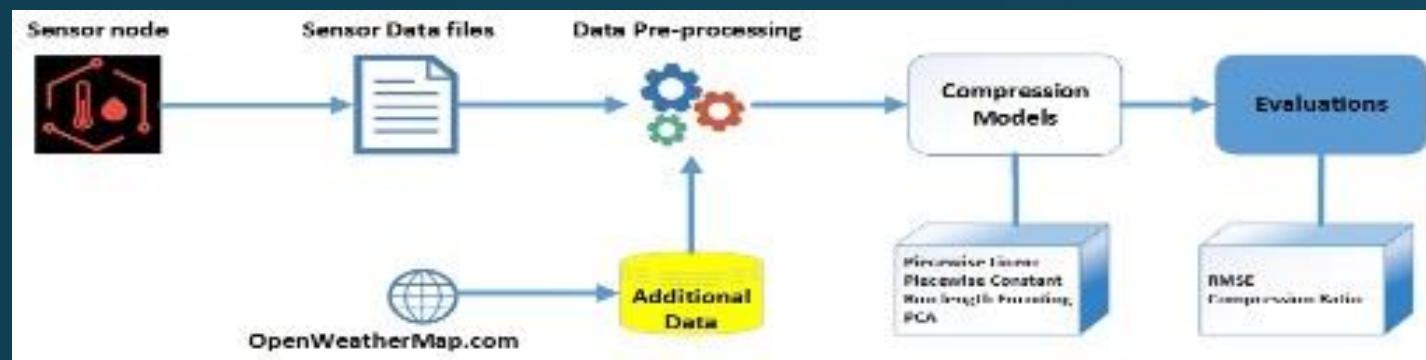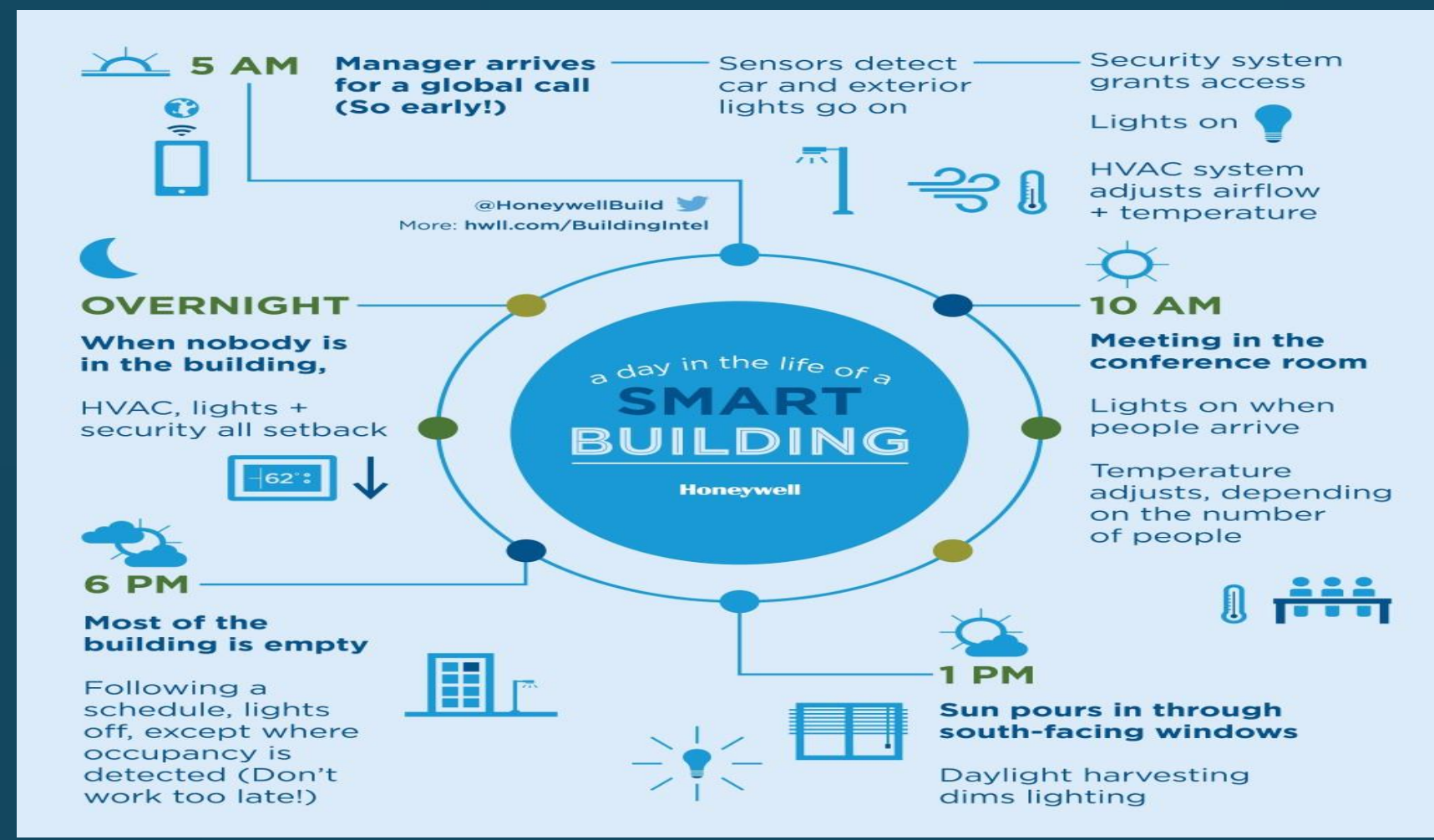# SMART