Applied Mathematics and Computational Intelligence
Volume 14, No. 2, 2025 [119-134]

The Development of CancerAtlas: Data Visualization on Number of Cancer Deaths Worldwide Analysis

Isyraf Iskandar[1], Norisan Abd Karim[2*], Fauziah Redzuan[3], Rogayah Abdul Majid[4]

[1,2,3,4]College of Computing, Informatics and Mathematics, MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia

[*] Corresponding author : norisan@tmsk.uitm.edu.my

Received: 1 October 2024
Revised: 21 February 2025
Accepted: 24 February 2025

ABSTRACT

Global mortality statistics demonstrate that cancer emerges as a lethal disease which affects mortality numbers extensively. The global understanding of this illness remains complicated by the multiple forms of cancer together with their variable death rates among different areas. Advanced data analytics form the basis of the CancerAtlas initiative as it uses visualization techniques to reveal worldwide cancer mortality patterns throughout various regions. The CancerAtlas dashboard enables users to perform interactive evaluation and comparison and analysis of cancer death rates among sequences of geographic areas and age groups and cancer types. The testing phase validates both the dashboard's intuitive interface and the predictive system which presents complex data information to multiple user types. The project highlights how data visualization methods together with analytics and predictive modeling increase public understanding about worldwide cancer pattern development. Future development should implement real-time data acquisition to show current visualizations and generate predictions that will improve the quality of extracted knowledge.

Keywords: Cancer mortality rates, Data visualization, Predictive analysis, Data analytics

1 INTRODUCTION

Cancer remains a significant global health issue, with millions of individuals affected by different types of cancer each year [1]. According to Bray et al. [2], cancer is among the leading causes of death around the world, with around 10 million deaths recorded in 2020 alone.

The use of Statistical Analysis even has limitations that affect the understanding insights of stakeholders especially when it comes to big data. Complex interactions between various variables, such as genetic, environmental, and lifestyle factors, which might influence cancer development and outcomes, may be difficult for statistical models to represent [3].

In order to solve these problems, CancerAtlas project elaborates Power BI dashboard which gives the user an opportunity to use a map, graphs, and tables for data analysis of global cancer mortalities. The specifics of the project's scope are as follows: Determining the needs and requests of the system regarding data visualization, as well as the design of adequate visualizations and the implementation of the CancerAtlas dashboard. This solution will thus seek to improve the comprehension of the data to aid in decision making and policies focusing on cancer control.

Thus, the application of data visualization with the help of tools such as Power BI can help overcome challenges and improve the work of many participants involved in the process. Converting the collected and entered data into more presentable charts and graphs in the form of dashboards enables users to understand what the data is conveying in detail and make intelligent decisions concerning cancer control and management.

1.1 Big Data

Big data is an expanding vast collection of information according to Kumar et al. [4]. Standard data management systems fail to handle or process such information because of its large size and complex nature. According to Segal [5] big data operates as a system which handles structured along with unstructured information. Creating orders from information occurs with structured data which companies traditionally kept in databases alongside spreadsheets as numeric values. Social media-derived data fits under the unstructured category because it lacks defined organizational standards yet provides organizations insight into customer behavior.

Researcher Favaretto [6] describes how scholars actively use vast datasets because of their ability to detect collective behavioral patterns together with promising data analytic potentials. Cancer caused 10 million worldwide deaths during 2020, and analysts expect this number to reach 13 million by 2030. Research with big data has shown that it has potential to decrease cancer fatality rates by enabling scientists to find novel cancer patterns which generate better disease treatments. These technological advancements have the potential to conserve numerous millions of human lives today. Using big data creates multiple substantial ethical challenges to be addressed.

The authors Vayena and Blasimme [7] highlight privacy breaches coupled with informed consent requirements along with data security challenges as core ethical matters affecting big data operations. The protection of patient privacy and General Data Protection Regulation (GDPR) compliance demands sensitive health information like cancer mortality rates to be properly anonymized before being processed. Protected data requires additional consent from participants whose information gets used while security frameworks need to establish strong protocols to block unauthorized data access.

1.1.1 Data Visualization

Visual big data analytics consists of transforming large data sets into graphical presentations that help improve understanding. Experts who specialize in SQL queries actually choose visual presentations over traditional tabular displays according to research [8]. The process of transforming intricate data through visualization methods enables users to interpret complex information effectively. The methods establish visual experiences that are both interactive and engaged for users.

The article by Rhee et al. [9] presents advanced visualization methods across different fields and Few [10] explains how data visualization delivers efficient information presentation combined with pattern detection capabilities. Interactive dashboards together with heatmaps and dynamic filtering form part of advanced techniques deployed by the CancerAtlas project to display worldwide cancer mortality rates. Users can gain deeper insights into data through these innovations because they detect hidden patterns that static charts normally miss.

Comparison with other available tools found in references [11] and [12] serves to demonstrate the project's accuracy for generating insights. The process of data visualization demands a high level of ethical awareness because health-related information requires special protection.

Kitchin [13] emphasizes ethical visualization relies on maintaining accurate data representations together with protection of privacy while preventing misleading results. The visual representations must maintain transparency along with inclusion values that protect stakeholders from bias and allow all users access to the data. The CancerAtlas project achieves better data understanding alongside improved decision-support abilities along with enhanced cancer research by including state-of-the-art approaches and ethical concerns.

2 MATERIAL AND METHODS

Both Agile Software Development and OSEMN framework share fundamental concepts which help in the design and assessment of data visualization systems. Agile defines working software while promoting developer-customers' communication until products achieve working status through a progression of 'sprints' or 'iterations' [14]. Based on Figure 1, use of OSEMN framework guides data science work by providing steps for Obtain, Scrub, Explore, Model and iNterpret which correspond with data acquisition and data cleaning and data analysis and data modeling and data interpretation processes. The combination of Agile and OSEMN methodologies makes development processes adaptable while producing numerous data refinement and analysis cycles. The method provides precise information which meets time requirements thus enhancing the total information quality. Data acquisition paired with preprocessing (Obtain and Scrub) happens first followed by assessment (Explore) to obtain insights before the data is used to build and validate models (Model) and ends with visual interpretation (Interpret). The fusion of OSEMN structured decision-making process with Agile's adaptive methodology enables the development of dependable user-centered data visualization solutions. The interrelation briefly states based on Table 1.

