Data Science Career

MING LI

Disclaimer: the opinions expressed in this presentation is Ming's own and do not represent the view of Ming's employers.

Science

Physics

Statistics

Engineering

Electrical Engineering

“Statistical Engineering”

+

Big Data & Software

=

Data Science

Statis-tician	Data Scientist
Relatively focus on modeling (i.e. science)	Mainly focus on business problem & result (i.e. engineering)
Bring data to model	Bring models to data
Data is relatively small in size and clean in text file formats	Need to work with messy and large amount data in various formats
Usually structured data	Both structured & unstructured data
Usually isolated from production system	Usually embedded in production system

Two Types of Data Science Career

Generalist

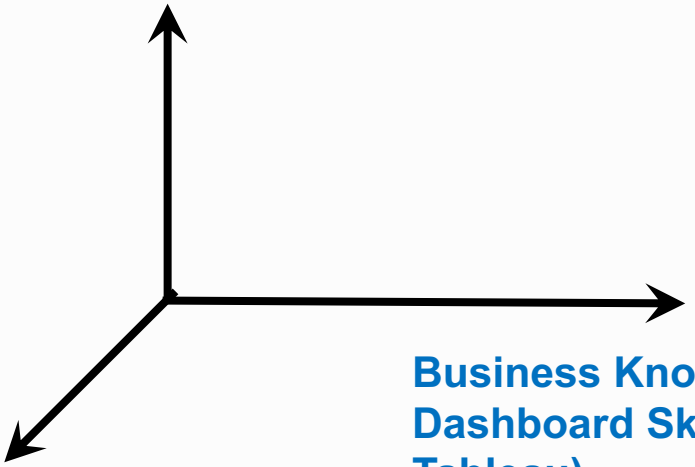
- Business problem formulation skills
- Data preprocessing skills
- Statistical and machine learning methods
- Deep learning applications knowledge
- Model building process experience

Focus on Generalist in this talk.

Specialist

- Natural language understanding
- Image and video analysis
- Voice recognition
- Language translation
- AI chat-bot

Three Pillars of Skills



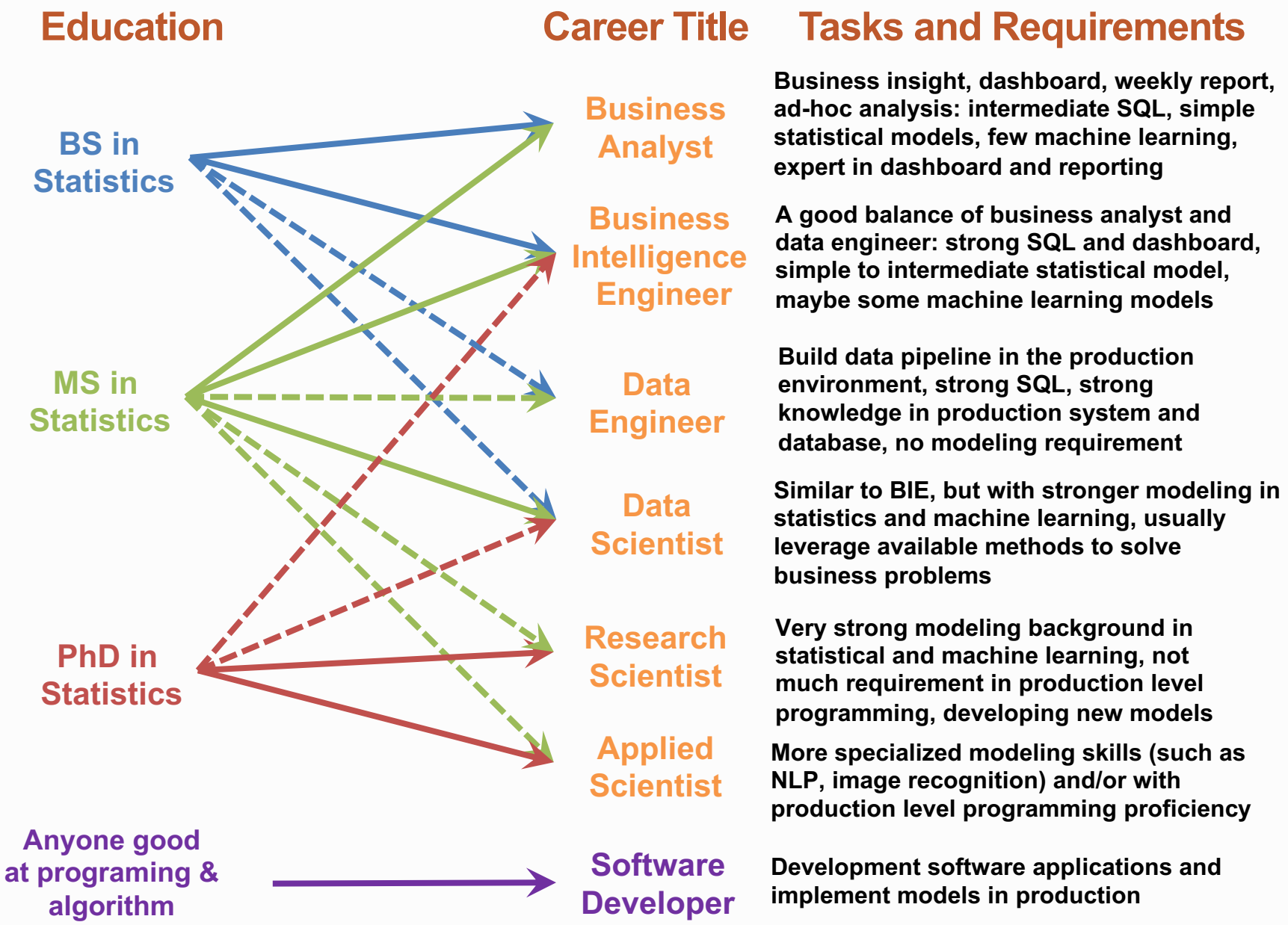
Recent trends of Cloud-Based Solution lower the barrier for statistician drastically in this area

If you are good at all three areas, you become a "***full-stack***" data scientist! The best of the best!

Two Types of Data Science Teams

- **Embedded within Business Team**
  - Majority of the data scientists belong to
  - Clear business objective for each project
  - Bridge business team and core machine learning team
- **Core Machine Learning Team**
  - Hard core machine learning scientists
  - Focus on specific machine learning algorithms
  - Provide models and tools for the production infrastructure

One of the key feature in Tech Companies is that DS are working closely with software developers. Often times DS/DE/SDE (i.e. the tech team) are in the same team to ensure a true collaboration with business team.



Each career path has its own promotion cycle.

Thank you!