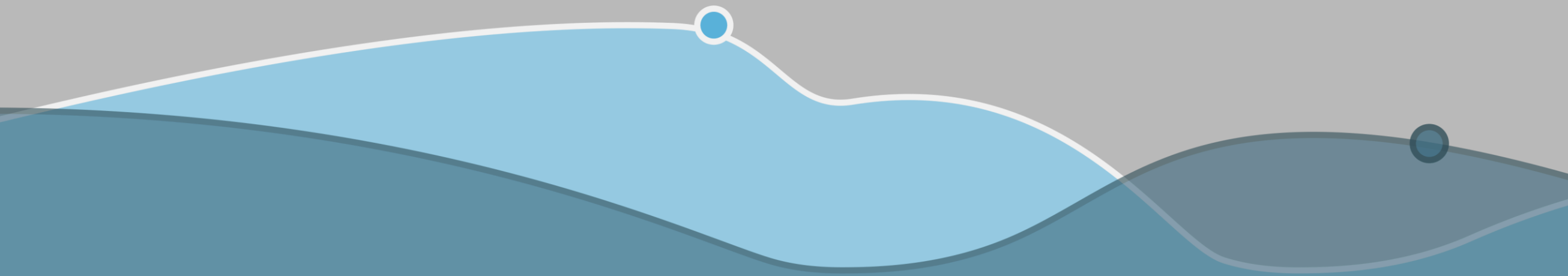




# Cortana Analytics Workshop

Sept 10 – 11, 2015 • MSCC



# Using the Cortana Analytics Process

Sachin Chouksey  
Principal Software Engineering Manager

Wee Hyong Tok  
Senior Program Manager



# Cortana Analytics Suite

## Transform data into intelligent action

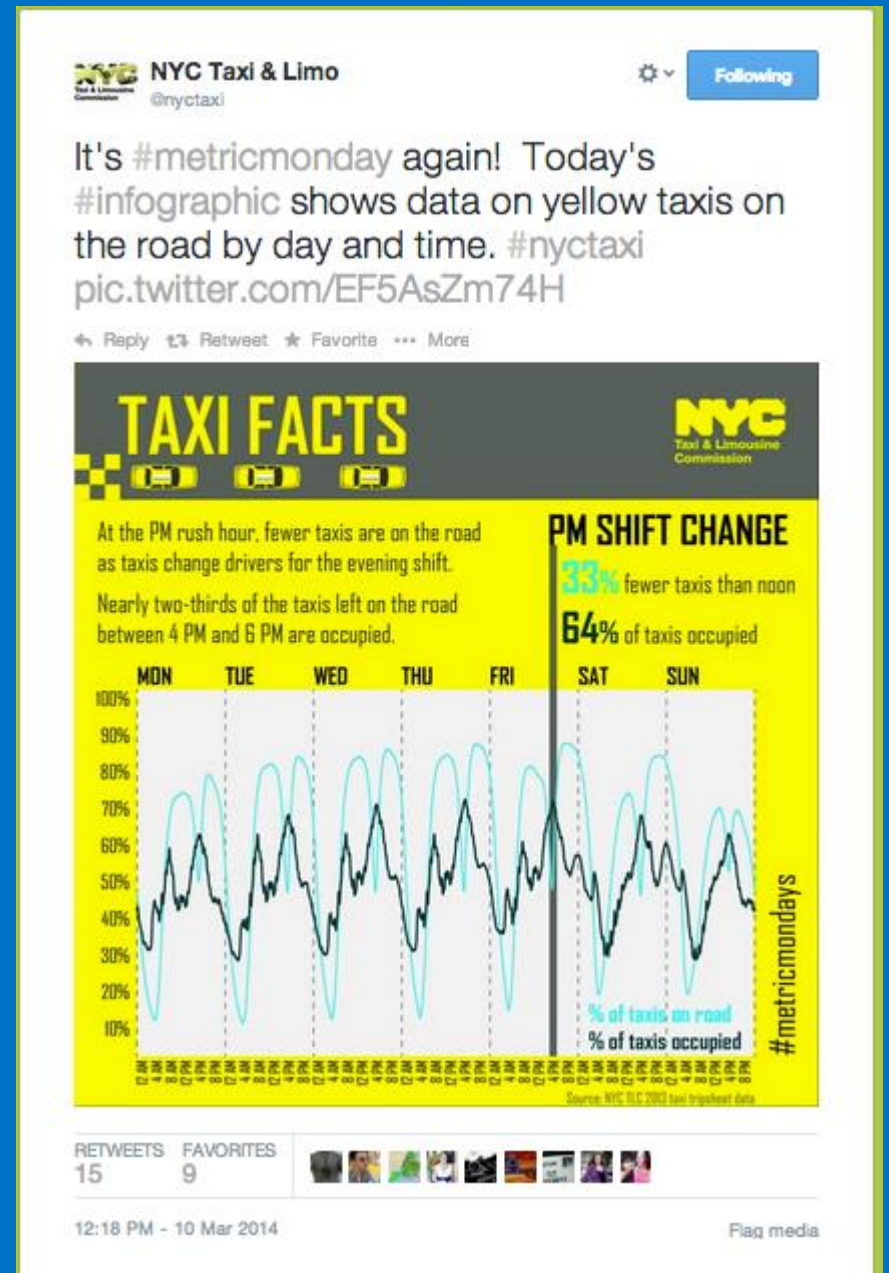


I have data.....

.....and I want to transform it into intelligent action

# The Data

Requested via a freedom of information law (FOIL) request to the NYC Taxi & Limo commission by Chris Whong, a data analyst in New York City.



# New York City Taxi Dataset

## Taxis taken and fares paid in NYC

For 12 months: January 2013 to December 2013

~20GB of compressed CSV data (~48GB uncompressed)

>173 million rows

## Each record includes

Pickup and drop-off location and time

Hack license number (driver's license/ID number)

Medallion number (a permit to operate a yellow taxi cab in New York City, i.e. a Car ID)

