Repository and Mining of Temporal Data

Jessica Nguy
Siomara Nieves

Dr. Philip Chan

Progress of Current Milestone

Task	Completion%	Jessica	Siomara	To-Do
Investigate/Select Tools	100%	50%	50%	none
Investigate additional plugins/sites	60%	20%	40%	Plotting software and web-to-app plugins
Create .CSV program, “Hello World”demos	100%	50%	50%	none
Requirements Document	100%	75%	25%	none
Design Document	100%	50%	50%	none
Test Plan	100%	60%	40%	none

Discussion of Each Accomplished Task

Investigate/Select Tools: Investigated and selected collaboration tools, programming languages, and algorithms to use for analysis. Currently are using a combination of Slack, Gmail, and Google Docs for collaboration and GitHub to host code. We are using Python for our main language and SQL for our database.

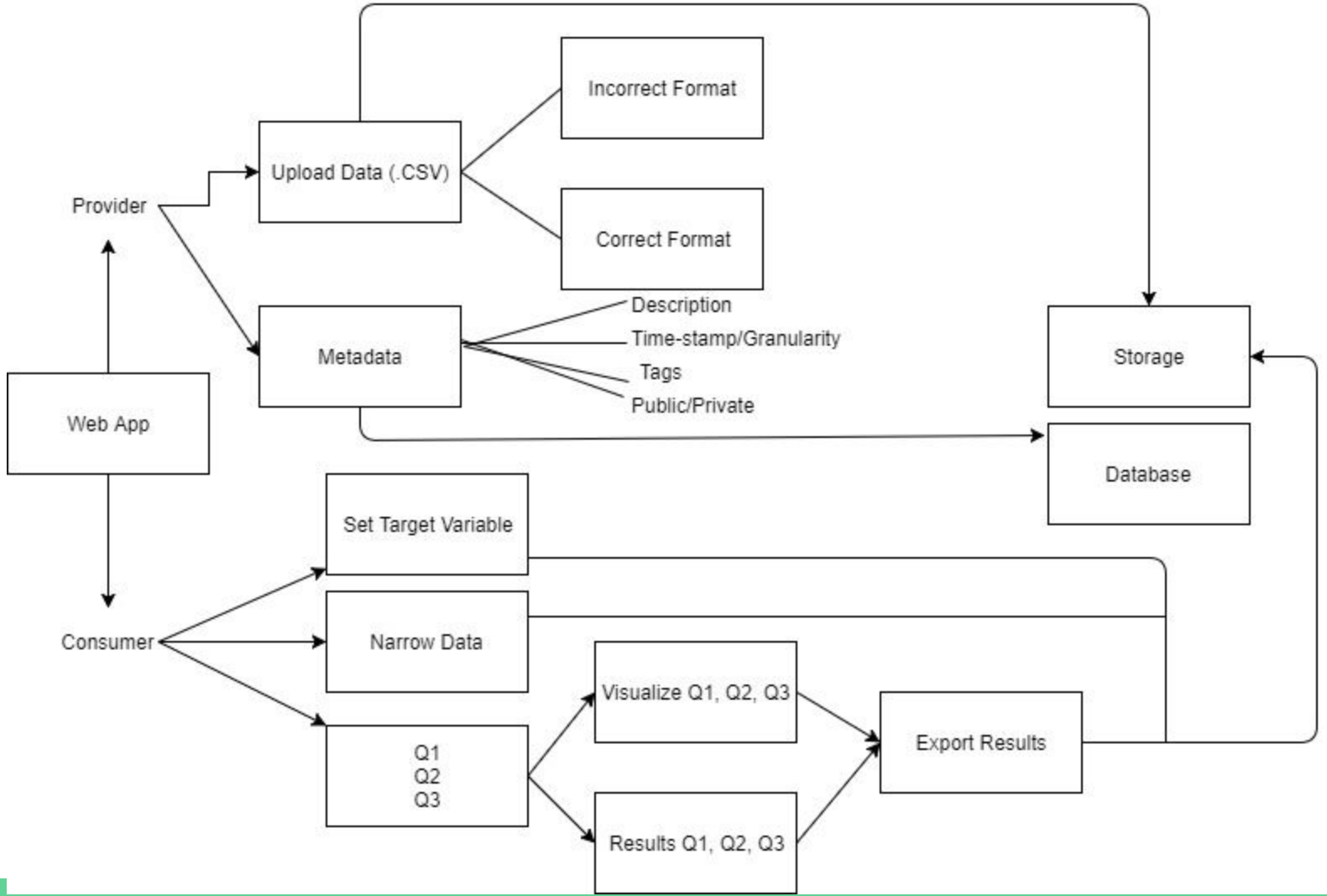
Investigate additional plugins/sites: Plugins were researched; currently looking at Django to implement and host our program on the web at the same time. Algorithms that we are using for the project is the calculations for z-score, standard deviation, Pearson Correlation, cross correlation, and linear regression.

