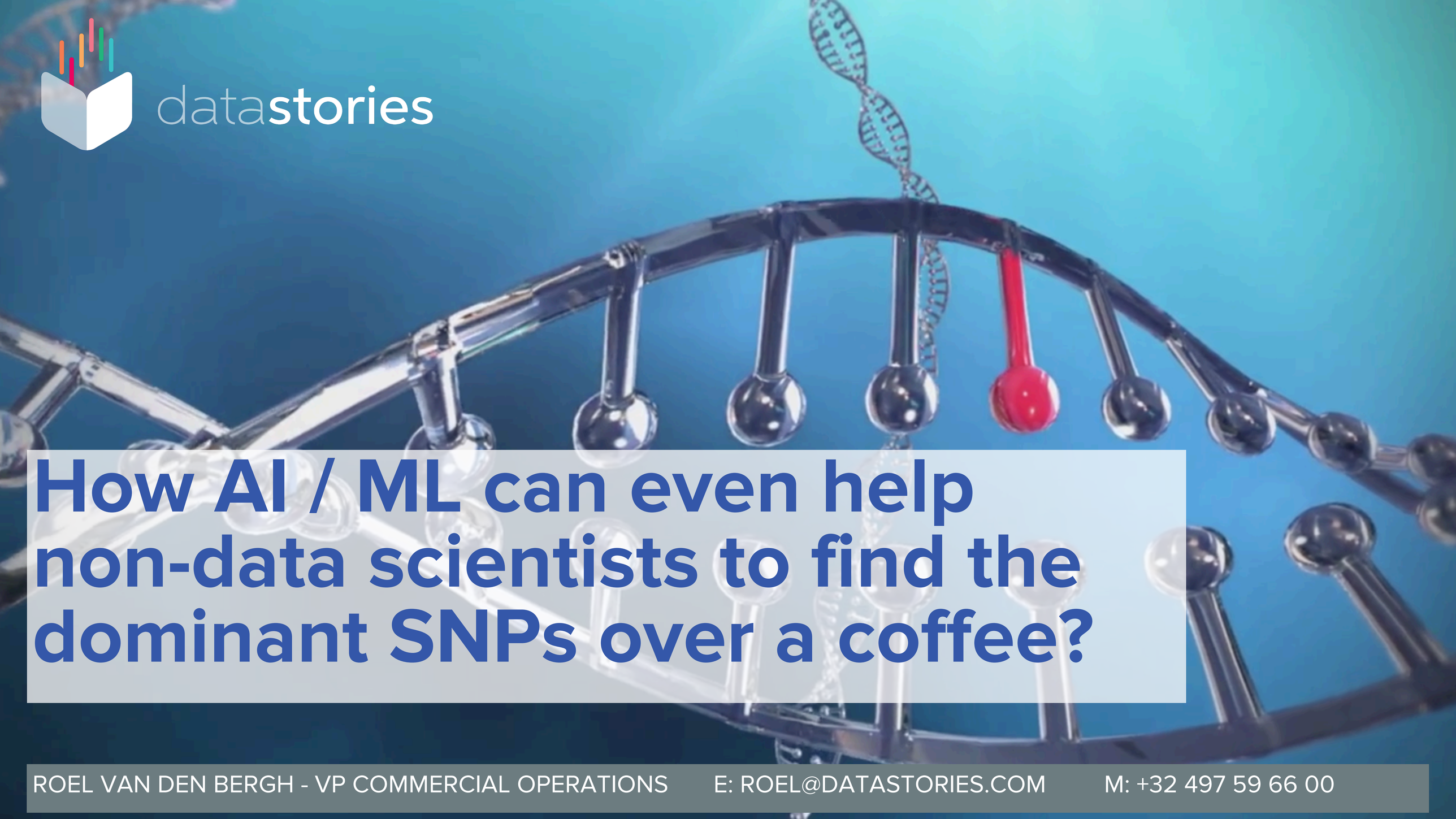
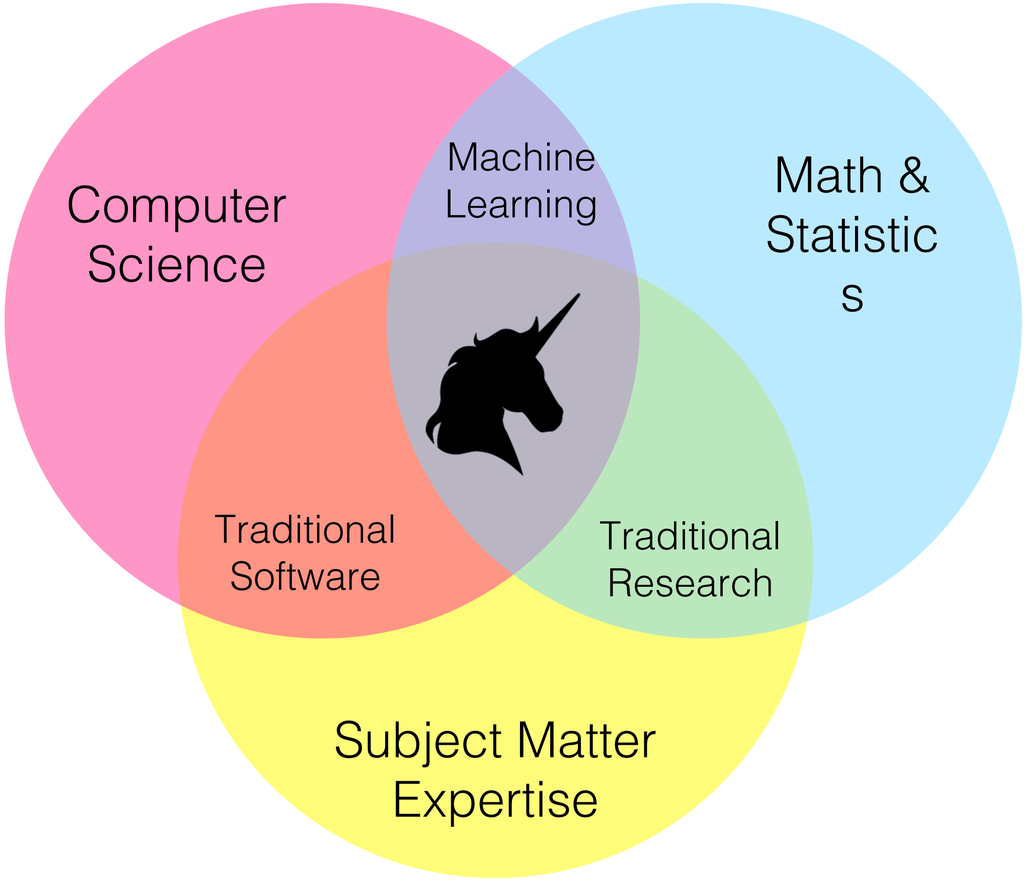


