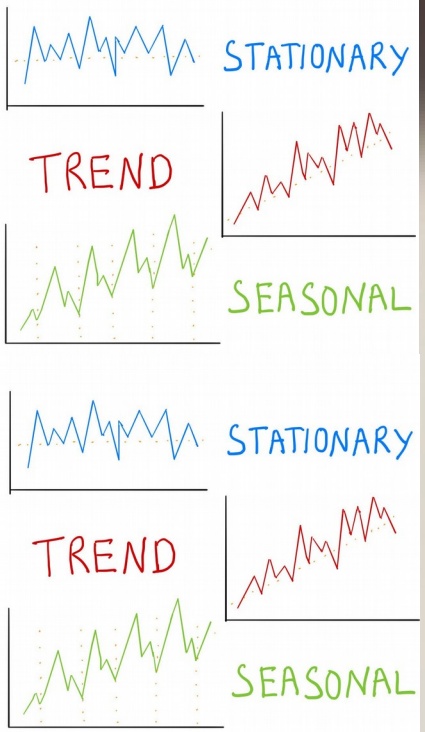
Hungarian Chickenpox Cases Forecasting



Presented by
Artem Ramus

Background

Abstract: A spatio-temporal data set of weekly chickenpox cases from Hungary. The data set consists of a county-level adjacency matrix and time series of the county-level reported cases between 2005 and 2015.

Source:
Benedek Rozemberczki, The University of Edinburgh, benedek.rozemberczki '@' gmail.com

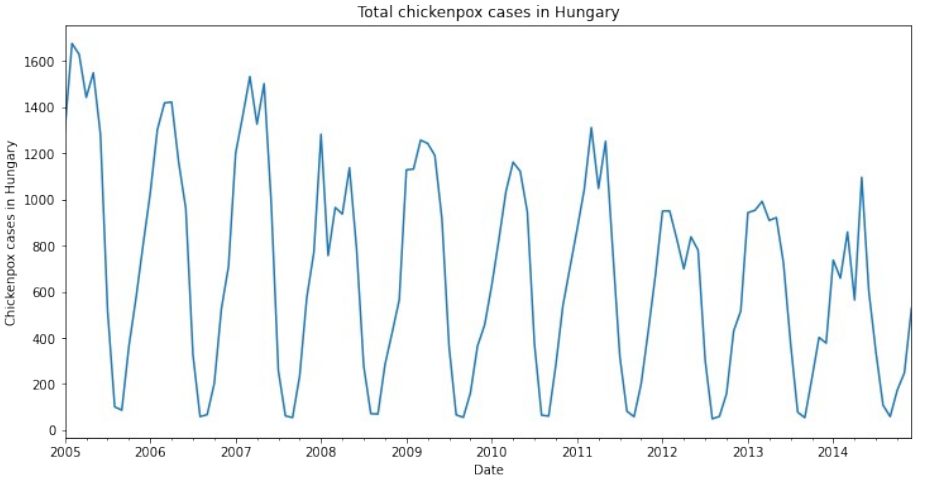
Link to hte dataset at UCI:
https://archive.ics.uci.edu/ml/datasets/Hungarian+Chickenpox+Cases

Introduction

The data set consists of a county-level adjacency matrix and time series of the county-level reported cases between 2005 and 2015. There are 2 specific related tasks: County level case count prediction and nation level case count prediction.

Attributes are weekly counts of chickenpox cases in Hungarian counties.

Data Set Characteristics: Time-Series
Number of Instances: 521
Area: Life
Attribute Characteristics: Real
Number of Attributes: 20



