WHO HEALTHCARE ANALYTICS: LEVERAGING MACHINE LEARNING TO UNRAVEL GLOBAL LIFE EXPECTANCY FACTORS

World Health Organization

DATA

Our research endeavors to provide valuable insights for international business and healthcare professionals to formulate effective strategies. By employing regression analysis and exploratory data analysis (EDA) to examine the relationship between key health indicators, income, region, and life expectancy across nations, we aim to uncover actionable insights. These insights will empower professionals to tailor strategies that address specific health challenges and socioeconomic factors, ultimately contributing to improved global health outcomes and facilitating informed decision-making in both the business and healthcare sectors.

DATASET & ATTRIBUTES

(WORLD HEALTH STATISTICS 2023)

- Countries and areas Region
- Income Group
- Life Expectancy
- Healthy Life Expectancy
- Cancer
- Suicide
- Alcohol Consumption

- Road Accidents
- Obesity
- Malaria
- Hypertension
- Tuberculosis
- Sanitation

RESEARCH OBJECTIVES

- Investigate the variations in factors influencing life expectancy across different regions and income groups.
- Determine the extent of differences in the impact of specific factors on life expectancy among various countries.
- Develop targeted interventions for international business and healthcare professionals based on country-specific findings to improve life expectancy outcomes.

LIBRARIES

Tools and Libraries Utilized:

The project utilizes Google Collab and Excel for data analysis, along with essential Python libraries such as Numpy Pandas, Seaborn, Matplotlib, Scikit-learn, and Statsmodels.

Data Collection and Preprocessing:

Before analysis, the dataset undergoes thorough preprocessing, including handling missing values, removing duplicates, and standardizing variables, ensuring data quality and consistency.

Exploratory Data Analysis (EDA):

Through EDA, the project uncovers variations and patterns in factors across countries, using Pandas, Seaborn, and Matplotlib for visualization and summarization.

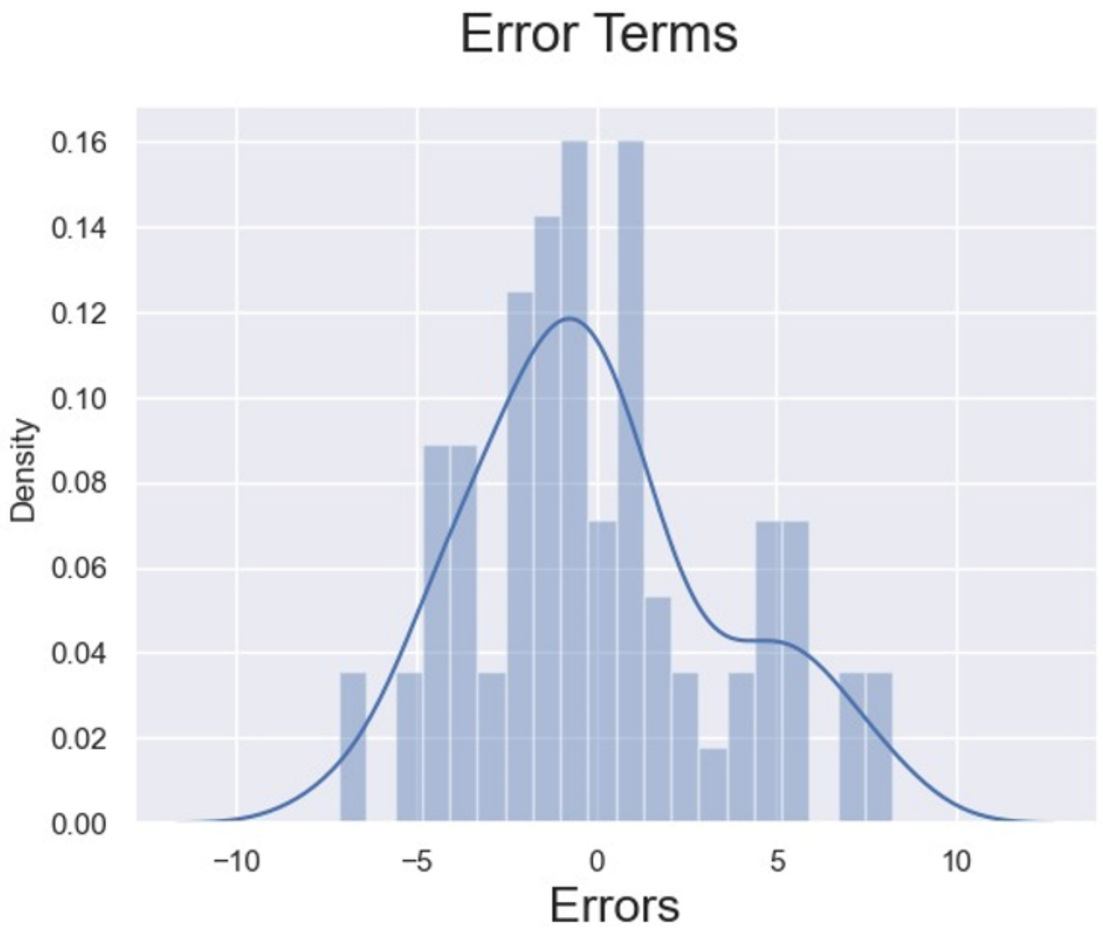
Multiple Regression Analysis:

