Seasonality

INTERMEDIATE PREDICTIVE ANALYTICS IN PYTHON

Nele Verbiest
Senior Data Scientist
@PythonPredictions

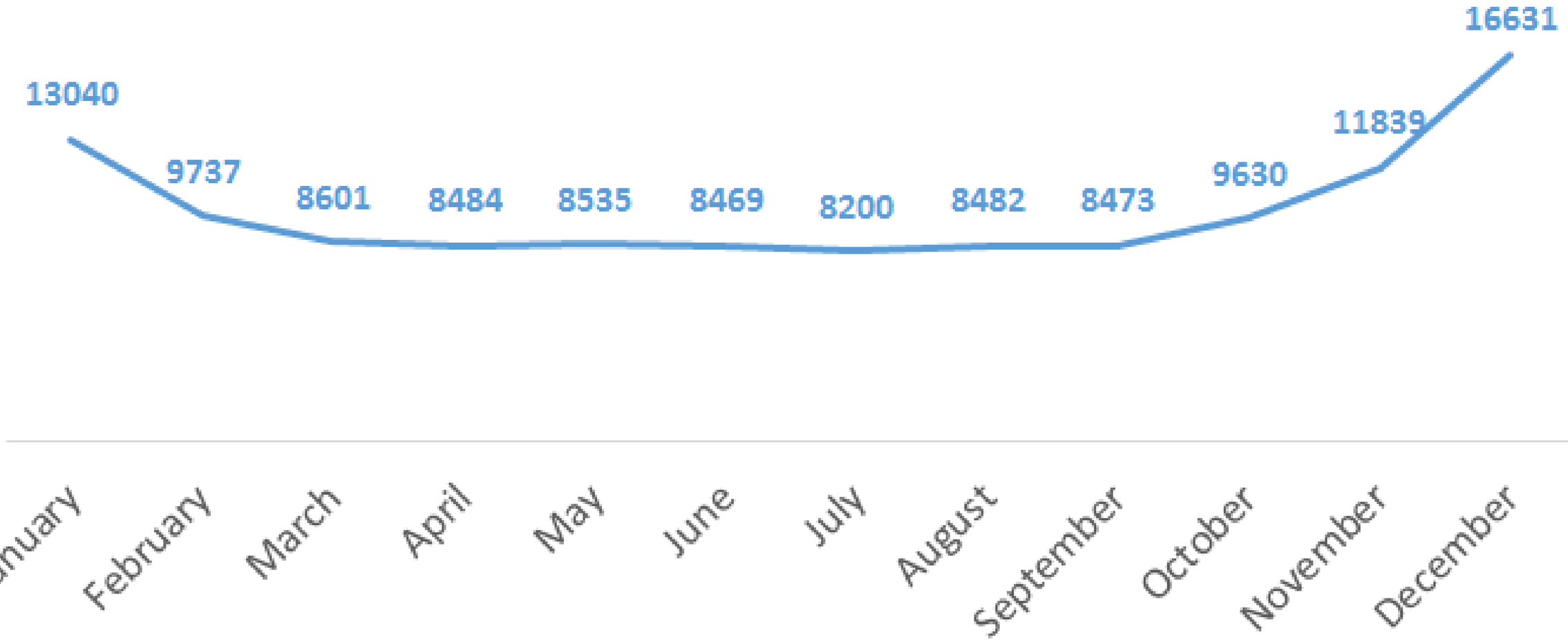
datacamp

Seasonal effects (1)

Mean donation last year, Number donations last year, ...	Target: > EUR 50 donations

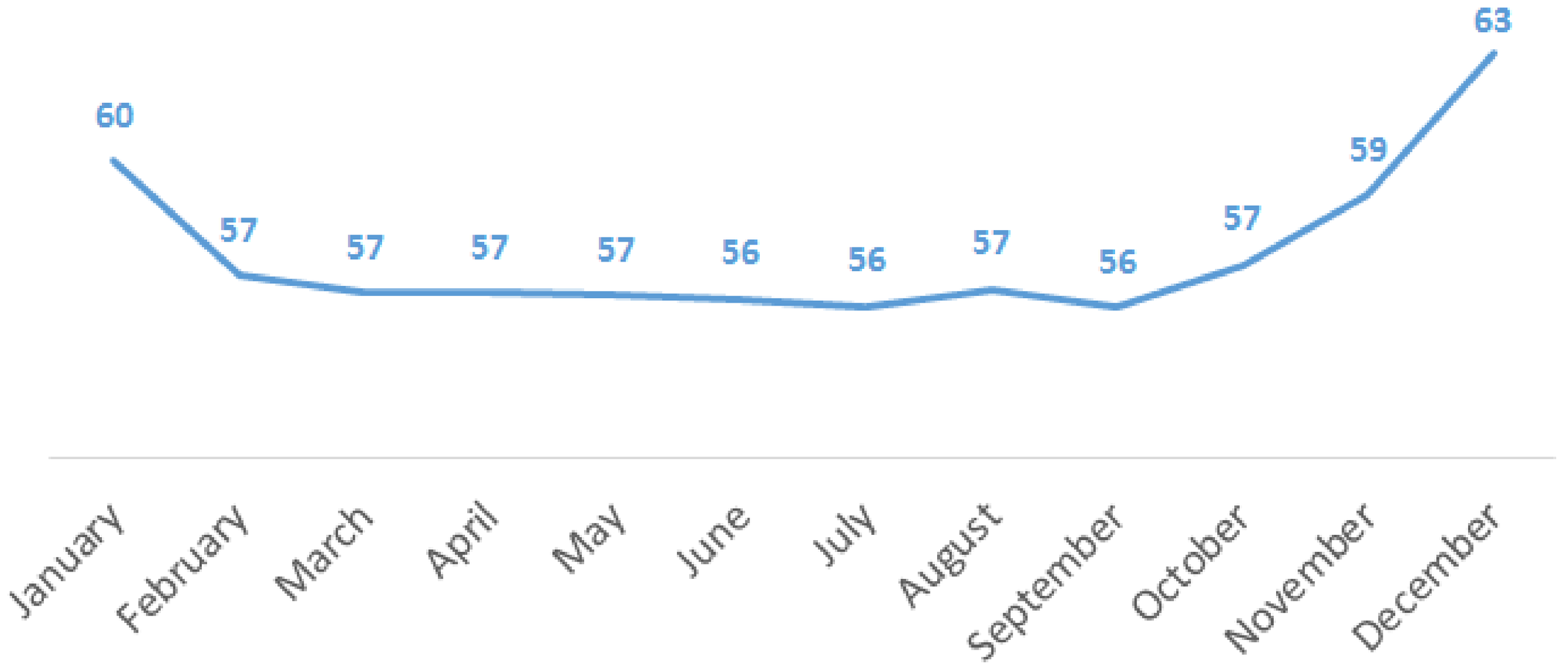
Seasonal effects (2)

