

INTERNATIONAL ONLINE ANALYTICS HACKATHON October-November 2021

PARTICIPANTS
REGISTRATION

ANALYZE DATASETS

Tech Talks
Abstract Reviews

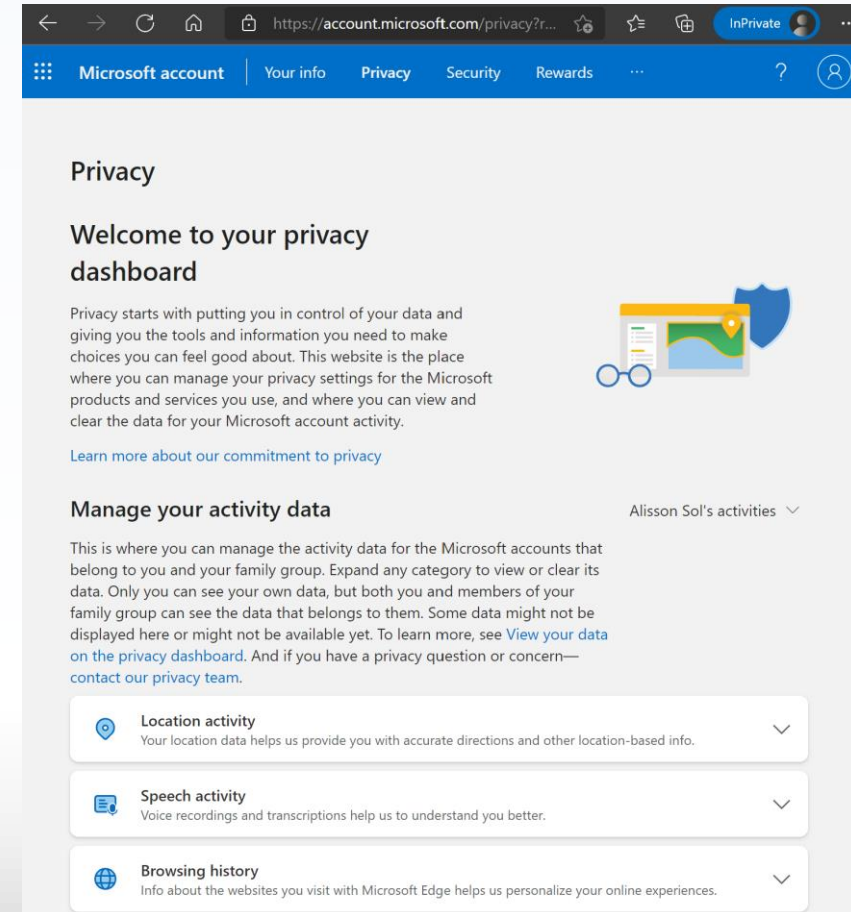
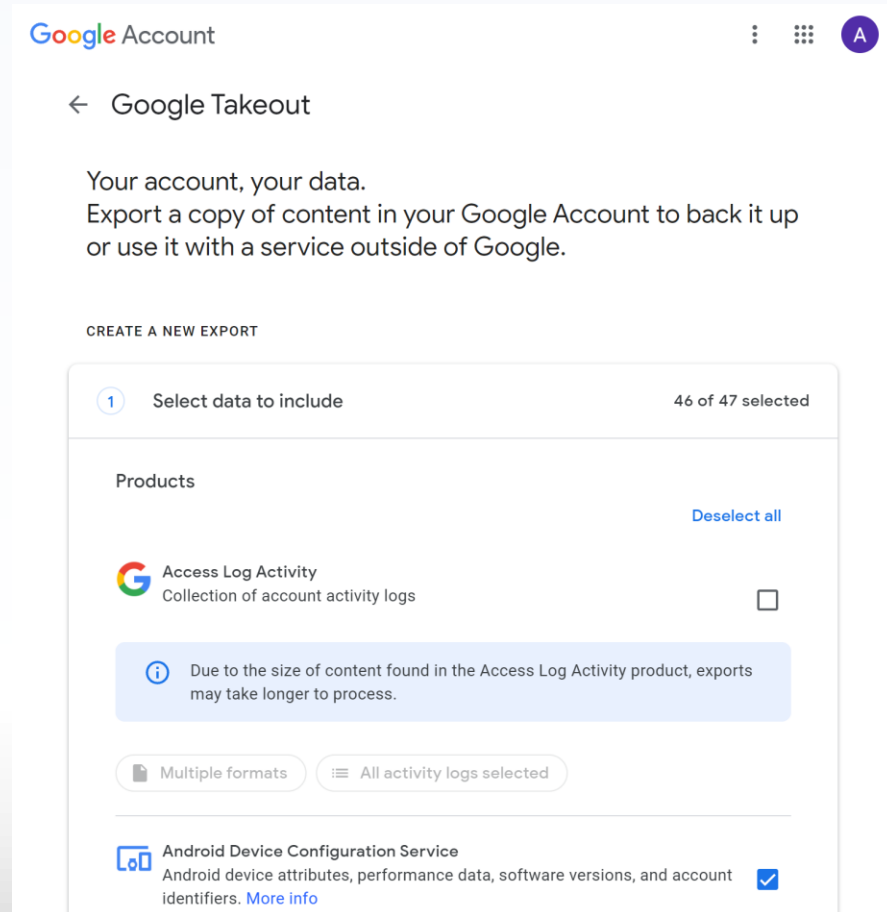
VISUALIZE RESULTS

