

Analysis of death rate of Diabetes after covid19 and before covid19 in NA

Introduction

The COVID-19 pandemic has profoundly affected global public health, with impacts extending beyond infection-related mortality to chronic conditions such as diabetes. Diabetes, a leading cause of death and disability worldwide.

This analysis focuses on diabetes-related death rates in North America, comparing trends before and after the onset of COVID-19. Using data visualization techniques, the study highlights a marked increase in mortality rates following the pandemic. This finding underscores the ripple effects of the pandemic on chronic disease outcomes, influenced by reduced access to medical care, delayed diagnoses, and changes in patient behavior during this period.

The insights from this study shed light on the urgent need for resilient healthcare strategies to protect vulnerable populations during health crises and mitigate long-term impacts on chronic disease outcomes.

Question

1. Did the rate of diabetes-related deaths increase in the USA after the onset of COVID-19?

Data source:

Death (2014-2019) data:

<https://data.cdc.gov/api/views/9dzkmvmi/rows.csv?accessType=DOWNLOAD>

Structured and Quality: The dataset is structured with columns representing different sectors and rows for each year and month. It has been verified to have consistent and clean data for the specified years and months.

Death (2020-2023) data:

<https://data.cdc.gov/api/views/9dzk-mvmi/rows.csv?accessType=DOWNLOAD>

Structure and Quality: This dataset complements the primary dataset by providing necessary additional information.

Data Structure: Tabular format with columns for year, month and death.

Reasons for Choosing These Data Sources

Both datasets are from the United State, making them highly relevant for regional Death rate analysis.

licenses: I used <https://data.gov/> to download dataset. As we all know it is open source data platform.

Data Pipeline:

1. High-Level Description of the Pipeline

The pipeline automates the process of downloading, cleaning, transforming, and storing diabetes-related data from multiple online sources into an SQLite database. It uses Python, Pandas for data manipulation, and SQLite for storage. The script ensures the pipeline is modular and can handle changing datasets dynamically.

2. Transformation and Cleaning Steps

- **Column Selection:** Retained Year, Month, and Diabetes Mellitus columns to focus on relevant data.
- **Missing Data Handling:** Removed rows with missing values using `dropna()` to ensure data quality.
- **Type Conversion:** Converted Year and Month to integers for consistency.
- **Shuffling:** Randomized rows to prevent bias in sequential data.
- **Dataset Combination:** Merged data from multiple sources into a unified format.

3. Problems Encountered and Solutions

- **Data Inconsistency:** Addressed by standardizing columns and filtering irrelevant data.
- **Missing Values:** Resolved by dropping incomplete rows.
- **URL Changes:** Designed the pipeline to easily accept new data URLs.
- **Duplicates:** Verified and removed duplicate entries during merging.

4. Meta-Quality Measures and Error Handling

- **Validation:** Ensured critical columns exist and dropped invalid rows.
- **Error Tolerance:** Handled missing or broken URLs gracefully, processing available data.
- **Dynamic Inputs:** Easily adaptable to new datasets with different schemas.
- **Resilience:** Logs errors and previews results for verification after saving to the database.

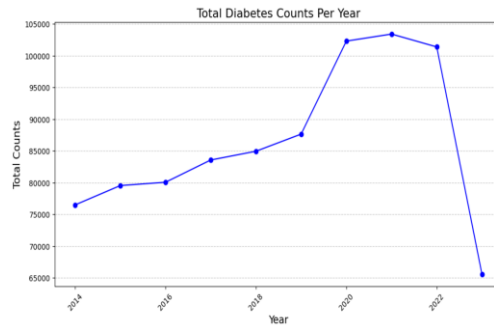
Analysis:

Data Summary:

This data shows different death for differ year and month for America over the period of 10 years.

```
Preprocessed data preview:
  Year  Month  Diabetes Mellitus
0  2015     10           6566
1  2019      2           7485
2  2016      6           6051
3  2016      5           6626
4  2018      5           6805
5  2017      4           6701
6  2015     11           6582
7  2019      4           7360
8  2019      7           6880
9  2019     11           7332
10 2016      4           6522
11 2017     10           6829
```

Trend of death 2014 to 2023



Pipeline Output

- The pipeline outputs diabetes-related death counts stored in an SQLite database with columns: Year, Month, and Diabetes Mellitus.
- The chart shows an increasing trend from 2014 to 2021, followed by a sharp decline in 2022 and dramatical decrease in 2023.

Data Structure and Quality

- Structure: Tabular, clean, and standardized data with no missing values.
- Quality: Follows reproducibility but lacks validation against external sources and needs better handling of potential outliers.

Output Format

- SQLite Database: Lightweight, query-friendly, and integrates well with analysis tools. Ideal for medium-sized datasets.

Critical Reflection

1. Data Limitations:
 - Steep drop in 2023 because incomplete (first 9 months) or maybe missing data.
2. Technical Issues: Column changes or data inconsistencies in source URLs could break the pipeline.

Chart Analysis

- **Observation:** The chart shows a significant increase in diabetes-related deaths from 2019 to 2022, followed by a sharp decline in 2023. This suggests that the **COVID-19 pandemic** may have had a major impact on the death rates of diabetic individuals in North America, either due to direct complications from the virus or disruptions in healthcare during the pandemic.