

STDA-project-proposal

Introduction:

The dataset comes from the open data provided by all the municipalities of Italy. This repository is available at dati.gov.it. From this repository I selected the data going from 2018 to 2019 about the number of connections of foreign cellular sims' to the repeaters of the Milan's city. The problem is very interesting from my point of view because can be found many and many interactions between the presence of people from other places of the world and for example (but not restricted to) fluctuations in the stock exchange prices. It's obvious that to find correlation between a certain fluctuation and a migration it should be very massive but we are lucky because the Italian stock exchange takes place in the Milan city so it's easier that a massive migration to that city provokes a fluctuation. Also are available many other relations such as the calculus of how many foreign people is attracted during the fashion's week. At the end of the day we can search for two different and interesting relations: the first is about how a migration can affect Italian's life and the reverse so how much the Italian behaviour can affect the life of people from other nations attracting them.

Data:

The data is composed by a date in which are present ~13 observations, the next columns have the total number of Italian sims, the total number of the foreign sims, the prefix of the foreign sims and the relative country which they belong to. The last column called num is the number of sims from the country written at its left.

	date <fctr>	total.ita.sim <dbl>	total.foreign.sim <int>	prefix <int>	country <fctr>	num <int>
1	2019-01-01	64	76	41	Svizzera	25
2	2019-01-01	64	76	7	Federazione Russa	17
3	2019-01-01	64	76	49	Germania	15
4	2019-01-01	64	76	33	Francia	8
5	2019-01-01	64	76	1	USA / Canada	3
6	2019-01-01	64	76	355	Albania	1
7	2019-01-01	64	76	44	Gran Bretagna	1
8	2019-01-01	64	76	91	India	1
9	2019-01-01	64	76	30	Grecia	1
10	2019-01-01	64	76	31	Paesi Bassi (Olanda)	1

1-10 of 30 rows

Previous 1 2 3 Next

Statistical Methods:

I would like to use the datasets about 2018 and 2019 to search for a pattern in the timeserie, this means to search for outliers but I would also like to observe the trend. Surely we will also be able to recognise a certain seasonality between fall/winter/spring/summer and this fact will be also confirmed by using the data from 2 different years instead of only one year. In particular probably I'm going to use the decomposition of the time serie through a moving average to search for seasonality and trends.