Eltecon Data Science Course by Emarsys

Measuring effect through experimenting

András Bérczi

October 7, 2020

About me

- Background in Economics
- Works as Data Scientist @ Emarsys

Homeworks from last week

Why talk about effect measurement in Data Science?

What is an effect?

Added value of a treatment (the new feature)

Why do we want to measure the effect?

To decide if the new feature works

What to experiment on?

How to decide?

- **Based on customer's need**
- Validated by data based research
  - Will the algo work?
  - Does it scale?
  - Cost of the feature?
- Make sure you understand your feature/algorithm!

How can we measure the effect?

- Simulation
- Based on historical data
- **Experimenting** (A/B test or something like that. . .)

How do we experiment?

Setup

- Define the goal
  - Measure one feature at a time
- Definte KPIs
  - How much added value does it bring to the user?
- Make sure you have enough data to measure significant results
  - check the minimum effect you can measure beforehand!
- Split contacts *randomly* into control and treatment groups
- Do not change parameters during the experiment

What are good KPIs?

Treatment receives personalized email (80% of contacs), control receives "standard" email (20% of contacs).

id	group	did_open	did_click	sales_amount
1	control	0	0	0
2	treatment	0	0	0
3	control	1	1	30
4	treatment	0	0	0
5	control	1	0	50
6	treatment	1	1	12
...	...	...	...	...
100	treatment	1	0	0

How to calculate the effect

$$Uplift_{KPI} = \frac{KPI_{treatment}}{KPI_{control}} - 1$$

Calculate the effect for the first period!

Use `experiment_results.csv`

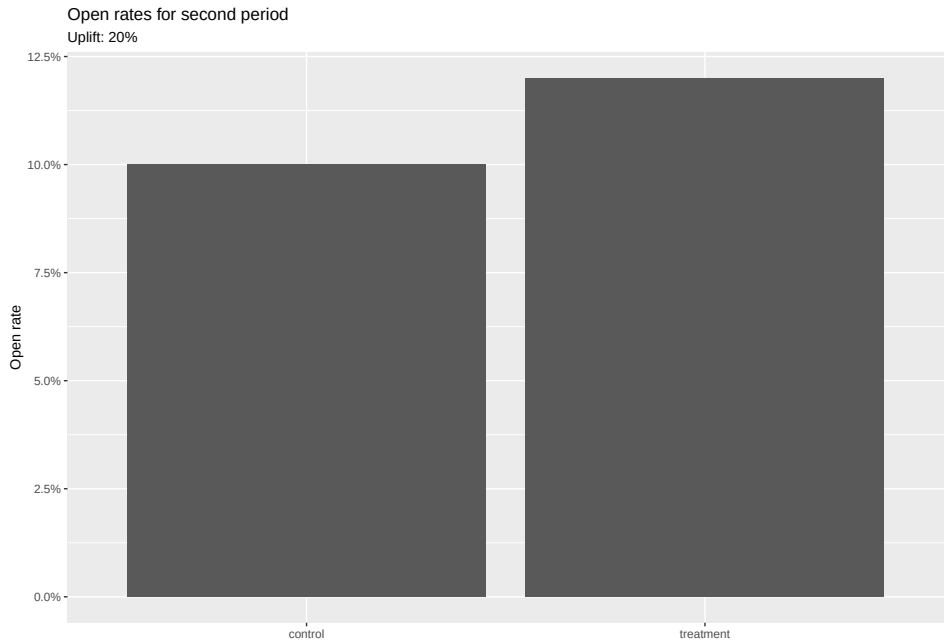
Enter your results in Socrative!

Calculate the effect for the second period and plot the results!

Use `experiment_results.csv`

When you are done with both the calculation and the plot, enter your results in Socrative!

The *design* should look something like this:



Calculate the effect for the whole period and plot the results!

When you are done with both the calculation and the plot, enter your results in Socrative!

Do you notice anything weird?

One way to aggregate the effect

$$KPI_{treatment} = \frac{\sum_{i=1}^{n} KPI_{treatment,i} * SampleSize_i}{\sum_{i=1}^{n} SampleSize_i}$$

$$KPI_{control} = \frac{\sum_{i=1}^{n} KPI_{control,i} * SampleSize_i}{\sum_{i=1}^{n} SampleSize_i}$$

$$Uplift_{KPI} = \frac{KPI_{treatment}}{KPI_{control}} - 1$$

, where n is the number of unit levels.

Calculate the effect for the whole period and plot the results!

