

# ThoughtWorks®

---

# YOTTABYTE 2016

---

*Introduction to Data Science*

# QUEM

---



## Felipe Martins

Consultant at ThoughtWorks Recife

[fmmartin@thoughtworks.com](mailto:fmmartin@thoughtworks.com)

[@ffmmjj\\_martins](#)

A woman with glasses and a light-colored blazer is pointing at a wall covered in various documents, charts, and sticky notes. The background is a collage of business-related graphics, including flowcharts, bar charts, and network diagrams. The entire image has a green tint.

ThoughtWorks®

# BI, STATISTICS, DATA SCIENCE...

---

# Big Data Changes Your Analysis

Area	BI Analyst	Data Scientist
Focus	Reports, KPIs, trends	Patterns, correlations, models
Process	Static, comparative	Exploratory, experimentation, visual
Data sources	Pre-planned, added slowly	On the fly, as-needed
Transform	Up front, carefully planned	In-database, on-demand, enrichment
Data quality	Single version of truth	"Good enough", probabilities
Data model	Schema on load	Schema on query
Analysis	Retrospective, Descriptive	Predictive, Prescriptive

EMC<sup>2</sup>

© Copyright 2013 EMC Corporation. All rights reserved.

24

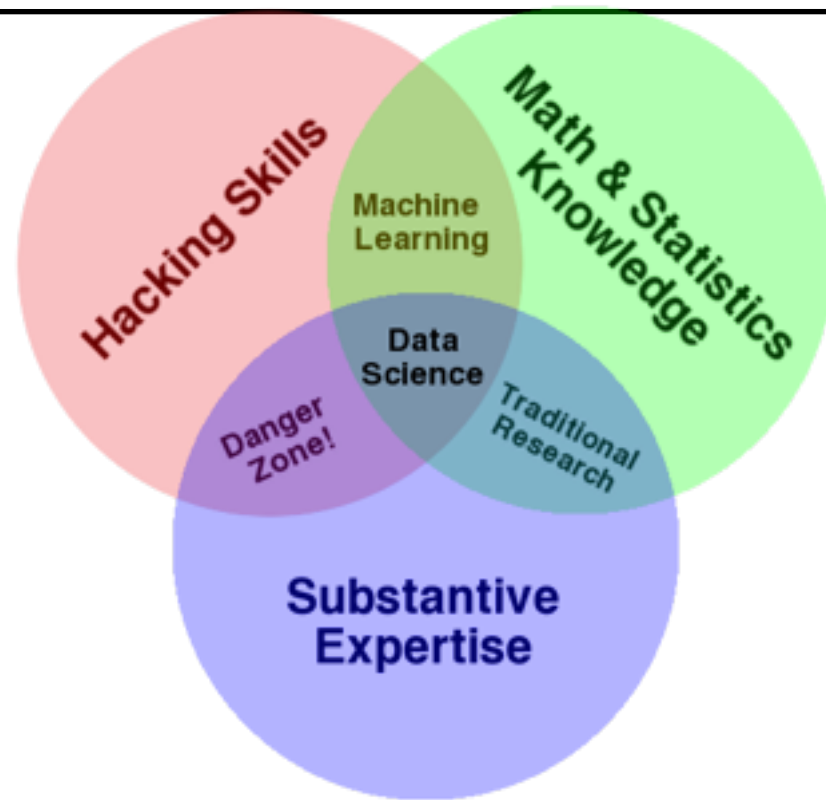
Source: <http://reflectionsblog.emc.com/business-intelligence-analyst-data-scientist-whats-difference/>