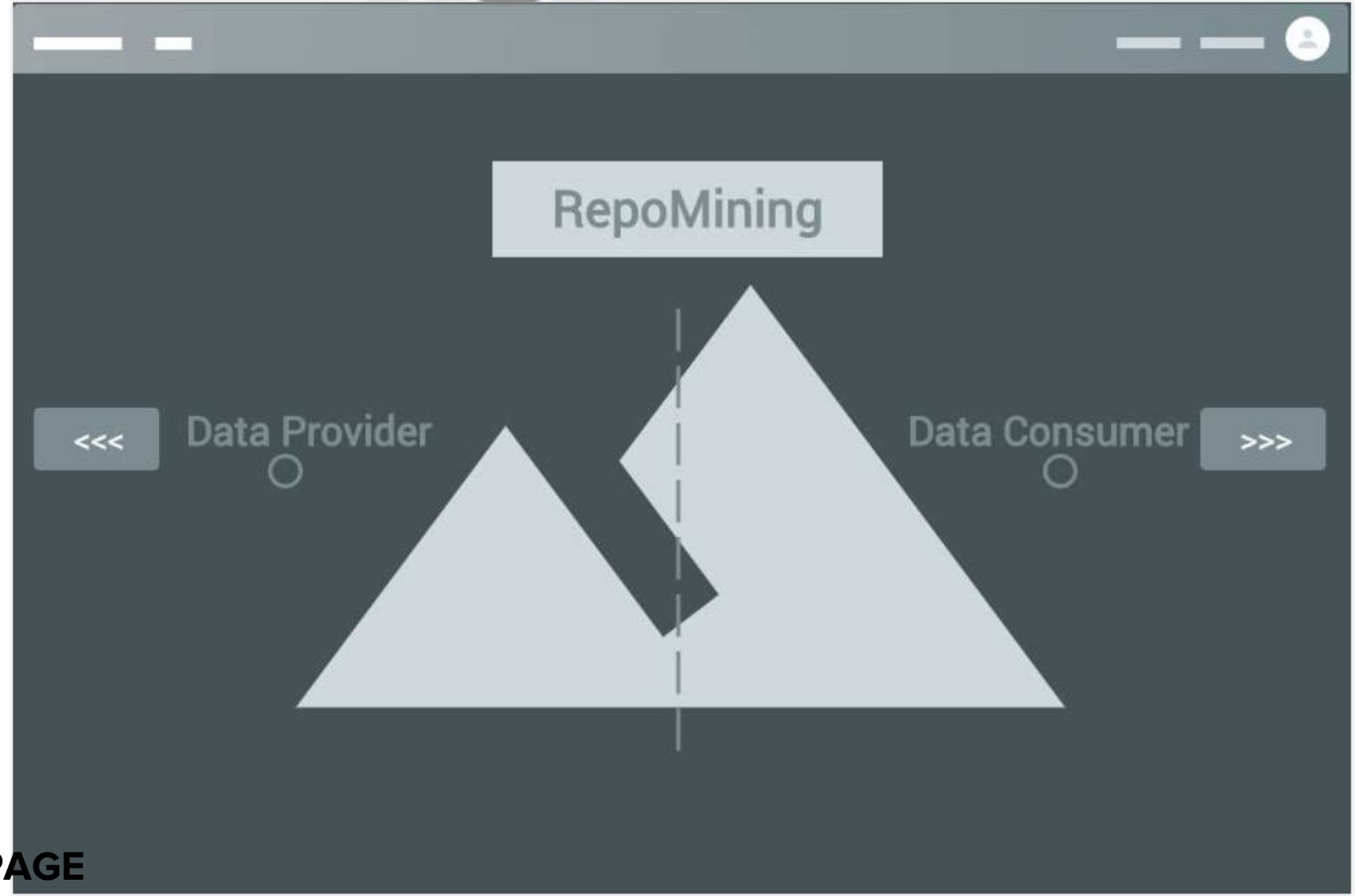
Create .CSV program, Hello World demos: a .CSV reader program was created as a demo to test the file input from the user and check if the file has the correct formatting with the data. The program simply takes the input file and prints it out on the console.

