\_\_\_\_*Національний університет харчових технологій*\_\_\_\_\_\_

(повне найменування вищого навчального закладу)

*Кафедра інформаційних технологій,*

*штучного інтелекту і кібербезпеки*

(повна назва кафедри, циклової комісії)

КУРСОВА РОБОТА

з дисципліни *Аналітика великих даних*

(назва дисципліни)

на тему:  *Збір аналітичних даних та побудова аналітичного звіту за темою: аналіз показників управління товарними запасами та перевірка стратегії для мережі алкомаркетів*

Здобувач(ка) 1 курсу групи КН-1-3М

Спеціальності \_\_*122 "Комп’ютерні науки"*

*Кучерявого М. В.*

(прізвище та ініціали)

Керівник *ст. викладач Грама М.П.*

(посада, вчене звання, науковий ступінь, прізвище та ініціали)

Національна шкала \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Кількість балів: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Члени комісії:

\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис) (прізвище та ініціали)

\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис) (прізвище та ініціали)

\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис) (прізвище та ініціали)

*Я як здобувач НУХТ розумію i підтримую політику університету з академічної доброчесності. Я не надавав i не одержував недозволеної допомоги під час підготовки цієї роботи. Використання ідей, результатів i текстів інших авторів мають посилання на вiдповiдне джерело.*

\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис) (прізвище та ініціали)

Київ-2024

*НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ХАРЧОВИХ ТЕХНОЛОГІЙ*

(назва вищого навчального закладу)

Кафедра*інформаційних технологій, штучного інтелекту і кібербезпеки*

Дисципліна *Аналітика великих даних*

Спеціальність*122 «Комп’ютерні науки»***\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Курс \_\_\_\_1\_\_\_\_\_Група КН-1-3М Семестр\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ЗАВДАННЯ**

**на курсову роботу здобувача**

***Кучерявого Максима Вікторовича***

1. Тема роботи : *Аналіз даних зведеної статистики показників щастя у світі;* \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2. Термін здачі здобувачем закінченого роботи *20 листопада 2024 \_р.*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

3. Вихідні дані до роботи веб ресурс з відкритою інформацією зведеної

статистики рітейл мереж \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

4. Зміст розрахунково-пояснювальної записки (перелік питань, які підлягають

розробці) Розділ 1 «Аналіз завдання»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Розділ 2 «Опис проекту звіту»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Розділ 3 «Опис реалізації звітів»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Розділ 4 «Аналіз результатів звіту»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

5. Перелік графічного матеріалу (з точним зазначенням обов’язкових креслень)

1. Схема алгоритму аналізу даних\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2. Знімки інтерфейсу користувача\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

6. Дата видачі завдання *06 жовтня 2024 \_р.*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**КАЛЕНДАРНИЙ ПЛАН**

|  |  |  |  |
| --- | --- | --- | --- |
| №  з/п | Назва етапів курсової роботи | Термін виконання  етапів курсової  роботи | Примітка |
| 1. | Аналіз поставленої задачі |  |  |
| 2. | Вибір методів вирішення задачі, опис  елементів звіту |  |  |
| 3. | Створення звіту |  |  |
| 4. | Оформлення пояснювальної записки |  |  |
| 5. | Захист курсової роботи |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Здобувач** \_\_\_\_\_\_\_\_\_\_\_\_ \_\_Кучерявий М. В.

(підпис) (прізвище та ініціали)

**Керівник курсової роботи** \_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис) (прізвище та ініціали)

**АНОТАЦІЯ**

**Кучерявий М. В.: Аналіз даних зведеної статистики показників щастя у світі;**

Курсова робота складається з \_\_ сторінок, \_\_ таблиць, \_\_ рисунків, \_\_ додатків та літературних джерел.

В даній курсовій роботі створено аналітичний звіт показників щастя населення країн світу і їх взаємозв’язок з іншими факторами, що впливають на якість життя соціуму. В якості основи для проведення аналізу були використані дані з відкритих джерел, що забезпечує отримання цілої низки різної довідкової інформації для обробкт. Під час виконання курсової роботи проведено пошук відкритої інформації щодо показників щастя на основі соціальних опитувань та об’єднання знайденої інформації в єдиний файл. Розроблено аналітичний звіт.

Аналітичний звіт створено у програмному комплексі PowerBI з

використанням мови DAX.

Ключові слова: LIFE CONDITION, ,....., ЗВІТ, ВІЗУАЛІЗАЦІЇ,

АНАЛІТИКА ДАНИХ.

**ANNOTATION**

**Rusyn D. K.: Collection of analytical data and construction of an analytical report on the topic: analysis of inventory management indicators and verification of strategy for a network of liquor stores.**

The term paper consists of \_\_ pages, \_\_ tables, \_\_ figures,\_\_ appendices and literary sources.

This term paper created an analytical report on inventory management indicators and strategy verification for a chain of liquor stores based on collected open data on the Internet, which provides a number of different background information about the work. During the course work, a search for open information on the sale of asynchronous motors in Ukraine was conducted, the information found was combined into a single file, and an analytical report was developed.

The analytical report was created in the PowerBI software complex with

using the DAX language.

Keywords: SKU, INVENTORY,....., REPORT, VISUALIZATIONS,

DATA ANALYTICS.

**ЗМІСТ**

[**ВСТУП 7**](#_fddfq1nloxhc)

[**РОЗДІЛ 1. “АНАЛІЗ ЗАВДАННЯ” 8**](#_mijvbn9i77iq)

[**РОЗДІЛ 2. “ОПИС ПРОЕКТУ ЗВІТУ” 9**](#_uylcwa50pr5i)

[**РОЗДІЛ 3. “ОПИС РЕАЛІЗАЦІЇ ЗВІТІВ” 10**](#_7tz8n7pxr96p)

[**РОЗДІЛ 4. “АНАЛІЗ РЕЗУЛЬТАТУ ЗВІТУ” 11**](#_axgxmm329zox)

[**ВИСНОВКИ 12**](#_to196qbc5iux)

[**СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 13**](#_lop8bgqenv7q)

[**ДОДАТОК А 14**](#_lswsgqjeovux)

[**ДОДАТОК Б 15**](#_f71gjpm2ungs)

# **ВСТУП**

У сучасному світі щастя стало одним із ключових показників добробуту суспільства. Воно визначає якість життя громадян, рівень соціальної гармонії та стійкість держав. На основі статистичних досліджень та міжнародних індексів, таких як World Happiness Report, можна оцінити вплив економічних, соціальних та політичних факторів на рівень щастя у різних країнах. Аналіз цих даних дозволяє виявити закономірності та тренди, які допомагають приймати ефективні рішення у сфері державної політики, економіки та соціального розвитку.

У даній курсовій роботі основна увага зосереджена на вивченні зведеної статистики показників щастя у світі. Зокрема, досліджуються ключові фактори, що впливають на рівень щастя, а також їх взаємозв’язок із економічними показниками, такими як ВВП на душу населення, рівень соціальної підтримки, очікувана тривалість життя та рівень корупції.

У ході роботи було використано описову аналітику для визначення основних характеристик даних та діагностичну аналітику для виявлення причинно-наслідкових зв’язків між факторами.

