Data Science Pathways

"Increasingly, businesses are recognizing data science as a tool for gaining valuable strategic insight. As a result, data science has a projected job growth of over 38% in the next 10 years." - Dice Tech Job Report 2020

Duane Rinehart & Wendy Porter
ICT-DM Regional Directors

Discussion Topics

1. Data Science Overview
2. Definitions: job titles, functions, KSA's
3. Existing CCC Curriculum and resources
4. Labor Market Info
5. Employer Engagement Opportunities

Data Science Overview

29-68% increase in total job openings[1]

- Most "data science" jobs concentrated in California (2019 job postings)
- Industries Growing at 23% y/y[2]
  - Most jobs in **San Francisco** (14% of statewide job postings), Los Angeles, San Diego, San José, Sunnyvale
  - Most growth in **San Diego** (37% growth in job postings), Sunnyvale, Los Angeles, Irvine
- Largest employers (statewide): Northrop Grumman, Amazon, Google, Walmart, Lockheed Martin, Accenture
- Key Drivers (based on usage in industry – 23-36% growth)[3] & Major Players:
  - "Big" Data in voice recognition – **Amazon: Echo & Alexa**, Google Home, Apple HomePod, Protonet
  - IoT (i.e., sensors, remote monitoring and analytics for "smart" cities) - **Cisco**, IBM, Intel, PTC, Schneider Electric, Honeywell
  - Military Defense and Intelligence - **Northrop Grumman**, Lockheed Martin, Boeing, Raytheon, BAE Systems, General Dynamics, L3Harris Technologies, Huntington Ingalls (Continental Maritime of San Diego), Booz Allen Hamilton
  - Telecommunications – **Qualcomm**, Verizon, AT&T, Sprint, Wipro, Comcast
  - Insurance & Advertising – Infosys, DXC
  - Financial Services – Capital One, Visa, American Express, JP Morgan
  - Healthcare - **United Healthcare Group**, Iqvia, Centene, Anthem Blue Cross
  - Integrators - **Jacobs Engineering Group**, Perficient, Calance

Sources: 1) 29% from Indeed.com (2013-2019), 68% from Burning Glass, "New Foundational Skills" (2012-2017), 50% (2019-2020 Y-o-Y), Dice Tech Report; 2) Labor Insight Jobs (Burning Glass Technologies), 2020; 23% CAGR (2015-2020) from MarketsandMarkets, 36% from Dresner Advisory Services, "2018 Big Data Analytics Market Study"

Top 3 Priority Areas for CCC's

1. Data collection / aggregation / storage (web scraping, database aggregation, file imports & manipulation)

2. Data Analysis (look at data; answer business questions)

3. Data Visualization

Key Functions & KSA's

Data collection / aggregation / storage (web scraping, database aggregation, file imports & manipulation)

- File systems (especially for large data such as ReFS, ZFS, BtrFS) & knowledge of data governance
- Programming (python[2], Java, C / C++ / C#, javascript, R[1])
- Ability to perform database queries, joins, etc. (SQL, JSON, MS Access)
- Cloud data storage/extraction (JSON, AWS, Google, Azure)
- Data cleaning (text manipulation)
- Apache Hadoop (hybrid filesystem / database)
- Familiarity with ETL (Extraction, Transformation & Loading) concepts

Representative Job titles: Data Engineer, Database Administrator, Cloud Technician, Data Warehouse Technician, Big Data Software Developer

Sources: 1) Ability to install libraries from CRAN is valued; 2) Use of pandas, numpy and related modules is required [for coding data]; 3) towardsdatascience.com/; 4) r4stats.com/; 5) Recently acquired by Salesforce – consolidation may spur additional skill requirements such as knowledge of Salesforce API

Key Functions & KSA's

Data analysis (look at data; answer business questions)[4]