Mean number of donations

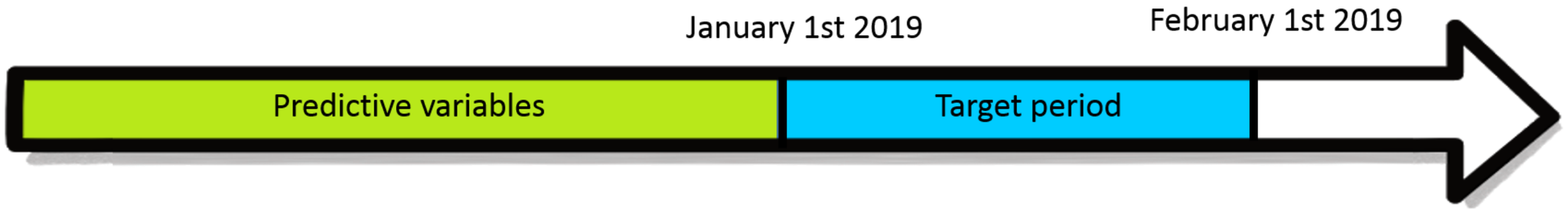


Seasonal effects (3)

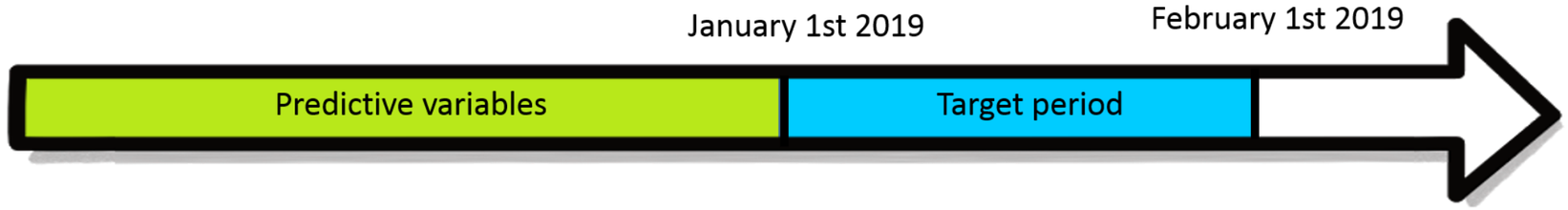
Mean donation amount

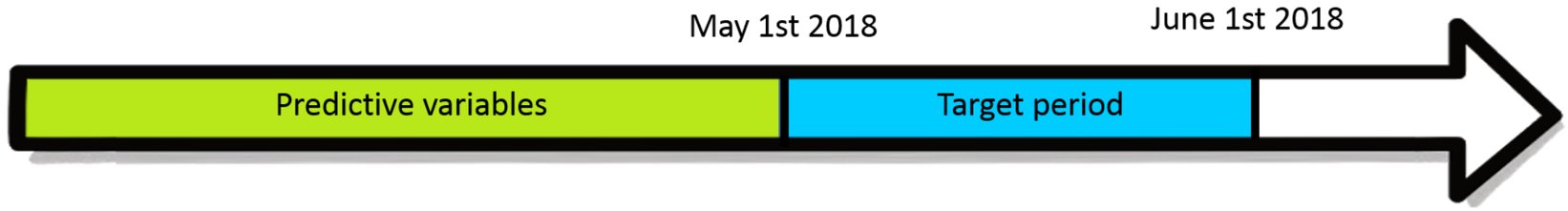


Seasonality and the timeline (1)



