

**Final Project Progress Update**  
**CIS 635-01**  
**Knowledge Discovery and Data Mining (F23)**

---

**Progress So Far:**

**Completed Tasks:**

**Data Acquisition and Preprocessing:**

I have successfully obtained the dataset titled 'trends and patterns of disparities in diabetes and chronic kidney disease mortality among US counties, 1984-2014' from data.gov data source.  
I have successfully initiated data preprocessing, addressing issues such as missing data.  
I successfully removed significant missing values for a refined dataset.  
I successfully conducted a thorough review of the original research paper by Mokdad et al. (2022) to understand the study's objectives, methodology, and findings.

APA Citation of the replicated article:

Mokdad, A. H., Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R. W., Morozoff, C., Shirude, S., Finegold, S. B., Callender, C., Naghavi, M., & Murray, C. J. L. (2022). Trends and patterns of disparities in diabetes and chronic kidney disease mortality among US counties, 1980-2014. Population Health Metrics, 20(1),9. <https://doi.org/10.1186/s12963-022-00285-4>

Dataset:

<https://catalog.data.gov/dataset?q=disparities+in+diabetes+and+chronic+kidney+disease+mortality+among+US+counties%2C+1980%E2%80%932014>

**Challenges**

**Encountered Difficulties:**

**1. Incomplete Data:**

Discovered that the acquired dataset lacked essential attributes, particularly the "Year" attribute crucial for determining the trends and patterns.

**Addressing Challenges**

**1. Incomplete Data:**

Engaged in exploratory data analysis to identify potential correlations that could supplement missing information.

**Collaboration**

**Group Collaboration:**

This project is an individual endeavor, and thus, there is no collaboration with a team.

## **Next Steps**

### **Remaining Tasks:**

Currently, the main outstanding task involves the documentation of the project.

### **Documentation**

#### **Plan:**

I have initiated the development of a comprehensive outline for the documentation, categorizing it into sections corresponding to each aspect of the project, including data acquisition, preprocessing, and analysis.

Subsequently, I will systematically document each step, offering clear explanations and providing the rationale behind decisions, along with the inclusion of relevant code.

To enhance the quality of the documentation, I plan to seek guidance from the professor, ensuring both clarity and completeness in conveying the intricacies of the replication process.

### **Potential Challenges:**

While progressing with the documentation phase, some challenges may arise, including:

#### **1. Data Completeness:**

Ensuring that all relevant data is included in the documentation, with the missing information in the original dataset.

#### **2. Code Reproducibility:**

Ensuring that the code included in the documentation is reproducible and understandable by others who may want to replicate or validate the project.