Low Level Design

Crime Analysis in Baton Rouge

|  |  |
| --- | --- |
| Written By | Prateek Jha |
| Document Version | 1.0 |
| Last Revised Date | 18th May 2022 |

**Document Control**

### Change Record:

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date** | **Author** | **Comments** |
| 1.0 | 15 – May -  2022 | Prateek Jha | Introduction & Architecture defined, Architecture & Architecture Description appended and updated |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

### Reviews:

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date** | **Reviewer** | **Comments** |
|  |  |  |  |

### Approval Status:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Version** | **Review**  **Date** | **Reviewed By** | **Approved By** | **Comments** |
|  |  |  |  |  |

Contents

1. [Introduction 1](#_bookmark0)
   1. [What is Low-Level design document? 1](#_bookmark1)
   2. [Scope 1](#_bookmark2)
2. [Architecture 2](#_bookmark3)
3. [Architecture Description 3](#_bookmark4)
   1. [Data Description 3](#_bookmark5)
   2. [Web Scrapping 3](#_bookmark6)
   3. [Data Transformation 3](#_bookmark7)
   4. [Data Insertion into Database 3](#_bookmark8)
   5. [Export Data from Database 3](#_bookmark9)
   6. [Data Pre-processing 3](#_bookmark10)
   7. [Data Clustering 3](#_bookmark11)
   8. [Model Building 4](#_bookmark12)
   9. [Data from User 4](#_bookmark13)
   10. [Data Validation 4](#_bookmark14)
   11. [User Data Inserting into Database 4](#_bookmark15)
   12. [Data Clustering 4](#_bookmark16)
   13. [Model Call for Specific Cluster 4](#_bookmark17)
   14. [Recipe Recommendation & Saving Output in Database 4](#_bookmark18)
   15. [Deployment 4](#_bookmark19)
4. [Unit Test Cases 5](#_bookmark20)

# Introduction

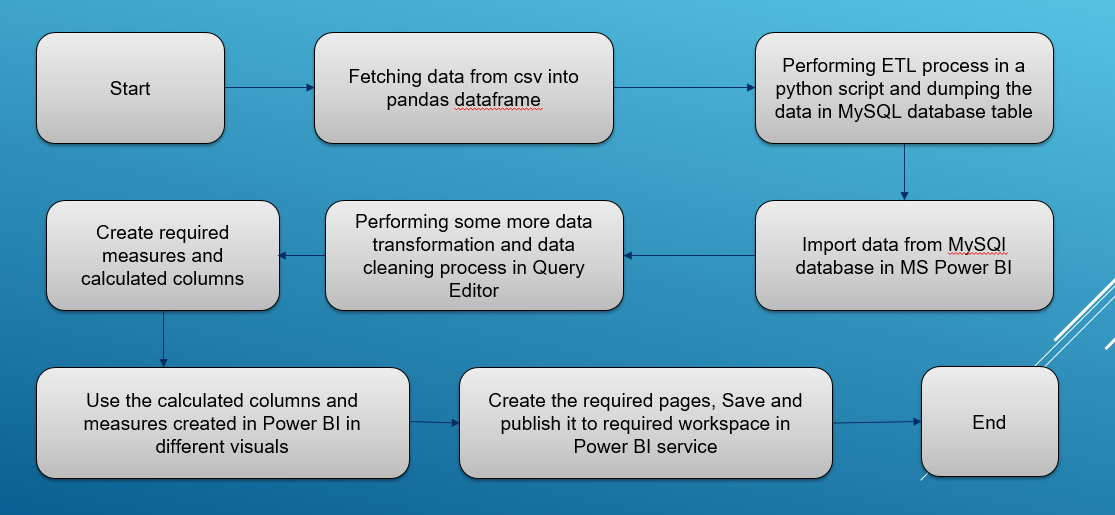
## What is Low-Level design document?

The goal of LLD or a low-level design document (LLDD) is to give the internal logical design of the actual program code for Crime Analysis report prepared for the given dataset. LLD describes the class diagrams with the methods and relations between classes and program specs. It describes the modules so that the programmer can directly code the program from the document.

## Scope

Low-level design (LLD) is a component-level design process that follows a step-by-step [refinement](https://en.wikipedia.org/wiki/Refinement_(computing)) process. This process can be used for designing data structures, required software architecture, source code and ultimately, performance of the final product. Overall, the data organization may be defined during requirement analysis and then refined during data design work

# Architecture



# Architecture Description

## Data Description

The dataset available with us contains 400k + rows for different incidents with different and unique file numbers for each offensive incidents across various categories, areas , different years and different directions in the city of Baton Rouge. This consolidated dataset is provided by multiple banks whose transactions is monitored and is updated every 1 month on this link:  <https://drive.google.com/file/d/1KxBXQZbAN-6kGrRx3yOwndoYeInnhyTV/view>

## Data Transformation

In the Transformation Process, we will be using a python script for loading data and making some minor conversions for easy use in Power Bi report construction

## Data Insertion into Database

1. Database Creation and connection - Create a database with name passed as “FinancialCrimeAnalysis”. If the database is already created, open the connection to the database using command in MySQL as “USE FinancialCrimeAnalysis”
2. Tables to be created in the database named “tAlerts”, “tTransactions” and “tAccounts”.
3. Insertion data from the available pandas dataframe in the python script from the source file named “Accounts.csv”, “Trasnsactions.csv” and “Alerts.csv”

## Data Pre-processing

Data Pre-processing includes Offense date rectification, office time rectification and filling up of some null values by city value of “Baton Rouge” as the data is available for this city only.

For this project, currently there was no data cleaning or data conversion was needed/ done as we are mostly numerical values and we didn’t observe and absurd values that needed cleaning initially for all the 3 datasets involved.

## Data from User

The data provided has various informative columns across 3 datasets for Accounts, Alerts and Transactions. As per our current understanding, we know that for each account id in Accounts, they can have multiple transactions on their name and each transaction can have one unique alert on their name if the fraudulent flag is true.

## User Data Inserting into Database

Collecting the data from the user and storing it into the database. The database is in local instance of MySQL database.

## Deployment

Once the data is imported after cleaning and stored in MySQL source database tables, we create reports with help of various insightful visuals using calculated columns and measures created in Power BI depending on the message we are trying to convey using the data and the most suitable visual.