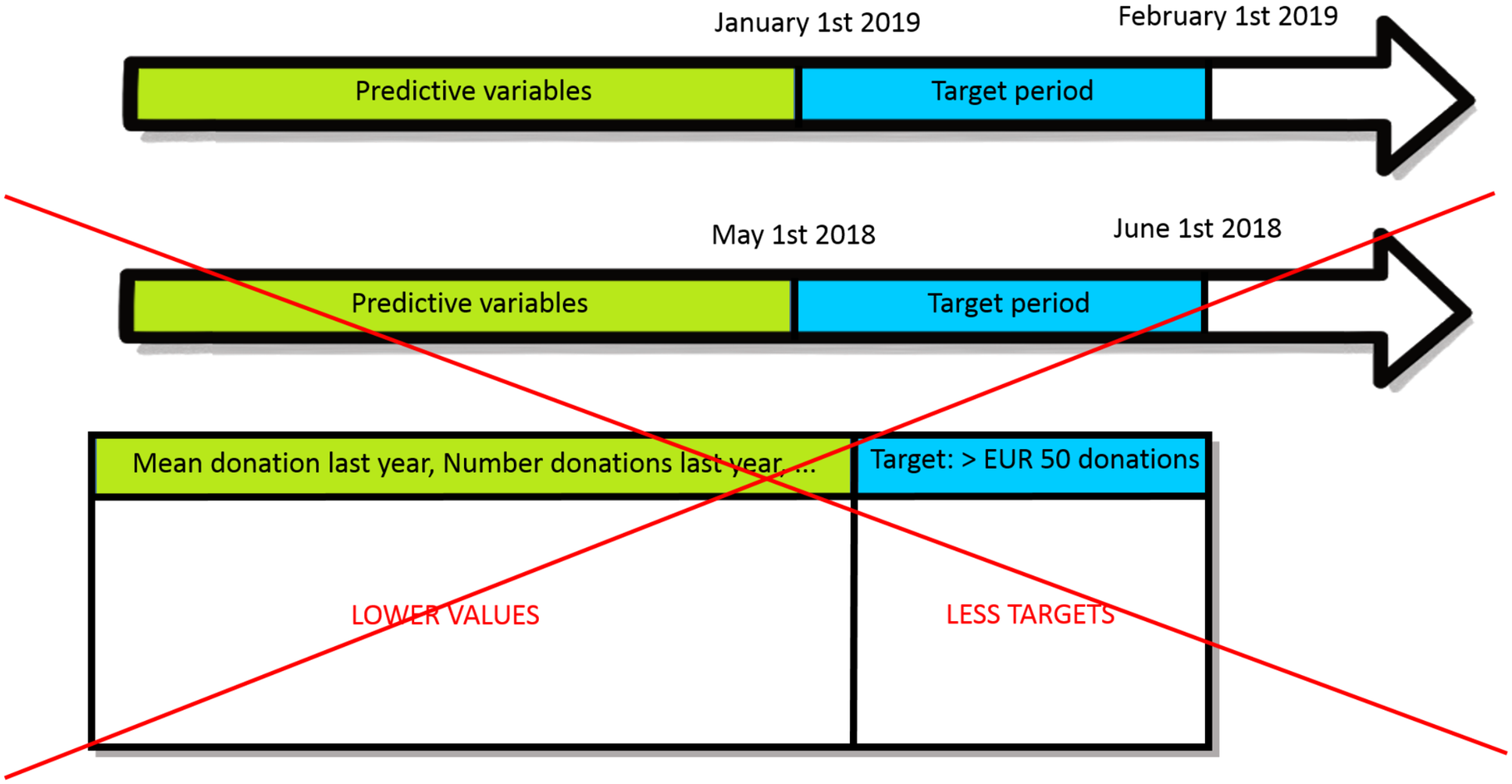
Seasonality and the timeline (2)





Mean donation last year, Number donations last year, ...	Target: > EUR 50 donations
LOWER VALUES	LESS TARGETS

Seasonality and the timeline (3)

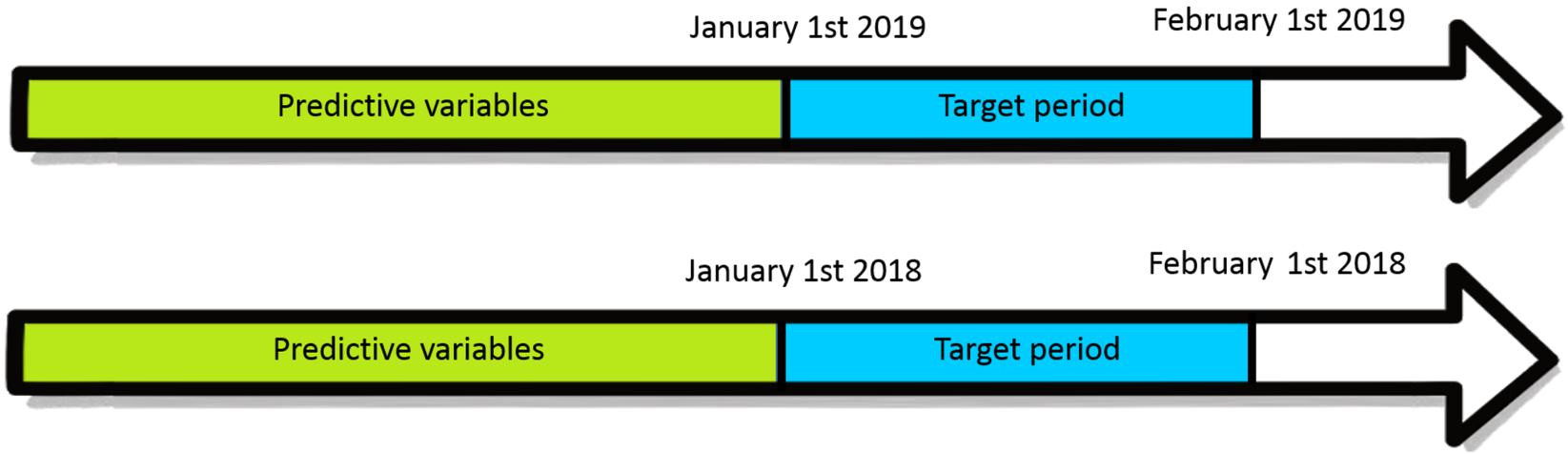


Dealing with seasonality

- Check for seasonality

```
gifts.groupby("month")["amount"].mean()
gifts.groupby("month").size()
```

- Use appropriate timeline in history



Seasonality and predictive models

Model timeline May 2018

```
logreg = linear_model.LogisticRegression()
logreg.fit(X_may2018, y_may2018)
predictions = logreg.predict_proba(X_jan2019)[:,1]
auc = roc_auc_score(y_jan2019, predictions)
print(round(auc,2))
```

```
0.53
```

Model timeline January 2018

```
logreg = linear_model.LogisticRegression()
logreg.fit(X_jan2018, y_jan2018)
predictions = logreg.predict_proba(X_jan2019)[:,1]
auc = roc_auc_score(y_jan2019, predictions)
print(round(auc,2))
```

```
0.56
```

Let's practice!

