Uber Data Science and Analytics: Case Study

Uber is one of the giant ride-hailing companies across the world. Uber serves approximately 8 million users in its platform. It makes over 1 billion trips across 449 cities spread over the globe. Therefore, this ride-sharing firm must be experiencing several challenges. For instance, Uber must deal with poor transportation infrastructures in certain cities, poor fulfillment, and unsatisfactory customer experience, among others. The company generates big data every day, relies on Data Science and Artificial Intelligence to decide and give you a pleasant experience. Uber has a massive database of drivers, so as soon as you request a car, Uber's algorithm goes right to work – in 15 seconds or less, it matches you with the driver closest to you. In the background Uber is storing data for every trip taken — even when the driver has no passengers. All this data is stored and leveraged to predict supply and demand, as well as setting fares. Uber also looks at how transportation is handled across cities and tries to adjust for bottlenecks and other common issues.

1. How Uber is leveraging AI and Machine Learning?

Uber ATG (Advanced Technologies Group) is dedicated towards advancing research and building technology to propel Uber in the Autonomous vehicle space. Uber is a two-sided marketplace of drivers(supply) and riders(demand). All the different kind of rides that Uber offers is a way to provide more options to add supply and demand. For Uber, the key KPIs include 'average pick up time' for riders and 'average gross $ per hour' for drivers. As Uber's execution and network effect optimizes the utilization in the marketplace, the wages for the driver increase while the average cost to the rider keeps reducing. Uber does all this while collecting data for each ride. This trove of data is a moat which it leverages to move horizontally and vertically in this space. The former improves the data in the network while the latter will further optimize the supply. These two components will have an exponential effect on the network in the marketplace. For the vertical move, data provides the direction for optimization (like improving average pick up time) while Artificial Intelligence and Autonomous technology are employed to achieve it.

National Highway Traffic Safety Administration (NHTSA) classified autonomous cars into five levels based on autonomous capabilities, which provides a framework for charting technological progress on a scale.

Level 0: No Automation: Driver controls all the functions of the car. No Automation. — These are most of the cars today without any automated functions.

Level 1: Function Specific Automation: One or more specific control systems are automated. For example, parking assist or lane assist. — The newer cars which came out recently have these functions.

Level 2: Combined Function Automation: At least two of the control systems are automated in tandem. For example, adaptive cruise control with lane centering. — This is where Uber and most of the other autonomous vehicle startups will start.

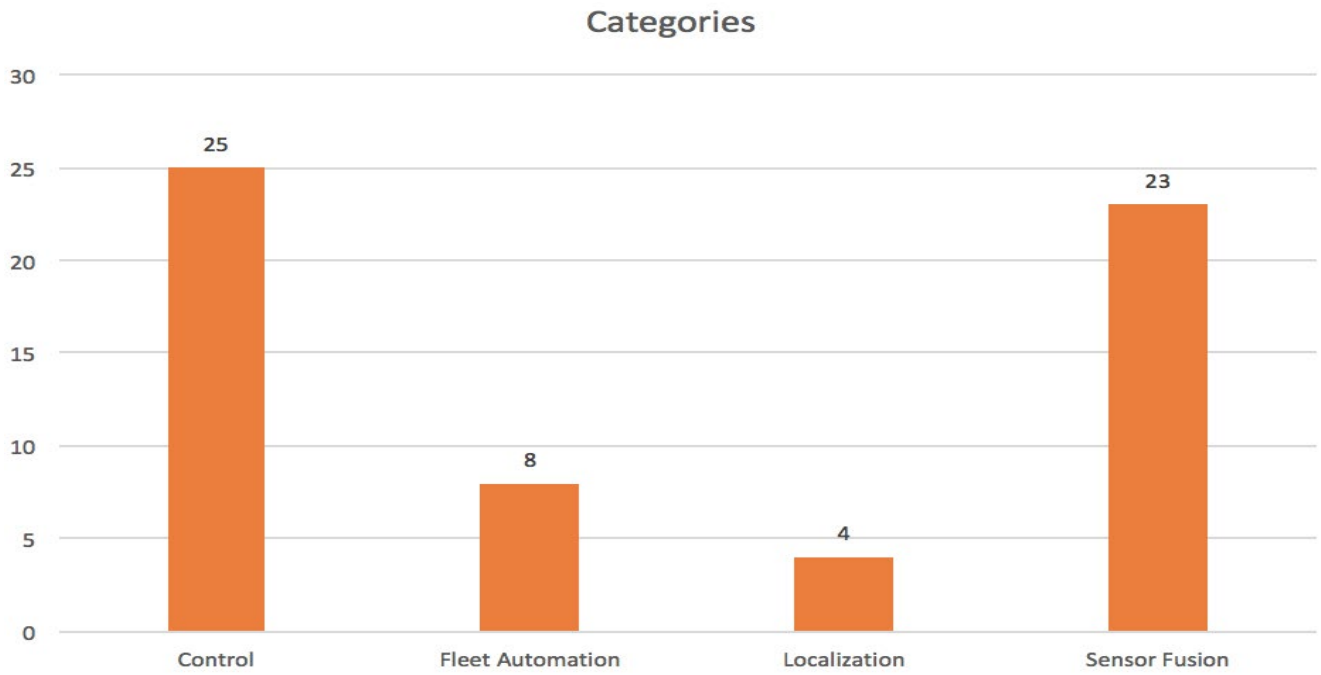
Level 3: Limited Self Driving Automation: All the control systems like steering, brakes and throttle are automated to work in tandem. The car monitors conditions to require relinquish control back to the driver. — Early versions of Cruise and other retro fit autonomous startups tried to approach this Level.