Варто зазначити, що в процесі аналізу виникли певні труднощі, пов’язані з обробкою даних. Однією з проблем є наявність пропущених даних, що ускладнює аналіз і вимагає застосування методів обробки відсутніх значень. Іншою проблемою є невідповідність форматів або структур даних у різних таблицях, що потребує додаткової підготовки даних для забезпечення їх коректної інтеграції. Вирішення цих питань є важливим етапом дослідження, оскільки забезпечує якість та достовірність отриманих результатів.

Актуальність дослідження обумовлена зростаючою роллю нематеріальних показників у визначенні якості життя та розробці стратегій сталого розвитку. Науковий інтерес полягає у вивченні мультидисциплінарного підходу до аналізу щастя, який охоплює економічні, соціальні, культурні та психологічні аспекти.

Метою роботи є проведення аналізу зведеної статистики показників щастя у світі, визначення основних факторів впливу та виявлення можливостей для підвищення рівня добробуту в різних регіонах, а також дослідження факторів впливу на якість життя населення в різних країнах. Визначення найбільш впливових аспектів на рівень щастя в країнах та порівняння цієї характеристики з іншими регіонами.

# [**РОЗДІЛ 1. “АНАЛІЗ ЗАВДАННЯ”**](#_mijvbn9i77iq)

Основним джерелом для проведення дослідження стали таблиці World Hapines Report. В ході пошуку було визначено, що найбільш актуальним і повним на сьогодні є звіт WHR за 2023 рік. Відповідно до визначних даних для аналізу сформована предметна область дослідження, що охоплює глобальні аспекти вимірювання та аналізу рівня щастя населення у різних країнах світу. Звіт World Hapines Report є міждисциплінарним та поєднує соціальні, економічні, політичні та культурні фактори, що впливають на добробут суспільства.

Таблиці відображають ключові показники, що впливають на щастя, серед яких економічний добробут, представлений ВВП на душу населення, соціальна підтримка, очікувана тривалість здорового життя, свобода вибору, щедрість громадян та рівень корупції. Кожен із цих параметрів забезпечує багатогранне уявлення про умови життя в різних країнах. Наприклад, ВВП на душу населення відображає економічний рівень добробуту, а соціальна підтримка вказує на можливість отримання допомоги у важких ситуаціях. Очікувана тривалість здорового життя є показником фізичного добробуту, тоді як свобода вибору, щедрість і рівень корупції вказують на особистісні та суспільні умови, які формують суб'єктивне сприйняття щастя.

Рівень щастя є інтегральним показником, що залежить від складної взаємодії соціальних, економічних, політичних та культурних умов. Взаємозв’язки між цими умовами дозволяють глибше зрозуміти, які фактори мають найбільший вплив на суб’єктивну задоволеність життям. Разом з тим, у процесі аналізу виникають труднощі, зокрема наявність пропущених даних у деяких показниках або країнах, що ускладнює повноцінний аналіз. Додатковою проблемою є невідповідність або несумісність структур даних у різних роках чи звітах. Це вимагає попередньої обробки та корекції даних для забезпечення їх коректного використання.

Таким чином, предметна область таблиці World Happiness Report має значну практичну цінність для розробки стратегій соціально-економічного розвитку, оцінки ефективності політики у різних країнах та визначення пріоритетних напрямів діяльності з метою підвищення добробуту населення на глобальному рівні.

Розглянемо конкретні таблиці звіту WHR за 2023 рік. Дані були взяті з відкритого веб ресурсу Kaggle, і подані у двох текстових файлах формату CSV. Оскільки powerBI дозволяє прямий імпорт файлів CSV відкриваємо їх і проводимо первинний аналіз стовпців таблиці та перетворення перетворення даних.

Первинний огляд таблиць дозволяє визначити що вона відображають дані щодо показника індексу щастя регіону дослідження та інші. Період даних в таблиці whr200522 з 2005 року по 2022 рік. Таблиця whr2023 відповідно містить дані за 2023 рік. Проведемо семантичний аналіз значень таблиці. Опис щодо семантики таблиць візьмемо з офіційної документації WHR. За ними зібрана інформація була отримана в ході Всесвітнього опитування Gallup «Gallup World Poll» далі GWT.

**Happiness score:** суб’єктивне благополуччя це відповідь на запитання про оцінку життя. За документацією WHR це є відповідь на питання: «Будь ласка, уявіть собі драбину зі сходинками, пронумерованими від 0 до знизу до 10 угорі. Вершина драбини символізує найкраще можливе життя для вас, а низ драбини означає найгірше можливе життя для вас. На якій сходинці, на вашу думку, ви особисто стоїте на даний час?»

****Logged GDP per capita**:** логарифм валового внутрішнього продукту (ВВП) країни на душу населення, показник економічного процвітання.

**Social support**: (або наявність на кого розраховувати у важкі часи) є національною середнє значення двійкових відповідей (0 або 1)

**Healthy Life Expectancy (HLE):** Очікувана тривалість здорового життя при народженні заснована

на основі даних, взятих із Всесвітньої організації охорони здоров’я (ВООЗ).

**Freedom to make life choices is the:** це середнє національне значення відповідей на GWP питання «Чи задоволені ви чи не задоволені своєю свободою вибирати, що ти робиш зі своїм життям?»

**Generosity:** це залишкова величина регресії середнього національного рівня відповіді на GWP запитання «Чи жертвували ви гроші на благодійність минулого місяця?» на ВВП на душу населення.

**Corruption Perception:** Показником є ​​середнє національне значення відповідей опитування на два запитання GWP: «Чи широко поширена корупція уряд чи ні» та «Чи поширена корупція в бізнесі чи ні?" Загальне сприйняття є лише середнім значенням двох відповідей 0 або 1.

**Positive affect:** визначається як середнє значення трьох показників позитивного впливу GWP: сміятися, насолоджуватися та робити цікаві речі в опитуванні Gallup World Poll. Ці заходи є відповідями на наступні три запитання: «Чи багато ви посміхалися чи сміялися вчора?» «Чи відчували ви наступні почуття протягом більшої частини дня вчора? Як щодо Задоволення?», «Чи навчилися ви або зробили щось цікаве вчора?»

**Negative affect:** визначається як середнє значення трьох показників негативного впливу в

GWP. Це занепокоєння, смуток і злість, відповідно відповіді на «Зробив ви відчуваєте наступні почуття протягом більшої частини дня вчора? Як щодо хвилювання?», «Чи відчували ви наступні почуття протягом більшої частини дня вчора? Як щодо печалі?» і «Чи відчували ви стежити за почуттями протягом більшої частини дня вчора? А як щодо Гніву?»

