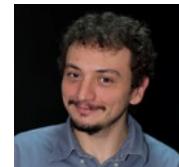




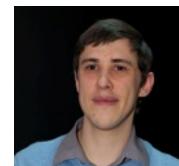
Dataiku

- ▶ Founded in 2013 by Data and Search Engine veterans



- ▶ From “data” and “haïku”

*“data can be big
solution would be small
feel the hot wind”*



- ▶ Data Science for Everyone

(big) data(s)

+ machine learning

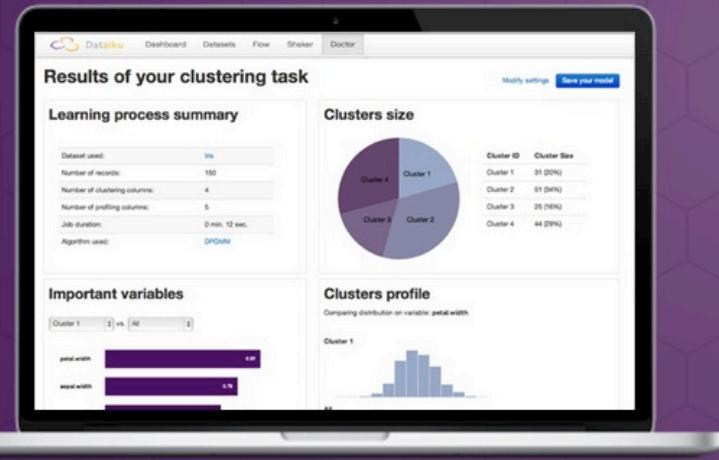
+ for practical applications

= Data Science

We build a Studio for Data Science

DATA SCIENCE FOR EVERYONE

The Data Science Studio makes the promise of big data a reality.



Our Data Science Studio is the only all-in-one platform designed specifically for building data science applications.

Curious business analyst?
Get started quickly with predictions that scale.

Data science team?
Share recipes and insights with the team, share visualizations with stakeholders.

Freelance Data Scientist?
Boost productivity with visual tools and pre-built transformations.

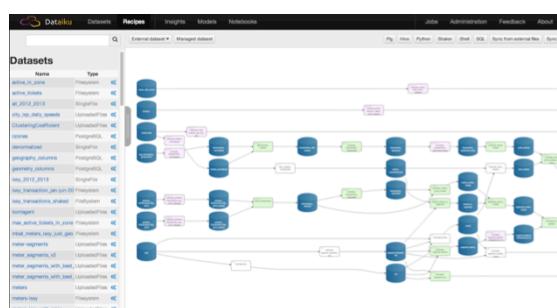
[Learn more...](#)

Data Science Studio



This screenshot shows the Dataiku Data Preparation interface. At the top, there's a navigation bar with tabs like Datasets, Recipes, Insights, Models, Notebooks, and a search bar. Below the navigation is a toolbar with various icons for operations like 'Script', 'Rename column', 'Delete', 'Extract date components', 'Flag French holidays', 'Remove rows where IPAddress is not a valid IP', 'Geo-locate IP in IPAddress', 'Split URL in URL', and 'Split sep. & on...'. The main area is a flow editor where data transformations are connected in a sequence. A preview table at the bottom shows the results of the transformation steps.

This screenshot shows the Dataiku Recipe app template selection screen. It features a grid of six boxes, each representing a different type of analysis or machine learning task: Exploration, Clustering, Profiling, Prediction, Recommendation, and Forecasting. Each box contains a brief description and a 'Select' button.



Visual and Interactive
Data Preparation
For Data Cleaners

Guided Machine
Learning
For non Machine Learning Experts

Production ready
For Data Leak Fixers

They use it



National Yellow Pages company
Using machine learning to optimize
the perceived relevance for 15 M
requests



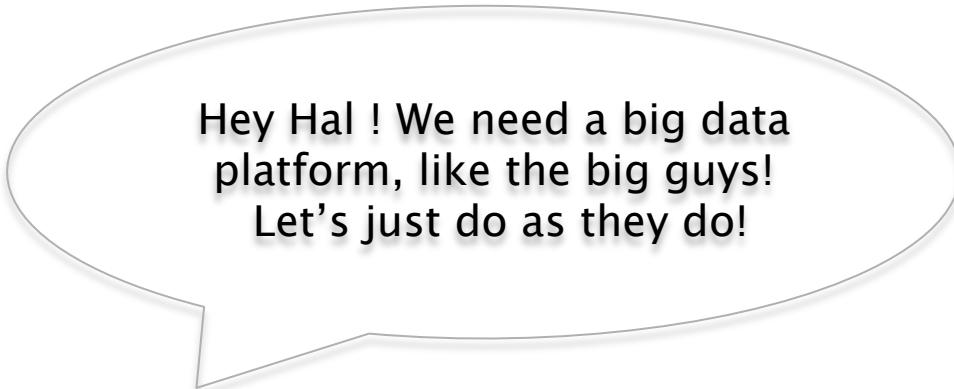
French E-commerce Leader
Rebuild BI Workflow, Increase E-mail
retargeting span



Global player in the urban
mobility sector, in Parking and
Transport Confidential, but around
Parking...

And many
others...

How it started?



Hey Hal ! We need a big data platform, like the big guys!
Let's just do as they do!



Dim Sum

CEO & Founder
Dim's Private Showroom
Medium size e-commerce
• 100M\$ revenue
• 1 Data Analyst



Hal Alowne

BI Manager
Dim's Private Showroom



Big Guys

\$10B + revenue
100+ Data Scientists

Hal's challenge



X00. 000 €/\$ *

6 Months

2 Terabytes

GENERATE

20%

UPLIFT

* (Actual Numbers may vary)

Should I Hire a Data **SCIENTIST** ?



This is not... a Data Scientist



is it a
MACHINE
LEARNING
EXPERT
?