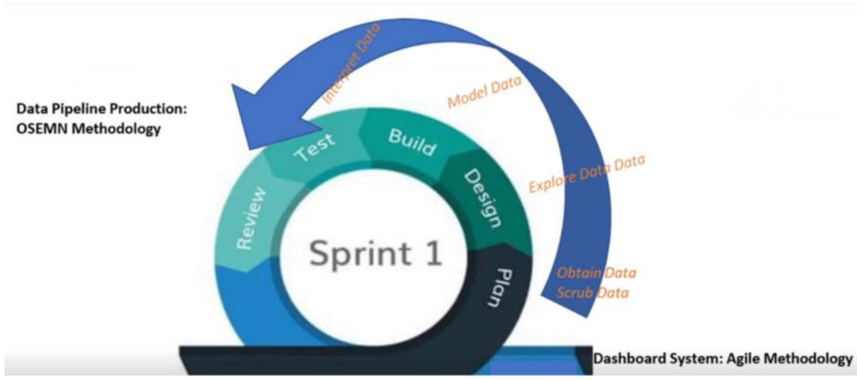


Figure 1: Shows the Agile combined with OSEMN Model Methodology.

Table 1: Methodology Phases

Agile	OSEMN	Description
Plan	Obtain Data (Identification of data requirement)	Collecting global cancer-related datasets from sources like WHO and IARC and identifying data requirements.
	Scrub Data (Data Cleaning)	Clean raw data by modifying and deleting data that is inaccurate, duplicated, or incomplete within the dataset.
Design	Explore (Storyboard Design)	To better comprehend the data and its story, explore the dataset using various techniques such as examining initial rows and identifying similar columns. Build a storyboard to show the usage of the CancerAtlas system.
Build	Model (Prototype of Data Visualization)	Use the cleaned dataset as model data to develop the CancerAtlas dashboard using Power BI.
Test	Interpret	Test the CancerAtlas dashboard system using feedback collected through surveys or user testing session and calculated via SUS score
Review		Review CancerAtlas based on user testing and feedback for further improvements.

2.1 Identification of data requirement

The 5V concepts (volume, velocity, veracity, variety, value) are closely related to big data discussions. This project deals with large amounts of data, especially related to cancer deaths cases.

This extensive dataset is provided by Belayet HossainDS through Kaggle Open-Source dataset platform, an initiative towards open data access through free online data sharing. For the purposes of this project, we used data collected from 1990-2019[15]. Data is provided in shapefile format in Comma-separated values (CSV). This dataset is comprehensive data of total number of cancer deaths by variety of type and age rates for every country around the world collected by World Health Organization (WHO) [16]. The range of data was collected from 1990 until 2019.

This first dataset is “Total Cancer Deaths by Type” that incorporates country, year, cancer type, and total deaths. It can be inferred that this dataset is a time-series type, with data collected in every from 1990-2019 despite the Year format is a character attribute. Besides, the total deaths value is cumulative, and it requires data cleaning and to find the actual deaths for every country. Similarly, the second dataset is “Cancer Deaths Rates by Age Category” that incorporates country, year, age category and deaths rates. To fulfill 5Vs (velocity, variety, value, and veracity), data cleaning needs to be implemented to make this data more understanding.

Volume, the first of the 5 Vs, signifies the amount of data to be used in this project. As described, the dataset was collected and posted to the Kaggle open-source dataset. The term 'velocity' refers to the speed at which the data has been collected every year by every country in the world and has 198 630 rows of data. This data is expected to continually grow and fluctuate yearly based on the cancer deaths cases that happen every year for each country.

'Variety' denotes whether the data is structured or unstructured. In this case, the dataset collected is structured in Excel format. Consequently, it needs to be converted to a CSV file to make it more readable. 'Value' is the third 'V', indicating how valuable the data can be to society and stakeholders. Apart from gain highest case death in every country, this data even is usable for trend insights and forecast value for research and knowledge for society.

'Veracity' pertains to the validity of the data source, ensuring the quality and credibility of the collected data. Given that data can be found from various sources, there's a chance that data collected from unknown sources may lack quality and accuracy. For this project, it is confirmed that the dataset comes from a credible platform.

Finally, 'variety' refers to the nature of the dataset, whether it's structured, semi-structured, or unstructured. From this project's dataset, it's determined that the dataset is structured, containing data collected in date format and total deaths to number.

2.2 Data Cleaning & Complexity Developing Cancer Atlas (Phyton, Power Query & DAX)

The initial step in the data cleaning process is executing the load operation. As indicated in the above figures, the 'read_excel' function is applied to the 2 datasets files to import the data from the local system. An alternative function, 'read_csv', can also be used for this purpose. Once the data is imported, the 'print' function generates an output presenting the first five rows, the last five rows, the total row count, and the total column count of the data.

```
In [3]: #load data file
        #Total worldwide cancer deaths from 1990-2019
        cancer_deaths=pd.read_csv('total_cancer_deaths_by_type.csv')
        print(cancer_deaths)

6836                                        40
6837                                        41
6838                                        43
6839                                        44

     Deaths - Mesothelioma - Sex: Both - Age: All Ages (Number)
0                                            6
1                                            7
2                                            7
3                                            7
4                                            8
...                                        ...
6835                                        16
6836                                        17
6837                                        17
6838                                        17
6839                                        18

[6840 rows x 32 columns]
```

Figure 2: Load Dataset-Total Cancer Deaths by Type

```
In [4]: #load data file
        #Total worldwide cancer deaths rates by age from 1990-2020
        cancer_age=pd.read_csv('cancer_death_rates_by_age.csv')
        print(cancer_deaths)

6836                                        40
6837                                        41
6838                                        43
6839                                        44

     Deaths - Mesothelioma - Sex: Both - Age: All Ages (Number)
0                                            6
1                                            7
2                                            7
3                                            7
4                                            8
...                                        ...
6835                                        16
6836                                        17
6837                                        17
6838                                        17
6839                                        18

[6840 rows x 32 columns]
```