Requirements Document: Requirements Document was created using Google Docs. Issues that arose during the writing of this document was the need to be specific for certain requirements. Concepts that the project needed were also brought up as requirements, added to the document, and edited as necessary.

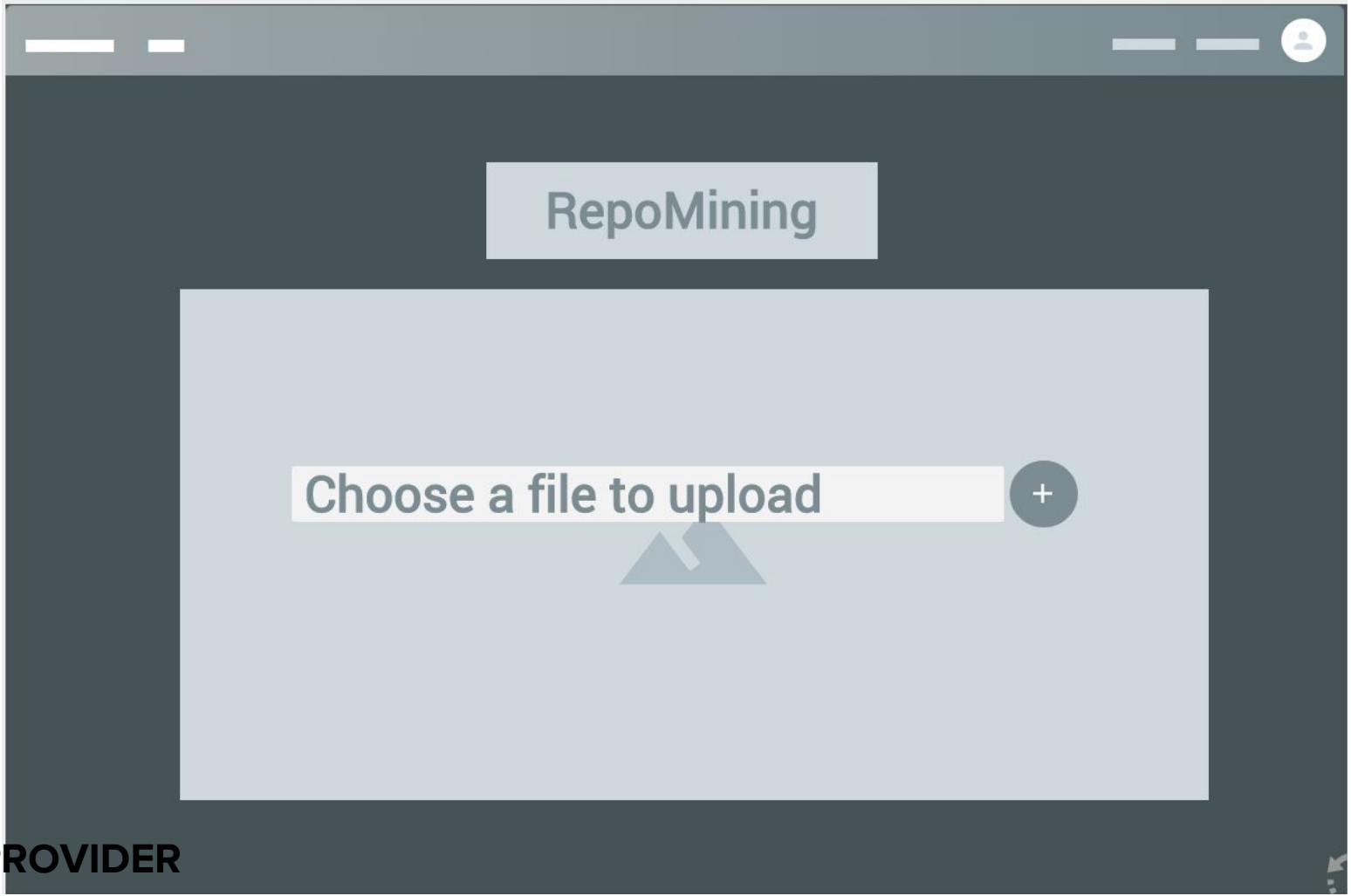
Design Document: Design Document was created using Google Docs. Includes the program's high level functional diagram and sketches for the GUI.

