NTT DATA

SAPIENZA
UNIVERSITÀ DI ROMA

# Predicting costumer complaints in an energy company

Laura Giuliano

Thesis Advisor
Anagnostopoulos Aristidis

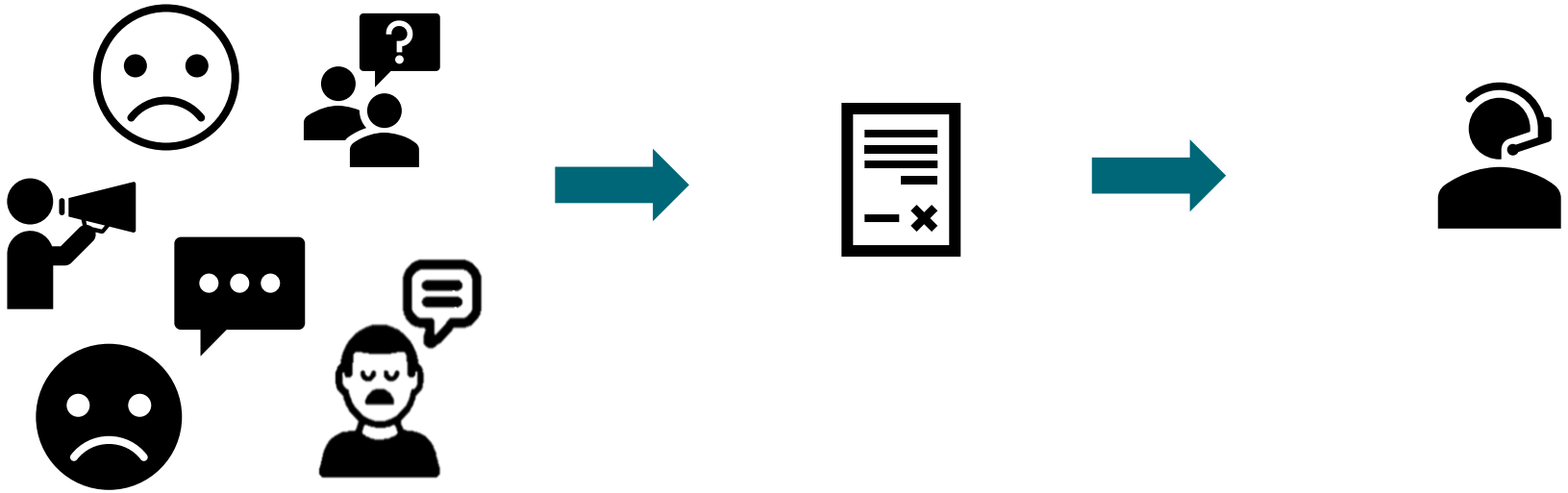
External Advisor
Olivieri Antonio

# Agenda

- Business challenge
- Goal
- Data
- Preprocessing
- Model
- Results
- Conclusions