**Country name:** відображає назву країни у якій проводилось дослідження

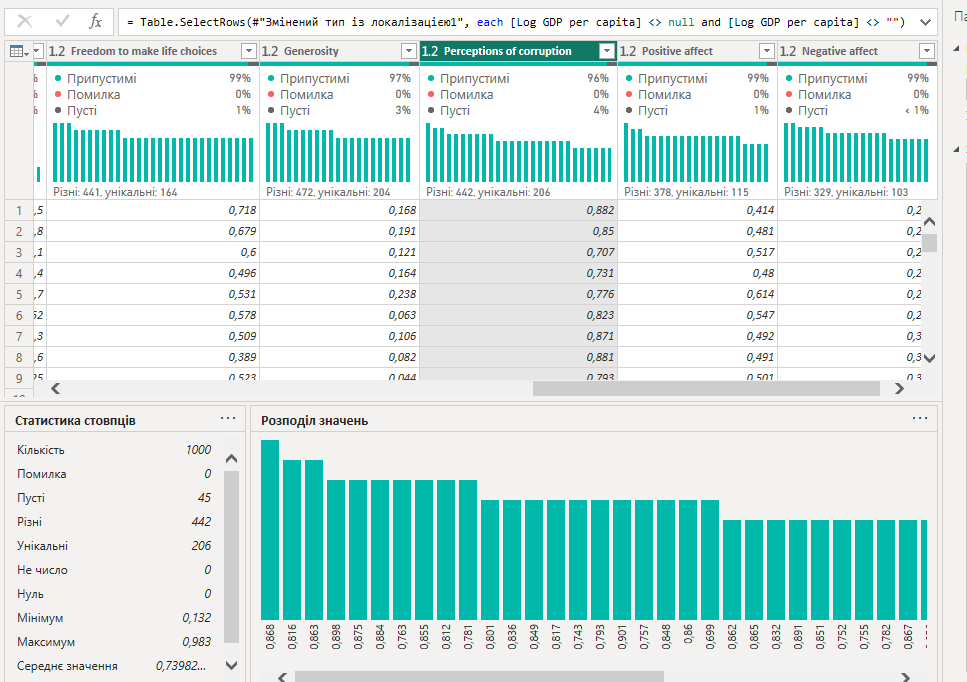
**ISO Alpha:** код країни.

**Regional indicator:** вказує на регіон в якій розташована країна.

Звернемо увагу на те що регіональний індикатор є тільки в таблиці whr2023 тому для створення звіту таблиці потрібно об’єднати та приготуванні дані до відображення.

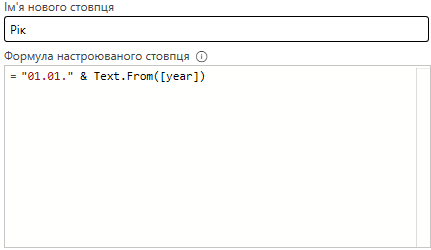
Перейдемо до підготовки таблиць. Почнемо з таблиці whr200522. Вона була перейменована на totalData. Аналізуючи дані в різних стовпцях можна побачити, що вони містять пропуски. Рисунок 1.

На жаль способи заповнення пропусків на основі існуючих даних не дали результатів, оскільки є країни в яких ці дані пропущені повністю і немає змоги визначити початкові значення. Тому було вирішено відфільтрувати пропущені дані адже їх кількість не є досить суттєвою і не перевищує 5%.

Рисунок 1

Після фільтрації дані проходили через приведення типів, оскільки більшість числових значень були інтерпретовані як рядки. Була визначена відмінність в локалізації. Через це зміна типу відбувалась з використанням локалізації English (World).

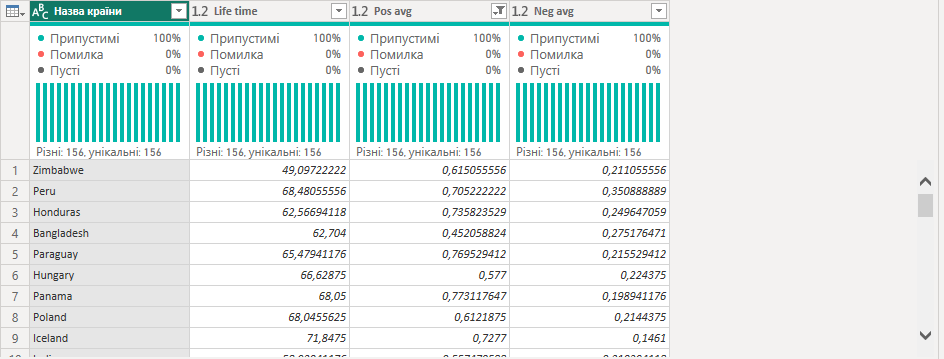
Для перетворення значення рік на «Дата» була додана додаткова колонка з настроюваним значенням. Рисунок 2

*Рисунок 2*

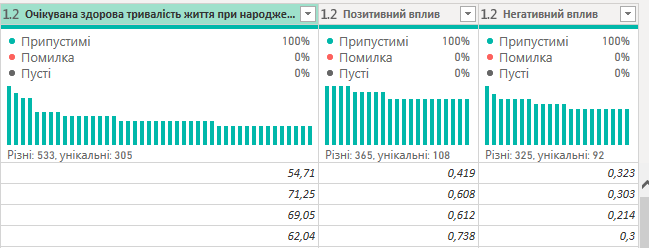
Після цього дані цієї колонки були перетворені на тип «Дата». Назви колонок були перекладені українською для подальшої зручності створення зівту.

Таблиця whr2023 пройшла схожі маніпуляції з типами даних і видалення значень, що не містяться в таблиці totalData. Для таблиці whr2023 було визначене нове значення «Рік» яке вствовлене «01.01.2023» і інтерпретоване як «Дата». Після цього дані з таблиці whr2023 були додані до таблиці totalData за допомогою додавання запитів.

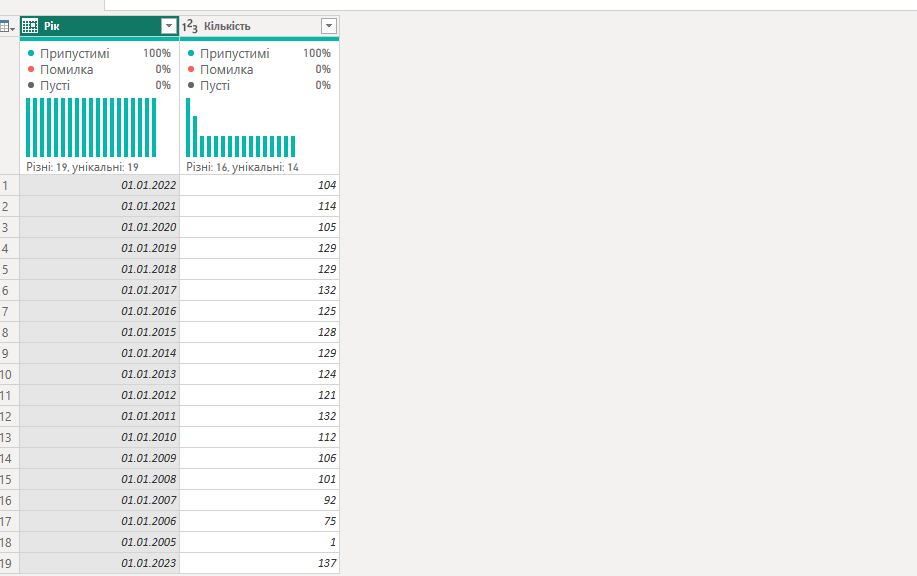
Звернемо увагу, що після об’єднання запитів в колонках «Очікувана здорова тривалість життя при народженні», «Позитивний вплив» та «Негативний вплив». Це було зумовлено тим, що в таблиці whr2023 були відсутні дані значення. Проста фільтрація знехтувала б об’єднанням таблиць, тому було вирішено заповнити їх середніми значеннями відповідно до країни. Для цього була створена додаткова таблиця «avg count fields» в якій були розраховані середні значення для вищезазначених полів відповідно до країни. Рисунок 3

