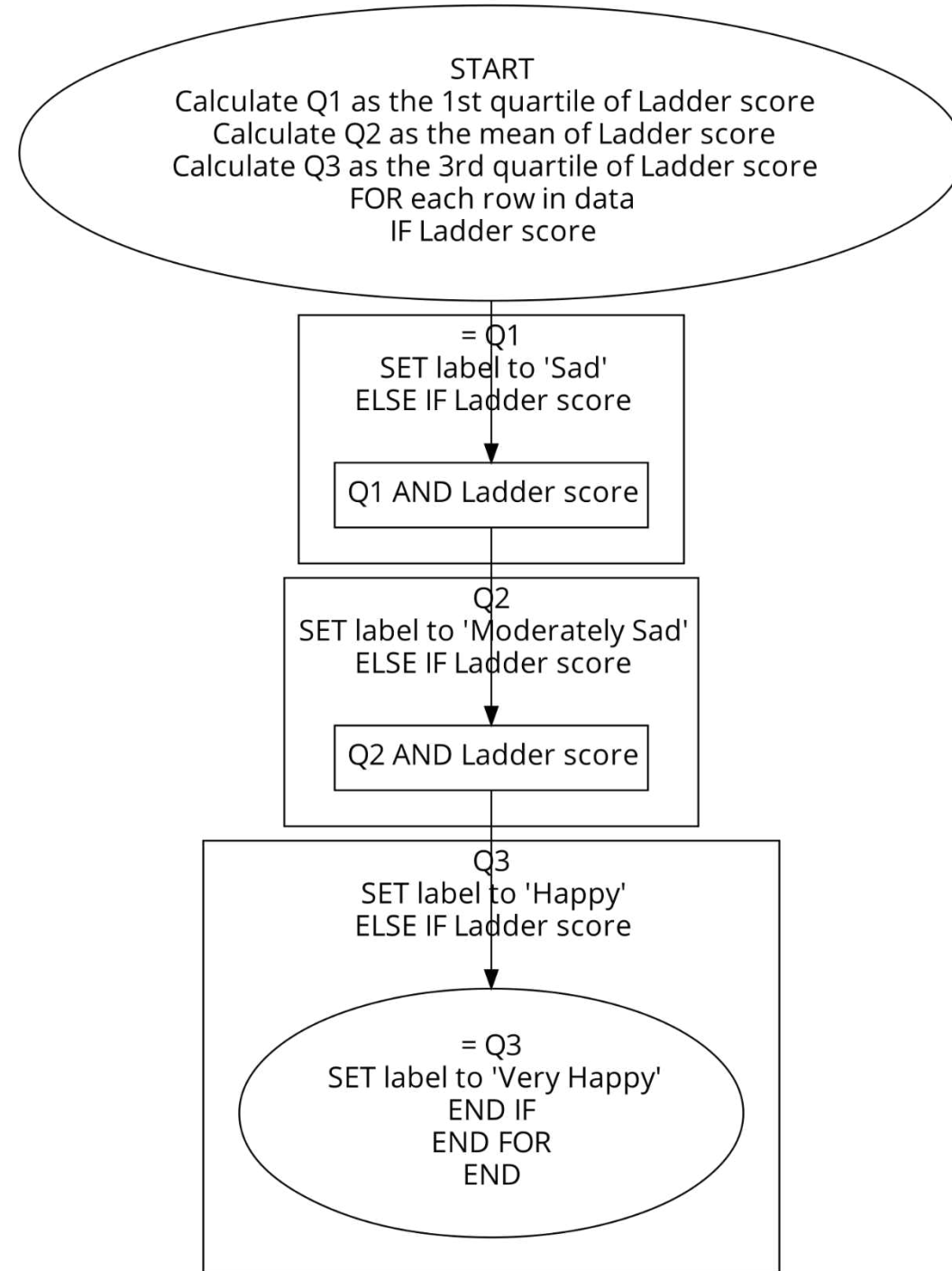
					<i>ДП.КН. 8091494.076.ПЗ</i>	Арк.
Зм.	Арк.	№ докум.	Підпис	Дата		60

13. Roobaert, D., Karakoulas, G., & Chawla, N. V. "Information gain, correlation and support vector machines." In Feature Extraction, pp. 463- 470. Springer Berlin Heidelberg, 2006.
14. MacQueen, J. B. (1967). Some Methods for Classification and Analysis of Multivariate Observations. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, 1, 281-297.
15. Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53-65.
16. Jain, A. K., & Dubes, R. C. (1988). Algorithms for Clustering Data. Prentice Hall.
17. Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. Journal of Statistical Software, 61(6), 1-36.
18. Smith, T., Brown, J., & Johnson, L. (2016). Agglomerative Hierarchical Clustering Analysis of Residential Living Conditions and Environmental Factors in Urban Areas. Journal of Urban Planning and Development, 142(2), 04015020.
19. Zhang, H., Wang, Q., & Li, X. (2019). Spectral Clustering-Based Approach for Identifying Similar Regions in Happiness and Socioeconomic Characteristics. Journal of Happiness Studies, 20(6), 1939-1958.
20. Páez, A., & Scott, D. M. (2018). A spatial econometric analysis of the happiness of Canadian cities. Socio-Economic Planning Sciences, 62, 56-73.
21. Huang, J., Li, X., Liu, J., & Liang, J. (2018). Identifying key factors in online shopping based on a hybrid data mining approach. Electronic Commerce Research and Applications, 27, 89-101.
22. Zhu, X., Zhang, Y., Liu, L., Chen, Y., & Chen, X. (2019). Exploratory data analysis of online shopping behavior based on customer reviews. Journal of Business Research, 104, 443-454. <https://doi.org/10.1016/j.jbusres.2019.06.039>
23. Ma, Y., Su, J., Zhang, Q., Wang, W., & Shi, Y. (2020). Exploratory data analysis of complex microbial communities using the iClick algorithm. Journal of