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

*Gil Press in this article*

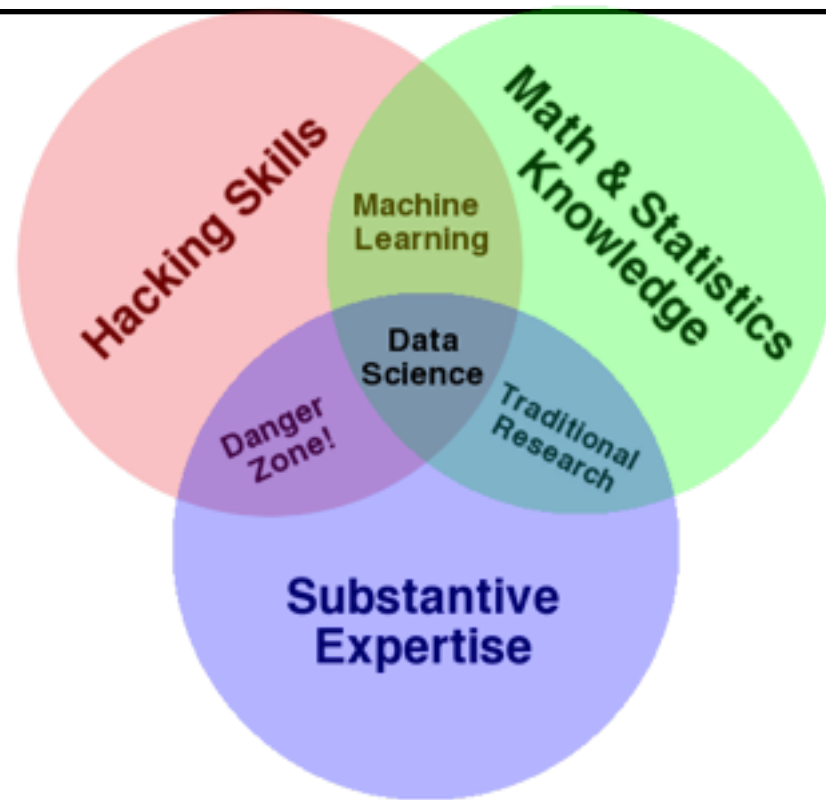
# WHAT IS A DATA SCIENTIST

---



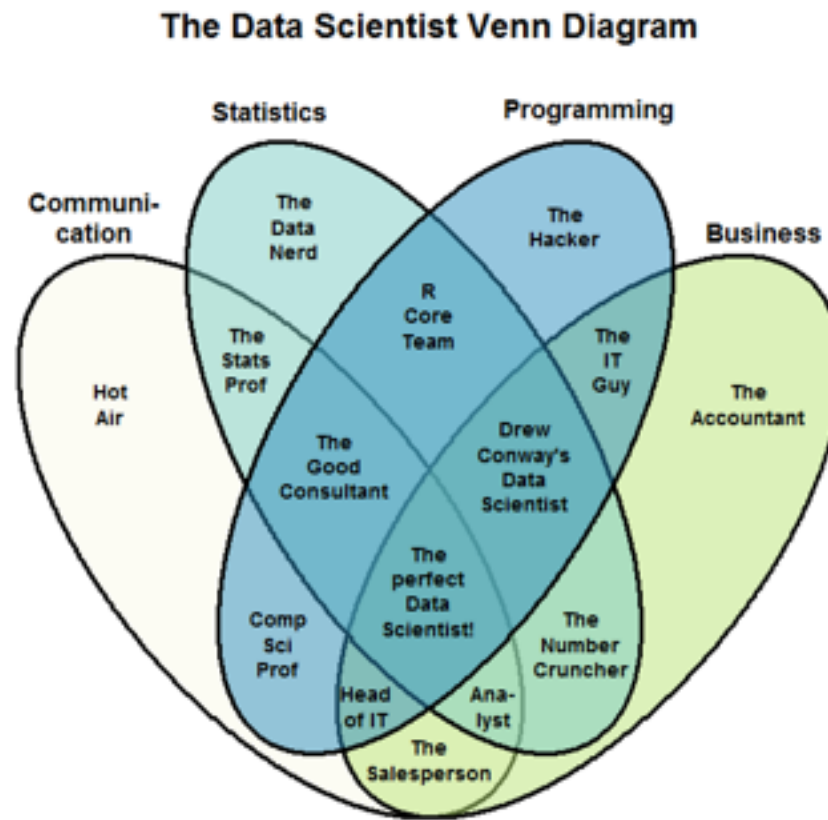
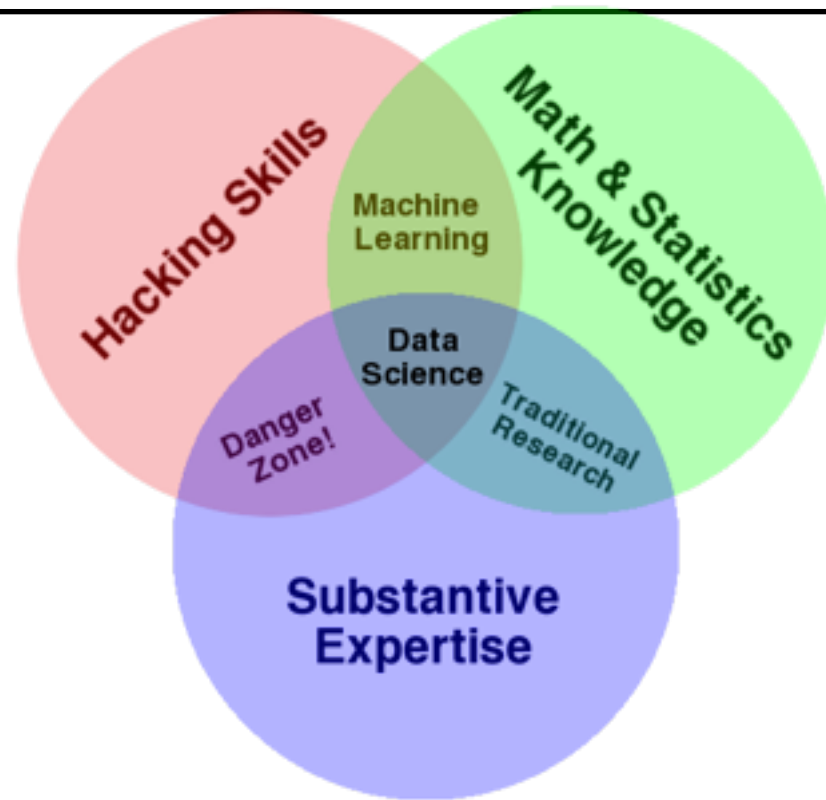
# WHAT IS A DATA SCIENTIST