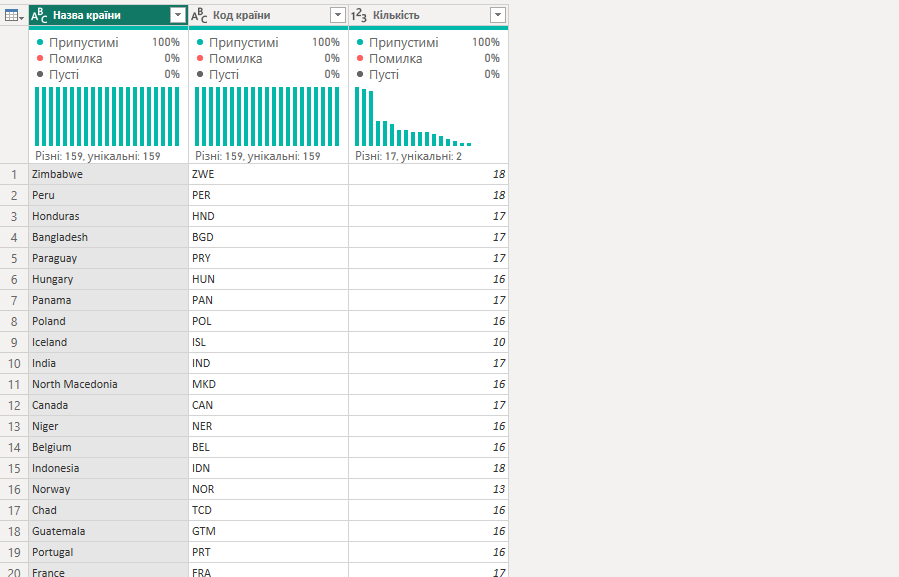
Рисунок 3

Відповідно таблиці totalData та avg count fields були об’єднані за допомогою «об’єднання запитів». Були вставлені відповідні колонки для кожлого поля яки обраховувались за умовним виразом «if[Негативний вплив] = null then[Neg avg] else[Негативний вплив]» (назви стовпців були змінені відповідно до обчислюваного поля). До отриманих стовпців було застосоване округлення до трьох знаків. В результаті змін пропуски були прибрані. Рисунок 4

Рисунок 4

Подальше перетворення було направлене на виокремлення назв країн та дати в окремі таблиці. За допомогою групування було створено таблицю «country» з полями: «Назва країни», «Код країни», «Кількість». Рисунок 5. Та таблицю year з полями: «Рік», «Кількість». Рисунок 6.

Рисунок 6

Рисунок 5

Для таблиці «country» була створена додаткова таблиця «Regional indicator» яка є видозміненою копією таблиці «whr2023». За допомогою об’єднання запитів поле «Regional indicator» було додано до таблиці «country». Дане поле було перейменовано в «Регіон» і значення null були замінені на «no region».

# **РОЗДІЛ 2. “ОПИС ПРОЕКТУ ЗВІТУ”**

Дані World Happiness Report є важливим джерелом для дослідження рівня щастя населення та аналізу впливу соціально-економічних факторів на його формування. Аналітичний звіт базується на використанні статистичних та математичних методів для оцінки ключових показників і встановлення взаємозв’язків між ними.

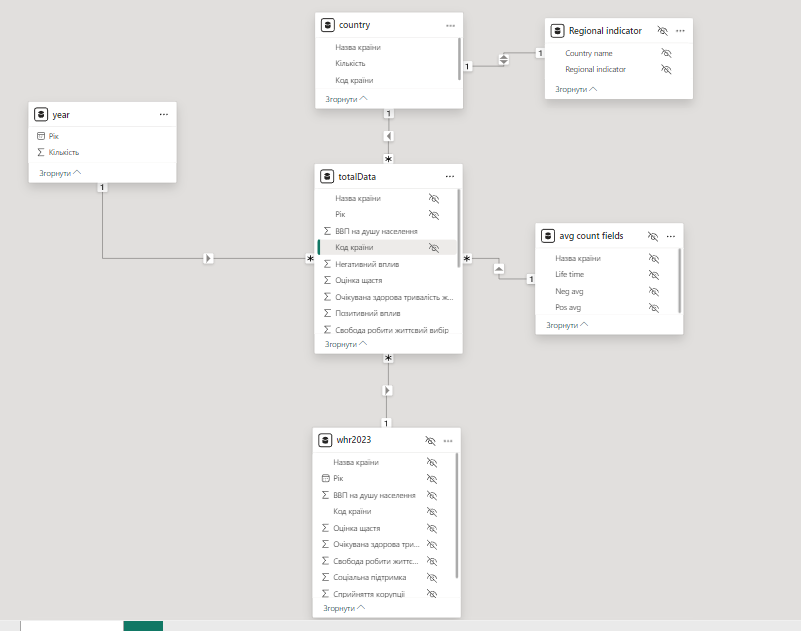
У дослідженні застосовуються методи аналізу середніх значень, що дозволяє встановити загальні тренди рівня щастя у світі. Зокрема, обчислюється середній рівень щастя, середній ВВП на душу населення та середній рівень соціальної підтримки. Ці метрики створюють основу для оцінки глобальної ситуації та порівняння регіонів.

Кореляційний аналіз є ключовим елементом звіту. Досліджується взаємозв’язок між рівнем щастя та соціально-економічними показниками, зокрема ВВП на душу населення, рівнем соціальної підтримки, тривалістю життя, свободою вибору та іншими змінними. Результати цього аналізу дають змогу виявити основні фактори, що впливають на суб’єктивне відчуття добробуту, та оцінити їхній відносний внесок.

Для сегментації даних за групами країн використовується фільтрація за регіонами, рівнями доходу та іншими категоріями. Це допомагає виявити відмінності між країнами з високим і низьким рівнем ВВП, а також зрозуміти вплив культурних та соціальних особливостей.

Отримані результати представлені у вигляді графіків і карт, що дозволяють візуально оцінити географічний розподіл рівня щастя, а також взаємозв’язки між різними показниками. Цей підхід забезпечує більш глибоке розуміння глобальних і регіональних тенденцій.

В результаті обробки та перетворень даних з таблиць World Happines Report ми отримали наступну модель даних. Рисунок 7.

Рисунок 7

Розглянемо цю модель. В якості основного джерела даних виступає таблиця totalData з відображеними полями ВВП на душу населення, Негативний вплив, Оцінка щастя, Очікувана здорова тривалість життя при народженні, Позитивний вплив, Свобода робити життєвий вибір, Соціальна підтримка, Сприйняття корупції. Таблиці year та country пов’язані з totalData і в подальшому використовуватимуться для формування роздільників. Інші таблиці та дані з них несуть допоміжний функціонал саме тому вони приховані від візуалізації.

**Візуалізації**

Розглянемо візуалізації які були створені для аналізу даних.

Перше подання дає змогу проаналізувати кількість даних які були включені в звіт. Включає в себе такі елементи:

1. Діаграму кількості зареєстрованих значень відповідно до року.

2. Мапу, на якій зображається кількість запитів відповідно до певної країни

3. Роздільник, що дозволяє відфільтрувати дані за певними роками.

Рисунок 8.

Друга візуалізація це мапа середнього рівня щастя Рисунок 9.