BUILDING DATA ANALYSIS
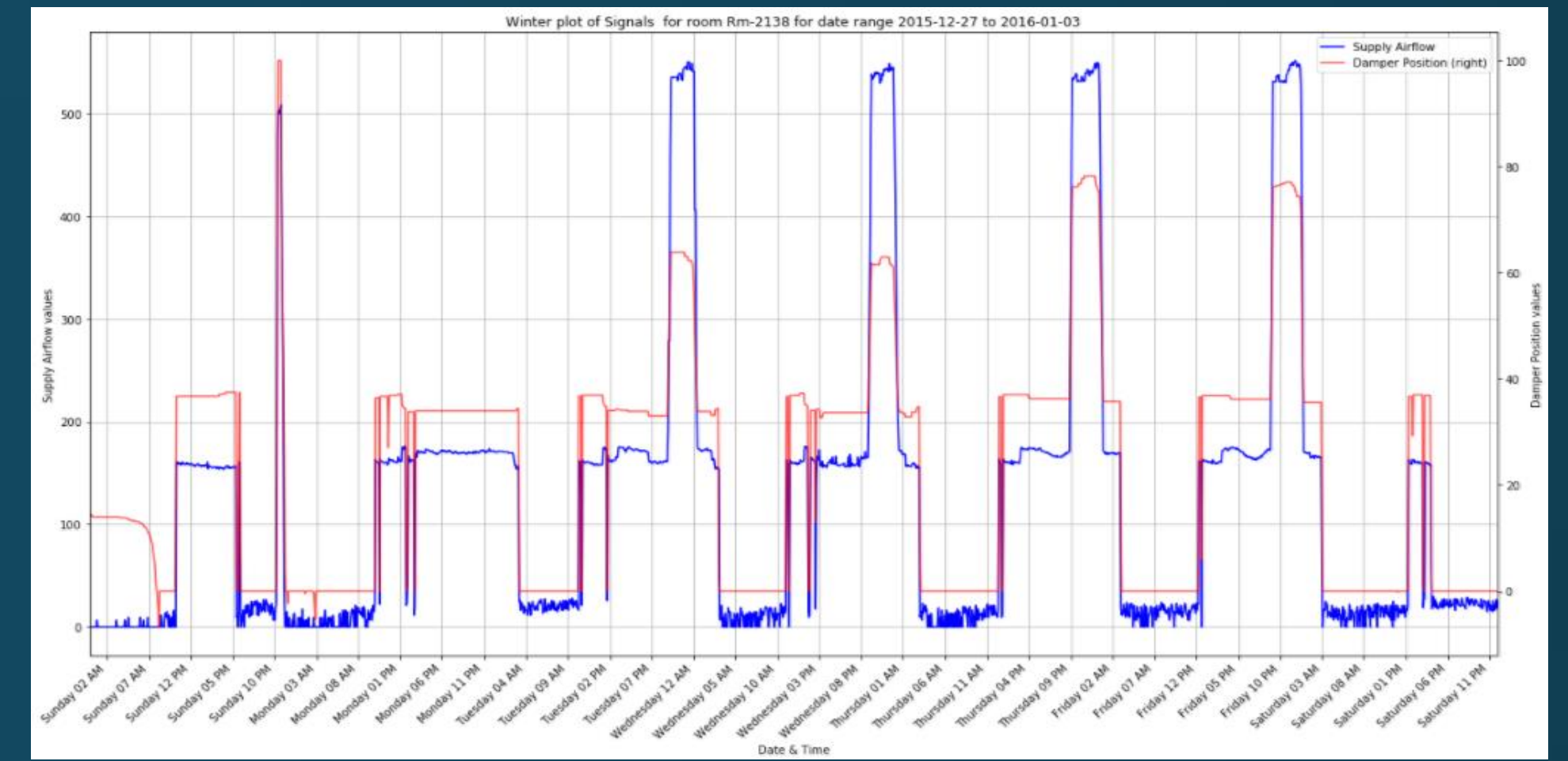
Architecture


