

Contents

1	Step 4 Machine learning	1
1.1	Step 0: Look at and Modify the dataset	1
1.2	Step 1: Explore the dataset	4
1.3	Step 2: Split sets, train a Machine Learning Model and Evaluate performance	5
1.4	Step 3: Plot results	8
1.5	Step 4: Improve models by changing the dataset	10

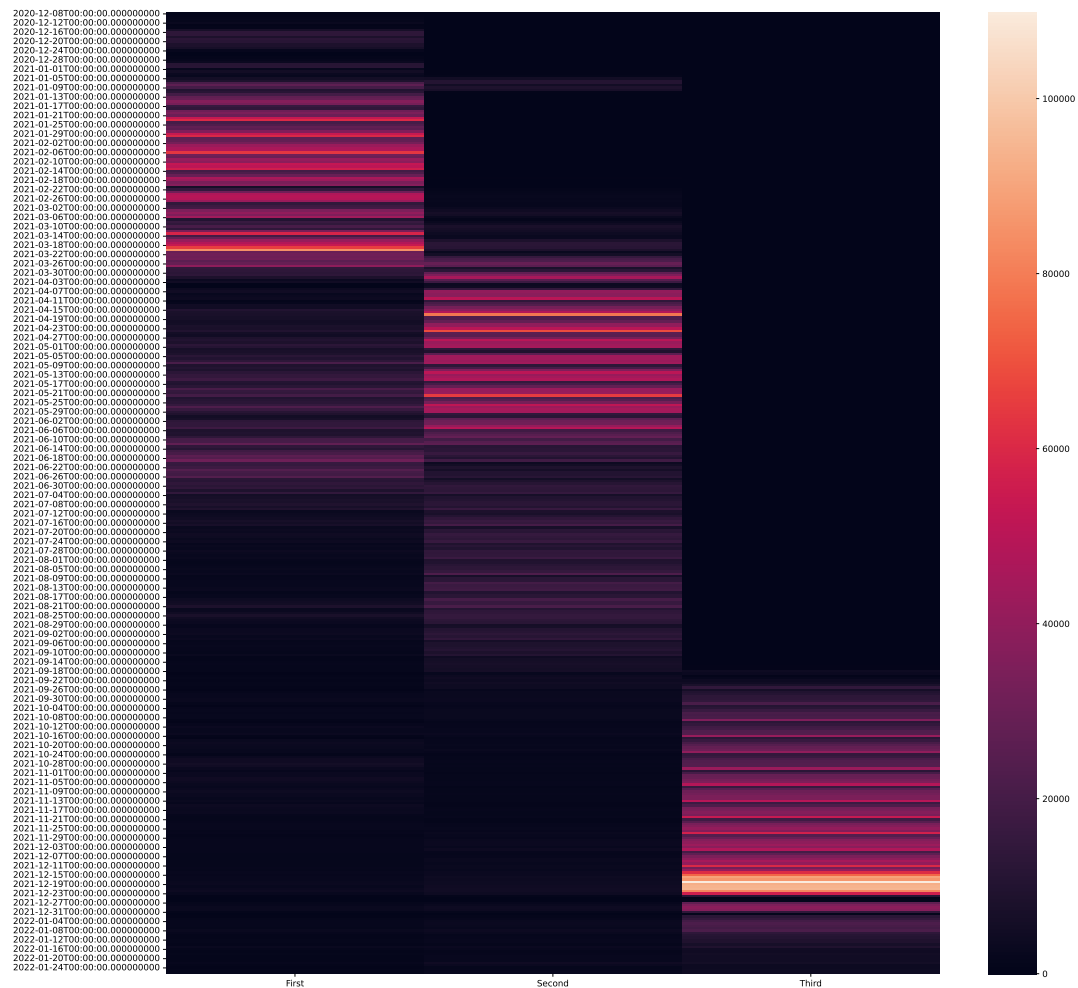
1 Step 4 Machine learning

1.1 Step 0: Look at and Modify the dataset

So, I am curious. Can I predict vaccination data?

I will work with the South West's vaccination data.

	First	Second	Third
2022-01-26	986	2520	4034
2022-01-25	899	1845	4283
2022-01-24	723	1445	3441
2022-01-23	1035	3007	3439
2022-01-22	1822	4709	5896



As we can see, there are waves. So, the count of jabs depends on dates.