DATA CLEANER



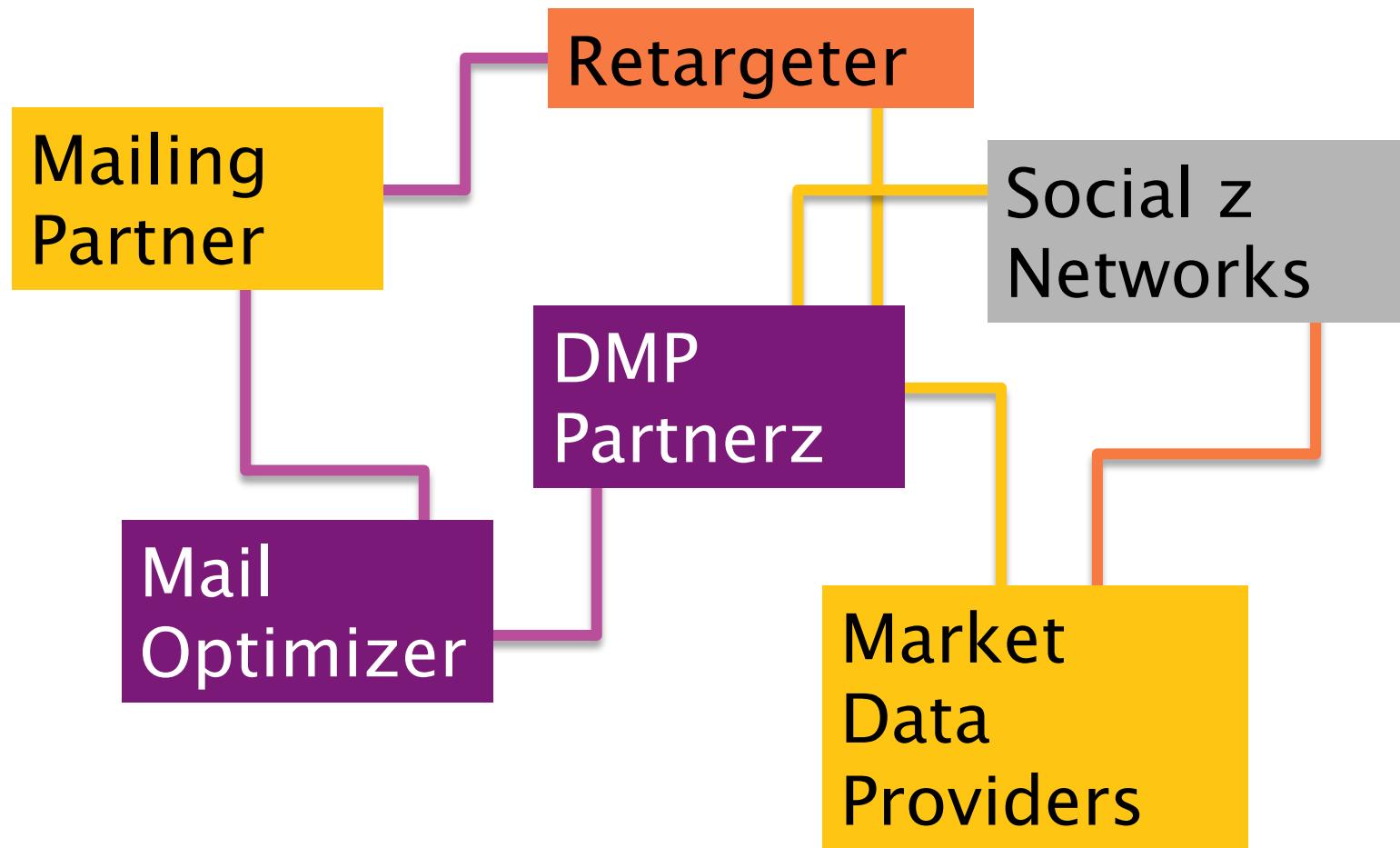
DATA LEAK FIXER



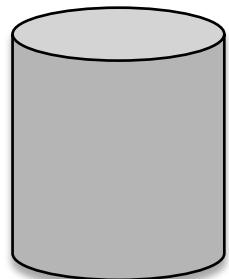


END OF
(HADOOP) JOB
DATA
WAITER

Partner Data Spaghetti

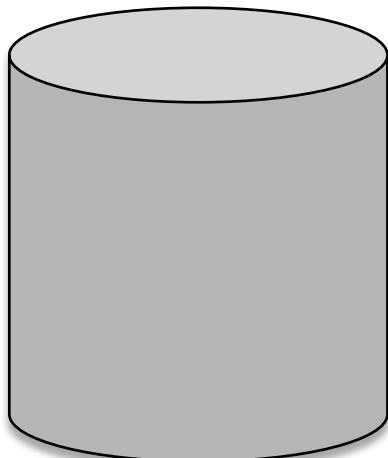


Database are Full



1 TB
BI Database

Any new computing job take
 > 1 day



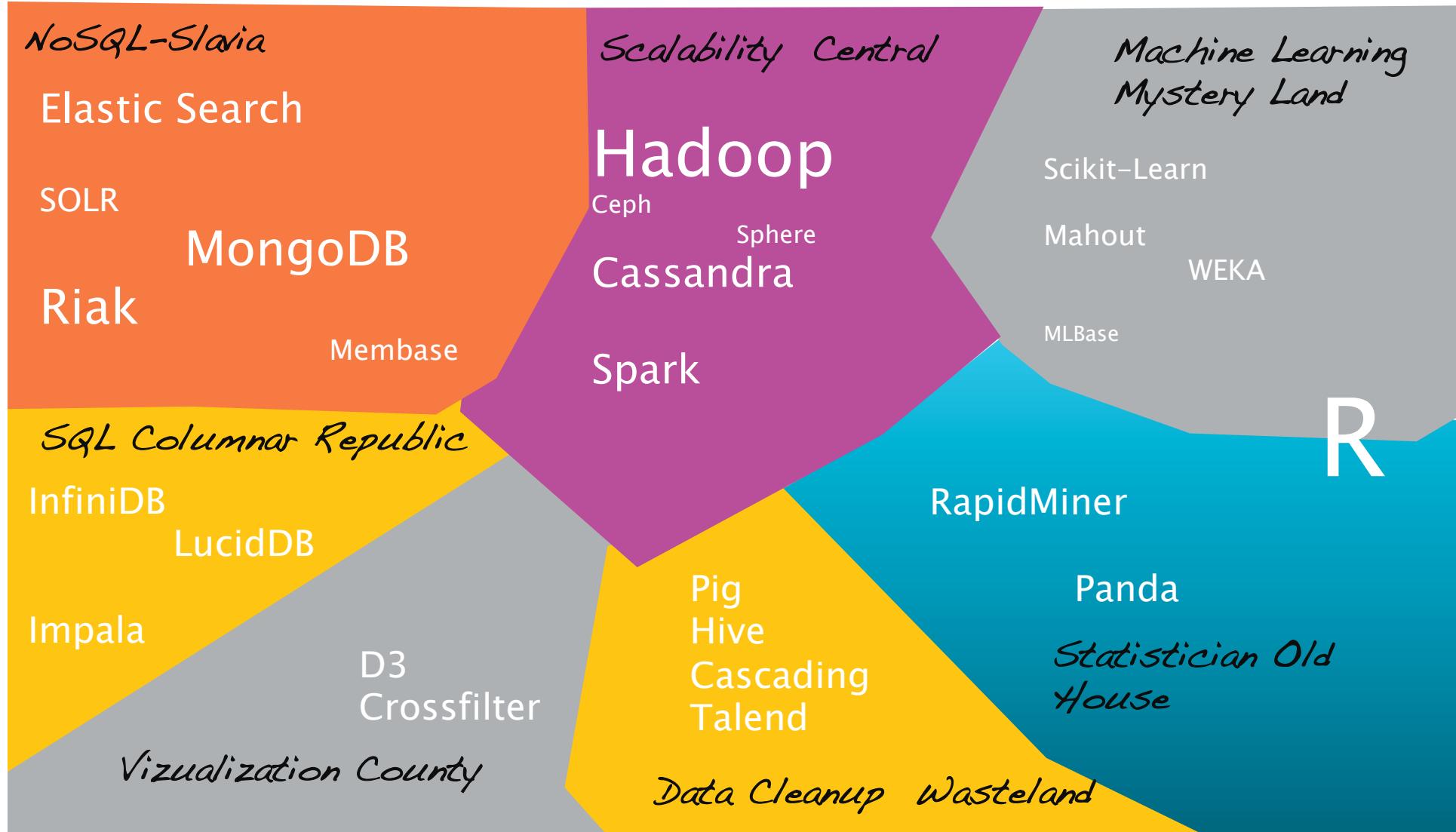
20 TB
BI Database

NEED FOR SCALE

Architecture Bingo

	Batch	BI Real-Time	Real Time	Hive	Cascading
Simple Queries				Pig	
Statistics					
Machine Learning				Spark	R
				MongoDB	ElasticSearch

