

# **Software Design Specification: Repository and Mining of Temporal Data**

## **Team Members:**

Jessica Nguy	jnguy2014@my.fit.edu
Siomara Nieves	snieves2014@my.fit.edu

## **Faculty Sponsor and Client:**

Philip Chan	pkc@fit.edu
-------------	-------------

## Document Outline

1. Introduction
2. Definitions
3. System Architecture Diagram
  - 3.1. UML Diagram
  - 3.2. Functionalities of Each Module
4. Sketch of General User Interface
  - 4.1. Illustrate Key Functionalities
5. Algorithms, Database, Protocols

## 1. Introduction

This document describes the design document for the RepoMining project, a web repository for temporal data intended for data mining and analysis. The system integrates data from different resources and allows users to upload, download, and visualize data featuring the projects GUI. The main focus of this GUI is to be able to verify that there was a change in the target variable specified by the user, come up with the variables that affected the target, find correlation onto why was there a change, and to predict the possible value after a certain period of time.

## 2. Definitions

### **GUI**

Graphical User Interface which allow users to interact with the repository,

### **UML Diagram**

Unified Modeling language which allows visualization of the key parts and method of the repository.

### **Question 1**

Is there a significant change in the target variable?

### **Question 2**

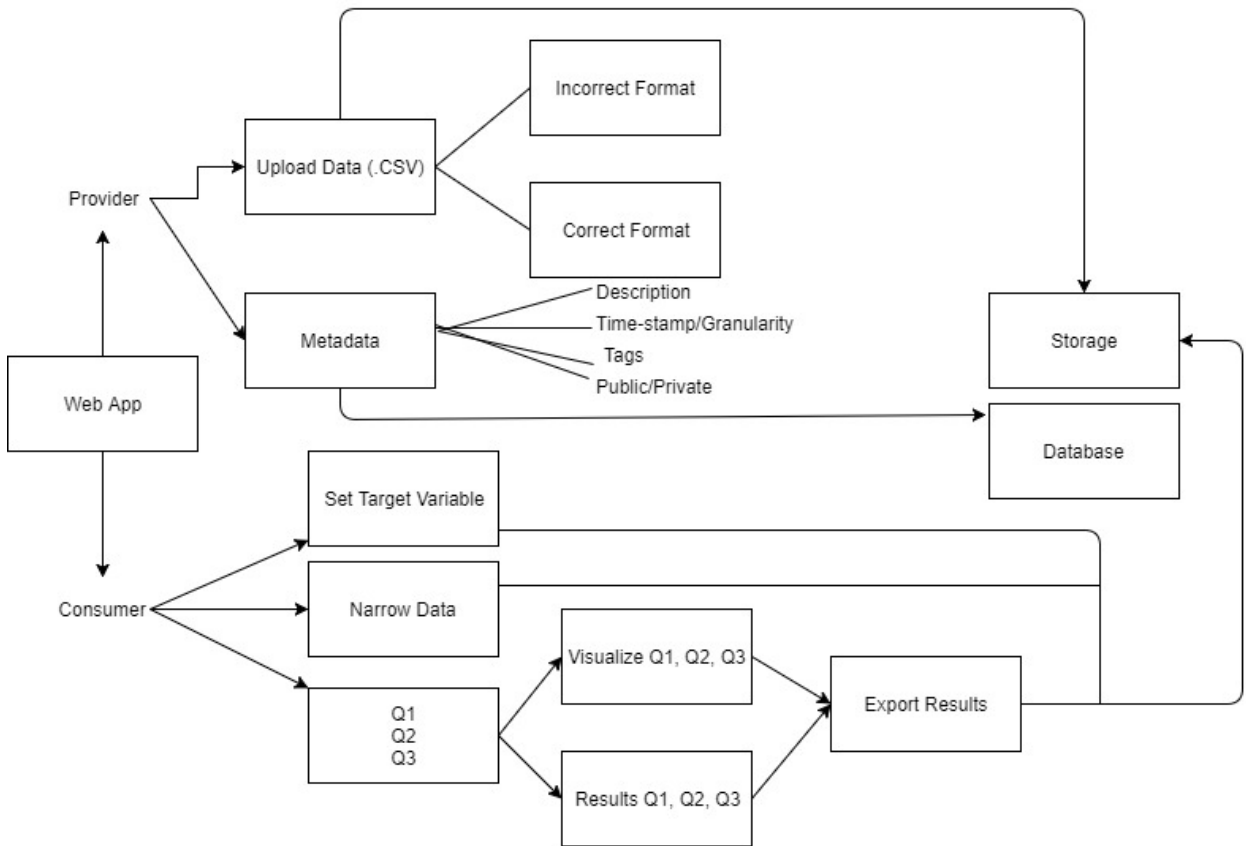
Why was there a significant change, and what are the top-k variables that are correlated to the target variable?