Level 4: Full Self Driving Automation: All control systems are fully automated requiring no human driver. — Final goal for most autonomous vehicle startups.

Uber is striving to go from a Level 2 to Level 4 with its data and technology prowess. We are exploring the autonomous landscape of Uber from the lens of their existing patents relating to autonomous technology.

Key Trends:

Deep Vertical Expansion: Uber is growing its stack of patents in different categories to control the entire stack.



Acquisitions: Uber acquired deCarta and Microsoft's mapping division to improve their mapping and navigation functionality. Uber also acquired Otto — a self-driving truck startup.

2. Describe the analytics process followed by the company?

Artificial intelligence powers many of the technologies and services underpinning Uber's platform, allowing engineering and data science teams to make informed decisions that help improve user experiences for products across our lines of business.

At the forefront of this effort is Uber AI, Uber's center for advanced artificial intelligence research and platforms. Uber AI powers applications in computer vision, natural language processing, deep learning, advanced optimization methods, and intelligent location and sensor processing across the company, as well as advancing fundamental research and engaging with the broader AI community through publications and open-source projects.

Big Data:

The company generates big data every day, relies on Data Science and Artificial Intelligence to decide and give you a pleasant experience. Some of Uber's big data infrastructures in the public domain are:

- **Hadoop Data Lake:** Uber leverages big data processing analytic capabilities to tackle its operational challenges that cannot be addressed using conventional data warehousing technologies.
- **Spark and Hadoop**: the two data frameworks can be used as alternatives. Both are mutually exclusive, but they work better when they are paired. Thus, Uber uses the two to process its data.
- Uber generates its data from databases and data types like database tables, SOA, event messaging system, or Apache Kafka and schema less data stores.

Data Science:

Uber has hired data analysts to study its back-end extract obtained from its application. They are behind its predictive models on its product front. These models can predict the exact time the driver will be at your door, or the estimated fares. The statistical data analysis can give Uber's customers a positive user-experience.

Data Science Tools:

Uber team uses Python in its programming language. Other third-party modules that the cab sharing company uses are SciPy, NumPy, Pandas, and Matplotlib. The data science team uses R programming language, MATLAB, and Octave intermittently for its single data science projects. It also uses data visualization tools like D3 and SQL frameworks such as Postgres.

