KPLER
DATA ENGINEER INTERNSHIP TEST

Yuksel Selen ADALI 13.12.2018

OVERVIEW

1. Project Background and Description

Create a model to track vessel trips which based on a vessel positions dataset and includes:

- *the moment when the vessel enters a port*
- *the moment when the vessel leaves the port*
- *the vessel's next destination*

2. Vessel Positions Dataset

We have information of vessel trips in the format of csv. I modelized a trip like:

vessels (vessel, timestamps, port_id)

ports (port_id, latitude, longitude)

For normalize vessels table we can create also date(date_id,day, month,year,hour,minute,second) and vessels(vessel, date_id, port_id) but for this implementation I use the version above.

And a **trip** can be defined like

Vessel 5291 : departure_time – departure_port → arrival_time – arrival_port2

3. Implementation Requirements

I use Anaconda Spyder Python3;

with libraries: pandas, gmplot (for plot results on google maps), pyqt5(for simple application interface), psycopg2 and configparser(for postgressql connection)

Please clean kernel/console before running app.py

Please change database.ini credentials and paths (paths of excel data files). * I tried to generalized paths with os.path.abspath and real_path but I had encoding problems between windows and unix.**

I have different implementations for this project:

1- vessel_track.py and test1.py

vessel_track.py is my first implementation with pytyhon pandas library, I extract data from given excel data file and I insert the data in my pandas tables.

Methods:

createVessels: creates dataFrame vessels from csv file

createPorts: creates dataFrame ports from the attitude and longitude of vessels and add ids

getVessels: returns unique vessel ids

getTrips: if only vessel_id is given returns all trips of vessel

if vessel_id and a period(start-end) is given returns trips of that vessel in given period

printTrips: for printing getTrips

getAllTripOfVessel: returns an ordered dataFrame for the trips of given vessel

printAllTripOfVessel: for printing getAllTripOfVessel like port1 -> port2 -> port5 …

getPortListForTrip : returns an ordered port list for the trips of given vessel

getLatLong: get list of (latitude,longitude) for given ports

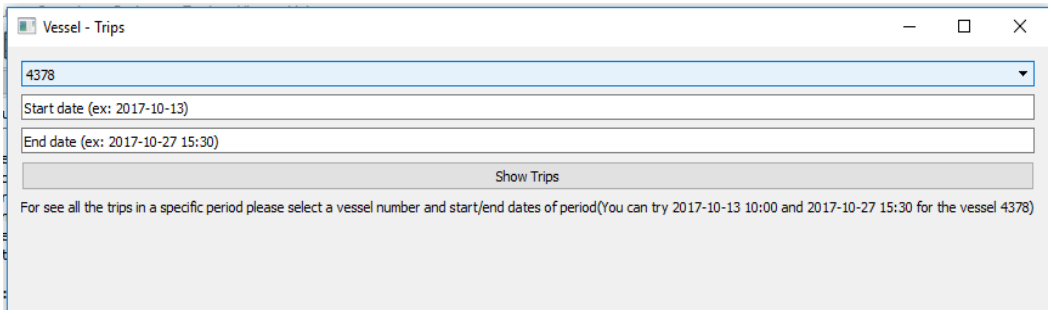
plotMap: plot the resuls on google maps (need to enter google maps apikey)

