UBER Data Science and Analytics
Case Study

Mohamed Megahed

4/22/2020

Introduction:

Uber has transformed the world. Indeed, its inconceivable to think of a world without the convenience of the innovative ride sharing service. Tracing its origins in a market which is constantly being deregulated, Uber has emerged triumphant. Operating in over 58 countries and valued roughly at US$ 66 billion, Uber has rapidly expanded to established branches in over 581 cities in over 82 countries with the United States, Brazil, China, Mexico and India being Uber's most active countries.

▪ How Uber is leveraging AI and Machine Learning?

To rapidly build models and algorithms that leverage the massive amounts of aggregated data processed from Uber's services, Uber has built several data science platforms. These platforms enable its data scientists to create technologies that increase the effectiveness and efficiency of products and operations.

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.

Another example of our ML-enabled technology includes Natural Language Processing (NLP) platform, which generates and deploys actionable responses for customer support tickets, chatbots to make driver onboarding easier, and suggested in-app replies. With Uber's commitment to its driver partners, using NLP platform along with deep learning models to improve the recommended actions and turnaround times for support tickets.

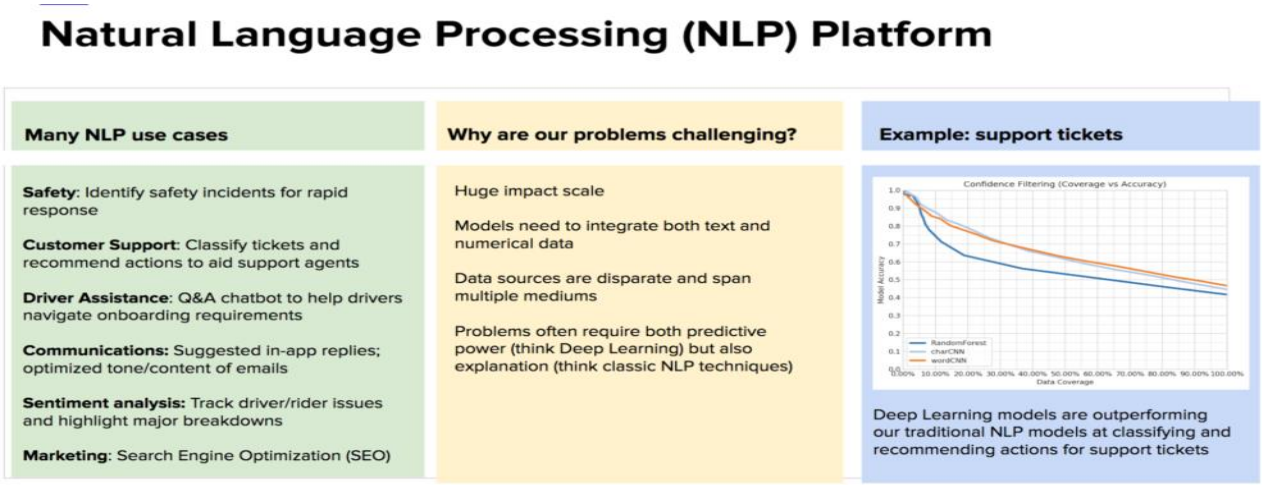


Figure 1: Customizable NLP platform enables Data Scientists to rapidly build models for chatbots, sentiment analysis, and rapid response for support tickets.

ML is also used to improve internal engineering production systems. To ensure reliability of our services during all hours, we use it to ease the on-call responsibilities of our engineers through our anomaly detection platform. This tool constantly monitors tens of thousands of service metrics to tease out any spurious alerts. We use a combination of recurrent neural networks and novel feature extraction techniques, so the system learns the patterns of these metrics, including day-night and weekday-weekend cycles.

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.

