

---

# Overview of DFE Run

Automated

December 25, 2022

## Abstract

This is an automatically generated report. It is meant to illustrate the kinds of analysis that can be performed in LightR.

## 1 Overview

- The report was generated on Jan 1, 2023.
- The number of data sets was 42.
- Summary statistics of number of rows in data set:
  - Minimum 1712
  - Average 2398.380952381
  - Maximum 3252
- The number of formulas per model was 5. Explanation of the formulas is in Table 6.

## 2 Basic Feature Engineering

- The number of times this was invoked was 42
- The number of errors reported was 0.
- The average time taken (in  $\mu$ seconds) was 55772.380952381

Distribution of occurrence of error codes is shown in Table 1  
Explanation of error codes is in Table 7.1

---

| Error Code | Count |
|------------|-------|
| 0          | 42    |
| 1          | 0     |
| 2          | 0     |
| 3          | 0     |

Table 1: Error Counts for basic feature engineering

| Formula | Attempts | Successes |
|---------|----------|-----------|
| 0       | 42       | 42        |
| 1       | 42       | 13        |
| 2       | 42       | 13        |
| 3       | 42       | 13        |
| 4       | 42       | 13        |

Table 2: Error Counts for formula specific feature engineering

### 3 Formula Specific Feature Engineering

- The number of times this was invoked was 42.
- The number of data sets that had at least one error was 29.
- The average time taken (in  $\mu$ seconds) was 44638.30952381

Explanation of error codes is in Table 7.2

Distribution of occurrence of error codes (by formula) is shown in Table 2

Distribution of errors by formula and type is shown in Table 3

### 4 Skipped Formulas

We do not build a model for a formula if it depends on a formula for which model building failed. A report for this situation is in Table 4.

### 5 Model Building

- The number of models attempted was 94
- The number of models that were built was 94
- Summary statistics of time (in seconds) to build a model
  - Minimum 2.001647

---

| Formula | Error Code | Count |
|---------|------------|-------|
| 0       | 0          | 42    |
| 0       | 1          | 0     |
| 0       | 2          | 0     |
| 0       | 3          | 0     |
| 0       | 4          | 0     |
| 1       | 0          | 13    |
| 1       | 1          | 0     |
| 1       | 2          | 0     |
| 1       | 3          | 0     |
| 1       | 4          | 29    |
| 2       | 0          | 13    |
| 2       | 1          | 0     |
| 2       | 2          | 0     |
| 2       | 3          | 0     |
| 2       | 4          | 29    |
| 3       | 0          | 13    |
| 3       | 1          | 0     |
| 3       | 2          | 0     |
| 3       | 3          | 0     |
| 3       | 4          | 29    |
| 4       | 0          | 13    |
| 4       | 1          | 0     |
| 4       | 2          | 0     |
| 4       | 3          | 0     |
| 4       | 4          | 29    |

Table 3: Error Counts, broken down by formula and type

| Formula | Formula Attempted | Model Built |
|---------|-------------------|-------------|
| 0       | 42                | 42          |
| 1       | 13                | 13          |
| 2       | 13                | 13          |
| 3       | 13                | 13          |
| 4       | 13                | 13          |

Table 4: Skipped model building by formula

---

| Decile | Value  |
|--------|--------|
| 1      | 3.162  |
| 2      | 4.050  |
| 3      | 4.300  |
| 4      | 5.853  |
| 5      | 10.561 |
| 6      | 12.722 |
| 7      | 16.772 |
| 8      | 21.458 |
| 9      | 28.013 |

Table 5: Deciles for model build times (seconds)

| Index | Key | Explanation                               |
|-------|-----|---|
| 0     | f0  | Basic formulas                            |
| 1     | f1  | Uses 2 lag components (week 1 and week 2) |
| 2     | f2  | Uses 2 lag components (week 2 and week 3) |
| 3     | f3  | Uses 2 lag components (week 3 and week 4) |
| 4     | f4  | Uses 2 lag components (week 4 and week 5) |

Table 6: List of Formulas

- Average 15.330078095238
- Maximum 81.555114
- Distribution of build times is in Table 5

## 6 Explanations of Terms

### 6.1 Formulas

## 7 Error Codes

### 7.1 Basic Feature Engineering

See Table 7.1

### 7.2 Formula Specific Feature Engineering

See Table 7.2

---

| Error Code | Explanation  |
|------------|--|
| 0          | No Error   |
| 1          | insufficient rows in input data frame                |
| 2          | insufficient rows in input data frame after cleaning |
| 3          | insufficient unique values in <b>toy</b> component   |

Table 7: Explanation of Error Codes

| Error Code | Explanation                             |
|------------|---|
| 0          | No Error                                |
| 1          | not enough rows after cleaning          |
| 2          | Range of <b>toy</b> component too small |
| 3          | too few uniques in <b>toy</b> component |
| 4          | too few uniques in lag component        |

Table 8: Explanation of Error Codes