The need for Uber to really use deep learning is not superficial. A company solely based on an application to interact with its customers on one hand and drivers on the other hand, operating in such a dynamic and complex environment, and working on a million variables at the same time is driven by its necessity to stay smart using machinery – not just finite computation but also high-order intelligence problems through Artificial General Intelligence (AGI is the ability to think more and more like humans, getting smarter with every single exposure revealing more about people's behavior

Creating more seamless communication with conversational AI

3. What are the big problems UBER is tying to solve using machine earning and AI?

Uber's problem space is new and rapidly evolving, especially at this scale and intersection of dimensions:

- **Spatial:** at both macro levels (global, regional, and city) and micro levels (riders, cars, and goods)
- **Temporal**: from seconds to years
- **Human:** involved at every stage, from decision making to decision receipt
- **Active:** immediate impact and response on the system being modeled
- **Scale:** billions of calculations and thousands of decisions made for millions of riders and drivers every minute

Designing Uber Maps:

To enable a more seamless user experience, Uber pairs mapping technologies with ML. In real time, they identify context-aware suggestions for destinations, considering the rider's current location, time of request, and historical information

Improving location accuracy with sensing and perception:

Uber AI's Sensing and Perception team worked on projects across our mobile and back-end stack to improve the coverage, accuracy, speed, and heading of vehicle locations on the Uber platform. Overcoming the limitations of GPS and having more precise locations makes it easier for riders and drivers to find one another, improves estimated times of arrival (ETAs), reduces rider and driver cancellations

Leveraging computer vision to make Uber safer and more efficient:

The Computer Vision Platform team has worked closely with product teams across Uber to enable scalable, reliable, and quick validation of driver identity when drivers go online. And, as Uber onboards a growing number of drivers and restaurants to our platform, we've built automated deep learning transcription technology that's suited to Uber's specific use case—documents with blocks of text that need to be output as structured data for downstream processing, rather than a text blob.

Enhancing real-time forecasting with neural networks:

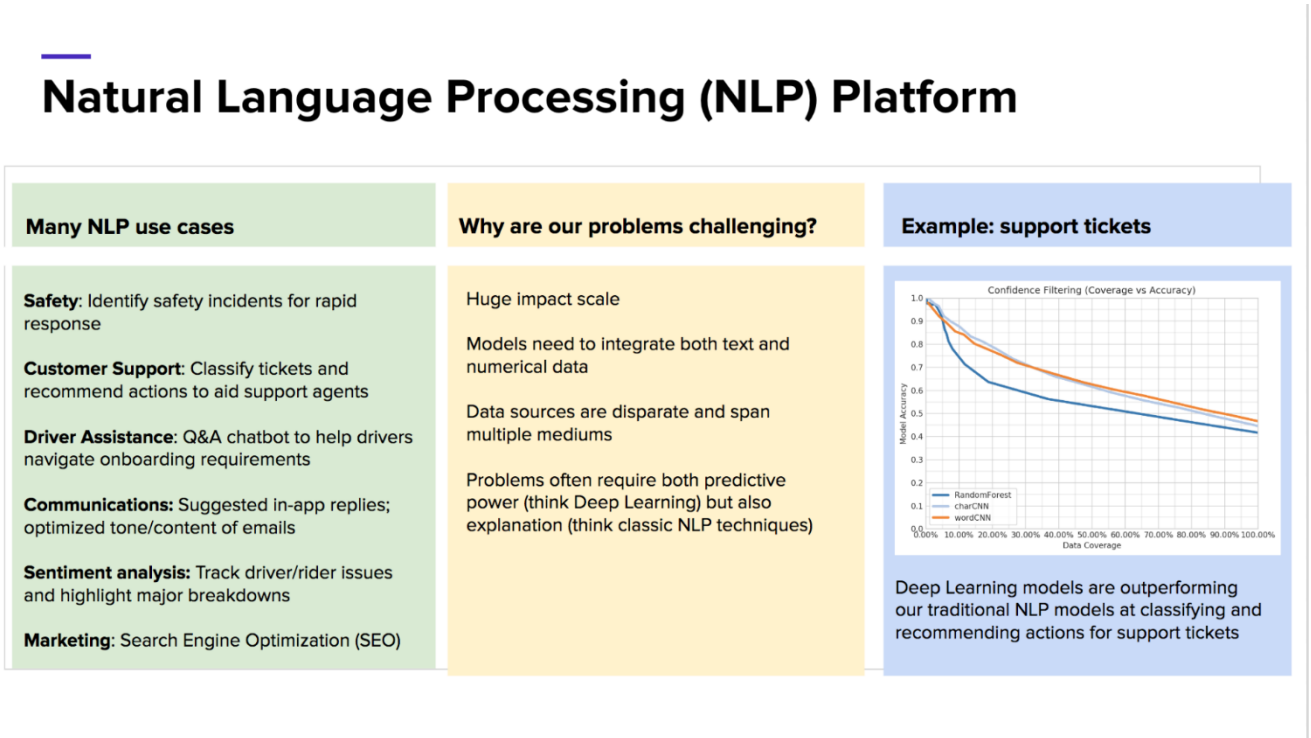
Uber leverages ML models powered by neural networks to forecast rider demand, pick-up and drop-off ETAs, and hardware capacity planning requirements, among other variables that drive our operations.