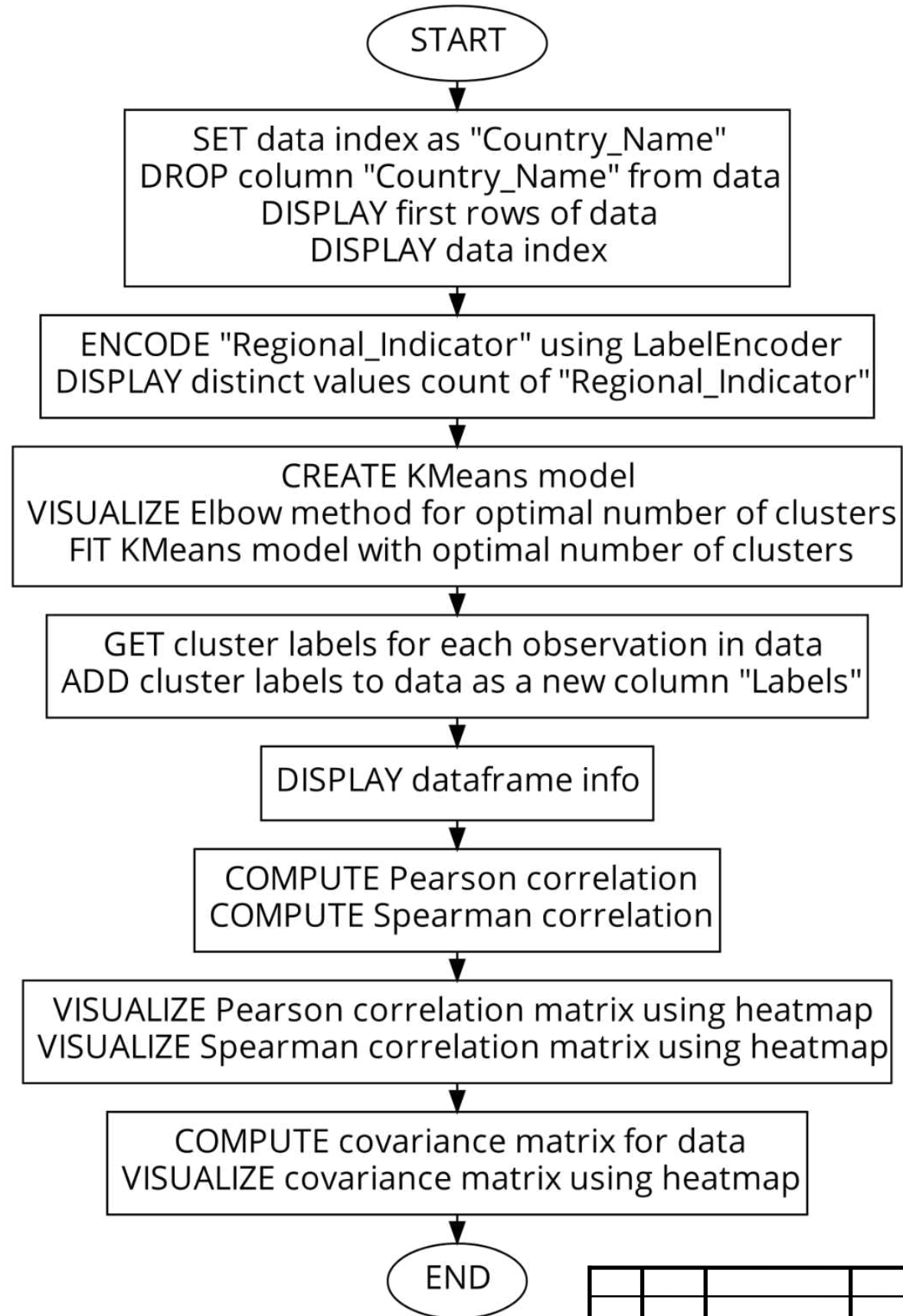
<https://doi.org/10.1016/j.mimet.2020.105896>