The project employs Multiple Regression Analysis, facilitated by Scikit-learn and Statsmodels, to understand the relationship between life expectancy and various factors, deriving meaningful insights for decision-making.

Global Life expectancy quality distribution

ERROR DISTRIBUTION

After conducting an error distribution analysis, we discovered a normal distribution of errors in our data. This finding indicates that our statistical model effectively captures the variability in the observed data.



REGRESSION OUTPUT

Dep. Variable:	Life Expectancy	R-squared:	0.925
Model:	OLS	Adj. R-squared:	0.916
Method:	Least Squares	F-statistic:	98.55
Date:	Thu, 11 Apr 2024	Prob (F-statistic):	5.43e-33
Time:	21:10:56	Log-Likelihood:	-145.35
No. Observations:	73	AIC:	308.7
Df Residuals:	64	BIC:	329.3
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	85.4687	1.100	77.719	0.000	83.272	87.666
Suicide	-0.0390	0.038	-1.026	0.309	-0.115	0.037
Cancer Deaths	-0.4029	0.045	-8.892	0.000	-0.493	-0.312
Alcohol Consumption	-0.0452	0.082	-0.551	0.583	-0.209	0.119
Road Accidents	-0.0997	0.026	-3.798	0.000	-0.152	-0.047
Obesity	-0.0060	0.033	-0.183	0.856	-0.071	0.059
Malaria	-0.0073	0.003	-2.763	0.007	-0.013	-0.002
Tuberculosis	-0.0072	0.002	-3.818	0.000	-0.011	-0.003
Sanitation	-0.1202	0.016	-7.488	0.000	-0.152	-0.088