### **Question 3**

What is the value of the target variable in the next timestamp?

### 3. System Architecture Design

#### 3.1. High-level Functional Diagram



#### 3.2. Functionalities of Each Module

##### .CSV File

Data is obtained in .CSV files and the program determines if the file is written in the correct format.

##### Upload Data

The data in the file is analyzed and if in the program determines that the data can be accepted, it stores a description of the file and its data for future use.

##### Set Tag

The program is able to create user-defined tags to the documents to allow a quickly and more organized search through the repository to later compare the data and calculate correlation between the uploaded data and the saved data.

##### Set Target

User defines the target variable to the program manually.

## Metadata

Includes a description of the variables, a tag for narrowing the results, start/end timestamp, and if the data is of public or private use.

## Narrow Data

The search of correlated variables is narrowed down to be able to answer Questions 1, 2, and 3. The software will have no more than 10 correlated variables found by performing Linear Regression and Pearson Correlation; search is implemented as well as the selection of variables.

## Question 1

Program determines if there was a significant change in the target variable defined by the user; the execution of the algorithm should take no more than 1 second.

## Question 2

Program compares the variables that mostly correlate and affect the target variable, answering Question 2. Execution of the algorithm should take no more than 1 second.

## Question 3

If there was a significant change in the target variable, and if there was correlation between the compared data, the program calculates and forecasts the next possible value for the target variable for some determined period of time.

## Visualize Data

The program outputs a graph showing the comparison between the target variable, the top-k variables, and the next possible value.

## 4. Sketch of General User Interface

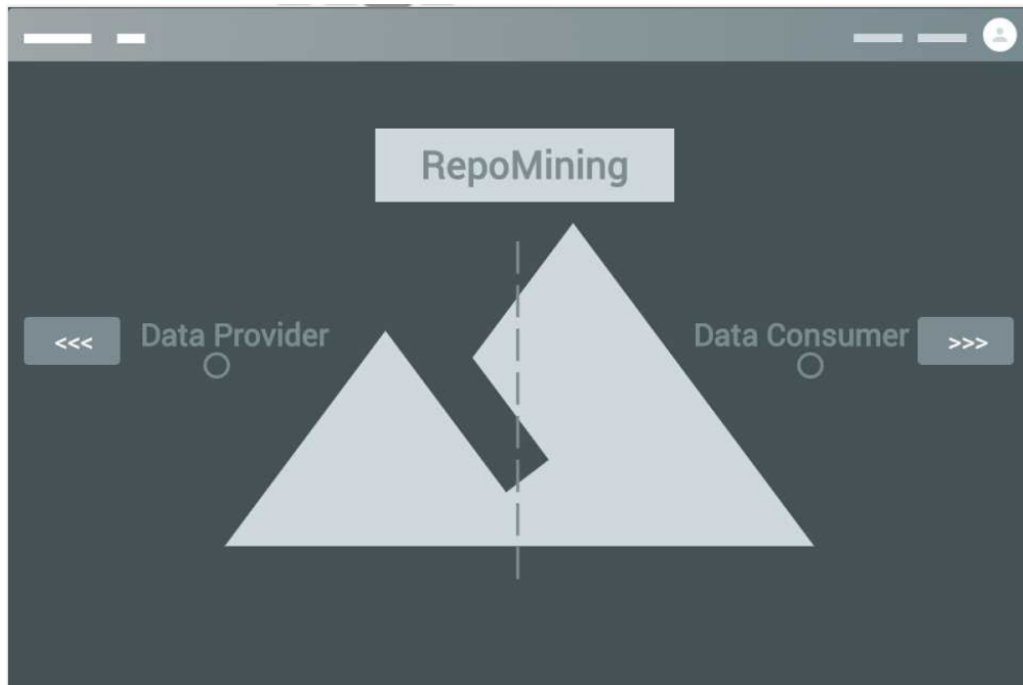
### 4.1. Illustrate Key Functionalities

There are seven pages on the Web Application planned: a Main Page, Data Provider Page, Choose of Metadata, Data Consumer Page, Question 1 Page, Question 2 Page, and a Question 3 Page.