datastories

# How AI / ML can even help non-data scientists to find the dominant SNPs over a coffee?

ROEL VAN DEN BERGH - VP COMMERCIAL OPERATIONS E: ROEL@DATASTORIES.COM M: +32 497 59 66 00

# SYSTEMATIC INNOVATION ORGANIZATION




**BEFORE**


**TODAY WITH DATASTORIES**

datastories

# AMBITION: #1 in R&D

datastories

# AI for domain experts...

**1.000 data scientists in-a-box**

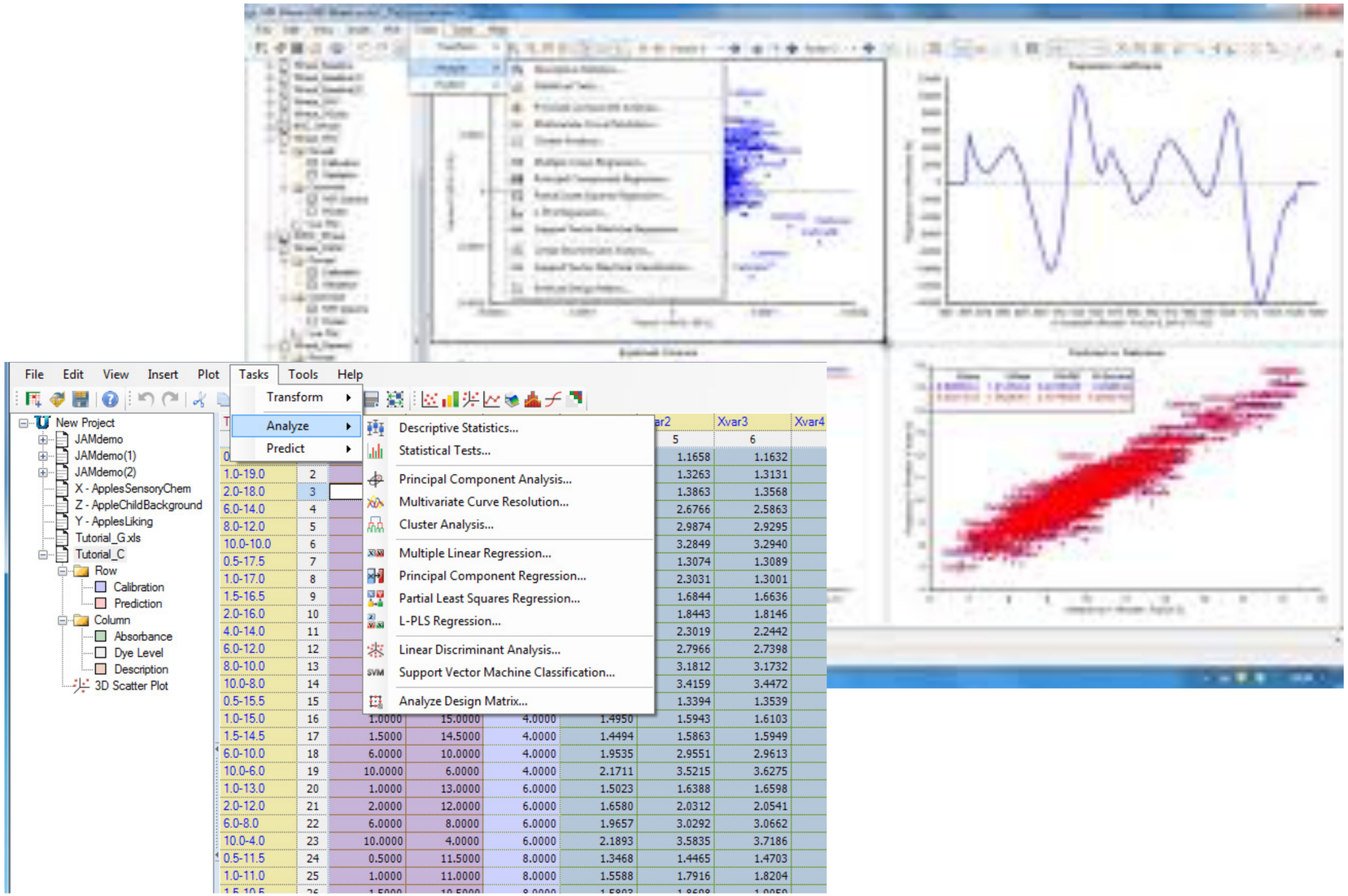
datastories

datastories

# Four main challenges

1. How does everything interconnects with each other?
2. What are the dominant variables?
3. How can I achieve optimal targets?
4. What are the exceptions to the rule?