- Basic math, statistics [and ability to use software application] (R, SPSS, Statistica, Matlab, SAP)
- Programming (python[2], Java, C / C++ / C#, javascript, R[1])
- Ability to perform database queries, joins, etc. (SQL, JSON)
- Excel (including pivot tables)
- Data portal usage (Qlik View, SAP Business Objects, SAS, Tableau)
- Critical thinking about the business implications (interpret the data), and telling the story with the data

  Representative Job titles: Data Analyst, Database Administrator, Business Intelligence Analyst, Systems Analyst, Marketing Analyst, Operations Analyst

Sources: 1) Ability to install libraries from CRAN is valued; 2) Use of pandas, numpy and related modules is required [for coding data]; 3) towardsdatascience.com/; 4) r4stats.com/; 5) Recently acquired by Salesforce – consolidation may spur additional skill requirements such as knowledge of Salesforce API

Key Functions & KSA's

Data Visualization (create presentations for decision making stakeholders)[3, 4]

- End-user portal creation (Tableau[5], GIS, Geo4w, d3.js, SAP Business Objects, PowerBI, Datatables)
- Ability to connect to databases or cloud storage [to setup visualization solution] (SQL, JSON, d3.js)
- Ability to format and customize interactive graphs, reports (HTML5/CSS, ECMA Script/javascript)

Representative Job titles: Data Visualization Specialist,

Other (more advanced)

- Amazon ML (machine learning)
- Hadoop, Apache Hive
- Apache Spark (clustered computing)
- Golang, Scala (programming)
- Understanding of machine learning / artificial intelligence concepts, (un)supervised models

Sources: 1) Ability to install libraries from CRAN is valued; 2) Use of pandas, numpy and related modules is required [for coding data]; 3) towardsdatascience.com/; 4) r4stats.com/; 5) Recently acquired by Salesforce – consolidation may spur additional skill requirements such as knowledge of Salesforce API

Current Known CCC Landscape

- Most CA Colleges: Some data classes in Math / MIS / CS / CIS
- OC Region 9 colleges K14 Data Analytics SWF funded program - Reyna Wolzinger loves our methodologies; she wants to collaborate. (Alteryx is industry contact)
- Coastline – Data Analytics (AS and Certificates), Data Science (Certificate)(1)
- San Bernardino College: CIS & Programming Certificates of Achievement include Database Concept & Design.
- Lassen College: Data Science Certificate
- Chaffey College has programs
- Grossmont – Data Analytics (Certificate)(2)
- Palomar – (AS – CIS with emphasis in Data Analytics)(2)
- Miracosta – GE statistics using R
- Fresno State: Data Analytics Pathway (good content available)

Sources: 1) "Science" refers to programming component; 2) Includes python programming component; Includes visualizations

Other Efforts

1. Ann Behler CTC Data Analytics Project
   ITSS Job Cluster List document: Data Management and Engineering and Data Analytics and Predictive Modeling

2. Duane hosting similar workshop in San Diego (in conjunction with National Convergence Technology Center, NFS)

3. **Industry-recognized certifications:**
   Certified Analytics Professional (CAP), Google Certified Professional Data Engineer, IBM Data Science Professional Certificate, Microsoft MCSE: Data Management and Analytics, Microsoft Certified Azure Data Scientist Associate, SAS Certified Big Data Professional and related[(3)]

Sources: 3) Additional: https://www.cio.com/article/3222879/15-data-science-certifications-that-will-pay-off.html

Learning Resources

Opportunities to learn outside classroom:

1. Kaggle.com sponsors data science competitions
2. Meetup Groups (Data Science for Non-Profits, She Loves Data)
3. LinkedIn Learning

LMI & Other Data Resources

1. Demand for Data jobs report (DICE):
2. CCN Member Sinclair Community College Data Visualization Program
3. COCI Program Export - Karen B
4. Course Title Spreadsheet 2016-19 - Karen B
5. ITSS Job Cluster List
6. Shared Drive Link to Docs & More
7. COE Data Science LMI Reports (Far North, Inland Empire, Bay Area, LA)
   Sara Philips provided a statewide report via the next slides.