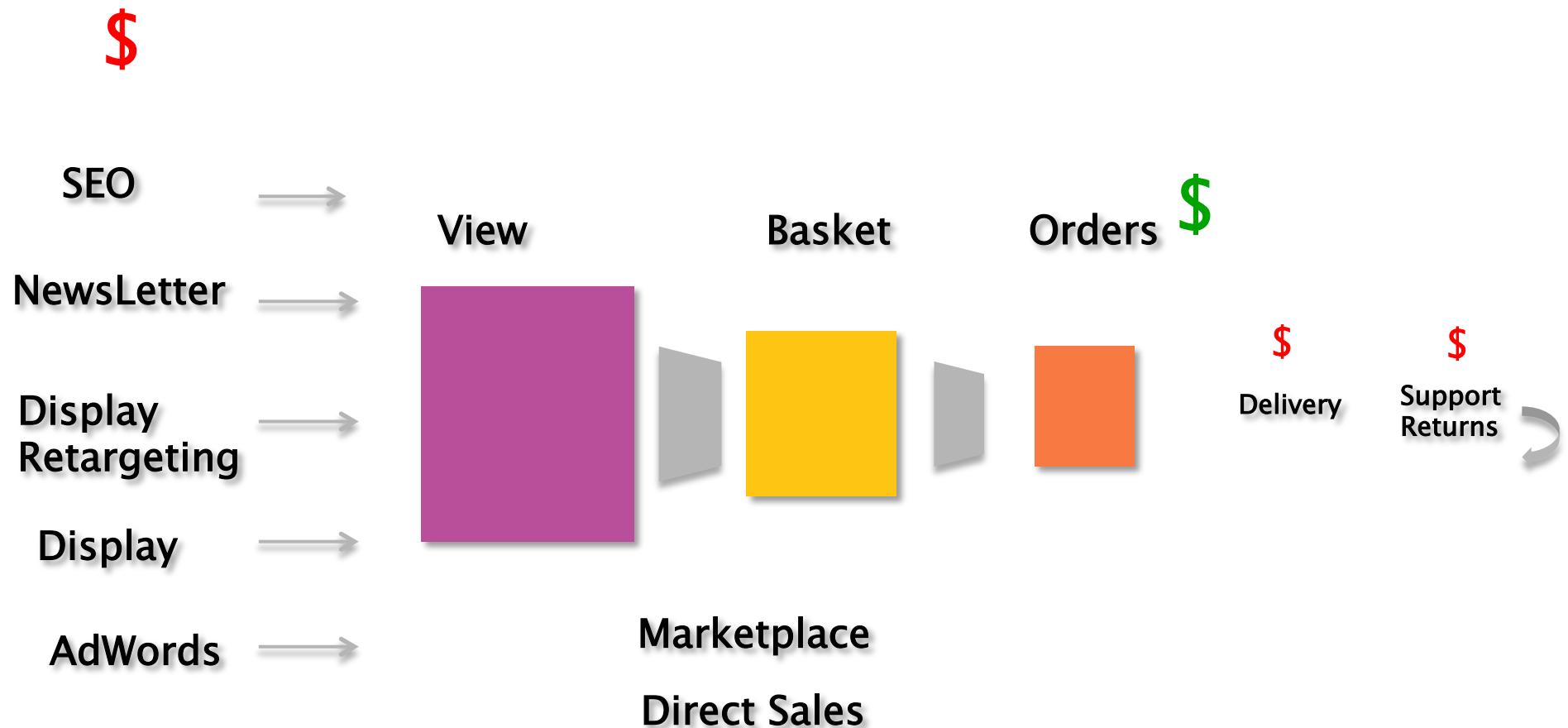
10/6/13



EXAMPLES

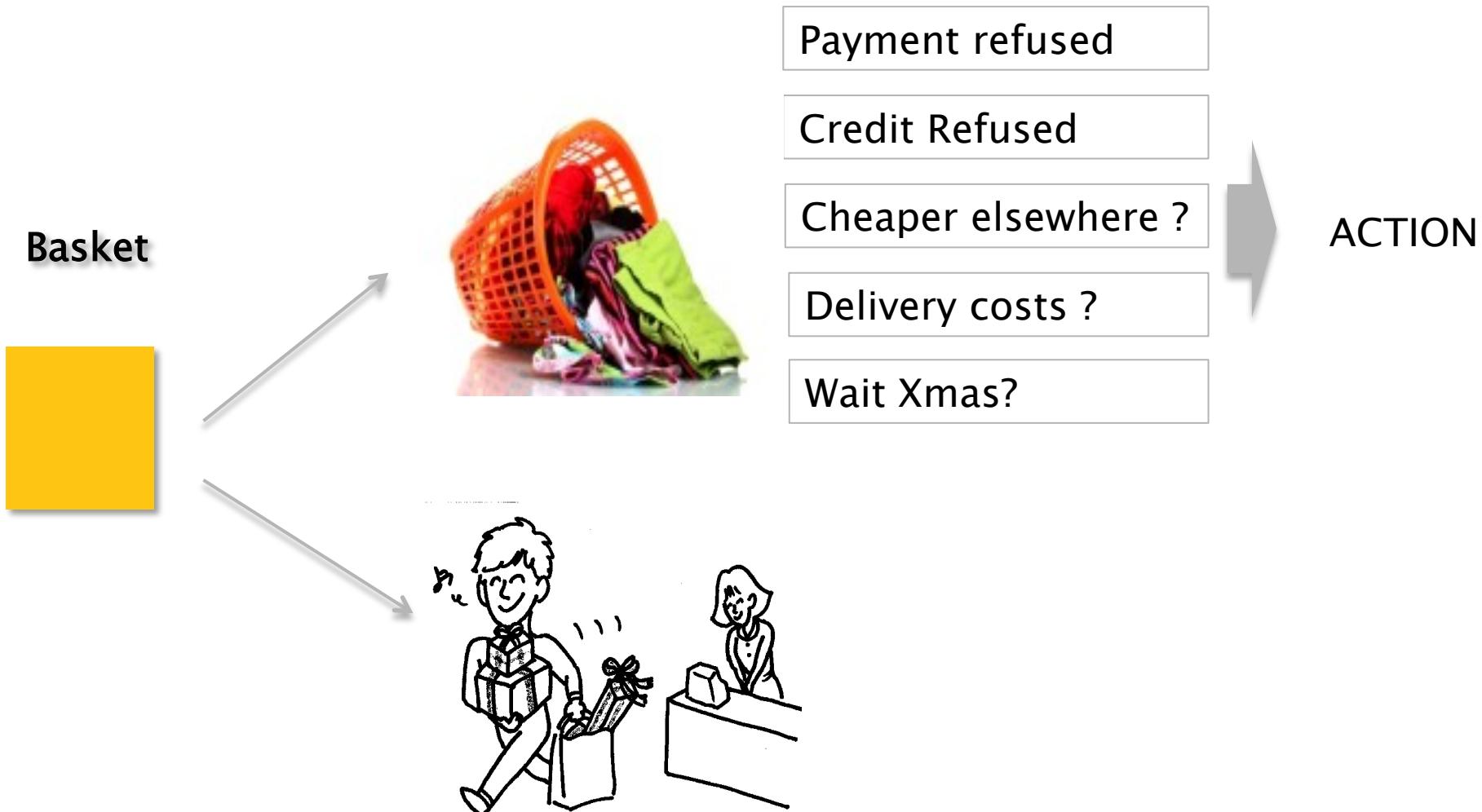
Hal Wish #1

Global Customer Value Funnel



Hal Wish #2

Why people drop basket ?



Hal Wish #3

What product to put on top ?

Original ★

Most Popular on top



Better ★★

Machine Learning Score
(age/discount/margin...)



