

KPLER DATA ENGINEER INTERNSHIP TEST

Yuksel Selen ADALI 13.12.2018

OVERVIEW

1. Project Background and Description

i 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 modeled 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 postgresql 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 python 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 results on google maps (need to enter google maps apikey)

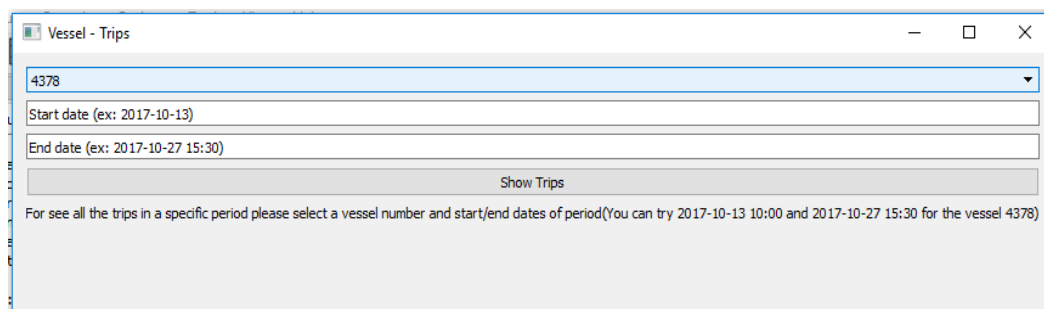
2- [app.py](#)

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.

The screenshot shows a web application window titled "Vessel - Trips". It has a dropdown menu for vessel selection (currently showing 4378), a text input for a date (2017-10-13), and another text input for a time range (2017-10-29 10:30). Below these is a "Show Trips" button. A message states: "For see all the trips in a specific period please select a vessel number and start/end dates of period(You can try 2017-10-13 10:00 and 2017-10-27 15:30 for the vessel 4378)". A table displays the following data:

	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

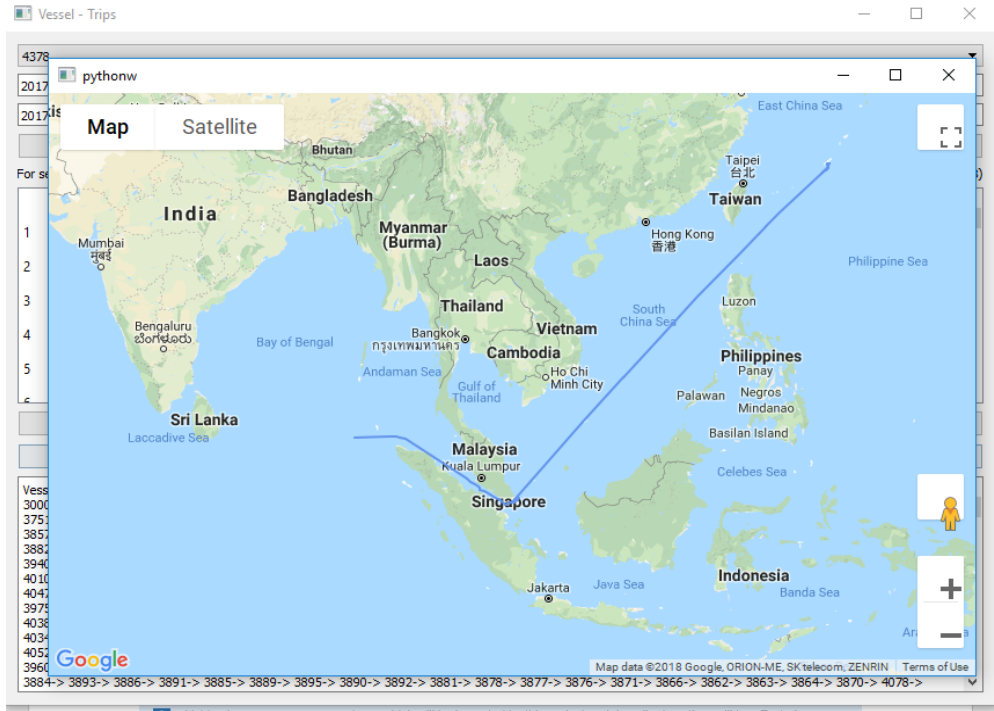
At the bottom, there are two buttons: "Show All Trip of Vessel 4378" and "Show This Trip On Map".