# TRADITIONAL APPROACH

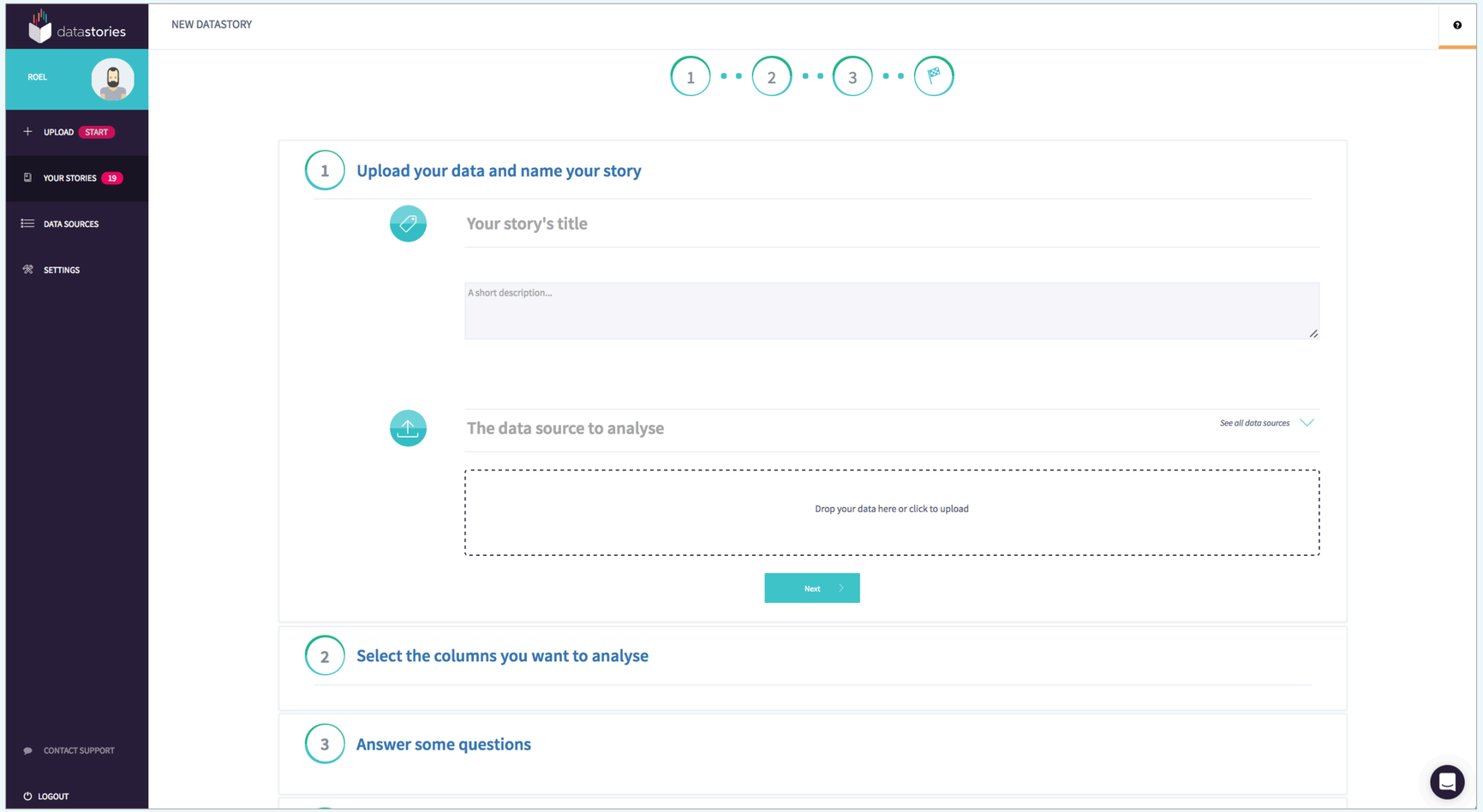



datastories

# FINDING THE RIGHT SNPs

## OVER A CUP OF

# FINDING THE RIGHT SNPs




## SIMPLE

1. Name
2. Link Dataset
3. Select Target
   eg. phenotype

**RUN**

datastories