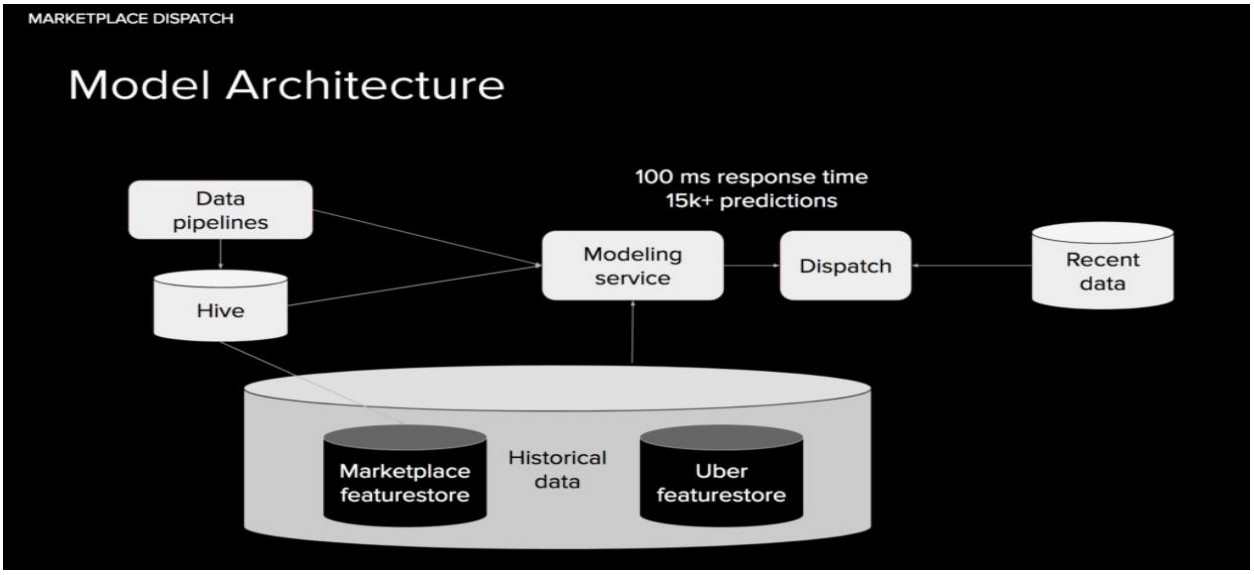


Figure 2: Dispatch matching leverages a thousand of features to generate thousands of predictions in sub-second timing.

To date, Uber's self-driving vehicles have completed over 30,000 real world passenger trips in places like Pittsburgh, PA and Tempe, AZ. In these environments, Uber's self-driving vehicles leverage ML to inform how they moves through space, helping our automated systems understand the differences between stationary and moving vehicles, pedestrians, and everything in between.

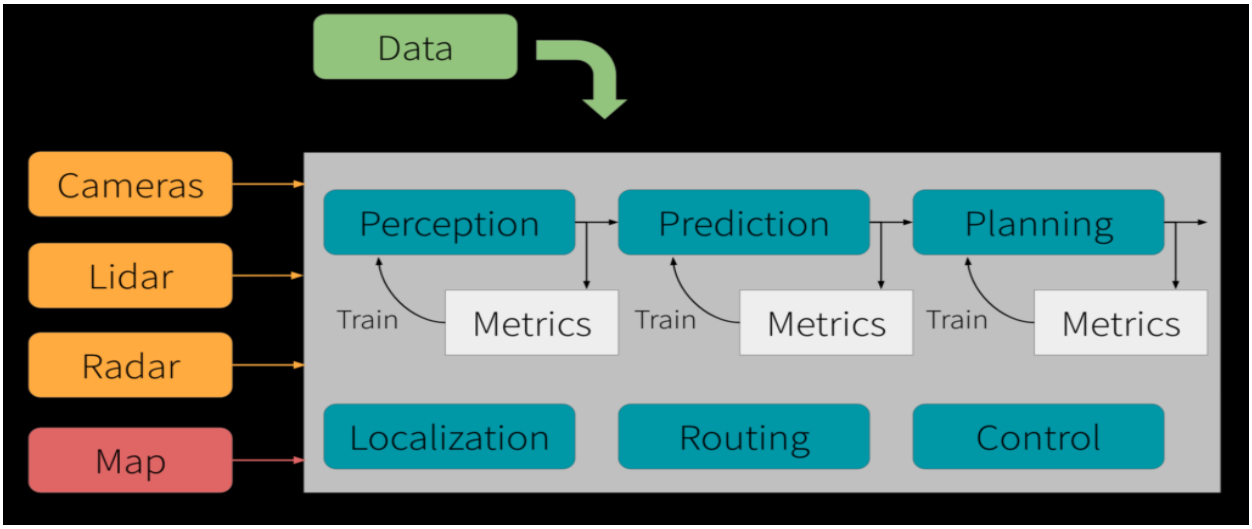


Figure 3: Multiple inputs are used for training along with intermediate feedback loops to build learning for end-to-end self-driving.

