

DATA PIONEERS SCIENCE

CONQUERING THE NEXT FRONTIER

BROUGHT TO YOU BY



RUNTIME:

1 HOUR

LANGUAGE:

ENGLISH WITH ENGLISH AND
FRENCH, SPANISH AND
PORTUGUESE SUBTITLES.

CONTACT:

IN NORTH AMERICA:

ALI DONZANTI, FOR DATAIKU
+1.646.428.0620
dataiku@allisonpr.com

IN FRANCE:

L'AGENCERP POUR DATAIKU
EVA GLÉLÉ | FLORENCE HANQUIEZ |
AUDREY VILFROY
+33 1 84 76 10 37
dataiku@lagencerp.com

IN THE UNITED KINGDOM:

COMMSGO FOR DATAIKU
LEAH JONES
+44 203 697 6680

IN THE NETHERLANDS:

CHANTAL SCHEPERS
ITSAREP
+31 6 235 099 23

chantal@itsarep.nl

SYNOPSIS



LOGLINE

With humor and humanity, DATA SCIENCE PIONEERS presents a fireside chat-style documentary about the passionate data scientists driving us towards technological revolution.

SYNOPSIS

Cut through the hype with DATA SCIENCE PIONEERS and see what it really means to be a data scientist. Even though the name is new, the role has a long lineage in statistics and information sciences. And while the challenges data scientists now face occur on an unprecedented scale today, they stand on the shoulders of technical and ethical quandaries that came before (after all, the question of self-driving cars is the trolley problem made real).

In DATA SCIENCE PIONEERS, real data scientists contextualize the role and why they love it while exploring the concerns and challenges to come. DATA SCIENCE PIONEERS is proof that data scientists aren't just data-crazy mathematicians that happen to work in a cutting-edge field; they enjoy their work and want to take on difficult ethical questions, too. They want to make a difference.

For the non-data scientist, DATA SCIENCE PIONEERS is still a critical piece of the puzzle that is the future of AI - it's about getting to know the people that are building systems that affect everyone's day-to-day lives. Accepting AI as an inevitability is complex and might start with data scientists themselves, but it's a fundamental change that will affect all and that cannot happen in isolation. DATA SCIENCE PIONEERS brings the story that will fuel discussions about not only how to construct a solid foundation for AI, but also a supportive community of those responsible for building it.

BACKGROUND

Data scientists are in extremely high demand as organizations leverage their skills to tackle all sorts of problems. Yet the demand (and the number of job postings) outpaces the supply. According to the Harvard Business Review, data scientists have the "Sexiest Job of the 21st Century."

This puts data scientists in a privileged position with a lot of job security, but curiously, what they actually do day-to-day remains a mystery to many. Data Scientists are at the front lines of data, privacy, and civil liberties debates, as they construct systems that take advantage of (or don't) personal user data. It's an incredibly challenging role, but data scientists are passionate, curious people who help generate meaning from sensors and stories from data.

ACCOLADES

**OVER
200**

screenings held in more than 40+ countries
worldwide with more than 6,000 viewers.



Selected for Snowdance Film Festival, Crystal Palace
International Film Festival, and more to come.

60,000+

Trailer views on
www.datascience.movie

8.8/10

IMDB rating.

“ There was a good compromise between
business and technical points of view. It is a real
challenge to understand the many roles involved
across such diverse domains ”

“ Dataiku has done a cool job with this new
documentary on what data science is all about ”

“ An amazing and grounded demystification of the data science
role, the people behind it, the challenges, the opportunities, the
technology, and the hype. A beautiful and important insightful story
that every business executive needs to see ”

A NOTE ON THE FILM



INSIDE THE CREATIVE PROCESS

We wanted to explore how data scientists perceive their work, impact, and the evolution of their careers. But since the data scientist role is so different across organizations, we prioritized speaking with data scientists at startups and large companies, small and large cities, GDPR countries and otherwise. Yet we found that these data scientists often shared perspectives on key issues, even if they didn't all arrive there from the same place.

Technologists, on the whole, don't care about competition that much; they focus on the next challenge and how they can overcome it together. The data science ecosystem is dominated by open source, collaborative tools, and libraries that enable all sorts of developers to contribute to solving these problems.