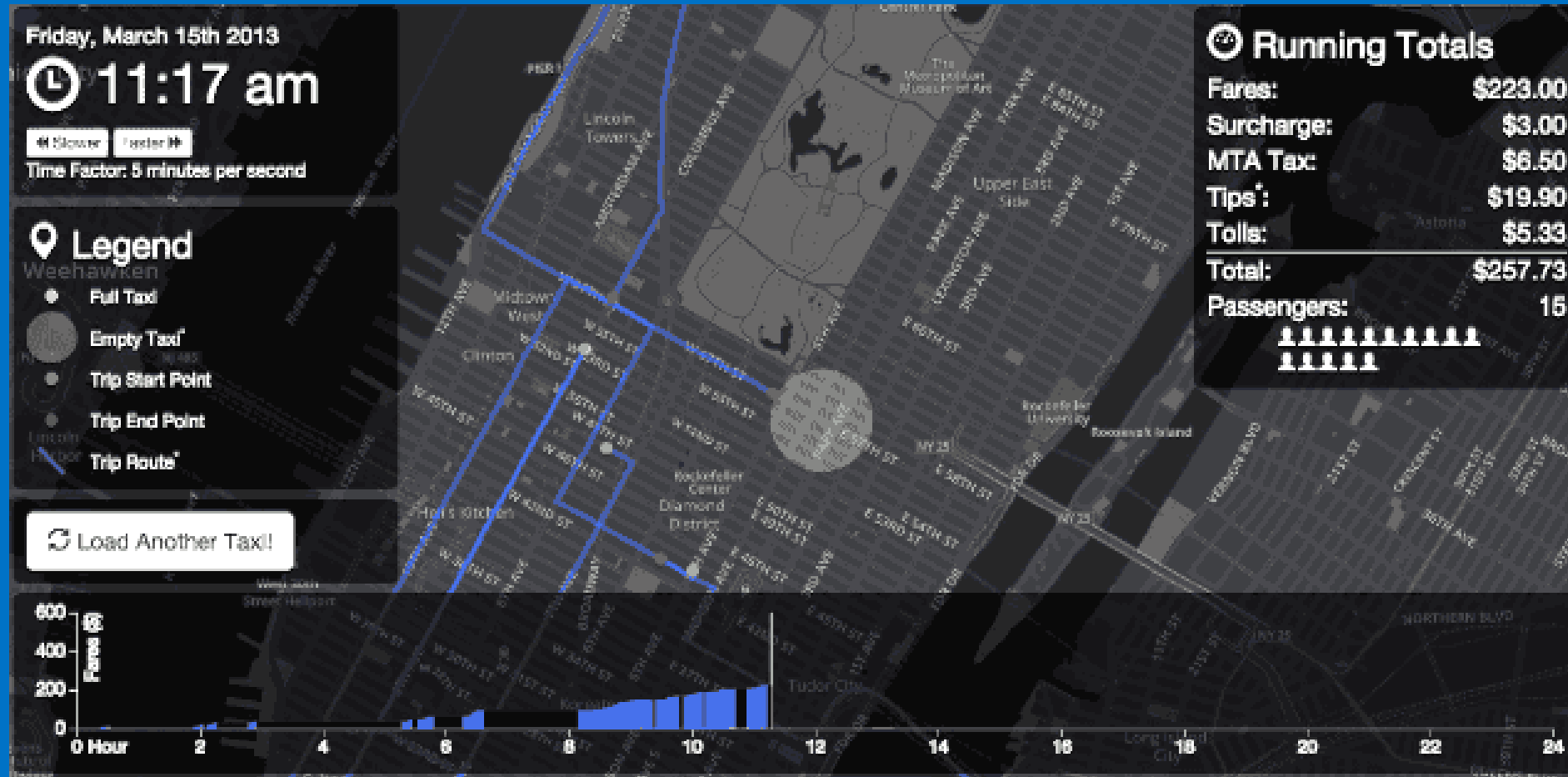
## Two sets of files

trip\_data CSVs contain trip details

trip\_fare CSVs contain details of fare paid

Unique key to join trip\_data and trip\_fare: medallion, hack\_licence, and pickup\_datetime

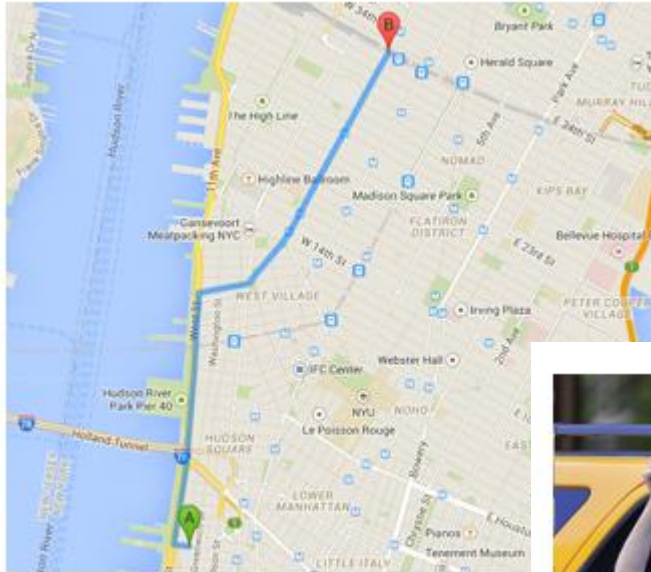
# New York City Taxi Dataset



# New York City Taxi Dataset



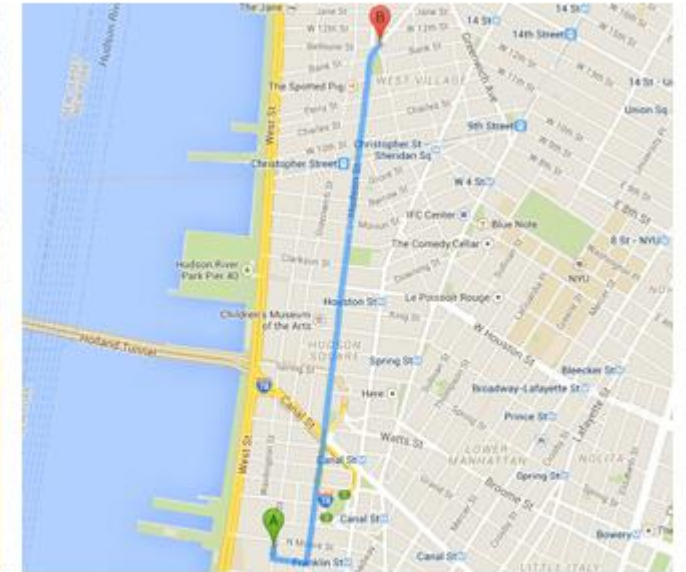
KATHERINE HEIGL



OCTOBER 4, 2013 • 1:21 PM - 1:40 PM  
80 N. MOORE ST. TO 421 8TH AVE  
\$14.50 FARE • \$3.62 TIP • ©WENN



JUDD APATOW  
LESLIE MANN



JUNE 21, 2013 • 11:28 AM - 11:35 AM  
376 GREENWICH ST. TO 1 ABINGDON SQUARE  
\$7.00 FARE • \$2.10 TIP • ©SPLASH



# New York City Taxi Dataset

## What can we predict?

Was a tip paid for the trip? (Binary classification)