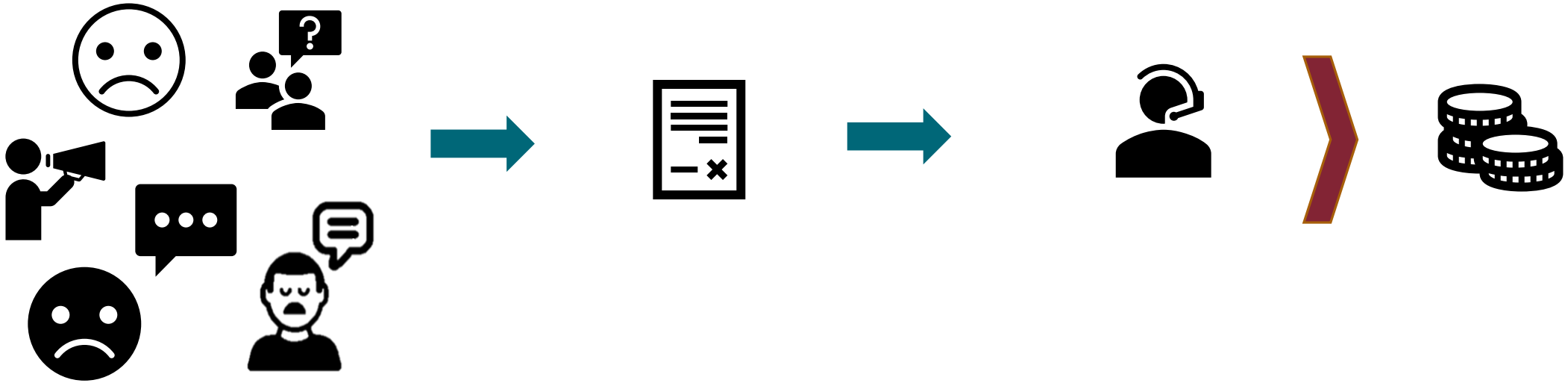
# Business challenge

*Even if you have a good business, customers complaints are unavoidable*



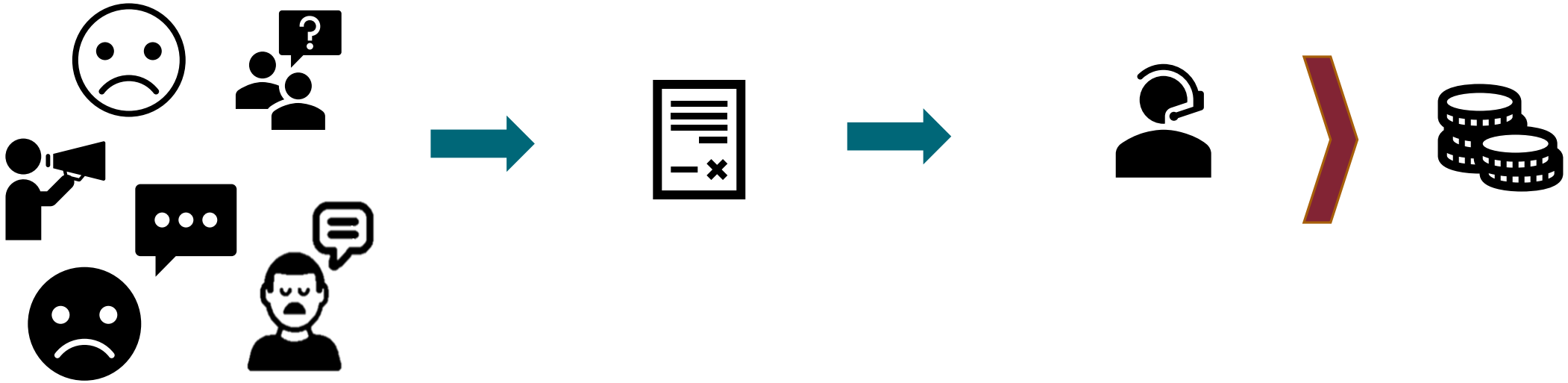
# Business challenge

*Even if you have a good business, customers complaints are unavoidable*



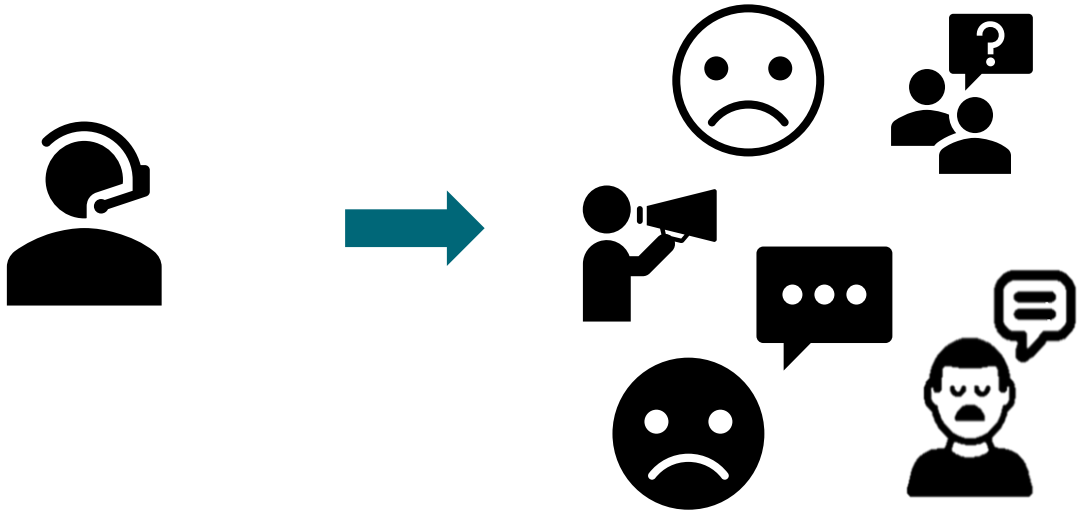
# Business challenge

*Even if you have a good business, customers complaints are unavoidable*



Understanding customers needs in order to prevent complaints

# Goal



# Goal

# Goal

Reducing by 12k customers complaints in a year

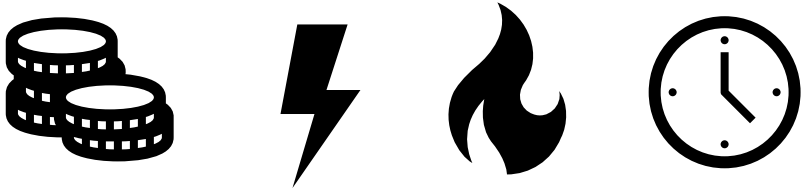
Satisfaction

# Data

- B2B

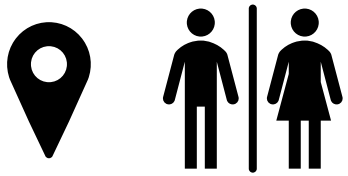
- B2C

- Billing & credit



- Interactions between customers and company

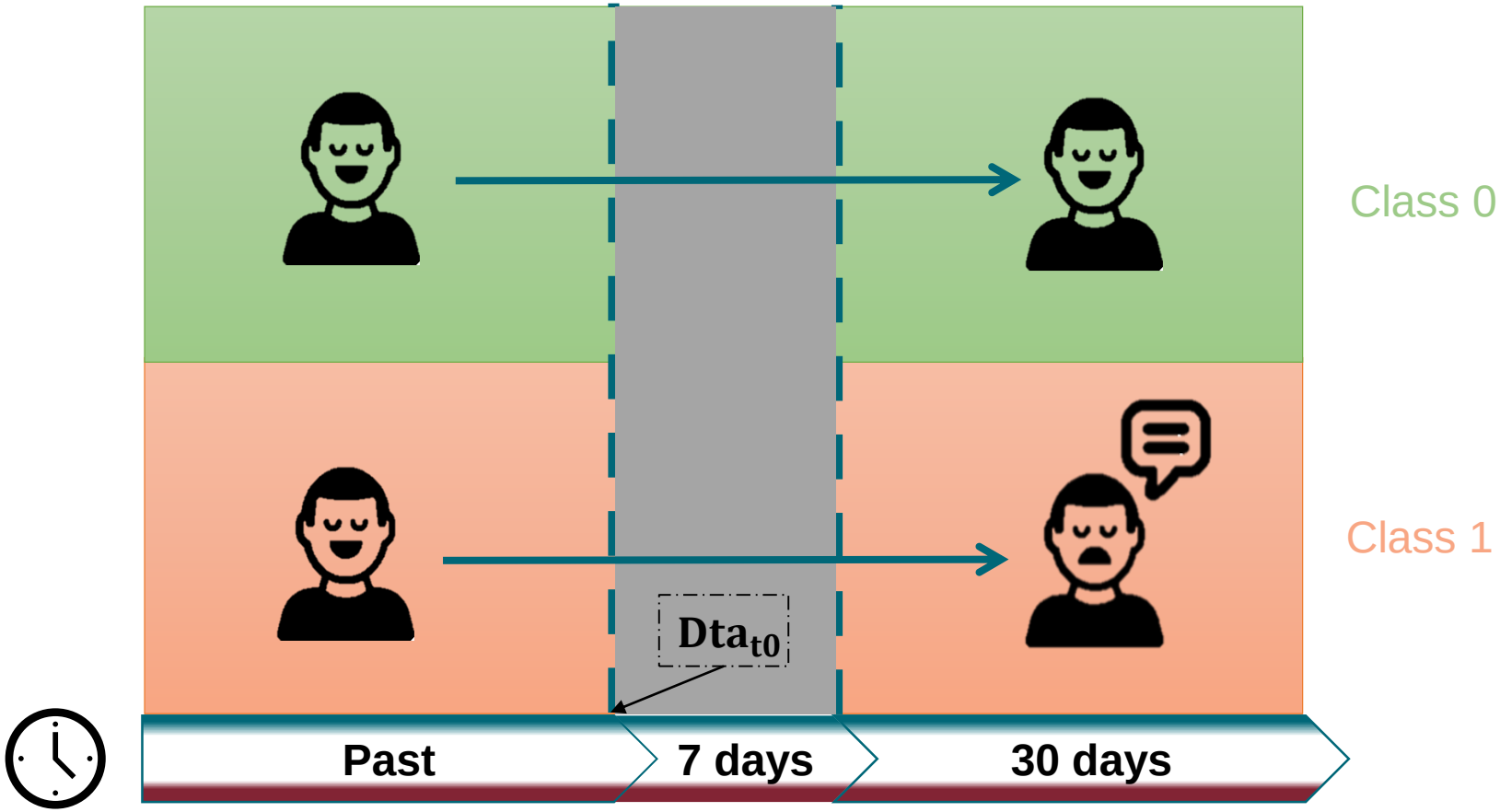
- Socio-Demographic information



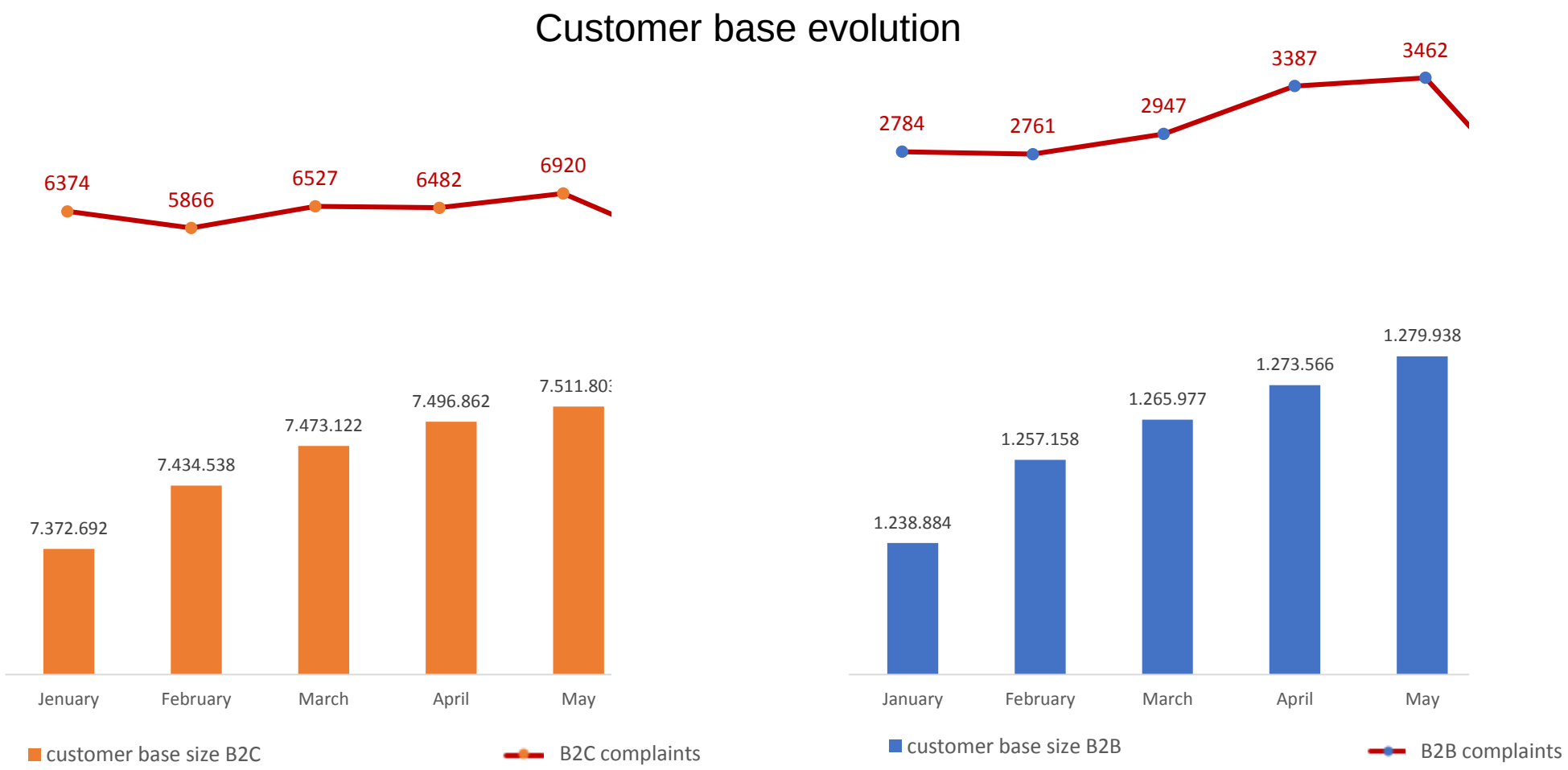
