

Data_Ingestion

December 20, 2021

1 Data Ingestion

```
[1]: from IPython.display import display, HTML
import pandas as pd
import sqlite3
from sqlite3 import Error

def create_connection(db_file, delete_db=False):
    import os
    if delete_db and os.path.exists(db_file):
        os.remove(db_file)

    conn = None
    try:
        conn = sqlite3.connect(db_file)
        conn.execute("PRAGMA foreign_keys = 1")
    except Error as e:
        print(e)

    return conn

def create_table(create_table_sql, conn):
    try:
        c = conn.cursor()
        c.execute(create_table_sql)
    except Error as e:
        print(e)

def execute_sql_statement(sql_statement, conn):
    cur = conn.cursor()
    cur.execute(sql_statement)

    rows = cur.fetchall()

    return rows
```

```
[2]: conn = create_connection("normalized_ppg6.db", delete_db=True)
```

```
[3]: import os
def loadFiles(foldername='Datasets', fileformat='.csv'):
    csvfiles = list(filter(lambda f: f.endswith(f'{fileformat}'), os.listdir(f'.
↪/{foldername}')))
    csvfiles = [os.path.abspath(f"{foldername}/{file}") for file in csvfiles]
    return(csvfiles)

csvfiles = loadFiles()
```

```
[4]: %%capture
def loadDatabase(file_list=csvfiles, connx=conn, overwrite=True):
    for filepath in file_list:
        df = pd.read_csv(filepath)
        filename = os.path.basename(filepath)
        table_name = filename.rsplit('.',1)[0]
        write_mode = 'replace' if overwrite else 'append'
        df.to_sql(table_name, connx, if_exists=write_mode, index=False)

loadDatabase()
```

```
[ ]:
```