Creating more seamless communication with conversational AI:

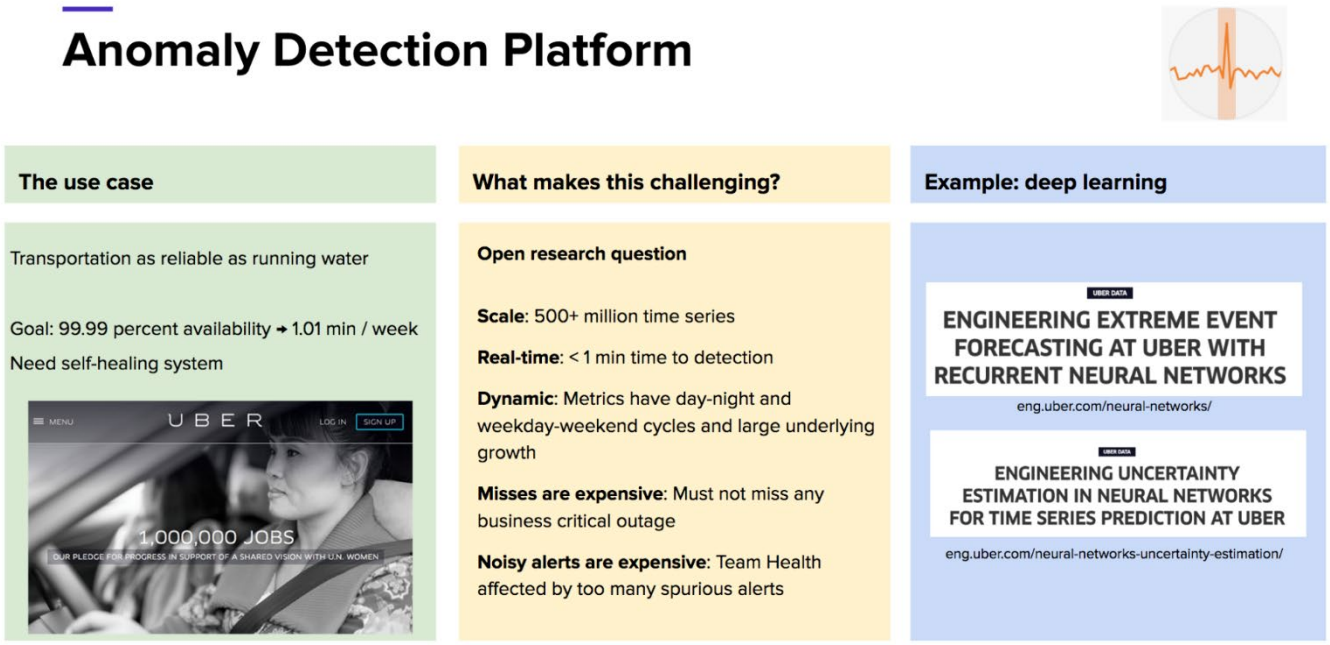
To facilitate the best end-to-end experience possible for users, Uber is committed to making communication with our customers easier and more accessible. In 2019, we leveraged Uber's conversational AI platform, empowering our support teams to resolve user issues as accurately and quickly as possible. Further, we used this platform to lessen the potential for distracted driving by allowing driver-partners to more seamlessly communicate with riders via hands-free pick-up and one-click chat.

Building data science platforms:

To rapidly build models and algorithms that leverage the massive amounts of aggregated data processed from Uber's services. Michelangelo, Uber's ML-as-a-service platform, allows users at the company to query data, generate features, and apply a host of ML models to solve problems in production. Advanced Technologies Group (ATG), who develop our self-driving vehicle technologies, UberEATS, Advertising, and Marketing are just a few of the teams that leverage this powerful platform. ML-enabled technology includes our Natural Language Processing (NLP) platform, which generates and deploys actionable responses for our customer support tickets, chatbots to make driver onboarding easier, and suggested in-app replies. we have been using our NLP platform along with deep learning models to improve the recommended actions and turnaround times for our support tickets.