Test Plan: Test Plan was created using Google Docs. The document uses requirements from the Requirements Document and elaborated on its function and expected outputs to be considered as correct. There were no issues that appeared during writing this document.

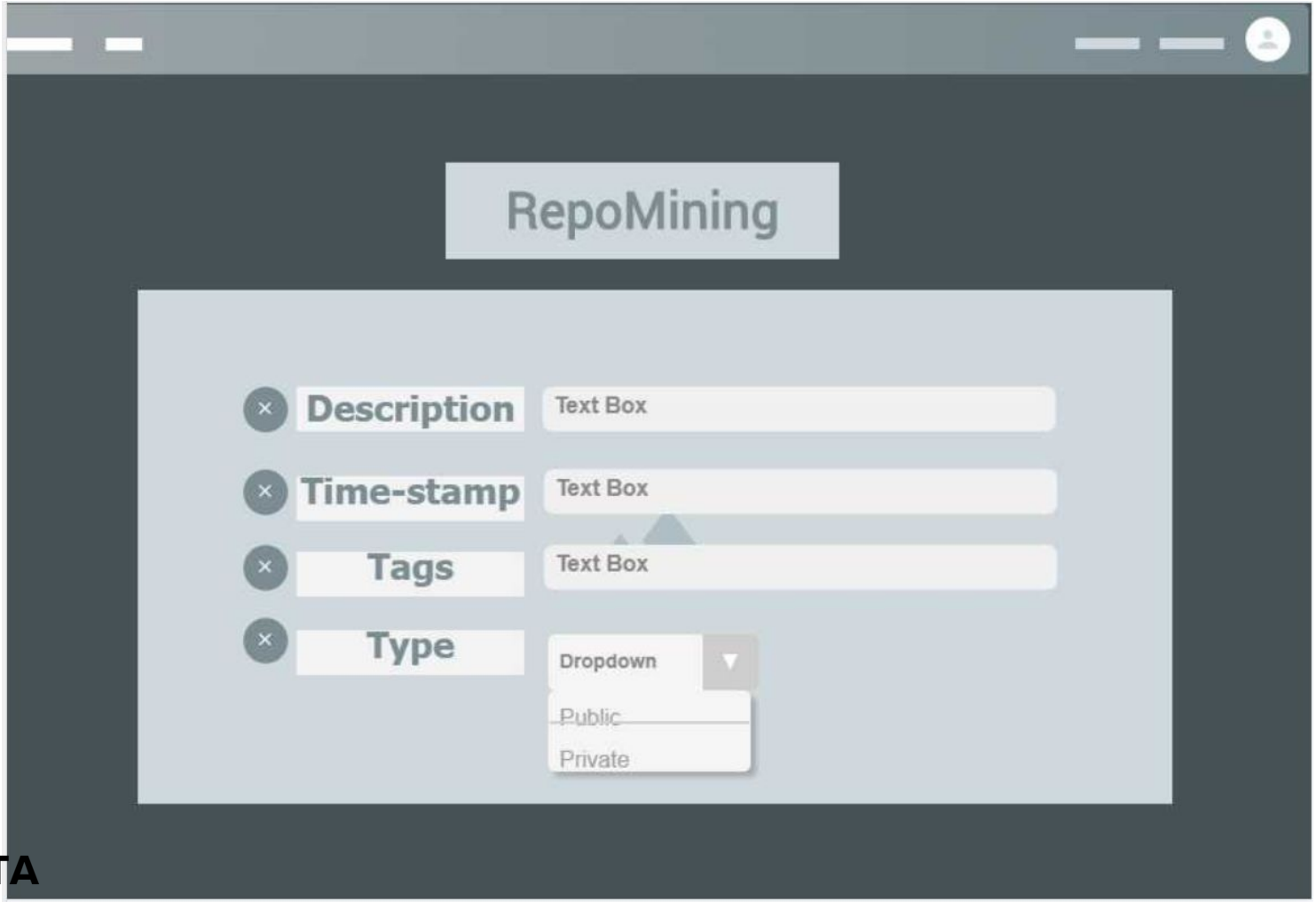

Incorrect Format
Upload Data (.CSV)
Provider
Correct Format
Description
Time-stamp/Granularity
Metadata
Tags
Public/Private
Storage
Web App
Database
Set Target Variable
Consumer
Narrow Data
Visualize Q1, Q2, Q3
Q1
Q2
Q3
Export Results
Results Q1, Q2, Q3