But as the field matures, there's this risk that it will become siloed around big businesses and that the spirit of innovation will be superseded by a preoccupation with profit. We want to ensure that at the end of the day, we're still tag-teaming these pressing issues like data ethics and the latest machine learning methods because data science impacts everyone, not just people buying and selling the tech. The data science community is still quite open and accessible, online and offline. That's why we supported this documentary: because we want to maintain the camaraderie, collaboration, and democratization that data science is fueled by.

—Florian Douetteau, CEO of Dataiku

BIOS AND CREDITS

• ALEXANDRA SUDOMOEVA, CISCO

Alexandra is a Data Scientist with Cisco's Business Strategy and Planning team in New York. She is originally from Moscow, Russia and moved to the United States to pursue her Bachelor's degree in Mathematical Economics from Santa Clara University. Alexandra also holds an Engineering Master's degree in Data Science from Columbia University and is one of the executive board members of Columbia's Data Science Society. She is an active member of the data science community and can often be spotted at a variety of AI and machine learning meetups specifically focused on women in tech.

• BEN FIELDS, LEAD DATA SCIENTIST, BBC

Ben Fields is a Lead Data Scientist at the BBC. There, his work focuses on recommenders, behavioral analytics, and data modeling. Previously, he was the Lead Data Scientist at FutureLearn where his work was focused on recommenders, learning analytics, and predictive modelling. He's also done post-doc research on music social networks and ran the data-centric agency Fun and Plausible Solutions. He holds a PhD from the Computing Department at Goldsmith University of London where his dissertation focused on automatic playlists for music recommendation; he is an expert on metadata, structured data, the semantic web and recommendation systems.

• BERNARDO NUNES, GROWTHTRIBE

Bernardo Nunes is the Head of Science at GrowthTribe Academy. He practices the combination of behavioural science, data analytics, and digital solutions to predict consumer behaviour.

• GIOVANNI LANZANI, GODATADRIVEN

Originally from Italy, Giovanni came to the Netherlands to study theoretical physics. He received a doctoral degree in Leiden by researching DNA mechanics. From there, he moved on to the consultancy world; in 2013, he joined GoDataDriven, where — first as a data whisperer and then as chief science officer — he started advising clients on how to extract value from (big) data and how to effectively bring data science models into production. Giovanni has also developed numerous courses in the machine learning and Python space and spoken at all-data conferences around Europe.

• JAN TEICHMANN, ZOOPLA

Jan Teichmann is a successful leader in the data transformation efforts of companies and has a track record of bringing data science into commercial production usage at scale. He previously co-founded Cambridge Energy Data Lab where they celebrated a successful exit with Enechange.jp, a utility comparison platform, which is now the market leader in Japan. Jan is a highly skilled data scientist, data engineer and solution architect and holds a PhD in Mathematics. He currently leads the data science team at Zoopla, driving innovation from vast amounts of property market data, behavioral data, geo data, property images, and text data sets.

• JOHN CALHOUN, AWS

John Calhoun is a machine learning specialist for AWS Public Sector. He works with customers and partners to provide leadership on machine learning, helping them shorten their time to value when using AWS. He holds a Masters in Mathematics, focused on high-performance computing, from Texas Tech University.

BIOS AND CREDITS

• JORIE KOSTER-HALE, FALLING COLORS

Jorie is a native New Mexican. After receiving her Bachelor's degree in linguistics, she got a PhD in cognitive neuroscience from MIT, where she studied the neural basis of human social reasoning. She continued that research as a postdoctoral fellow at Harvard University. Before joining Falling Colors as their Chief Data Officer, Jorie lived in Paris, working as a lead data scientist for an Enterprise AI software company. Jorie's primary goal at Falling Colors is to leverage data and technology to improve quality of care, efficacy of service delivery, and transparency of spending in behavioral health.

• JULIO PEIRONCELY, FREELANCE

Julio Peironcelly received his PhD from Leiden University for his research in computational life sciences. After graduating, he worked as a data science consultant at Deloitte and Microsoft. He is the founder of Data Fitter, a data and AI boutique agency. He has also served as Head of Data Science team at the Schiphol Amsterdam Airport and as Lead Data Scientist at VodafoneZiggo.