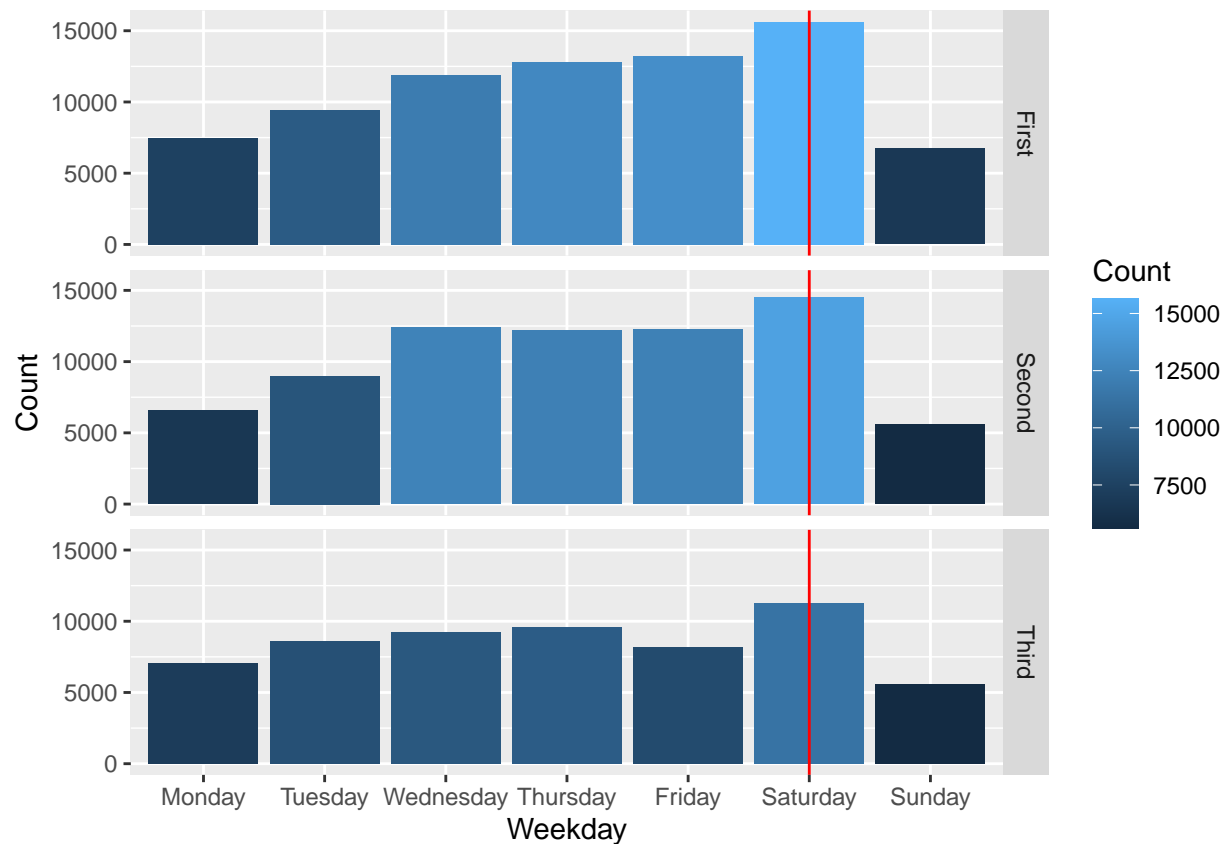
Let's get features: 1) Year 2) Month 3) Day etc.

	First	Second	Third	Year	Month	Day	DayOfYear	Weekday	Quarter	IsMonthStart	IsMonthEnd
2022-01-26	986	2520	4034	2022	1	26	26	2	1	FALSE	FALSE
2022-01-25	899	1845	4283	2022	1	25	25	1	1	FALSE	FALSE
2022-01-24	723	1445	3441	2022	1	24	24	0	1	FALSE	FALSE
2022-01-23	1035	3007	3439	2022	1	23	23	6	1	FALSE	FALSE
2022-01-22	1822	4709	5896	2022	1	22	22	5	1	FALSE	FALSE

1.2 Step 1: Explore the dataset

1.2.1 Weekdays

As you remember, I have a question.



Let's answer.

So, most of South West's people prefer to get a jab on Saturdays.

1.2.2 Missing values

Calculate a count of dates in the dataset.

```
## 415
```

Calculate a count of dates between maximum and minimum dates.

```
## 415
```

There are no missing dates.