Advanced ★★★

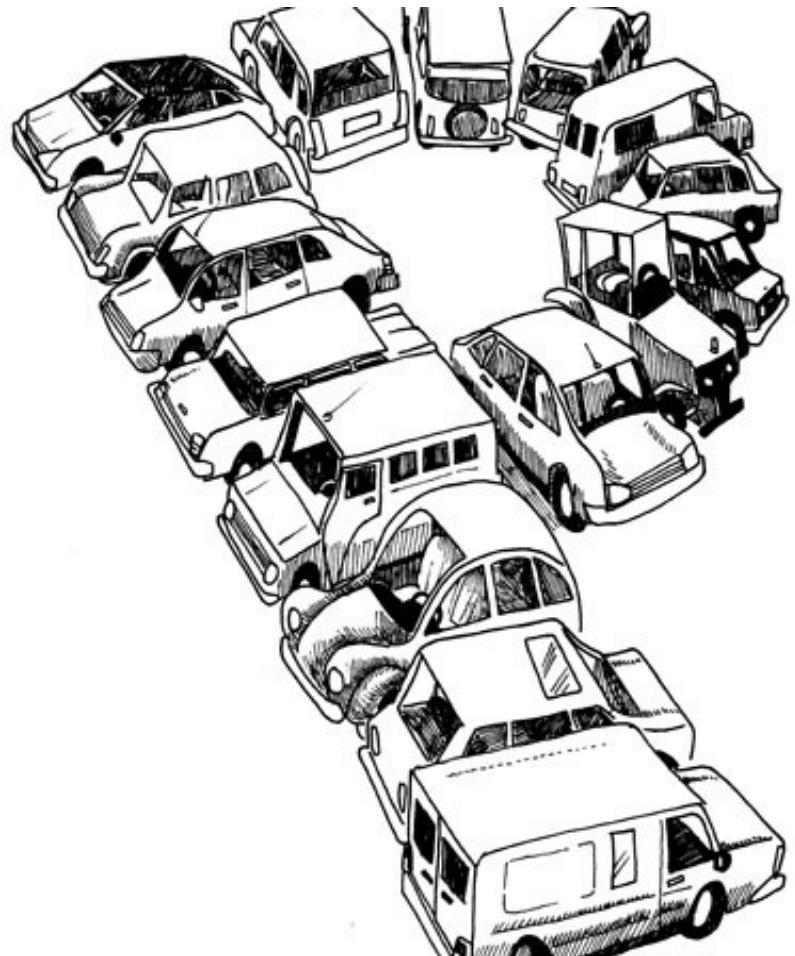
Machine Learning Score
+ Personalization



10/6/13

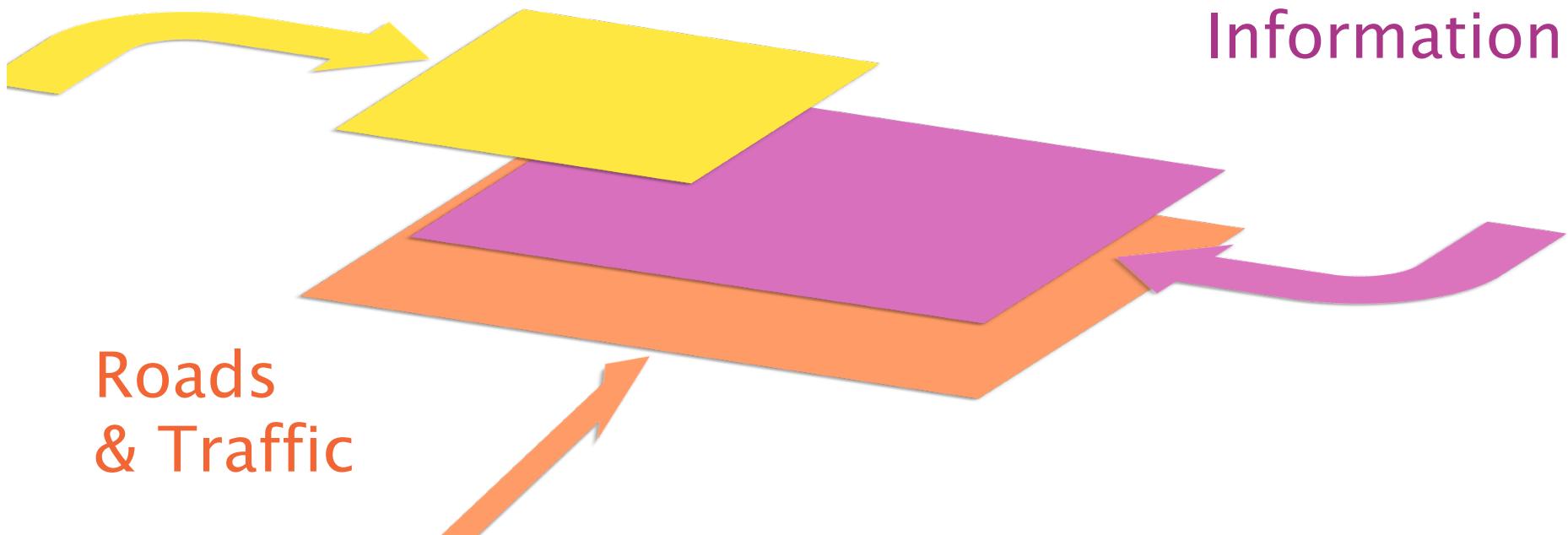
20

Parking Pressure



Where
should I
park my car
?