Visualization Workshops
Presentation Reviews

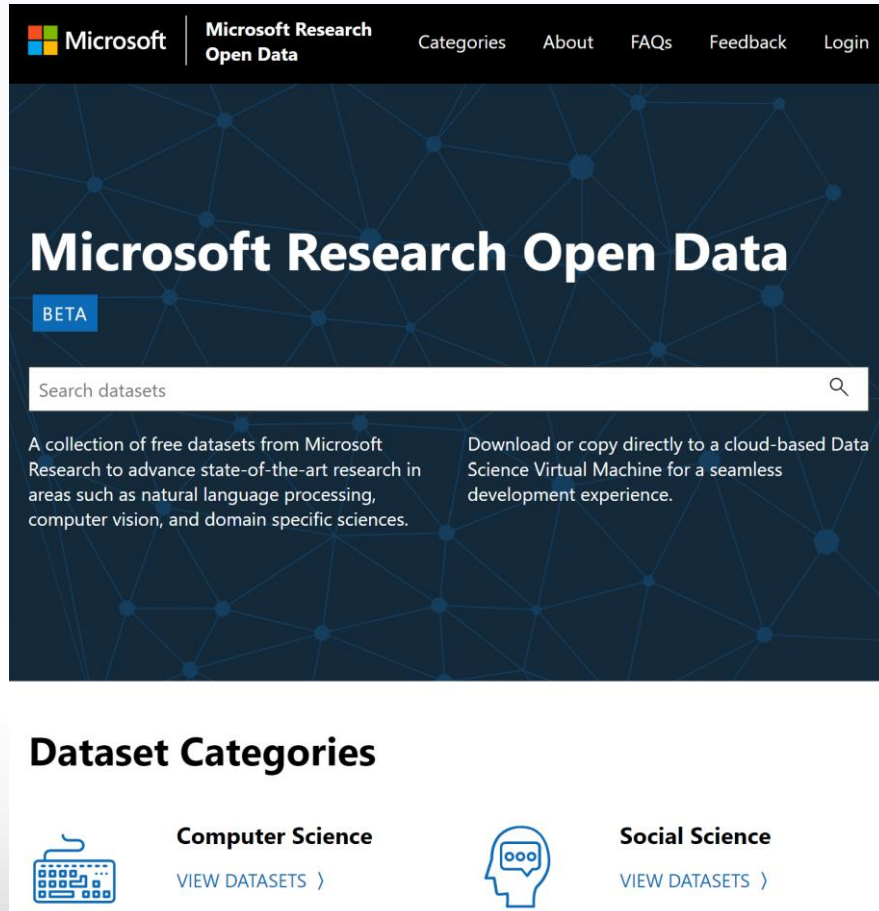
SUBMIT PROJECTS

Presentation Promotions
Public Voting

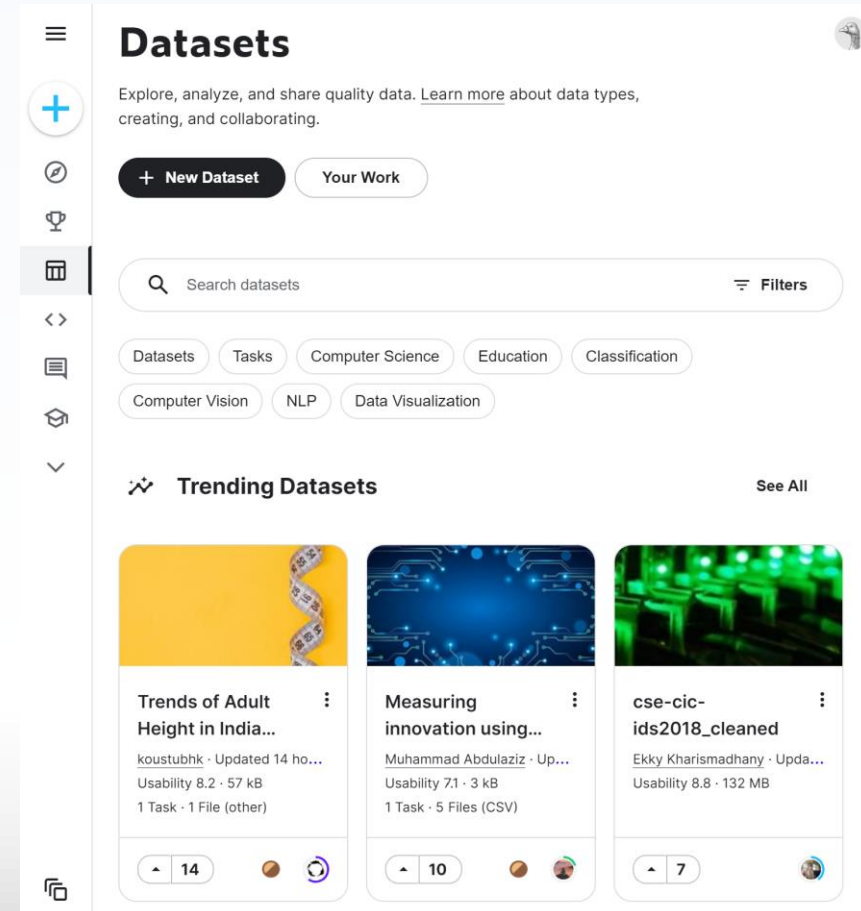
Data: starting with “You” ...



Datasets



The screenshot shows the Microsoft Research Open Data website. The header includes the Microsoft logo, "Microsoft Research Open Data", and navigation links for Categories, About, FAQs, Feedback, and Login. The main section features a dark blue background with a network diagram and the text "Microsoft Research Open Data" with a "BETA" badge. Below this is a search bar labeled "Search datasets". A descriptive paragraph states: "A collection of free datasets from Microsoft Research to advance state-of-the-art research in areas such as natural language processing, computer vision, and domain specific sciences." Another paragraph mentions: "Download or copy directly to a cloud-based Data Science Virtual Machine for a seamless development experience." The bottom section, titled "Dataset Categories", highlights "Computer Science" and "Social Science" with icons and "VIEW DATASETS" links.