#### Main Page

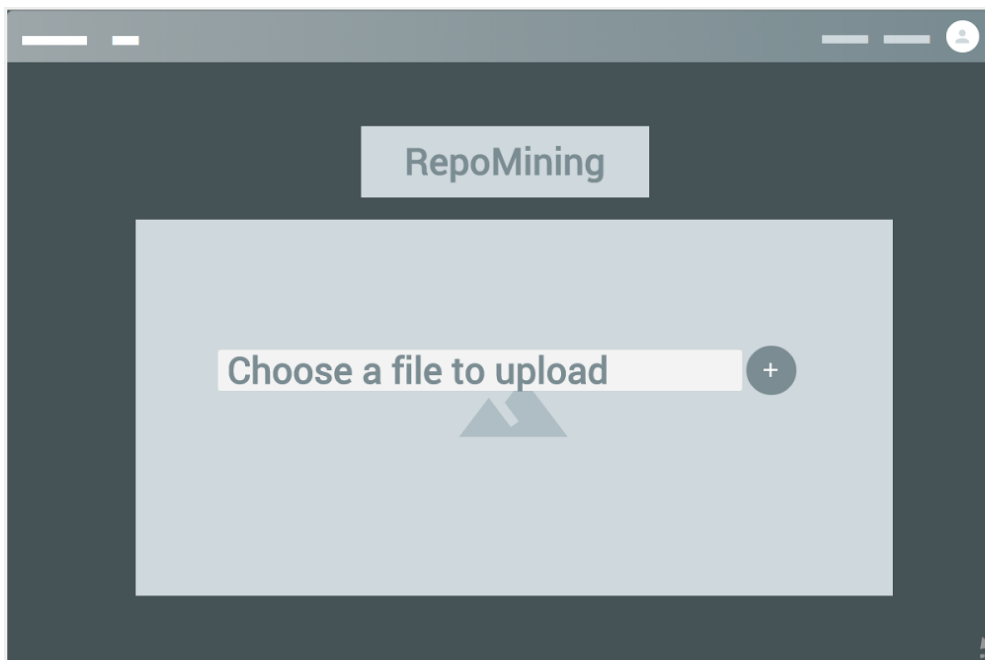
Features the name of the project, RepoMining, as well as a quick welcome message. The page is split into halves: one half has a display of 'Data Provider', indicating that if the user is

providing data that they should click on the double-arrow button to be redirected. The other side is the 'Data Consumer' side, so if the user is looking to use data they should click on the double arrows on the right side to be redirected to the Data Consumer Page.



