

Images haven't loaded yet. Please exit printing, wait for images to load, and try to print again.

How Airbnb Democratizes Data Science With Data University



Jeff Feng [Follow](#)

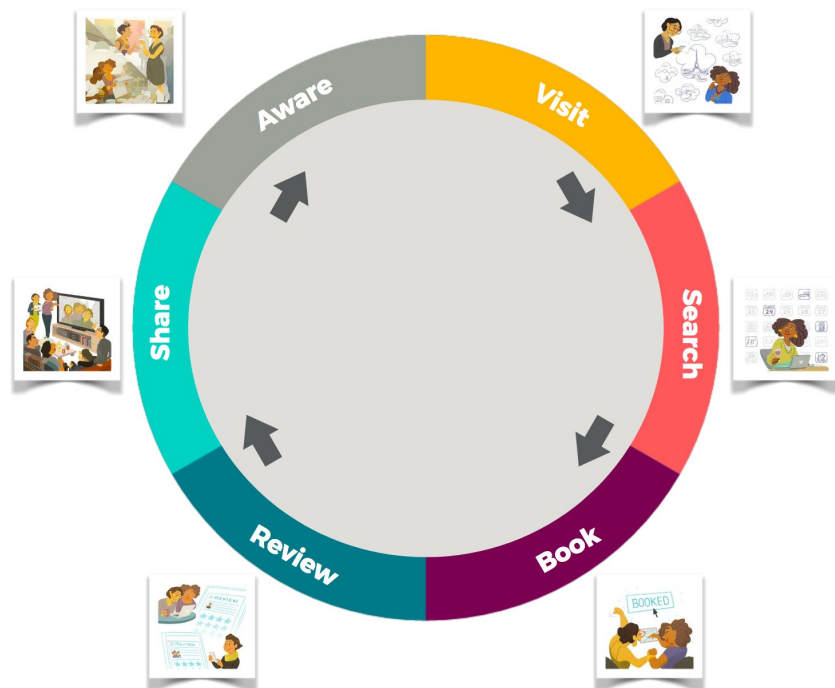
May 24, 2017 · 7 min read

By [Jeff Feng](#), [Erin Coffman](#) & [Elena Grewal](#)

Introduction

Data is essential to us at Airbnb. We characterize data as the voice of our users at scale. Thus, data science plays the role of an interpreter—we use data and statistics to understand our users and translate it to a voice that people or machines can understand. We leverage these quantitative insights, paired together with qualitative insights (e.g. in-person user research) to make the best possible decisions for both the business and our community of hosts and guests.

Data is the voice of our customer



To that, we have built a world-class Data Science Team that has scaled to nearly 100 people strong working on everything from experimentation to data analysis & visualization to modeling & machine learning. We also built a stable, reliable and scalable data infrastructure to serve as the foundation for our data, as well as a powerful suite of data tools to empower data scientists and knowledge workers all across Airbnb:

- Apache Airflow (Incubating) [[Blog post](#) | [Github](#)]—Data workflow management platform
- Apache Superset (Incubating) [[Blog post](#) | [Github](#)]—Data visualization and access
- ERF [[Blog post](#)]—Experiment reporting framework for A/B testing
- Dataportal [[Blog post](#)]—Data search, exploration and trust
- Knowledge Repo [[Blog post](#) | [Github](#)]—Scaling knowledge and insights