The screenshot shows the "Datasets" page on the Microsoft Research Open Data platform. The header includes a hamburger menu, a "+" icon, and a search bar labeled "Search datasets" with a "Filters" button. Below the search bar are category tags: Datasets, Tasks, Computer Science, Education, Classification, Computer Vision, NLP, and Data Visualization. The "Trending Datasets" section displays three dataset cards:

- Trends of Adult Height in India...**
koustubhk · Updated 14 ho...
Usability 8.2 · 57 kB
1 Task · 1 File (other)
- Measuring innovation using...**
Muhammad Abdulaziz · Up...
Usability 7.1 · 3 kB
1 Task · 5 Files (CSV)
- cse-cic-ids2018_cleaned**
Ekky Kharismadhany · Upda...
Usability 8.8 · 132 MB

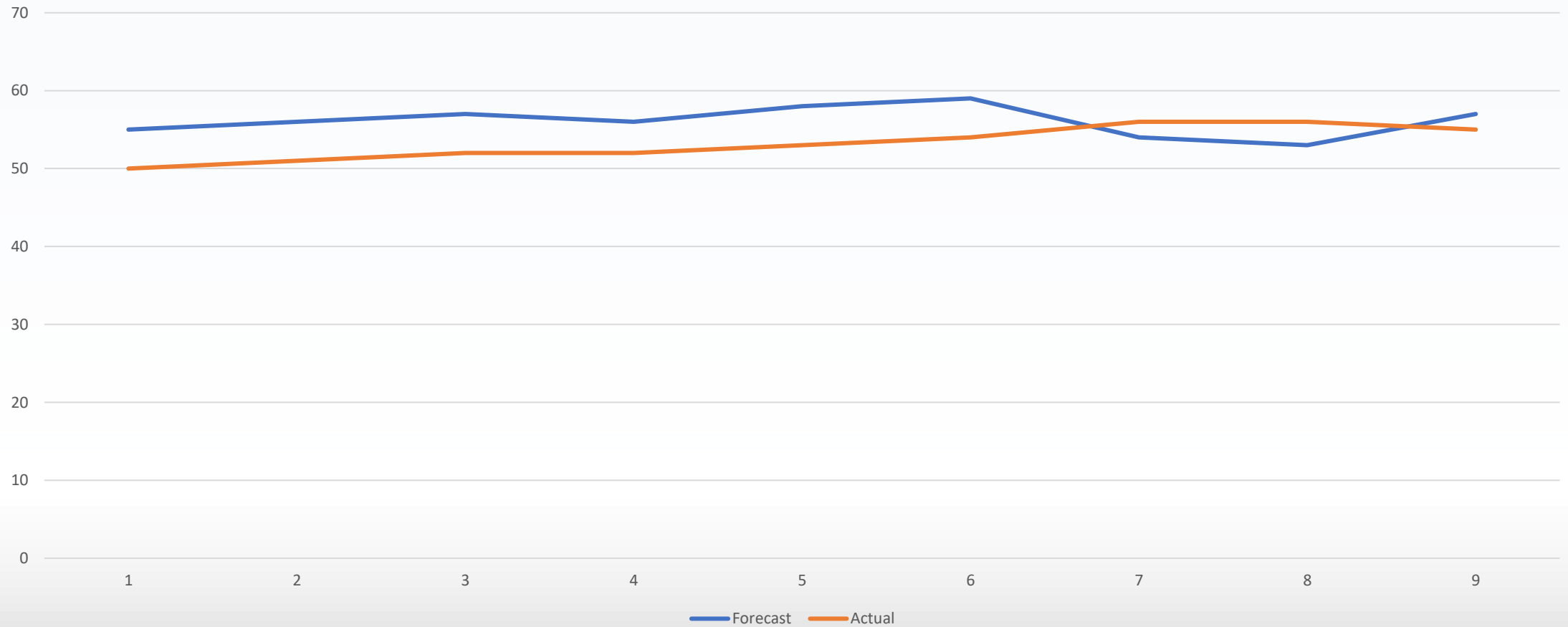
Each card includes a thumbnail image, a title, author information, usability score, file size, and task/file count. Navigation controls like arrows and a "14", "10", or "7" indicator are at the bottom of each card.