# RESULTS

## ACCURACY IMPROVEMENT WITH 20% TO 200%

## WITH ONLY DOMINANT VARIABLES > 4 TO 10 SNPs in different cases

datastories

datastories

# FINDING THE RIGHT QTL's

**OVER A**

datastories

# CHALLENGE

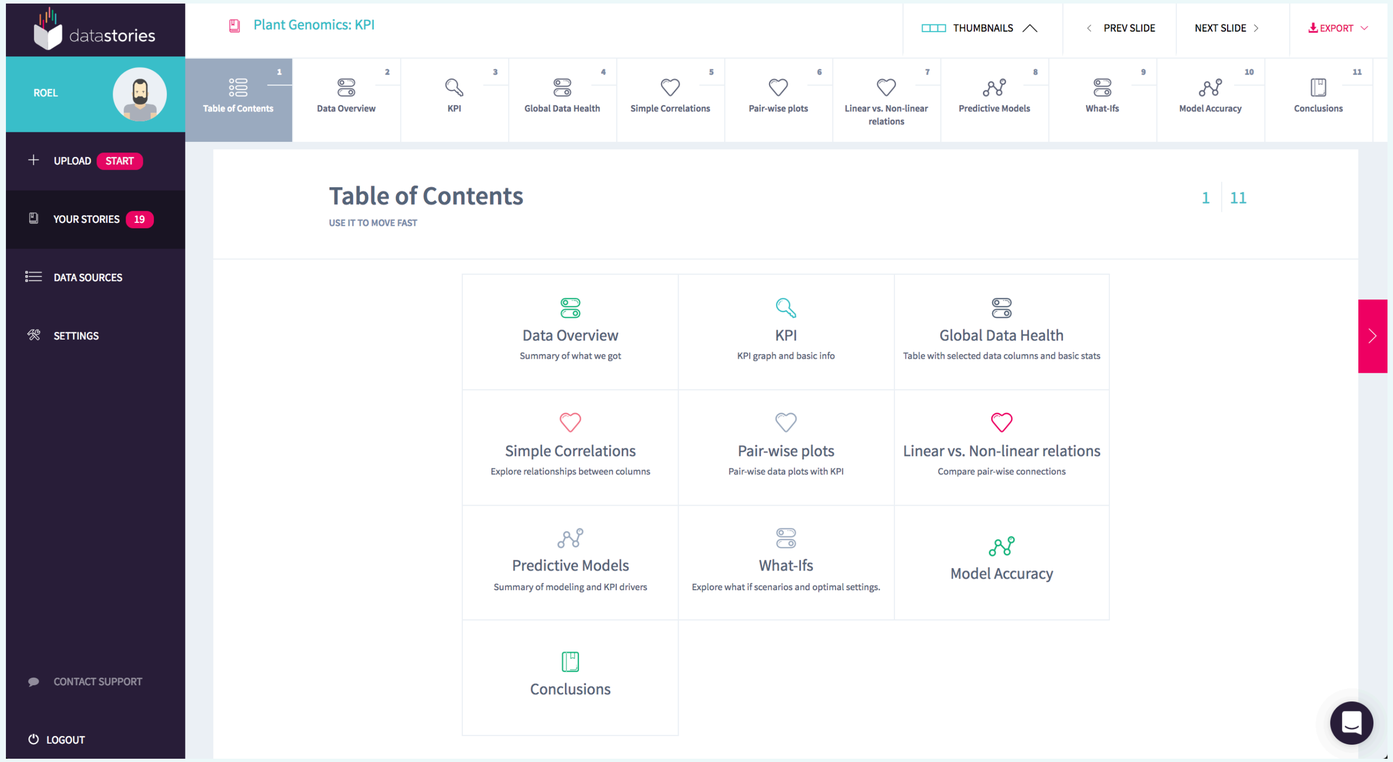
- +8000 QTLs
- ONLY 150 MEASUREMENTS
- RESULT OF WEEKS: 45% accuracy

