PROMENADE DES ANGLAIS PRELIMINARY STEPS:

The data at disposition are collected from 9 detectors positioned on the Promenade in both directions. 3 detectors collect the traffic in the north direction and 6 detectors collect the traffic in the south direction (facing the sea). Two period of time are registered from 01/06/2019 to 31/12/2019 and from 06/02/2020 to 30/12/2020. Two variables are reported for each lane of the road: the flow (n° veh/h) and the occupancy rate of the detector in percentage. The data are recorded every minute. Rough data needs to be treated and aggregated. Here a list of operation computed in order to study the dynamics in both intervals time 2019 and 2020:

1. Duplicate removal.
2. Mark as missing observations the occupancy rates that are greater than 50%.
3. Mark as missing observations the flow that greater than 60 veh/h and have an occupancy rate equal to 100%.
4. For every detector all lanes are aggregated together (any lane containing NaN, cause the resulting output column flow or occupancy to be NaN as well)
5. By looking at the fundamental diagram outlier observations are removed.
6. Fill jumps in the series due to the omission of data with Nans values for both flow and occupancy rate.
7. 6-minute and 12-minute averages aggregation to mitigate the impact of traffic lights.
8. Compute the percentage of missing values in the series in order to decide which detectors can be used in the analysis.
9. Fill Nans values with an imputation procedure.

Imputation:

To apply the clustering procedure the NaNs values must be filled. The imputation procedure takes in account the temporal order of observations, thus if there are too many consecutive NaNs to fill the procedure is not applied.

The procedure reads as follow:

1. If an isolated Nan value is recognized (both previous and next observations are present), the observation before the Nan value is propagate forward to fill the missing value.
2. If the number of consecutive Nan values is lower than 20 (2 hours). The consecutive Nans are filled with a linear method based on the temporal alignment of the previous observations available. Otherwise if the number of consecutive Nans exceed 2 hours the imputation is not performed.

In order to decide which detectors using in the analysis. I report the percentage of NaN values and the subperiods where disorders are registered for each detector after having applied the steps above.

2019: 1430 Observations are not registered from the initial dataset (mostly jumps of 1 minute in the initial series).

* C602: 4.32% of missing values for both flow and occupancy rate. No data from 19/8 at 6 to 8:54. No data from 29/8 at 21:12 to 30/8 14:48. No data from 3/11 at 9 to 11:36. No data from 23/12 at 17:42 to 31/12.
* C601: threshold values 100%, 6000 veh/h.
* C077: 56.72% of missing values. No data registered from 1/6 to 30/9 at 9:06.
* C009 ﻿: threshold values 100%, 4000 veh/h.
* C094 : threshold values 100%, 4000 veh/h.
* C614: 7.05% of missing values. No data available from 4/6 to 6/6 and from 21/8 to 3/9.
* C615: 46.81% of missing values for the occupancy rate. No data registered from 1/6 to 30/9 at 9:06. 7.04 % of missing values for the flow. No data registered from 4/6 at 17:00 to 6/6 at 15:00 and from 21/8 at 13:12 to 3/9 at 16:18.
* C599: No data available from 12/7 to 15/7 and from 11/10 to 30/10.
* C598: 64.39% of missing values for both flow and occupancy.

Three detectors are been used for the analysis. Despite after the pre-process these detectors do not contain missing values, for some days aberrant measurements are registered, here the list of the days:

C009: 30/6, 13/7, 14/7, 15/8, 7/9, 8/9, 3/11, 23/11, 24/11, 20/12.

C601: 30/6, 7/9, 8/9, 3/11, 24/10, 1/11, 22/12, 25/12.

C094: 10/30, 30/6, 13/7, 14/7, 6/9, 7/9, 31/10, 1/11.

2020:

All detectors have problem in the measurements from 11/11 to 24/11.