Data disconnected from storage

864 servers
27.6 petabytes of storage



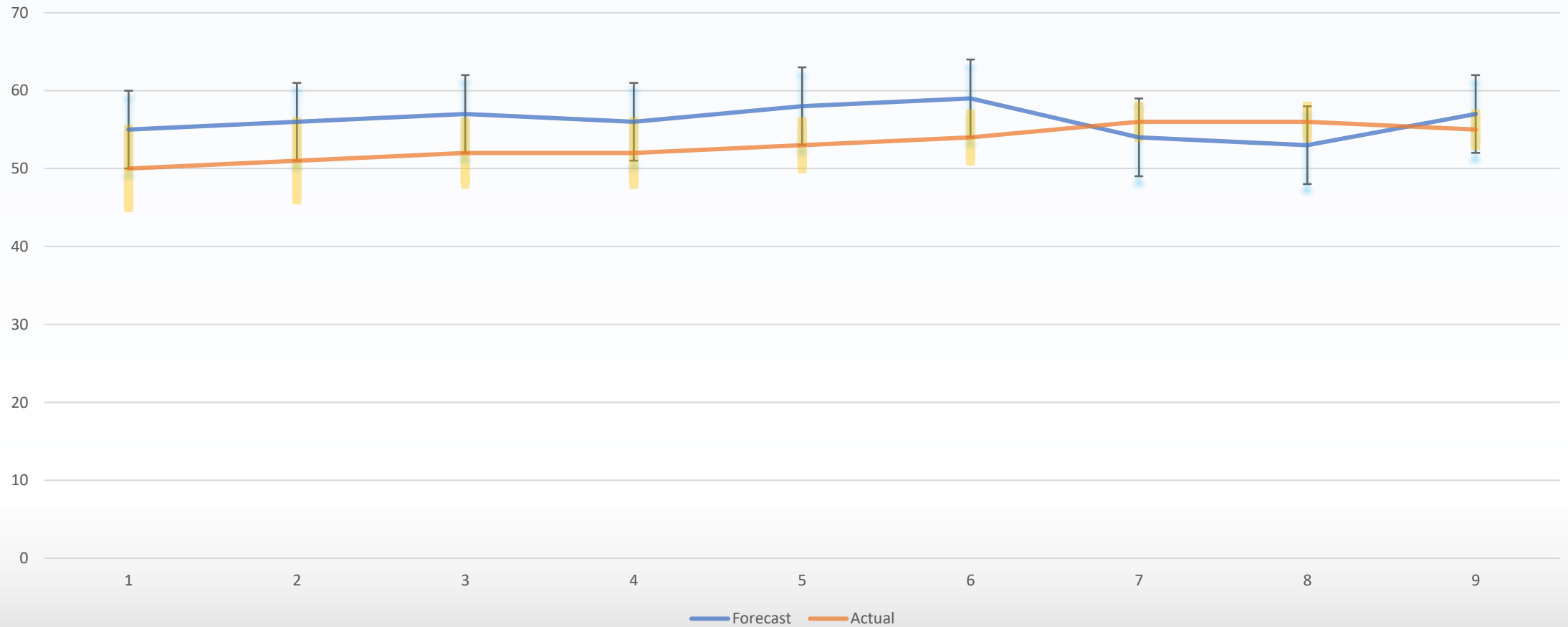
Analysis



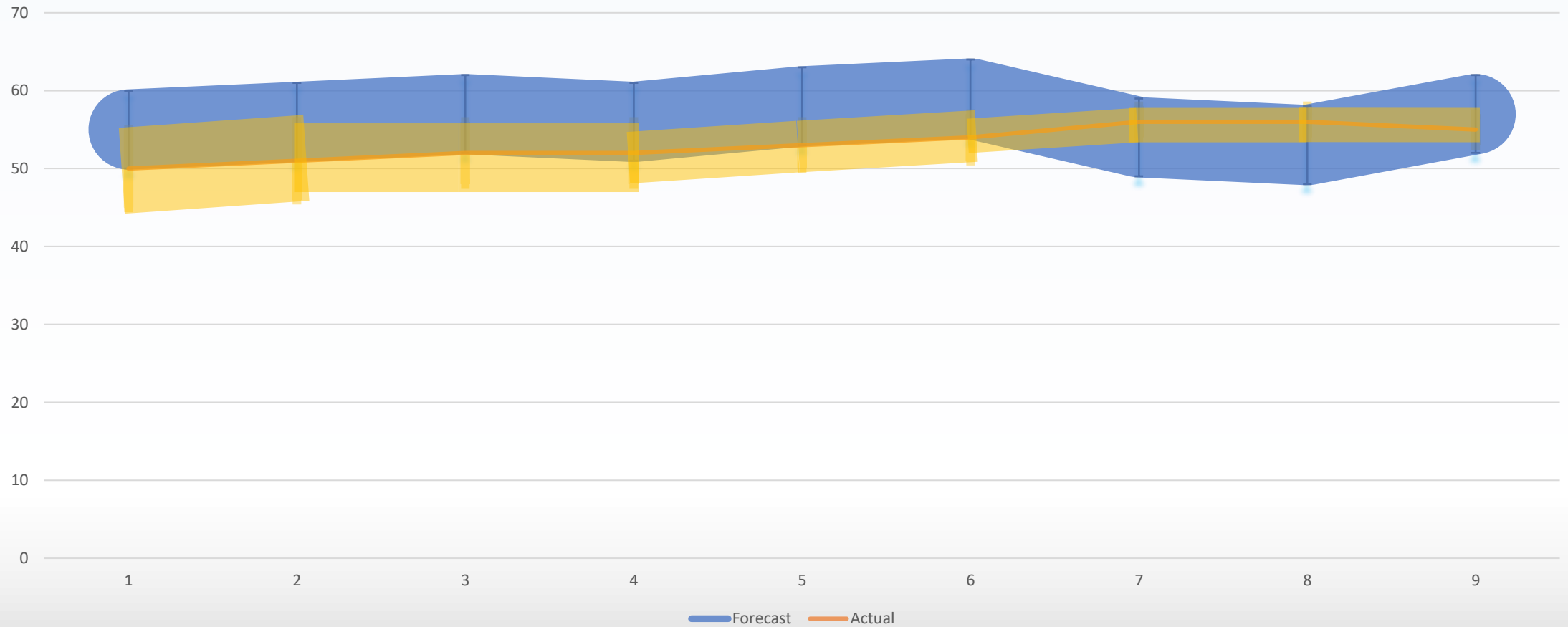
Raw Data

Time	Forecast	Actual	Forecast Error	Actual Error
1	55	50	5	5
2	56	51	5	5
3	57	52	5	4
4	56	52	5	4
5	58	53	5	3
6	59	54	5	3
7	54	56	5	2
8	53	56	5	2
9	57	55	5	2

Error Margin in Graph



Error Margin in Line



Spreadsheet Statistics: Service Cost/Unit

Service	Cost (\$MM)	Units (1,000,000s)	Cost/Unit (\$/unit)
Service1	6.00	2.00	3.00
Service2	6.00	3.00	2.00

Your “average cost” (mean average)

Service	Cost (\$MM)	Units (1,000,000s)	Cost/Unit (\$/unit)
Service1	6.00	2.00	3.00
Service2	6.00	3.00	2.00
Average	AVERAGE () = 6.00	AVERAGE () = 2.50	AVERAGE () = 2.50

Cannot see the problem? Scale services differently...

