

IS DATA SCIENCE STILL SEXY?

Ahmed Sherif

Tampa Bay Data Science Group
November 7, 2022

AGENDA

- We answer the question is Data Science still Sexy?
 - a. Why or Why Not?
 - b. What has changed in the last 10 years
- How do we adapt as a Data Science Community?
- How and where do you upskill?

A LITTLE ABOUT ME

- Sr. CSA at Microsoft since 2018
- Data Scientist since 2014
- Started as a Data Analyst in 2005
- Currently as a Software Engineer
- Author of two books:
 1. Practical Business Intelligence (2016)
 2. Apache Spark Deep Learning Cookbook (2018)



REMEMBER THIS ARTICLE?

Harvard
Business
Review

Analytics And Data Science | **Data Scientist: The Sexiest Job of the 21st Century**

Subscribe

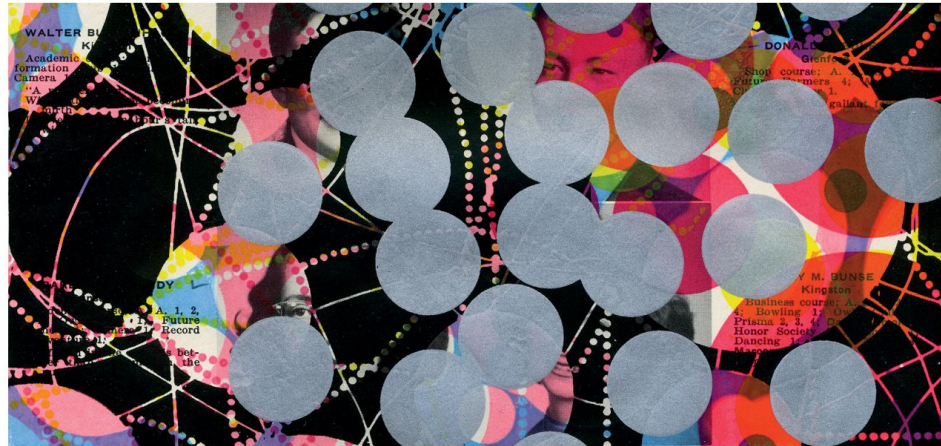
Sign In

Analytics And Data Science

Data Scientist: The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data. by Thomas H. Davenport and DJ Patil

From the Magazine (October 2012)



Andrew J. Buboltz, silk screen on a page from a high school yearbook, 8.5" x 12", 2011 Tamar Cohen

SO, IS IT STILL SEXY?



THE TERM IS NOT AS SEXY, BUT THE SKILLS ARE MORE NECESSARY

Data Science has matured since 2012

- More collaborative
- Less Siloed

Once a model is an API, it's just software!

THE STATS DON'T LIE

- 85% of big data projects fail (Gartner, 2017)
- 87% of data science projects never make it to production (VentureBeat, 2019)
- “Through 2022, only 20% of analytic insights will deliver business outcomes” (Gartner, 2019)

REASONS FOR FAILURE

- Within the last 10 years, there have been many failed “Data Science” projects and initiatives
- The field has taken a hit in reputation
- Why?

WHAT HAS GONE WRONG?

1. Lack of Data or worse
2. Lack of Business requirements
3. Lack of Deployments
4. Lack of Deployment Monitoring
5. Lack of Ethics
6. Slow Pipelines
7. Lack of reusability, explainability, or transparency