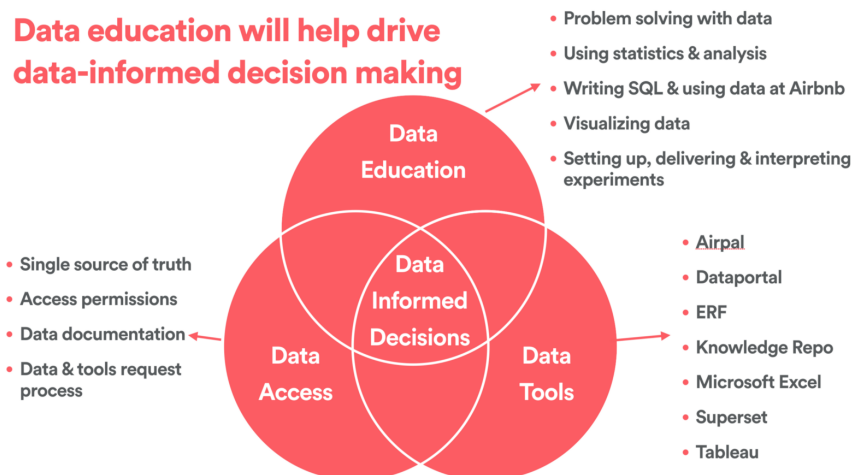
Scaling & Democratizing Data Science

Another one of our fundamental beliefs is that *every* employee should be empowered to make **data informed** decisions. This applies to all parts of Airbnb's organization—from deciding whether to launch a new product feature to analyzing how to provide the best possible employee experience. Our Data Science team firmly believes that part of our goal is to empower the company to understand and work with data. In order to inform every decision with data, it wouldn't be possible to have a data scientist in every room—we needed to scale our skillset.

Additionally, our rapid international growth made the situation even more challenging. We expanded from one office in San Francisco in 2011 to 22 offices internationally today, many of which do not have Data Science presence. Furthermore, we believe that people have the capability to think critically and understand the data on their own, and we wanted to give them the tools to do it.

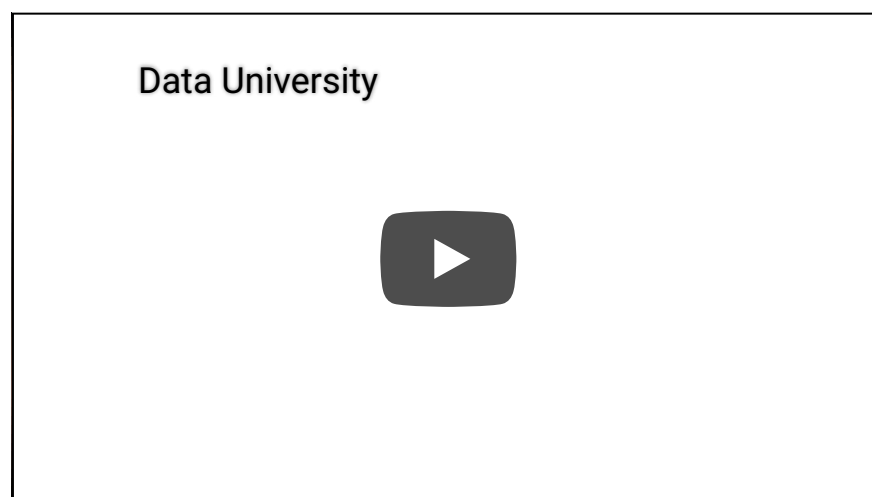
To address this challenge, we thought deeply about how to democratize data science and scale data informed decision making during the 2nd half of 2016. We used a metric of weekly active users (WAUs) of our data platform as a proxy to how “data informed” we were as an organization. At the beginning of Q3 2016, only about 30% of Airbnb employees were a WAU of our data platform, which was significantly lower than other hypergrowth internet company peers we benchmarked with like Facebook and Dropbox.

We then thought about what might be holding back our company from looking at data themselves.



The key ingredients needed for data informed decision making included having accessibility to data, a comprehensive set of data tools and user knowledge of how to utilize the data and tools. As we spoke to people throughout Airbnb, it became more and more apparent that the bottleneck to scaling data informed decisions was actually data education for users. Our data tools were serving Data Scientists well. Also, we had already made huge strides in making data more accessible through efforts like Core Data, our single source of truth for product data, as well as [SQL Lab](#), a new SQL editor we built into Superset. The gap was that we didn't have any formal programs to equip employees with the knowledge to use our tools and how to work with Core Data. Thus, we decided to create Data University.

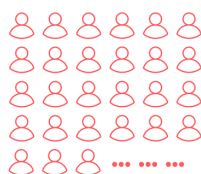
Our Solution: Data University



Video credit: ReadyState

Data University is data education for anyone at Airbnb that scales by role and team. Our vision is to empower **every employee** to make **data informed decisions**. Our approach is unique since organizations offering data education typically focus just on their technical employees. Our approach is also intentional because we believe that every person at Airbnb should and can utilize data in his/her role to make better decisions. Thus, we designed the program to make it accessible and relevant to anyone at Airbnb.