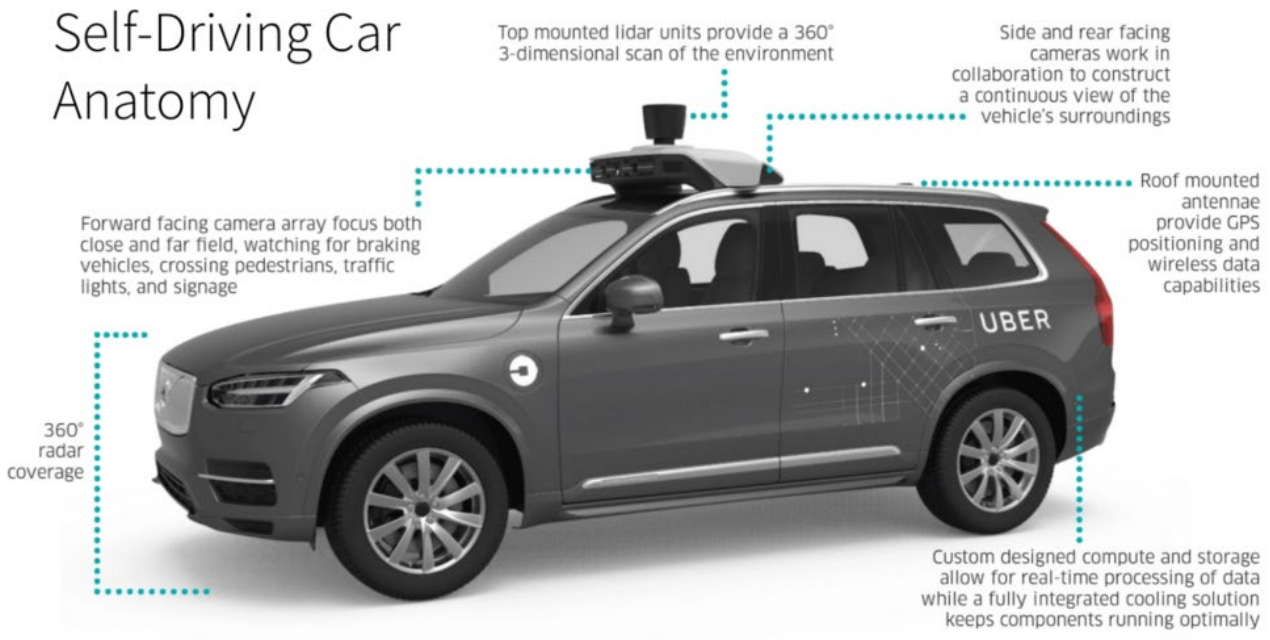


Alerting thresholds are constantly adjusted without human intervention, so we are always ahead of any potential business critical outage. When ensuring safe and reliable transportation for millions of people daily, a system outage can have a huge impact.



Engineering self-driving vehicles with Uber ATG:

Globally, 1.3 million people die in car accidents each year; in the U.S. alone, 94 percent of fatal accidents are a result of human error. Uber is committed to developing technologies that put a dent in these statistics by increasing safety for both current and future users. Powered by ML and AI, self-driving vehicles will play a significant role in making the roads safer.



4. Name all the data technologies UBER is using?

Backend Development: Java, Node.js, PHP, Python, Ruby on Rails

Cloud Servers: Adobe, Amazon Web Services (AWS), Google Cloud Platform (GCP), Microsoft Azure

Database: Firebase, MongoDB, MySQL

Email Notification: Mandrill, SMTP

GPS API/Geo-location: Google Maps, OpenLayers, HERE, Mapbox, TOMTOM, MapFit, Jawgmaps

Mobile- Based Frontend: Java, Kotlin for Android & Swift for ios

Payment Gateway: Braintree, PayPal, Paytm, Stripe

Push Notification: Google Firebase Cloud Messaging (FCM), Apple Push

Real-Time Connection: Socket.IO, WebRTC (Voice & Video call)