THE UNBEARABLE
WEIGHT OF MASSIVE
TALENT OF ROLES IN
DATA SCIENCE

ML knowledge

ML researcher
data scientist

ML-heavy role

data scientist

software-heavy role

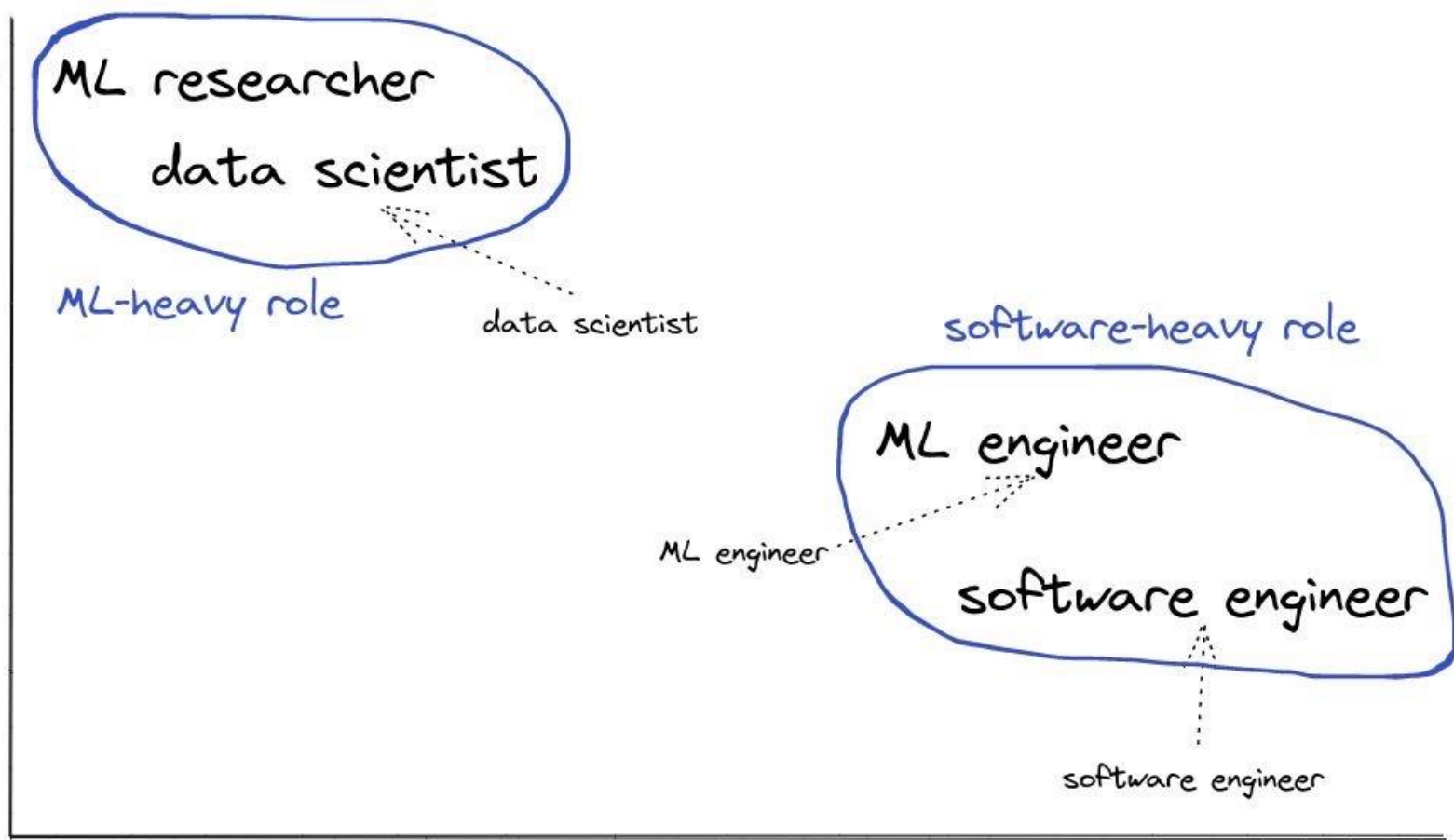
ML engineer

ML engineer

software engineer

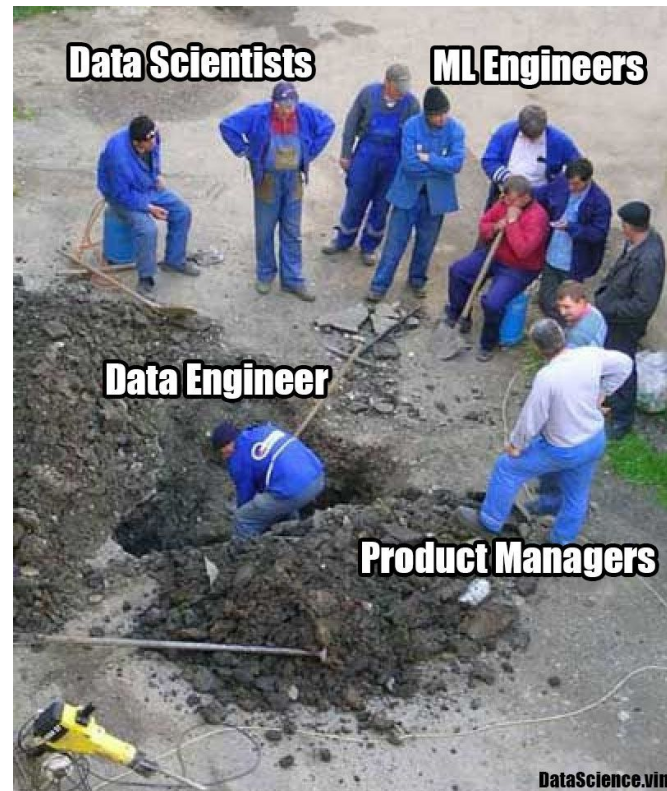
software engineer

software engineering knowledge