Service	Cost (\$MM)	Units (1,000,000s)	Cost/Unit (\$/unit)
Service1	6.00	20.0	0.30
Service2	6.00	30.0	0.20
Service3	60.0	1.00	60.0
Average	AVERAGE () = 24.0	AVERAGE () = 17.0	AVERAGE () = 20.2

Don't use mean average with ratios

Service	Cost (\$MM)	Units (1,000,000s)	Cost/Unit (\$/unit)
Service1	6.00	2.00	3.00
Service2	6.00	3.00	2.00
Average	AVERAGE () = 6.00	AVERAGE () = 2.50	AVERAGE () = 2.50

Service	Cost (\$MM)	Units (1,000,000s)	Cost/Unit (\$/unit)
Service1	6.00	2.00	3.00
Service2	6.00	3.00	2.00
Average	SUM () = 12.0	SUM () = 5.00	12.0/5.00 = 2.40

Vaccine trials

Trial 1	Test	Success	Success Ratio
Vaccine	100	66	66%
Placebo	40	24	60%

Trial 2	Test	Success	Success Ratio
Vaccine	200	180	90%
Placebo	500	430	86%

Trial 1+2	Test	Success	Success Ratio
Vaccine	300	246	82%
Placebo	540	454	84%

Simpson's Paradox

Analysis of rare events...

