

ThoughtWorks®

YOTTABYTE 2016

Introduction to Data Science

QUEM



Felipe Martins

Consultant at ThoughtWorks Recife

fmmartin@thoughtworks.com

[@ffmmjj_martins](#)

A woman with glasses and a light-colored blazer is pointing at a wall covered in numerous sticky notes and charts. The wall is a collaborative workspace. The image has a green tint.

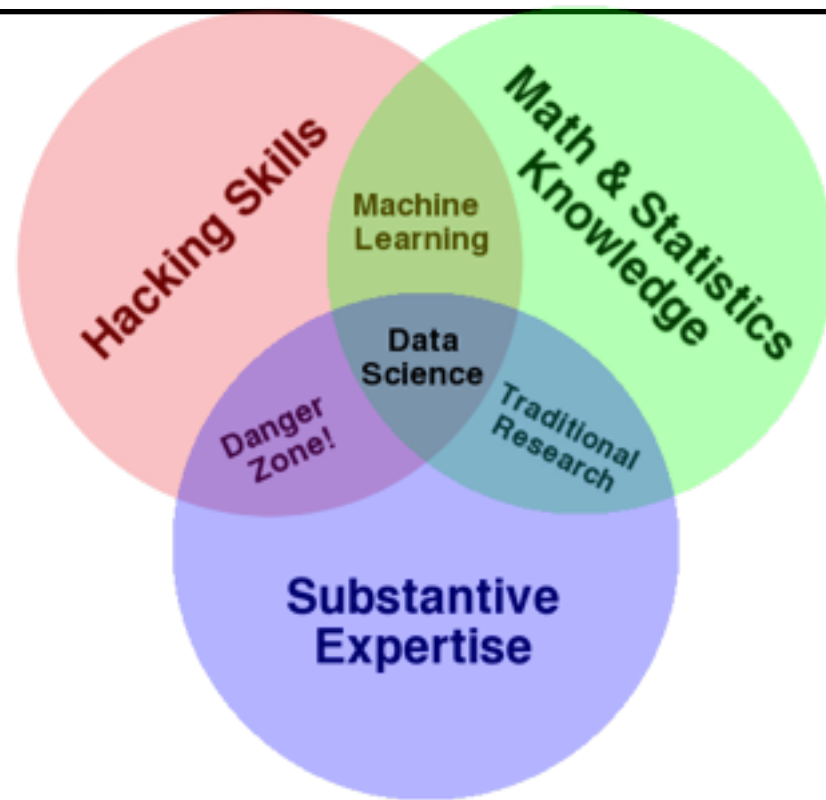
ThoughtWorks®

BI, STATISTICS, DATA SCIENCE...

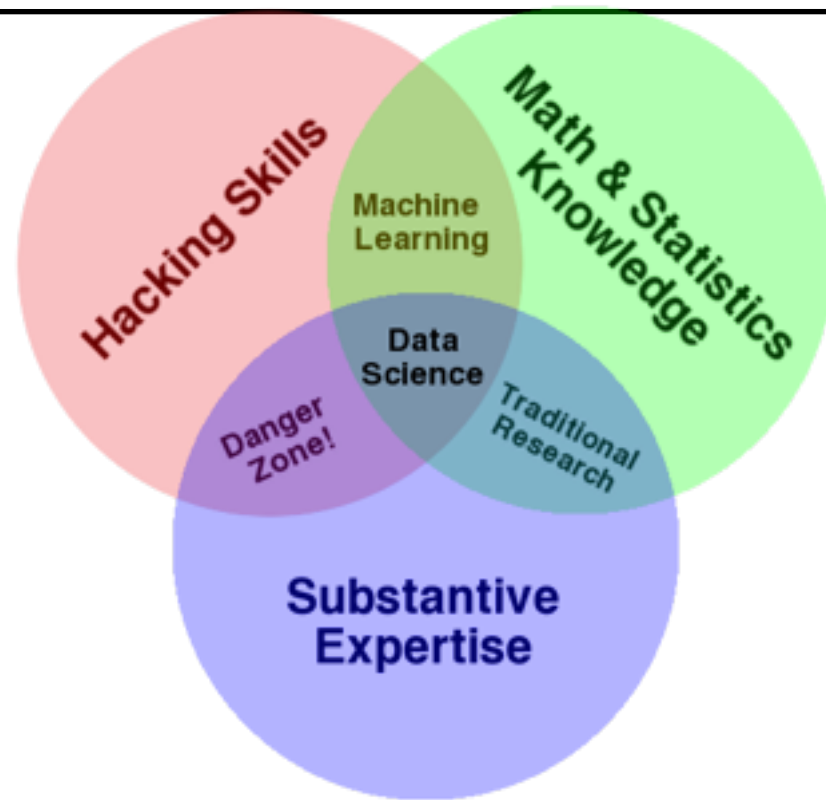
*"There is more or less a consensus
about the lack of consensus regarding a
clear definition of data science."*