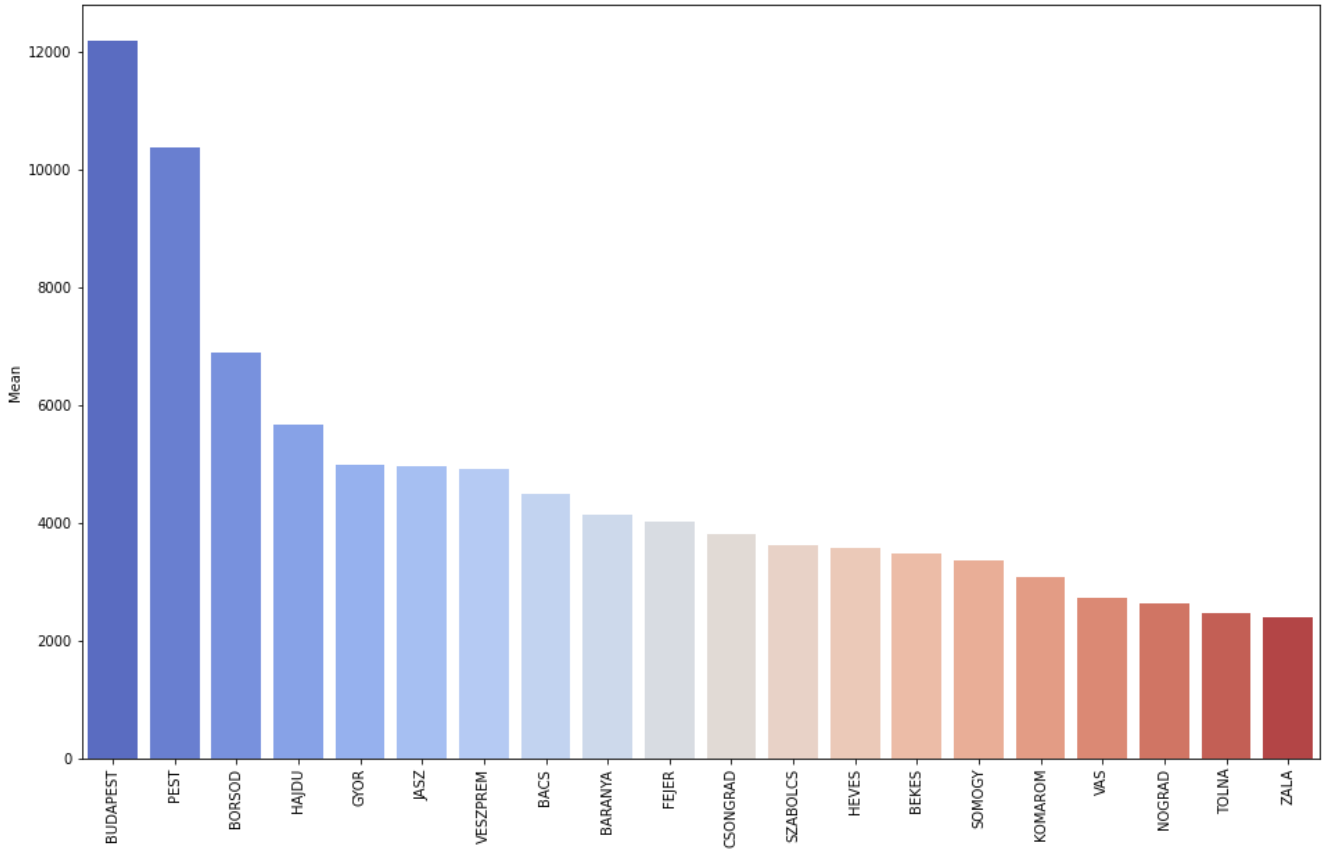
Exploratory Data Analysis

Distribution of total cases per areas in Hungary

Budapest and Pest areas cases count is prominently higher than others.

Tolna and Zala shows least cases.

Order of cases sum is consistent with mean for all the areas.