Figure 3: Load Dataset-Total Cancer Deaths by Type

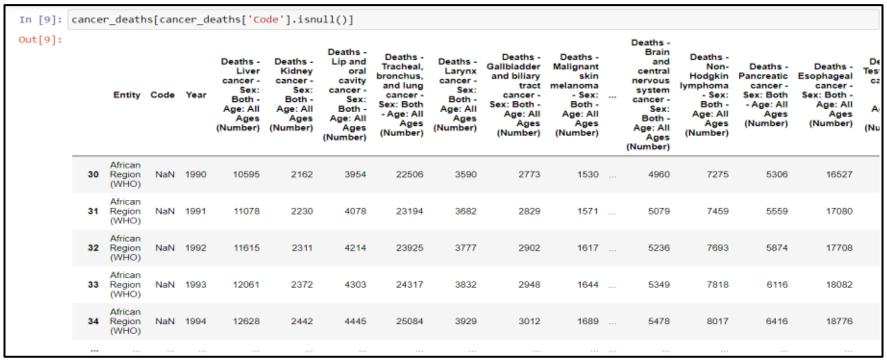
The detection process should follow with null value replacement during this step. Finding null values in the data relies on the 'isnull()' function while the 'sum()' function determines the number of rows containing null data. değerli missing codes from regions and countries are identified then substituted by the 'fillna' function utilizing ISO country codes. Verification occurs for the substituted values. The vital step enhances dataset precision and avoids biases through proper handling of incomplete information. Figure 4,65,6 and 7 explain briefly the method of processing.

```
In [5]: #Find missing values in th dataframe
        cancer_deaths.isnull().sum().sum()
        print('Missing Values in cancer deaths by type dataset:',cancer_deaths.isnull().sum().sum(), 'rows\n')
        cancer_age.isnull().sum().sum()
        print('Missing Values in cancer deaths by age dataset:',cancer_age.isnull().sum().sum(), 'rows\n')

        Missing Values in cancer deaths by type dataset: 690 rows

        Missing Values in cancer deaths by age dataset: 690 rows
```

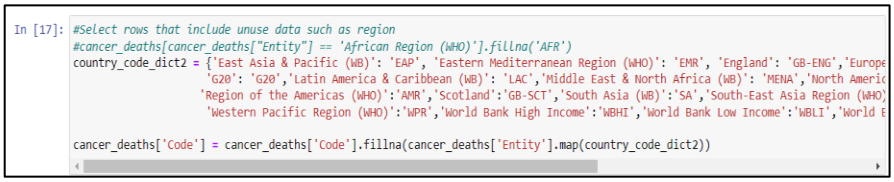
Figure 4: Sum of Null Values



```
In [9]: cancer_deaths[cancer_deaths['Code'].isnull()]
Out[9]:
```

	Entity	Code	Year	Deaths - Liver cancer - Sex: Both - Age: All Ages (Number)	Deaths - Kidney cancer - Sex: Both - Age: All Ages (Number)	Deaths - Lip and oral cavity cancer - Sex: Both - Age: All Ages (Number)	Deaths - Tracheal, bronchus, and lung cancer - Sex: Both - Age: All Ages (Number)	Deaths - Larynx cancer - Sex: Both - Age: All Ages (Number)	Deaths - Gallbladder and biliary tract cancer - Sex: Both - Age: All Ages (Number)	Deaths - Malignant skin melanoma - Sex: Both - Age: All Ages (Number)	...	Deaths - Brain and central nervous system cancer - Sex: Both - Age: All Ages (Number)	Deaths - Non-Hodgkin lymphoma - Sex: Both - Age: All Ages (Number)	Deaths - Pancreatic cancer - Sex: Both - Age: All Ages (Number)	Deaths - Esophageal cancer - Sex: Both - Age: All Ages (Number)
30	African Region (WHO)	NaN	1990	10595	2162	3954	22506	3590	2773	1530	...	4960	7275	5306	16527
31	African Region (WHO)	NaN	1991	11078	2230	4078	23194	3682	2829	1571	...	5079	7459	5559	17080
32	African Region (WHO)	NaN	1992	11615	2311	4214	23925	3777	2902	1617	...	5236	7693	5874	17708
33	African Region (WHO)	NaN	1993	12061	2372	4303	24317	3832	2948	1644	...	5349	7818	6116	18082
34	African Region (WHO)	NaN	1994	12628	2442	4445	25084	3929	3012	1689	...	5478	8017	6416	18776
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

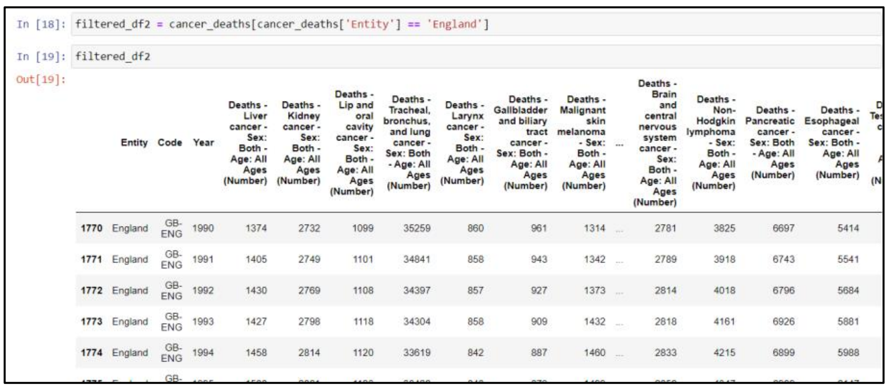
Figure 5: Trace Null Values Column



```
In [17]: #Select rows that include unuse data such as region
         #cancer_deaths[cancer_deaths["Entity"] == 'African Region (WHO)'].fillna('AFR')
         country_code_dict2 = {'East Asia & Pacific (WB)': 'EAP', 'Eastern Mediterranean Region (WHO)': 'EMR', 'England': 'GB-ENG','Europe
                               'G20': 'G20','Latin America & Caribbean (WB)': 'LAC','Middle East & North Africa (WB)': 'MENA','North Americ
                              'Region of the Americas (WHO)':'AMR','Scotland':'GB-SCT','South Asia (WB)':'SA','South-East Asia Region (WHO)
                               'Western Pacific Region (WHO)':'WPR','World Bank High Income':'WBHI','World Bank Low Income':'WBLI','World E

         cancer_deaths['Code'] = cancer_deaths['Code'].fillna(cancer_deaths['Entity'].map(country_code_dict2))
```

Figure 6: Replace Null Values



```
In [18]: filtered_df2 = cancer_deaths[cancer_deaths['Entity'] == 'England']

In [19]: filtered_df2
Out[19]:
```

