

# 7 roads to data-driven value creation

Clément Levallois

2017-09-10

# Table of Contents

|   |   |
|---|---|
| 7 roads to data-driven value creation .....       | 1 |
| 1. PREDICT .....                                  | 1 |
| Prediction: The ones doing it                     | 1 |
| Prediction: the hard part                         | 2 |
| 2. SUGGEST .....                                  | 2 |
| Suggestion: The ones doing it                     | 2 |
| Suggestion: the hard part                         | 2 |
| 3. CURATE .....                                   | 2 |
| Curation: The ones doing it                       | 3 |
| Curation: the hard part                           | 3 |
| 4. ENRICH .....                                   | 3 |
| Enrichment: The ones doing it                     | 4 |
| Enrichment: the hard part                         | 4 |
| 5. RANK / MATCH / COMPARE .....                   | 4 |
| Ranking / matching / comparing: The ones doing it | 5 |
| Ranking / matching / comparing: the hard part     | 5 |
| 6. SEGMENT / CLASSIFY .....                       | 5 |
| Generating: The ones doing it                     | 6 |
| Segmenting / classifying: the hard part           | 6 |
| 7. GENERATE / SYNTHETIZE(experimental!) .....     | 7 |
| Generating: The ones doing it                     | 7 |
| Generating: the hard part                         | 8 |
| Combos! .....                                     | 8 |
| The end .....                                     | 8 |



## 7 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




#### Prediction: The ones doing it

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

2. Predicting crime  **PredPol**  
Predict Crime in Real Time™

3. Predicting deals  **TILKÆE**

4. Predictive maintenance  **CAT**




## Prediction: the hard part

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

## 2. SUGGEST



### Suggestion: The ones doing it

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

### Suggestion: the hard part

1. The [cold start problem](#), managing serendipity (see review: [paying version](#), free version not available) and "filter bubble" effects (review: [paying version](#), [free version here](#)).
2. Finding the value proposition which goes beyond the simple "you purchased this, you'll like that"

## 3. CURATE



## Curation: 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



## Curation: the hard part

1. Slow progress: curation needs human labor to insure high accuracy, it does not scale the way a computerized process would.
2. Must maintain continuity: missing a single year or month hurts the value of the overall dataset disproportionately.
3. Scaling up / right incentives for the workforce: the workforce doing the curation should be paid fairly, which is [not the case yet](#).
4. Quality control

## 4. ENRICH



## Enrichment: The ones doing it

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



## Enrichment: the hard part

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

# 5. RANK / MATCH / COMPARE



## Ranking / matching / comparing: 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

## Ranking / matching / comparing: the hard part

1. Finding emergent, implicit attributes (imagine: if you rank things based on just one public feature: not interesting nor valuable)
2. Insuring consistency of the ranking (many rankings are less straightforward than they appear)
3. Avoid gaming of the system by the users (for instance, [companies try to play Google's ranking of search results at their advantage](#))

# 6. SEGMENT / CLASSIFY

## Chihuahua or Muffin?



### Generating: 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  
Medical Imaging Cloud Platform

### Segmenting / classifying: the hard part

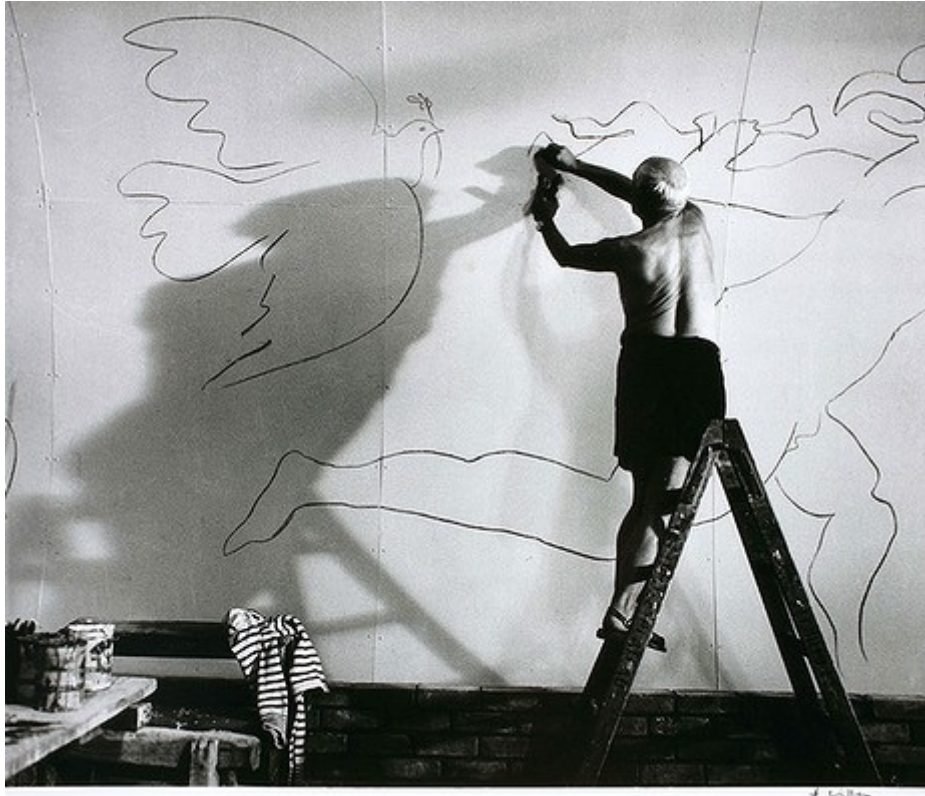
1. Evaluating the quality of the comparison

2. Dealing with boundary cases

3. Choosing between a pre-determined number of segments (like in the k-means) or letting the number of segments emerge



## 7. GENERATE / SYNTHETIZE(experimental!)



### Generating: The ones doing it

(click on the logos to get to the relevant web page)

1. Intelligent BI <a href="https://www.aiden.ai/">" tmp="false">]

2. wit.ai, the chatbot by FB <a href="https://wit.ai/">" tmp="false">]

3. Virtual assistants <a href="https://www.cxcompany.com/digitalcx/">" tmp="false">]

4. Image generation <a href="https://deepart.io/">" tmp="false">]

5. Close-to-real-life speech synthesis <a href="https://deepmind.com/blog/wavenet-generative-model-raw-audio/">" tmp="false">]

## Generating: 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

## The end

Find references for this lesson, and other lessons, [here](#).



This course is made by Clement Levallois.

Discover my other courses in data / tech for business: <http://www.clementlevallois.net>

Or get in touch via Twitter: [@seinecle](#)