

Instructions: More information on expectations for the report can be found in the general directions for the data analysis sets. In this document, the specifics for the individual assignment will be discussed. Students are responsible for both the requirements in the general directions, and for the specific directions discussed below.

Topic 1: Time Series

Information on the nybirths dataset and loading it can be found here: <https://rpubs.com/ryankelly/ts6>

```
births <- scan("http://robjhyndman.com/tsdldata/data/nybirths.dat")

births <- ts(births, frequency = 12, start = c(1946, 1))
births
```

Create appropriate plots of the data. Discuss the kind of model that seems appropriate from your analysis using moving averages, acf/pacf graphs, differencing, seasonal properties etc. You may also apply smoothing functions to the data and discuss the results. Which method seems to produce the best forecast results? You may find it helpful to review our time series analysis examples from 325.

Topic 2: Validation

This topic is a bit more open-ended. You may choose from one of the datasets we used for either classification or clustering. Apply an appropriate classification/clustering algorithm to the data and test appropriate hyperparameters to select the best settings to obtain the best results. If you want, you may also use a regression set if you prefer. Describe the process.

Create appropriate graphs to illustrate the validation process. And then also include plots of your final model. If you are using a dataset you used before, you may include examples of those graphs to illustrate how the model performance has improved over the default settings.

You may include your code in an appendix or separate file, but the report of approximately 10 pages should focus on the analysis. It should look professionally formatted. Raw code and raw output is frowned upon.