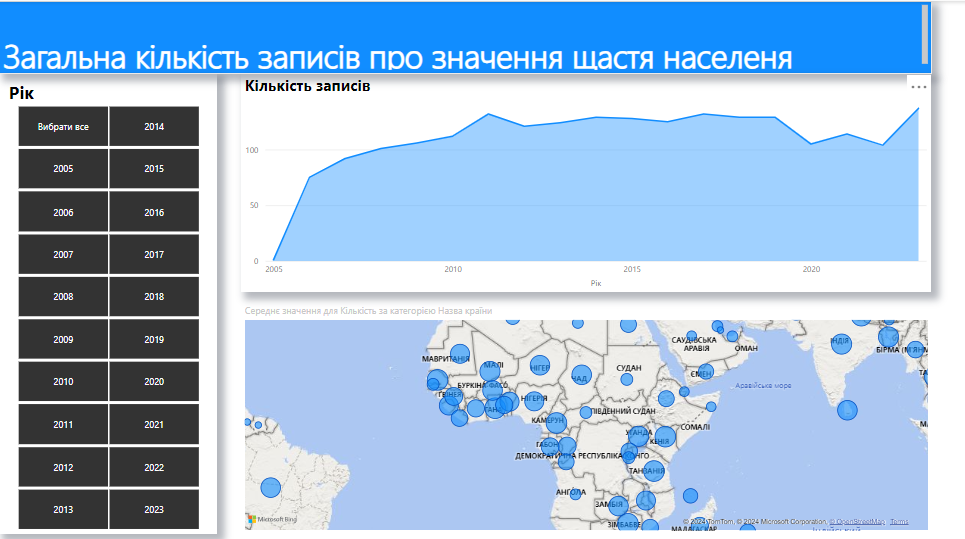
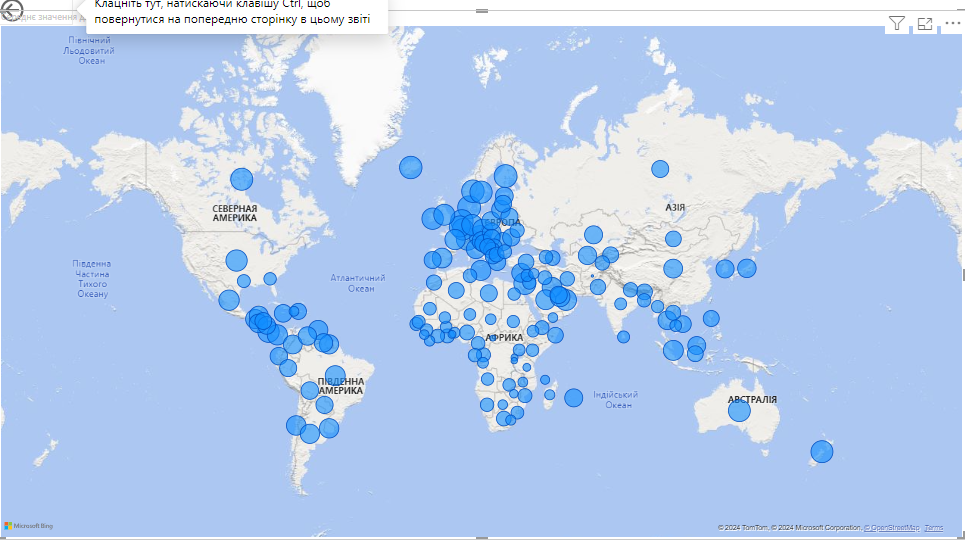
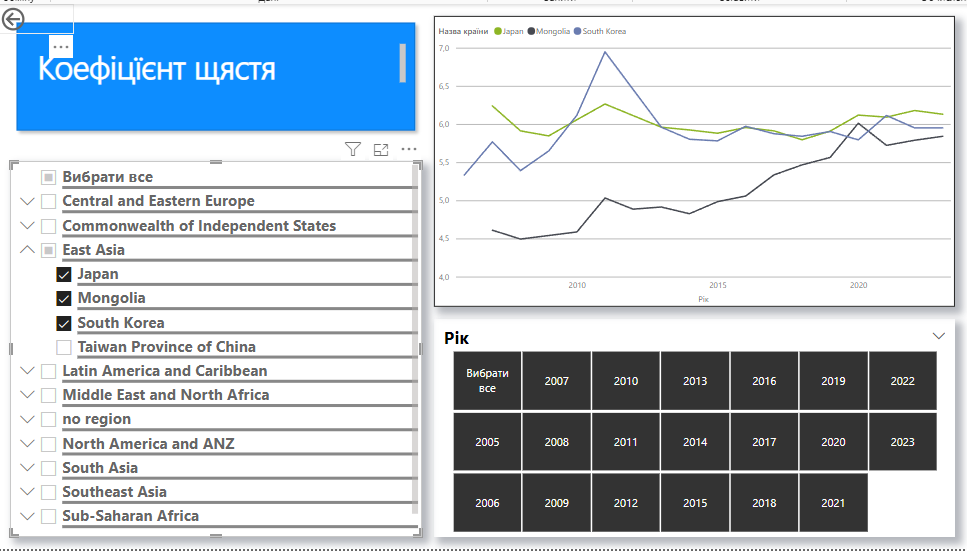
Рисунок 8

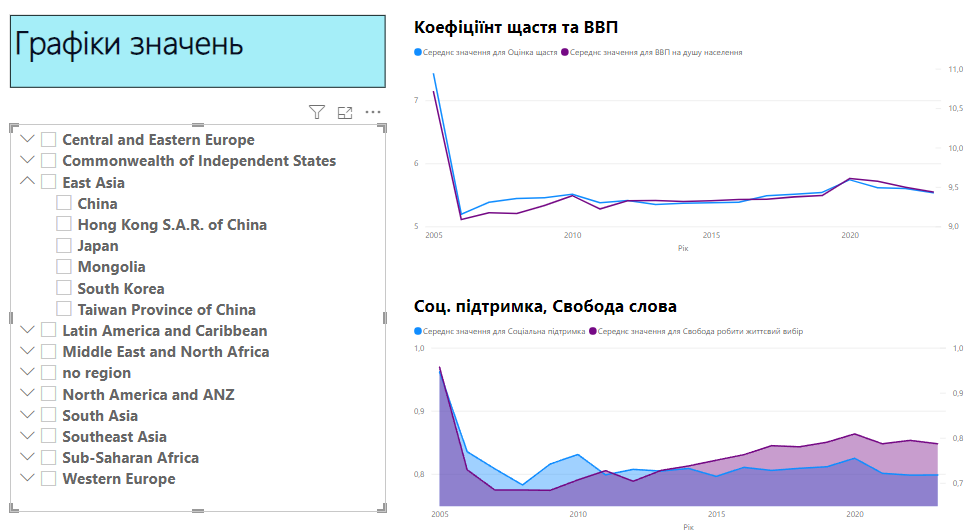
Рисунок 9

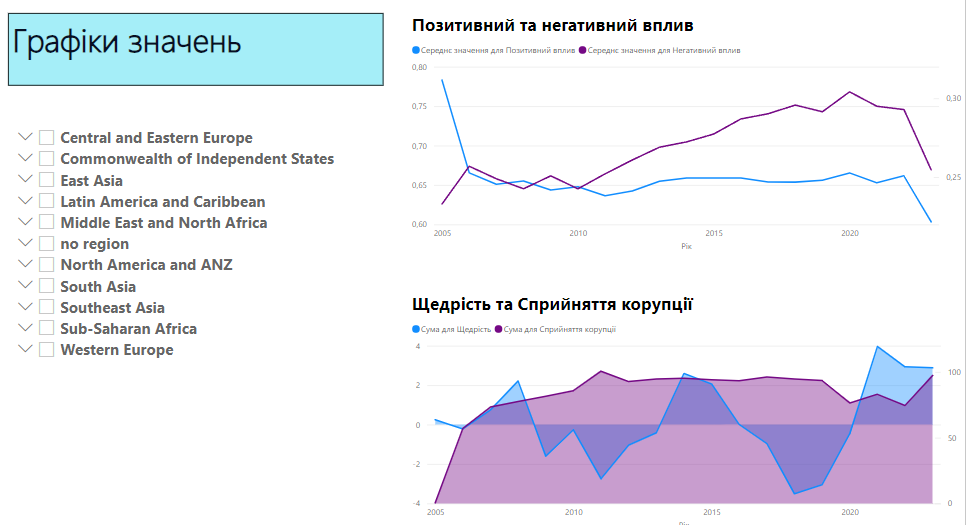
Рисунок 9

Третя візуалізація «Коефіцієнт щастя» дає змогу прослідкувати зміну рівня щастя з плином часу. За рахунок роздільників ми можемо відфільтрувати значення відповідно до часових меж потрібних для візуалізації а також певних країн та регіонів. Що будуть включені до візуалізації. Рисунок 10