Introduction

Data Exploration

Purpose

➢ Time Series Sensor Data is being collected at high frequency.
➢ Build a compression model for rooms applying different compression techniques.
➢ The compression of the raw sensor must be done in a way that makes accurate reconstruction of the signal possible.

Design

Compression Methods

Piecewise Constant : compression achieved by approximating a continuous signal to a several piecewise step functions

Piecewise Linear : function is a real-valued function defined on the real numbers or a segment thereof, whose graph is composed of straight-line sections.

PCA : Statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called **principal components**
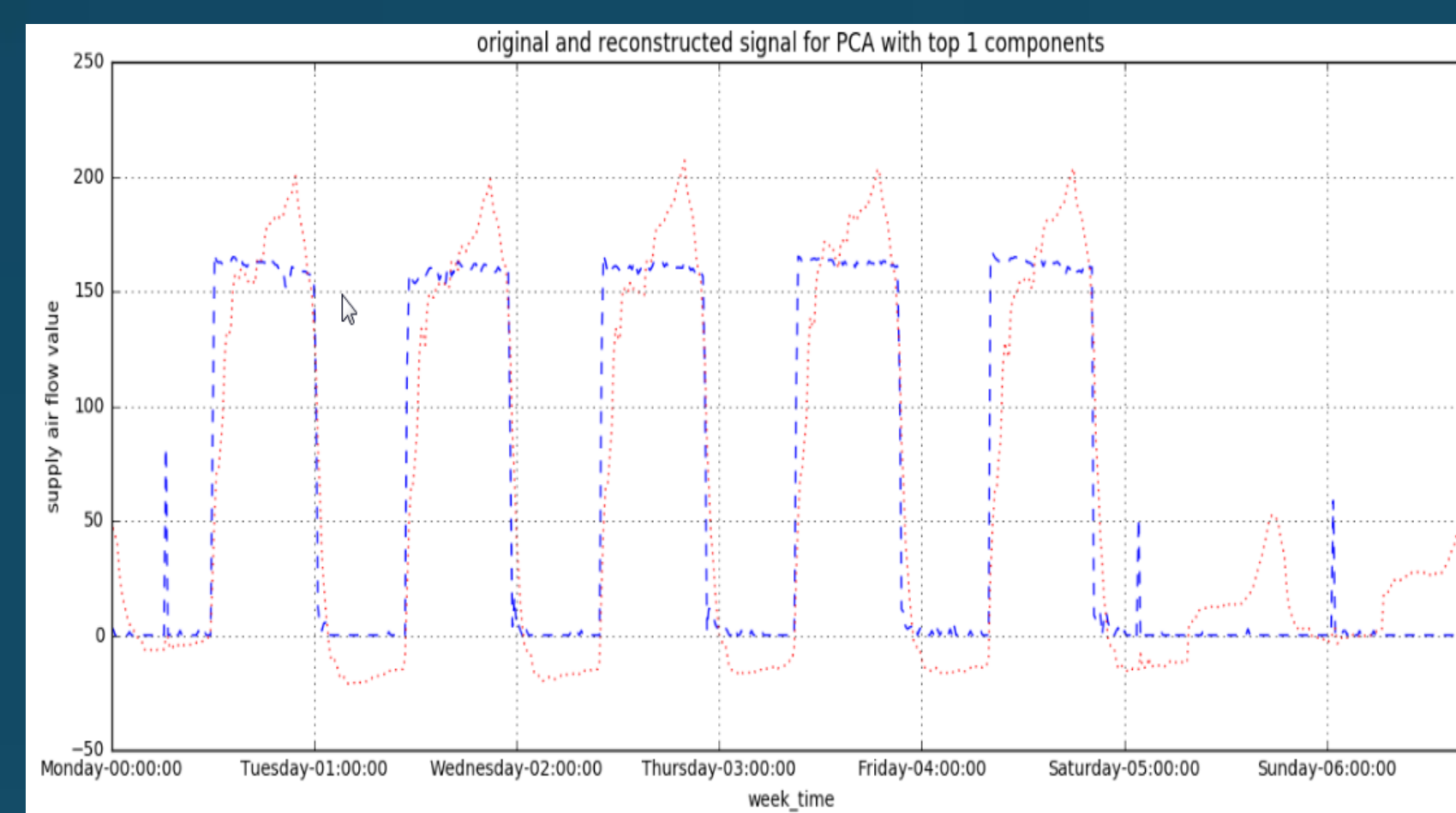