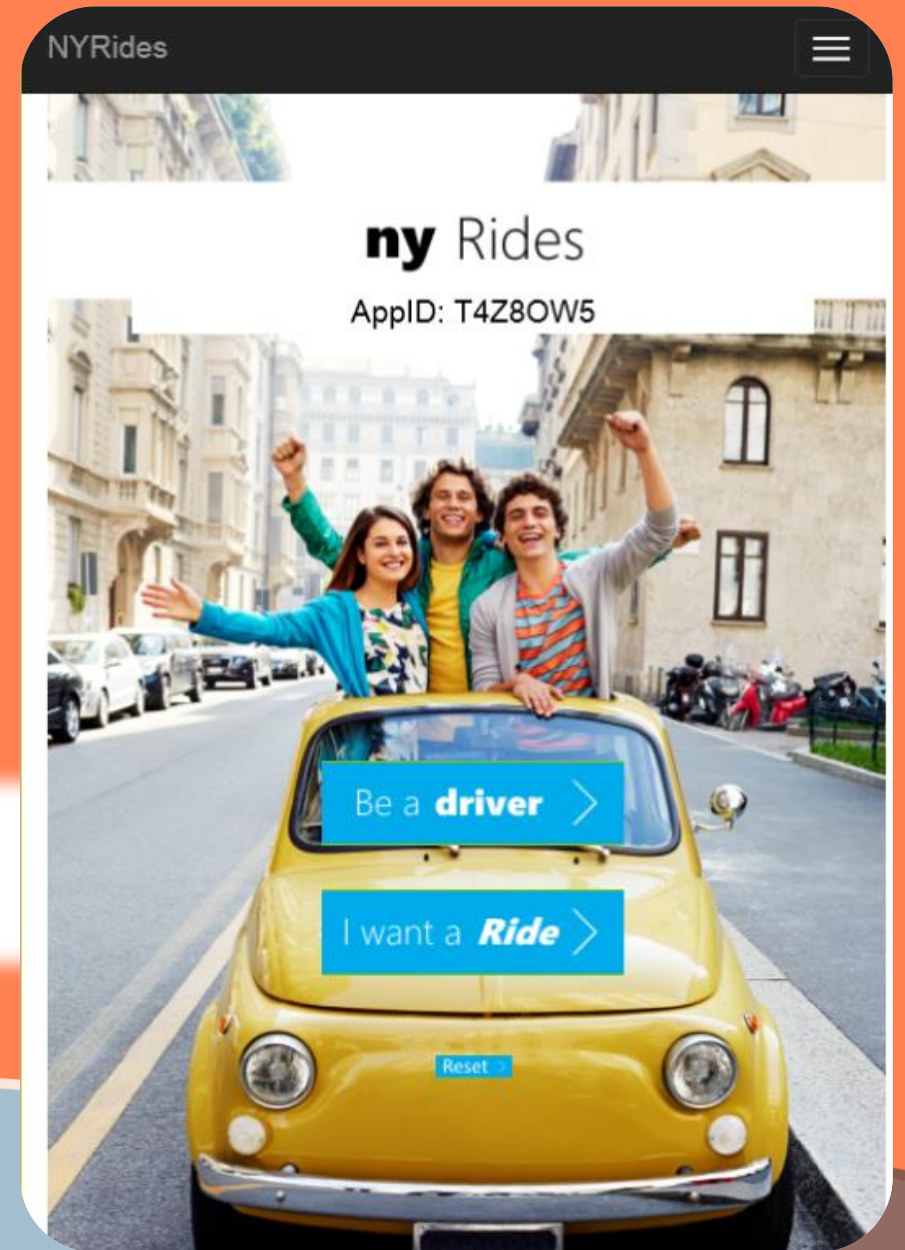
What was the tip amount range? (Multiclass classification)

What was the tip amount? (Regression)

# Demo

Let's make some data!