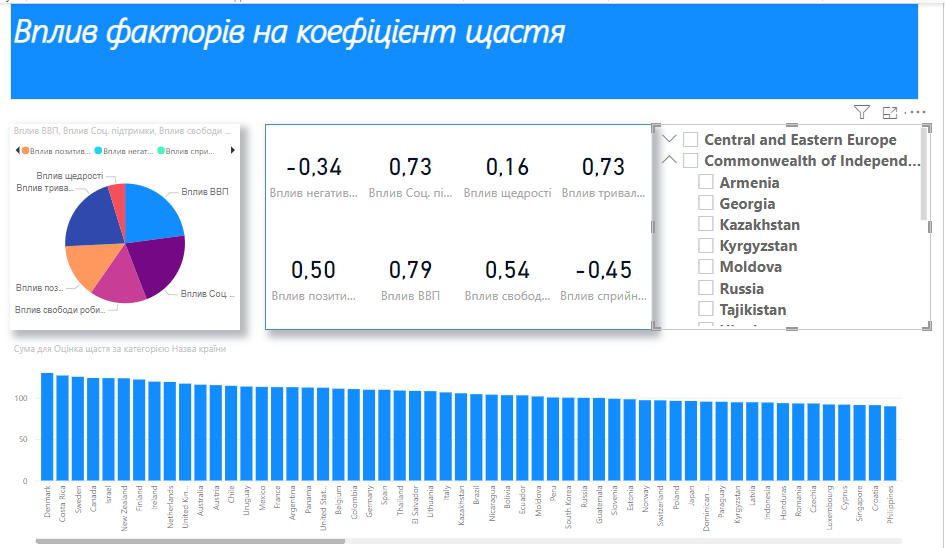
Наступна візуалізація дозволяє відслідкувати зміни значень коефіцієнту щастя, ВВП, та інших вище наведених показників в певні країні протягом певного періоду Рисунок 11, Рисунок 12. Слід зауважити, що подаються середні значення і в разі вибору роздільника як регіону значення будуть подаватись за узагальненням регіону.

Рисунок 10

Рисунок 11

Рисунок 12

# П’ята візуалізація характеризує країну за рівнем впливу різних факторів на коефіцієнт щастя. Також подає ранжування країн за рівнем щастя від найбільшої до найменшої. Рисунок 13.

Рисунок 13

# **РОЗДІЛ 3. “ОПИС РЕАЛІЗАЦІЇ ЗВІТІВ”**

Зауважимо, що рівень щастя, представлений у World Happiness Report, є виключно суб’єктивним показником, оскільки він базується на самооцінці респондентами власного життя за шкалою від 0 до 10. Цей показник відображає індивідуальне сприйняття щастя, яке формується під впливом особистих цінностей, емоційного стану, очікувань і культурного контексту.

На відміну від об’єктивних показників, таких як ВВП на душу населення чи тривалість здорового життя, рівень щастя не залежить напряму від зовнішніх умов, а натомість показує, як люди відчувають себе в цих умовах. Наприклад, навіть за високих економічних і соціальних стандартів, суб’єктивний рівень щастя може бути низьким через стрес, відсутність особистої гармонії чи незадоволеність життям. Таким чином, рівень щастя не є універсальним критерієм добробуту, а радше відображає внутрішній стан і унікальність людського досвіду. Саме тому є важливо розглянути які саме показники якості життя напряму впливають на формування суспільної думки. Зробити ми це можемо за допомогою кореляції показників добробуту та рівня щастя. Для цього використаємо міри з використанням мови DAX.

Адаптована фомула Пірсона для розрахунку кореляції

Сov – коваріація змінних X та Y

σX​ та σY​ - стандартне відхилення змінних

Формула розрахунку коваріації:

є середніми значеннями змішних при спостереженні.

Запишемо розрахунок кореляції мовою DAX. Оскільки ми хочемо розрахувати кореляцію всіх значень до рівня щастя то формула для розрахунку міри буде відрізнятись лише вхідними полями. Тому для прикладу наводимо лише одну формулу.

**Формула кореляції ВВП та рівня щастя:**

GDPCorelation =

VAR MeanX = AVERAGE(totalData[Оцінка щастя])

VAR MeanY = AVERAGE(totalData[ВВП на душу населення])

VAR cov =

SUMX(

totalData,

(totalData[Оцінка щастя] - MeanX) \* (totalData[ВВП на душу населення] - MeanY)

) / (COUNT(totalData[Оцінка щастя]) - 1)

VAR StdDevX = STDEV.P(totalData[Оцінка щастя])

VAR StdDevY = STDEV.P(totalData[ВВП на душу населення])

RETURN

IF(

StdDevX \* StdDevY = 0,

BLANK(),

cov / (StdDevX \* StdDevY)

)

За аналогічною формулою формуємо всі подібні міри.

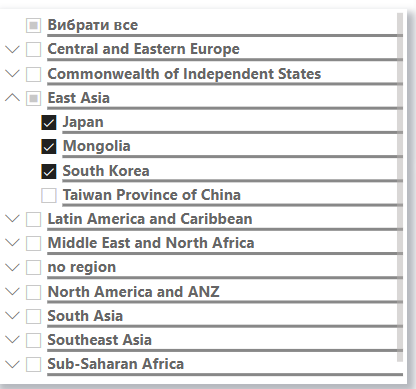
**Використання роздільників**

У процесі аналітичної роботи з даними World Happiness Report важливим етапом є оптимізація структури моделі даних. Зокрема, винесення інформації про країни та їхні регіони в окрему таблицю є доцільним як з точки зору організації даних, так і для підвищення ефективності подальшого аналізу.

Однією з основних причин такого підходу є усунення дублювання даних. Оскільки інформація про країну та її регіон часто повторюється у вихідному наборі даних, створення окремої таблиці, що містить унікальні зв’язки між країною та її регіоном, дозволяє зменшити обсяг даних у основній таблиці. Це не лише знижує навантаження на модель, але й підвищує швидкість обробки інформації під час виконання запитів чи побудови звітів.

Структурування даних за допомогою окремої таблиці "Країна-регіон" також сприяє гнучкості аналітики. Така таблиця спрощує внесення змін у класифікацію країн або додавання нових даних без потреби модифікації основних джерел. Це особливо актуально для динамічних звітів, де є потреба у швидкому оновленні або адаптації структури даних.