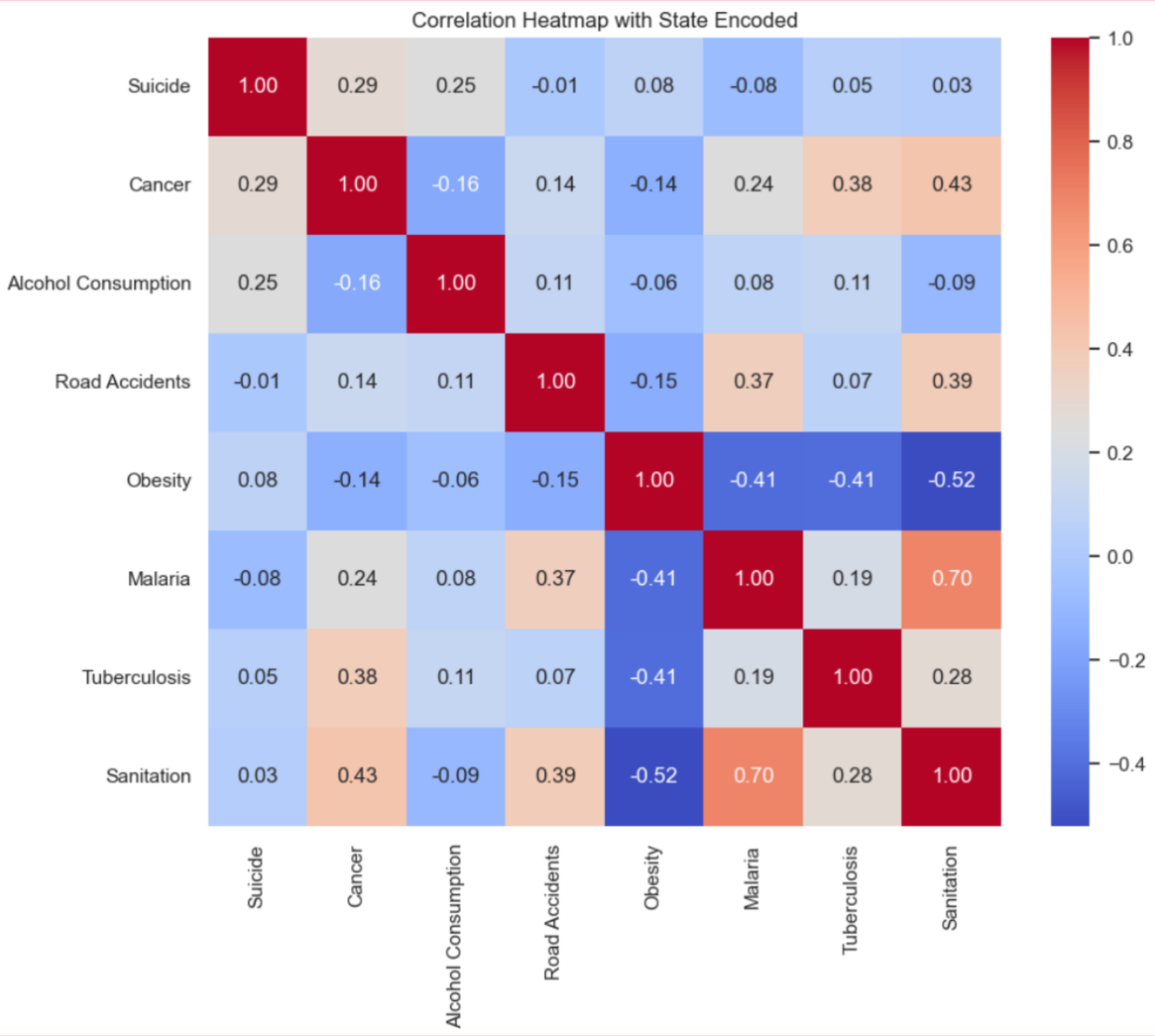
Omnibus:	2.605	Durbin-Watson:	2.098
Prob(Omnibus):	0.272	Jarque-Bera (JB):	2.003
Skew:	0.394	Prob(JB):	0.367
Kurtosis:	3.192	Cond. No.	1.25e+03

REGRESSION RESULTS

- Mean Squared Error (MSE): The MSE is approximately 3.75, indicating that, on average, the squared difference between observed and predicted life expectancy values is around 3.75 units.
- R-squared: With an R-squared value of approximately 0.82, this indicates that approximately 82% of the variability in life expectancy is accounted for by the independent variables in the model.