COE LMI Report Data Science CA

The following Standard Occupational Classification (SOC) codes were analyzed:

15-1132.00 - Software Developers, Applications

Develop, create, and modify general computer applications software or specialized utility programs. Analyze user needs and develop software solutions. Design software or customize software for client use with the aim of optimizing operational efficiency. May analyze and design databases within an application area, working individually or coordinating database development as part of a team. May supervise computer programmers.

Sample of reported job titles: Application Developer, Application Integration Engineer, Applications Developer, Computer Consultant, Information Technology Analyst (IT Analyst), Software Architect, Software Developer, Software Development Engineer, Software Engineer, Technical Consultant

15-1151.00 - Computer User Support Specialists

Provide technical assistance to computer users. Answer questions or resolve computer problems for clients in person, or via telephone or electronically. May provide assistance concerning the use of computer hardware and software, including printing, installation, word processing, electronic mail, and operating systems.

Sample of reported job titles: Computer Specialist, Computer Support Specialist, Computer Technician, Desktop Support Technician, Help Desk Analyst, Help Desk Technician, Information Technology Specialist (IT Specialist), Network Technician, Support Specialist, Technical Support Specialist

15-1152.00 - Computer Network Support Specialists

Analyze, test, troubleshoot, and evaluate existing network systems, such as local area network (LAN), wide area network (WAN), and Internet systems or a segment of a network system. Perform network maintenance to ensure networks operate correctly with minimal interruption.

Sample of reported job titles: Computer Network Specialist, IT Consultant (Information Technology Consultant), Network Engineer, Network Specialist, Network Support Specialist, Network Technical Analyst, Network Technician, Personal Computer Network Analyst, Senior IT Assistant (Senior Information Technology Assistant), Systems Specialist

43-9111.00 - Statistical Assistants

Compile and compute data according to statistical formulas for use in statistical studies. May perform actuarial computations and compile charts and graphs for use by actuaries. Includes actuarial clerks.

Sample of reported job titles: Actuarial Analyst, Actuarial Assistant, Actuarial Technician, Administrative Analyst, Assistant Statistician, Data Analyst, Research Analyst, Research Assistant, Statistical Clerk, Statistical Technician

Employment and projected occupational demand

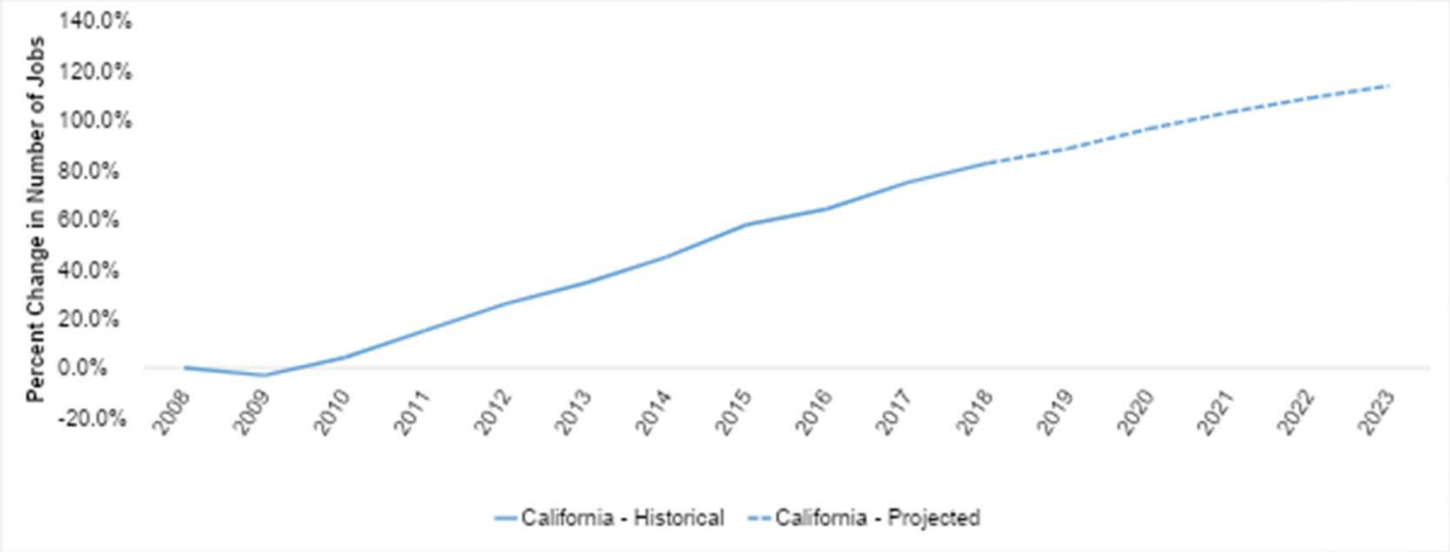
This table summarizes job trends for occupations related to data science in California.