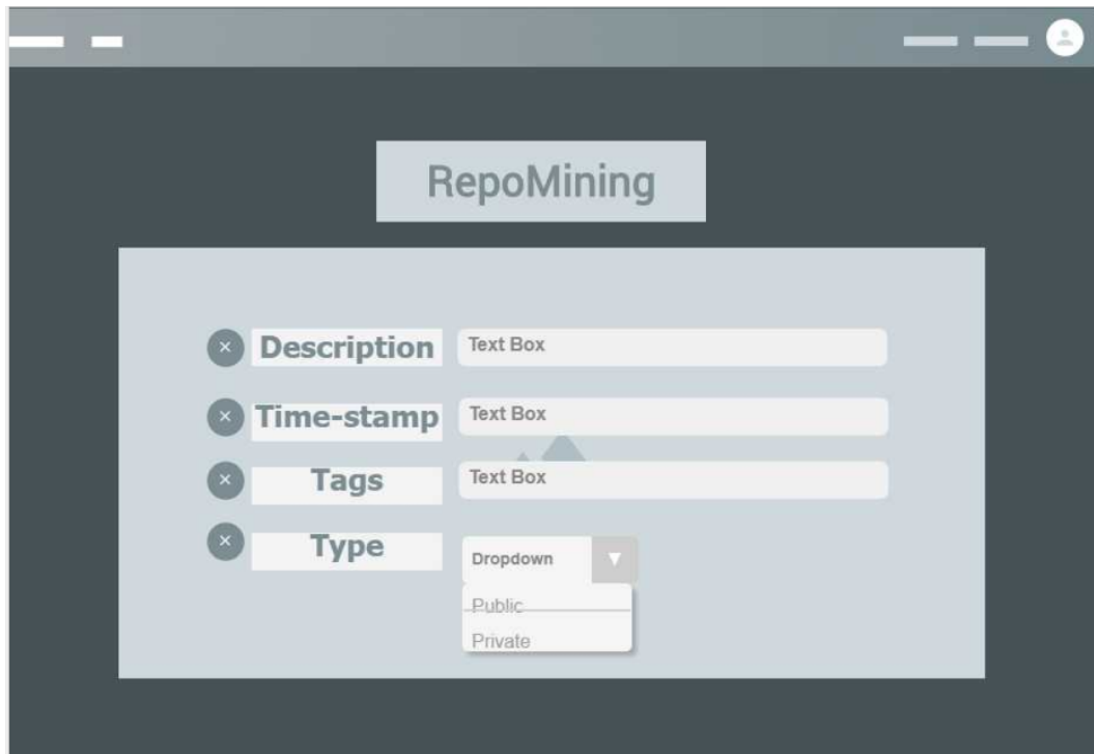
### Data Provider Page

Features an area to upload a file, and an upload button. Once file has been accepted, prompt user to input tags. Error message is planned to appear underneath the 'RepoMining' text.



## Choose of Metadata

Add description, timestamps, tags, and type of the data uploaded.

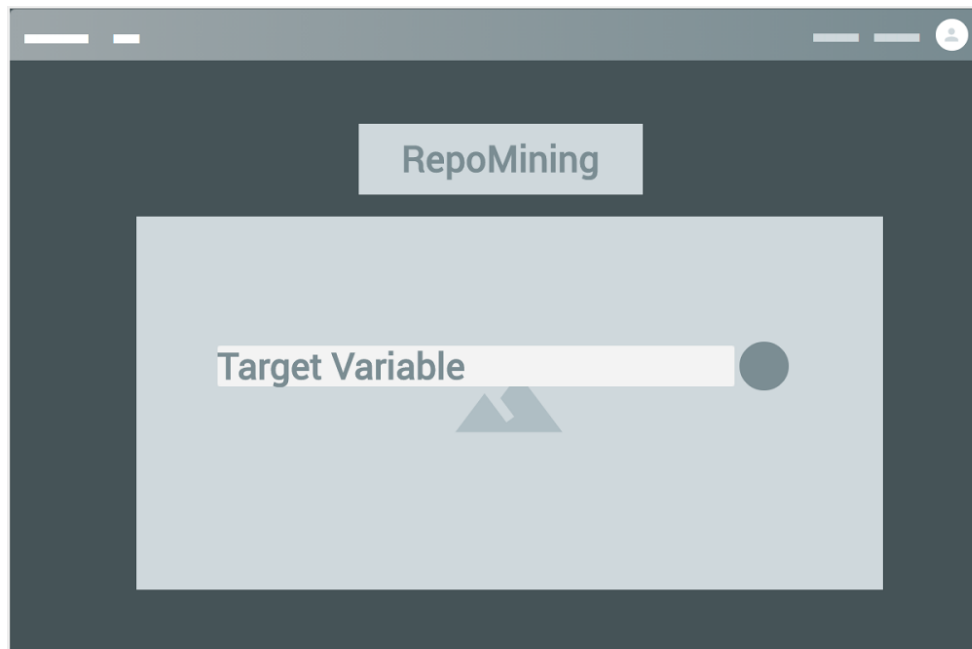


The screenshot shows a web application window titled "RepoMining". Inside the window, there is a light gray panel with four metadata fields, each with a close button (an 'x' in a circle) on the left:

- Description**: A text box labeled "Text Box".
- Time-stamp**: A text box labeled "Text Box".
- Tags**: A text box labeled "Text Box".
- Type**: A dropdown menu labeled "Dropdown" with a downward arrow. The dropdown is open, showing two options: "Public" and "Private".

