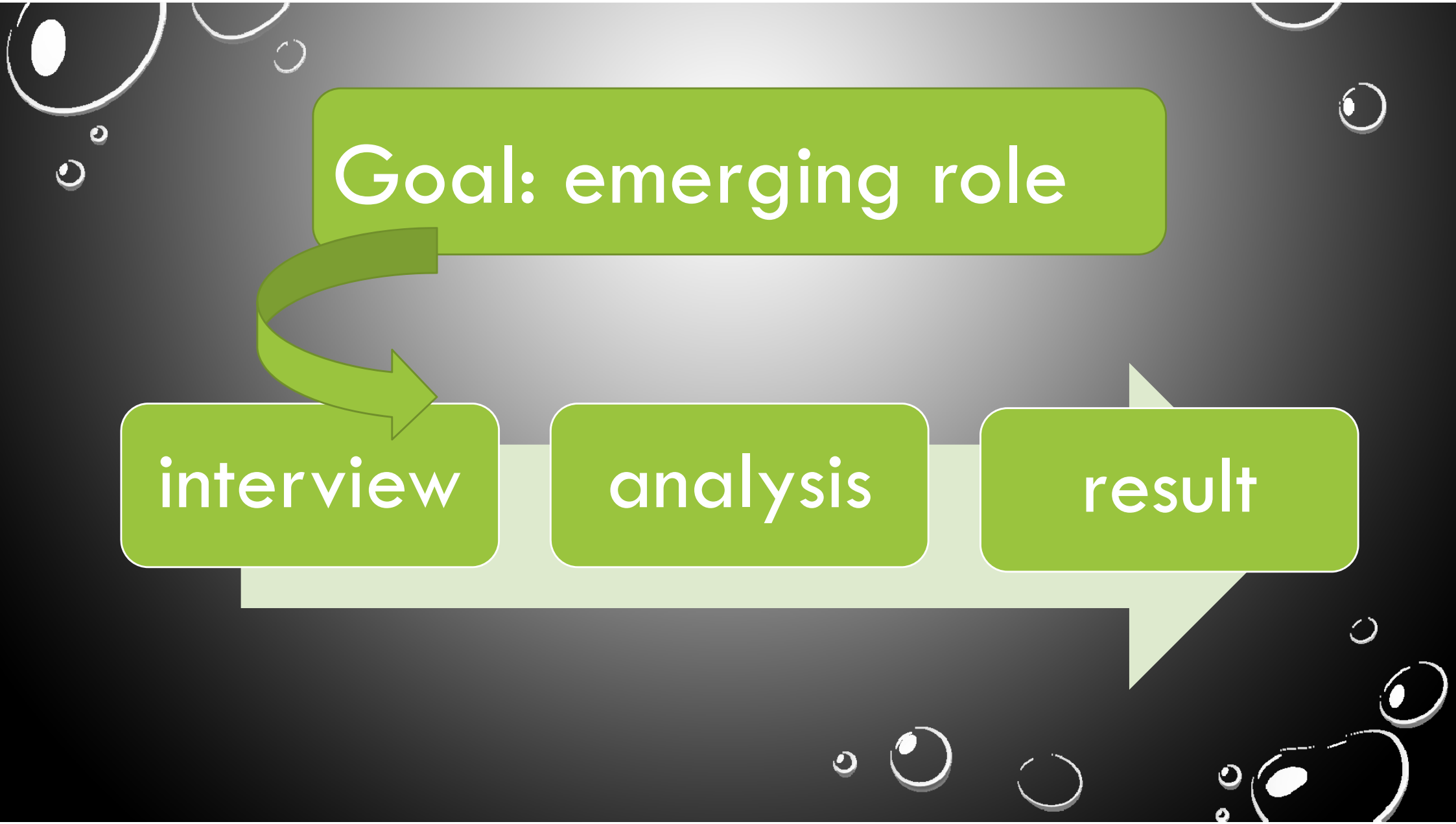
THE EMERGING ROLE OF DATA SCIENTISTS ON SOFTWARE DEVELOPMENT TEAMS

AUTHOR: MIRYUNG KIM UCLA

PRESENTER: SI CHEN SFU (301308915)