---



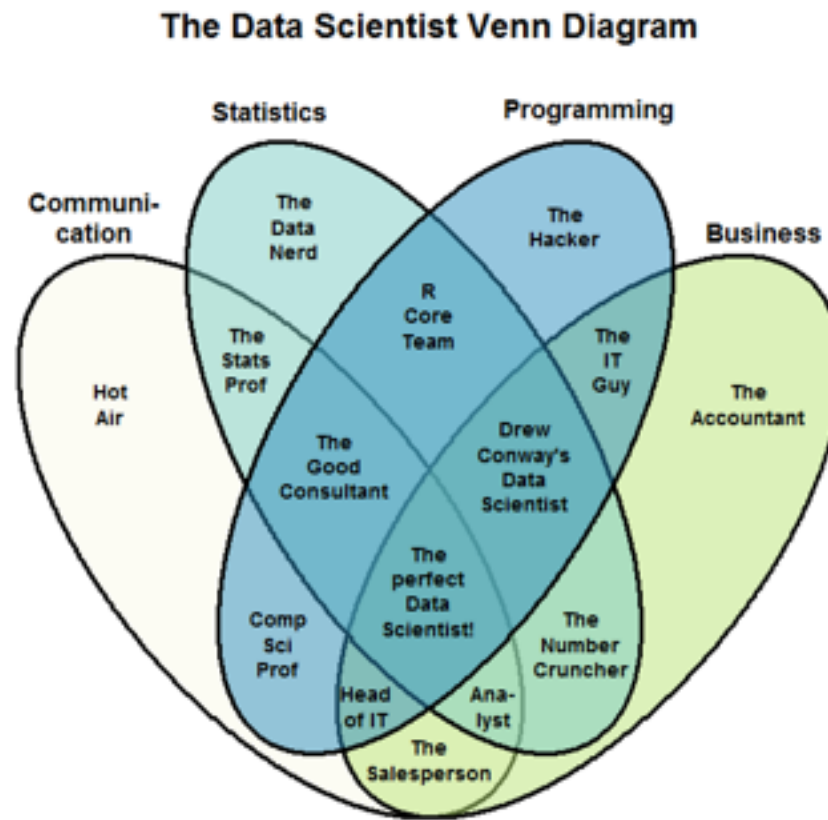
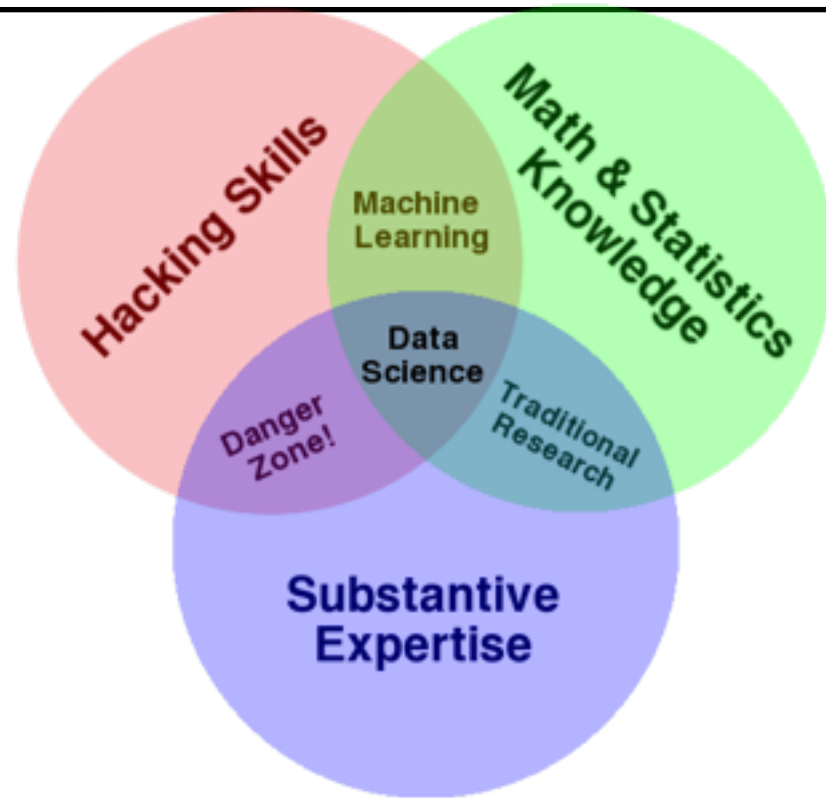