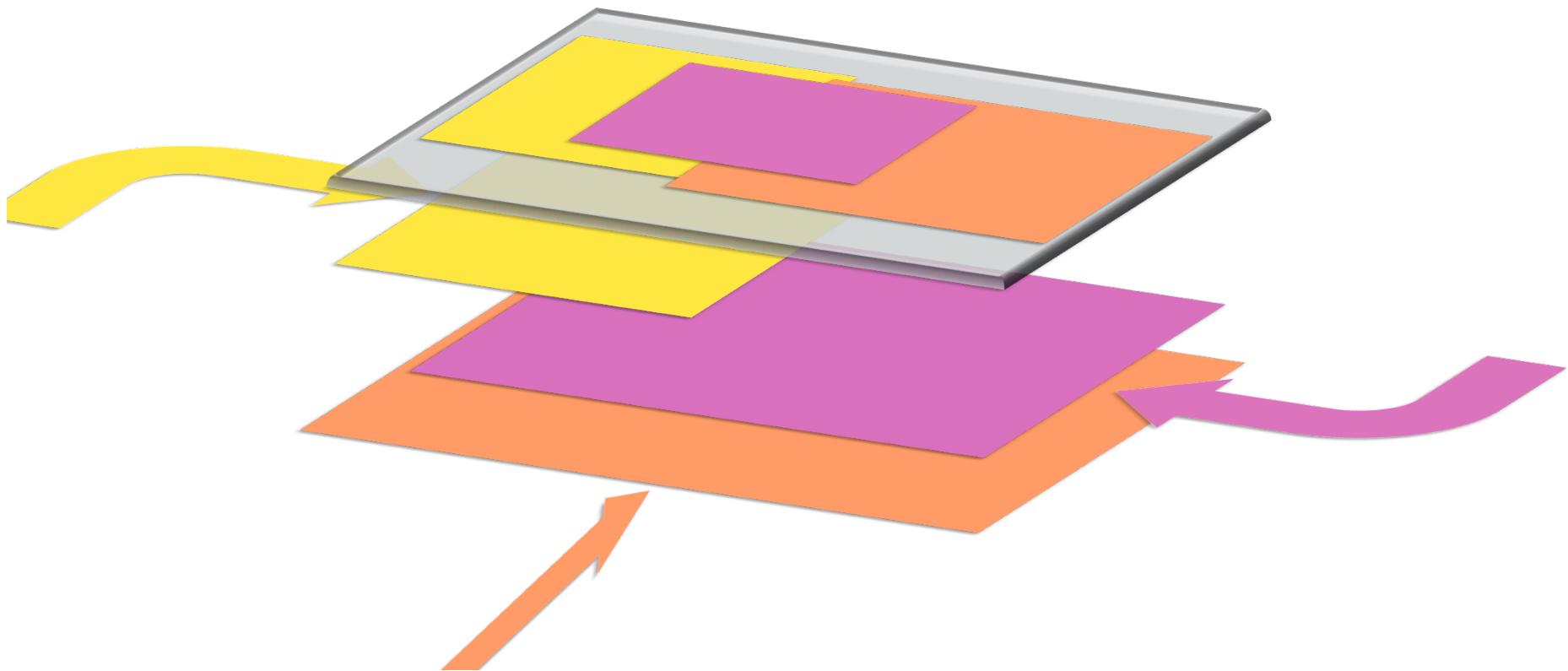
Parkmeters



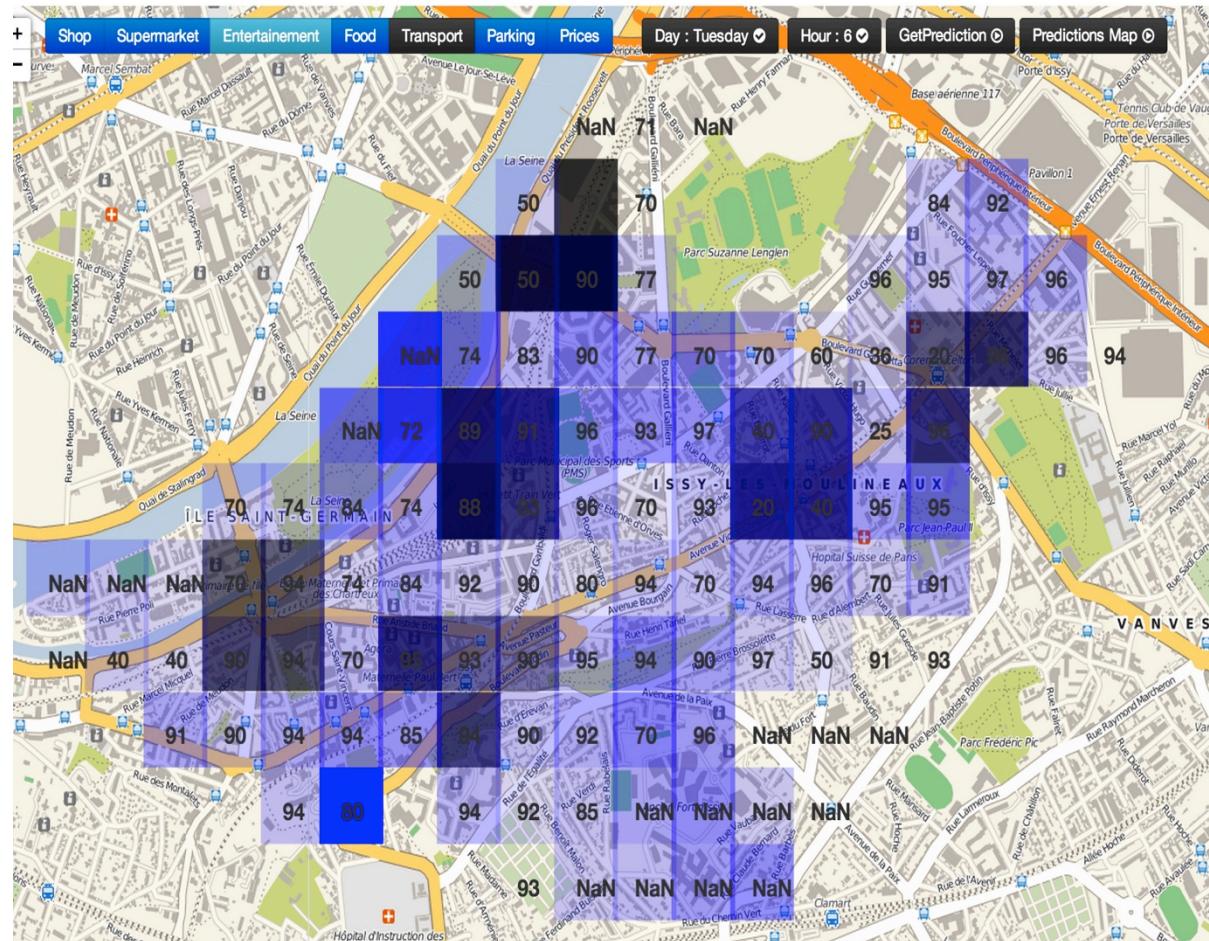
Roads
& Traffic

Geographic
Information

Is the parking full ?
How many people are likely to leave ?
What are “good” and “bad” streets?



