1.INTRODUCTION

1.1 Purpose

This software design document describes the architecture and the system design of the Stack overflow Analysis software system version 1.0 for analyzing and visualizing survey data collected from developers who use the stack overflow website.

1.2 Scope

The system analyses and visualizes developer data to determine the patterns and trends as well as the insights in the data collected.

Its based in RStudio, a free and open source integrated development environment for R programming

language installed on a computer with the Shiny packages and other relevant packages.

The objective of the system is to provide a visual and statistical representation of responses from Stack overflow users to assist these the website’s management team in making

operational plans, strategic planning and decision making to make the stack overflow more useful.

The goals of the system are:

The goal of this system is to critically analyse developer survey data so as to find a way on how to improve on developer marketing, technical recruiting, market research and enterprise knowledge sharing at Stack Overflow.

i. To determine the patterns in a given country to help the management team deploy resources more effectively and also assist the technical team in developing better products for the website

ii. To determine the levels of coding experience among site contributors to better assess authenticity of answers to questions on the website

iii. To determine the trends at which the open source projects are occurring . This will provide a scientific basis for decisions and actions aimed at improving operations.

This project will aid in the analysis process and also provide a scientific basis for making

decisions and taking actions aimed at improving operations of the Stack overflow website

1.3 Overview

This document has 8 sections that explain its use to the user

The Introduction defines the system’s objective and the summary of the system functionality

so as to give the reader a good understanding of the system goals.

The System Overview explains to the reader the general system functionality and its design.

The System Architecture has the detailed view of the different conceptual requirements for

setting up the system in terms of hardware and software.

The Data Design defines the data storage techniques and the format of the data for each of

the stored data files.

The Component Design describes how the different components of the system interact to

satisfy the user needs.

The Human machine interface explains to the reader the needed skills of the system’s

interface in order to be able to interact freely with the system.

The Requirements matrix shows the system components that satisfy each of the functional

requirements from the SRS document.

Appendices contains the glossary where all the key terms used in this document are defined

1.5 Definitions and Acronyms

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Crime Data: The original data records from the various provinces and stations in South Africa.

Crime Trends: A recognizable general tendency regarding recurring patterns of crime which is

revealed over a period of time. A trend may involve any one of the crime pattern factors or

any combination of the factors.

Crime Pattern - The occurrence of similar offenses in a defined geographic area, usually

defined by administrative boundaries.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

2. S YSTEM OVERVIEW

Due to the rise in demand for software developers across the world ,and the need for a website that can be a platform for interaction for them (developers), the Stack oveflow Analysis system has been developed with an aim of analyzing and visualizing developer statistics collected from the users of the website using the

various statistical tools of analysis, visualization and prediction to assist in planning and

decision making. The system context diagram below shows the system’s functionalities.

System Context Diagram

3. S YSTEM ARCHITECTURE

3.1 Architectural Design

The diagram below shows how the system was decomposed giving a general understanding of how the individual sub systems work together.

Conceptual Diagram

3.2 Decomposition Description

The system functionality has been broken down and is represented as a functional

decomposition diagram and data flow diagrams.

Functional Decomposition Diagram

3.3 Design Rationale

The chosen architecture in 3.1 is good for authentication purposes, and it is cheaper to

construct.

4. D ATA DESIGN

4.1 Data Description

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

The system will not have any database. The user will be required to upload an excel sheet having a data format explained in section 4.2 This file will be uploaded for every user session and will only be stored temporarily.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

4.2 Data Dictionary

The table below shows the expected contents of each work sheet in the data store.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

|  |  |  |
| --- | --- | --- |
| COLUMNS | DESCRIPTION | DATA TYPE |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

5. C OMPONENT DESIGN

Module 1: User Verification module

Flow Chart Diagram

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Module Inputs:

 Google account login details (password and email).

Module outputs

 Verification result

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Module 2: Validate data file

The uploaded file must match the file template which is described in section 4.2. When a

user uploads a file, it will be checked to ensure that it matches the template. If it doesn't

exactly match the template, it will be rejected and the user will be prompted upload another

file. This module will be executed for every session of a user. It will only run after module 1

has granted access to the user.

Flow Chart Diagram

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Module Inputs:

 Crime excel data file

Module outputs

 File Validation Result

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Module 3: Data Cleaning

This module involves reading the crime data file and making valid variable names from the

columns in the excel sheet. The spaces in the column names must be replaced with default

values, for instance; if the column name is Education level, its variable name

should be Education.level. This happens to every sheet that has is stored in the

data store.

This module will be called once after execution of module 1 and 2 for every successfully

loaded file.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Module Inputs

 Crime data file

Module Outputs

 Cleaned data

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Module 4: Data Analysis

This module is the core part of the system and will be called frequently and its results will

depend on the user commands. The user will choose which parameters to be analyzed and how they should be analyzed using the graphical interface.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Module inputs:

 User selections/choices

 Crime data file

Module outputs:

 Analysis results

 Textual reports

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Module 5: Graphical Representation of results

The analytical results from module 4 will be presented in a graphical view based on the user

selections. A number of graphical features including but not limited to line graphs, bar

graphs, pie charts, box plots will be used by this module to display data to the user. The user

will select the tool they wish to use to view the results.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Module inputs

 User selections/choices

 Analysis results

Module outputs

 Graphs

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

6. H UMAN INTERFACE DESIGN

6.1 Overview of User Interface

Interface Architecture

This system will interface with the Google Verifier to validate the user credentials so as to

ensure that only authorized users are accessing the system. When the user has entered

his/her Google account login credentials, the details will be forwarded to the Google

verifier which responds by giving validations results.

Interface Architecture Diagram

Interface Detailed Design

The system users will be required to have Google email accounts which will be added to

the system by the system administrator. On attempt to access the system, the user will be

authenticated by entering his/her Google credentials. The credentials will be sent to the

Google verifier to check the validity of the user, the verifier then returns the results to the

system which then decides whether to grant or deny access to the user.

6.2 Screen Images

Log in Screen Image

6.3 Screen Objects and Actions

The screen image figure 6.0.2 will have a provision for logging in via Google.

7. R EQUIREMENTS M ATRIX

Requirements Table

|  |  |
| --- | --- |
| FUNCTIONALITY REQUIREMENTS | REFERENCE SECTION IN SRS |
|  |  |
|  |  |
|  |  |

Table 3: Requirements table

8. A PPENDICES

8.1

GLOSSARY

|  |  |
| --- | --- |
| TERM | DEFINITION |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

Table 4: Glossary

9.REFERENCE MATERIAL