<http://aka.ms/nyrides>

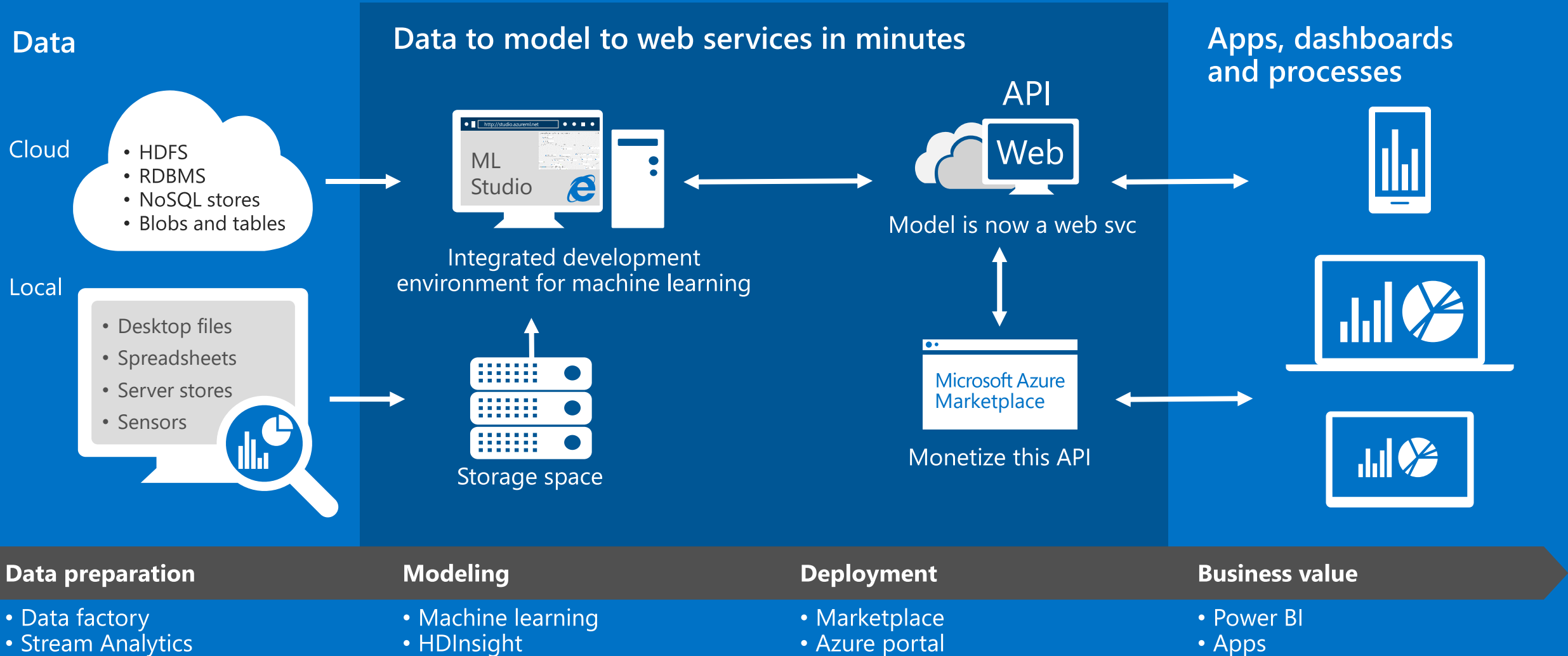


# Cortana Analytics – Demo Recap

## Transform data into intelligent action



# Cortana Analytics Solution Architecture

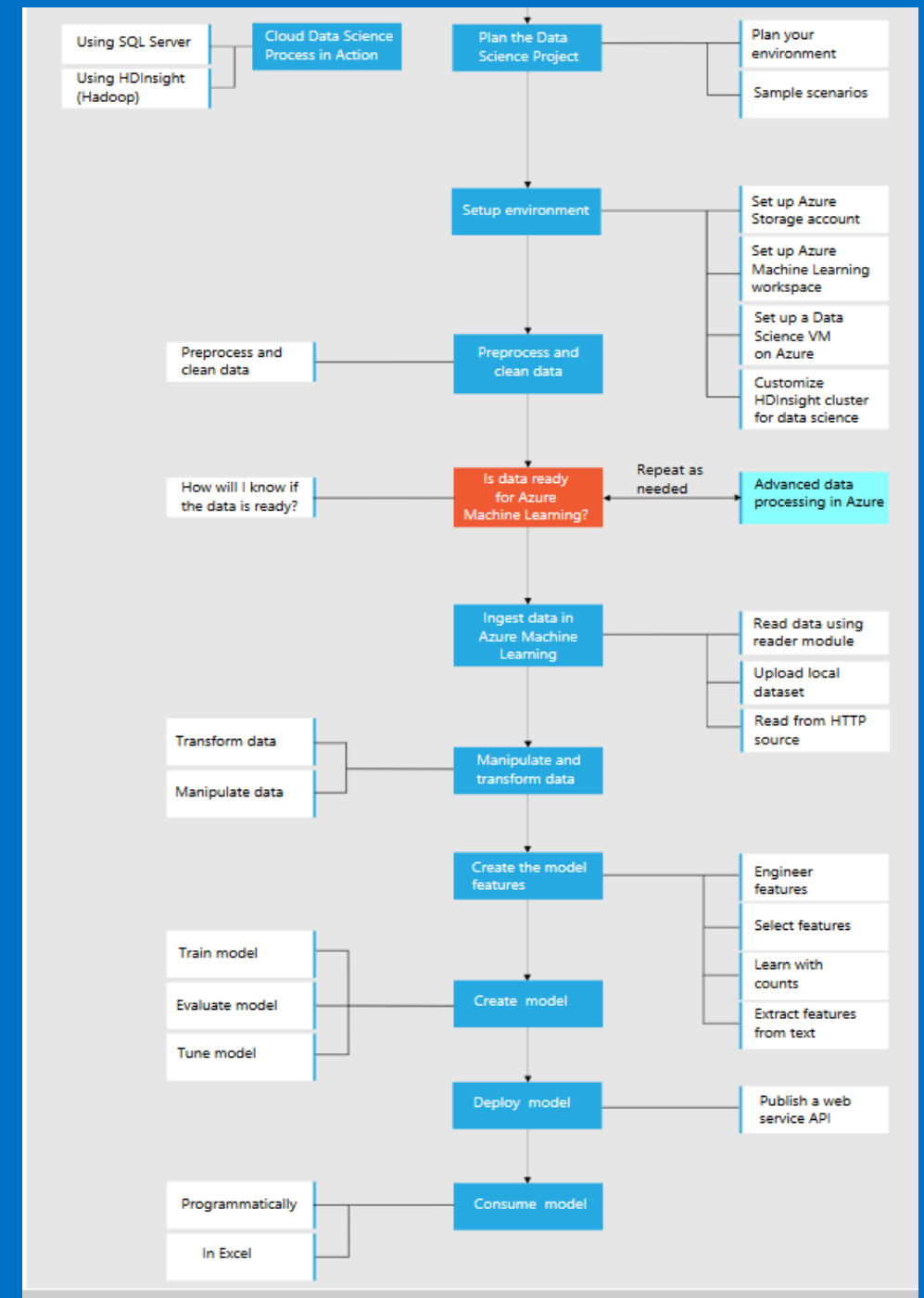


How do I stitch this  
