Project Statement for Milestone 1

Coffee & T

Charles Nguyen, Dylan Hodge, Alanna Lee, Xinyu Liu, Noah Howell

1. Problem Statement:

   a. Give a formal description of the project. What's the input and output of the problem?

In this project, we want to investigate the inequities potentially present in different demographics' health care payout, premiums, and overall cost. Through this project we hope to discover whether the implementation of certain policies in different states have had any noticeable impact on these disparities.

Input = Health insurance premiums, payouts, policy implementation categories, policy implementation date

Output = Top ranked states, Lowest ranked states, highest impact policies

   b. Why is the problem you want to address important? What's its application?

It is a reality that different socio-economic groups experience a varying degree of insurance costs. For example, pregnancypregnancy typically plays a large role in higher premiums. These disparities can contribute significantly to damaging minority groups and affecting their economic mobility. The goal is to see whether it is possible to reduce these inequalities through existing or new policies.

   c. Specify the goal you want to achieve (a new problem, improvement for algorithms, and/or a thorough experimental evaluation of existing algorithms).

Discover the most effective policies (if a causality has been proven).

2. Team:

a. Who are the team members?

Charles Nguyen, Dylan Hodge, Alanna Lee, Xinyu Liu, and Noah Howell are the team members.

b. What knowledge and skills do the team have from previous courses, projects or internships?

Charles Nguyen: C++, C#, Java, Python and some SQL.
Dylan Hodge: Data structuring/searching/parsing using Python and Java, data cleaning with Excel, modeling using SQL, analysis using R, Tableau and/or PowerBI.
Alanna Lee: Data analysis using R, Java, and Python, website creation with Visual Basic, and database creation and manipulation using PSQL.
Xinyu Liu: Good software engineering skills from internships and work experience. I can code.
Noah Howell: C++, C#, Python, SQL, Go. Software engineering experience from current internship at SEL.

c. What will be each team member's roles and responsibilities in the current project?

Charles Nguyen: Writes code.
Dylan Hodge: Data cleaning, modeling, collection.
Alanna Lee: Data analysis using Python and/or SQL
Xinyu Liu: Code them like a monkey
Noah Howell: Coding / software engineering. Can do SQL.

3. Dataset and Tools:

a. Give the link and description of the dataset.

Census health data:

2018 CPS ASEC Split-Panel Test (census.gov)

This dataset is made up of data collected by the US Census Bureau on the use of health insurance among American citizens. 100,000 addresses were used, with 5000 of these randomly given a set of questions from previous years; the rest were given a new set to complete.

The data collected with these two sets were slightly different, with the old set giving a total of 11 percent uninsured. The new set of questions returned a total rate of 8.8 percent.

US Health Insurance Dataset | Kaggle

This is a general database with over 1300 entries, with insurance costs listed along with the insured individual's age, sex, BMI, the area they live in, whether or not they smoke, and how many children they have.

b. Briefly state what tools you are considering for implementing the project.

The languages we're interested in are Python, SQL, or a mixture of both to analyze the data, since those are the two the team is most familiar with. The other possible tools that we're considering are:

- a parser to capture and clean the input data
- a DBMS to store the dataset
- a query tool for extracting data tables
- Parser-DBMS and Query DBMS interfaces
- Others as needed

4. Project Progress and Contributions:

- What have you done to prepare the data sets?

We've done some individual data exploration in order to sharpen the focus of our project, and to determine the best method of manipulating the data.

- List each team member's contribution during Milestone 1.

Charles Nguyen: Reviewed milestone 1 doc.
Dylan Hodge: Discovered project idea, found relevant data sets, gave a basic idea of project scope.
Alanna Lee: Added dataset descriptions, explored dataset in Python, helped flesh out and format milestone paper, and worked together with the team to decide project direction and scope.
Xinyu Liu: Review datasets
Noah Howell: Reviewed data, suggested ideas, helped edit

- What's your plan for Milestone 2?

We plan to use a small set(s) of data as the starting point as we consider this project to be a prototype in the spirit of learning. If the prototype proves functional, we will consider the possibility of generalizing it out for large sets.

More specifically on the technical side, we are currently planning to use Python and Pandas to clean up the data (if needed). From there, the data will be mapped into a SQL database. The specific SQL database technology will be decided when we begin to work on Milestone 2.

We are also open to adjusting our tools and datasets as the project continues to develop and evolve.