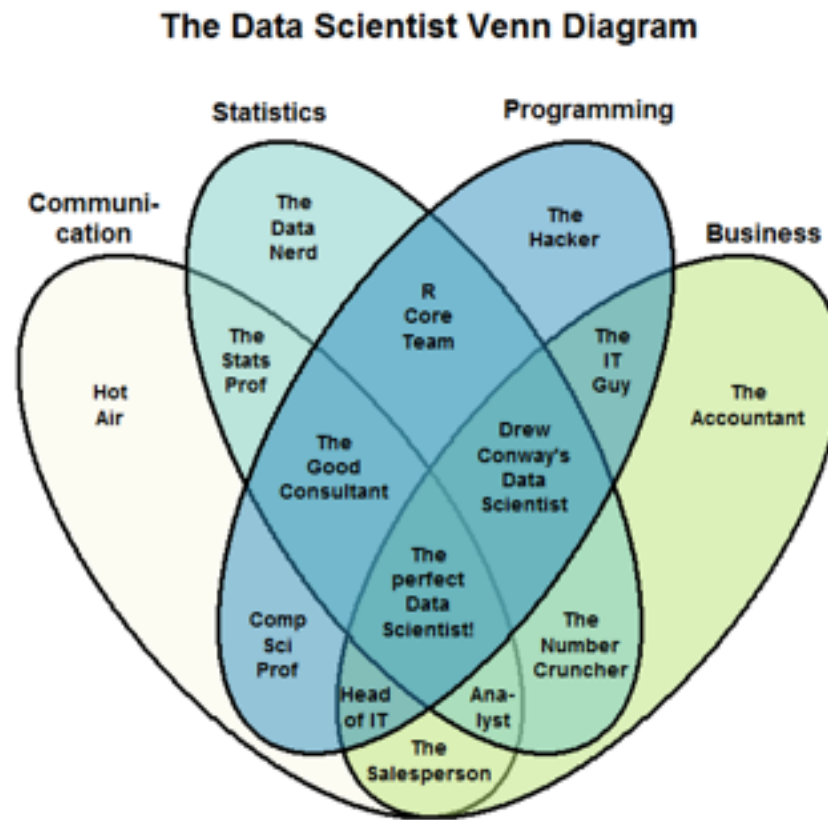
Gil Press in this article