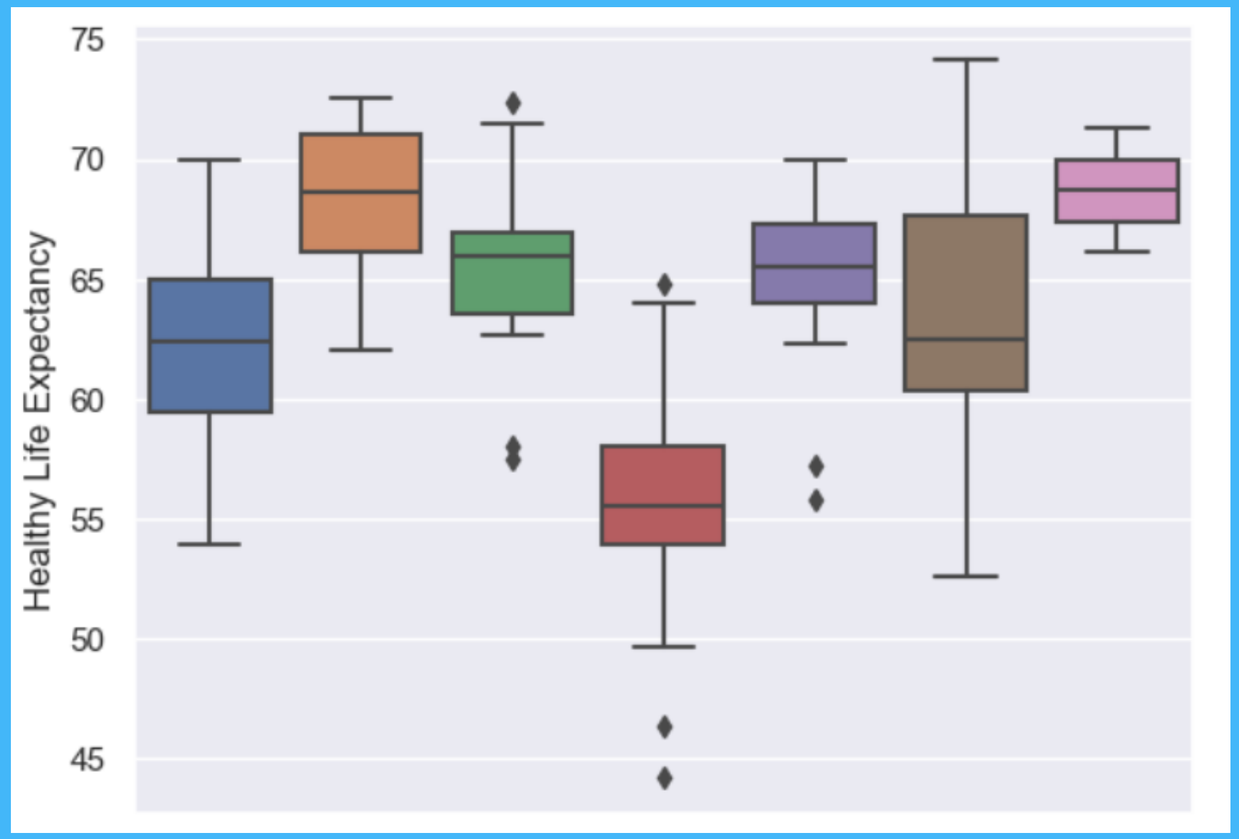
- Regression Equation Coefficients: Each coefficient represents the change in life expectancy associated with a one-unit change in the corresponding independent variable.
- Regression Equation Intercept: , the intercept is approximately 85.47, indicating that, in the absence of any other factors, the estimated life expectancy is around 85.47 years.

HEATMAP



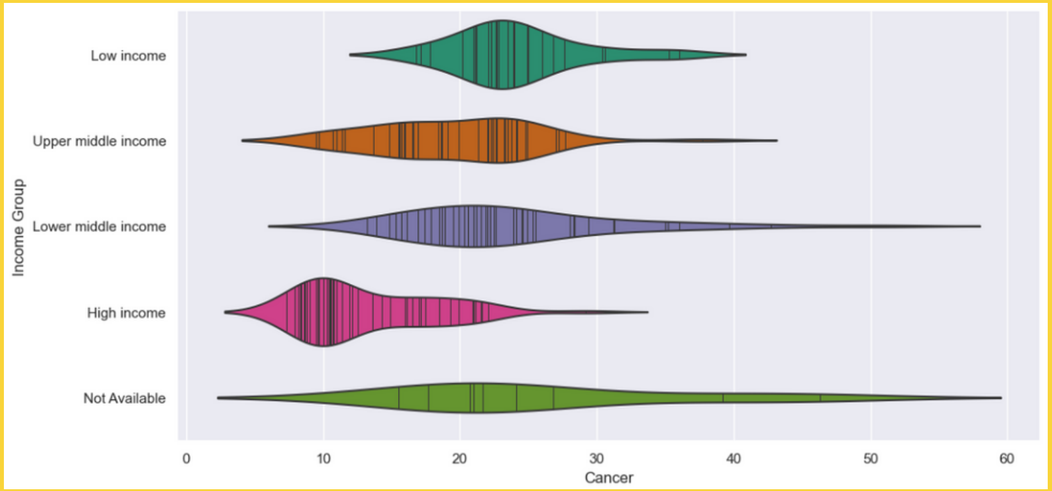
From the correlation plot, we can see that none of the variables are highly correlated, which indicates that the model does not have multi-collinearity issues.