• KENNY NING, BETTER MORTGAGE

Kenny is currently a data engineer at better.com, an online mortgage lender. Previously, he was at Spotify as a data scientist on the content team. Outside of data, he performs music under the aliases FYK and more recently, k-means, a tech/music mashup project. He lives in Brooklyn, NY.

• KIA EISINGA, TOMTOM

Kia Eisinga is a senior data scientist at TomTom. She has been working in the field of data science for around three years. Previously, she worked as a data scientist at ABN AMRO bank in the company's center of expertise for data and AI. Today, she enjoys working with TomTom's different sources of structured and unstructured data and being confronted with practical, complex problems. Kia is convinced that data is an integral part of every business and that data science can bring a huge competitive advantage when embraced in the right way.

• LONGHOW LAM, FREELANCE

As an experienced data scientist, LongHow's interests are in applying data science, machine learning and programming in business environments. He has executed various projects in the telecommunications, banking, and retail industries. LongHow enjoys hobby projects on, for example, data scraped from web sites to inspire others and to show others that data science can be fun!

• LUCA BERNARDI, BOOKING.COM

Luca is a Principal Data Scientist for Booking.com based in Amsterdam. He focuses on the use of Data Science in product development, recommender systems, and productionalized machine learning models.

• LUCA MARIA AIELLO, NOKIA BELL LABS

Luca Maria Aiello is a Senior Research Scientist in the Social Dynamics team at Nokia Bell Labs Cambridge, UK and Research Fellow of the ISI Foundation in Torino, Italy. Formerly, he has been a Research Scientist at Yahoo Labs for almost 5 years. He conducts interdisciplinary research in network science, computational social science, and urban informatics. His work has been covered by hundreds of news articles published by news outlets worldwide including Wired, WSJ, and BBC. He is a founding member of GoodCityLife.org, a global network of scientists with the goal of giving a good life to city dwellers.

• MARC JANSEN, GOUSTO

Marc is a data scientist at the UK-based recipe kit company Gousto. Since joining in early 2017, he has worked on development and production deployment of algorithms for personalization, stock management, and warehouse optimisation. Marc currently leads a team implementing machine learning and optimization in the company's operations. Prior to joining Gousto, he completed a PhD in Operations Management at the University of Cambridge and spent time as a visiting researcher at the MIT-SUTD International Design Centre in Singapore.

BIOS AND CREDITS

• **MARIEM AYADI, CREDIT SUISSE**

Mariem's background is in computer science and engineering, beginning her career as a software developer and later discovering a passion for data science. Mariem works on various data sets and her projects touch upon BI, text mining, automation, sentiment analysis, and recommendation systems. In addition to her day-to-day role, Mariem has a strong interest in tech for social good, volunteering at hackathons for non-profits. She is also an advocate for getting more people into technology: she has previously created and taught Python and web programming workshops as well as been featured in the book, *Her STEM Career: Adventures of 51 Remarkable Women*.

• **MARYAM JAHANSHAH, TAP RECRUIT**

Maryam Jahanshahi is a research scientist at TapRecruit, a software company that uses AI and automation tools to bring efficiency and fairness to the recruiting process. The goal of her research program is to find ways to reduce bias in decision-making by using insights from different disciplines including computational linguistics, machine learning, and behavioral science. Maryam previously worked as a research scientist at the intersection between cancer and developmental biology, characterizing molecular regulators of organ size control as part of her graduate research.

• **PERRY BEAUMONT, DISTINGUISHED PROGRAMS**

Perry serves as Head of Data Science and Actuary at Distinguished Programs (a national insurance program manager) in New York City, he is also a Lecturer at Columbia University and has published books, articles, and blogs on a variety of financial service topics

• **SHAUN MCGIRR, HALF STACK DATA SCIENCE**

Shaun has been working with data in one way or another for about 15 years, starting out as an analyst and now leading a team of Data Scientists and BI Analysts for Cox Automotive UK. In between, Shaun earned a PhD in Political Science from the University of Michigan, where he learned the art of the good question. He is co-host of the Half Data Science Podcast, which discusses all the problems data scientists must solve beyond what they type in to the computer (the hosts will launch a course offering that fills this critical gap in data science training in late 2019).