INTERMEDIATE PREDICTIVE ANALYTICS IN PYTHON

Using multiple snapshots

INTERMEDIATE PREDICTIVE ANALYTICS IN PYTHON

Nele Verbiest
Senior Data Scientist
@PythonPredictions

datacamp

Not enough data

- Small population

```
print(len(basetable))
```

```
4738
```

- Small number of targets

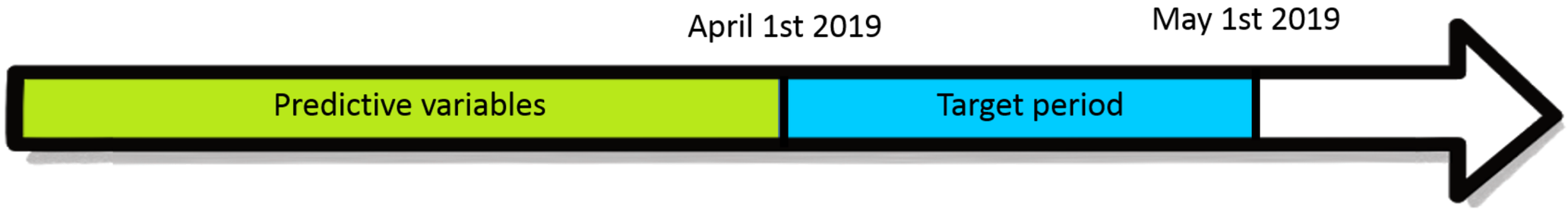
```
print(len(basetable))
```

```
394010
```

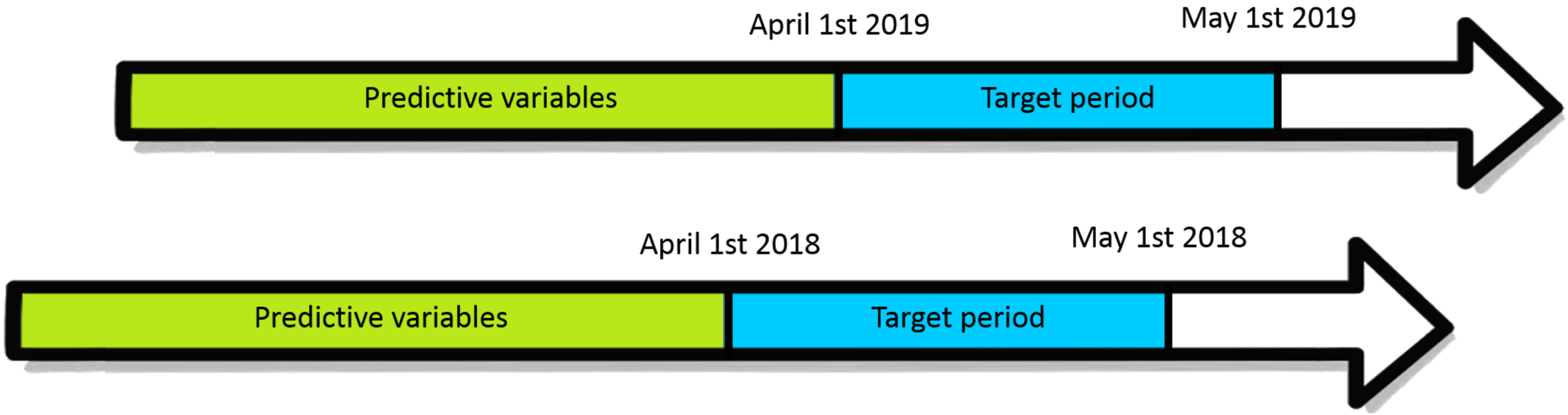
```
print(sum(basetable["target"]))
```

```
230
```

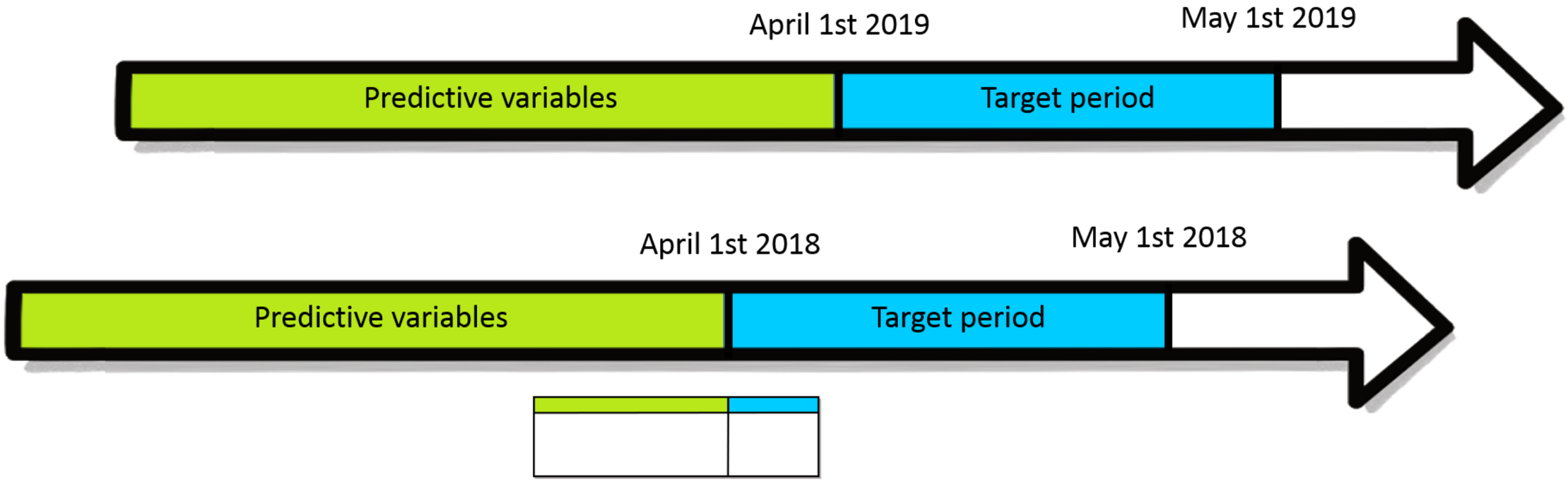
Using multiple snapshots (1)