- **Describe the Analytics process followed by the company?**
  - 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 of 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.

  - Uber also gathers data on its drivers. In addition to collecting non-identifiable information about their vehicle and their location, Uber also monitors their speed and acceleration, and checks to see if they are working for a competing company as well (such as Lyft).

  - *Uber uses your personal data in an anonymized and aggregated form to closely monitor which features of the Service are used most, to analyze usage patterns and to determine where we should offer or focus our Service. We may share this information with third parties for industry analysis and statistics.*

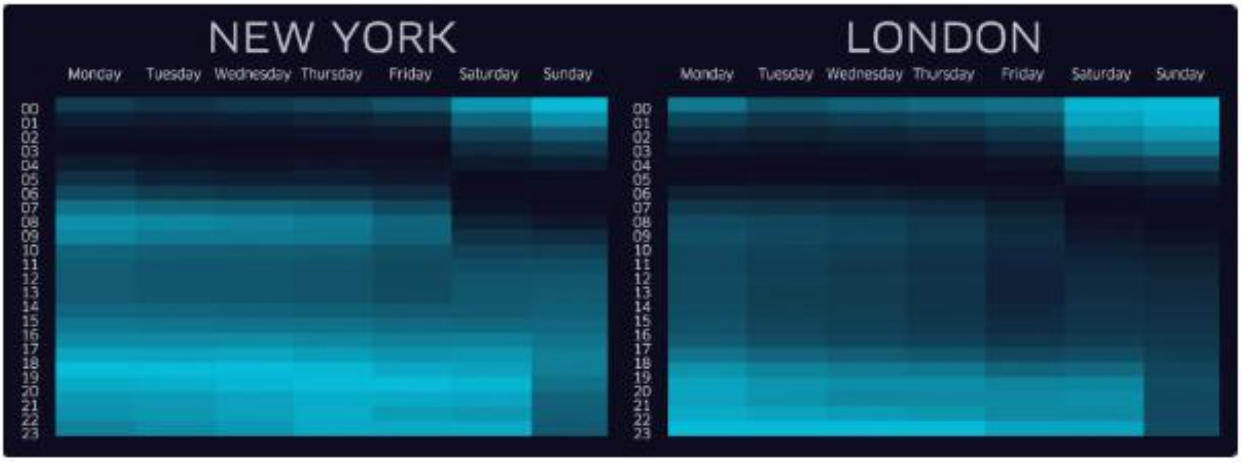
  - All this data is collected, crunched, analyzed and used to predict everything from the customer's wait time, to recommending where drivers should place themselves via heatmap in order to take advantage of the best fares and most passengers. All of these items are implemented in real-time for both drivers and passengers alike.

Monitoring Supply and Demand:

One of Uber's biggest uses of data (and likely the one that draws the greatest ire from passengers) comes in the form of surge pricing, a model nicknamed "Geosurge" at Uber. If you're running late to an appointment and you need to book a ride in a crowded downtown.

In the short term, surge pricing substantially affects the rate of demand, while long-term use could be the key to retaining or losing customers. Customer backlash on rate-hiking is strong, so Uber has considered using machine-learning algorithms to predict where demand will be strong, so that drivers can adequately prepare to meet that demand, and surge pricing will be significantly reduced. This new system has not yet been released, but Uber knows that in order to get and maintain a strong customer and driver base, it needs to put data to work for it in new and innovative ways.

However, that supply and demand data are not the same from city to city, so Uber engineers devised a way to map the "pulse" of a city to connect drivers and riders more efficiently. And if you think all major metropolitan cities are alike – think again. Just look at how New York City compares to London:



When Uber trips occur throughout the week in New York City and London. The brightness levels per hour and day are compared to the city itself. All times are standardized to the local time zone and expressed in military time (i.e. 20 is 20:00, or 8 pm).