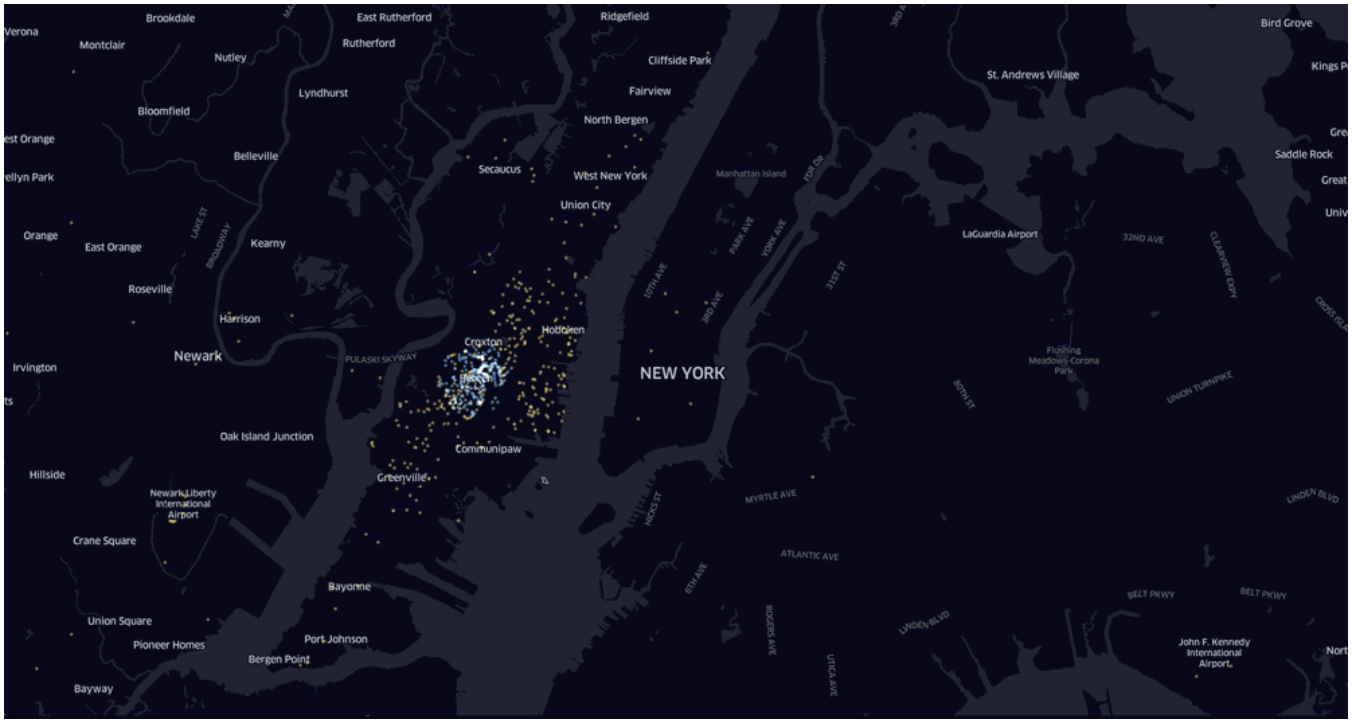
SMS notifications: Nexmo, Plivo, Snich, Tropo and Twilio