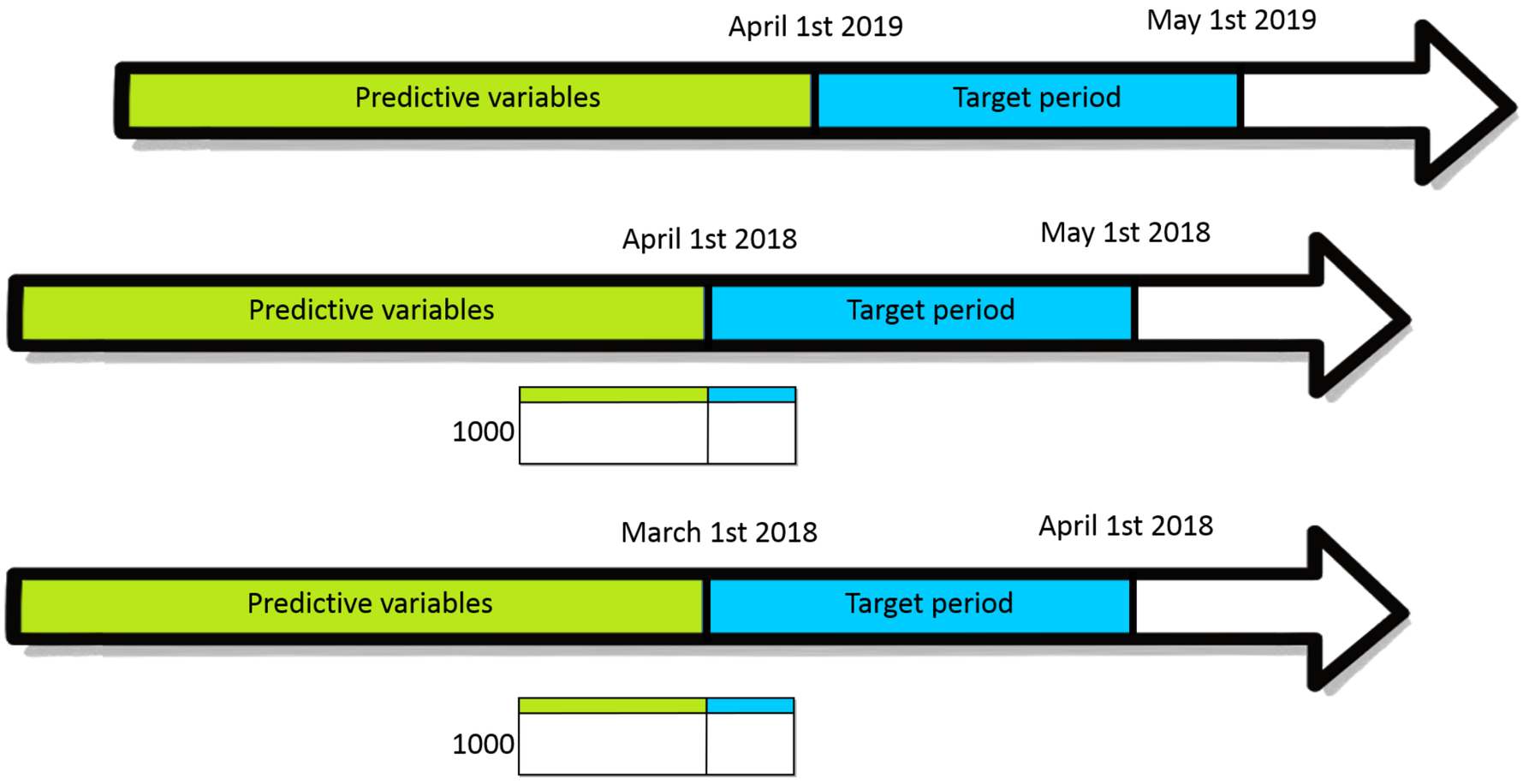
Using multiple snapshots (2)



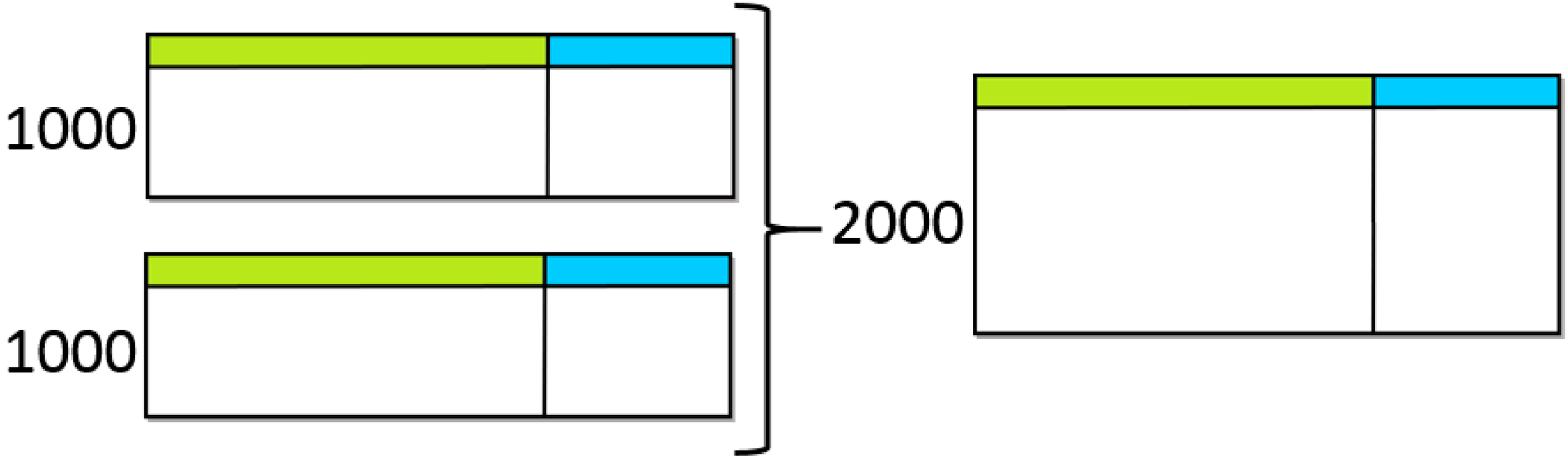
Using multiple snapshots (3)



Using multiple snapshots (4)