January 2019 – June 2019

# Data: target

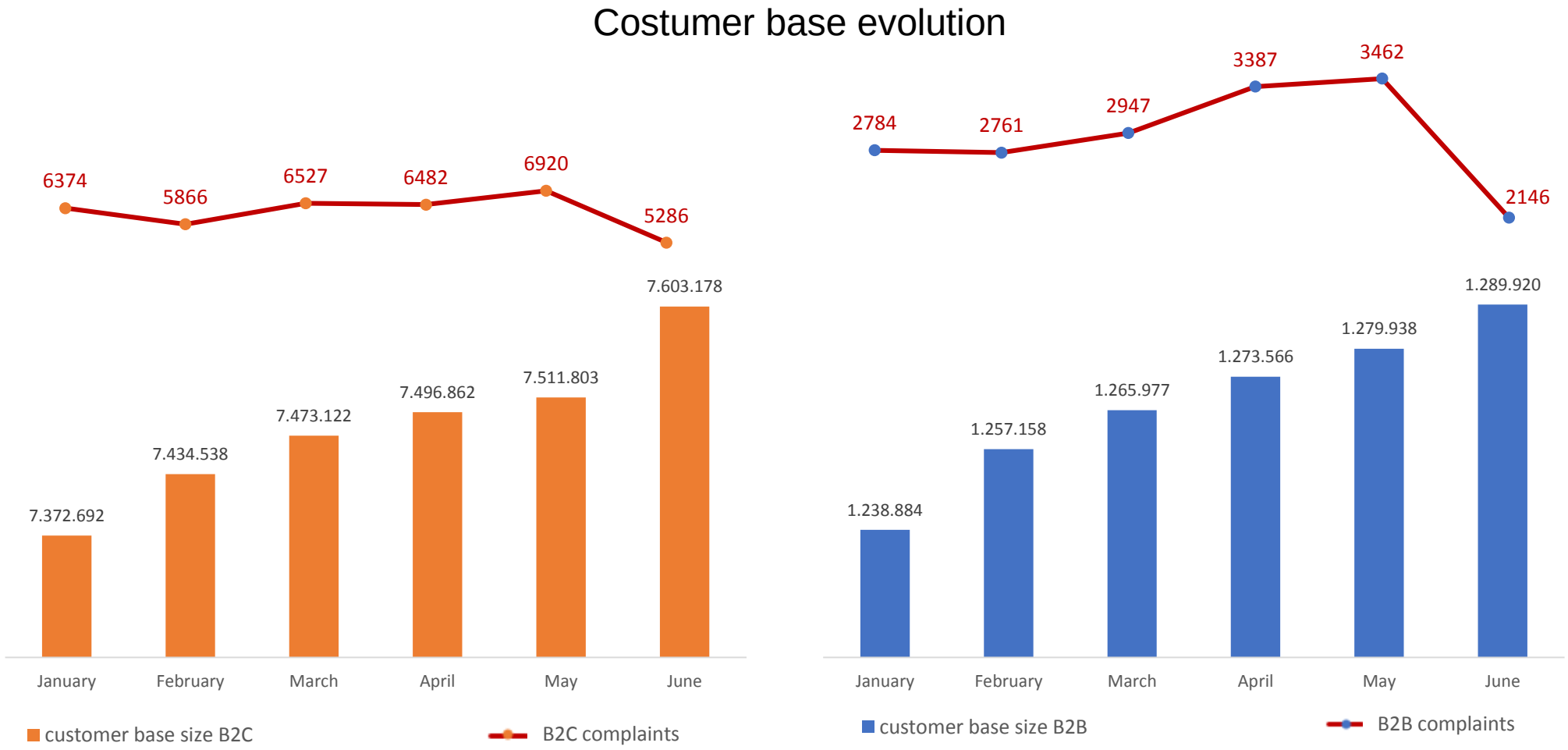
Time window constraint



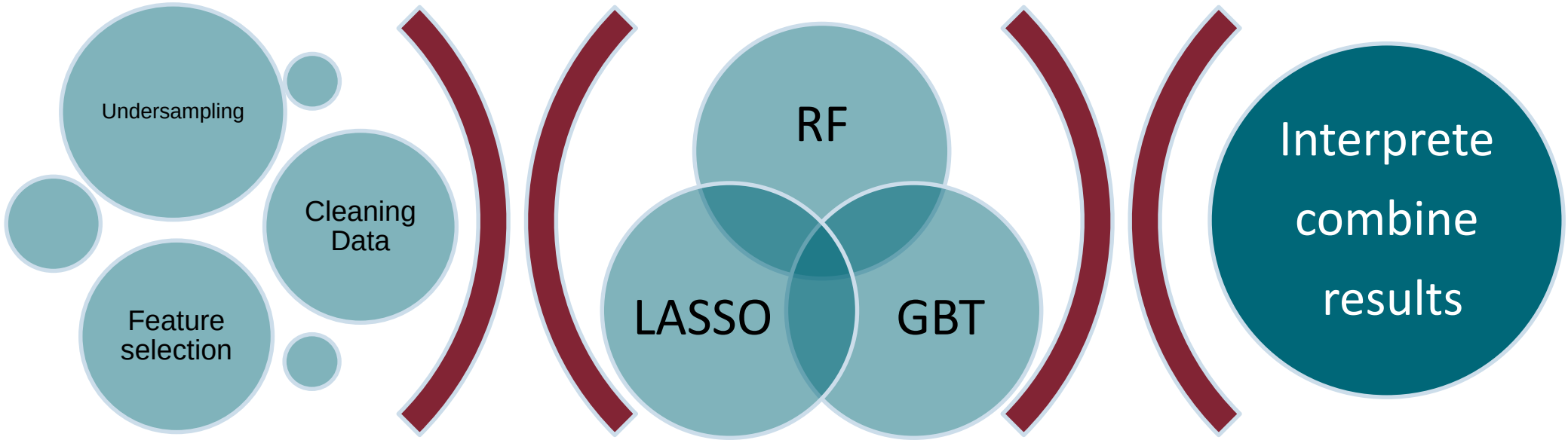
# Data: target


Customer base evolution
6374
5866
6527
6482
6920
2784
2761
2947
3387
3462
7.372.692
7.434.538
7.473.122
7.496.862
7.511.803
1.238.884
1.257.158
1.265.977
1.273.566
1.279.938
Jenuary
February
March
April
May
January
February
March
April
May
customer base size B2C
B2C complaints
customer base size B2B
B2B complaints

# Data: target


Costumer base evolution
6374
5866
6527
6482
6920
5286
7.603.178
7.511.803
7.496.862
7.473.122
7.434.538
7.372.692
January
February
March
April
May
June
customer base size B2C
B2C complaints
2784
2761
2947
3387
3462
2146
1.289.920
1.279.938
1.273.566
1.265.977
1.257.158
1.238.884
January
February
March
April
May
June
customer base size B2B
B2B complaints