Parking Pressure

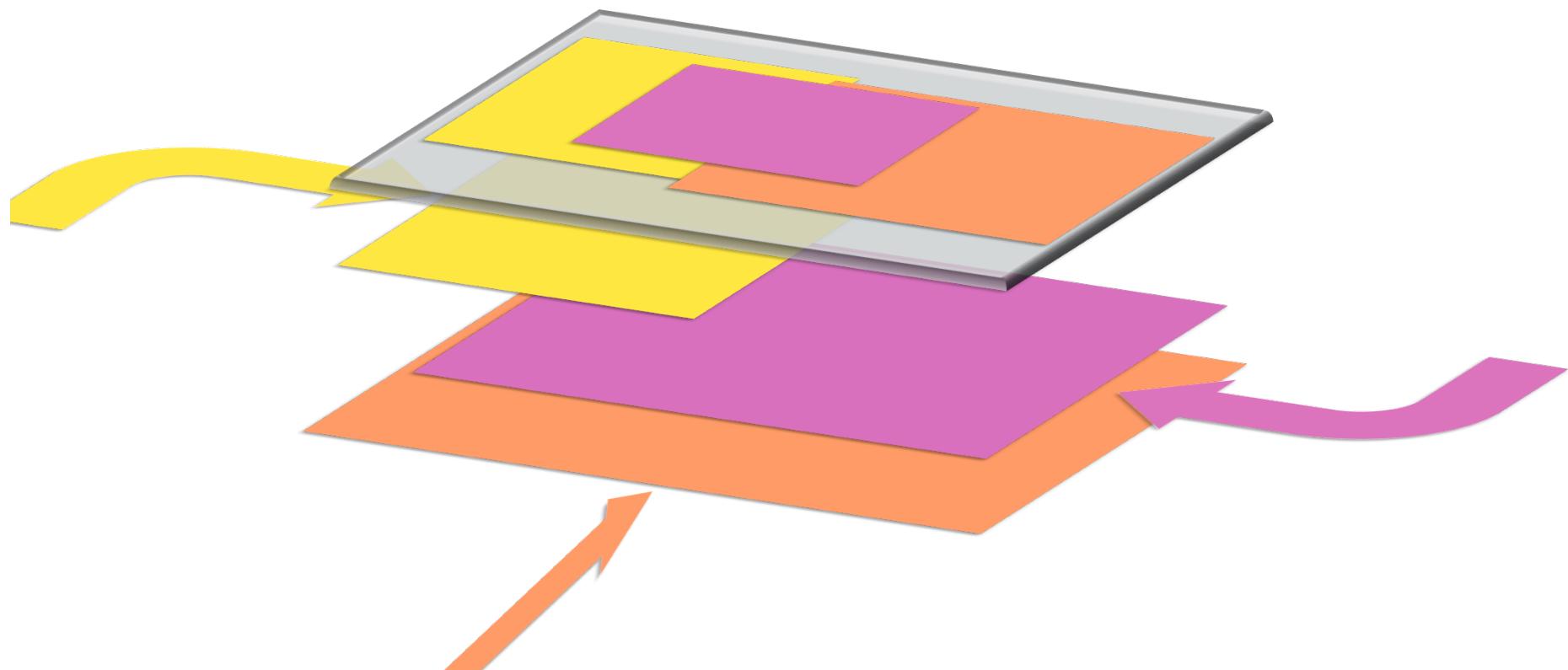


10/6/13

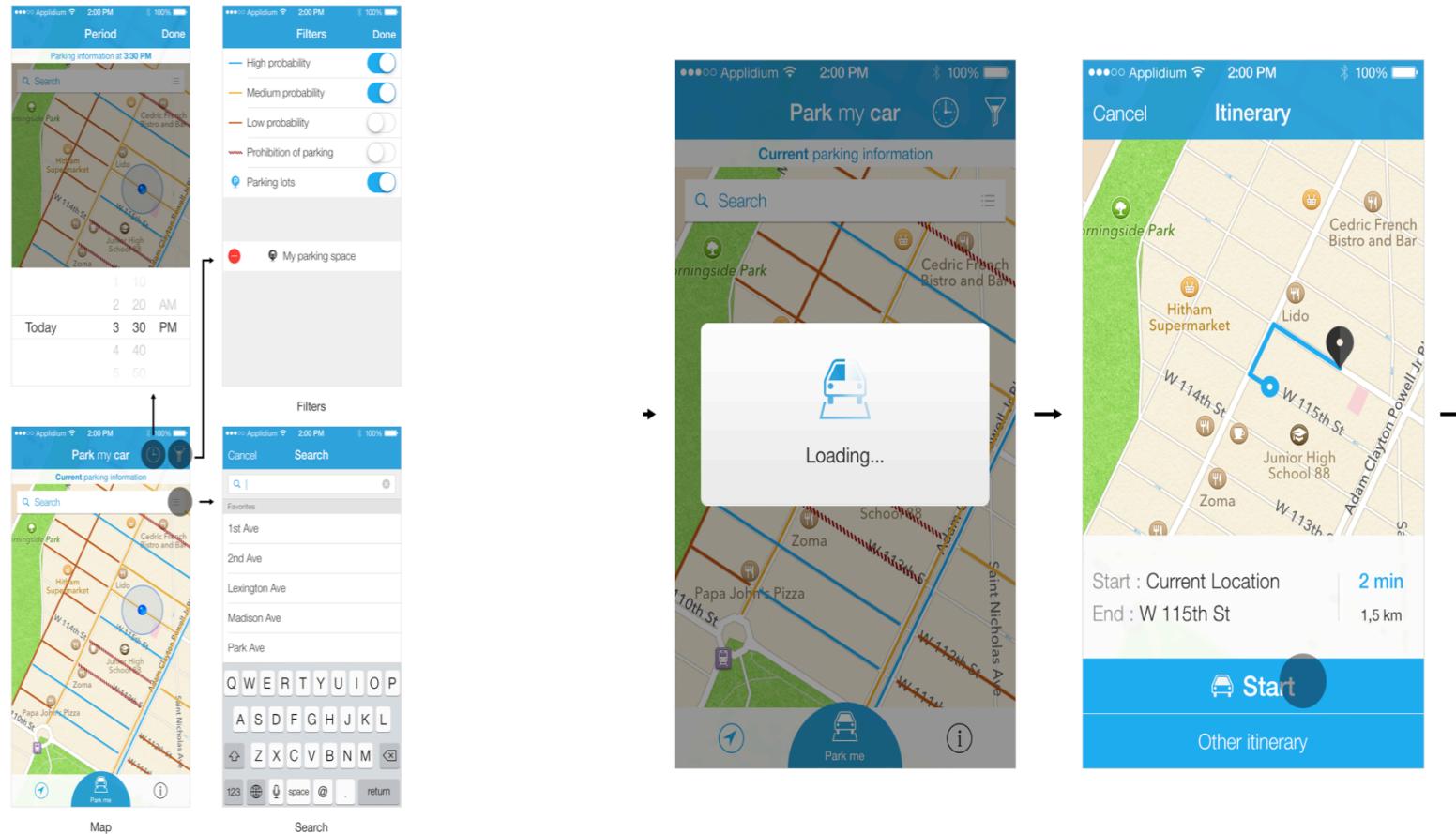
Practical Use:

Help People
Park their Car !

Help Cities put parking
places at the right location



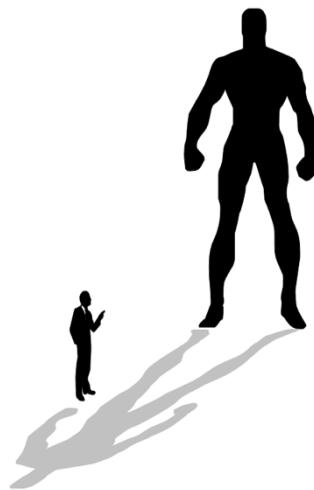
Parking Guidance



10/6/13

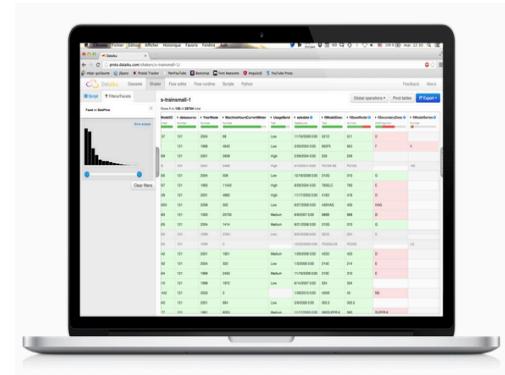
2
c

(This is a targeted ad)



Dataiku's **Data Science Studio**

Data Analysts → Data Scientists
Data Scientists → More Productive
MORE FUN
SEXIER
STRONGER

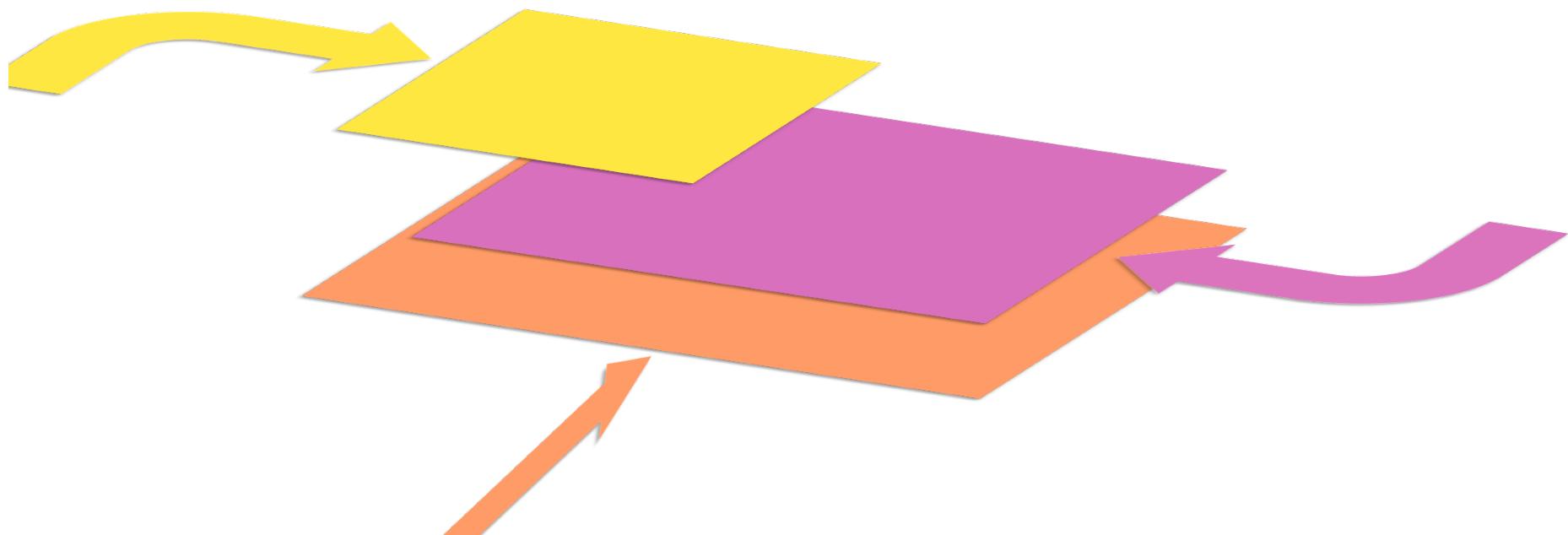


Rent a data scientist **or** Subscribe to the Studio

Big Data(s)