**WHICH ARE THE DOMINANT QTLs?**

datastories

datastories

# RESULT AFTER LESS THAN 1 HOUR

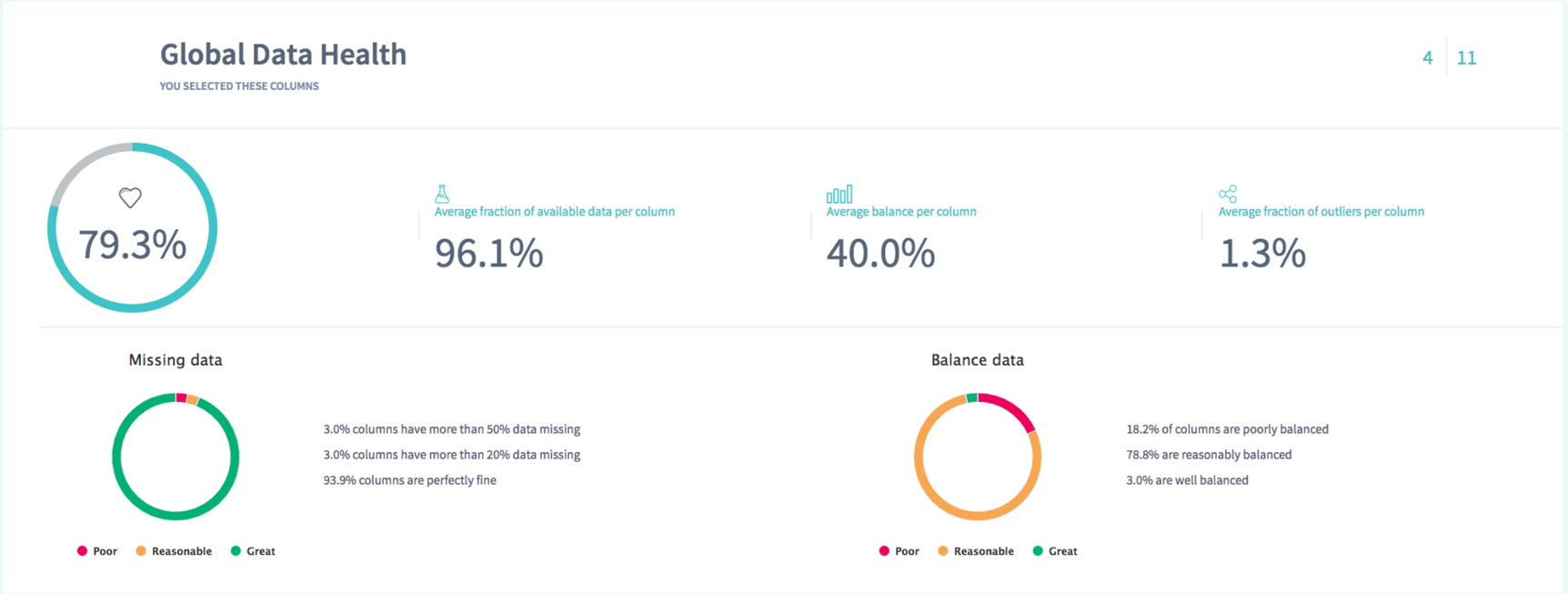



The story behind your data.

