

0_data_collection

February 7, 2020

1 Assignment1, CMPT826

1.1 Data Collection

- Seyedeh Mina Mousavifar
- 11279515
- sem311

1.1.1 Battery table

```
In [ ]: import pandas as pd
import sqlalchemy as db

engine = db.create_engine('mysql://sem311:p^HA89/h@crepe.usask.ca:3306/SHED10')

connection = engine.connect()
metadata = db.MetaData()

# Equivalent to 'SELECT * FROM battery
battery = db.Table('battery', metadata, autoload=True, autoload_with=engine)
query = db.select([battery])

# getting data by executing the query above
BatteryResultProxy = connection.execute(query)
BatteryResultSet = BatteryResultProxy.fetchall()

# converting data to data frame
battery_data = pd.DataFrame(BatteryResultSet)
battery_data.columns = BatteryResultSet[0].keys()

# removing index column from data - 0 for rows and 1 for column
battery_data = battery_data.drop('index', 1)

# saving data frame as pickle object for future use
battery_data.to_pickle('data/battery.pkl')

print('fetching battery data completed')
```

1.1.2 GPS table

```
In [ ]: # Equivalent to 'SELECT * FROM gps
gps = db.Table('gps', metadata, autoload=True, autoload_with=engine)
query = db.select([gps])

# getting data by executing the query above
GPSResultProxy = connection.execute(query)
GPSResultSet = GPSResultProxy.fetchall()

# converting data to data frame
gps_data = pd.DataFrame(GPSResultSet)
gps_data.columns = GPSResultSet[0].keys()

# removing index column from data - 0 for rows and 1 for column
gps_data = gps_data.drop('index', 1)

# saving data frame as HDF5 object(because of its size) for future use
gps_data.to_pickle('data/gps.pkl')

print('fetching gps data completed')
```