	Entity	Code	Year	Deaths - Liver cancer - Sex: Both - Age: All Ages (Number)	Deaths - Kidney cancer - Sex: Both - Age: All Ages (Number)	Deaths - Lip and oral cavity cancer - Sex: Both - Age: All Ages (Number)	Deaths - Tracheal, bronchus, and lung cancer - Sex: Both - Age: All Ages (Number)	Deaths - Larynx cancer - Sex: Both - Age: All Ages (Number)	Deaths - Gallbladder and biliary tract cancer - Sex: Both - Age: All Ages (Number)	Deaths - Malignant skin melanoma - Sex: Both - Age: All Ages (Number)	...	Deaths - Brain and central nervous system cancer - Sex: Both - Age: All Ages (Number)	Deaths - Non-Hodgkin lymphoma - Sex: Both - Age: All Ages (Number)	Deaths - Pancreatic cancer - Sex: Both - Age: All Ages (Number)	Deaths - Esophageal cancer - Sex: Both - Age: All Ages (Number)
1770	England	GB-ENG	1990	1374	2732	1099	35259	860	961	1314	...	2781	3825	6697	5414
1771	England	GB-ENG	1991	1405	2749	1101	34841	858	943	1342	...	2789	3918	6743	5541
1772	England	GB-ENG	1992	1430	2769	1108	34397	857	927	1373	...	2814	4018	6796	5684
1773	England	GB-ENG	1993	1427	2798	1118	34304	858	909	1432	...	2818	4161	6926	5881
1774	England	GB-ENG	1994	1458	2814	1120	33619	842	887	1460	...	2833	4215	6899	5988

Figure 7: Check The Replace Value

Next, as the third step, we modify the names of the cancer type column to simpler and more appropriate terms, facilitating easier plotting during visualization on the dashboard. The same approach is applied to the age categories.

```
In [25]: cancer_deaths=cancer_deaths.rename(columns={'Deaths - Kidney cancer - Sex: Both - Age: All Ages (Number)':'Kidney',
                                              'Deaths - Lip and oral cavity cancer - Sex: Both - Age: All Ages (Number)':'Oral',
                                             'Deaths - Tracheal, bronchus, and lung cancer - Sex: Both - Age: All Ages (Number)':'
                                             'Deaths - Larynx cancer - Sex: Both - Age: All Ages (Number)':'Larynx',
                                             'Deaths - Gallbladder and biliary tract cancer - Sex: Both - Age: All Ages (Number)':'
                                             'Deaths - Malignant skin melanoma - Sex: Both - Age: All Ages (Number)':'Malignant Ski
                                             'Deaths - Leukemia - Sex: Both - Age: All Ages (Number)':'Leukemia',
                                             'Deaths - Hodgkin lymphoma - Sex: Both - Age: All Ages (Number)':'Hodgkin Lymphoma',
                                             'Deaths - Multiple myeloma - Sex: Both - Age: All Ages (Number)':'Multiple Myeloma',
                                             'Deaths - Other neoplasms - Sex: Both - Age: All Ages (Number)':'Other Neoplasms',
                                             'Deaths - Breast cancer - Sex: Both - Age: All Ages (Number)':'Breast',
                                             'Deaths - Prostate cancer - Sex: Both - Age: All Ages (Number)':'Prostate',
                                             'Deaths - Thyroid cancer - Sex: Both - Age: All Ages (Number)':'Thyroid',
                                             'Deaths - Stomach cancer - Sex: Both - Age: All Ages (Number)':'Stomach',
                                             'Deaths - Bladder cancer - Sex: Both - Age: All Ages (Number)':'Bladder',
                                             'Deaths - Uterine cancer - Sex: Both - Age: All Ages (Number)':'Uterine',
                                             'Deaths - Ovarian cancer - Sex: Both - Age: All Ages (Number)':'Ovarian',
                                             'Deaths - Cervical cancer - Sex: Both - Age: All Ages (Number)':'Cervical',
                                             'Deaths - Brain and central nervous system cancer - Sex: Both - Age: All Ages (Number)
                                             'Deaths - Non-Hodgkin lymphoma - Sex: Both - Age: All Ages (Number)':'Non-Hodgkin Lymp
                                             'Deaths - Pancreatic cancer - Sex: Both - Age: All Ages (Number)':'Pancreatic',
                                             'Deaths - Esophageal cancer - Sex: Both - Age: All Ages (Number)':'Esophageal',
                                             'Deaths - Testicular cancer - Sex: Both - Age: All Ages (Number)':' Testicular',
                                             'Deaths - Nasopharynx cancer - Sex: Both - Age: All Ages (Number)':'Nasopharynx',
                                             'Deaths - Other pharynx cancer - Sex: Both - Age: All Ages (Number)':'Other Pharynx',
                                             'Deaths - Colon and rectum cancer - Sex: Both - Age: All Ages (Number)':'Colon and Rec
                                             'Deaths - Non-melanoma skin cancer - Sex: Both - Age: All Ages (Number)':'Non-melanoma
                                             'Deaths - Mesothelioma - Sex: Both - Age: All Ages (Number)':'Mesothelioma'})
```

Figure 8: Rename Column

To consolidate all the cancer types from the first dataset into one column, we apply the 'unpivot column' technique in Power BI's Power Query. The 'unpivot' function essentially converts multiple columns into rows, yielding a more streamlined and succinct view of the data. The same approach is used for the age category data in the second dataset. The 'unpivot column' operation can be performed by selecting one or multiple columns that don't need to be unpivoted, then choosing the 'unpivot other columns' option in the transform tab.

Before:

```
= Table.TransformColumnTypes(#"Promoted Headers",{{"Entity", type text}, {"Code", type text}, {"Year", Int64.Type}, {" Liver ", Int64.Type}
```

	Entity	Code	Year	Liver	Kidney	Oral	Tracheal
1	Afghanistan	AFG	1990	851	66	89	
2	Afghanistan	AFG	1991	866	66	89	
3	Afghanistan	AFG	1992	890	68	91	
4	Afghanistan	AFG	1993	914	70	93	
5	Afghanistan	AFG	1994	933	71	94	
6	Afghanistan	AFG	1995	946	71	94	
7	Afghanistan	AFG	1996	962	72	95	
8	Afghanistan	AFG	1997	976	73	95	
9	Afghanistan	AFG	1998	985	73	96	
10	Afghanistan	AFG	1999	994	73	96	
11	Afghanistan	AFG	2000	1002	73	96	
12	Afghanistan	AFG	2001	1018	74	96	
13	Afghanistan	AFG	2002	1019	75	96	
14	Afghanistan	AFG	2003	1036	80	97	
15	Afghanistan	AFG	2004	1054	83	98	
16	Afghanistan	AFG	2005	1064	85	99	
17	Afghanistan	AFG	2006	1057	87	100	
18	Afghanistan	AFG	2007	1066	89	101	
19	Afghanistan	AFG	2008	1079	91	103	
20	Afghanistan	AFG	2009	1092	94	105	
21	Afghanistan	AFG	2010	1111	99	107	
22	Afghanistan	AFG	2011	1130	103	109	
23	Afghanistan	AFG	2012	1155	107	112	
24	Afghanistan	AFG	2013	1181	112	115	
25	Afghanistan	AFG	2014	1210	117	118	
26	Afghanistan	AFG	2015	1238	122	122	
27	Afghanistan	AFG	2016	1271	127	126	