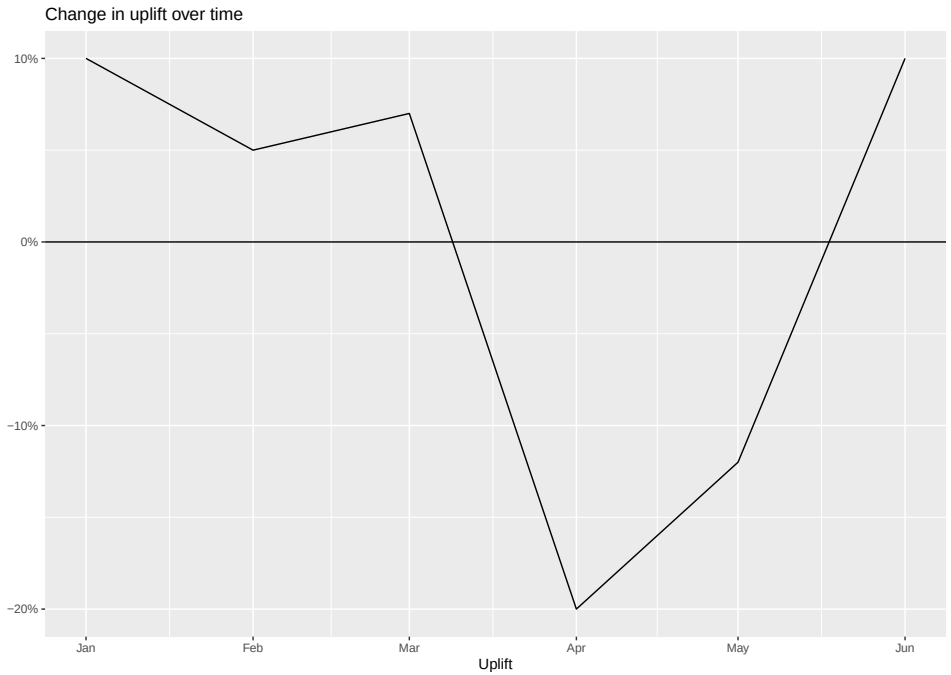
Weight the open rates by the number of contacts present in the period!

When you are done with both the calculation and the plot, enter your results in Socrative!

Plot the effect (uplift) over time!

Use `experiment_results_over_time.csv`

The *design* should look something like this:

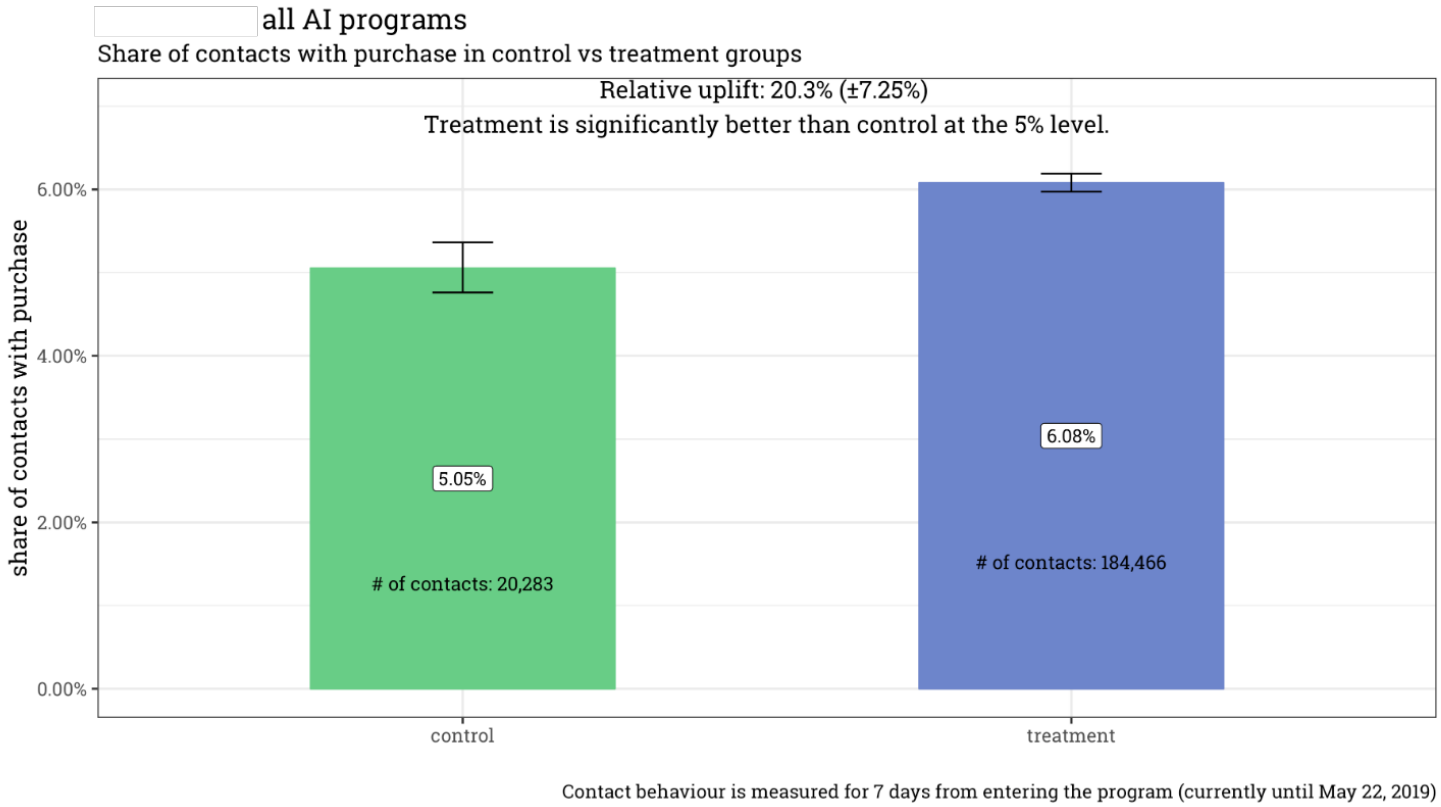


How to present

How to present

What is missing from the previous plots?

How we do it in Emarsys



How we do it in Emarsys

Shiny app from Emarsys

Homework for next week and presenters