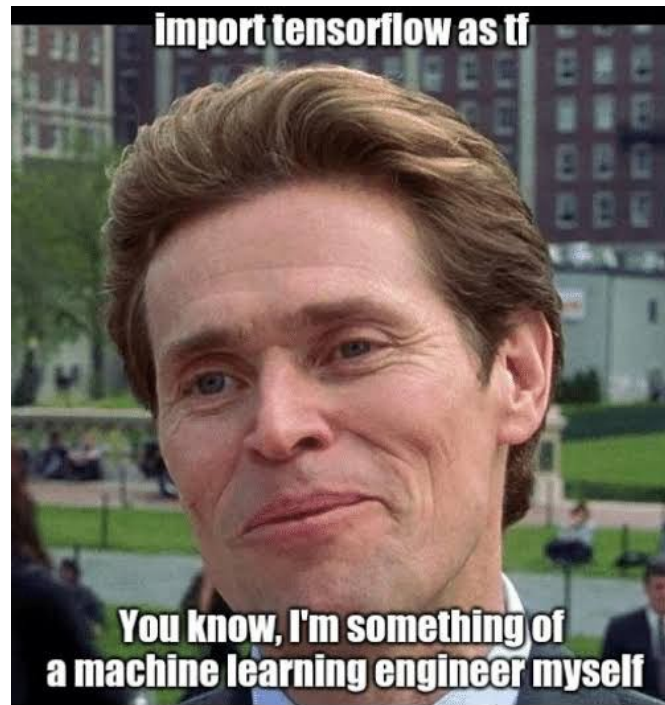
DATA ENGINEER

- Responsibility
 - Creating Data Pipelines to transform data from raw to procured
- Closest Adjacent Role
 - ETL Developer
- Skills
 - Data Quality / Data Architect
 - Programming in SQL and Python
 - PySpark / Pandas / SparkSQL
- Technologies
 - Distributed Computing i.e. Spark
 - Data Lake
 - Data Warehousing