=

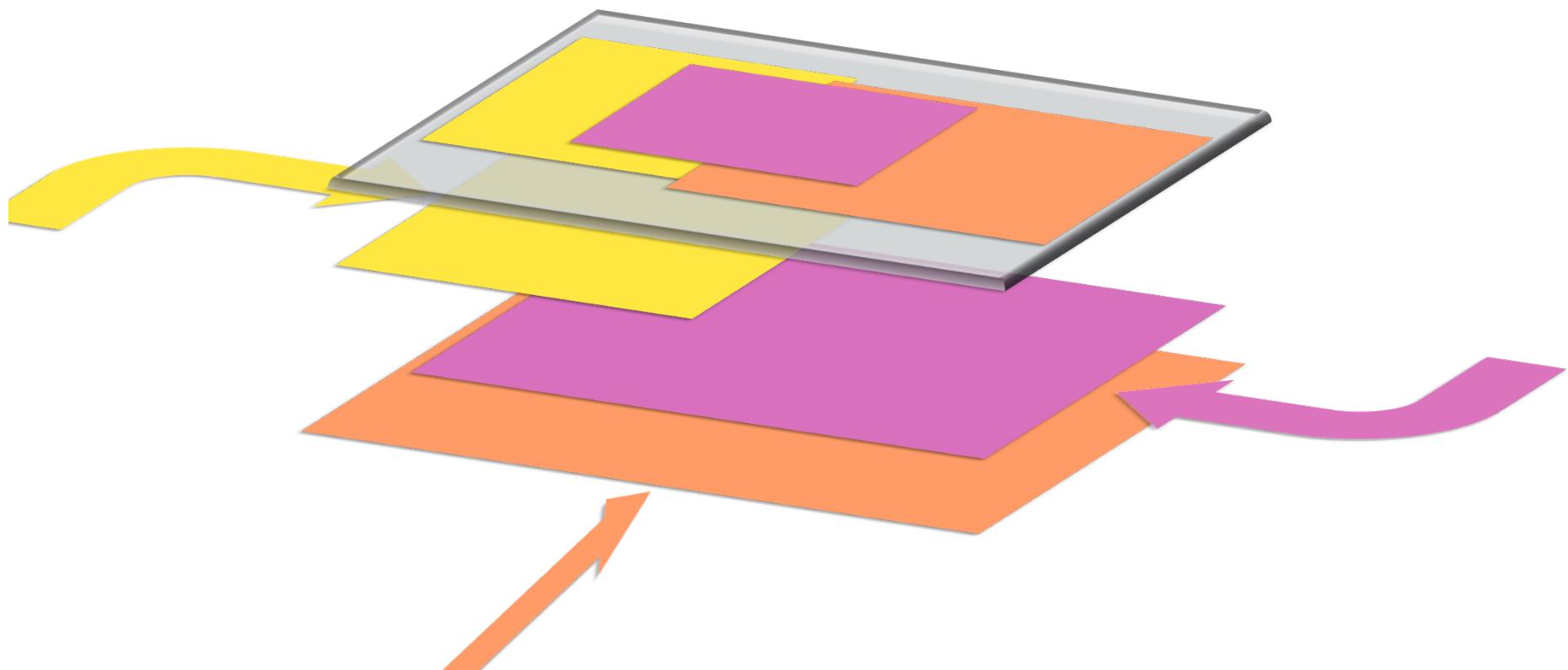
Put Together Data
Some small
some big



Machine Learning

=

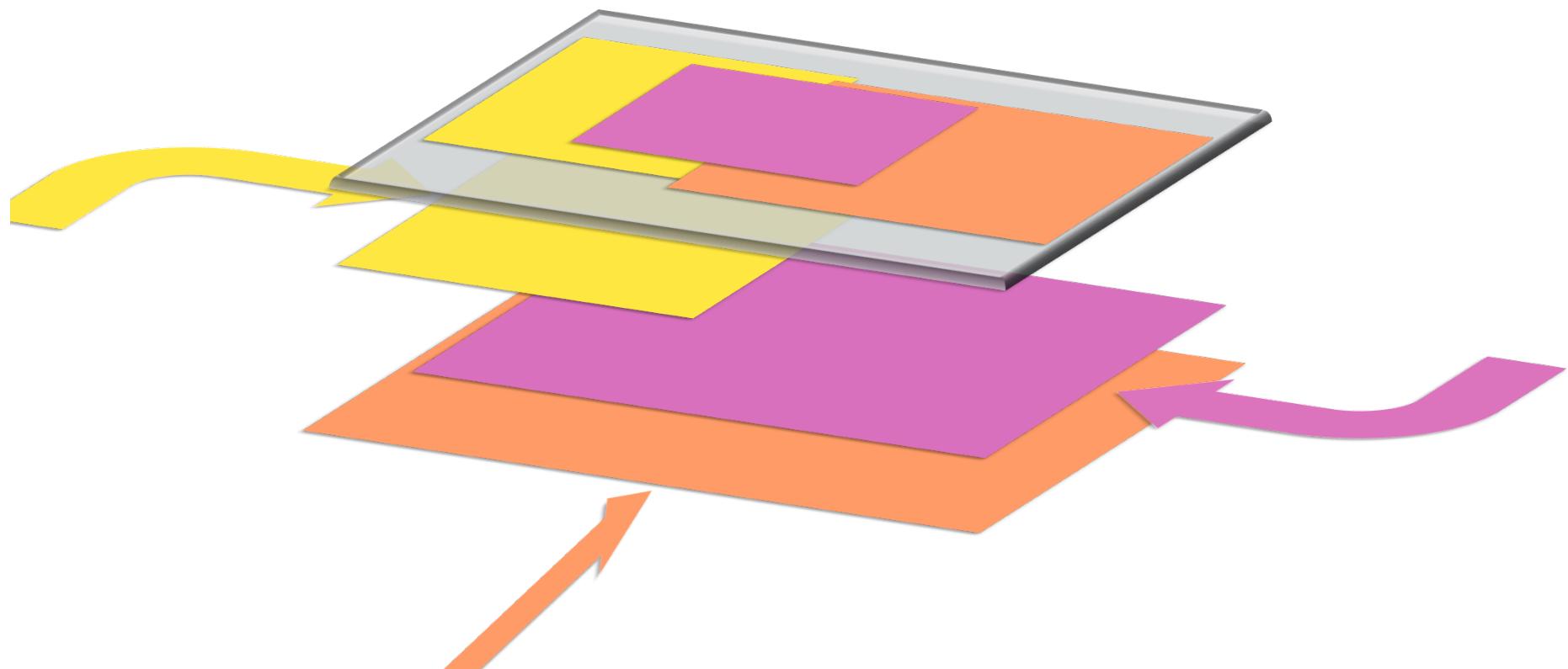
Create a layer that predicts
or explains the phenomena



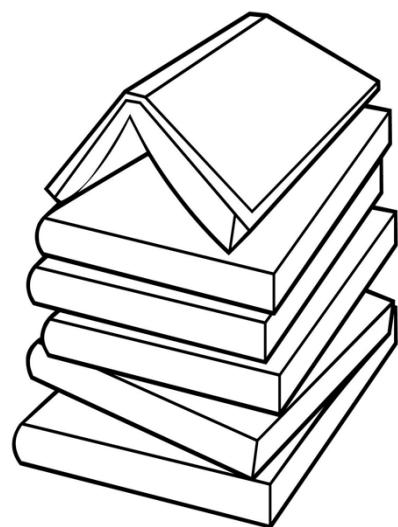
Practical Use

=

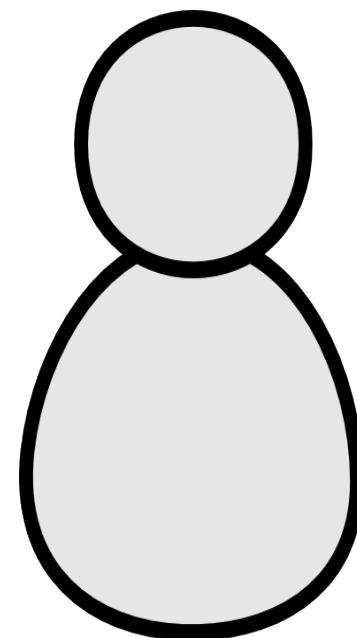
Mobile App – Advertising – Avoid Manual Work – Optimize Sales – ...
Automate Risk Calculations –