- Are you blinking?
 - Average: 10 blinks per minute, for about ~100-150 milliseconds
 - Total = ~1 to 1.5 seconds for 60 seconds
 - Typical camera: 30 frames per second
 - Raw data: 30×60 frames = 1,800 frames
 - Frames with “blink”: $30 \times 1.5 = 45$

		Actual	
		Negative	Positive
Predicted	Negative	TN: 1,755	FN: 45
	Positive	FP: 0	TP: 0

Accuracy

- Accuracy: $(TP+TN)/(TP+TN+FP+FN) = 0.975$
- Seek other metrics: F-Score or F1Score

		Actual	
		Negative	Positive
Predicted	Negative	TN: 1,755	FN: 45
	Positive	FP: 0	TP: 0

Analysis as a team



MLOps: Model management, deployment, lineage, and monitoring with Azure Machine Learning

07/08/2021 • 9 minutes to read • 

In this article, learn about how to use Azure Machine Learning to manage the lifecycle of your models. Azure Machine Learning uses a Machine Learning Operations (MLOps) approach. MLOps improves the quality and consistency of your machine learning solutions.

What is MLOps?


Machine Learning Operations (MLOps) is based on [DevOps](#) principles and practices that increase the efficiency of workflows. For example, continuous integration, delivery, and deployment. MLOps applies these principles to the machine learning process, with the goal of:

- Faster experimentation and development of models
- Faster deployment of models into production
- Quality assurance and end-to-end lineage tracking

Azure Machine Learning provides the following MLOps capabilities:

- **Create reproducible ML pipelines.** Machine Learning pipelines allow you to define repeatable and reusable steps for your data preparation, training, and scoring processes.
- **Create reusable software environments** for training and deploying models.
- **Register, package, and deploy models from anywhere.** You can also track associated metadata required to use the model.
- **Capture the governance data for the end-to-end ML lifecycle.** The logged lineage information can include who is publishing models, why changes were


Think!

[GDPR.EU](#) 

Search...

Search

[Home](#) [Checklist](#) [FAQ](#) [GDPR](#) [News & Updates](#)



Everything you need to know about the “Right to be forgotten”





Also known as the right to erasure, the GDPR gives individuals the right to ask organizations to delete their personal data. But organizations don't always have to do it. Here we explain when the right to be forgotten applies and when it doesn't.

The General Data Protection Regulation ([GDPR](#)) governs how personal data must be collected, processed, and erased. The “right to be forgotten,” which received a lot of press after the [2014 judgment from the EU Court of Justice](#), set the precedent for the right of erasure provision contained in the GDPR. Of course, given competing interests and the hyper-connected nature of the Internet, the right to be forgotten is much more complicated

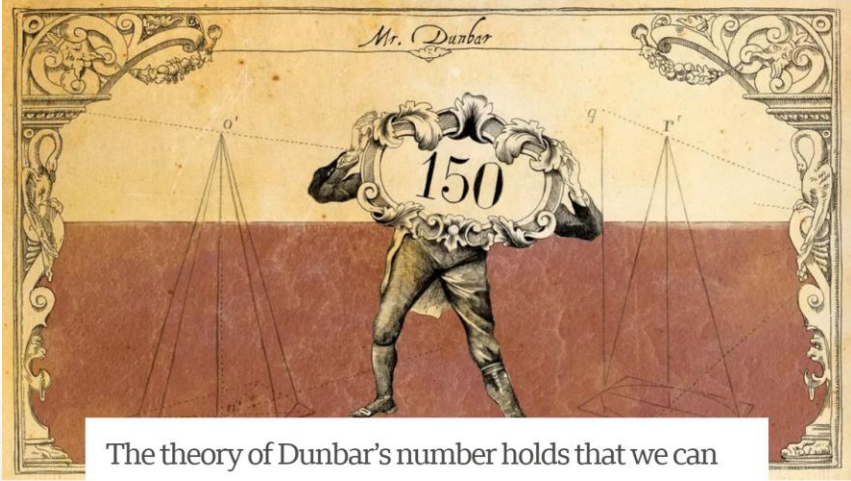
FUTURE [What is BBC Future?](#) [Future Planet](#) [Bright Sparks](#) [More](#)