WHAT IS A DATA SCIENTIST

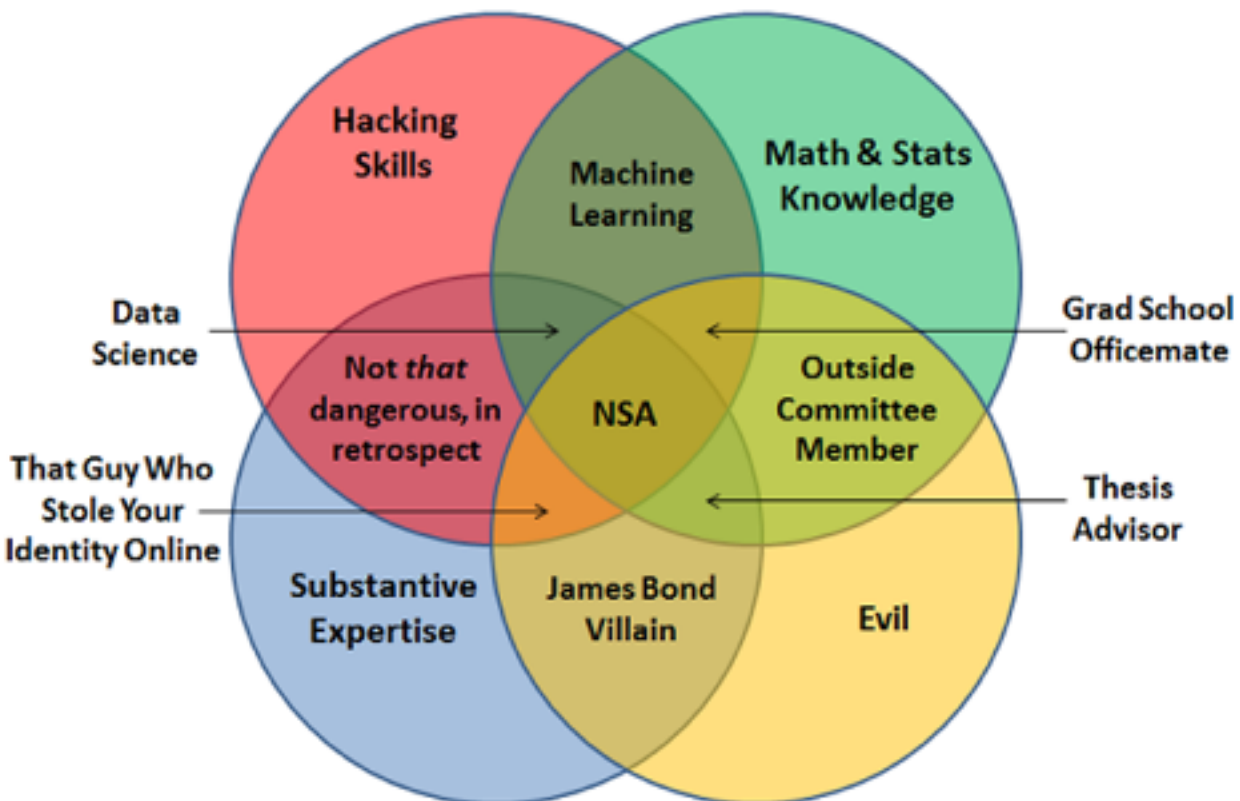