1.3 Step 2: Split sets, train a Machine Learning Model and Evaluate performance

Define necessary variables

Prepare sets and train models using parameters.

Compare the score with the mean value of the column that we predicted.

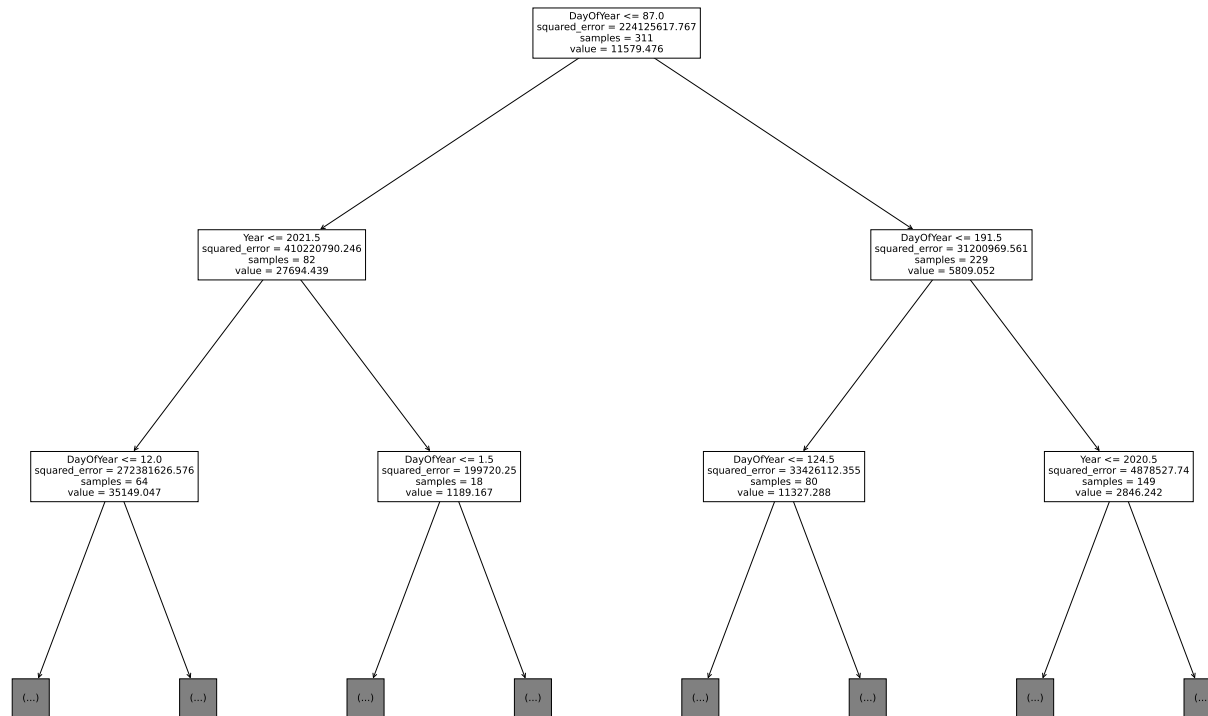
```
## 0.719657929335243
```

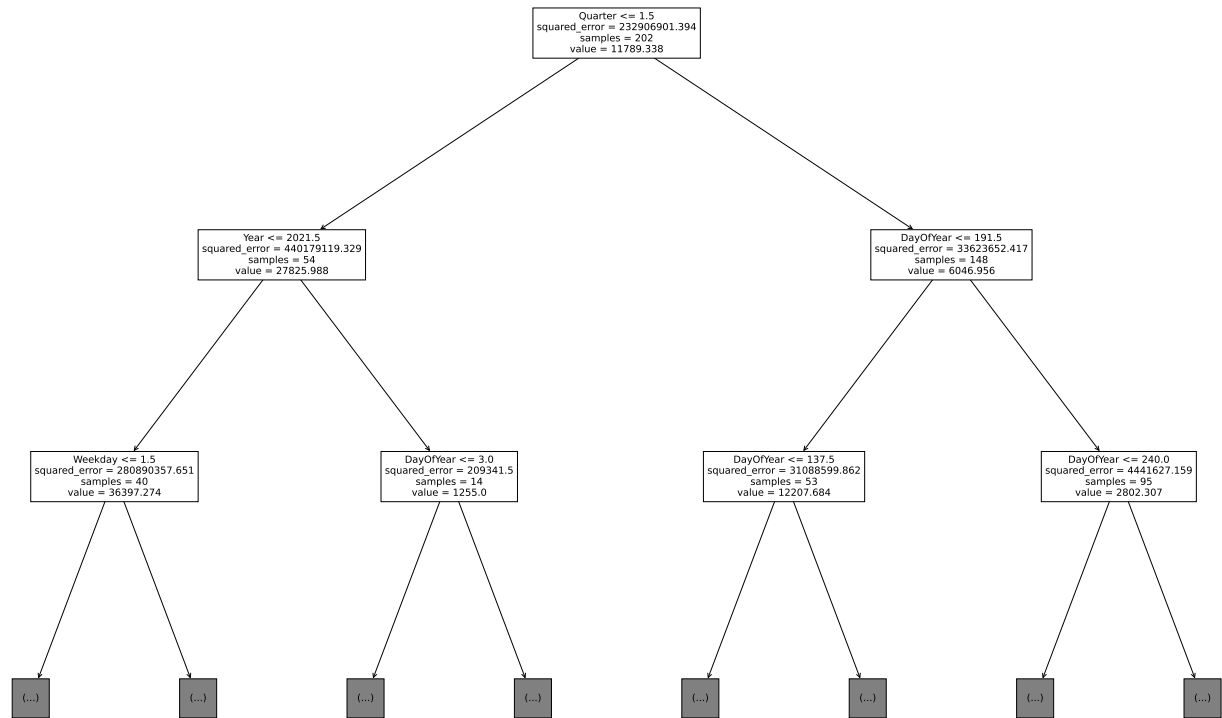
```
## 0.774580856609961
```

```
## 0.7445280765098062
```