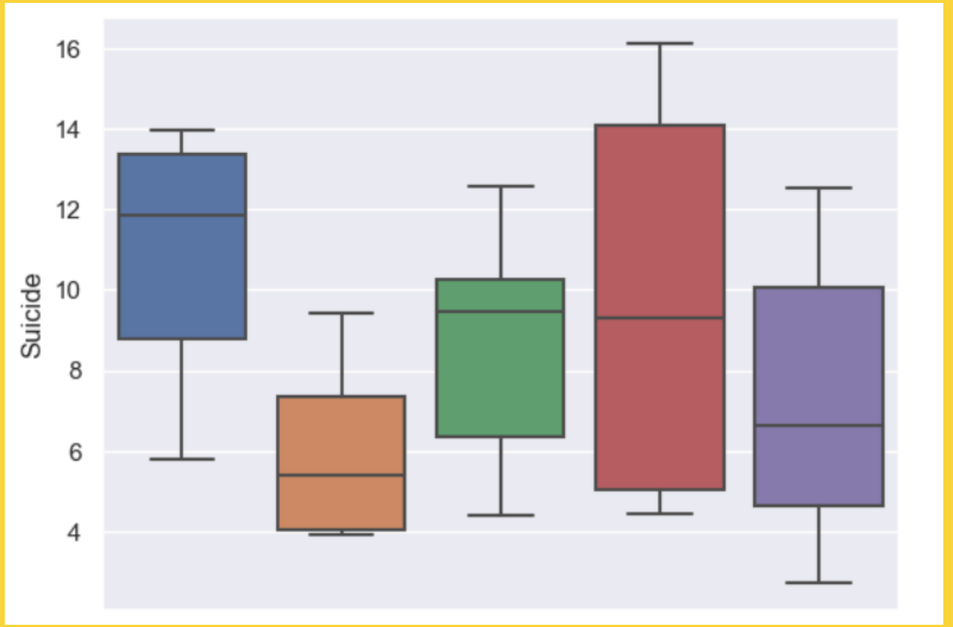
VISUALIZATIONS

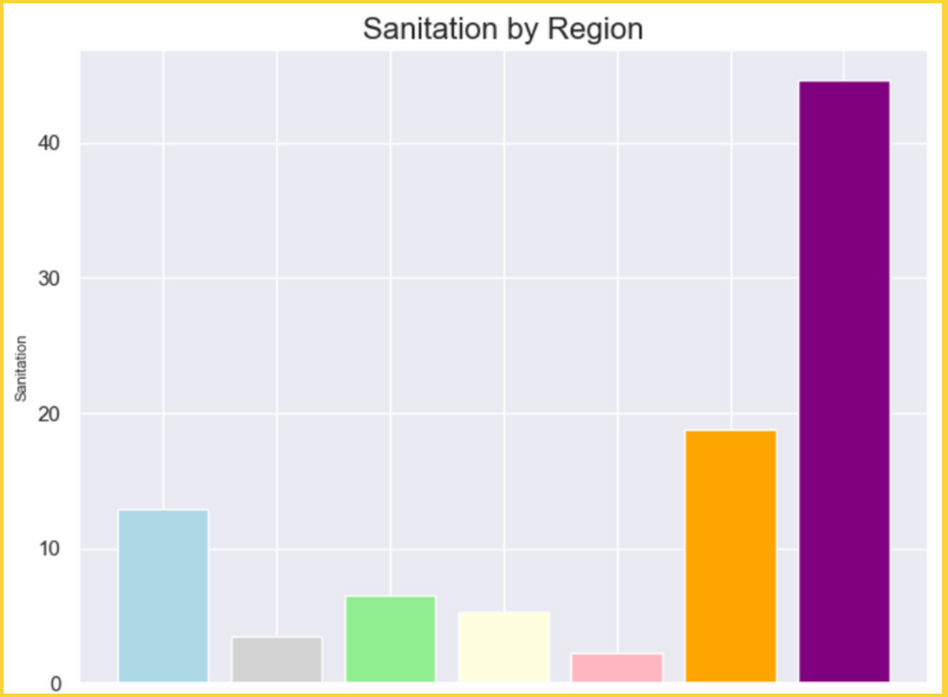


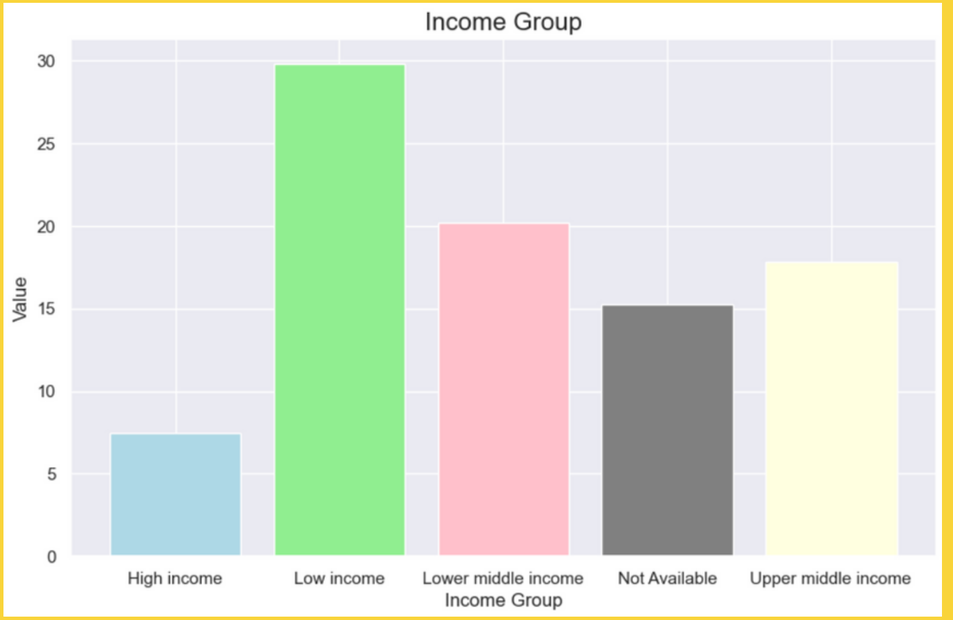
Life Expectancy vs Region

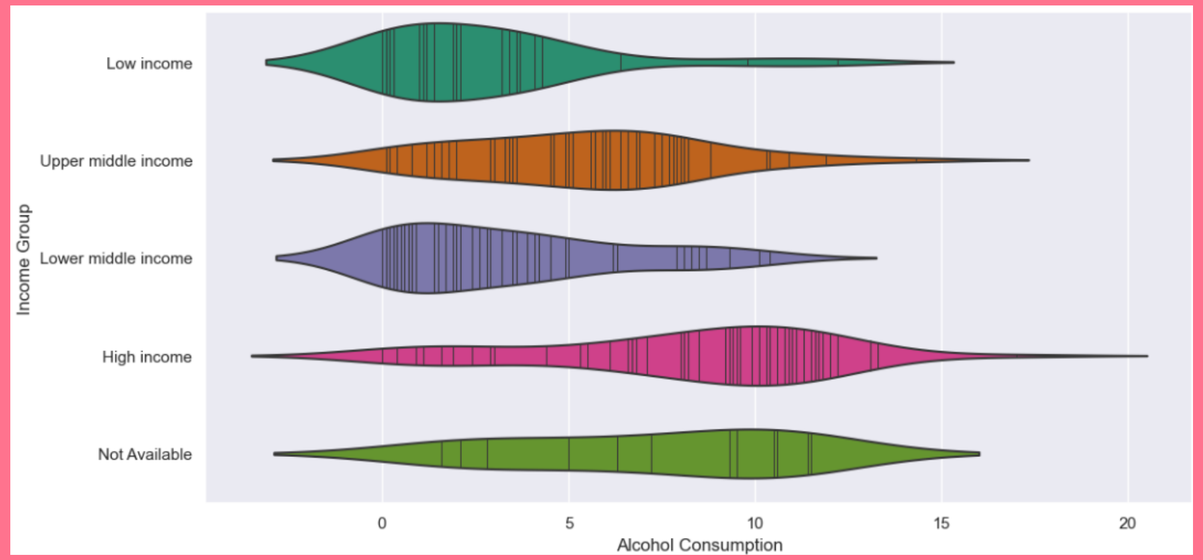
Income group vs Region


Low income
Upper middle income
Lower middle income
High income
Not Available
Income Group
0
10
20
30
40
50
60
Cancer


16
14
12
10
8
6
4
Suicide


Sanitation by Region
40
30
20
10
0
Sanitation


Income Group
30
25
20
15
10
5
0
Value
High income
Low income
Lower middle income
Not Available
Upper middle income
Income Group




```
Region
East Asia & Pacific                25.600000
Europe & Central Asia                    NaN
Latin America & Caribbean           4.555556
Middle East & North Africa         11.660000
North America                            NaN
South Asia                          1.771429
Sub-Saharan Africa                179.068889
Name: Malaria, dtype: float64
```

```
Region
East Asia & Pacific               181.550000
Europe & Central Asia              22.362745
Latin America & Caribbean          35.158824
Middle East & North Africa         32.390476
North America                       3.950000
South Asia                        172.250000
Sub-Saharan Africa                208.416667
Name: Tuberculosis, dtype: float64
```

KEY IMPLICATIONS -GLOBAL BUSINESSES

1. Personalized Marketing and Customer Segmentation:
   - Utilize consumer health data to tailor marketing campaigns and promotions for specific customer segments, increasing engagement and loyalty.

2. Health and Wellness Product Development:
   - Develop products based on consumer health trends and preferences to address specific health concerns and resonate with health-conscious consumers.

3. Supply Chain Optimization and Inventory Management:
   - Apply analytics to healthcare supply chain data to improve inventory forecasting, streamline logistics, and enhance operational efficiency for cost savings and improved product availability.

4. Customer Experience Enhancement:
   - Offer personalized health-related recommendations and services based on consumer health behaviors and preferences to meet the evolving needs of health-conscious consumers.

KEY IMPLICATIONS (CONTD.)

5. Risk Management and Product Liability:

- Proactively identify risks related to product liability and consumer health concerns using health data and consumer feedback to ensure product safety, compliance, and maintain consumer trust.

6. Data Privacy and Security Compliance:

- Implement data protection practices from healthcare analytics to safeguard consumer health data, ensure regulatory compliance, and build trust with customers regarding sensitive health information.

7. Market Trend Analysis and Competitive Intelligence:

- Analyze health-related market trends, consumer behaviors, and competitive landscape to identify opportunities, anticipate consumer preferences, and make informed strategic decisions to stay ahead in the market.

KEY IMPLICATIONS - ANALYTICS

1. Predictive Analytics for Early Disease Detection:
   - Research in predictive analytics can focus on developing models to detect diseases at an early stage based on patient data, biomarkers, and risk factors.
   - By leveraging predictive analytics, healthcare providers can intervene proactively, leading to early diagnosis, timely treatment, and improved health outcomes for patients.

2. Precision Medicine and Genomic Analytics:
   - Further research in genomic analytics can enable the advancement of precision medicine by tailoring treatment plans based on individual genetic profiles.
   - By integrating genomic data with clinical information, researchers can identify personalized treatment strategies, optimize drug selection, and enhance therapeutic outcomes for patients.

3. Telehealth and Remote Monitoring:
   - The application of analytics in telehealth and remote monitoring can facilitate real-time data analysis for virtual patient consultations and remote health management.
   - By leveraging analytics tools, healthcare providers can monitor patient health remotely, deliver personalized care, and improve access to healthcare services, especially in underserved areas.