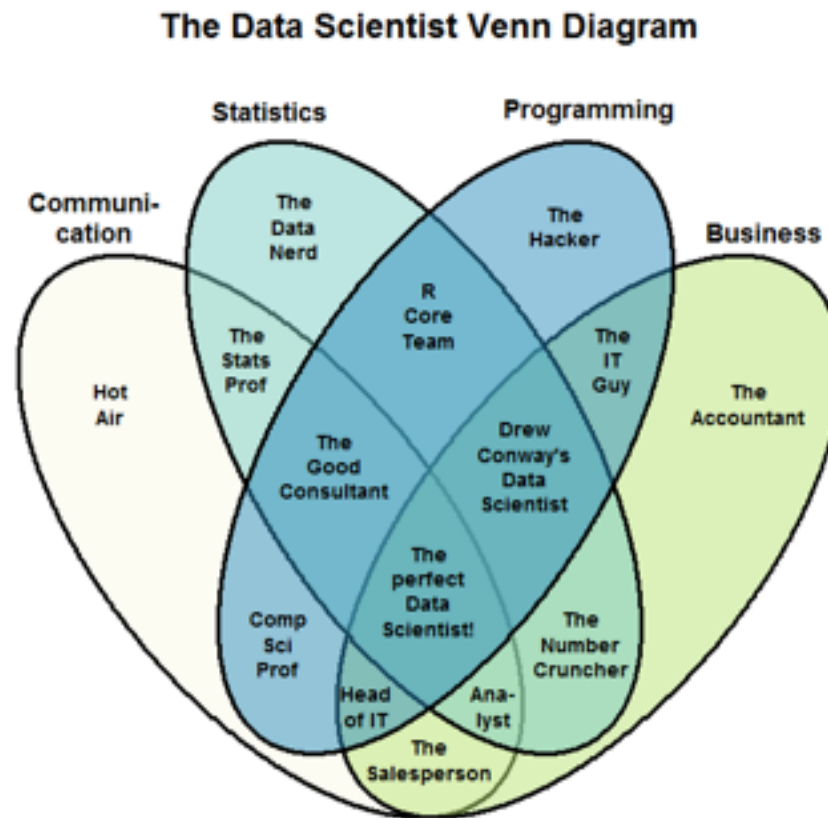
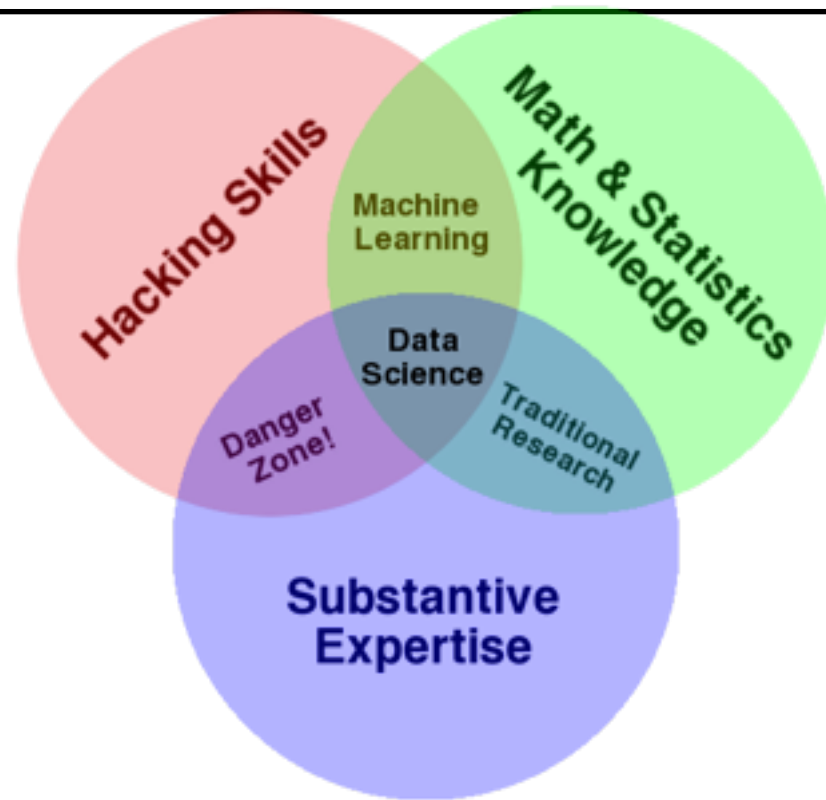