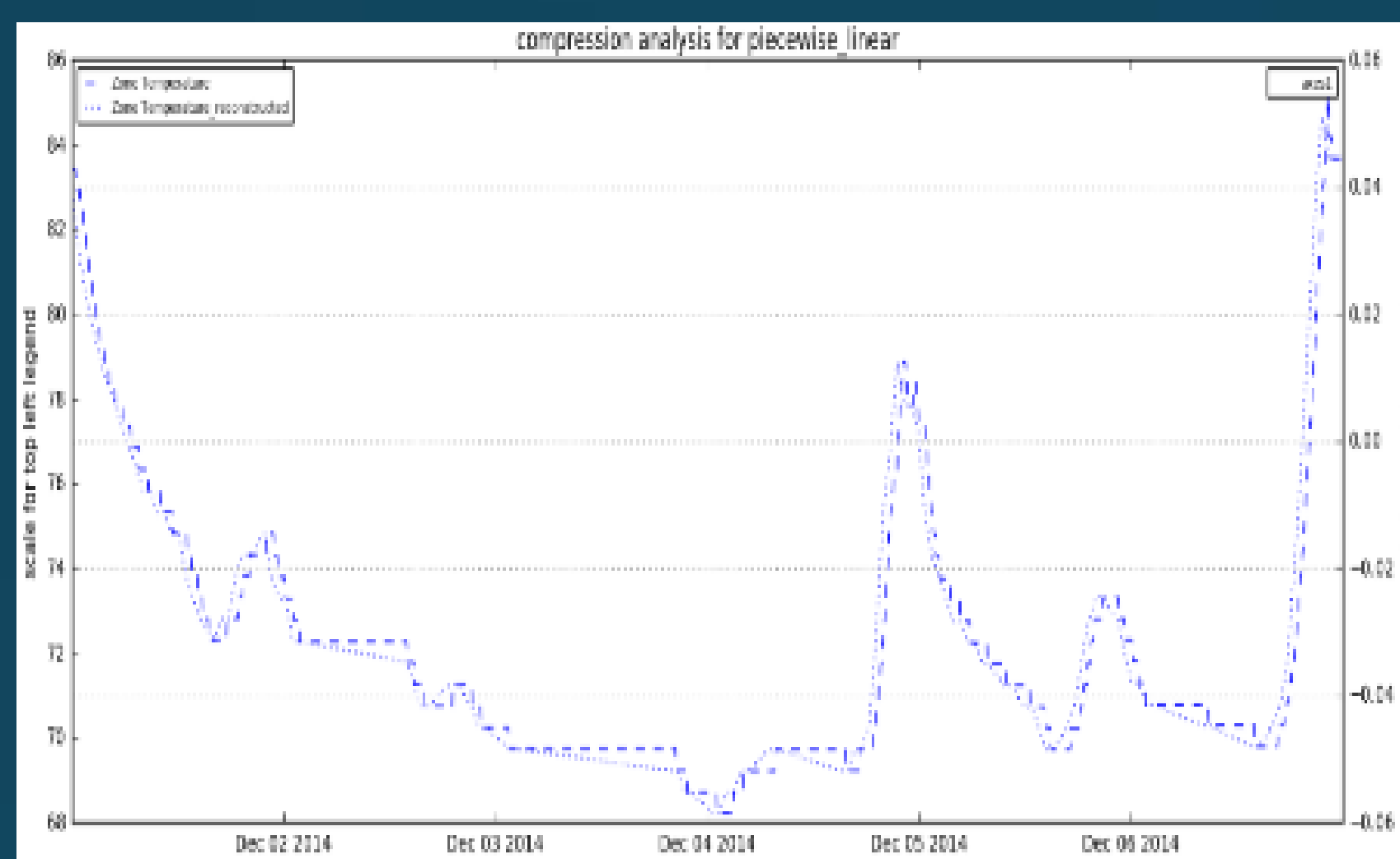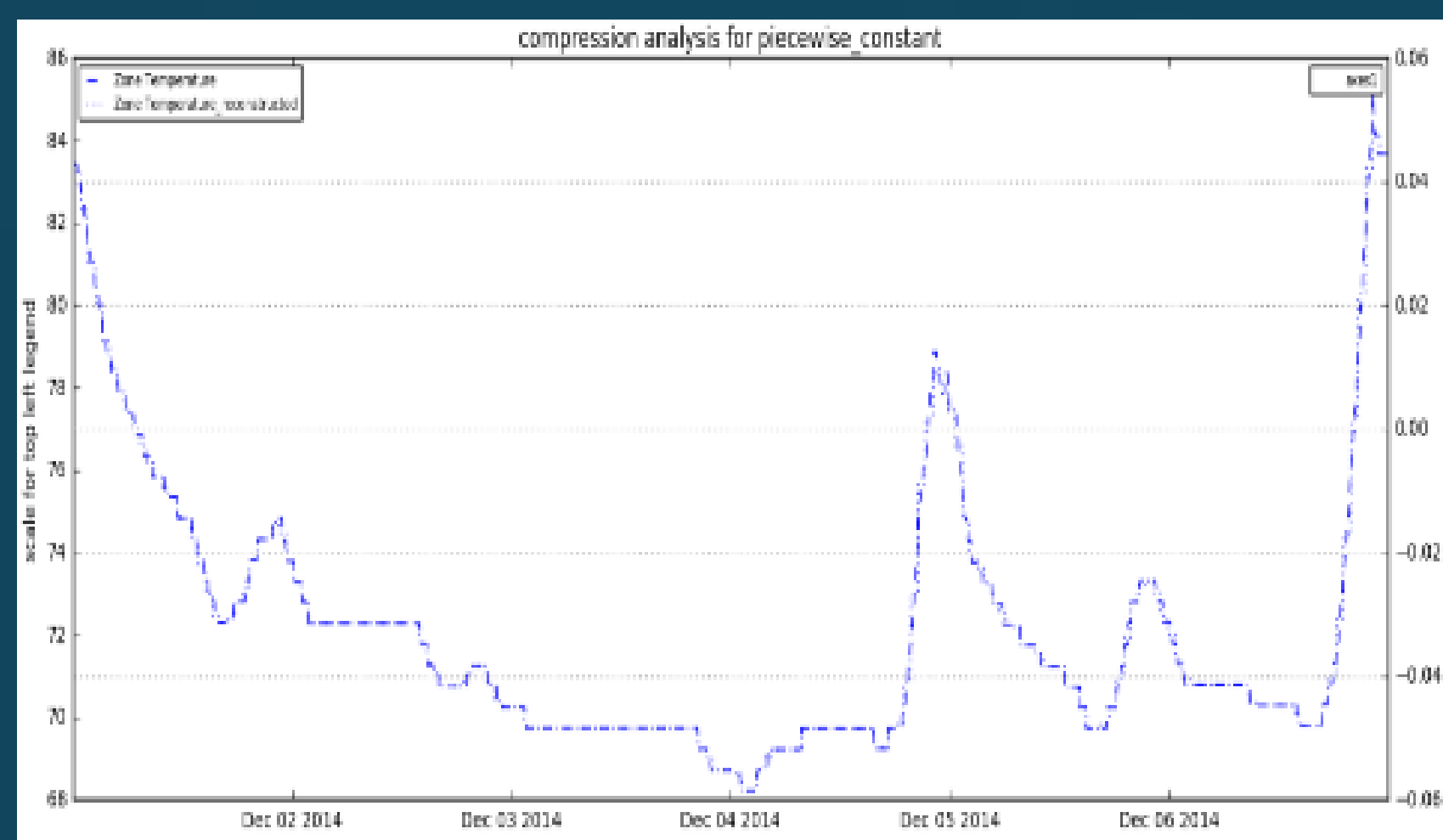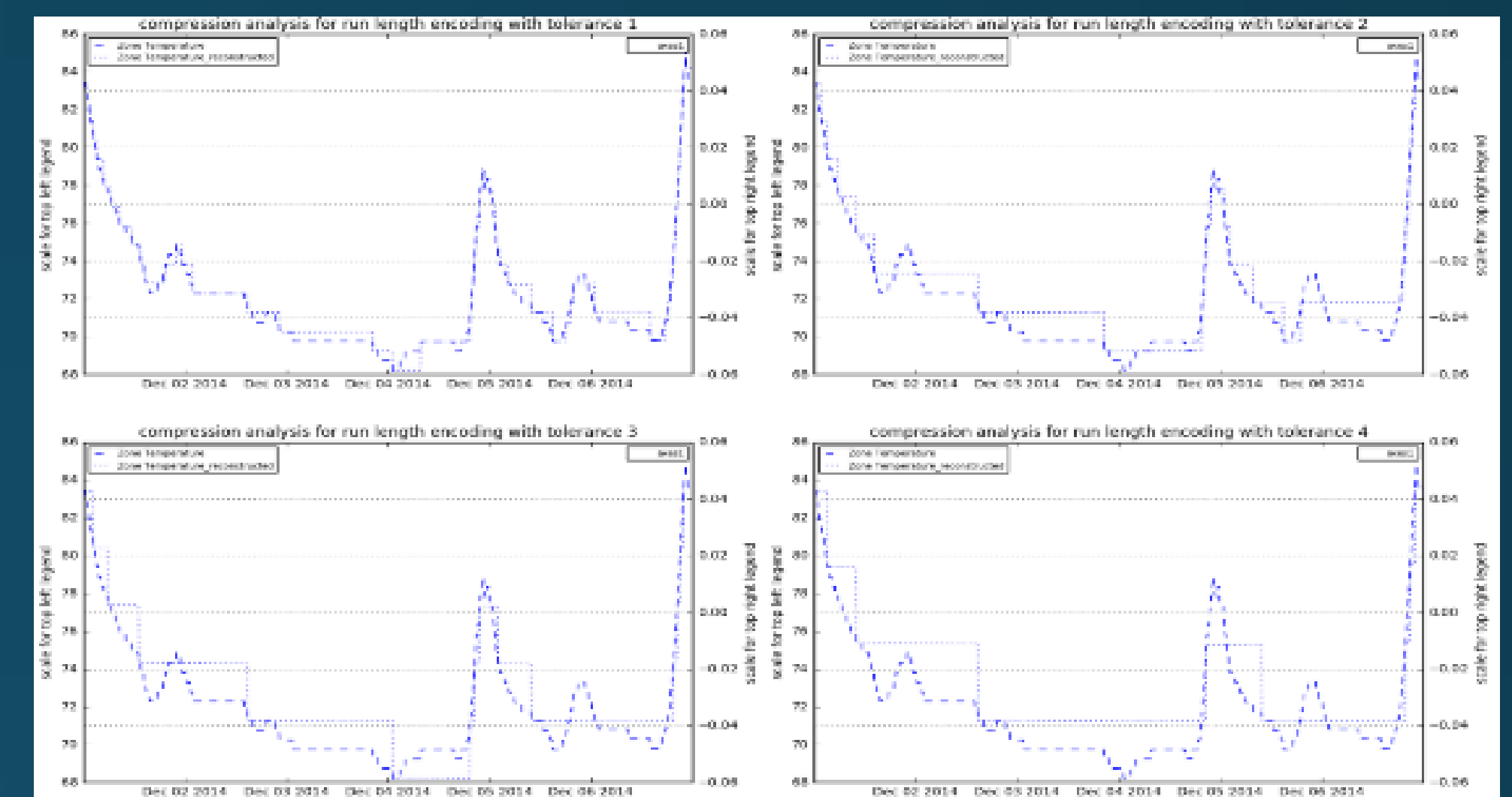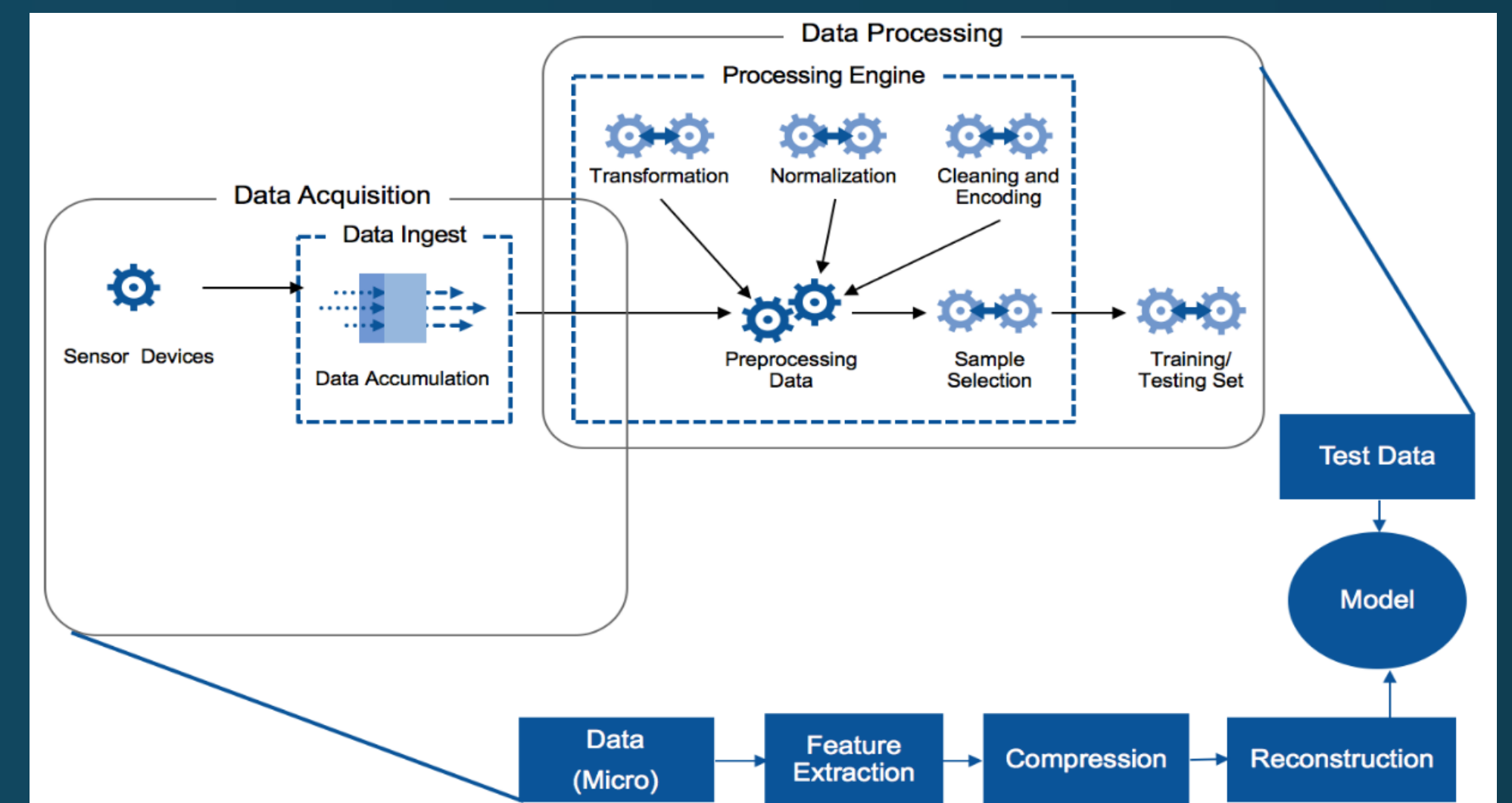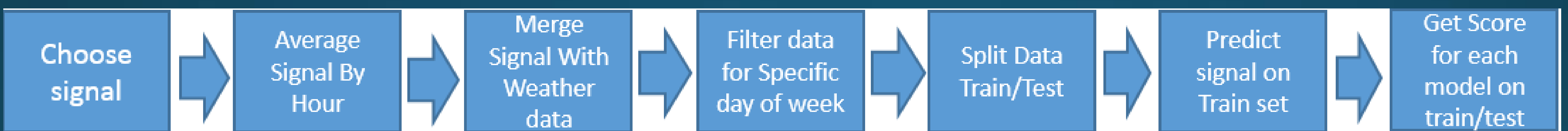