Figure 9: The Data Structure Before the Unpivot Process

Process:

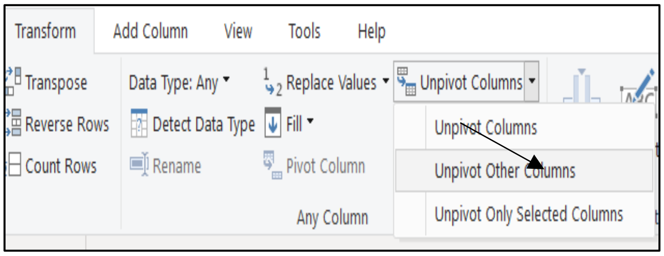


Figure 10: Shows the Process to Unpivot Column

After:

	Entity	Code	Year	Cancer Type	Deaths
1	Afghanistan	AFG	1990	Liver	851
2	Afghanistan	AFG	1990	Kidney	66
3	Afghanistan	AFG	1990	Oral	89
4	Afghanistan	AFG	1990	Tracheal, Bronchus and Lung	983
5	Afghanistan	AFG	1990	Larynx	260
6	Afghanistan	AFG	1990	Gallbladder and Biliary Tract	180
7	Afghanistan	AFG	1990	Malignant Skin Melanoma	47
8	Afghanistan	AFG	1990	Leukemia	1055
9	Afghanistan	AFG	1990	Hodgkin Lymphoma	102
10	Afghanistan	AFG	1990	Multiple Myeloma	108
11	Afghanistan	AFG	1990	Other Neoplasms	16
12	Afghanistan	AFG	1990	Breast	526
13	Afghanistan	AFG	1990	Prostate	351
14	Afghanistan	AFG	1990	Thyroid	56
15	Afghanistan	AFG	1990	Stomach	2200
16	Afghanistan	AFG	1990	Bladder	263
17	Afghanistan	AFG	1990	Uterine	61
18	Afghanistan	AFG	1990	Ovarian	79
19	Afghanistan	AFG	1990	Cervical	332
20	Afghanistan	AFG	1990	Brain and Central Nervous system	422
21	Afghanistan	AFG	1990	Non-Hodgkin Lymphoma	996
22	Afghanistan	AFG	1990	Pancreatic	138
23	Afghanistan	AFG	1990	Esophageal	529
24	Afghanistan	AFG	1990	Testicular	3
25	Afghanistan	AFG	1990	Nasopharynx	66
26	Afghanistan	AFG	1990	Other Pharynx	37
27	Afghanistan	AFG	1990	Colon and Rectum	539
28	Afghanistan	AFG	1990	Non-melanoma skin	25

Figure 11: Table Structures After the Unpivot Process

Given that the raw dataset only provides cumulative total deaths for each category year by year, a measure is created in Power BI using DAX (Data Analysis Expressions) query to compute the actual total deaths for each type of cancer. Based on Figure 12 and 13, a measure in DAX is a formula that is created specifically for data analysis and is used in the Values area of a PivotTable or PivotChart for aggregated calculations. Consequently, this allows for the analysis and visualization of differing patterns and trends. As a result of this process, an additional column named "Actual Deaths" is introduced.

```
Actual Deaths =
    'cancer_deaths_type'[Total Deaths] -
    CALCULATE(
        SUM('cancer_deaths_type'[Total Deaths]),
        FILTER(
            'cancer_deaths_type',
            'cancer_deaths_type'[Entity] = EARLIER(cancer_deaths_type[Entity]) &&
            'cancer_deaths_type'[Code] = EARLIER('cancer_deaths_type'[Code]) &&
            'cancer_deaths_type'[Year] = EARLIER('cancer_deaths_type'[Year]) - 1 &&
            'cancer_deaths_type'[Cancer Type] = EARLIER('cancer_deaths_type'[Cancer Type])

        )
    )
```

Figure 12: Formula of Data Value In "Actual Deaths" Column

Entity	Code	Year	Cancer Type	Total Deaths	Year Date	Actual Deaths
Malaysia	MYS	2000	Stomach	1109	2000	-47
Malaysia	MYS	2000	Leukemia	942	2000	-46
Malaysia	MYS	2000	Nasopharynx	906	2000	-40
Malaysia	MYS	2000	Cervical	761	2000	-40
Malaysia	MYS	2000	Colon and Rectum	2052	2000	-31
Malaysia	MYS	2000	Breast	1617	2000	-30
Malaysia	MYS	1998	Prostate	417	1998	-25
Malaysia	MYS	2013	Stomach	1285	2013	-19
Malaysia	MYS	1996	Stomach	1057	1996	-18
Malaysia	MYS	2013	Tracheal, Bronchus and Lung	4166	2013	-17
Malaysia	MYS	2005	Nasopharynx	874	2005	-15
Malaysia	MYS	1996	Cervical	734	1996	-14
Malaysia	MYS	2013	Leukemia	1160	2013	-12
Malaysia	MYS	2001	Nasopharynx	894	2001	-12
Malaysia	MYS	2002	Nasopharynx	882	2002	-12
Malaysia	MYS	2000	Oral	335	2000	-11
Malaysia	MYS	2000	Brain and Central Nervous system	307	2000	-10
Malaysia	MYS	1997	Stomach	1047	1997	-10
Malaysia	MYS	1998	Stomach	1037	1998	-10
Malaysia	MYS	2000	Larynx	177	2000	-10
Malaysia	MYS	1998	Bladder	283	1998	-9
Malaysia	MYS	2000	Non-Hodgkin Lymphoma	597	2000	-9
Malaysia	MYS	2000	Esophageal	344	2000	-9
Malaysia	MYS	2010	Stomach	1269	2010	-8
Malaysia	MYS	2004	Nasopharynx	889	2004	-8
Malaysia	MYS	2006	Nasopharynx	866	2006	-8
Malaysia	MYS	2010	Liver	1362	2010	-8
Malaysia	MYS	2004	Cervical	768	2004	-8

Figure 13: New Column (Actual Deaths) Added to The Database

Output:

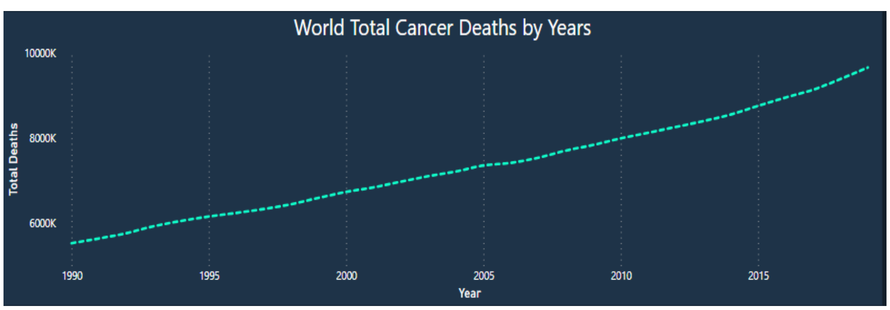


Figure 14: Line Graph of World Total Cancer Deaths by Years (1990-2019)

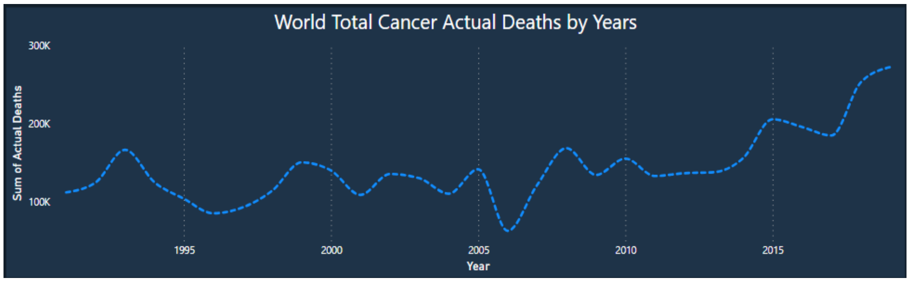


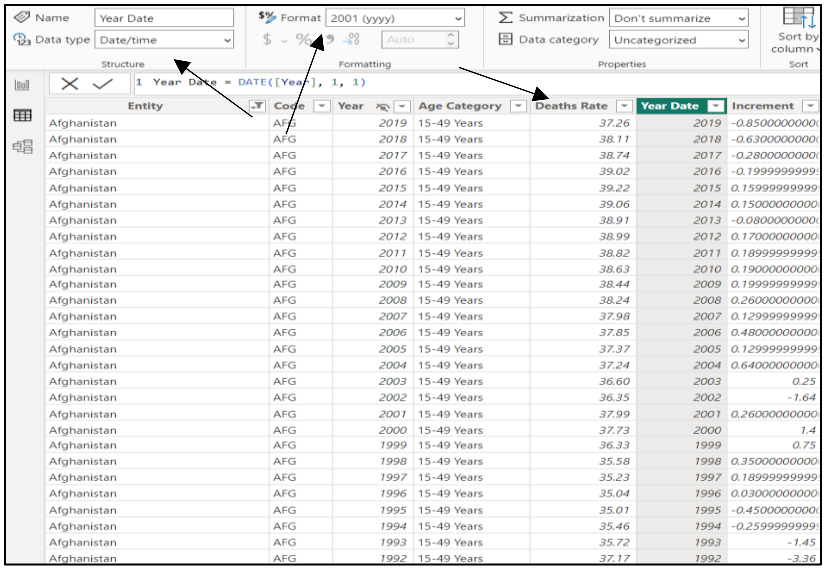
Figure 15: Line Graph of World Total Cancer Actual Deaths by Years (1990-2019)

The outputs in Figures 14 and 15 reveal unique patterns. Figure 14 shows a consistent annual increase in cumulative deaths, while Figure 15 displays an erratic pattern with fluctuating increases and decreases in actual deaths. This allows users to gain more extensive information about total cancer deaths by type.

Based on Figure 16 and 17, considering that the 'Year' column in the raw data is in numerical format, it is converted to a date format to enhance dashboard visualization. This is achieved using DAX in Power BI, with the addition of a new column representing years in date format.

```
Year Date = DATE([Year], 1, 1)
```

Figure 16: DAX Formula to Add a New Column That Represents the Year Column.



Entity	Code	Year	Age Category	Deaths Rate	Year Date	Increment
Afghanistan	AFG	2019	15-49 Years	37.26	2019	-0.8500000000
Afghanistan	AFG	2018	15-49 Years	38.11	2018	-0.6300000000
Afghanistan	AFG	2017	15-49 Years	38.74	2017	-0.2800000000
Afghanistan	AFG	2016	15-49 Years	39.02	2016	-0.1999999999
Afghanistan	AFG	2015	15-49 Years	39.22	2015	0.1599999999
Afghanistan	AFG	2014	15-49 Years	39.06	2014	0.1500000000
Afghanistan	AFG	2013	15-49 Years	38.91	2013	-0.0800000000
Afghanistan	AFG	2012	15-49 Years	38.99	2012	0.1700000000
Afghanistan	AFG	2011	15-49 Years	38.82	2011	0.1899999999
Afghanistan	AFG	2010	15-49 Years	38.63	2010	0.1900000000
Afghanistan	AFG	2009	15-49 Years	38.44	2009	0.1999999999
Afghanistan	AFG	2008	15-49 Years	38.24	2008	0.2600000000
Afghanistan	AFG	2007	15-49 Years	37.98	2007	0.1299999999
Afghanistan	AFG	2006	15-49 Years	37.85	2006	0.4800000000
Afghanistan	AFG	2005	15-49 Years	37.37	2005	0.1299999999
Afghanistan	AFG	2004	15-49 Years	37.24	2004	0.6400000000
Afghanistan	AFG	2003	15-49 Years	36.60	2003	0.25
Afghanistan	AFG	2002	15-49 Years	36.35	2002	-1.64
Afghanistan	AFG	2001	15-49 Years	37.99	2001	0.2600000000
Afghanistan	AFG	2000	15-49 Years	37.73	2000	1.4
Afghanistan	AFG	1999	15-49 Years	36.33	1999	0.75
Afghanistan	AFG	1998	15-49 Years	35.58	1998	0.3500000000
Afghanistan	AFG	1997	15-49 Years	35.23	1997	0.1899999999
Afghanistan	AFG	1996	15-49 Years	35.04	1996	0.0300000000
Afghanistan	AFG	1995	15-49 Years	35.01	1995	-0.4500000000
Afghanistan	AFG	1994	15-49 Years	35.46	1994	-0.2599999999
Afghanistan	AFG	1993	15-49 Years	35.72	1993	-1.45
Afghanistan	AFG	1992	15-49 Years	37.17	1992	-3.36

Figure 17: New Column “Year Date” That Represents Year In Date Format- 2001 (yyyy)

3 RESULTS AND DISCUSSION

The development of CancerAtlas serves as a vital to display worldwide cancer mortality data which helps fulfill project requirements. This project requires equal importance from its front-end together with its back-end components. A detailed analysis of the CancerAtlas will be provided regarding its frontend and backend systems that handle worldwide cancer fatality assessments.