2401 observations are missing in the range 06/02- 31/12. There are jumps in the series, mostly of 1 minute, 1 hour and 5 minutes. Here the days that present jumps:

1. February 2nd, 24th, 25th, 26th
2. March 5th, 6th, 7th, 8th, 17th, 18th, 19th, 20th, 21st, 23rd, 26th, 28th, 29th, 30th.
3. April 2nd, 5th, 21st.
4. May 2nd, 5th, 11th, 12th, 14th,15th, 16th,17th, 20th, 22nd, 25th, 26th.
5. June 2nd, 3rd, 19th, 22nd, 28th.
6. July 3rd, 5th, 10th, 31st.
7. August 5th,10th,12th,25th.

* C602: threshold value 100%, 4000 veh/h. 5.30% of missing values both for occupancy rate and flow. No data from 24/2 at 22:24 to 25/2 at 8:30. No data from 2/5 at 17:54 to 19:24. No data from 29/8 at 18:48 to 19:06. No data from 2/9 at 12:48 to 4/9 9:48. No data 12/9 from 13:06 to 14:42. No data from 20/9 at 13:54 to 21/9 at 14:36. No data from 11/10 at 10:24 to 12/10 at 9:24.
* C601: threshold value 100%, 6000 veh/h. 5.30% of missing values for the flow,

6.80% of missing for the occupancy rate. No data from 11/12 at 21:06 to 31/12. No data from 24/2 at 22:24 to 25/2 at 8:30. No data from 2/05 at 17:54 to 19:24. No data from 27/7 at 16:00 to 20:06. No data 08/04 at 12:00 to 21:06. No data from 29/8 18:48 to 19:06. No data from 3/9 at 15:00 to 16:36. No data from 12/9 at 13:06 to 14:42. No data for the month of December.

- C615: 30.15 % of missing values for the occupancy rate. No data registered for the occupancy from 6/4 to 21/6.

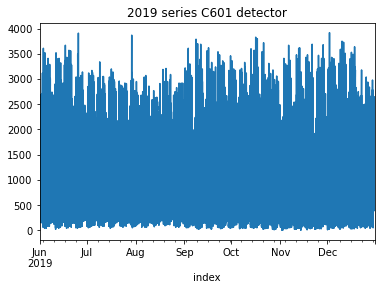
- C614: threshold values 100%, 4000 veh/h. 4.13% of missing values. No data from 24/2 at 22:24 to 25/2 at 8:30. No data from 2/5 at 17:54 to 19:24. No data from 3/9 at 15:00 to 16:36. No data from 12/9 at 13:06 to 14:42. No data from 12/12 at 4:36 to 5:54.

- C077 threshold 100%, 4000 veh/h. 4.735% of missing values. No data from 24/2 at 22:24 to 25/2 at 8:30. No data from 2/5 at 17:54 to 19:24. No data from 3/9 at 15:00 to 17:06. No data from 8/9 at 15:24 to 10/9 at 9:48. No data from 12/9 at 13:06 to 14:42. No data from 11/10 at 10:00 to 13:54. No data 12/12 from 4:36 to 5:54.

- C009 C094 threshold 100%, 4000 veh/h. 4.13% of missing values. No data from 24/2 at 22:24 to 25/2 at 8:30. No data from 2/5 at 17:54 to 19:24. No data from 3/9 at 15:00 to 16:36. No data from 3/9 at 15:00 to 16:36. No data from 12/9 at 13:06 to 14:42. No data 12/12 from 4:36 to 5:54.

- C599 threshold 100%, 4000 veh/h. 5.39 % of missing values. No data from 24/2 at 22:24 to 25/2 at 8:30. No data from 2/5 at 17:54 to 19:24. No data from 3/9 at 15:00 to 16:42. No data from 12/9 at 13:06 to 14:42. No data from 2/10 at 17:54 to 6/10 at 7:36. No data from 4/12 at 2:30 to 4:00. No data 12/12 from 4:36 to 5:54.

The C601 detector for the year 2019 are reported.



The C601 detector for the year 2020