Occupation	2008 Jobs	2018 Jobs	2023 Jobs	2018-23 Jobs Change	2018-23 Jobs % Change	Annual Openings
Software Developers, Applications	73,593	155,664	187,661	31,997	21%	18,252
Computer User Support Specialists	51,164	81,981	92,139	10,158	12%	9,125
Computer Network Support Specialists	14,949	19,493	21,487	1,994	10%	2,073
Statistical Assistants	1,759	1,237	1,352	115	9%	182
Total	141,465	258,375	302,639	44,264	17%	29,632

Source: Emsi 2020.1 - QCEW Employees, Non-QCEW Employees and Self-Employed

Rate of change for selected occupations

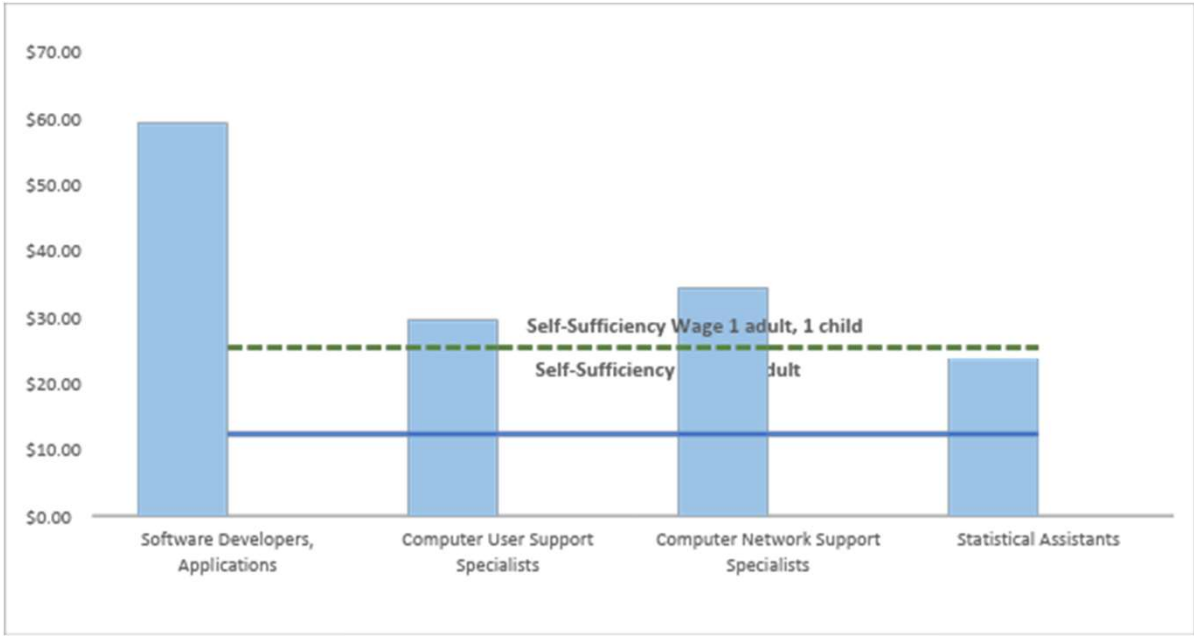
This chart displays the percentage change in number of jobs between 2008 through 2018 and occupational projections from 2018 through 2023. The rate of change is indexed to the total number of jobs in 2008 as the base year.