datastories

datastories

# RESULT AFTER LESS THAN 1 HOUR

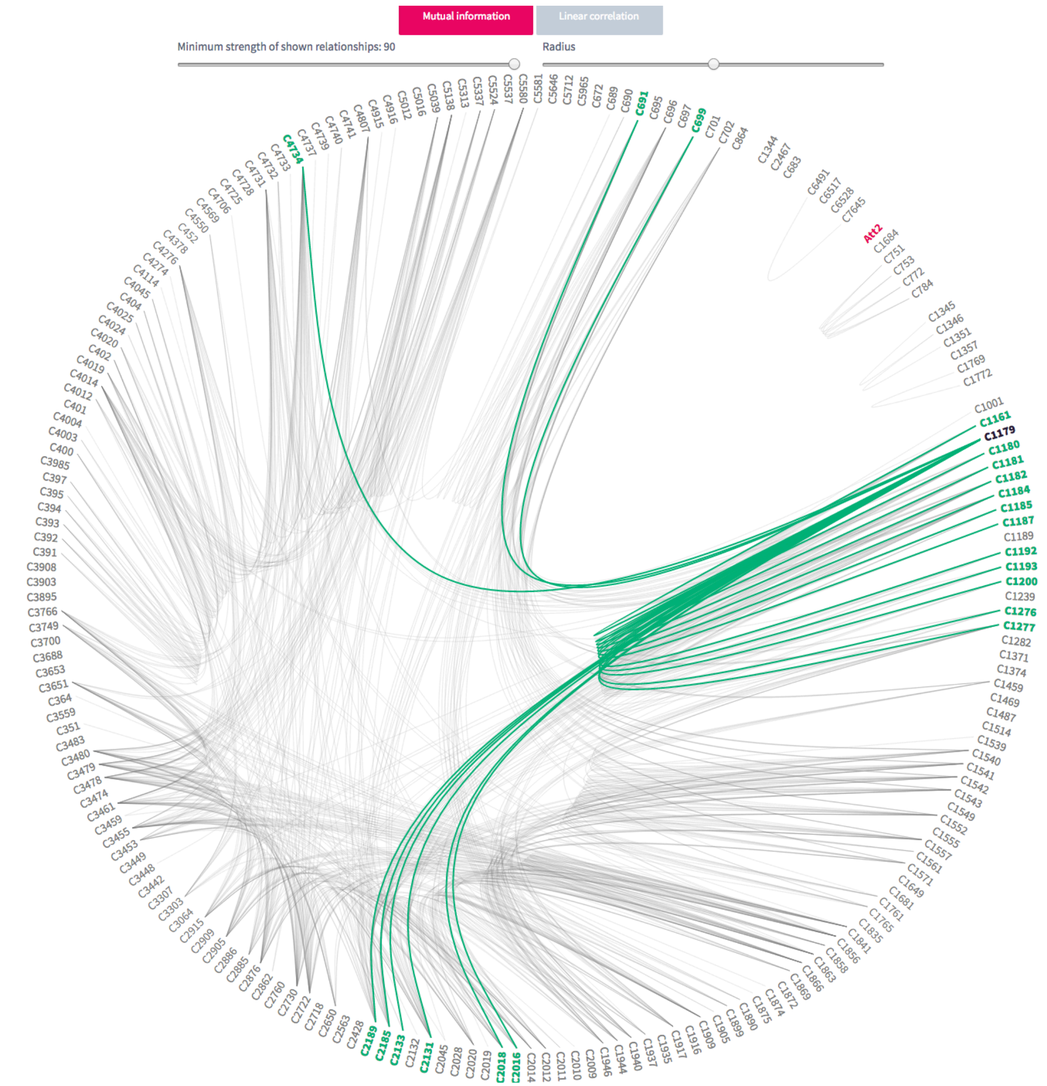



datastories

**Robust & Automatic correction for: missing data, unbalance, binary, categorical, etc...**

datastories

# RESULT AFTER LESS THAN 1 HOUR




## Non-linear correlation linkage map

datastories

datastories

# RESULT AFTER LESS THAN 1 HOUR

datastories

PREDICTION ACCURACY

81.3%

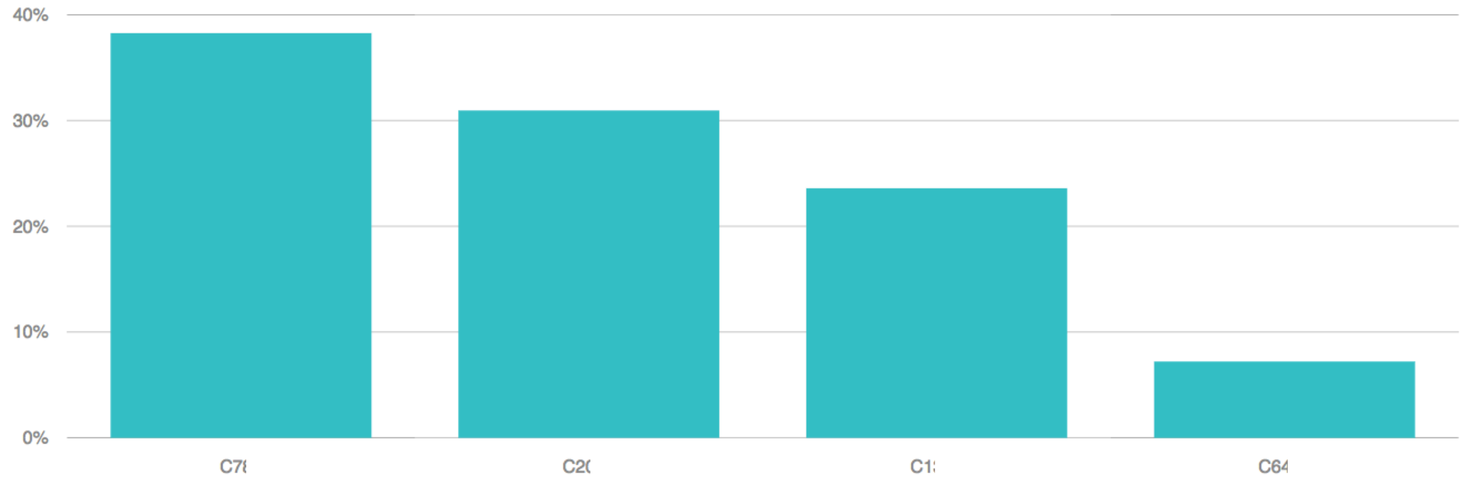
NUMBER OF VARIABLES:

4

NUMBER OF VARIABLES WE STARTED WITH:

8076

**Important drivers:**




datastories

# RESULT

## AFTER LESS THAN 1 HOUR

- IDENTIFIED 4 DOMINANT QTLs
- ACCURACY IMPROVED FROM 45% TO 81%

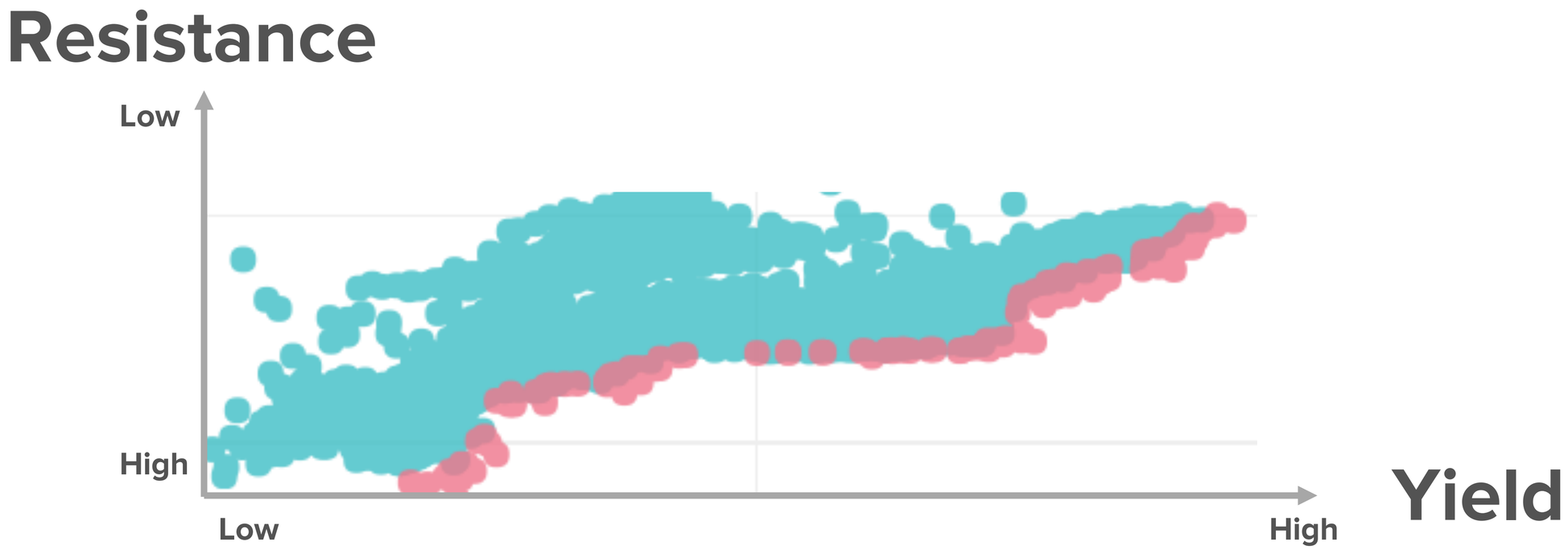

datastories

# MULTI-OBJECTIVE OPTIMISATION

## eg. Yield & Resistance

datastories

# IN-SILICO EXPERIMENTS




datastories

# WORK IN PROCESS

## PROPRIETARY TECHNOLOGY

## DATASTORIES CRISPR

For millions of variables

datastories

datastories

# APPROACH

- **APPLIED OUR IoT BEST PRACTICES FROM THE INDUSTRY**

- **INSPIRED BY GENOMICS PROPRIETARY TECHNOLOGY FOR VIRTUAL MARKERS**

datastories

# RESULT DATASTORIES CRISPR

datastories

**ACCURACY IMPROVED FROM 45% (PCA) TO +84% (DataStories)**

datastories

# OFFERING

- AI SOFTWARE FOR RESEARCHERS
- MULTI OBJECTIVE OPTIMISATION
- SMART DESIGN OF EXPERIMENTS
  **100x to 500x faster**
- UNICORN ACADEMY