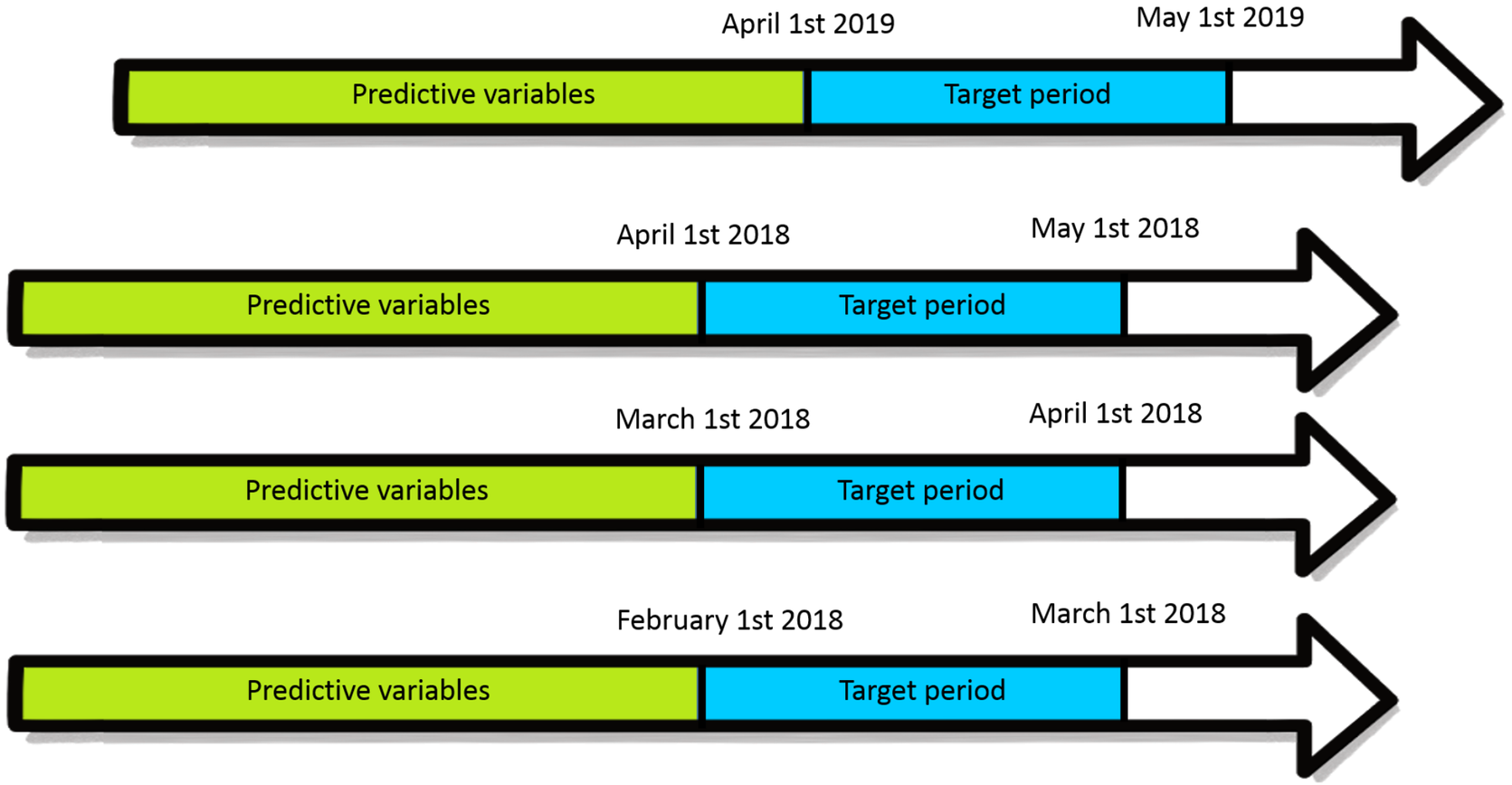


Stacking basetables




```
basetable = basetable_april2018.append(basetable_march2018)
```

Snapshots and seasonality



Let's practice!

INTERMEDIATE PREDICTIVE ANALYTICS IN PYTHON

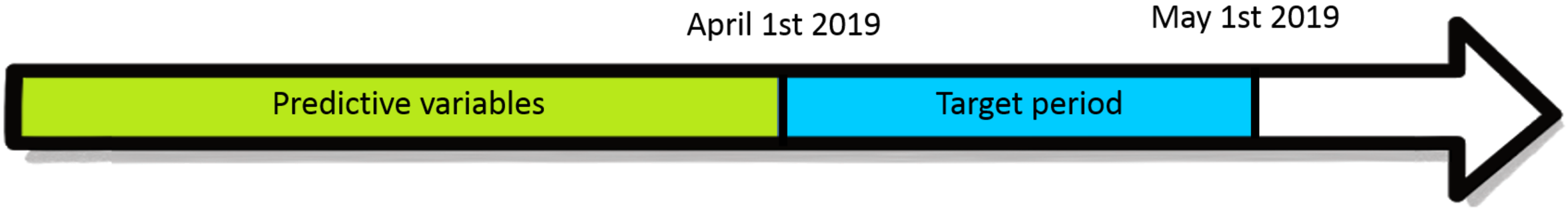
The timegap

INTERMEDIATE PREDICTIVE ANALYTICS IN PYTHON

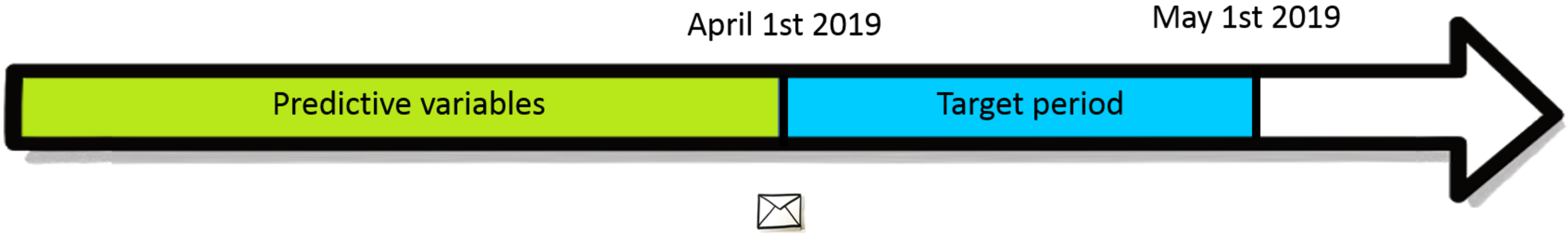
Nele Verbiest
Senior Data Scientist
@PythonPredictions