Goal: emerging role
interview
analysis
result

THE SEXIEST JOB IN 21ST CENTURY

Top one of the 25 Best Jobs in America 2016 released by Glassdoor

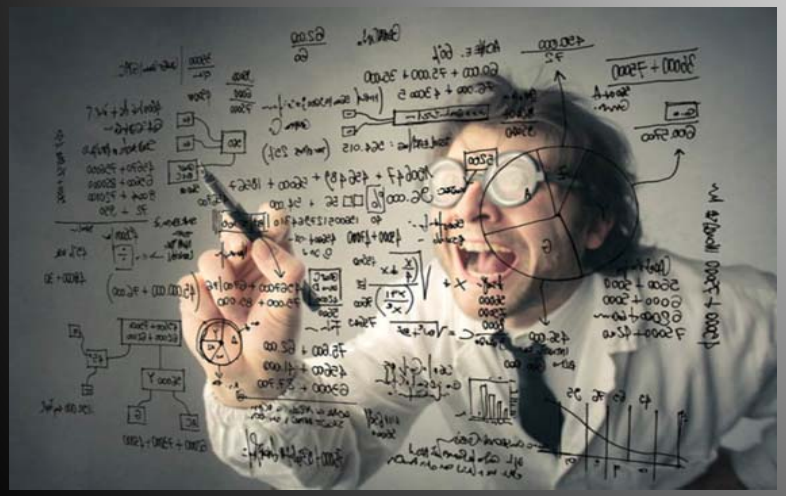
overwhelming demand 60% greater

rising salaries **$116,840 A YEAR**

top Glassdoor survey for best work-life balance

really challenges

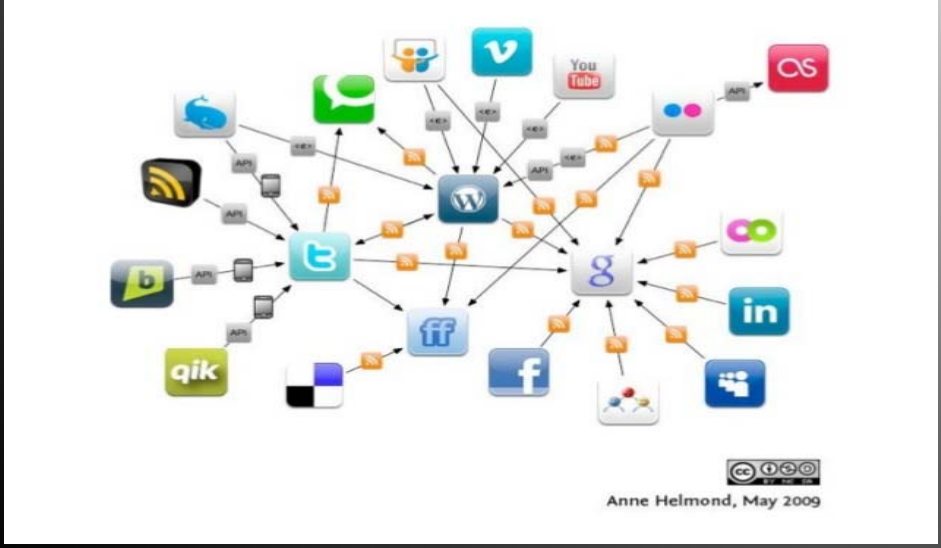
WHO DATA SCIENTISTS ARE ?



- Definition: The people who do *COLLECTION* & *ANALYSIS*
- Main mission : transform data into insight, providing guidance for leaders to take action
- Example: the use of user telemetry data to redesign Windows Explorer (a tool for file management) for Windows 8.