The Overview Dashboard displays worldwide cancer fatality patterns through cancer type rankings together with worldwide cancer death counts and Malaysia's five most common cancer types. Users can perform cancer case and mortality analysis through the Explore Dashboard by choosing particular country selections along with preferred years and cancer types. The Death Rates Dashboard showcases mortality patterns across exclusive age groups which reveals both cancer-related death numbers and nations that experience maximum occurrences. The Summary Page includes essential statistical information about the chosen cancer such that users can easily understand critical characteristics along with trends.

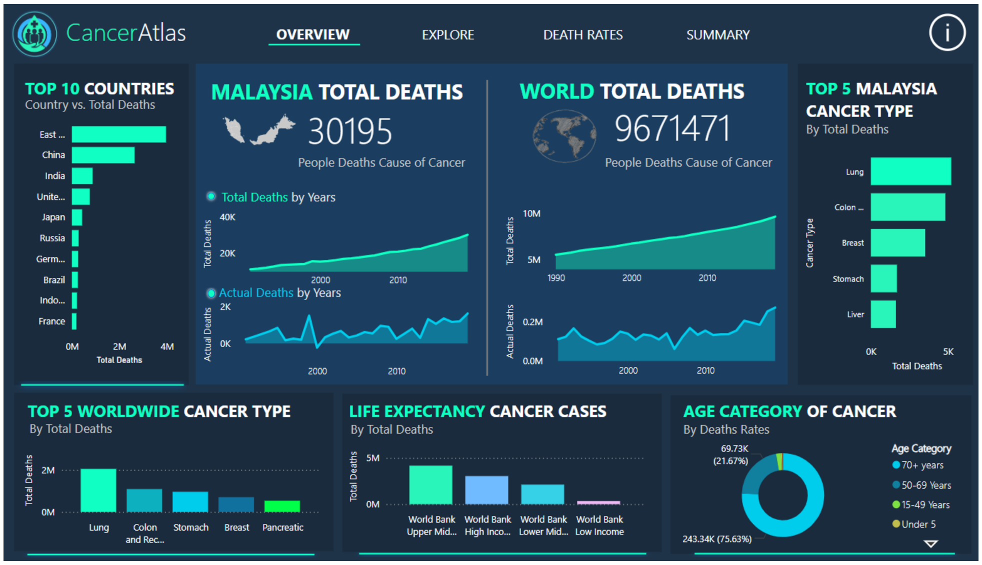


Figure 18: Overview for CancerAtlas

The principal view of the CancerAtlas Dashboard System appears in the following image (Figure 18). This system allows users to examine and compare cancer mortality statistics throughout the world and within Malaysia regarding death rate totals and yearly growth. Users can access broad cancer mortality information on worldwide death statistics which includes both global and national rankings of the deadliest countries. Furthermore, it presents elaborate case statistics about cancer occurrences alongside detailed life expectancy indicators and age-specific cancer mortality rates.

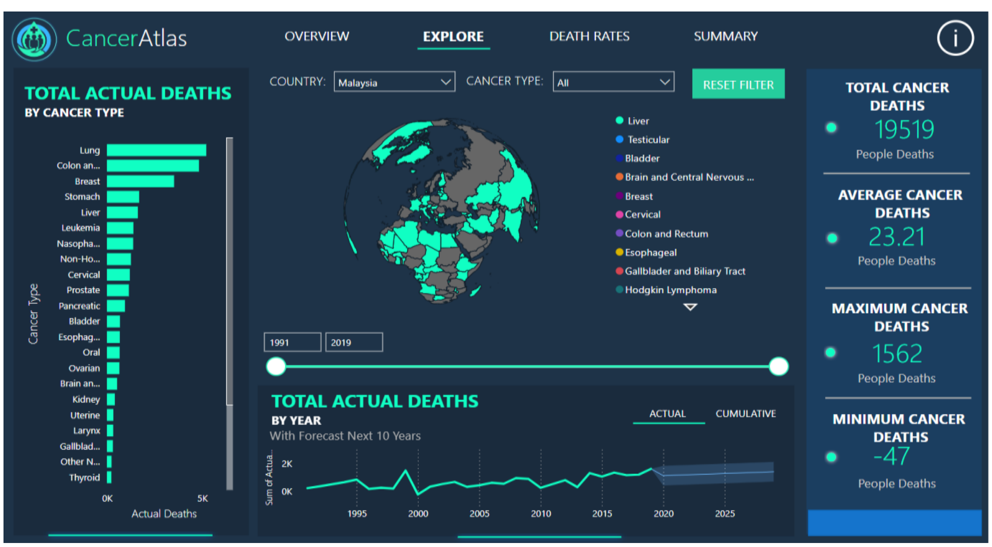


Figure 19: Explore for CancerAtlas

Based on Figure 19, the CancerAtlas Dashboard enables users to create statistic reports from chosen countries through this page. Hence the map reveals the chosen area selection while death data shows cancer-type-specific actual fatalities based upon current filters. Total actual cancer deaths by type are projected to enlarge through the next five years according to the displayed line chart. The right section of the interface presents four cards which display both the total number and average and highest and lowest recorded cancer death counts for the selected country.