# Machine learning process

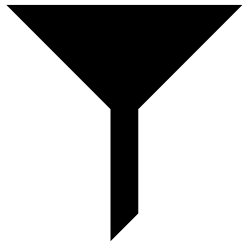



Data preprocessing Models Evaluation

# Preprocessing

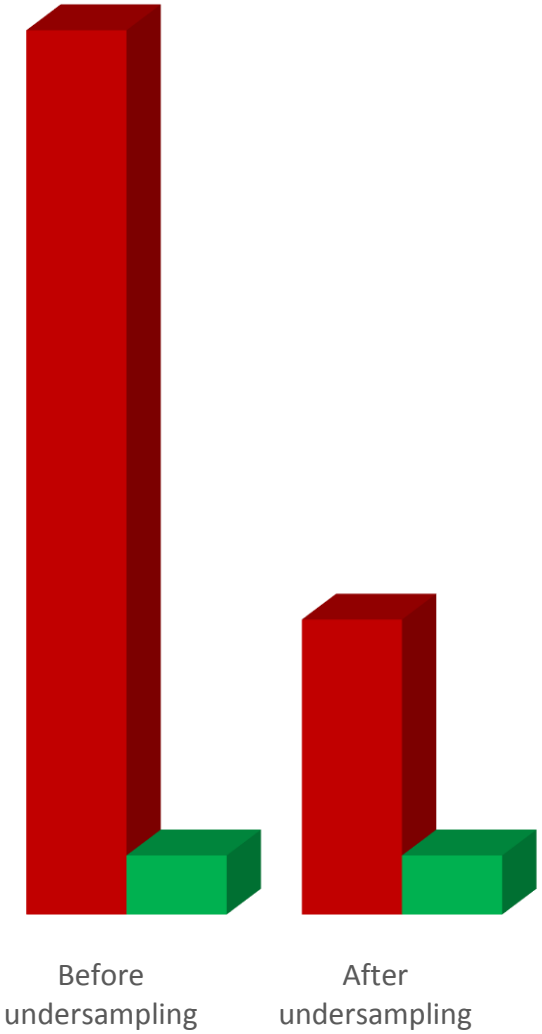
Cleaning data

Undersampling

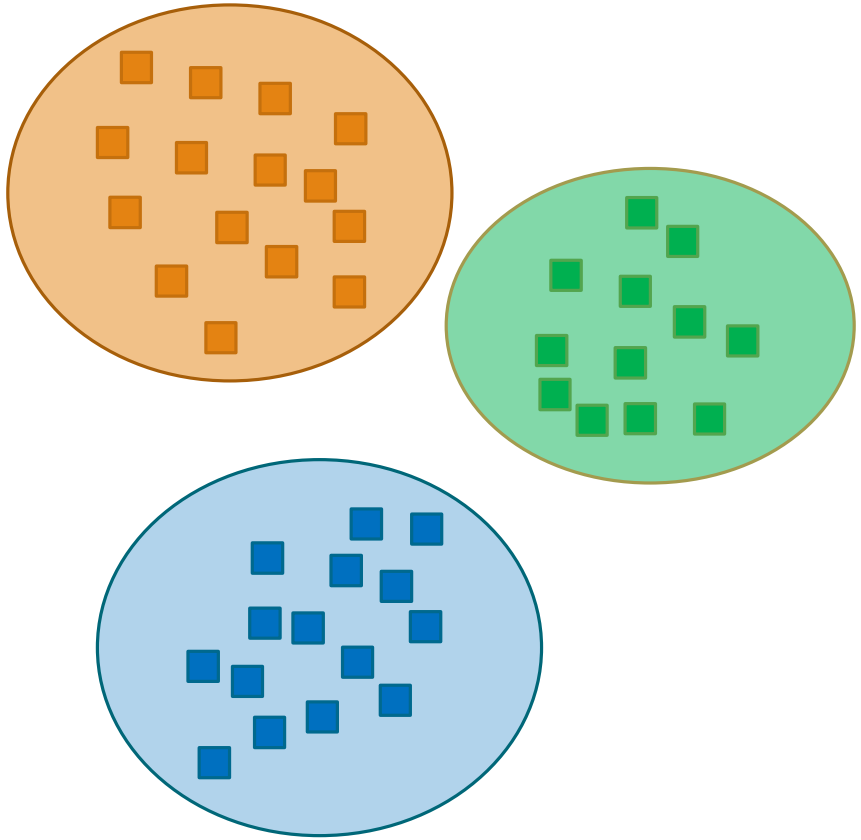


Feature selection

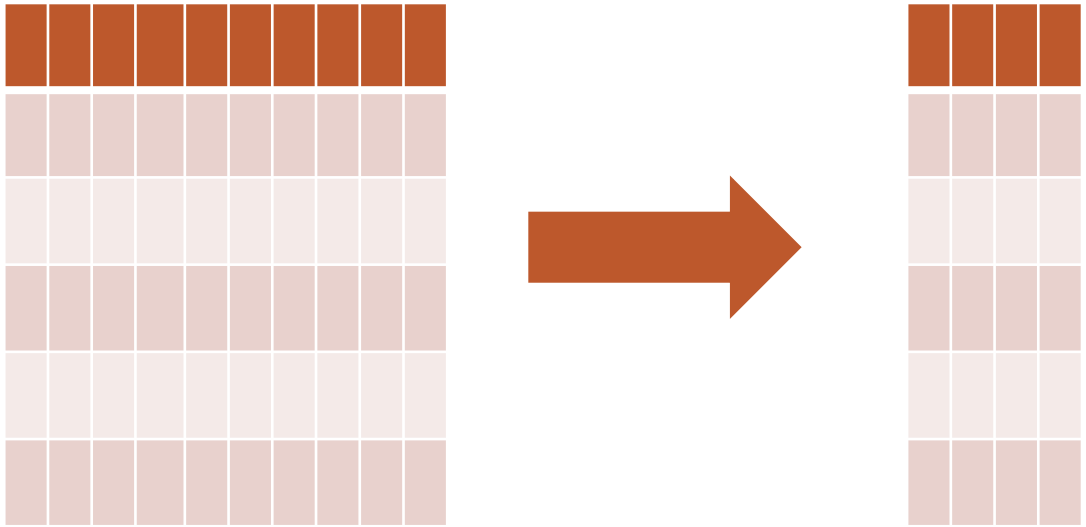
# Undersampling




K-means clustering

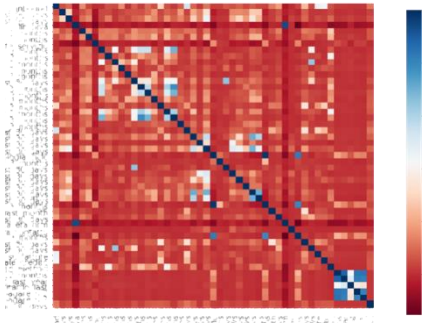


# Feature selection

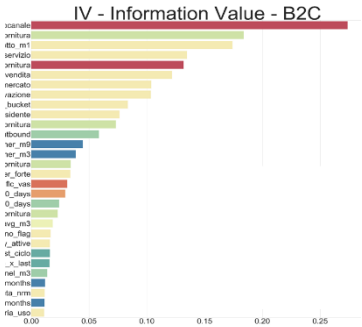


Reducing number of fields in the dataset

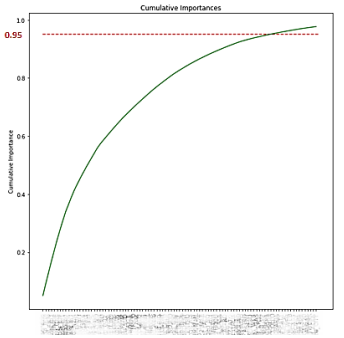
Correlation



WOE & IV




Cumulative importance



# Model selection

April

| Models | LASSO | RF | GBT |
|---|---|---|---|
| Lift | 21.8 | 32.53 | 24.2% |
| service5k | 7.8% | 15.9% | 13.26% |
| customer5k | 22.6% | 33.9% | 20.3% |