Passengers are out and about in New York City during the evenings, while London is bright well into the night — demonstrating that what works well for Uber in one city doesn't necessarily correspond to another.

▪ What are the big problems UBER is trying to solve using ML and AI?

The complexity of Uber's problem space goes beyond clicking web links to purchase a car—Uber deal with the car itself.
Massive complexity exists when modeling interactions from the movement and routing of cars to optimizing thousands of interconnected entities on each side of a marketplace in real time.

As Uber's network of users grows, so too does the number of sensors that help Uber understand the physical world. they use this information to build maps, optimize their marketplace, and train our self-driving cars.

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.

- **Name all the data technologies UBER is using?**

The tech side of the App is written largely in JavaScript which is also used to calculate supply and predict demand. With the real time dispatch systems being built on Node.js and Redis. Java, as well as Objective-C is used for the iPhone and Android apps. Twilio is the force behind Uber's text messages, and push notifications are implemented through Apple Push Notifications Service on the iOS platform and Google Cloud Messaging (GCM) for the Android App.

This is a list of technologies used by Uber

- Node.js
- AngularJS
- Machine learning
- Redis
- Mysql
- Mango DB
- Python
- Java
- Backbone.js
- ExpressJS
- Nylas
- RequireJS

So, to build an App like Uber, we need to have:

1. Registering/Log-in features: Uber allows you to register with your first name, last name, phone number and preferred language. Once you've signed up, they'll send you an SMS to verify your number, which will then allow you to set your payment preferences.
2. Booking features: This allows drivers the option to accept or deny incoming ride requests and get information on the current location and destination of the customer.
3. The ability to Identify a Device's location: Uber, via CoreLocation framework (for iOS platforms) obtains the geographic location and orientation of a device to schedule location and delivery. Understanding iOS and Android geolocation features is crucial for this step, because that's what your App is running on.
4. Point to Point Directions: The Uber App provides directions to both the driver and the user. Developers of the Uber App use MapKit for iOS and Google Maps Android API for Android to calculate the route and make directions available. They further implemented Google Maps for iPhone and Android, but cleverly adapted technology from other mapping companies to solve any logistical issues that might come up.
5. Push Notifications and SMS: You get up to 3 notifications instantly from Uber when you book a ride.

- A notification telling you when the driver accepts your request
- One when the driver is close to your location
- One in the off chance your ride has been cancelled

6. Price Calculator: Uber offers a cashless payment system, paying drivers automatically after every ride, processed through the user's credit card. Uber takes 25% of the driver's fare, making for easy profit. They paired with Braintree, a world leader in the mobile payment industry, but other good options avaible are Stripe, or Paypal, via Card.io.