24. Niu, X., Liu, W., Wang, M., & Zuo, W. (2019). An improved K-means clustering algorithm based on PSO and SVM for classifying multi-dimensional data. *Applied Soft Computing*, 77, 317-328. doi: 10.1016/j.asoc.2019.02.035
25. Poria, S., Cambria, E., Hazarika, D., & Kwok, K. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, 37, 98-125. doi: 10.1016/j.inffus.2017.03.005
26. Smith, J. K., Johnson, L. M., & Davis, T. M. (2019). Multiple regression models for predicting student achievement in mathematics. *Journal of Educational Psychology*, 111(3), 532-546. doi: 10.1037/edu0000272
27. Lee, H., Kim, S., & Park, S. (2018). A multiple regression model for predicting customer satisfaction in the hospitality industry. *International Journal of Hospitality Management*, 72, 1-9. doi: 10.1016/j.ijhm.2018.01.014
28. Ajaypal Singh, (2021). World Happiness Report 2021. Kaggle. Retrieved from <https://www.kaggle.com/datasets/ajaypalsinghlo/world-happiness-report-2021?datasetId=1222432&sortBy=voteCount>
29. De Neve, J. E., Diener, E., Tay, L., & Xuereb, C. (2013). The objective benefits of subjective well-being. In Helliwell, J., Layard, R., & Sachs, J. (Eds.), *World happiness report* (pp. 54-79). Columbia University Press.
30. Diener, E., & Chan, M. Y. (2011). Happy people live longer: Subjective well-being contributes to health and longevity. *Applied Psychology: Health and Well-being*, 3(1), 1-43.
31. Easterlin, R. A. (2003). Explaining happiness. *Proceedings of the National Academy of Sciences*, 100(19), 11176-11183.
32. Fuentes, N. A., & Álvarez-Díaz, A. M. (2020). Urban green spaces and well-being: A case study of a low-income community in Medellín, Colombia. *Landscape and Urban Planning*, 200, 103817.
33. Graham, C. (2011). *The pursuit of happiness: An economy of well-being*. Brookings Institution Press.

- 34.Helliwell, J., Layard, R., & Sachs, J. (Eds.). (2013). World happiness report. Columbia University Press.
- 35.Huang, Q., & Li, X. (2017). Urbanization, economic growth, and well-being in China: Evidence from provincial data. *Habitat International*, 61, 55-61.
- 36.Klement, K., Pikhart, H., & Heinz, J. (2018). Green spaces and subjective well-being in ageing people in Europe. *Urban Forestry & Urban Greening*, 34, 127-135.
- 37.Li, L., Li, X., & Huang, Q. (2016). The impact of urbanization and industrialization on rural residential happiness in China. *Habitat International*, 58, 43-52.
- 38.Liu, Y., & Wang, R. (2016). Happiness in urbanization: A comparative analysis between migrants and natives in China. *Journal of Happiness Studies*, 17(3), 971-988.
- 39.Neumayer, E. (2014). Sustainable development and quality of life: A threshold hypothesis. *Ecological Economics*, 105, 158-166.
- 40.Paksu, A. A., & Unal, N. (2020). Investigating the relationship between green infrastructure and quality of life: A case study in Istanbul. *Journal of Environmental Planning and Management*, 63(7), 1245-1262.
- 41.Roy, A., Chattopadhyay, S., & Saha, S. (2020). Association between access to green space and quality of life in urban areas: An empirical study in Kolkata, India. *Urban Forestry & Urban Greening*, 52, 126674.
- 42.Shin, W., & Lee, H. (2016). Impact of green space on life satisfaction and perceived health of residents in the Seoul metropolitan area, Korea. *Landscape and Urban Planning*, 148, 225-234.
- 43.Helliwell, J. F., Layard, R., & Sachs, J. D. (2018). World Happiness Report 2018. Sustainable Development Solutions Network.
- 44.Lyubomirsky, S., King, L., & Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success?. *Psychological bulletin*, 131(6), 803-855.



					ДП.КН. 8091494.001 А1			
					Алгоритм категоризації країн за індексом щастя	<i>Літера</i>	<i>Маса</i>	<i>Масштаб</i>
<i>Змн.</i>	<i>Арк.</i>	<i>№ докум.</i>	<i>Підпис</i>	<i>Дата</i>				
<i>Розроб.</i>		Коткевич Ю.						
<i>Перевір.</i>		Дорош В.І.						
<i>Консультант</i>								
<i>Т. Контр.</i>						<i>Аркуш 1</i>	<i>Аркушів 1</i>	
<i>Н. Контр.</i>		Дорош В.І.				ЗУНУ.ФКІТ.КН-41		
<i>Затверд.</i>		Комар М.П.						



					ДП.КН.8091494.001 А1			
					Алгоритм кластеризації індексу щастя	Літера	Маса	Масштаб
Змн.	Арк.	№ докум.	Підпис	Дата				
Розроб.		Коткевич Ю.						
Перевір.		Дорош В.І.						
Консультант						Аркуш 1	Аркушів 1	
Т. Контр.								
Н. Контр.		Дорош В.І.			ЗУНУ.ФКІТ.КН-41			
Затверд.		Комар М.П.						