May

| Models | LASSO | RF | GBT |
|---|---|---|---|
| Lift | 30.16 | 38.6 | 24.4 |
| service5k | 9.3% | 14.2% | 8.7% |
| customer5k | 19.4% | 22.8% | 10% |

# Model selection

April

| Models | LASSO | RF | GBT |
|---|---|---|---|
| Lift | 21.8 | 32.53 | 24.2% |
| service5k | 7.8% | 15.9% | 13.26% |
| customer5k | 22.6% | 33.9% | 20.3% |

May

| Models | LASSO | RF | GBT |
|---|---|---|---|
| Lift | 30.16 | 38.6 | 24.4 |
| service5k | 9.3% | 14.2% | 8.7% |
| customer5k | 19.4% | 22.8% | 10% |

# Model selection

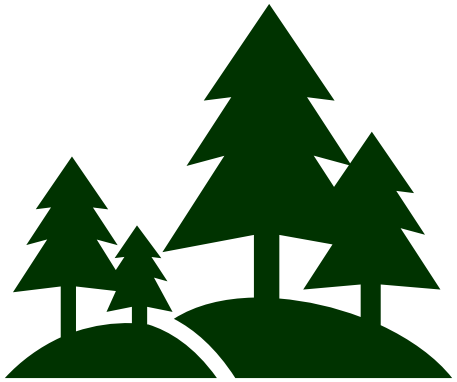
April

| Models | LASSO | RF | GBT |
|---|---|---|---|
| Lift | 21.8 | 32.53 | 24.2% |
| service5k | 7.8% | 15.9% | 13.26% |
| customer5k | 22.6% | 33.9% | 20.3% |

May

| Models | LASSO | RF | GBT |
|---|---|---|---|
| Lift | 30.16 | 38.6 | 24.4 |
| service5k | 9.3% | 14.2% | 8.7% |
| customer5k | 19.4% | 22.8% | 10% |

Random Forest model



# Model selection

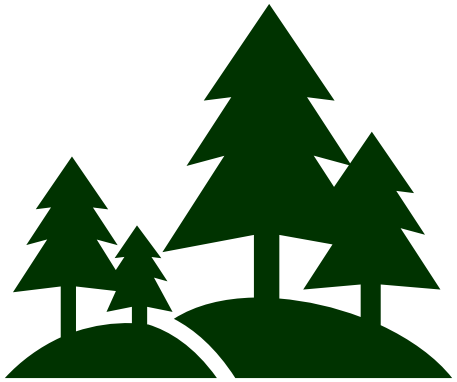
April

| Models | LASSO | RF | GBT |
|---|---|---|---|
| Lift | 21.8 | 32.53 | 24.2% |
| service5k | 7.8% | 15.9% | 13.26% |
| customer5k | 22.6% | 33.9% | 20.3% |

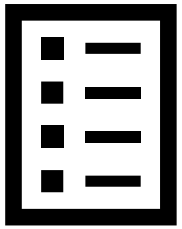
May

| Models | LASSO | RF | GBT |
|---|---|---|---|
| Lift | 30.16 | 38.6 | 24.4 |
| service5k | 9.3% | 14.2% | 8.7% |
| customer5k | 19.4% | 22.8% | 10% |

Random Forest model



Propensity of complain



# Model selection

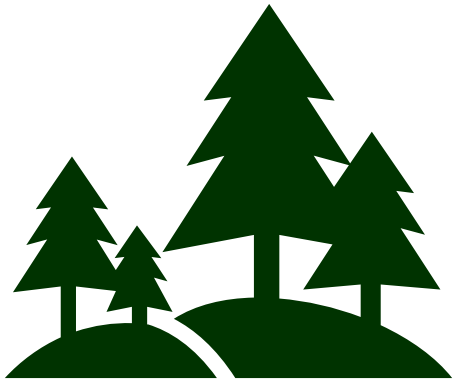
April

| Models | LASSO | RF | GBT |
|---|---|---|---|
| Lift | 21.8 | 32.53 | 24.2% |
| service5k | 7.8% | 15.9% | 13.26% |
| customer5k | 22.6% | 33.9% | 20.3% |

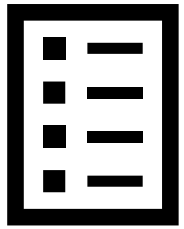
May

| Models | LASSO | RF | GBT |
|---|---|---|---|
| Lift | 30.16 | 38.6 | 24.4 |
| service5k | 9.3% | 14.2% | 8.7% |
| customer5k | 19.4% | 22.8% | 10% |