WHAT IS A DATA SCIENTIST





WHAT IS A DATA SCIENTIST



WHAT IS A DATA SCIENTIST



"A data scientist is someone who blends, math, algorithms, and an understanding of human behavior with the ability to hack systems together to get answers to interesting human questions from data."

Hilary Mason, data scientist at Accel

SOME APPLICATIONS



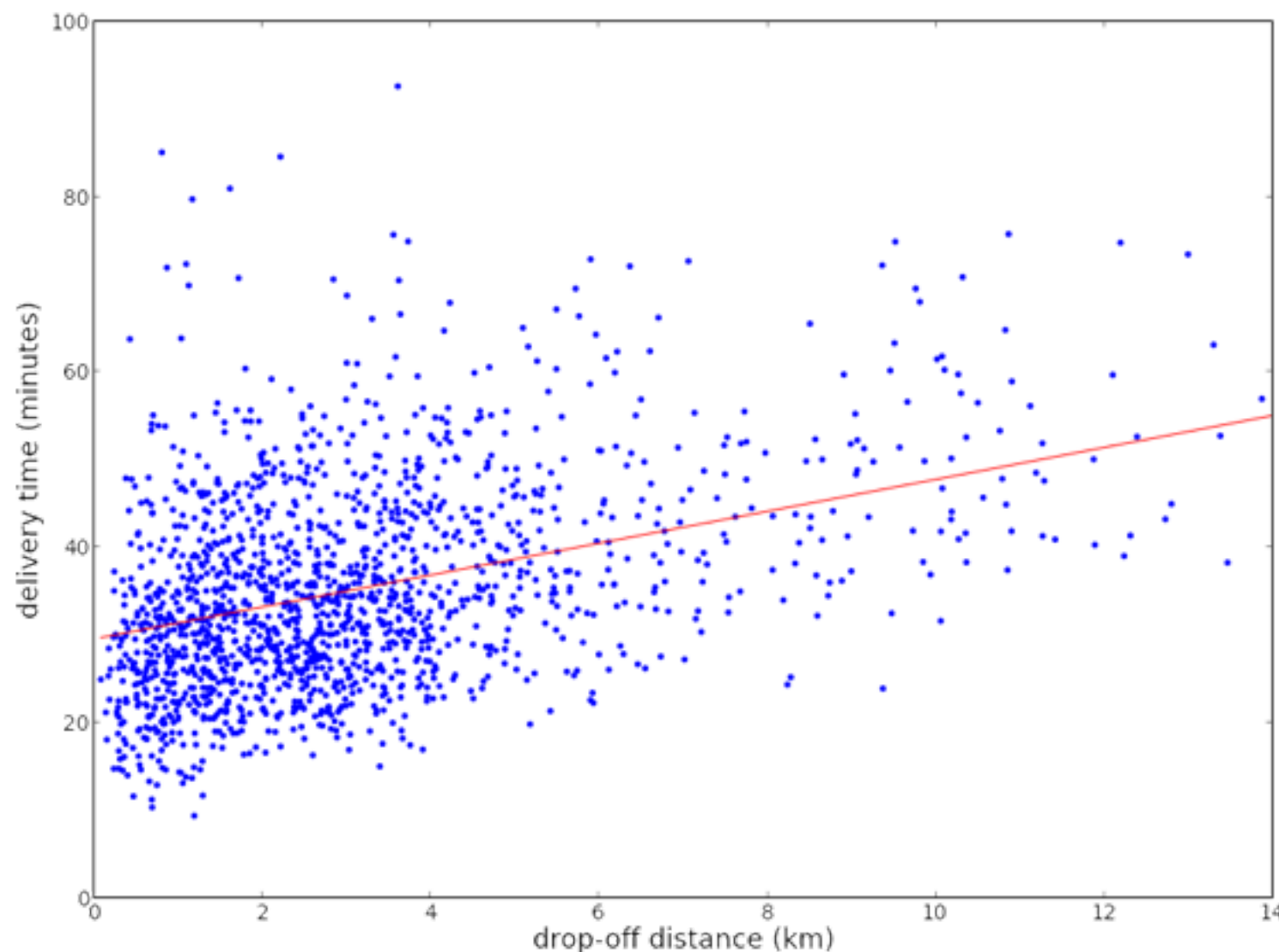
POSTMATES

SOME APPLICATIONS

Estimating Delivery Times: A Case Study In Practical Machine Learning



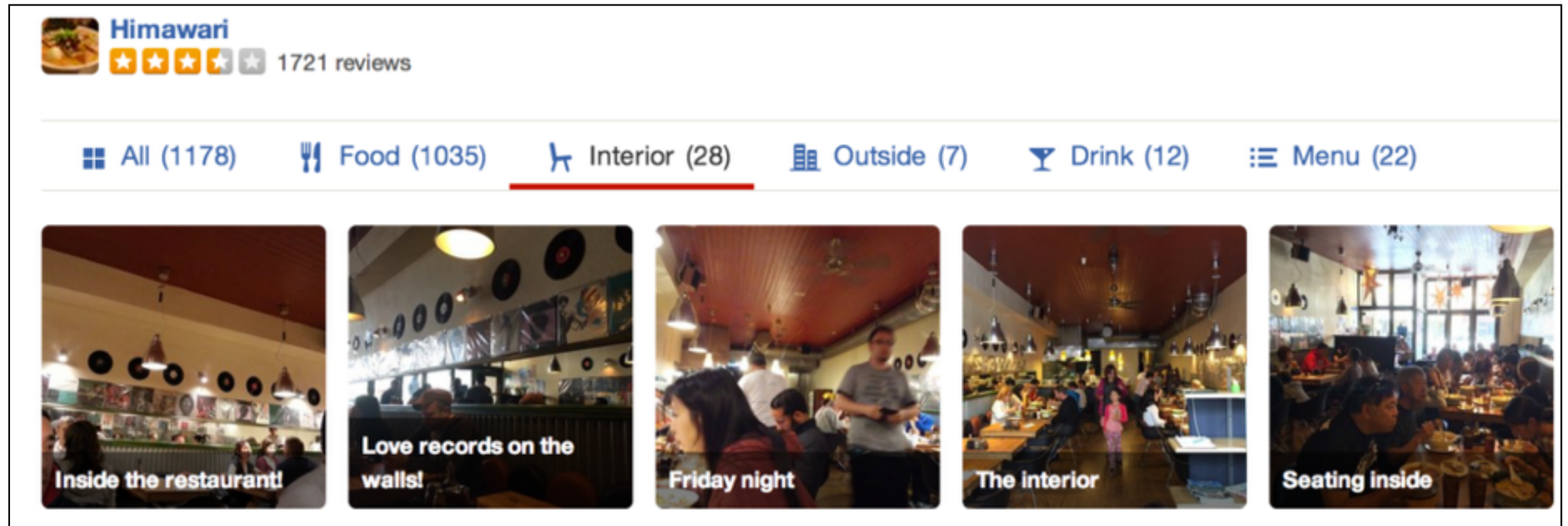
POSTMATES



SOME APPLICATIONS



SOME APPLICATIONS



How We Use Deep Learning to Classify Business Photos at Yelp

A close-up photograph of a smiling woman with dark hair, wearing a patterned scarf and a dark jacket. She is holding a smartphone in her hands. The image has a green tint. The text 'ThoughtWorks®' is in the top left, and 'LET'S GET SOME DATA!' is in the center in large white letters.

ThoughtWorks®

LET'S GET SOME DATA!

JUPYTER NOTEBOOKS

[https://github.com/ffmmjj/
intro_to_data_science_workshop](https://github.com/ffmmjj/intro_to_data_science_workshop)

WRAP-UP

- Regression, Classification, Clustering;
- Dimensionality reduction;
- Training and testing datasets;

WHAT NOW?

- Andrew Ng's MOOC on machine learning;
 - <https://www.coursera.org/learn/machine-learning>

- Book: [Data Science from Scratch](#);

- [Scikit-learn's cheatsheet](#);

WHAT NOW?

- Data Skeptic podcast;
 - [MINI] series;

- Higher-level tools:
 - Tableau, WEKA, gretl...

THANK YOU

ThoughtWorks®