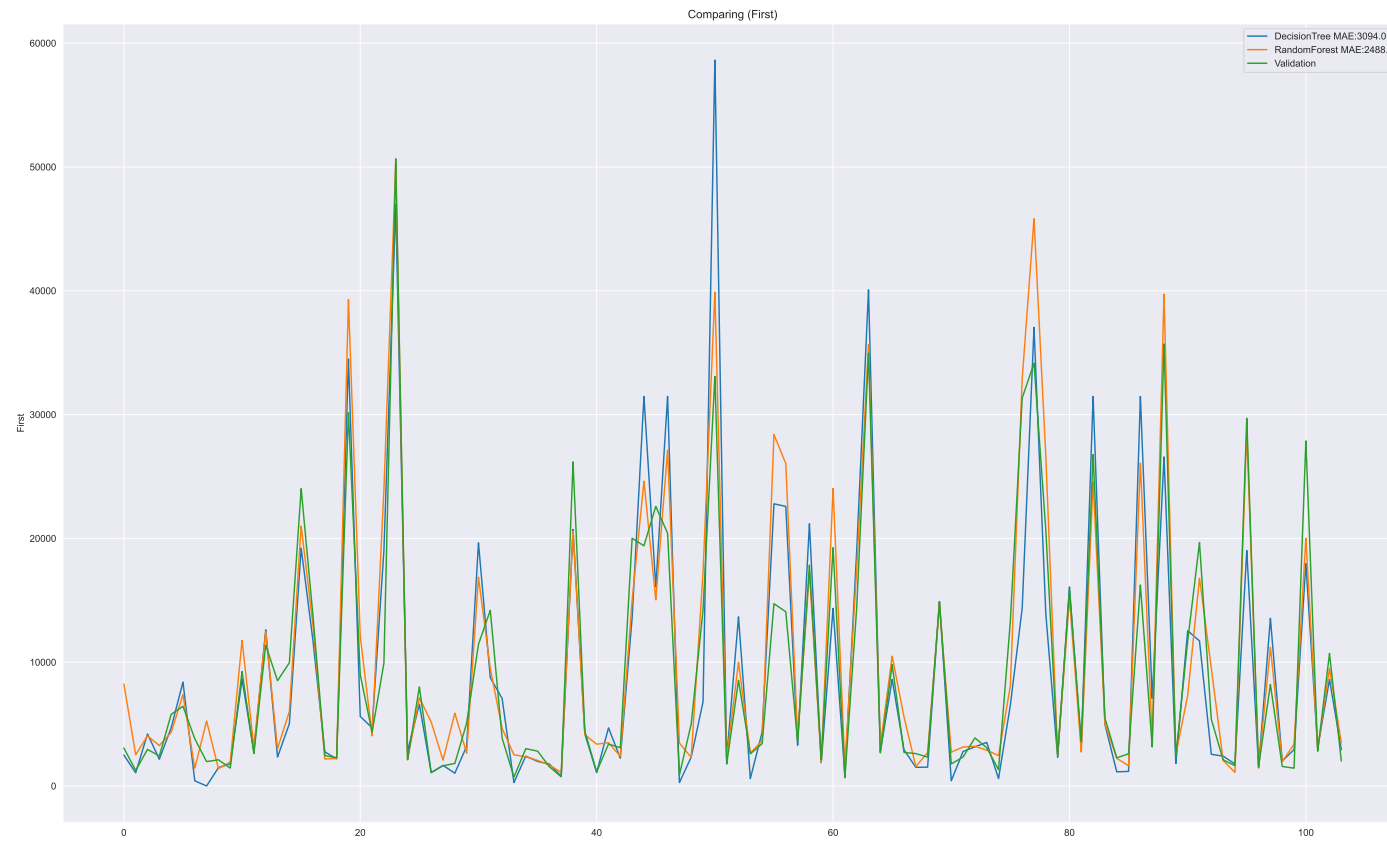
```
## 0.8144473412230163
```

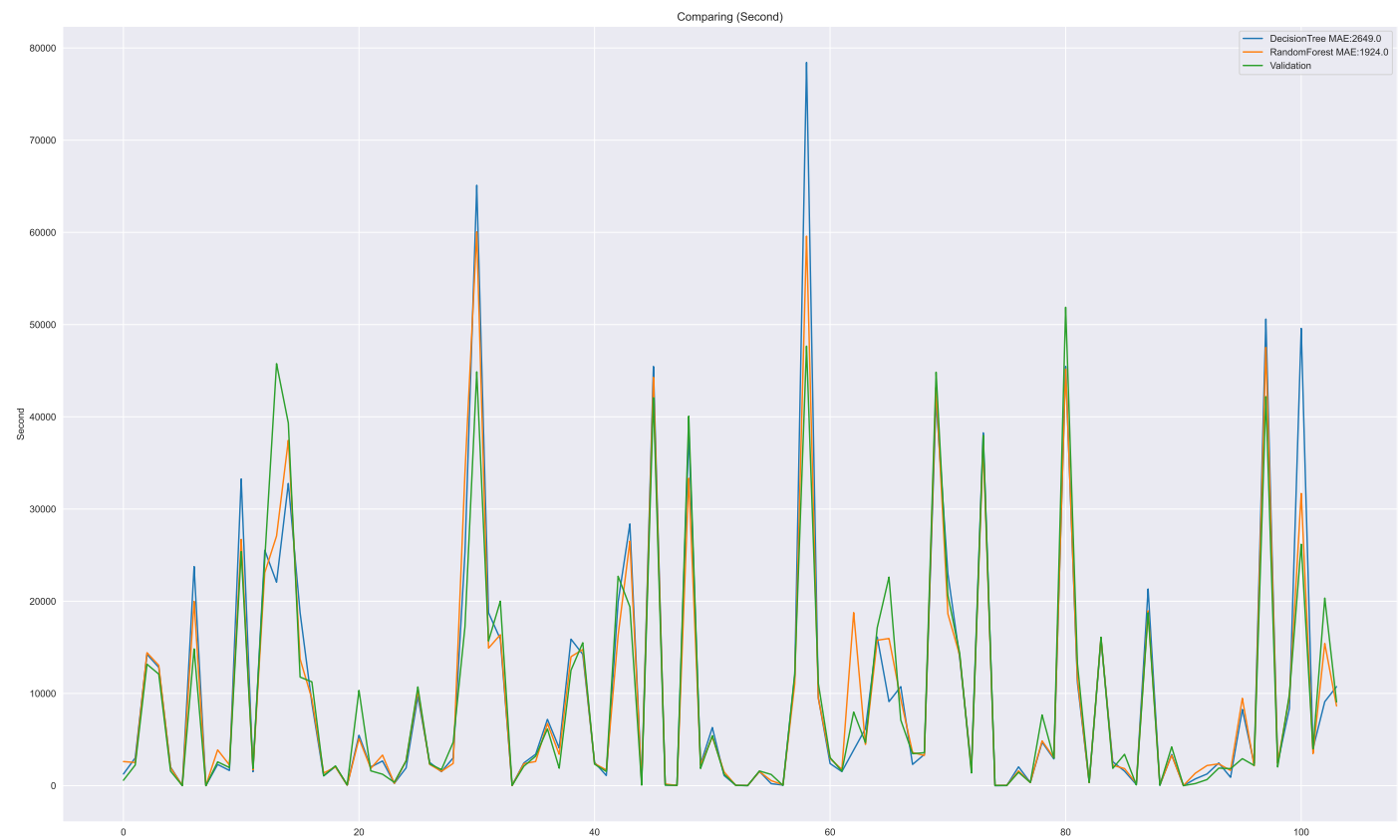
Look at the tree





1.4 Step 3: Plot results





1.5 Step 4: Improve models by changing the dataset

I am going to work with features.

Define necessary variables

Prepare sets and Train models

Compare the score with the mean value of the column that we predicted.