datacamp

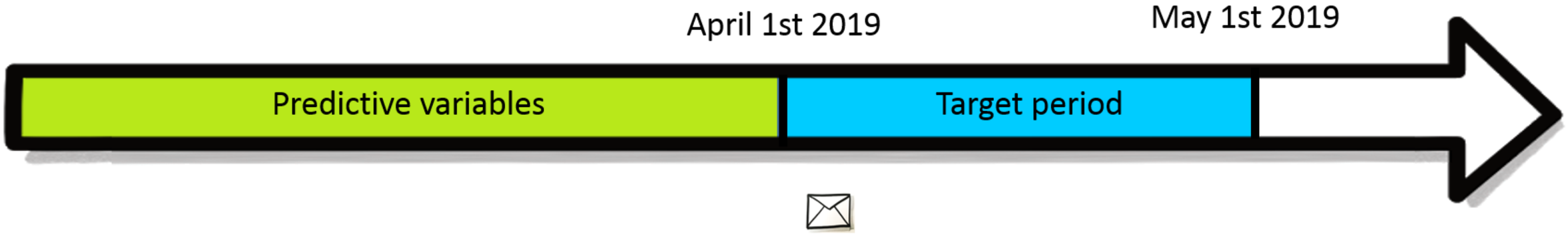
Timegap: motivation (1)



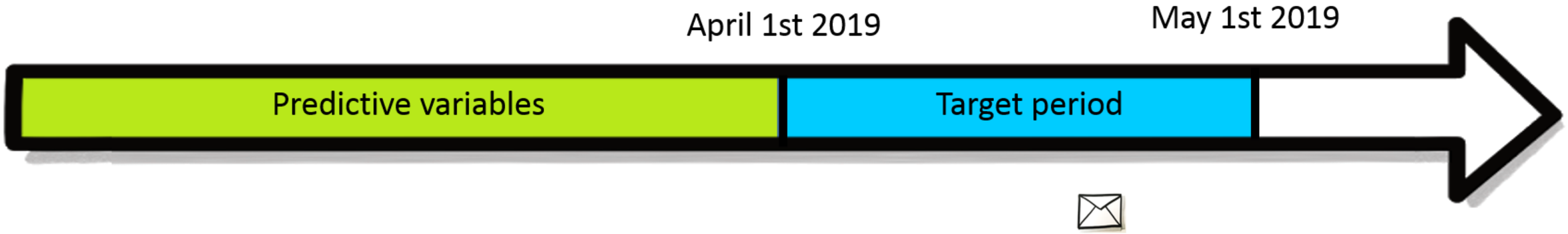
Timegap: motivation (2)



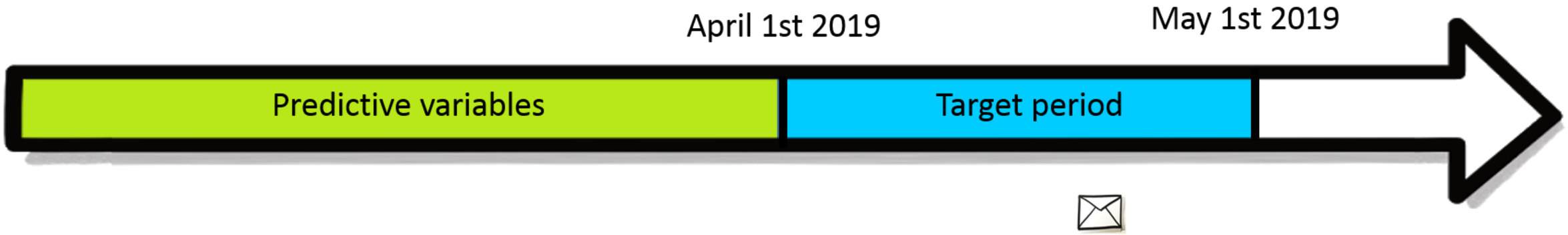
Timegap: motivation (3)



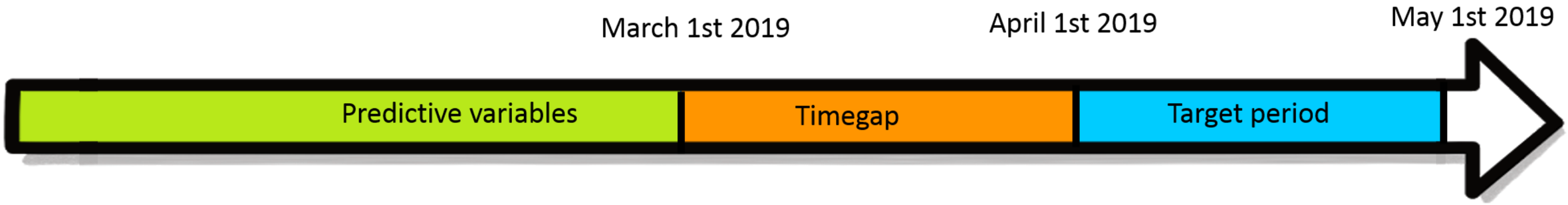
Timegap: motivation (4)



Timegap: motivation (5)