2- app.py

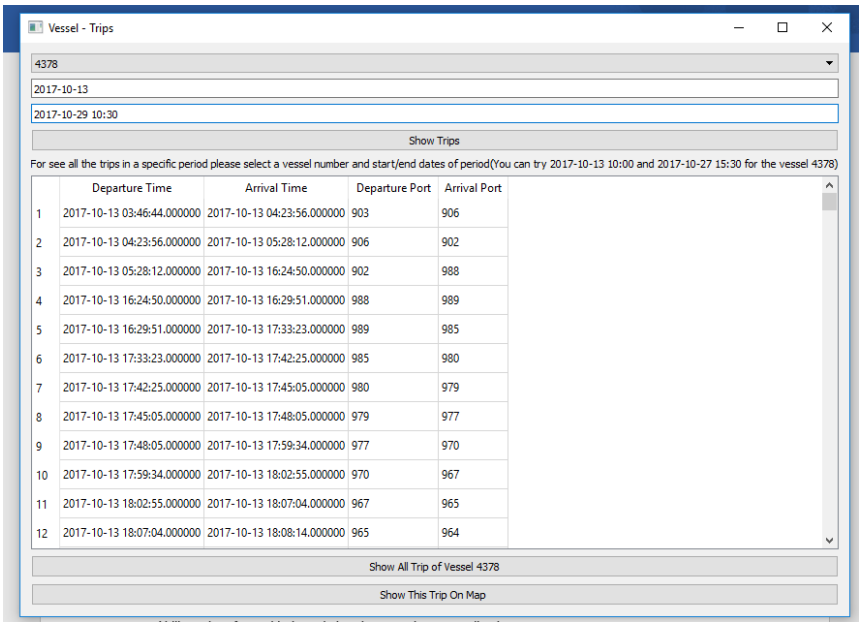
If app doesn't start please check pyqt5 library and clear kernel and try again!

A simple application to simplify the usage of the methods in vessel_track.py

For I used PYQT5 library in python and I followed the tutorial in http://zetcode.com/gui/pyqt5/

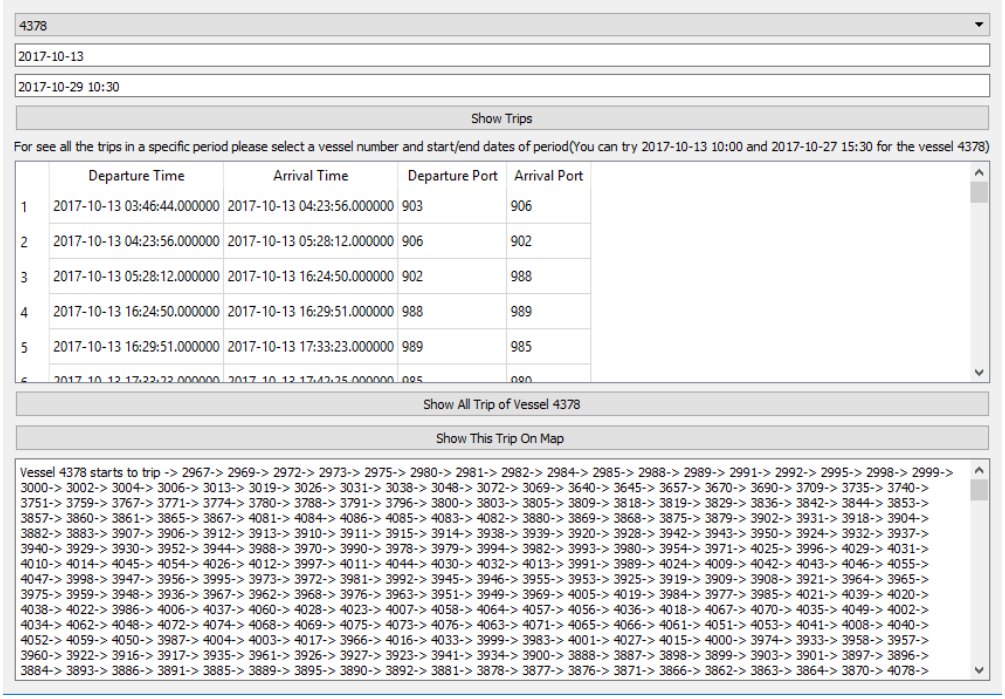


- You need to choose a vessel and enter a start and end date/datetime (2017-10-13 or 2017-10-13 13:30) for see the trips. And then click the button Show Trips.