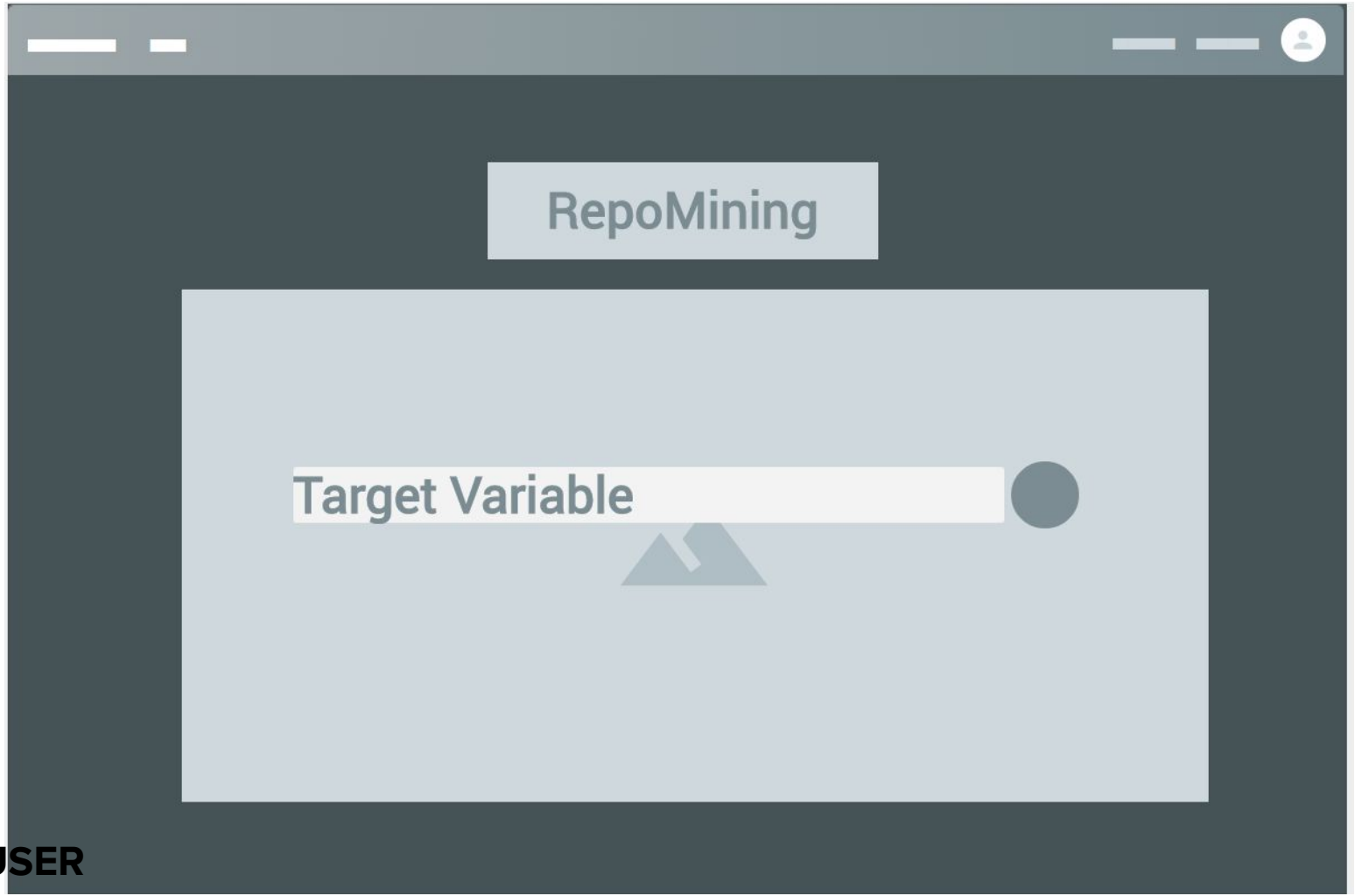
MAIN PAGE



DATA PROVIDER

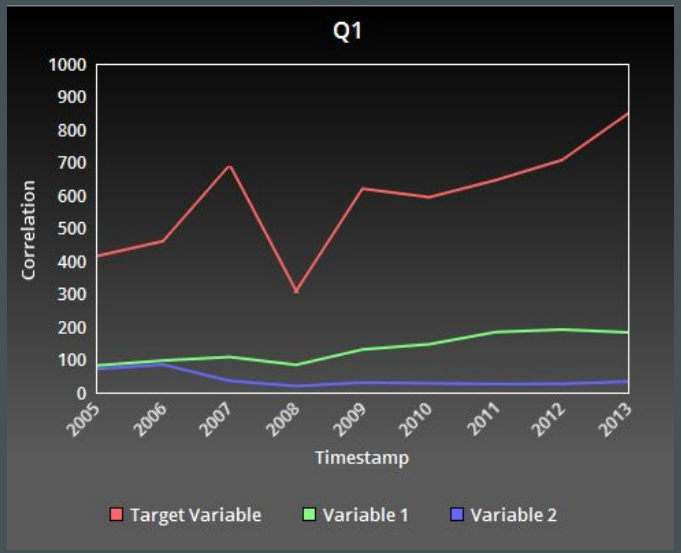

RepoMining
Description
Text Box
Time-stamp
Text Box
Tags
Text Box
Type
Dropdown
Public
Private