MACHINE LEARNING ENGINEER

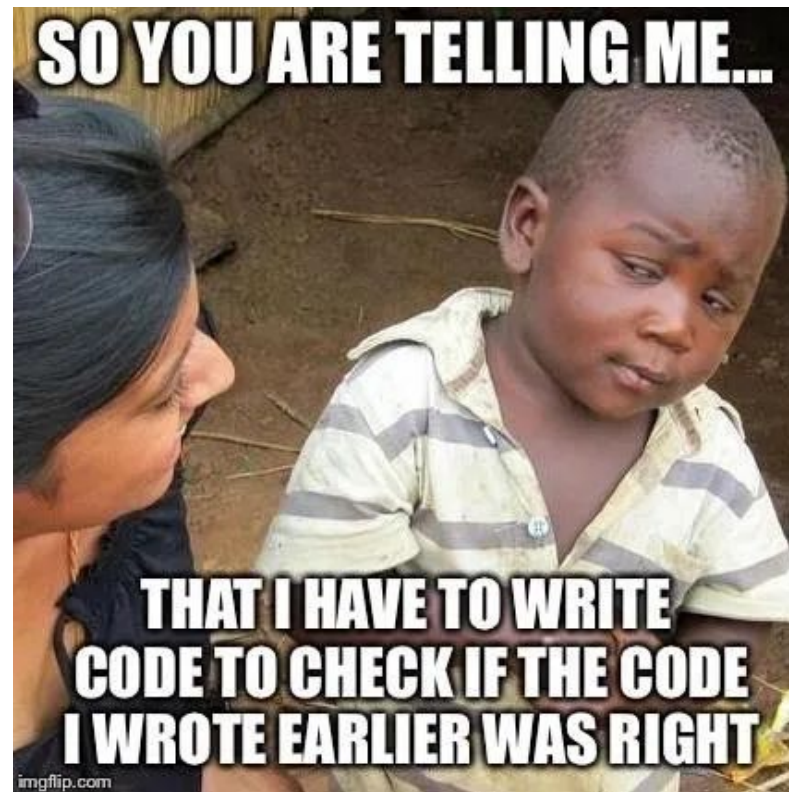
- Responsibility
 - Create Machine Learning pipelines
 - Scalability
 - Feature Engineering
- Closest Adjacent Role
 - Software Engineer
 - MLOps Engineer
- Skills
 - ML Algorithms
 - Model training
 - Programming in Python, R, Scala, Java
- Technologies
 - Cloud
 - Spark
 - Containers & Kubernetes
 - API



You know...