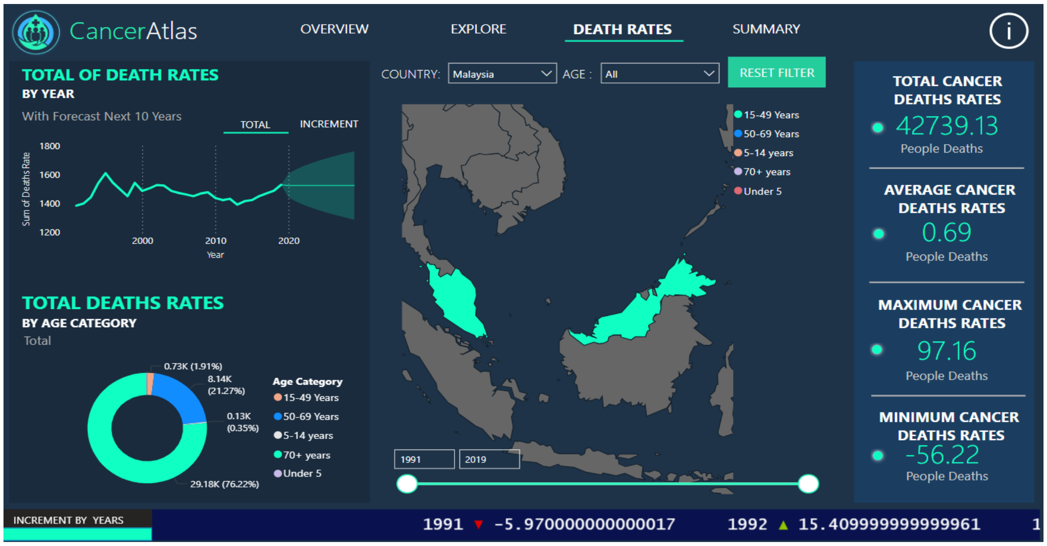


Figure 20: Death Rates for CancerAtlas

In the case represented in Figure 20, The displayed data includes cancer death rate information just like the earlier page showed data about predicted total death rates during the following five years. Through its breakdown the dashboard displays age-specific death rates each day. Viewers can find the total and average cancer death rates alongside minimum and maximum values in the cards on the right section.

4 CONCLUSION

A project named "CancerAtlas: The Data Visualization on Worldwide Cancer Death Cases Analysis" emerged to solve issues related to analyzing international cancer mortality patterns. Sophisticated data visualization methods enabled the project to organize A complex health dataset which contained global cancer mortality rates. Cancer mortality data showed substantial complexity while several nations lacked complete cancer risk factor data along with difficulties in data understanding among different stakeholders. The project established a dynamic processing system to handle substantial data quantity while boosting user interaction for more advanced analysis. The system adopted this method TO make information available for people from different backgrounds. Analysis of system needs related to global cancer death visualization resulted in the development of purposeful visual representations. Unified in CancerAtlas users receive comprehensive knowledge about global cancer death statistics across different types during the 1990-2019 period through validated Kaggle data. Various components within the dashboard display data through a combination of graphs charts and maps that help make the information more appealing and engaging to viewers. The method proved successful for identifying trends along with patterns in worldwide cancer fatalities

ACKNOWLEDGEMENT

The authors would like to thank Universiti Teknologi MARA for the research support and valuable support.

REFERENCES

[1] World Health Organization, "Cancer," World Health Organization, 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer. [Accessed: Apr. 16, 2023].

[2] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," CA: A Cancer Journal for Clinicians, vol. 71, no. 3, 2021, doi: 10.3322/caac.v71.i3.

[3] D. M. Parkin, F. Bray, J. Ferlay, P. Pisani, and A. Znaor, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: A Cancer Journal for Clinicians, vol. 70, no. 6, pp. 359-386, 2020.

[4] K. S. A. Kumar, S. M. Hababa, B. Worku, G. Tadele, Y. G. Mengistu, and A. Y. P., "Big Data Characteristics, classification and Challenges - A Review," May 16, 2021. [Online]. Available: https://turcomat.org/index.php/turkbilmat/article/download/8316/6490/14910. [Accessed: Apr. 16, 2023].

[5] T. Segal, "What is Big Data? Definition, how it works, and uses," Investopedia, Nov. 29, 2022. [Online]. Available: https://www.investopedia.com/terms/b/big-data.asp#:~:text=Big%20data%20refers%20to%20the,v’s%22%20of%20big%20data. [Accessed: Apr. 20, 2023].

[6] M. Favaretto, E. De Clercq, C. O. Schneble, and B. S. Elger, "What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade," PLoS One, vol. 15, no. 2, p. e0228987, Feb. 2020, doi: 10.1371/journal.pone.0228987.

[7] E. Vayena and A. Blasimme, "Health Research with Big Data: Time for Systemic Oversight," The Journal of Law, Medicine & Ethics, vol. 45, no. 1, pp. 7-11, 2017.

[8] G. Cai and Y. Zhu, "Understanding Graphical Symbols: An Examination Of User Performance And Preference Between Textual And Graphical Passwords In Mobile Environments," International Journal Of Human-Computer Interaction, 2020, doi: 10/1080/i1942-390x.

[9] S. Y. Rhee, J. Lee, and B. C. Kwon, "Immersive analytics and visualization: Design considerations and applications," IEEE Transactions on Visualization and Computer Graphics, vol. 26, no. 1, pp. 146-163, 2020.

[10] S. Few, Data visualization: A visual guide to communicating data, 2nd ed. Thousand Oaks, CA: Sage Publications, 2018.

[11] Martin Luenendonk, "23 Best Data Visualization Tools of 2025 (with Examples)," FounderJar, Dec. 6, 2022. Available: https://www.founderjar.com/best-data-visualization-tools/. [Accessed: Feb. 1, 2025].

[12] Portable, "Top 50 Data Visualization Tools: Complete List for 2024," Portable, 2024. Available: https://portable.io/learn/data-visualization-tools. [Accessed: Feb. 1, 2025].

[13] R. Kitchin, "Ethical Data Visualization: Accuracy, Transparency, and Inclusivity," Data Ethics Journal, vol. 7, no. 1, pp. 45-60, 2021.

[14] AltexSoft, "Agile methodology overview," 2019. [Online]. Available: https://www.altexsoft.com/. [Accessed: Apr. 16, 2023].

[15] B. Hossain**,** "Cancer and deaths dataset (1990-2019) globally," *Kaggle,* 2022. [Online]. Available: https://www.kaggle.com/datasets/belayethossainds/cancer-and-deaths-dataset-19902019-globally?select=02+total-cancer-deaths-by-type.csv. [Accessed: Feb. 20, 2023].

[16] ARC Global Cancer Observatory**,** "Global cancer statistics by continent (2020)," *International Agency for Research on Cancer,* 2020. [Online]. Available: https://gco.iarc.fr/today/online-analysis-map?v=2020&mode=population&mode_population=continents&population=900&populations=900&key=asr&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&nb_items=10&group_cancer=1&include_nmsc=0&include_nmsc_other=0&projection=natural-earth&color_palette=default&map_scale=quantile&map_nb_colors=5&continent=0&show_ranking=0&rotate=%255B10%252C0%255D. [Accessed: Jan. 10, 2023].