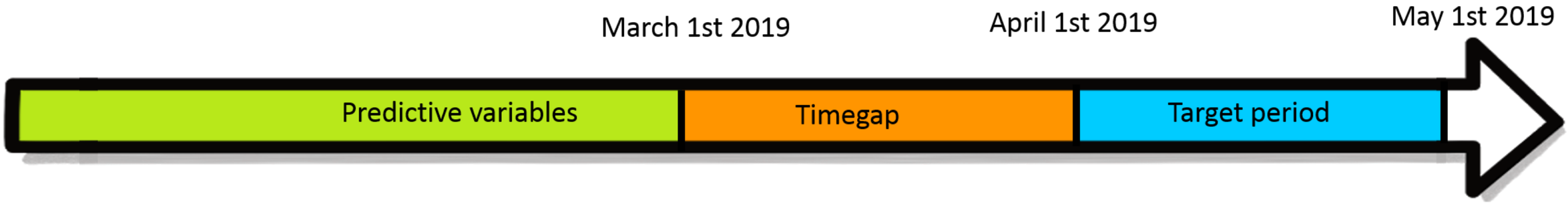
Adding a timegap



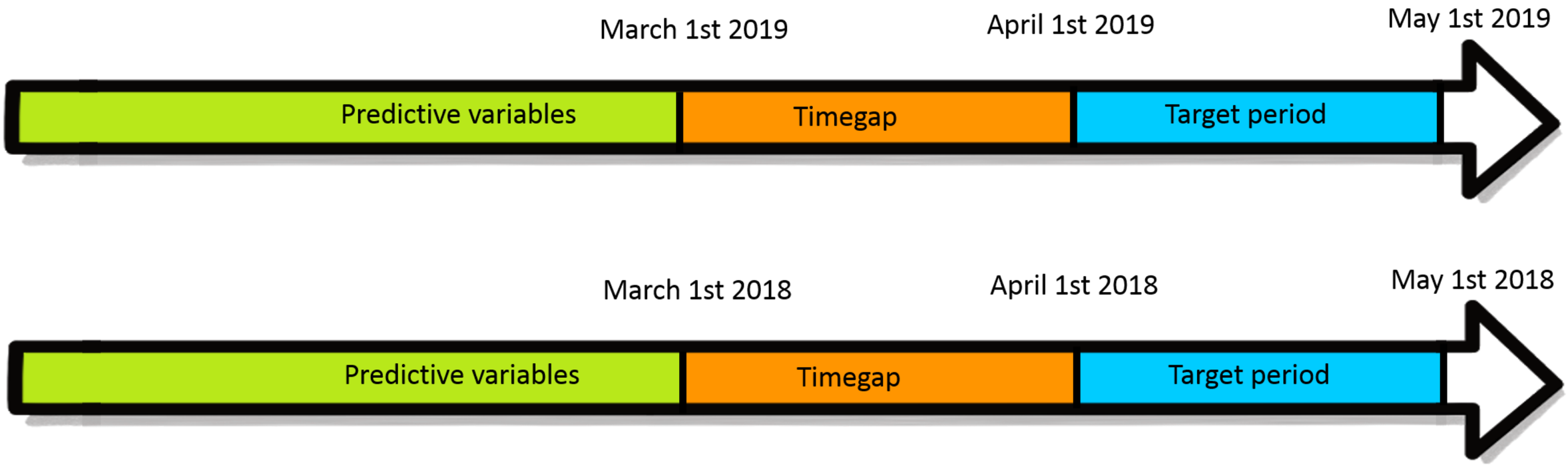
Timegap:

- Gather data
- Run the model
- Prepare the campaign

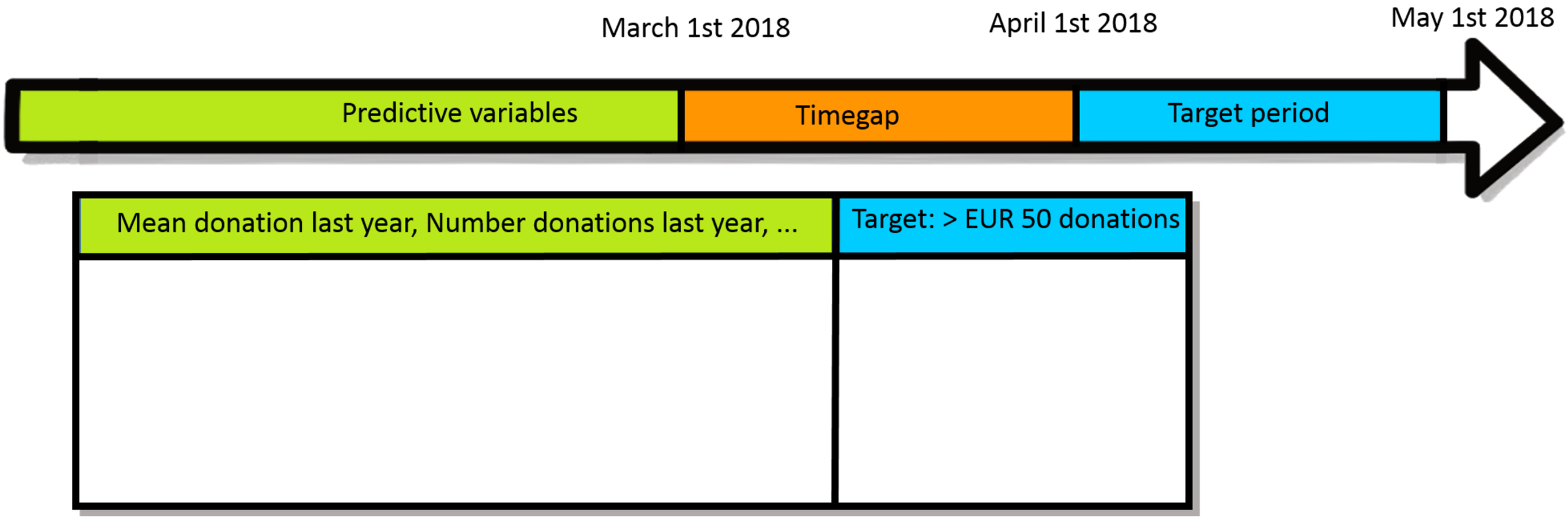
Reconstructing the timeline in history (1)



Reconstructing the timeline in history (2)



Constructing the basetable



Let's practice!

INTERMEDIATE PREDICTIVE ANALYTICS IN PYTHON

Congratulations!

INTERMEDIATE PREDICTIVE ANALYTICS IN PYTHON

Nele Verbiest
Data Scientist @PythonPredictions

What you learned...

1. Draw the timeline:
   - Timegap
2. Reconstruct timeline in history:
   - Seasonality
   - Multiple snapshots
3. Determine the population
4. Calculate the target values
5. Add candidate predictors
6. Clean the data

Congratulations!

INTERMEDIATE PREDICTIVE ANALYTICS IN PYTHON