INVISIBLE NUMBERS | PSYCHOLOGY

Dunbar's number: Why we can only maintain 150 relationships

(Image credit: Emmanuel Lafont)



The theory of Dunbar's number holds that we can only really maintain about 150 connections at once. But is the rule true in today's world of social media?

Act

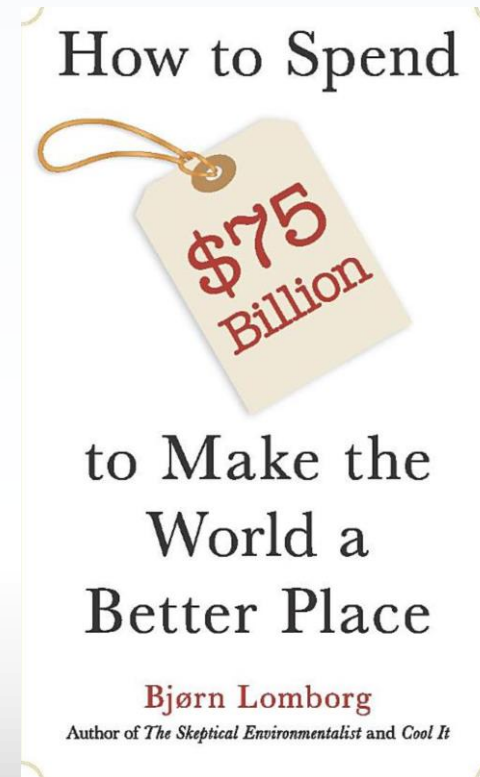


Action and priorities

2006



2014



Call to Action

- Data
- Analysis
- Think!
- Act

Education

To Educate Children, We Have To Teach Their Parents

We can improve education in poor countries by showing parents the importance of schooling.

Thank you!



Alisson Sol

- Email: email@AlissonSol.com
- Resume: PDF [file](#)
- GitHub: <http://GitHub.com/alissonsol>
- Curriculum Vitae: PDF [file](#)
- Social networks: [LinkedIn](#), [Twitter](#), [Instagram](#)
- Personal [blog](#)

Alisson Sol has many years of experience in software development, having hired and managed several software teams that shipped many applications, services and frameworks, with focus on image processing, computer vision, ERP, business intelligence, big data, machine learning, AI, and distributed systems.

He currently works as a Group Engineering Manager for the Microsoft [AI Frameworks](#) team. Previous, he worked from Sep/2019 to Oct/2020 in the Microsoft 365 (M365) Intelligent Conversation and Communications Cloud (IC3) team. From 2017 to 2019, he was a Senior Development Manager for Amazon, working in the AWS WorkSpaces and the retail direct fulfillment (dropship) teams. He worked for Microsoft and Microsoft Research in the USA and UK from 2000 to 2017, and was previously a co-founder for 3 software companies. He has published several technical papers and has several patent applications and granted patents. He has a B.Sc. in Physics and a M.Sc. in Computer Science by the Federal University of Minas Gerais in Brazil, and General Management training at the University of Cambridge-UK. When not coding, he likes to run half-marathons, play soccer or disassemble hardware, put it back to work and reuse the spare parts elsewhere!

Tweets by @alissonsol



Alisson Sol
@alissonsol

There is no need of a cargo cult for leadership principles... lnkd.in/gu7K5U7d



Oct 6, 2021

Instagram pictures: [alisson.a.sol](#)