Поділ на країни та регіони створює ефективний механізм для сегментації та фільтрації даних. У аналітичних звітах регіони виступають як ключовий роздільник, що дозволяє виконувати порівняльний аналіз між різними частинами світу. Наприклад, можна легко оцінити середній рівень щастя, ВВП або соціальної підтримки для окремих регіонів або груп країн. Це надає змогу дослідникам зосередитися на певних регіональних аспектах без необхідності вручну сортувати дані Рисунок 14.

Рисунок 14

Крім того, таблиця "Країна-регіон" сприяє ефективній географічній візуалізації. У звітах можна використовувати інтерактивні карти, де кожен регіон або країна виділені відповідно до рівня щастя чи інших показників. Такі карти дозволяють миттєво оцінити глобальні та регіональні тенденції, виявляючи відмінності між регіонами та країнами. Це підвищує доступність аналітики для широкого кола користувачів і полегшує представлення результатів.

Таким чином, виділення таблиці "Країна-регіон" у структурі даних є важливим підходом для покращення організації даних, оптимізації обчислень та підвищення якості аналітичних звітів. Це рішення забезпечує ефективну обробку даних, підтримує точність розрахунків і сприяє створенню зрозумілих та інформативних візуалізацій.

Аналогічно до таблиці "Країна-регіон", винесення інформації про роки в окрему таблицю є логічним продовженням оптимізації структури моделі даних. Цей підхід дозволяє ефективно працювати з часовим виміром, що є невід’ємною складовою аналізу змін рівня щастя та інших показників у динаміці Рисунок 15 .

Рисунок 15

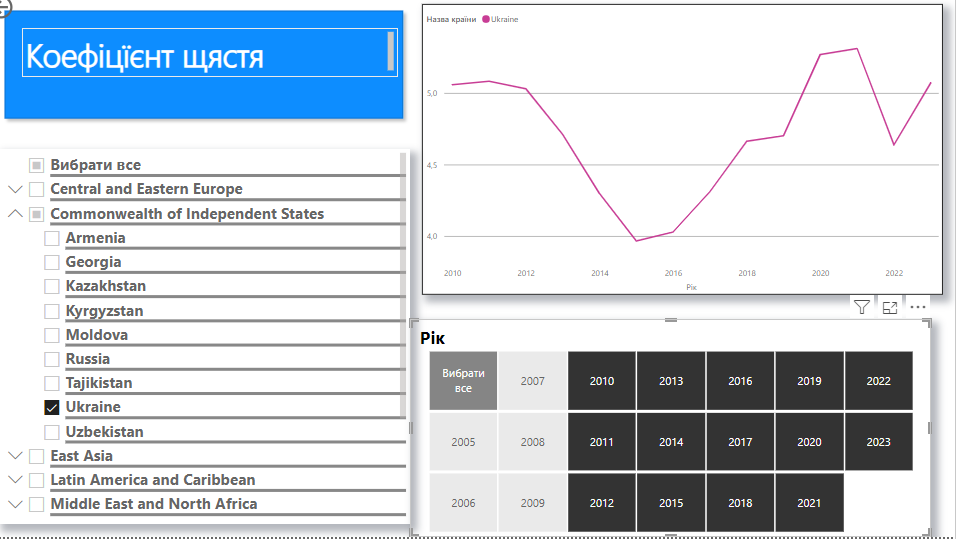
Часовий вимір відіграє ключову роль у трендових аналізах та дослідженні змін у соціально-економічних показниках. Завдяки окремій таблиці "Роки" можна легко виконувати порівняння між різними періодами, виявляти довгострокові тенденції та оцінювати вплив подій у конкретні роки. Наприклад, це дозволяє зосередитися на змінах у рівні щастя внаслідок економічних чи соціальних трансформацій.

**Робота зі звітом**

Для розуміння порядку роботи зі звітом створена діаграма діяльності (Додаток 1). Вона відображає відображає основні ключові етапи, за допомогою яких користувач або група користувачів взаємодіють із звітами, забезпечуючи прозорість та зрозумілість робочих потоків. Діаграма наочно показує, як рухається інформація через різні фази обробки та дозволяє зрозуміти, які саме дії повинні бути виконані на кожному етапі.

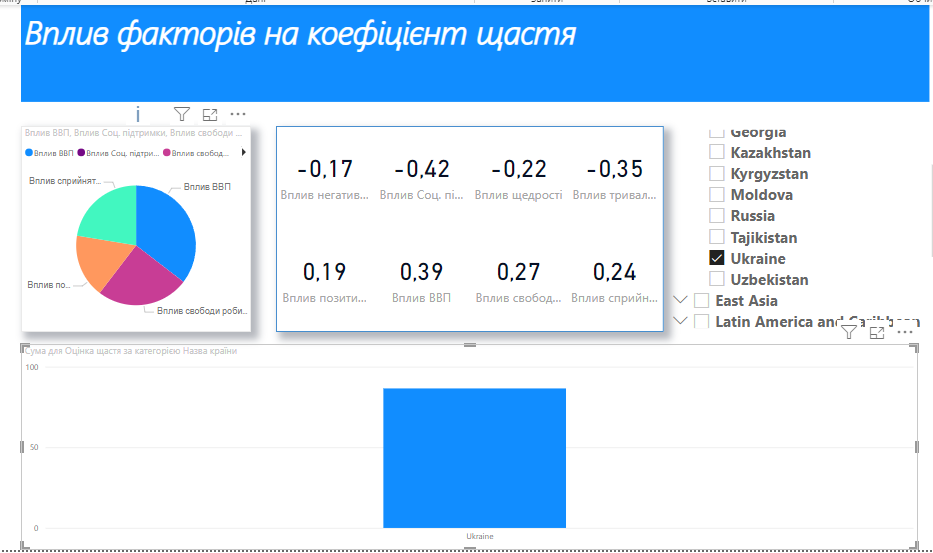
Спробуємо зробити певний аналіз за цією діаграмою. В якості задачі візьмемо аналіз показнику щастя в Україні з 2010 по 2023 роки та фактори впливу на нього. Перші пункти щодо оновлення та перевірки цілісності можемо опустити, так як дані вже підготовлені.

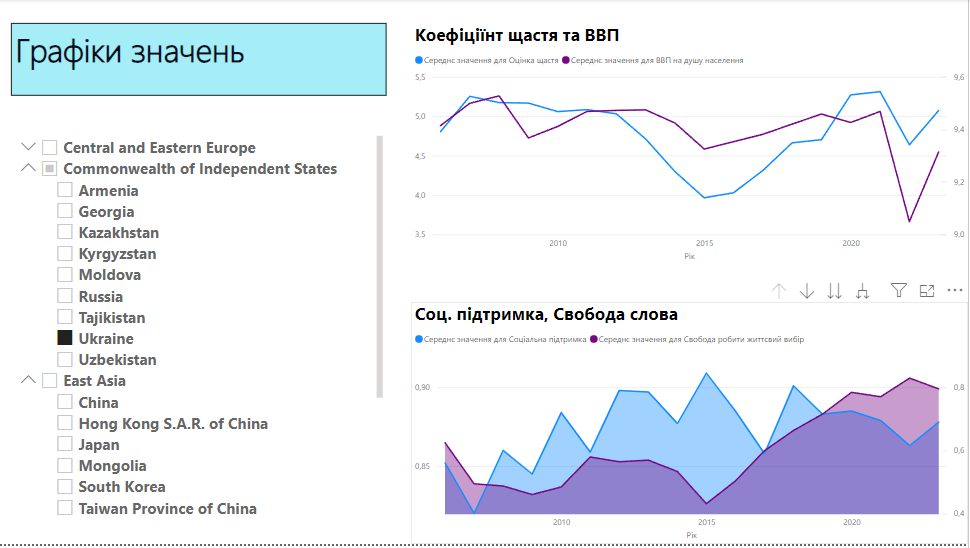
Аналізуємо діаграму рівня щастя за вказаний період Рисунок 16:

Рисунок 16

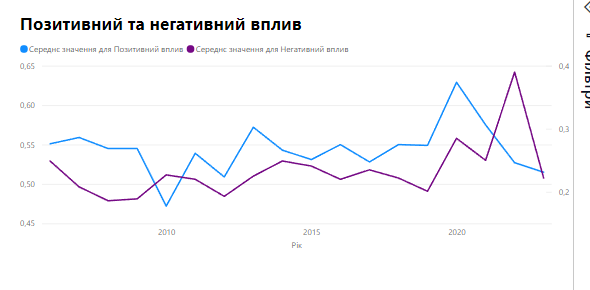
Відповідно до отриманих даних можемо сказати що рівень щастя змінювався за досить об’єктивними чинниками: Російська агресія 2014 року та повномасштабне вторгнення 2022 дають чіткі провали.

Однак 2020-2021 роки за графіком характеризуються найвищими показниками за період. Спробуємо дослідити чому. Визначимо найбільш впливові фактори на рівень щастя для України Рисунок 17:

Рисунок 17

Інтерпретуючи звіт можна сказати що найбільш впливовим на значення є значення ВВП. А збільшення соціальної підтримки навпаки зменшує рівень щастя. Перевіримо це на графіках:

Судячи з графіку існує певний зв’язок між рівнем щастя та ВВП на душу населля, але в 2020, ймовірно рівень щастя збільшився не від нього. Щоправда графік впливу настрою населення підтверджує стрибок як негативних емоцій так і позитивних. Тому, можливо причину слід шукати саме в цьому. Рисунок 18

Рисунок 18

# **РОЗДІЛ 4. “АНАЛІЗ РЕЗУЛЬТАТУ ЗВІТУ”**

**Характеристика отриманого звіту**

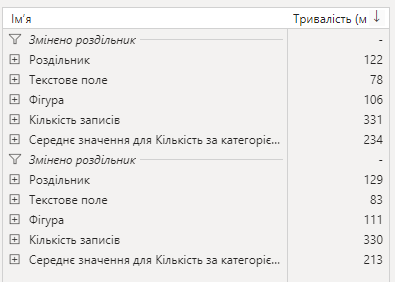
Отриманий звіт є досить актуальним на сьогоднішній час оскільки він дозволяє не тільки дозволяє розглянути рівень щастя населення в різних країнах а також зробити певні висновки щодо факторів впливу на цього. Тако звіт дозволяє зробити певні висновки щодо вхідних даних, їх кількості і розподілу по часовим межам, що дозволяє оцінити їх загальну вагомість. До прикладу можемо розглянути кількість зареєстрованих значень в період з 2018 — по 2023 роки. Даний аналіз показує різке зменшення кількості записів, що зумовлене об’єктивними причинами впливу пандемії COVID19. Рисунок 19.

Рисунок 19

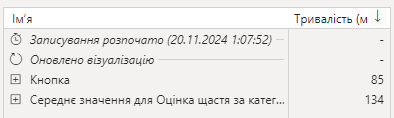
**Аналіз продуктивності звіту**

Сучасні тенденції в комп’ютерних технологіях зумовлюють розробників дуже ретельно відноситись до швидкодії свого продукту. Тому аналіз продуктивності звіту є досить важливим фактором впливу на отримання позитивного UX користувача під час його використання. Проведемо аналіз продуктивності кожної сторінки звіту.

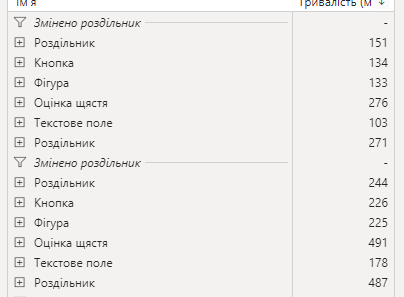
Результат зміни роздільника на сторінці «Кількість записів»



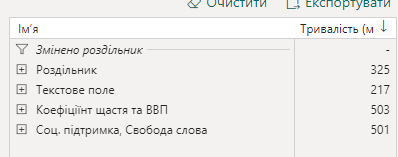
Результат оновлення візуалізації мапи рівня щастя.



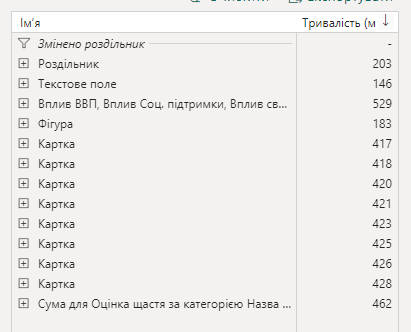
Результат зміни роздільника на графіку Коефіцієнт щастя



Результат зміни роздільника на сторінці Графіки значень



Результат зміни роздільника на сторінці Кореляція значень



Можна встановити, що час виконання дій не перевищує середнє значення в 500мс, що є досить непогано, і швидше за все залежить від продуктивності ПК.

# **ВИСНОВКИ**

У ході дослідження було проведено комплексний аналіз зведеної статистики показників щастя у світі, що дозволило визначити основні фактори, які впливають на рівень добробуту населення. Дослідження показало, що рівень щастя у країнах значною мірою залежить від таких економічних показників, як ВВП на душу населення, рівень соціальної підтримки, очікувана тривалість життя, а також соціальних чинників, зокрема рівня корупції та соціальної гармонії.

Аналіз продемонстрував, що існують тісні причинно-наслідкові зв’язки між економічними, соціальними та політичними факторами, які визначають якість життя у різних регіонах світу. Використання описової та діагностичної аналітики дозволило глибше зрозуміти структуру даних, виявити закономірності та тренди, а також встановити взаємозв’язок між ключовими показниками.

Під час дослідження виникли певні труднощі, такі як пропущені дані та невідповідність форматів, які потребували додаткових зусиль для підготовки та обробки інформації. Застосування сучасних методів обробки даних забезпечило якість та достовірність отриманих результатів.

Отримані висновки підкреслюють важливість врахування нематеріальних показників у розробці стратегій сталого розвитку. Актуальність цього підходу зумовлена необхідністю підвищення рівня добробуту через мультидисциплінарний аналіз, що включає економічні, соціальні, культурні та психологічні аспекти.

У результаті дослідження був сформований звіт за даними World Happiness Report, який дозволить подальший продуктивний аналіз показників щастя в країнах. Звіт також дозволяє визначити певні залежності коефіцієнту щастя від інших показників соціального добробуту і дозволить зосередитись на найбільш впливових чинниках задля забезпечення покращення проживання суспільства.

# **СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ**

1. Kaggle. World Happiness Report 2023. URL: <https://www.kaggle.com/datasets/atom1991/world-happiness-report-2023> .

2. World Happiness Report 2023. Statistical Appendix. URL: <https://happiness-report.s3.amazonaws.com/2023/WHR+23_Statistical_Appendix.pdf>.

**Додаток 1 Діагама роботи зі звітом**

