

# CSCI 721 Group - Data Miners

June 28, 2020

This is the documentation for installation and usage of CSCI 721: Group Project

## 1 Installation

Use the package manager [pip] to install pymysql, mysql.connector, sqlalchemy, pandas and NumPy to perform necessary operations

```
pip install pymysql
pip install MySQL.connector
pip install sqlalchemy
pip install Pandas
pip install NumPy
```

For data loading cleaning and merging the file SQL.py is used. For structuring the dataset in MySQL queries from the file SQL\_Queries\_to\_make\_2\_databases\_uniform,SQL are used.

## 2 Usage

From the file SQL.py the function setSQLCredentials is used to set MySQL credentials from the config file

```
setSQLCredentials()
```

The function load\_datasets are used to load all the datasets into MySQL database. Where we take the take of data quality and dimensionality issues.

```
load_datasets()
```

Then the data is loaded into Python for performing data cleaning, analysis and visualization.using the function GetDataset

```
GetDataset(query)
```