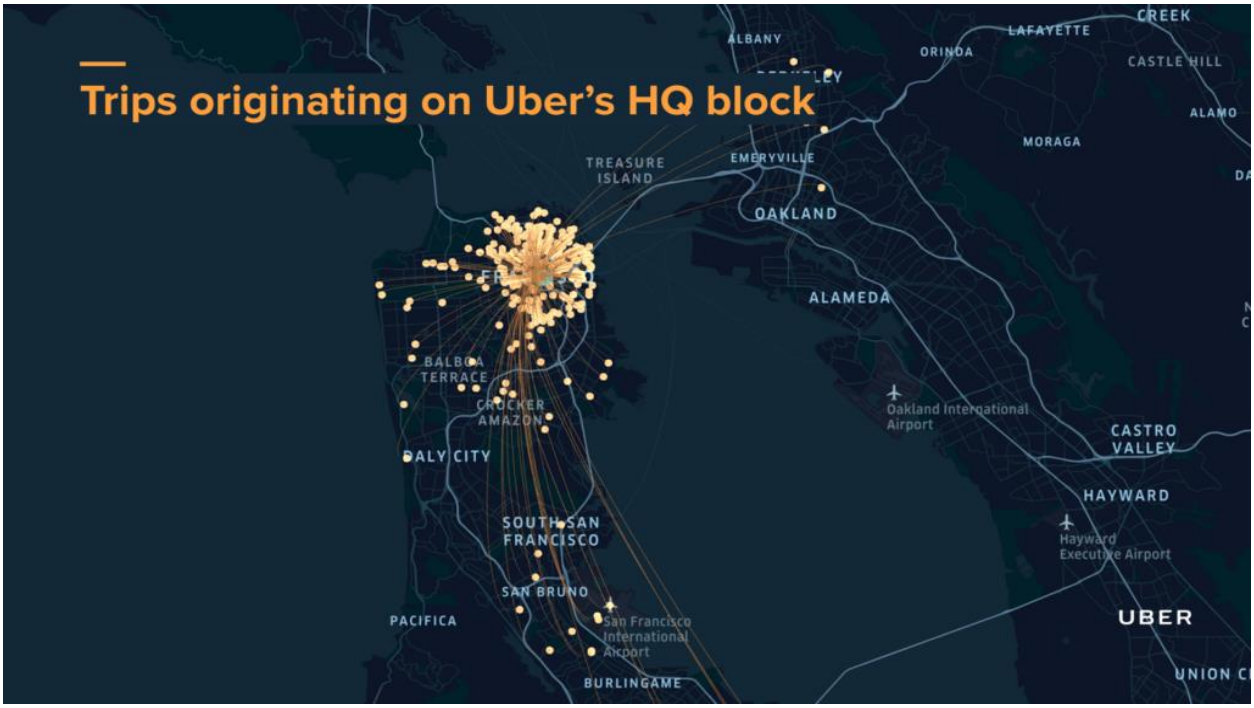
- **Why Data Visualization is key Analytics area for Uber?**

Of course, collecting all this information is just one step in the big data journey. The real question is — how does Uber determine the best way to make decisions using this information? How do they glean actionable points out of the data they collect?
For example, Uber manages billions of GPS locations. Every minute, their platform juggles millions of events. How do they leverage these details into a way to better manage moving people and things from place to place?

The answer is data visualization.

- 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.

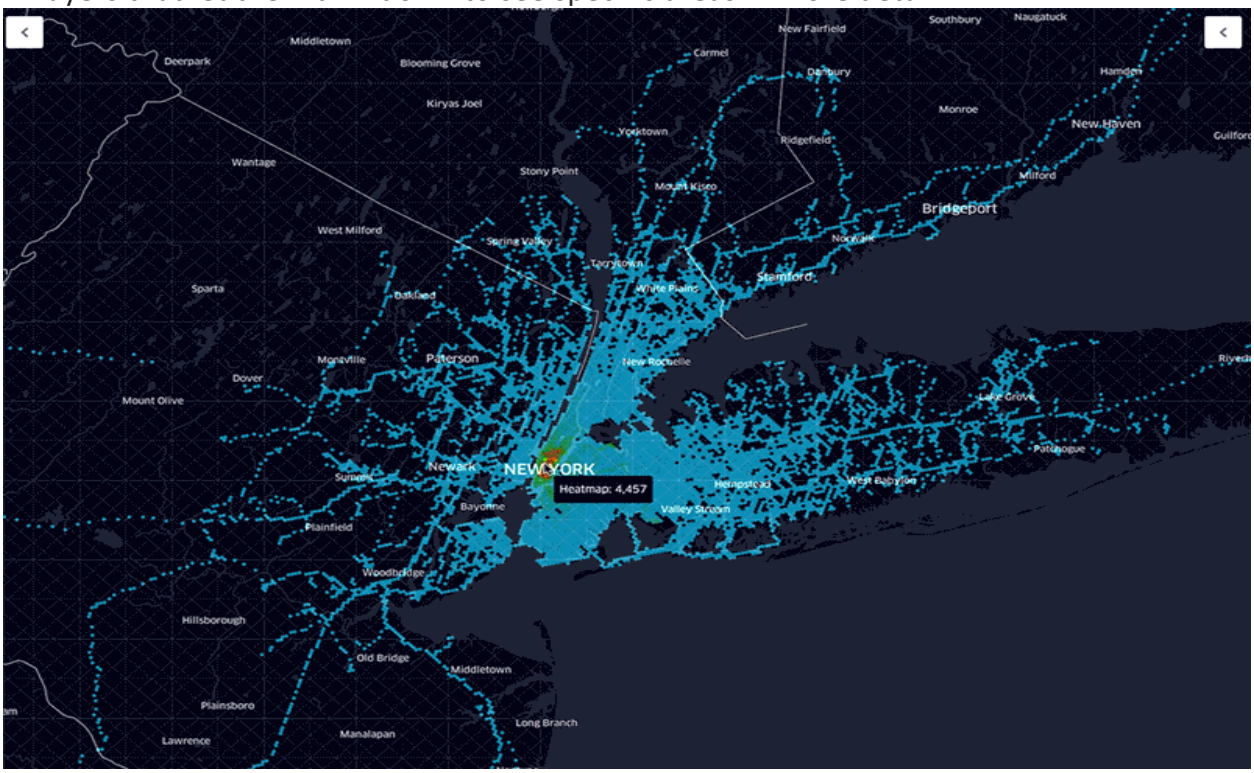
- Without getting too technical, some of the many applications for their data visualization challenges include:
- Mapping Applications for City Ops Teams and General Managers