datastories

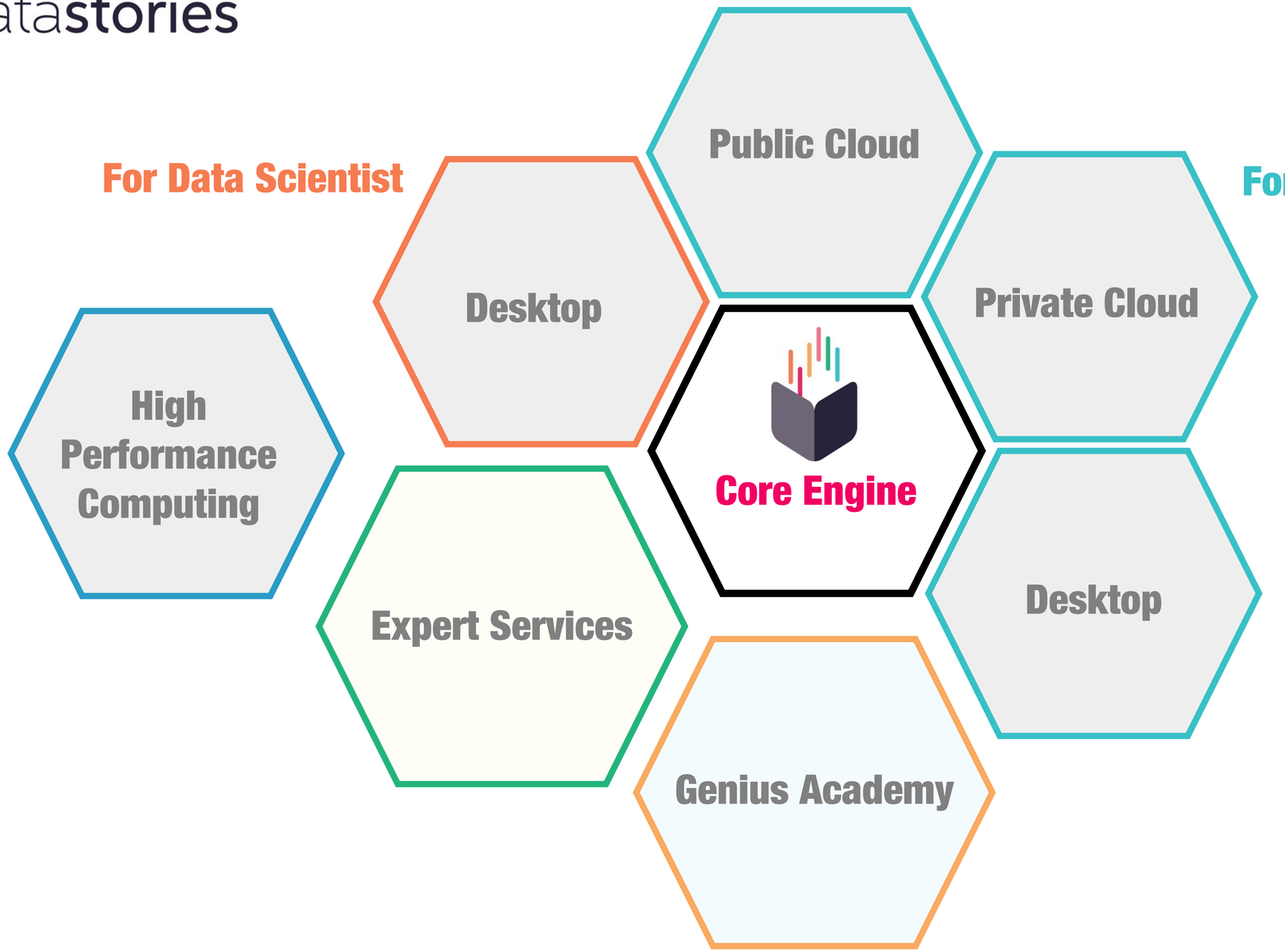
datastories

# Unicorn Academy

- **Fundamentals of Data Science**
- **Advanced Data Science**
- **Business-driven Data Science**
- **Data science for R&D**
- **Systematic Innovation Organisation**
- **The Data infrastructure and Data engineering behind Data Science / Big Data**


datastories

datastories




datastories