METADATA

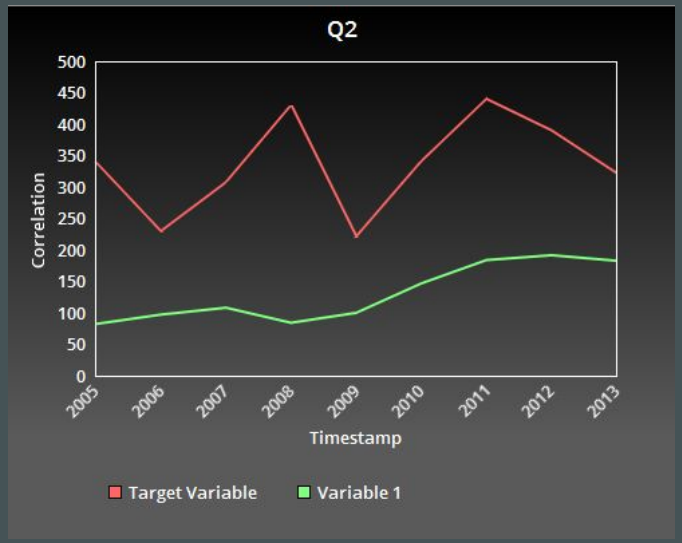


DATA USER

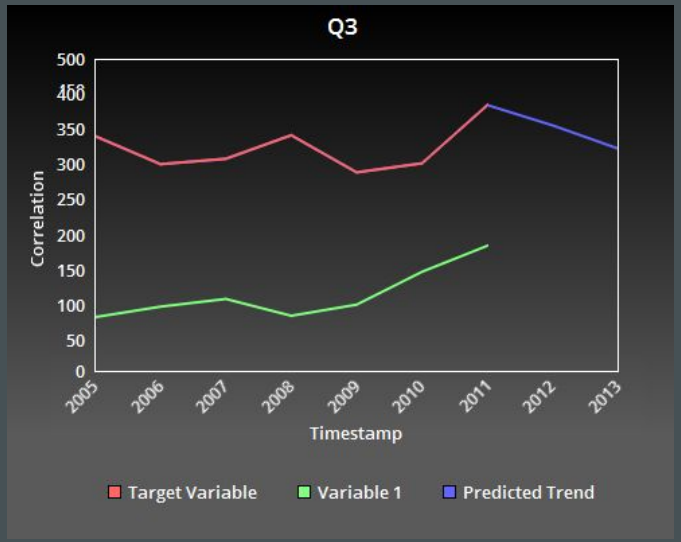
Q1


Q1
1000
900
800
700
600
500
400
300
200
100
0
Correlation
2005
2006
2007
2008
2009
2010
2011
2012
2013
Timestamp
Target Variable
Variable 1
Variable 2

Q2


Q2
500
450
400
350
300
250
200
150
100
50
0
Correlation
2005
2006
2007
2008
2009
2010
2011
2012
2013
Timestamp
Target Variable
Variable 1

Q3


Q3
500
450
400
350
300
250
200
150
100
50
0
Correlation
2005
2006
2007
2008
2009
2010
2011
2012
2013
Timestamp
Target Variable
Variable 1
Predicted Trend

Plan for Next Milestone

Task	**Jessica**	**Siomara**
.CSV User Input catch cases	Write and test catch cases	Write and test catch cases
Q1	Coding math-equations. visualizations	Client input for target variable, export .CSV file
Website	Host web site	Website vulnerabilities
Data Processing	Set Target Variable. Seach	Narrow Data, find data to use
Database setup for metadata, metadata inputs	SQL, database setup	SQL. database setup

Discussion for Each Planned Task

.CSV User Input catch cases: Need to add a ‘catcher’ code to indicate if users uploaded incorrect formats for csv - csv files need to have timestamps going across the top row with variables in the first column, or timestamps going down in the first column with variables names in the first row with subsequent data in columns going down.

Q1: Client requires that Question 1 be completed by the next Milestone. Target Variable needs to be set. Other variables need to be narrowed down using another algorithm. Question 1 needs to be able to visualize the data and create a csv export of results.

Website: Needs at least 3 pages. A working main page, a page for data providers, and a page for data consumers. Buttons need to be working.

Data Processing: Set Target Variable, Search, and Narrow Data functions need to be programmed.

Database setup for meta-data: Database creation using SQL to save the description of the target variable, its tags, initial and final timestamps, and if data is of public or private use.