KEY IMPLICATIONS (CONTD.)

4. Healthcare Operations Optimization:
   - Research in healthcare operations analytics can focus on optimizing resource allocation, workflow efficiency, and patient flow within healthcare facilities.
   - By analyzing operational data, researchers can identify bottlenecks, streamline processes, and enhance the overall efficiency of healthcare delivery, leading to improved patient experiences and cost savings.

5. Drug Safety and Pharmacovigilance Analytics:
   - Further exploration in pharmacovigilance analytics can enhance drug safety monitoring, adverse event detection, and post-market surveillance of pharmaceutical products.
   - By analyzing real-world data and adverse event reports, researchers can identify potential safety concerns, assess drug effectiveness, and support regulatory decision-making to ensure patient safety.

6. Population Health Analytics and Epidemiological Studies:
   - Research in population health analytics can focus on analyzing health trends, disease patterns, and social determinants of health to inform public health interventions.
   - By conducting epidemiological studies and leveraging population health data, researchers can identify health disparities, predict disease outbreaks, and design targeted interventions to improve population health outcomes.

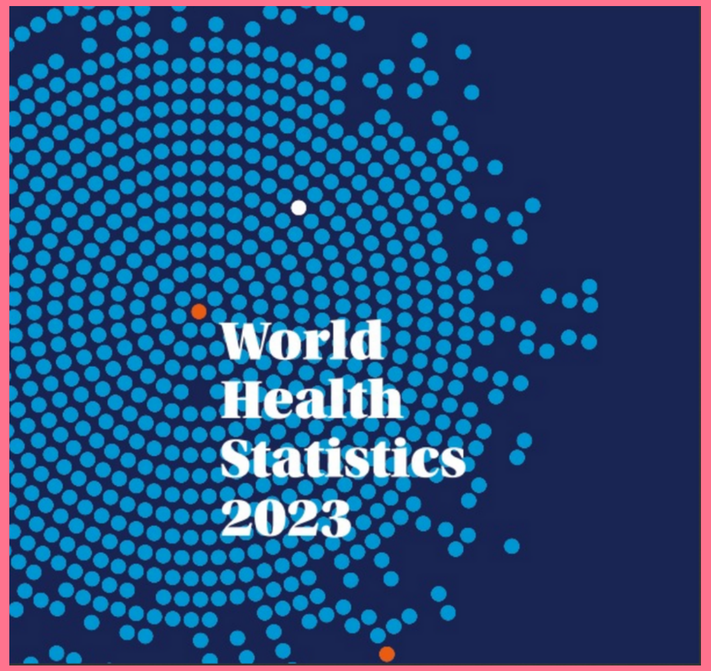
7. Patient Engagement and Behavioral Analytics:
   - Exploring behavioral analytics in healthcare can help understand patient preferences, behaviors, and engagement with healthcare services.
   - By analyzing patient interactions, feedback, and engagement metrics, researchers can design personalized interventions, improve patient adherence to treatment plans, and enhance overall patient satisfaction and outcomes

CONCLUSION

In conclusion, the findings of studies and reports in healthcare analytics provide valuable insights that can guide future research and innovation in the healthcare industry. By exploring new areas and developing novel applications based on data-driven insights, researchers can drive advancements in healthcare delivery, personalized medicine, patient outcomes, and public health initiatives, ultimately leading to improved quality of care and better health outcomes for individuals and communities.

Bibliography:

1. https://www.who.int/publications/i/item/9789240074323
2. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups
3. https://repository.gheli.harvard.edu/repository/11242/#:~:text=World%20Health%20Statistics%202023%20summarizes,policy%20and%20health%20systems%20strengthening
4. https://online.sbu.edu/news/how-business-analytics-is-changing-the-healthcare-industry
5. https://www.linkedin.com/pulse/retail-healthcare-shifting-landscape-david-jensen-vdhnc/


World
Health
Statistics
2023


analytic Steps
www.analyticssteps.com


World Health
Organization

Mentor:

Prof.Muhammad Al-Abdullah

Bhavya Priya Akula

Sathvika Kokku

Vishnu Mannam