

# Value creation with data

Clément Levallois

2017-09-10

# Table of Contents

Seven roads to data-driven value creation. ....	1
1. PREDICT .....	1
2. SUGGEST .....	2
3. CURATE .....	2
4. ENRICH .....	3
5. RANK / MATCH / COMPARE .....	4
6. SEGMENT / CLASSIFY .....	5
7. GENERATE (experimental!) .....	6
Combos! .....	7



## Seven roads to data-driven value creation



Not a closed list, not a recipe!

Rather, these are essential building blocks for a strategy of value creation based on data.

### 1. PREDICT



Table 1. Predict: The ones doing it

1. Predictive churn / default / ... (banks / telco)

2. Predicting crime



3. Predicting deals



4. Predictive maintenance






Table 2. Predict: The hard part

1. Collecting data ("cold start problem")
2. Risk missing the long tail, algorithmic discrimination, stereotyping
3. Neglect of novelty

## 2. SUGGEST



*Table 3. Suggest: The ones doing it*

1. Amazon's product recommendation system 
2. Google's "Related searches..." 
3. Retailer's personalized recommendations 

*Table 4. Suggest: The hard part*

1. Managing serendipity and bubble effects
2. Finding the value proposition which goes beyond the simple "you purchased this, you'll like that"

## 3. CURATE



Table 5. Curate: the ones doing it

1. Clarivate Analytics curating metadata from scientific publishing



2. Nielsen and IRI curating and selling retail data



3. IMDb curating and selling movie data



Table 6. Curate: the hard part

1. Slow progress
2. Must maintain continuity
3. Scaling up / right incentives for the workforce
4. Quality control

## 4. ENRICH



*Table 7. Enrich: the ones doing it*

1. Selling methods and tools to enrich datasets	
2. Selling aggregated indicators	
3. Selling credit scores	

*Table 8. Enrich: the hard part*

1. Knowing which cocktail of data is valued by the market
2. Limit replicability
3. Establish legitimacy

## 5. RANK / MATCH / COMPARE



Table 9. Rank / Match / Compare: the ones doing it

- |  |   |
|--|---|
| 1. Search engines ranking results  |    |
| 2. Yelp, Tripadvisor, etc... which rank places   |  |
| 3. Any system that needs to filter out best quality entities among a crowd of candidates |   |

Table 10. Rank / Match / Compare: the hard part

- |  |
|--|
| 1. Finding emergent, implicit attributes   |
| 2. Insuring consistency of the ranking     |
| 3. Avoid gaming of the system by the users |

## 6. SEGMENT / CLASSIFY



## Chihuahua or Muffin?



Table 11. Segment / classify: the ones doing it


1. Tools for discovery / exploratory analysis by segmentation
2. Diagnostic tools (spam or not? buy, hold or sell? healthy or not?)  medimsight

Table 12. Segment / classify: the hard part

1. Evaluating the quality of the comparison
2. Dealing with boundary cases
3. Choosing between a supervised and unsupervised approach (how many categories?)

## 7. GENERATE (experimental!)



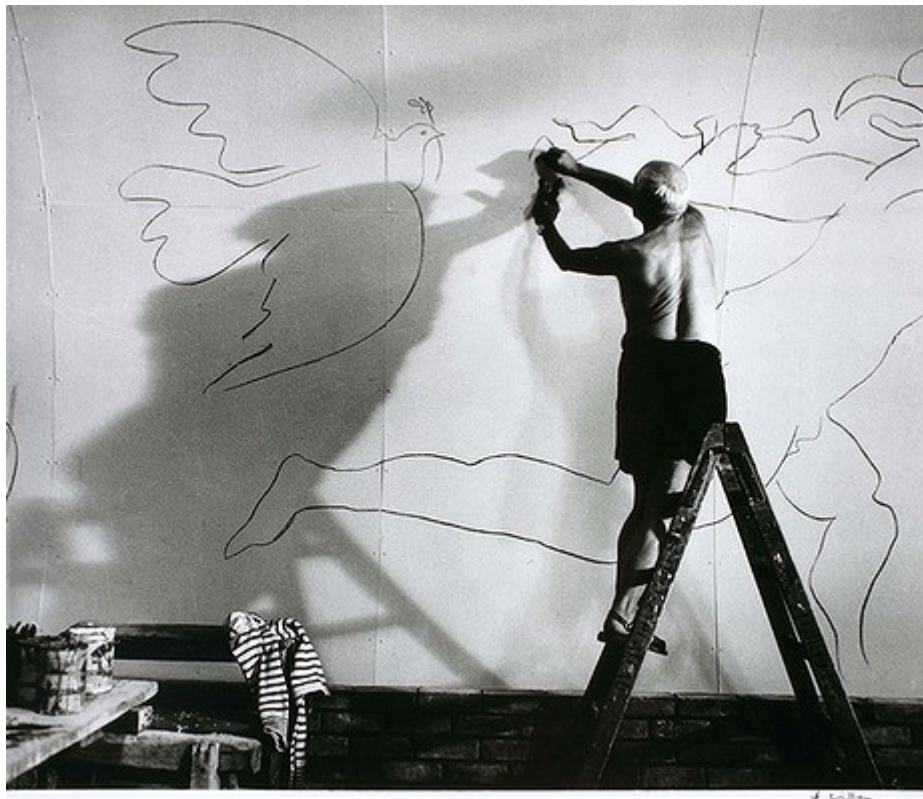



Table 13. Generate (experimental!): the ones doing it

1. Intelligent BI 
2. wit.ai, the chatbot by FB 
3. Virtual assistants company 
4. Image generation 

Table 14. Generate (experimental!): the hard part

1. Should not create a failed product / false expectations
2. Both classic (think of ) and frontier science: not sure where it's going

## Combos!



Figure 1. Combinations

- [levallois@em-lyon.com](mailto:levallois@em-lyon.com)
- get in touch via Twitter: [@seinecle](https://twitter.com/seinecle)