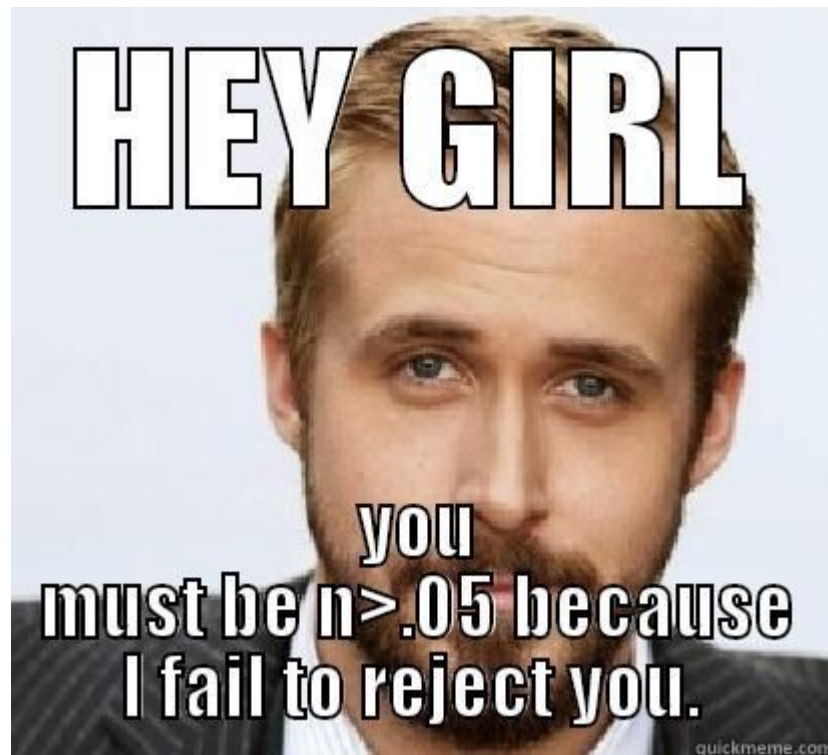
MLOPS ENGINEER

- Responsibility
 - Monitor Data Drift
 - Monitor Feature Degradation
 - Reliability
- Closest Adjacent Role
 - DevOps/Software/ML Engineer
- Skills
 - Modeling Monitoring & Tracking
 - ML Algorithms
 - CI / CD Pipelines
 - Unit Testing
- Technologies
 - mlFlow
 - Kubeflow
 - Azure DevOps
 - GitHub Actions



MACHINE LEARNING SCIENTIST

- Responsibility
 - Model Testing and Tuning
 - Algorithm Selection
 - Explainability, Transparency, and Interpretability
 - Ethics
 - Feature Engineering
- Closest Adjacent Role
 - Research Scientist
 - Statistician
- Skills
 - Python or R Programming
 - Sklearn, PyTorch, TensorFlow, Pandas
 - Deep Statistics knowledge
- Technologies
 - SHAP / FAIRLearn
 - Open Source DataViz / Tableau / D3
 - Spark
 - Grid Search
 - Model Metrics



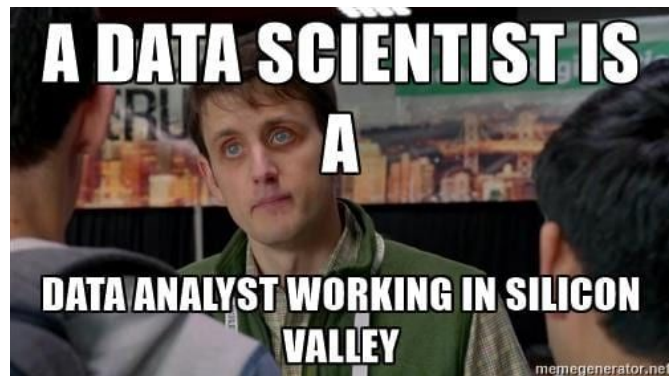
BONUS: FULL STACK DATA SCIENTIST

- Responsibility
 - A little bit of everything
 - Data Engineer / ML Engineer
 - Data Analyst / Data Scientist
- Closest Adjacent Role
 - Machine Learning or Data Science Manager
- Skills
 - Everything prior
 - Web Development
- Technologies
 - Everything Prior
 - JavaScript / HTML / D3 / React
 - Streamlit for Python
 - Shiny for R or Python?
 - Flask, FastAPI, Django



BONUS: DATA ANALYST

- Responsibility
 - Collects, cleans, and analyzes data
 - Prescriptive analytics
 - Creates actionable steps from business requirements
- Closest Adjacent Role
 - Business Analyst
- Skills
 - SQL
 - Pandas or Tidyverse
 - Documentation
- Technologies
 - Tableau / Power BI
 - Spreadsheets



ADDITIONAL ROLES

- AI Engineer
- Data Science Manager

IS IT FAIR TO LAY THE BURDEN OF ALL OF THESE SKILLS ON ONE PERSON

- Data Quality / Data Architect
- PySpark / Pandas / SparkSQL
- ML Algorithms
- Model training
- Programming in Python, R, Scala, Java
- Modeling Monitoring & Tracking
- CI / CD Pipelines
- Unit Testing
- Sklearn, PyTorch, TensorFlow
- Deep Statistics knowledge
- Web Development
- Business Domain knowledge

ONLINE RESOURCES

FrontendMasters.com

AlgoExpert.io

Oreilly.com

FreeCodeCamp.org

MAINTAIN YOUR BRAND

Online Website / Blog

Linkedin Profile

GitHub.com Repo

THANK YOU!

Ahmed Sherif

Twitter: **@thatAhmedSherif**

LinkedIn: **<https://www.linkedin.com/in/thatahmedsherif>**