Practical Machine Learning Exercise

Introduction: This is a short exercise to test your ability to construct and evaluate machine learning models. You should produce a machine learning model and an evaluation metric. The goal of this exercise is not to produce a state-of-the-art machine learning model. Which model you use, and how you evaluate, is up to you. The choice of model is not important (although we will assume that when you choose a model, you understand what it is and how it works). Your solution should be simple, but sensible: you should be able to explain why it tests something of impact to the problem.

Dataset description:

In the "training_sales.csv" you will find the hourly sales of a store in the US. The following is the description of the dataset:

Date: Date Time of the Sales
Value: Sales in cents

In the "training_traffic.csv" you will find the hourly traffic of the same store in the US. The following is the description of the dataset:

Date: Date Time of the Traffic
Value: Traffic data in person measured using store sensors

We are looking to identify the sales and traffic per hour for the following month. You can use any external data you wish (PS: you might want to look for holidays and national days).

The dataset has some missing points, this is deliberate to understand your assumptions about the missing data, and how you handle them.

Submission: Please submit the following:

1- A report discussing your methods, and the final results
2- All your code
3- Any external dataset used

Please feel free to use either R, Python, or Scala, and please feel free to use any out of the box functionality. We are mainly looking for your thinking process, and how you would approach the problem.