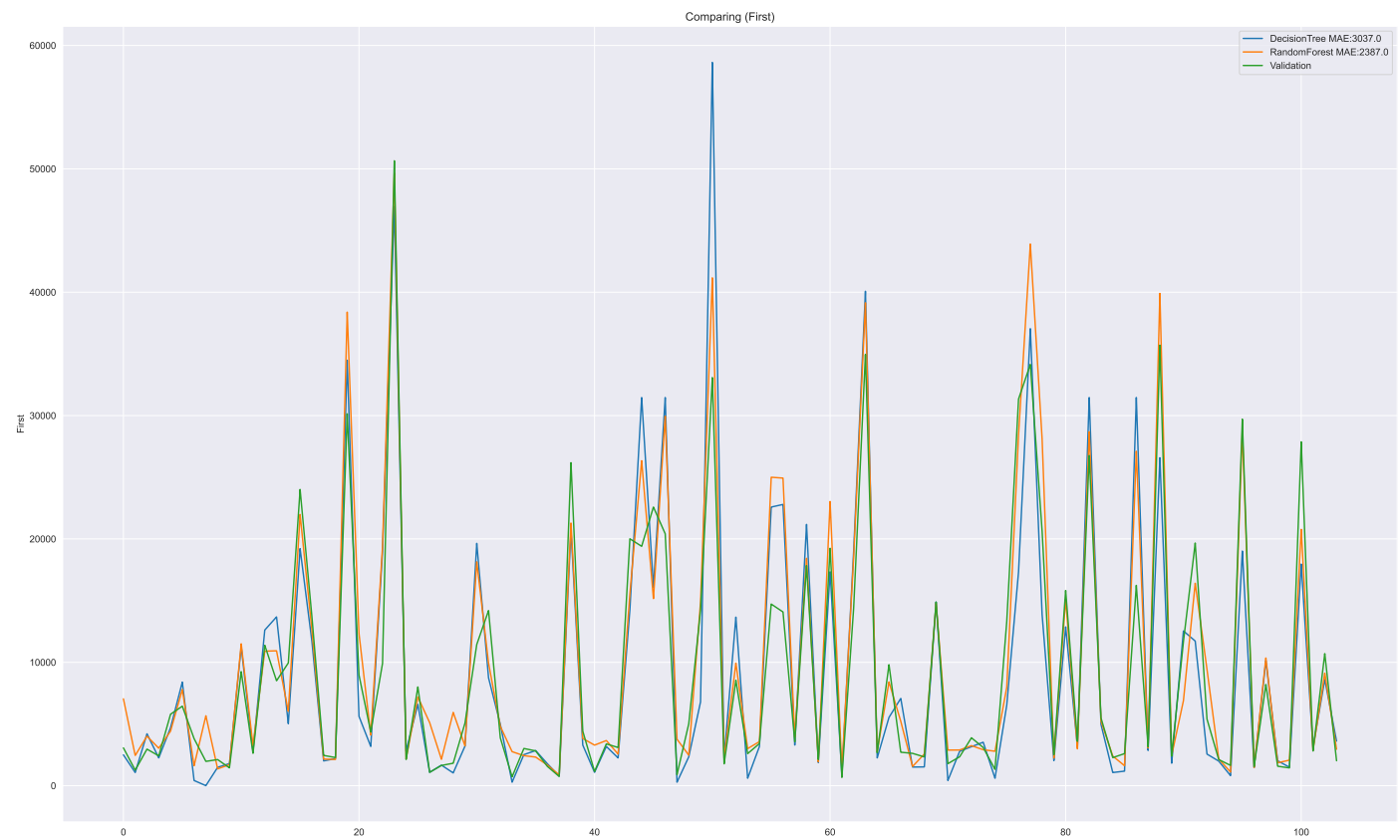
```
## 0.7248630326024768
```

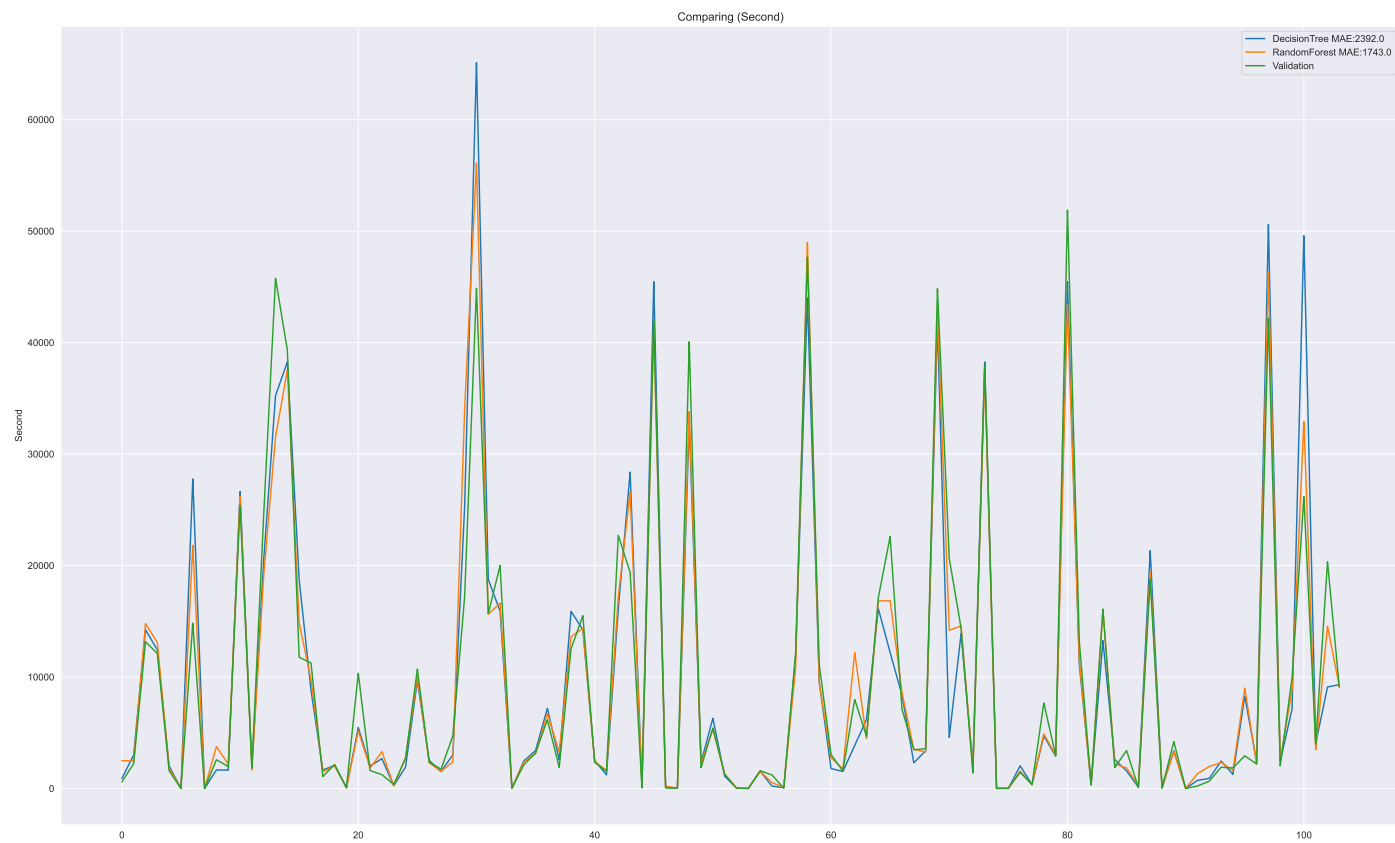
```
## 0.7837038702898657
```

```
## 0.7692636418171874
```

```
## 0.8318561653086721
```

Plot the result.





A combination of the following features give us the best result: * Weekday, * Year, * DayOfYear