HEALTH ISSUES ANALYSIS REPORT

-Fazaluddin Mohd

INTRODUCTION

The aim of this report is to provide an analysis of health data sets to examine the impact of socioeconomic and demographic factors on health outcomes, with a particular focus on the period 2020 to the present. The datasets used for this analysis are sourced from the Centers for Disease Control and Prevention (CDC). We loaded these datasets into Google BigQuery, filtered, transformed, and merged them to create a combined dataset for comprehensive analysis.

DATASETS USED

The research analysis in this report is based on a combination of the following data sets, each of which provides important insights into various aspects of health care and mortality in the United States:

1. ***Excess Deaths Associated with COVID-19 (Dataset 1):***
   Source: https://data.cdc.gov/NCHS/Excess-Deaths-Associated-with-COVID-19/xkkf-xrst
   Screenshot: This database provides detailed information on excess mortality related to the COVID-19 pandemic, including observed rates, threshold values, and estimates, to provide a deeper understanding of the impact of the pandemic on mortals who have been weakened.

2. ***VSRR - State and National Provisional Counts for Live Births, Deaths, and Infant Deaths*** *(Dataset 2):*
   Source: https://data.cdc.gov/NCHS/VSRR-State-and-National-Provisional-Counts-for-Liv/hmz2-vwda
   Screenshot: This database provides a comprehensive breakdown of live births, deaths, and infant mortality rates by state and country, and is an important resource for analyzing trends in demographic indicators.

3. ***Weekly Counts of Death by Jurisdiction and Select Causes of Death (Dataset 3):***
   Source: https://data.cdc.gov/NCHS/Weekly-Counts-of-Death-by-Jurisdiction-and-Select-/u6jv-9ijr
   Screenshot: Providing weekly mortality rates by control and cause of death, this database helps to identify mortality patterns and trends based on a variety of health-related factors.

4. **Weekly Counts of Deaths by Jurisdiction and Age (Dataset 4):**
   Source:https://data.cdc.gov/NCHS/Weekly-Counts-of-Deaths-by-Jurisdiction-and-Age/y5bj-9g5w
   Screenshot

5. **COVID-19 deaths by week and state:**
   Source:https://data.cdc.gov/NCHS/Provisional-COVID-19-Death-Counts-by-Week-Ending-D/r8kw-7aab
   Screenshot

6. **Provisional COVID-19 Death Counts by Week Ending Date and State:**
   Source:https://data.cdc.gov/NCHS/Provisional-COVID-19-Death-Counts-by-Week-Ending-D/r8kw-7aab
   Screenshot

7. **COVID-19 Weekly Cases and Deaths by Age, Race/Ethnicity, and Sex:**
   Source:
   source:https://data.cdc.gov/Public-Health-Surveillance/COVID-19-Weekly-Cases-and-Deaths-by-Age-Race-Ethni/hrdz-jaxc
   Screenshot: Providing insight into weekly mortality rates stratified by jurisdiction and age group, this dataset enables microscopic analysis of how mortality rates vary by age group and geographic area

HEALTH ISSUES ANALYSIS REPORT

-Fazaluddin Mohd

These datasets form the basis of our comprehensive analysis, which will provide a multidimensional view of the interactions between health, demographics, mortality, and COVID-19 outcomes.

DATA PREPARATION

1. Filtering the Datasets:

We began by getting the data from 2020 onwards. This is crucial to study the pandemic's effects on health. We used SQL queries in Google BigQuery to filter the datasets and select the most recent and relevant data. Here are the queries we used for each dataset. Filtering datasheets to get data from year 2020: Screenshot

2. Data Transformation:

We kept the data as is after filtering to maintain accuracy and ensure an impartial analysis. This approach preserves the dataset's original structure, ensuring reliability. We transformed the data to get unique columns for merging.

3. Merging the Datasets:

We combined data by matching characteristics like "weekend date" and "situation" from four datasets. This made it easier to understand the links between health outcomes, demographics, and COVID-19 impact. The merged dataset, named integrated_dataset, offers diverse health context perspectives, laying a strong groundwork for our analysis and research inquiries. Screenshot

4. filtering the dataset_5, dataset_6, dataset_7:

New_Dataset_5:
Objective: Focus on deaths from various diseases.
Columns: Data_As_Of, Jurisdiction_of_Occurrence, MMWR_Year, All_Cause, Natural_Cause, etc.
Dataset_6_Clean:
Objective: Focus on COVID-19 deaths and total deaths.
Columns: Data_as_of, Start_Date, State, County_of_Occurrence, Fips_Code, COVID_19_Deaths, Total_Deaths, etc.
Dataset_7_Cleaned:
Objective: Focus on COVID-19 deaths, total deaths, and associated factors.
Columns: Data_as_of, Start_Date, State, Age_group, Race_and_Hispanic_Origin_Group, COVID_19_Deaths, Total_Deaths, Pneumonia_Deaths, etc. Screenshot

5. Merging these datasets with combined_dataset: Screenshot

6. Preview of combined_data_final_v2:

Screenshot 1
Screenshot 2
Screenshot 3

RESEARCH QUESTIONS: To better understand how socioeconomic factors, demographics, and health outcomes interact. Our study aims to answer these crucial questions.

1. How do socioeconomic factors correlate with the number of excess deaths associated with COVID-19 in different states?
2. Are there any significant trends in infant mortality rates in recent years based on demographic and geographical factors?
3. What is the impact of age and demographics on causes of death and excess mortality in different states during the COVID-19 pandemic?

ANALYSIS

1. Impact of Socioeconomic Factors on COVID-19 Excess Deaths

HEALTH ISSUES ANALYSIS REPORT

-Fazaluddin Mohd

The merged data lets us look at how socioeconomic factors relate to the number of COVID-19 cases in various countries. We can check variables like a country's economy, healthcare, population, and vaccination rates to see how they affect COVID-19 death rates in different areas. It also helps us find gaps in healthcare and resource use, showing how the pandemic affects vulnerable communities.

2. Trends in Infant Mortality Rates and Demographic Patterns

Analyzing data on live births, deaths, and infant mortality helps us understand trends in infant health. We look at factors like maternal age, race, and location to see where there are differences in infant mortality. This information helps us focus on improving the health of mothers and children in specific areas.

3. Impact of Age and Demographics on Causes of Death

Weekly mortality data includes information on how age and demographics impact various causes of death, not just COVID-19. By breaking the data down by age, location, and cause of death, we can spot trends and differences in health conditions at the population level. This helps us learn about chronic diseases, the success of preventive health measures, and how public health issues change in different populations.

POTENTIAL IMPLICATIONS

This research can guide smart policies and actions in public health to address ongoing epidemics and health disparities. By understanding what factors affect health, we can create specific rules to improve healthcare, prevent diseases, and ensure everyone gets equal access to good health. Additionally, the insights from this research help us allocate resources effectively in public health, create personalized health plans, and use data to enhance our public health efforts, especially for marginalized and vulnerable communities.

IMPLICATIONS AND RECOMMENDATIONS

The findings highlight the critical role of proactive public health programs and interventions in addressing multifaceted challenges including health outcomes, socioeconomic factors, and demographics emphasizing the role of networks. In light of these findings, we recommend the following key considerations.

1. Equitable Healthcare Access: Prioritize access to health care, especially in underserved communities, to ensure that all individuals have access to essential health goods and services.
2. Targeted Public Health Interventions: Develop tailored public health interventions that target specific population groups and address their specific health needs, promote disease prevention strategies, manage it, and promote overall welfare.
3. Data-Driven Policy Implementation: Advocate for evidence-based policymaking that is grounded in comprehensive data analysis, fostering the implementation of informed policies that prioritize health equity, disease prevention, and community well-being.

CONCLUSION

This report offers a comprehensive analysis of combined health data, shedding light on how socioeconomic and demographic elements influence health outcomes both during and post the COVID-19 pandemic. The synthesized CDC data serves as a valuable source for conducting in-depth research and gaining critical insights into public health.

HEALTH ISSUES ANALYSIS REPORT

-Fazaluddin Mohd

References:

Barrera, T. (2023, January 12). BigQuery 101: A Beginner's Guide to Google's Cloud Data Warehouse. Airbyte.com. https://airbyte.com/blog/bigquery-guide

Cloud, G. (n.d.). BigQuery documentation. Google Cloud. https://cloud.google.com/bigquery/docs

Haridass, T., & Brown, E. (2017). Learning Google BigQuery : a beginner's guide to mining massive datasets through interactive analysis. Packt Publishing.

Khan, A. (2021, July 12). Google BigQuery Beginner's Guide – How to Analyze Large Datasets. https://www.freecodecamp.org/news/google-bigquery-beginners-guide/

Lakshmanan, V., & Tigani, J. (2019). Google BigQuery: The Definitive Guide. O'Reilly Media.

Mucchetti, M. (2020). BigQuery for data warehousing : managed data analysis in the Google cloud.