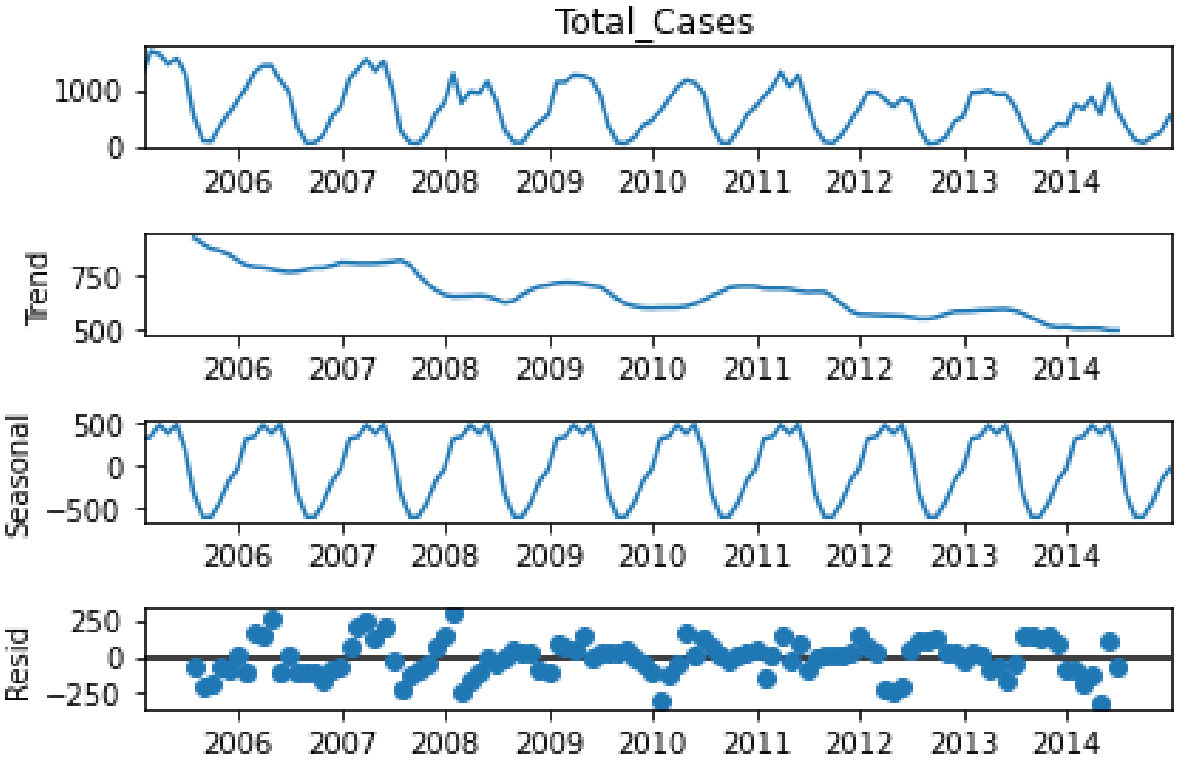


Decomposition

Yearly seasonality

Downward trend

Noise

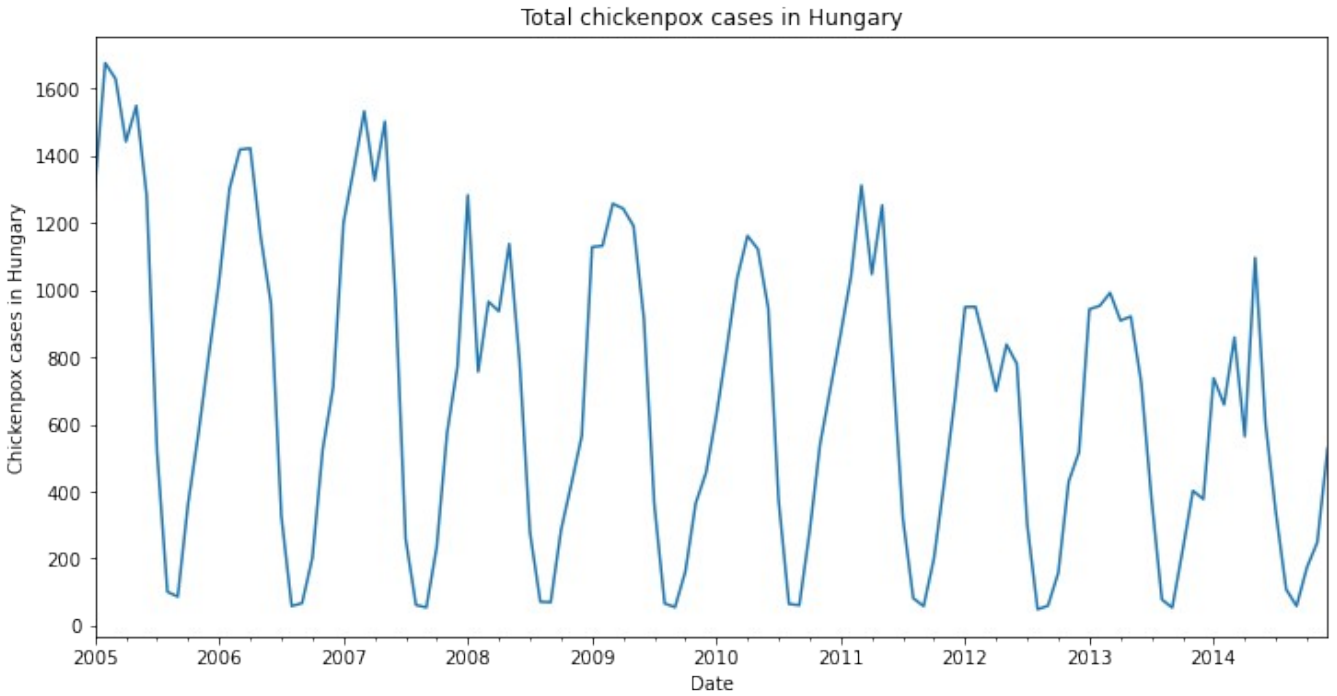


Methodology

Optimal sampling frequency - monthly

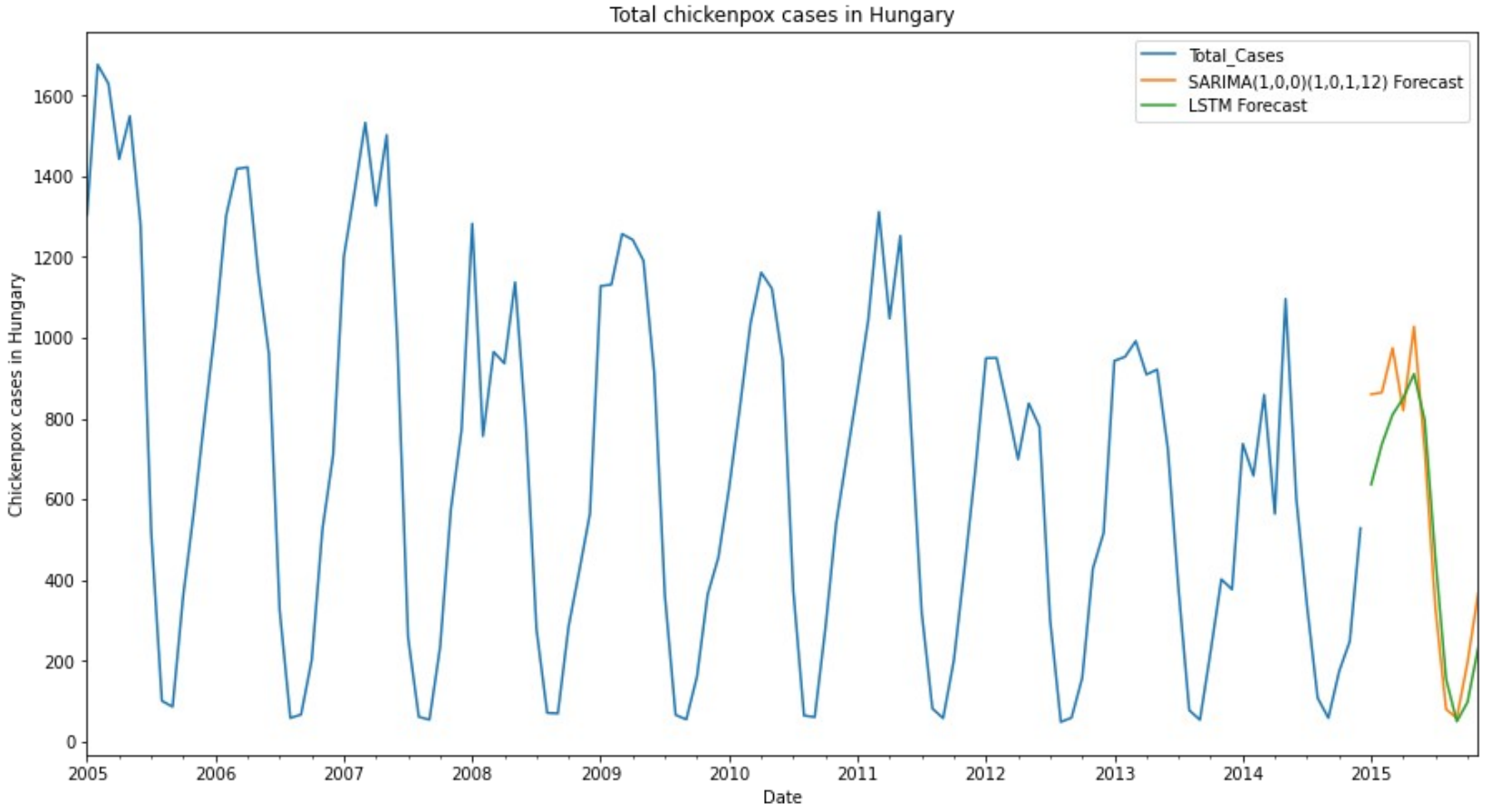
SARIMA and RNN LSTM models trained on 9 years train-set

1 year prediction



Performance of the Model

Predictions along with the original data



Summary and Conclusions

SARIMA and LSTM RNN demonstrated similar performance.

RMSE of SARIMA is 140, LSTM - 165.

The end

Thank you for your attention!