Source: Emsi 2020.1; QCEW Employees, Non-QCEW Employees and Self-Employed.

Wages

This chart displays median hourly wages for selected data science occupations in California compared to median self-sufficiency wages for a one-adult, one-child household and for one-adult households.



Sources: Median Hourly Wages: Emsi 2020.1; QCEW Employees, Non-QCEW Employees and Self-Employed.
Self-Sufficiency Wage: *The Self-Sufficiency Standard for California 2018*, Center for Women's Welfare, University of Washington. More information at http://www.selfsufficiencystandard.org/california or contact pearce@uw.edu.

Wages

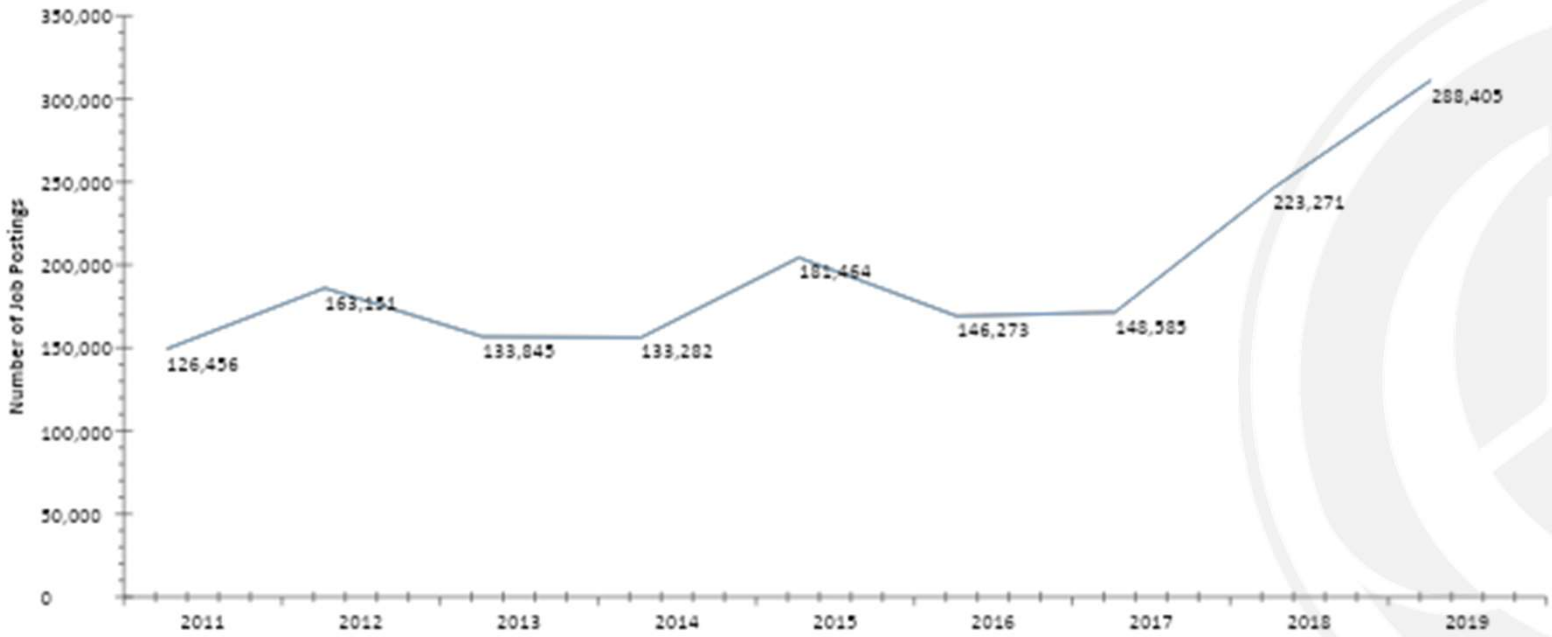
This table displays entry level, median, and experienced hourly wages for selected data science occupations in California compared to the average living wages for a one-adult, one-child household and for one-adult households.