- 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:

This screenshot shows the same application window, but with the "Show All Trip of Vessel 4378" button clicked. The table below the message displays a long, ordered list of ports visited by the vessel, starting from 2967 and ending at 4078. The list is as follows:

Vessel 4378 starts to trip -> 2967-> 2969-> 2972-> 2973-> 2975-> 2980-> 2981-> 2982-> 2984-> 2985-> 2988-> 2989-> 2991-> 2992-> 2995-> 2998-> 2999-> 3000-> 3002-> 3004-> 3006-> 3013-> 3019-> 3026-> 3031-> 3038-> 3048-> 3072-> 3069-> 3640-> 3645-> 3657-> 3670-> 3690-> 3709-> 3735-> 3740-> 3751-> 3759-> 3767-> 3771-> 3774-> 3780-> 3788-> 3791-> 3796-> 3800-> 3803-> 3805-> 3809-> 3818-> 3819-> 3829-> 3836-> 3842-> 3844-> 3853-> 3857-> 3860-> 3861-> 3865-> 3867-> 4081-> 4084-> 4086-> 4085-> 4083-> 4082-> 3880-> 3869-> 3868-> 3875-> 3879-> 3902-> 3931-> 3918-> 3904-> 3882-> 3883-> 3907-> 3906-> 3912-> 3913-> 3910-> 3911-> 3915-> 3914-> 3938-> 3939-> 3920-> 3928-> 3942-> 3943-> 3950-> 3924-> 3932-> 3937-> 3940-> 3929-> 3930-> 3952-> 3944-> 3988-> 3970-> 3990-> 3978-> 3979-> 3994-> 3982-> 3993-> 3980-> 3954-> 3971-> 4025-> 3996-> 4029-> 4031-> 4010-> 4014-> 4045-> 4054-> 4026-> 4012-> 3997-> 4011-> 4044-> 4030-> 4032-> 4013-> 3991-> 3989-> 4024-> 4009-> 4042-> 4043-> 4046-> 4055-> 4047-> 3998-> 3947-> 3956-> 3995-> 3973-> 3972-> 3981-> 3992-> 3945-> 3946-> 3955-> 3953-> 3925-> 3919-> 3909-> 3908-> 3921-> 3964-> 3965-> 3975-> 3959-> 3948-> 3936-> 3967-> 3962-> 3968-> 3976-> 3963-> 3951-> 3949-> 3969-> 4005-> 4019-> 3984-> 3977-> 3985-> 4021-> 4039-> 4020-> 4038-> 4022-> 3986-> 4006-> 4037-> 4060-> 4028-> 4023-> 4007-> 4058-> 4064-> 4057-> 4056-> 4036-> 4018-> 4067-> 4070-> 4035-> 4049-> 4002-> 4034-> 4062-> 4048-> 4072-> 4074-> 4068-> 4069-> 4075-> 4073-> 4076-> 4063-> 4071-> 4065-> 4066-> 4061-> 4051-> 4053-> 4041-> 4008-> 4040-> 4052-> 4059-> 4050-> 3967-> 4004-> 4003-> 4017-> 3966-> 4016-> 4033-> 3999-> 3983-> 4001-> 4027-> 4015-> 4000-> 3974-> 3933-> 3958-> 3957-> 3960-> 3922-> 3916-> 3917-> 3935-> 3961-> 3926-> 3927-> 3923-> 3941-> 3934-> 3900-> 3888-> 3887-> 3898-> 3899-> 3903-> 3901-> 3897-> 3896-> 3884-> 3893-> 3886-> 3891-> 3885-> 3889-> 3895-> 3890-> 3892-> 3881-> 3878-> 3877-> 3876-> 3871-> 3866-> 3862-> 3863-> 3864-> 3870-> 4078->

- 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

Functions (5)

- get_port_by_id(value_id integer)
- get_port_by_period(start_time timestamp without time zone, end_time timestamp without time zone)
- get_port_by_period_vessel_id(value_id integer, start_time timestamp without time zone, end_time timestamp without time zone)
- get_trip_by_vessel_period(value_id integer, start_time timestamp without time zone, end_time timestamp without time zone)
- get_vessel_by_id(value_id integer)

Tables (2)

- ports**
 - Columns (3)
 - port_id
 - latitude
 - longitude
 - Constraints
 - Indexes
 - Rules
 - Triggers
- vessels**
 - Columns (3)
 - vessel
 - vessel_timestamp
 - vessel_port_id
 - Constraints