# WHAT IS A DATA SCIENTIST





# WHAT IS A DATA SCIENTIST



# 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

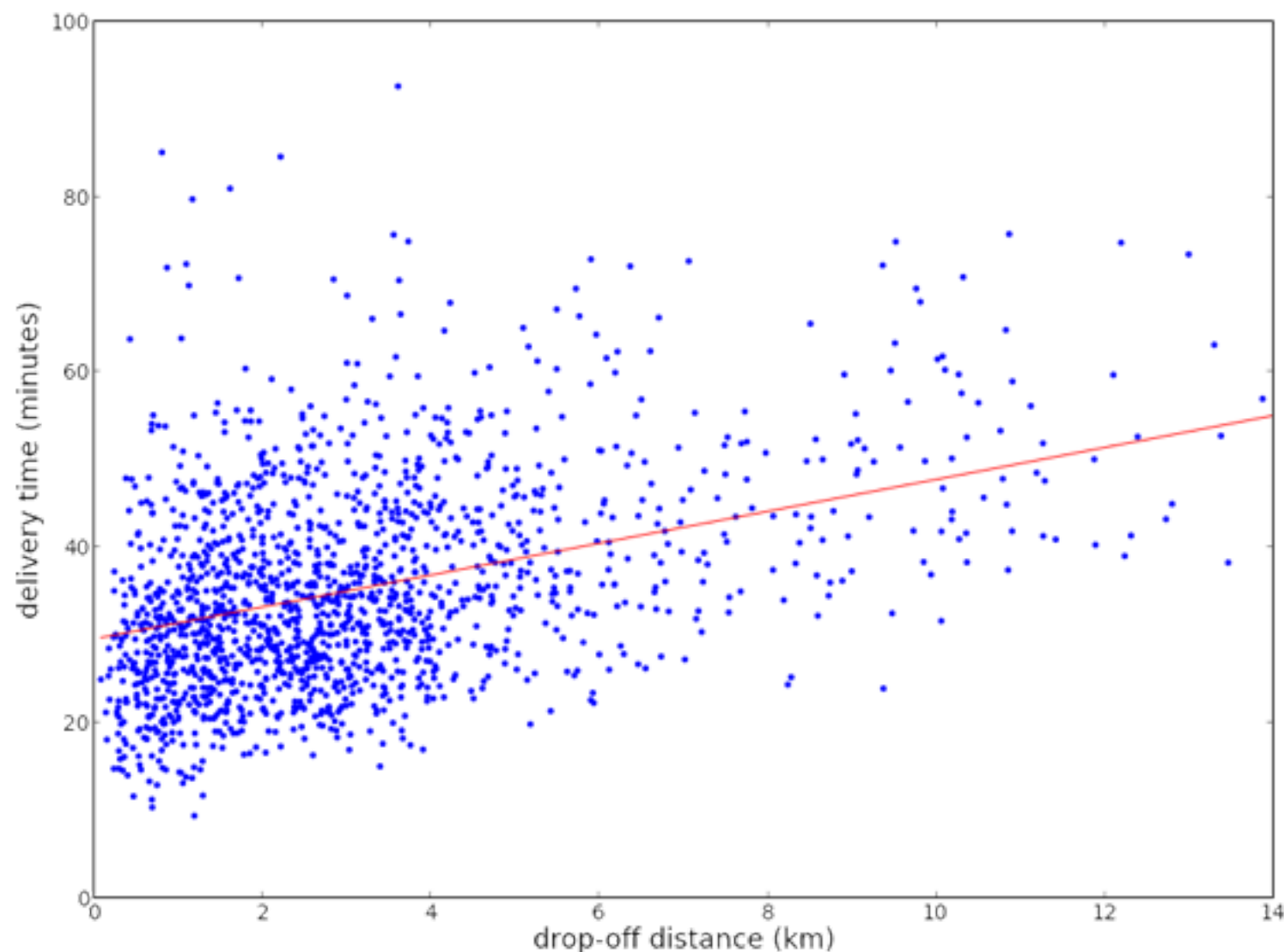
# SOME APPLICATIONS

---

## Estimating Delivery Times: A Case Study In Practical Machine Learning



# POSTMATES



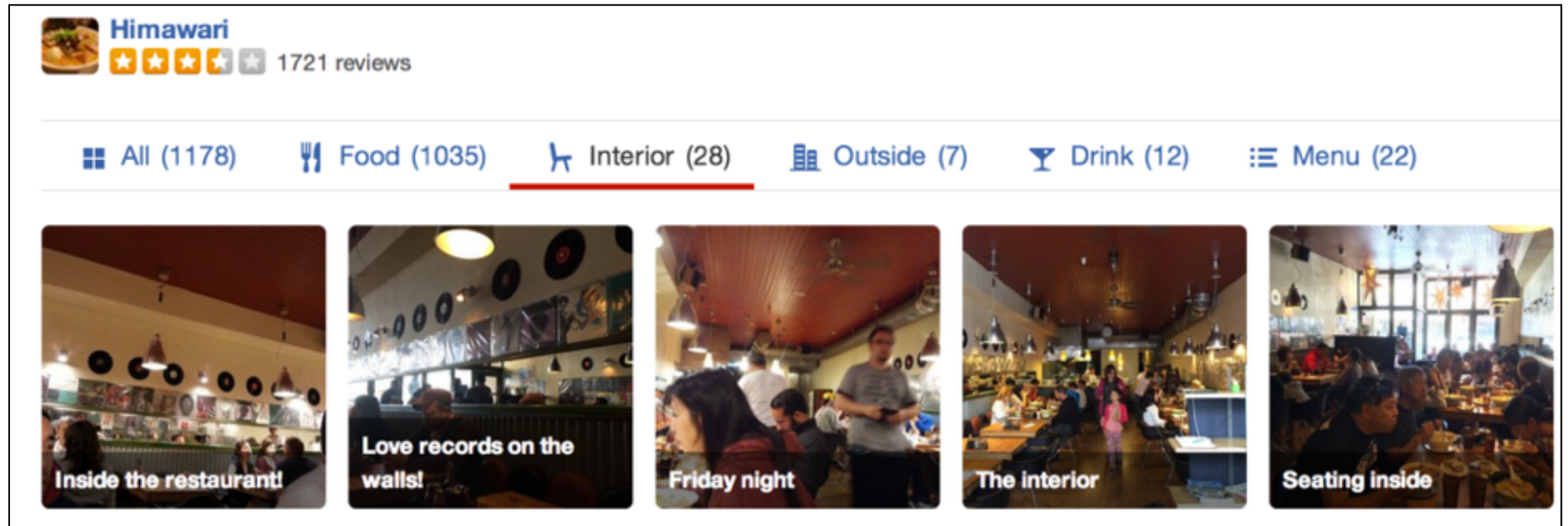
# SOME APPLICATIONS

---





# SOME APPLICATIONS



How We Use Deep Learning to Classify Business Photos at Yelp

A close-up photograph of a woman with a warm smile, looking down at a smartphone in her hands. She is wearing a dark jacket and a light-colored scarf with a bold, geometric pattern. The background is blurred, suggesting an office or indoor setting. The entire image has a greenish tint.

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>
  
- Higher-level tools:
  - [Tableau](#), [WEKA](#), [gretl](#)...

# THANK YOU

---

ThoughtWorks®