**README FILE**

**PROJECT TITLE:**  Parse and load sample statement to the database

**TOOLS AND REQUIREMENTS:**

**DATABASE:** MYSQL

**PROGRAMMING LANGUAGE:** PYTHON

**IDE’S:** JUPYTER NOTEBOOK(PYTHON), MYSQL WORKBENCH(DATABASE)

**STEP BY STEP EXPLINATION(Jupyter Notebook):**

**Step 1**: Import tabula

Tabula is a package which is used to liberate data tables which is inside a pdf file.

**Step 2**: Read the pdf file.

**Step 3**: Converting the pdf file to csv format and printing it on the console.

**Step 4**: Import pandas

Pandas is a python package which is used to read file formats such as csv,xlsx etc..

**Step 5**: Reading the converted pdf file.

The file is then displayed on the console.

Note: Since the package is not that reliable, it has printed a lot of garbage values. So, I directly downloaded the tabula tool and converted the pdf to csv file.

**Step 6**: Reading the given “bank reconciliation file” and printing it on the console.

**Step 7**: I converted the given “xls” file to “csv” format because the database only accepts csv format.

**Step 8**: Since the given data has a lot of “NaN” values, I filled it with blank spaces using fillna() command.

**Step 9**: Reading the “summary of investment” csv file which was extracted from the pdf file.

**Step 10**: As there are some garbage values, I removed both the columns using del function.

**Step 11**: Reading the “"EFG Investment Bond Fund” csv file and printing it on the console.

**Step 13**: Reading the “QRS Funds Money Market FE” csv file and printing it on the console.

Note: I used fillna() on all these files to remove “NaN” values.

**Step 14**: Import mysql.connector

This package is used to connect to the mysql database.

**Step 15**: After connecting to database, I have read all the data from the database to the console

Additionally, I have added an output screenshot file for your reference.