## Data Consumer Page

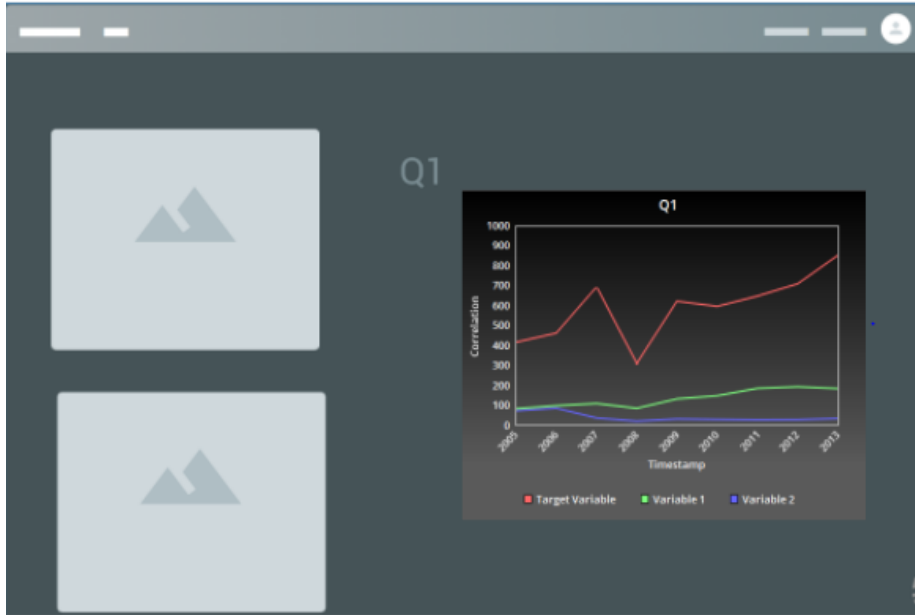
Prompts user for a target variable. Future enhancements is to allow user the option to insert additional search tags to narrow data down.



The screenshot shows a web application window titled "RepoMining". Inside the window, there is a light gray panel with a single input field labeled "Target Variable". The field has a dark gray circular button on its right side. A large, faint watermark of a triangle is visible in the background of the panel.

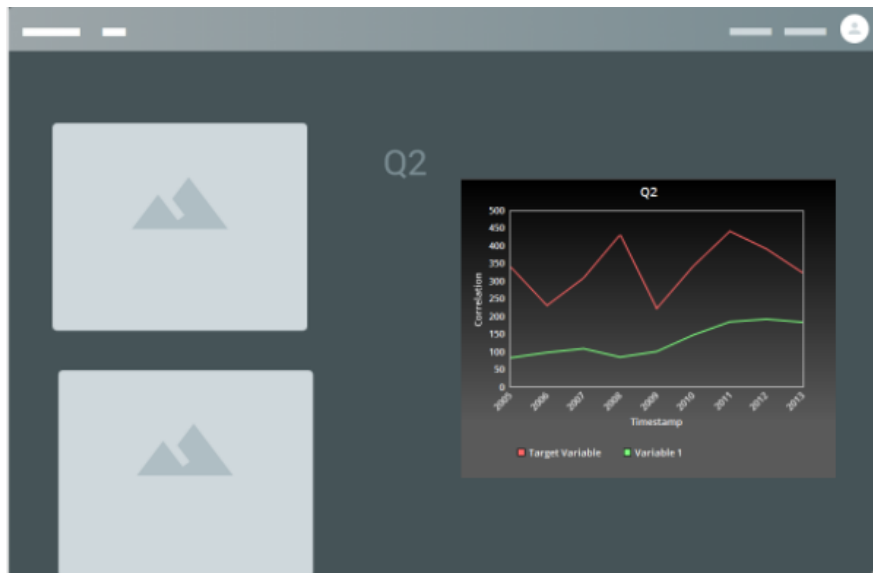
## Question One Page

Displays the results of answering Q1 in visual/graphic form. There is a double arrow button on the right side to move to the next question. Also featured on the page is a button to download the csv file for Q1.



## Question Two Page

Displays the results of answering Q2 in visual/graphic form. There is a double arrow on both sides to go backwards to Question One Page or forwards to Question Three Page. Also featured on the page is a button to download the csv file for Q2.





### Question Three Page

Displays the results of answering Q3 in visual/graphic form. There is a double arrow on both sides to go backwards to Question Two Page or forwards to the Main Page. Also featured on the page is a button to download the csv file for Q3.



## 5. Algorithms, Database, Protocols

- Algorithms for z-Score, Linear Regression, Pearson Correlation.
- SQL for Database setup for Metadata.
- Search Algorithms.
- Narrowing Data.
- Data Organization in Database.
- Files Repository.