Run-length encoding (RLE) is a very simple form of underlined data compression in which runs of data are stored as a single data value and count, rather than as the original run.
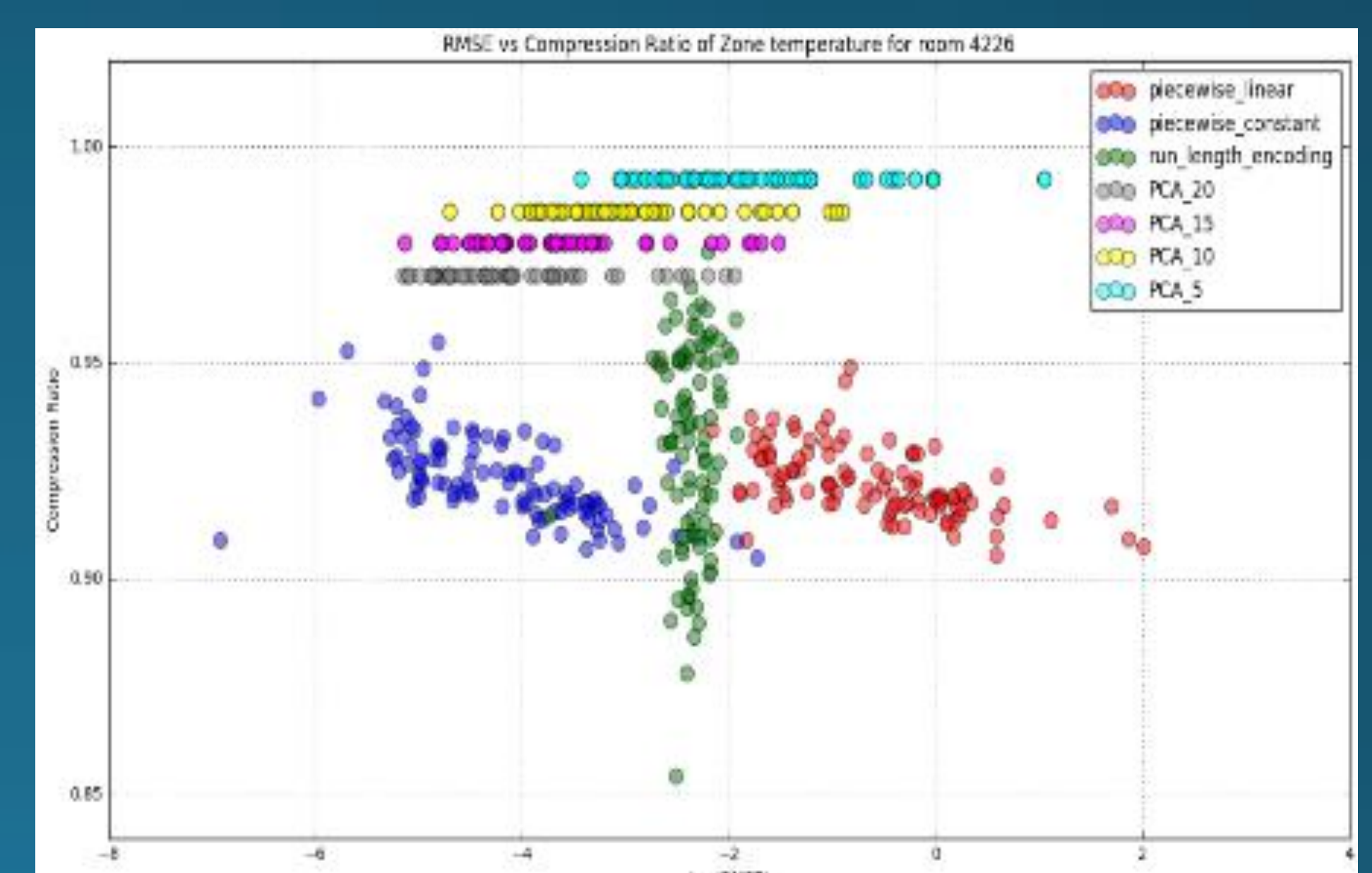
Modelling

Choose signal ⇨ Average Signal By Hour ⇨ Merge Signal With Weather data ⇨ Filter data for Specific day of week ⇨ Split Data Train/Test ⇨ Predict signal on Train set ⇨ Get Score for each model on train/test

Results

model	test score	train score
DecisionTreeRegressor	0.777367	0.783654
LinearRegression	0.803248	0.740135
Lasso	0.813416	0.738179
Ridge	0.803258	0.740146
AdaBoostRegressor	0.820211	0.781733

$$compression\_ratio = \frac{Original\ size - compressed\ size}{Original\ size}$$

➢ External Weather Features were vital features in modelling Zone temperature

➢ PCA – Highest compression Least RMSE
➢ Piecewise Linear & Constant compression is very similar for Zone Temperature

Project By : Rajesh Bondugula, Karunakaran T Nadadur & Manoj Jaiswal
Advisor : Dr Yoav Freund