WHY IT IS NEEDED IN SOFTWARE DEVELOPMENT TEAMS

THE TREND OF DATA-DRIVEN ENGINEERING DECISIONS (RATHER THAN RELYING ON GUT FEEL)



DATA==PROFITS

THE VALUE OF USERS' DATA

EMERGING ROLE

- Experimental design
- Statistical reasoning
- Data collection

EXPERIMENTAL DESIGN

DESIGNING EXPERIMENTS WITH REAL USER DATA :

Customer usage data is easier to obtain and more authentic

DEMAND FOR STATISTICAL RIGOR

- formal hypothesis testing,
- report confidence intervals,
- determine baselines through normalization.

DEMAND FOR DATA COLLECTION RIGOR

- Data cleaning
- data shaping

80% of data science work requires "janitor work"

WHAT DO THEY DO IN DAILY LIFE

- PROBLEMS THAT THEY WORK ON:

Performance Regression

Server Anomaly Detection

Requirements Identification

Fault Localization and Root Cause Analysis

Failure Rate Estimation

Bug Prioritization

Cost Benefit Analysis

Customer Understanding

WHAT DO THEY DO IN DAILY LIFE

- MAIN ACTIVITIES:

data collection

data analysis

data use and dissemination

TABLE 2. ACTIVITIES THAT PARTICIPANTS STATED THEY DID THEMSELVES (■) OR MANAGED (□)

		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Collecting	Building the data collection platform	■			■				■			■		■	□		
	Injecting telemetry	■	□		■				■		□	■		■	□		
	Building the experimentation platform	■													□		
Analyzing	Data merging and cleaning	■	■	■	■	■	■	■	■	■	□	■		■	□		
	Sampling	■	■	■	■	■	■	■		■	□	■	■	■	□	■	■
	Shaping, feature selection		■	■	■	■	■	■			□		■	■	□	■	■
	Defining sensible metrics	■			■	■	■	■			□	■		■	□		■
	Building predictive models		■	■		■	■	■		■	□		■	■	□	■	■
	Defining ground truth							■		■			■	■	□	■	■
	Hypothesis testing		■	■		■	■				□				□		■
Using and Disseminating	Operationalizing models						■	■		■	□		■	■	□	■	
	Defining actions and triggers									■	■	■		■	□		
	Applying insights/models to business	■	■	■	■	■				■	■	■	■	■	□	■	■

HOW DO THET EVALUATE THEIR OWN WORK?

- new features
- deprecate unused features
- releasing a product two weeks earlier than the expected schedule
- reduce crash rates
- reducing development operation cost

A UNIQUE PATH TO THEIR EMERGING ROLE

- INSIGHT PROVIDER
- MODELING SPECIALISTS
- PLATFORM BUILDERS
- POLYMATHS
- TEAM LEADERS

INSIGHT PROVIDER

- interstitial role between managers and engineers
- Statistic knowledge
- communication and coordination skills

MODELING SPECIALISTS

- build predictive models
- strong background in machine learning
- “translate” findings into business values

PLATFORM BUILDERS

- build a data engineering platform
- strong background in big data systems

POLYMATHS

- “do it all”
- Unclear Boundary

TEAM LEADERS

- senior data scientists who run their own data science teams
- act as data science “evangelists,”
- work with senior company leaders

DO WE HAVE A CHANCE?

- **BY 2018, THE U.S. MAY FACE A SHORTAGE OF AS MANY AS 190,000 PEOPLE WITH ANALYTICAL EXPERTISE AND OF 1.5 MILLION MANAGERS AND ANALYSTS WITH THE SKILLS TO MAKE DATA-DRIVEN DECISIONS, ACCORDING TO A REPORT BY THE MCKINSEY GLOBAL INSTITUTE**

WHO THEY ARE HIRING?

- REQUIREMENTS OF EDUCATION AND TRAINING BACKGROUND:

4 years cs degree(11/16): not necessary

interdisciplinary backgrounds : statistics, physics, math…

higher education degrees(11/16): MS & PhD

Skill sets: self-trained

strong statistics background: some requires

NEW WORLD IS RULED BY DATA !