Data University Vision



To empower
EVERY
employee at
Airbnb...



to make
data-informed
decisions...



by providing data
education...



that scales by
role & team.

Creating “citizen data scientists” is powerful—not only does it help ensure that decisions are grounded in data, but it enables people to make decisions autonomously. This is important because the person asking the question always has the best context on the question they are trying to answer, and it reduces the feedback loop to answering questions. This also has the side benefit of freeing up some of the Data Science Team’s time. We had considered leveraging existing resources from MOOCs such as Coursera and Udacity, however many of our data tools are unique, and we believe there is tremendous value in educating people in the context of Airbnb’s data.

The Data University Curriculum

The curriculum consists of over 30 classes covering an array of different topics. The 100-level series provides the foundation for data informed decision making at Airbnb and was designed to be accessible to everyone. The 200-level series equips people with the applied skills for accessing data using SQL, or analyzing and visualizing data using tools such as Superset, Tableau and ERF in the context of Airbnb data. Then, the 300-level series is targeted primarily towards engineers and data scientists. It exposes people to advanced data techniques such as machine learning and tools such as Airflow for writing data pipelines. We also cover popular languages such as R, Python and Hive for analyzing and manipulating data.



Data University Curriculum Overview



100-Level: Data Awareness

- Data-Informed Decision Making, Intro to Stats & Data Resources



200-Level: Data Collection & Visualization

- Core Data, SQL, Superset, Tableau, ERF, Knowledge Repo



300-Level: Data at Scale

- Hive, Airflow, R, Python, Machine Learning, Data Logging

The Data University Faculty

Many of the initial classes were developed and taught by [Erin Coffman](#), our most tenured Data Scientist at Airbnb. However, since then we have amassed more than 30 volunteer faculty (many pictured below) from across the Data Science and Engineering organizations to help create course content as well as teach classes. We are incredibly grateful for all the volunteers who make Data University possible!

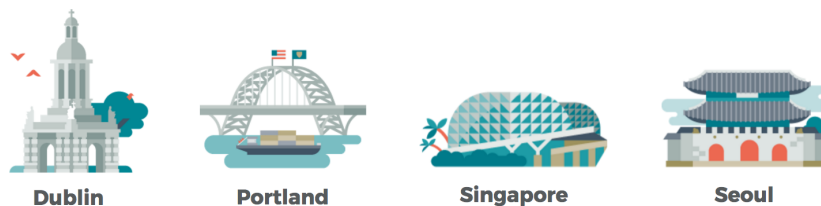


The Impact: Democratizing Data Science

Data University has been a huge success thus far at Airbnb. In the first half year since it launched, more than 500 unique people have participated in at least one class (or about 1/8th of Airbnb). Depth of engagement is high as each employee who has participated has taken more than 4 classes on average as we have had a total of over 2,100 “butts in seats.” Every class offered thus far has an NPS score of +55 or higher.

Furthermore, it has completely transformed Airbnb’s data culture as 45% of Airbnb is now a WAU of the data platform. Ad hoc data requests that used to go to Data Scientists or Analysts are now often being self-serviced or addressed by other Data University graduates. We routinely hear anecdotes about employees being empowered with data, from employees in recruiting creating Tableau dashboards to Product Managers writing their own SQL and interpreting their own experiments. Most recently, we have begun scaling the program to other offices including Dublin, Portland, Singapore and Seoul.

Data University is going global!



Looking Ahead

Our team is really encouraged about the initial results of Data University, and we will continue to iterate on the content as well as scale the program both in terms of breadth of classes (the 300-level series is next) and locations where the curriculum is offered. We will also be experimenting with different learning formats such as online and/or streamed courses.

In sharing our experience, we hope to inspire other organizations working on the same kinds of problems of scale and data democratization that we are trying to solve, as well as share learnings so that we can collaboratively produce best practices. If you're

interested in exchanging notes, or have follow up questions about our approach, please [reach out!](#)



A related article appeared on [Tech Crunch](#).

Check out all of our open source projects over at [airbnb.io](#) and follow us on Twitter: [@AirbnbData](#) + [@AirbnbEng](#).