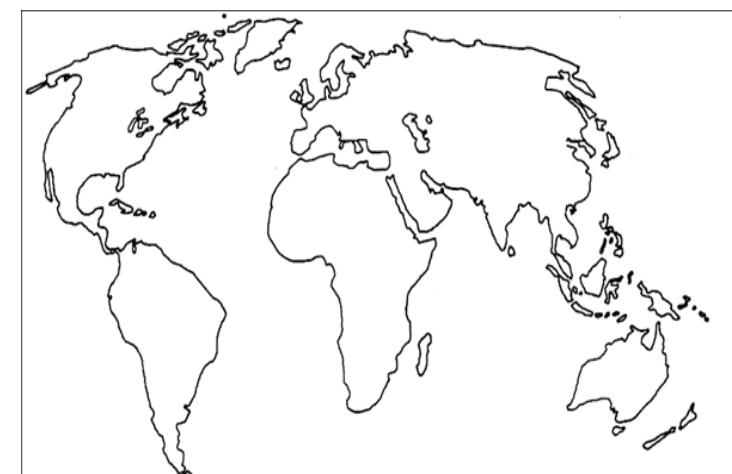
Knowledge Centric



User Centric



Geo Centric



Knowledge Centric Search Optimization

- ▶ How to optimize Search Relevance to provide better search for local very specific services ?

E.g.
“ssn old people”
in 1km around me”



Knowledge Worker Dilemma



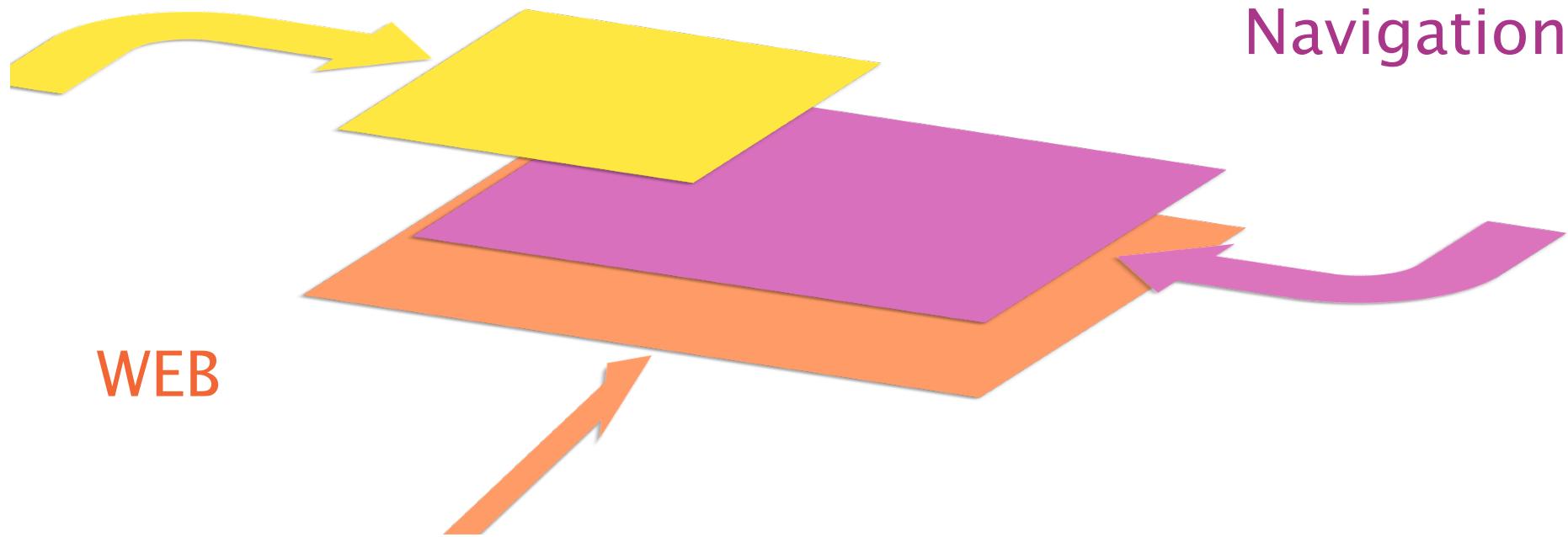
“I can't manage every word manually “

I already did 200.000 by hand !

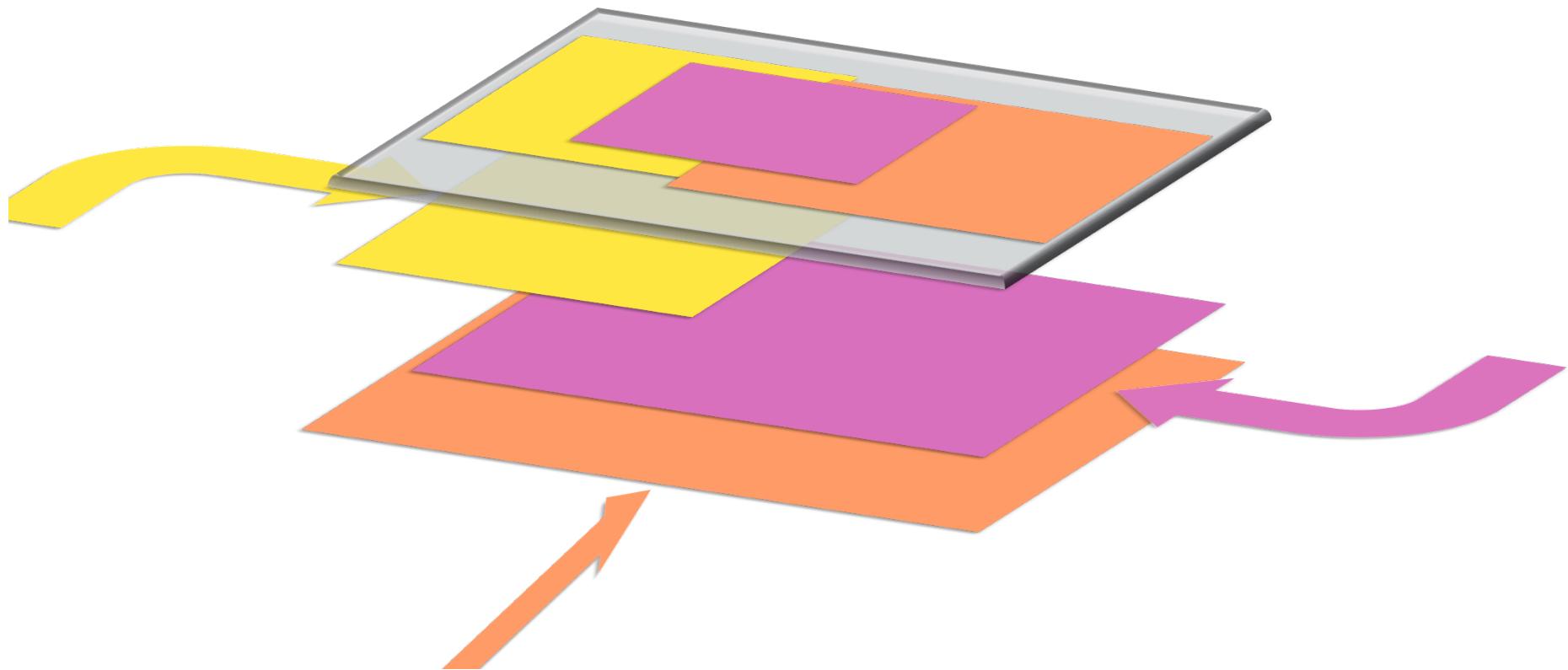
Allons!

User Feedbacks

User
Navigation



When is my user satisfied by response ?
Why is he dissatisfied ?
What corrections can I make automatically ?



Practical Use:
Automate the work
of knowledge Workers

ssn old people

=

**state-sponsored nurse
specialized in old people care”**