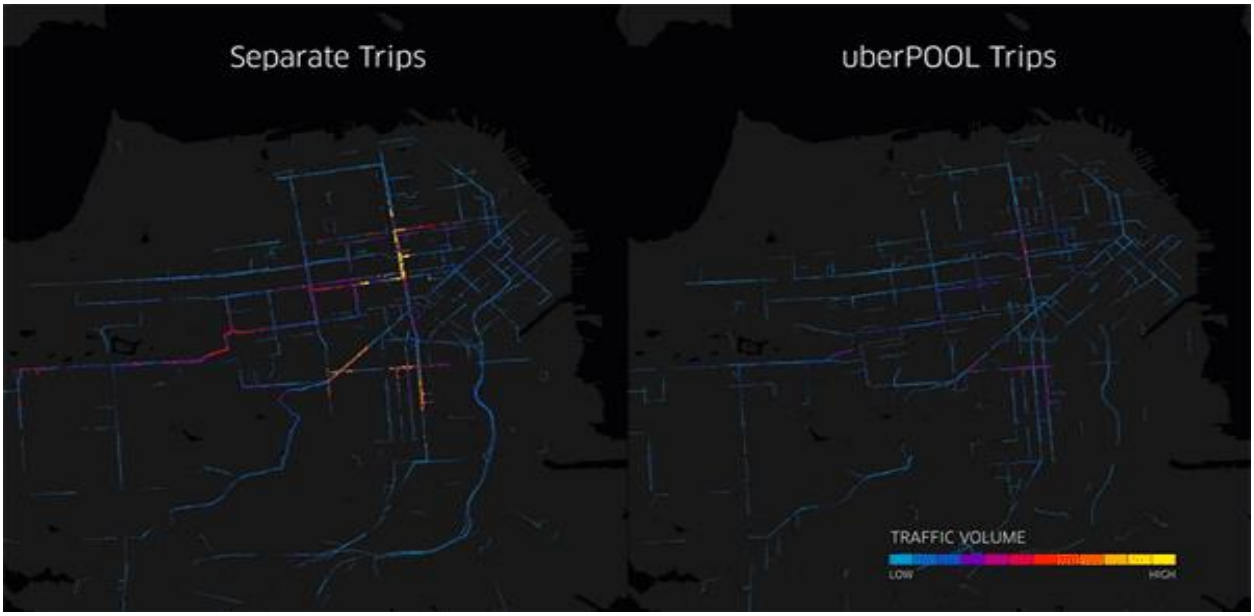
- Figure 4: A visualization highlights trips originating from Uber's HQ block in San Francisco.

- 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 aren't 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, such as this visualization of how uberPOOL helps reduce traffic — a visualization that was shared during founder Travis Kalanick during his TED Talk:



- Figure 5: A visualization highlights the traffic volume in city streets.

References:

https://www.uber.com/us/en/careers/teams/data-science/

https://neilpatel.com/blog/how-uber-uses-data/

https://eng.uber.com/machine-learning/?_ga=2.91919834.1533279776.1587940474-2022902129.1587940474

https://www.quora.com/What-is-the-technology-stack-behind-Uber

https://www.freecodecamp.org/news/how-uber-was-made-da3c631066d0/