Data-Driven Decision Making: It's Not Just For Experts

Jennie Duggan

MIT CSAIL

jennie@csail.mit.edu

DBg Database Group
MIT Computer Science and Artificial Intelligence Lab

Data-Driven Decision Making

- Big data has shown promise in a lot of fields
  - …but the benefits have been slow to reach everyday life
- Search is much better for global facts than specific situations
- Everyone deals with domains where they are not experts

Walter White

How much money do I need to save to provide for my wife and kids?

What is the expected yearly income for a car wash business in Albuquerque?

If I am caught, what is the likelihood that I will get leniency?
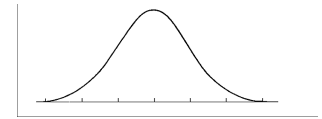
Data Sources

- Self
  - Curate data from web, smartphones, physical docs, etc.
    - Data includes hobbies, home, and so on
- External
  - Structured public info
  - Companies may sell coarse-grained queries
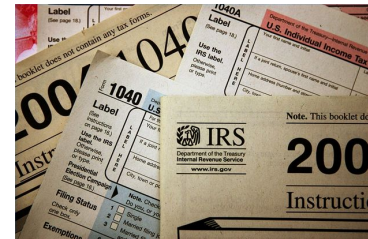- Middle roads, i.e., a data cooperative

Query Structure

Model its cash flow

SELECT EXPECTED_INCOME
FROM CARWASH
WHERE
LOCATION SIMILAR
TO 'Albuquerque';

Structured data source

Anchor: gives query context

Challenges

- Efficiently collecting & storing personal data
- Getting privacy right for each data source
  - Users desire privacy
  - External sources may need to protect IP
- Query language
- Easy model creation and configuration
  - Has hints for query optimizer
- Making statistical query output accessible