Random Forest model



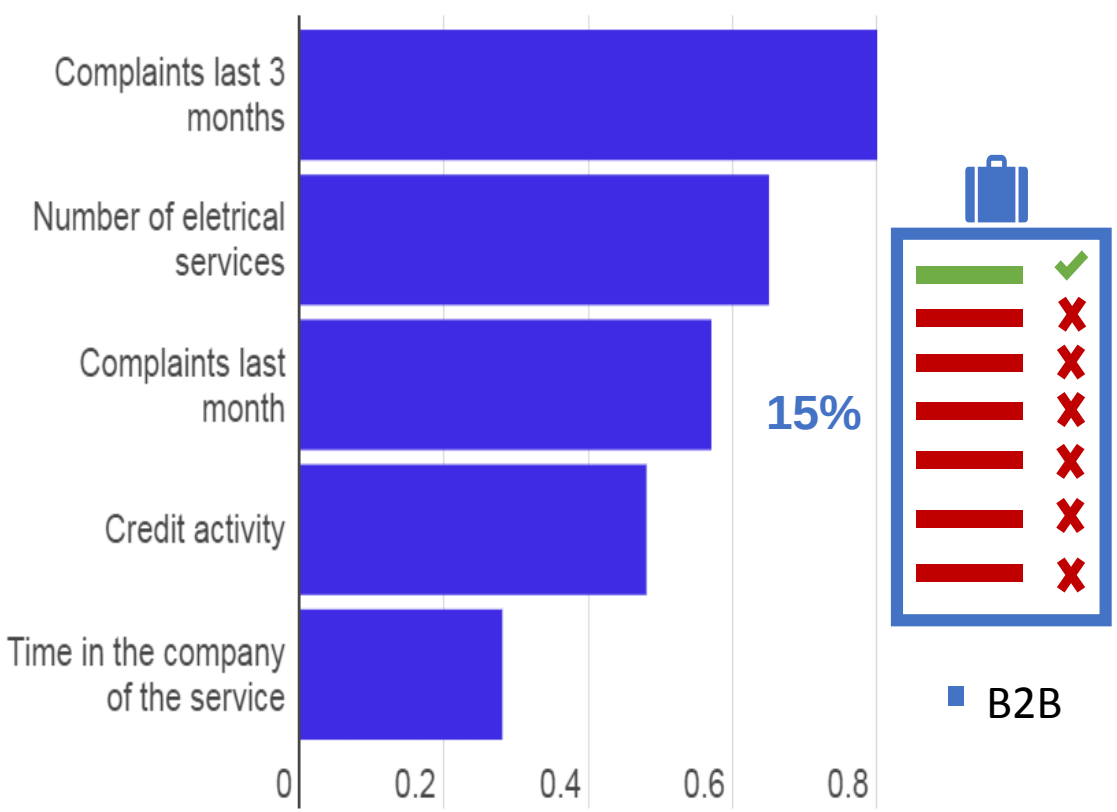
Propensity of complain



Understanding reasons

# Results

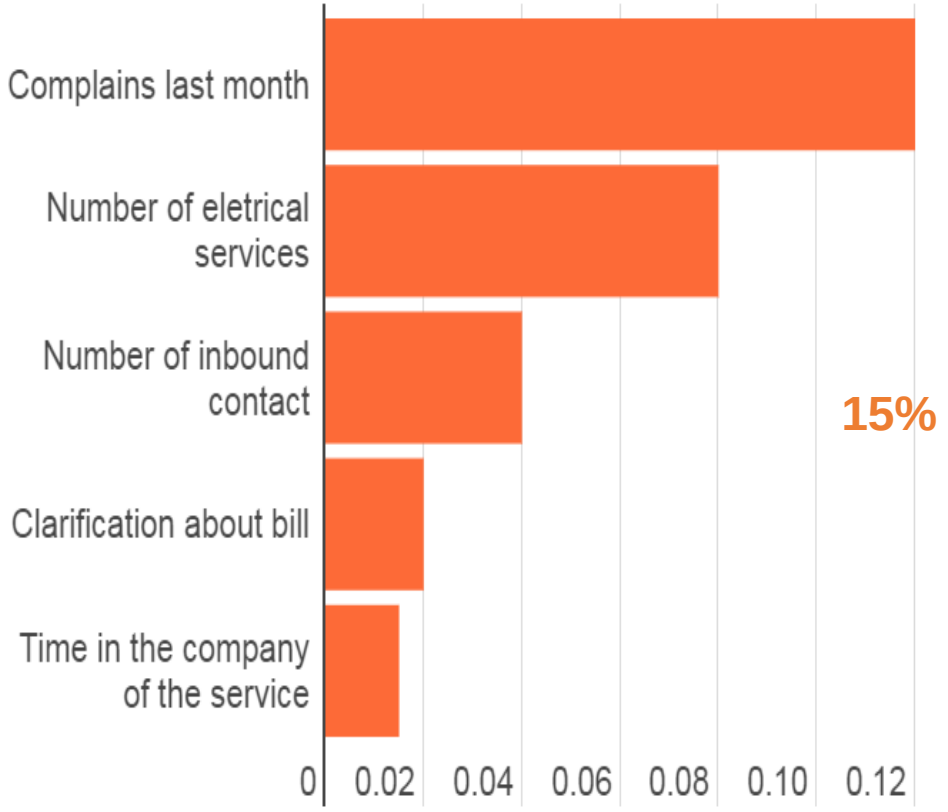
Example of May validation's result



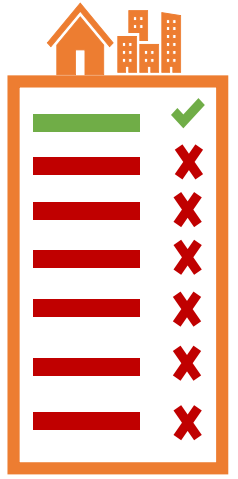

15%



B2B



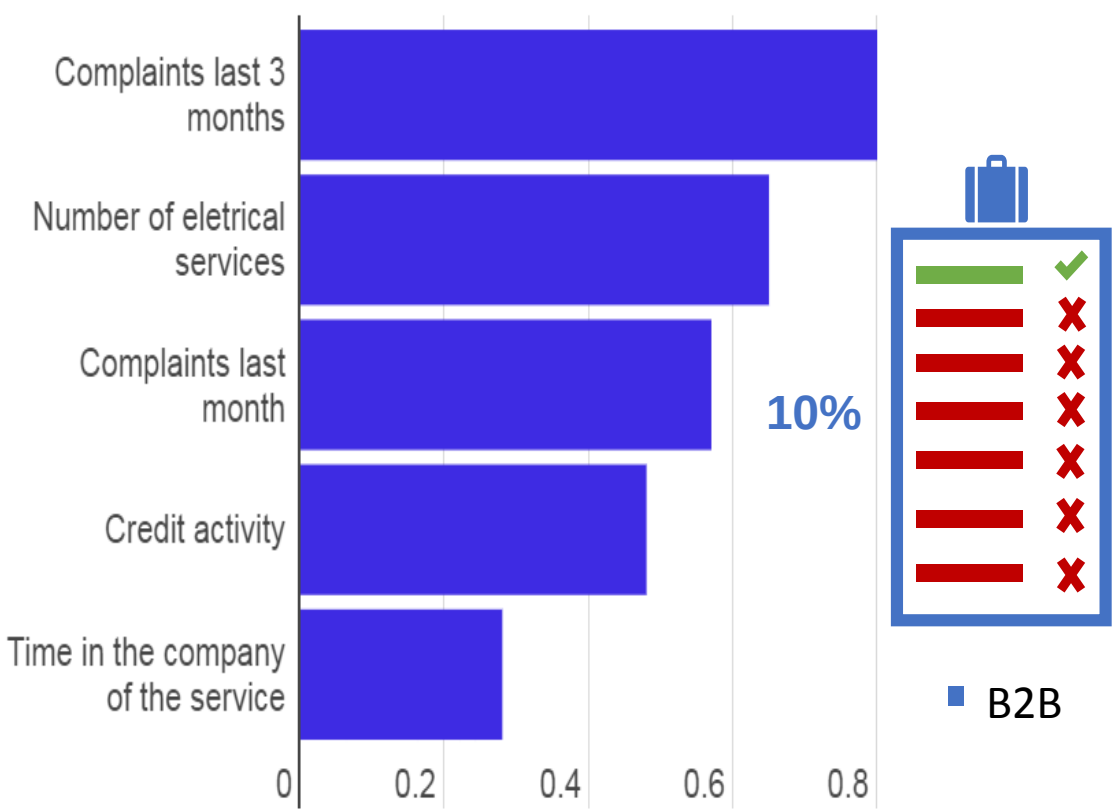

15%



B2C

# Results

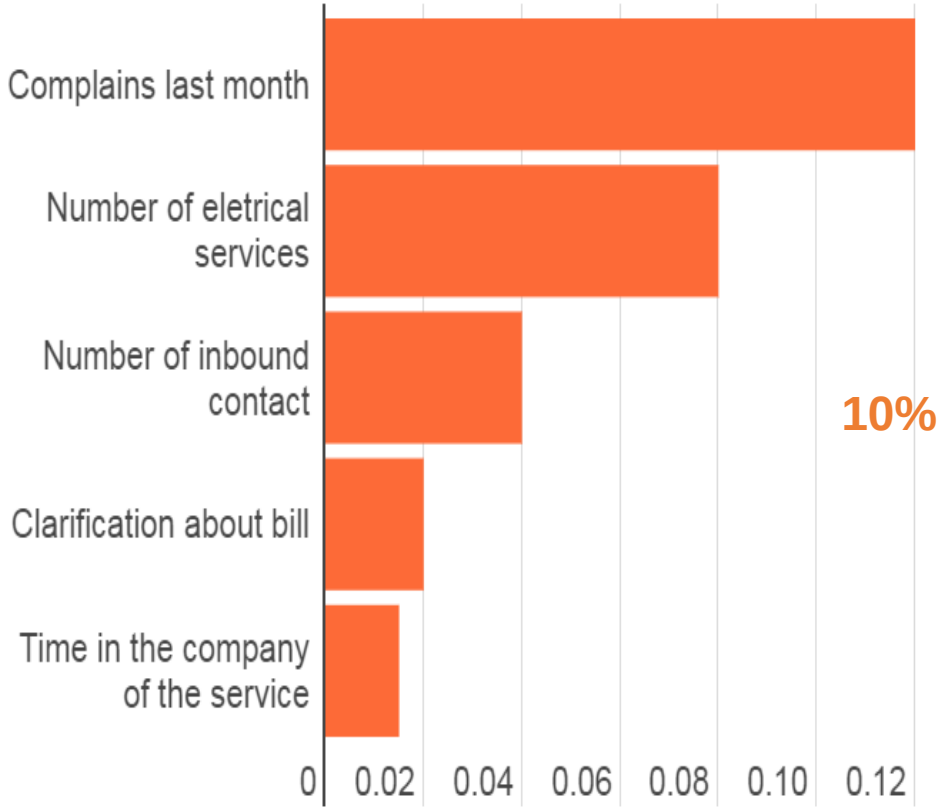
Example of June validation's result



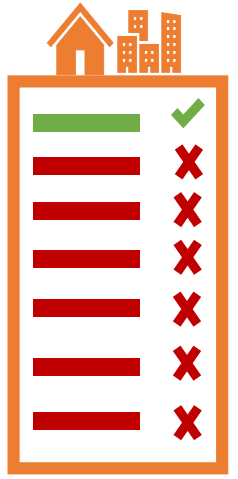

10%



B2B

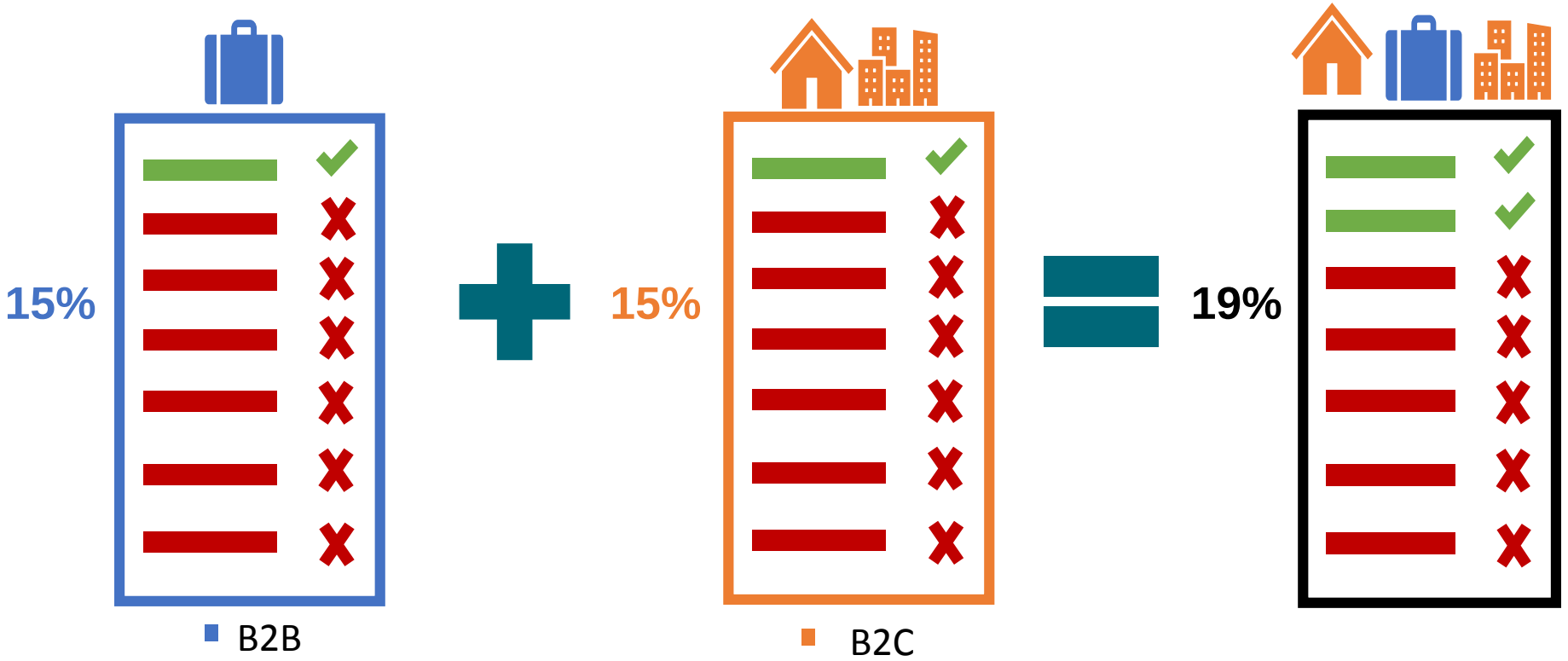