• **TIMO GEMMECKER, DAIMLER**

Together with his team, Timo is driving Data & AI Solutions in Mercedes-Benz Cars Finance as well as promoting data culture and organizing data trainings. He has a passion for ramping up teams for the future, which he already did in China where he established a data team in finance for Mercedes-Benz.

• **TRIVENI GANDHI, DATAIKU**

Triveni Gandhi is a Data Scientist at Dataiku where she works with clients to deploy custom AI solutions and find meaning from complex data science pipelines. She enjoys learning about new technologies to solve whatever challenge comes her way and is a proud self-taught data scientist. Prior to working at Dataiku, she served as a data analyst for a large education non-profit in NYC, where she developed data pipelines and analyses to support the work of educators across the city. Triveni holds a Ph.D in Political Science from Cornell University.

FAQ

A woman with blonde hair is shown in profile, looking out a window. The background is a blurred view of a city street with buildings and a car. The overall tone is contemplative and professional.

WHY HAS DATAIKU INITIATED THIS DOCUMENTARY?

A Data Scientist's role comes in many shapes and sizes, and we believe that generating more transparency into the diverse paths data scientists take to their role and the variety of questions they seek to answer will increase understanding and inclusion in the field.

WHO IS DATA SCIENCE PIONEERS FOR?

DATA SCIENCE PIONEERS is for data scientists and everyone who works with them. Whether you want to be a data scientist or are one or work with them, this documentary will help you better understand data culture, history, and the variety of work experiences data scientists have. Ultimately, DATA SCIENCE PIONEERS is for anyone interested in data, whether professionally or as an amateur.

WHAT DID WE LEARN THROUGH FILMING?

Data scientists enjoy the challenge and variety of their work are ready to make AI more user-friendly and ethical. We've found that even without a physical community, data scientists are mostly in sync; the community is there, it just needs to be connected.

WHAT ARE SOME STATS ABOUT THE FILM?

It involved 22 Data Scientists in 4 cities representing 20 companies. There were 193 days of filming and prep that were fueled by 47 liters of coffee.

FAQ

WHEN WILL IT BE RELEASED?

DATA SCIENCE PIONEERS will be released in late September 2019.

WHAT IS THE DISTRIBUTION PLAN?

To bring together an array of data scientists, we will be hosting several premiere events globally in major cities (New York, London, Paris). Exclusive screenings will also be held at corporate offices for companies as a whole or data science teams. DATA SCIENCE PIONEERS will also be available for online streaming.

HOW CAN I GET A SCREENING OF DATA SCIENCE PIONEERS IN MY OFFICE?

Please reach out to our PR team if you are interested in bringing DATA SCIENCE PIONEERS to your team!

WHO IS DATAIKU?

Dataiku is the centralized data platform that moves businesses along their data journey from analytics at scale to Enterprise AI. The company's mission is to provide organizations with the technological environment that enables all people throughout an enterprise to use data by removing friction surrounding data access, cleaning, modeling, deployment, and more that allow for successful operationalization. Learn more at www.dataiku.com.



— LEARN MORE



WWW.DATASCIENCE.MOVIE



TWITTER.COM/PIONEERSDATA



INFO@DATASCIENCE.MOVIE

CONTACT:

IN NORTH AMERICA:

ALI DONZANTI, FOR DATAIKU
+1.646.428.0620
dataiku@allisonpr.com

IN FRANCE:

L'AGENCERP POUR DATAIKU
EVA GLÉLÉ | FLORENCE HANQUIEZ |
AUDREY VILFROY
+33 1 84 76 10 37
dataiku@lagencerp.com

IN THE UNITED KINGDOM:

COMMSO FOR DATAIKU
LEAH JONES
+44 203 697 6680

IN THE NETHERLANDS:

CHANTAL SCHEPERS
ITSAREP
+31 6 235 099 23
chantal@itsarep.nl