Table of Contents

[Assignment 1 1](#_Toc115111045)

[Report Submission Instructions: 1](#_Toc115111046)

[1. Report: Problem Definition (5 marks) 1](#_Toc115111047)

[2. Report: Exploratory Data Analysis (10 marks) 1](#_Toc115111048)

[3. Model Development (20 marks) 1](#_Toc115111049)

[Report: Model Development 2](#_Toc115111050)

[4. Report: Model Evaluation (10 marks) 2](#_Toc115111051)

[5. Future Predictions (5 marks) 2](#_Toc115111052)

[6. Code Appendix 2](#_Toc115111053)

# Assignment 1

Perform a time series analysis to predict icecream sales.

## Report Submission Instructions:

* Please submit either a Jupyter notebook or a word document with your report and code in it (no pdf’s please).

In your report please include the following:

### Report: Problem Definition (5 marks)

Explain the problem you are trying to solve.

### Report: Exploratory Data Analysis (10 marks)

Please provide a thorough exploratory data analysis. Ensure the report is informative for a non-technical audience. Use visualizations. Identify trends, cycles or seasonal movements if they exist. Identify attribute time steps that appear to be correlated with the target variable.

### Model Development (20 marks)

Important: Build a stacked model which uses the outputs from at least three separate time series models. Each model in the time series model should use a significantly different algorithm or features than the other. The models should include at least an **OLS model with more than one feature and** an ARIMA model. You might also try other machine learning algorithms.

Hold back the 30 most recent observations for test and validation. Use 15 of these observations for testing and 15 of these observations for validation of the stacked model. See the week 4 notes from COMP4948 (other course) for how to use test\_train\_split() to do this.

### Report: Model Development

In the **report**, discuss how you prepared the data for modelling. List any variables that you created or treated and explain the modifications. Explain any back shifting and alternative treatments that you have done such as scaling or outlier handling. Explain which variables where manufactured, binned or imputed.

### Report: Model Evaluation (10 marks)

Present statistics and summaries in a professional reader-friendly manner for a non-technical audience to explain the differences in performance for each of the individual models and your stacked model. Compare the results of each model.

While this step is not required and it is likely too much for the scope of this project, you might consider how you could create a back test to simulate and measure your current model’s success over an extended historic period. If you do not implement a back test, in your report at least suggest how you might evaluate your model over a much longer term.

### Future Predictions (5 marks)

Using the most current observation, use your model to make a prediction for t+1 where the outcome is unknown. Highlight this prediction in your report.

### Code Appendix

If you submit your document using Word, please place your code in the appendix. Keep it neat and tidy. Use a mono-spaced font for code. Please remember do not use pdf.