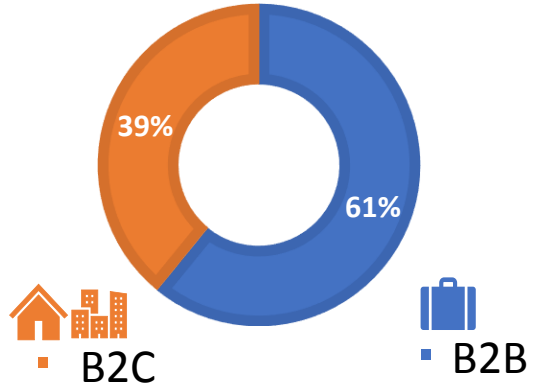

10%



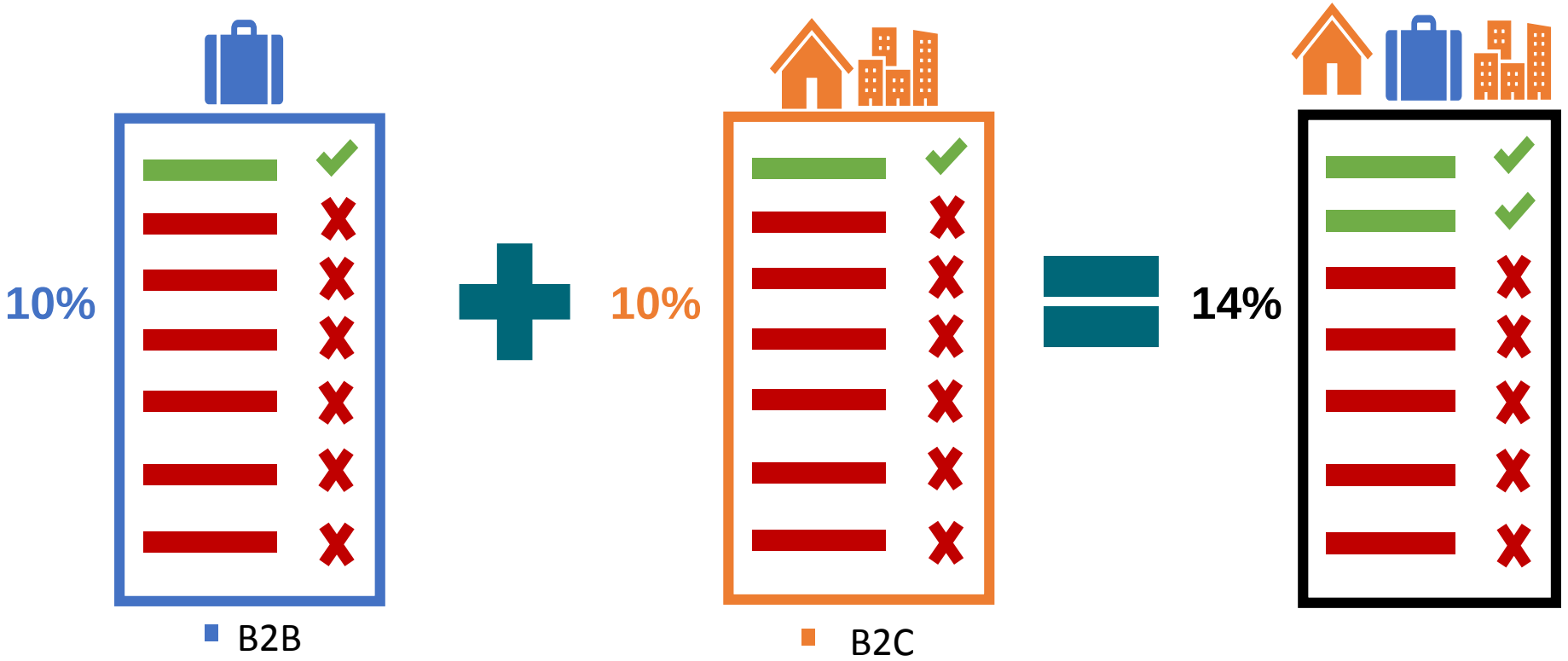
B2C

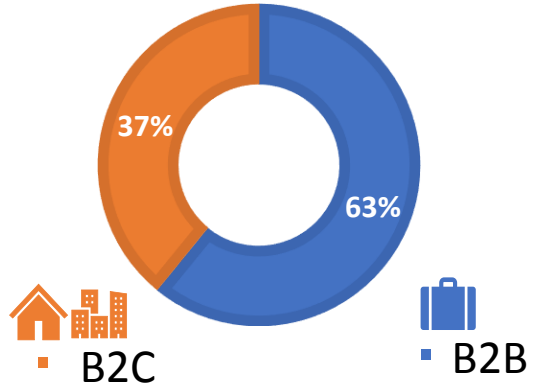
# Results: Combined May



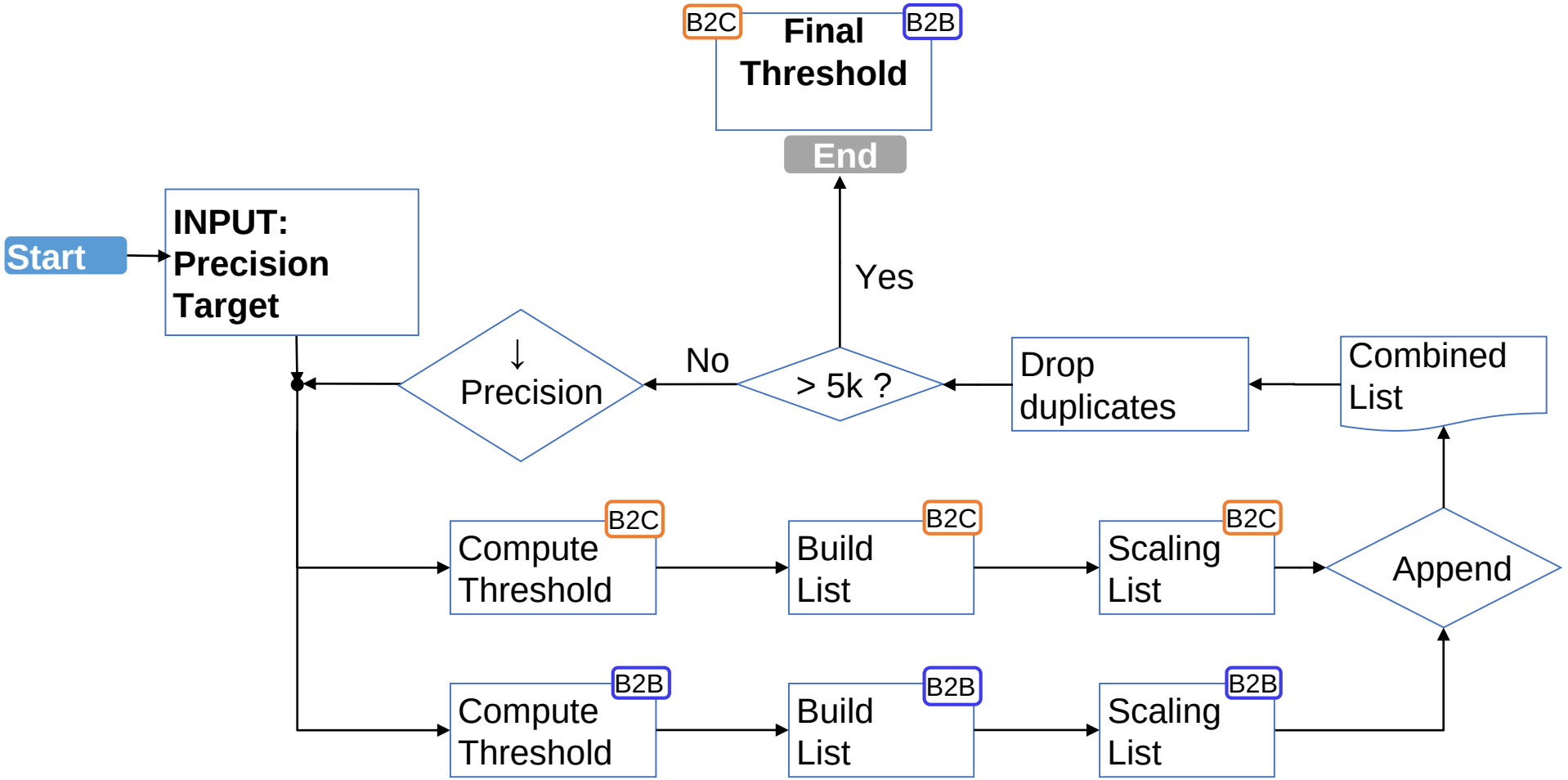
# Results: Combined June



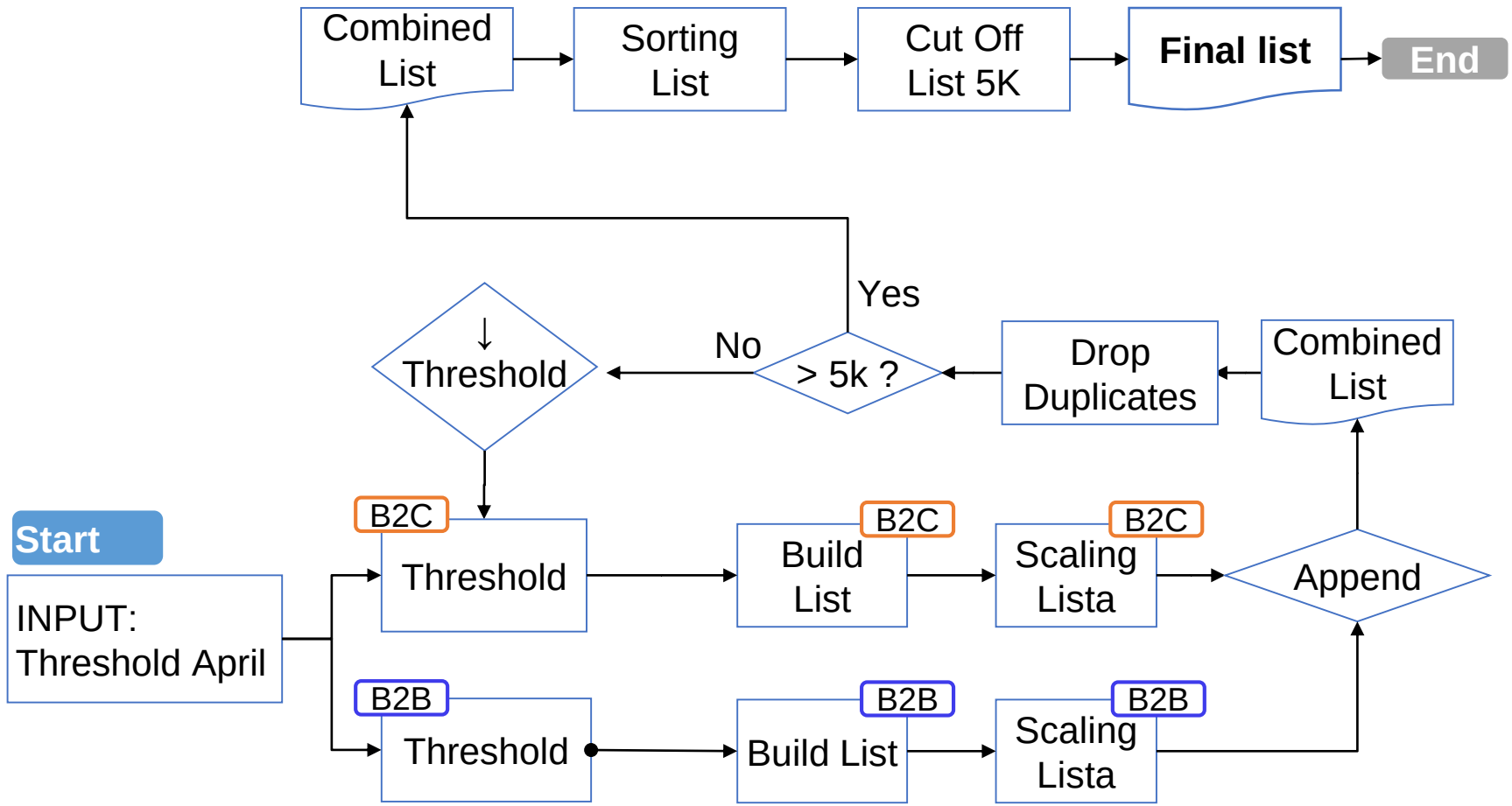
# Results: strategy to combine results

Process to find best Threshold for B2C & B2B applied to April



# Results: strategy to combine results

Definition of combined list applied to May and June



# Conclusions

- The model was deployed in September and gained 12% of precision

# Conclusions

- The model was deployed in September and gained 12% of precision

- 11% of benefits in term of costs

# Conclusions

- The model was deployed in September and gained 12% of precision

- 11% of benefits in term of costs

- Developing of a model to predict reasons of claims

SAPIENZA
UNIVERSITÀ DI ROMA

# Thank you
# for your attention

**Predicting costumer coplaints in an energy company**

Laura Giuliano