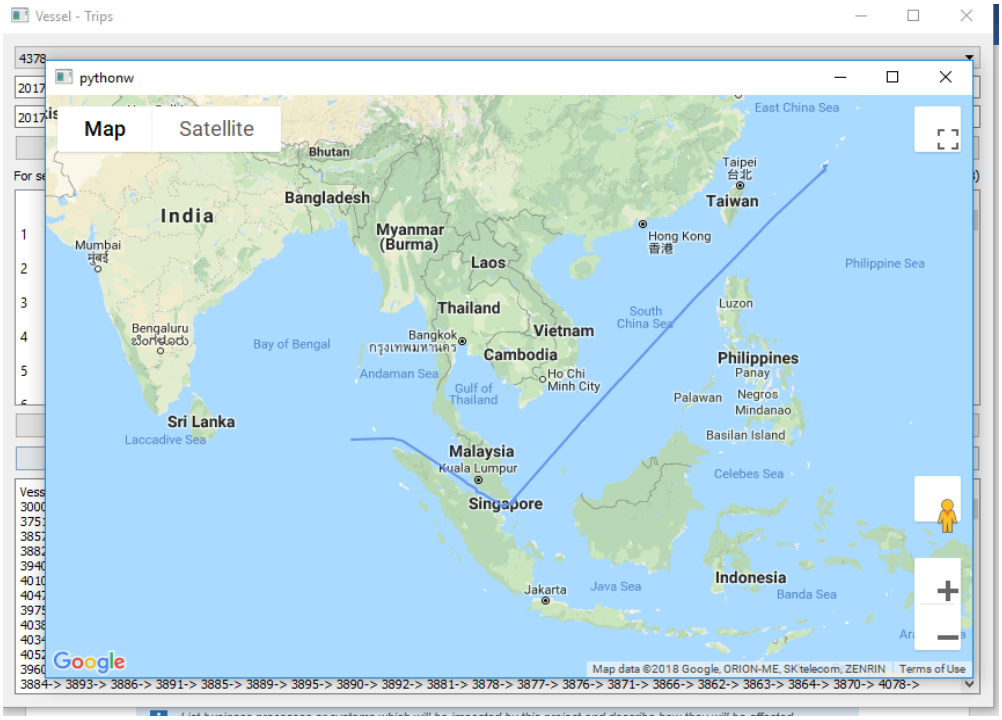


	Departure Time	Arrival Time	Departure Port	Arrival Port
1	2017-10-13 03:46:44.000000	2017-10-13 04:23:56.000000	903	906
2	2017-10-13 04:23:56.000000	2017-10-13 05:28:12.000000	906	902
3	2017-10-13 05:28:12.000000	2017-10-13 16:24:50.000000	902	988
4	2017-10-13 16:24:50.000000	2017-10-13 16:29:51.000000	988	989
5	2017-10-13 16:29:51.000000	2017-10-13 17:33:23.000000	989	985
6	2017-10-13 17:33:23.000000	2017-10-13 17:42:25.000000	985	980
7	2017-10-13 17:42:25.000000	2017-10-13 17:45:05.000000	980	979
8	2017-10-13 17:45:05.000000	2017-10-13 17:48:05.000000	979	977
9	2017-10-13 17:48:05.000000	2017-10-13 17:59:34.000000	977	970
10	2017-10-13 17:59:34.000000	2017-10-13 18:02:55.000000	970	967
11	2017-10-13 18:02:55.000000	2017-10-13 18:07:04.000000	967	965
12	2017-10-13 18:07:04.000000	2017-10-13 18:08:14.000000	965	964

- You can see ordered list of ports which have been visited by the selected vessel since the beginning by clicking the button Show All Trips of Vessel 4378:



- You can also the trip of selected vessel in selected period on the map by clicking the Show This Trip On Map button:



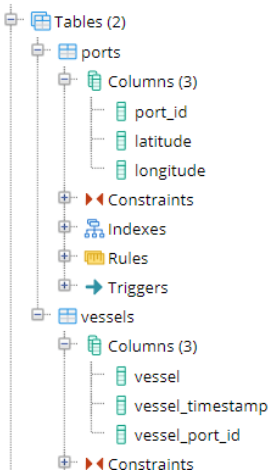
4. PostgreSQL

App_postgres.py

I use my local PostgreSQL server for create database and tables, insert excel data to the tables and create/call functions. I followed http://www.postgresqltutorial.com tutorial.

In database.ini file I have databasename:vessels_kpler, username and password. (***please change these parts before run the code***)

I created 2 tables vessels and ports, then I created two excel files from the given data for the vessels(vessel,vessel_timestap,vessel_port_id) and ports(port_id,latitude,longitude). Then I use these two excel files to insert data in my db tables. The scripts are written in python code. (app_postgres.py)



After that I create some functions for select and format data from the table