Occupation	Entry-level (Pct. 10 Hourly Earnings)	Median	Experienced (Pct. 90 Hourly Earnings)
Software Developers, Applications	$34.33	$59.25	$91.03
Computer User Support Specialists	$17.83	$29.55	$49.87
Computer Network Support Specialists	$20.22	$34.39	$57.50
Statistical Assistants	$11.60	$23.75	$36.04

Sources: Median Hourly Wages: Emsi 2020.1; QCEW Employees, Non-QCEW Employees and Self-Employed.
Self-Sufficiency Wage*: The Self-Sufficiency Standard for California 2018*, Center for Women's Welfare, University of Washington. More information at http://www.selfsufficiencystandard.org/california or contact pearce@uw.edu.

Job Posting Trends

This chart shows job postings since 2011 for selected occupations related to data science in California.



Source: Burning Glass Technologies, "Labor Insights Real-Time Labor Market Information Tool" 2020.

Data science skill taxonomy

The data science skill cluster comprises 24 skills. Skills are sorted based on demand.

Data Science	45,179	80.7%
Predictive Models	6,628	11.8%
Predictive Analytics	4,464	8.0%
Pandas	3,844	6.9%
Cluster Analysis	2,448	4.4%
Pattern Recognition	1,117	2.0%
Time Series Models	659	1.2%
Ontologies	596	1.1%
Factor Analysis	563	1.0%
Information Extraction	407	0.7%
Time Series Forecasting	407	0.7%
Monte Carlo Simulation	286	0.5%
Naive Bayes	280	0.5%
Graph-Based Algorithms	247	0.4%
Bayesian Methods	244	0.4%
Social Network Analysis	221	0.4%
K-Means	186	0.3%
Bayesian Networks	157	0.3%
Principal Component Analysis (PCA)	136	0.2%
Bayesian Modeling	76	0.1%
Chi Square Automatic Interaction Detection (CHAID)	74	0.1%
Markov Chains	44	0.1%
Stochastic Optimization	40	0.1%
Kernel Methods	36	0.1%

Data science related skills

This table shows the top co-occurring skills when searching for job postings requiring “data science” as a primary skill. The skills below come from job postings.

55,986
job postings requiring “data science” as a skill.
(Feb 1, 2019 – Jan 31, 2020)

Specialized Skills	Share
Data Science	81%
Python	51%
SQL	41%
Machine Learning	37%
Data Analysis	21%
Big Data	20%
Java	20%
Apache Hadoop	18%
Software Engineering	17%
Tableau	16%

Baseline Skills	Share
Communication skills	41%
Teamwork/collaboration	37%
Research	31%
Problem solving	23%
Creativity	17%
Writing	16%
Detail-oriented	13%
Microsoft Excel	13%
Planning	12%
Presentation skills	10%

Software & Programming Skills	Share
Python	51%
SQL	41%
Java	20%
Apache Hadoop	18%
Software engineering	17%
Tableau	16%
Microsoft Excel	13%
Software development	12%
Apache Hive	10%
Data visualization	10%

C·O·E CENTERS OF EXCELLENCE
Inform Connect Advance

For more information, please contact:

Sara Phillips

Director, Far North Center of Excellence for Labor Market Research

sphillips@shastacollege.edu or 530.242.7635

Employer Engagement Ideas

1. Partner with OC Regional Efforts (Renah - she has SWF resources to help)
   a. They have already done 2 years of research
   b. We can utilize their surveys etc.

2. Regional Directors work with their local industry contacts
   a. Validate & update info [LMI, job titles, KSA] as appropriate

3. Add topic to ICT Webinar Series

Data Science Pathways

Open Discussion with team…

- Data Science Specifics
- Other employer engagement ideas?
- Process on pathway development?

Duane Rinehart & Wendy Porter
ICT-DM Regional Directors