Web-Based Frontend: Angular, Backbone, Ember, React.JS

5. Why Data Visualization is key Analytics area for UBER?

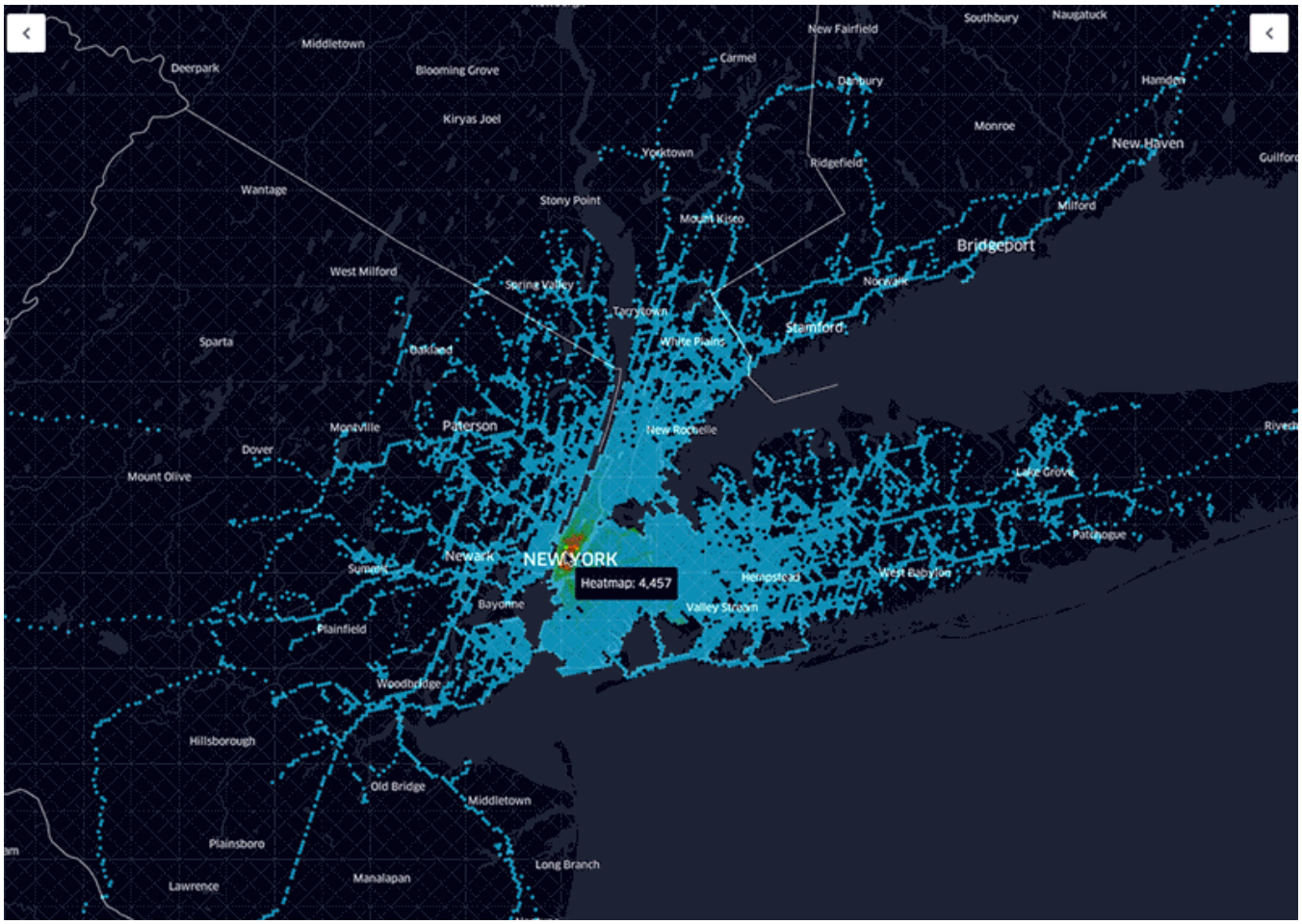
According to Uber's own data intelligence blog, data visualization specialists range from computer graphics professionals to information design. They handle everything from mapping and framework developments to data that the public (such as drivers) sees. And a lot of these data extrapolations and visualizations have never been done before, which has created a need for tools to be developed in-house. Without getting too technical, some of the many applications for their data visualization challenges include:

Mapping Applications for City Ops Teams and General Managers:



These are the teams at Uber who need up-to-the-minute details of current supply and demand. In the same vein, marketing professionals might need aggregate data to plan a campaign. So the Uber engineering team built a system that would show distributions of Uber drop-offs in real-time as you drag your cursor over a given area.

Another example is of importance in big cities, where understanding the density of a given area may lead to dynamic pricing changes. Uber demonstrates this with a combination of layers that let them drill-down to see specific areas in more detail:



But these are not just data visualizations for engineers and data scientists to pore over. Data visualization also helps the public better understand what Uber does and how it works.

References:

1. https://eng.uber.com/uber-ai-blog-2019/
2. https://eng.uber.com/machine-learning/
3. https://eng.uber.com/machine-learning/
4. https://medium.com/acing-ai/uber-case-study-the-state-of-autonomous-transport-42bf2da90fb3
5. https://datascienceai365.com/case-study-how-uber-uses-data-science-machine-learning/
6. https://neilpatel.com/blog/how-uber-uses-data/