together?

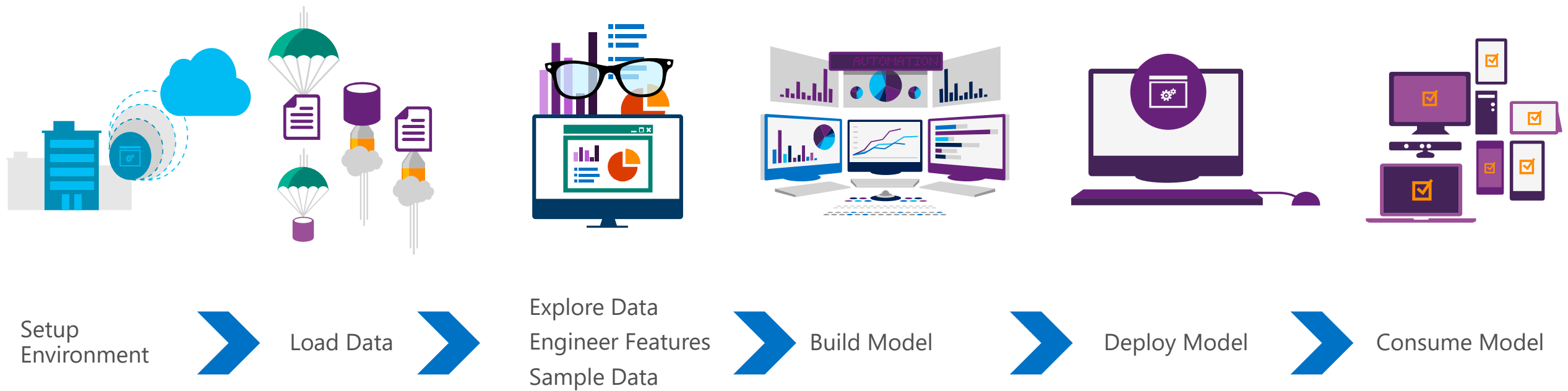
# Cortana Analytics Process

Establish mechanisms to conduct data science activities end-to-end in the cloud or on premises, friction free.

<http://aka.ms/CortanaAnalyticsProcess>



# Cortana Analytics Process

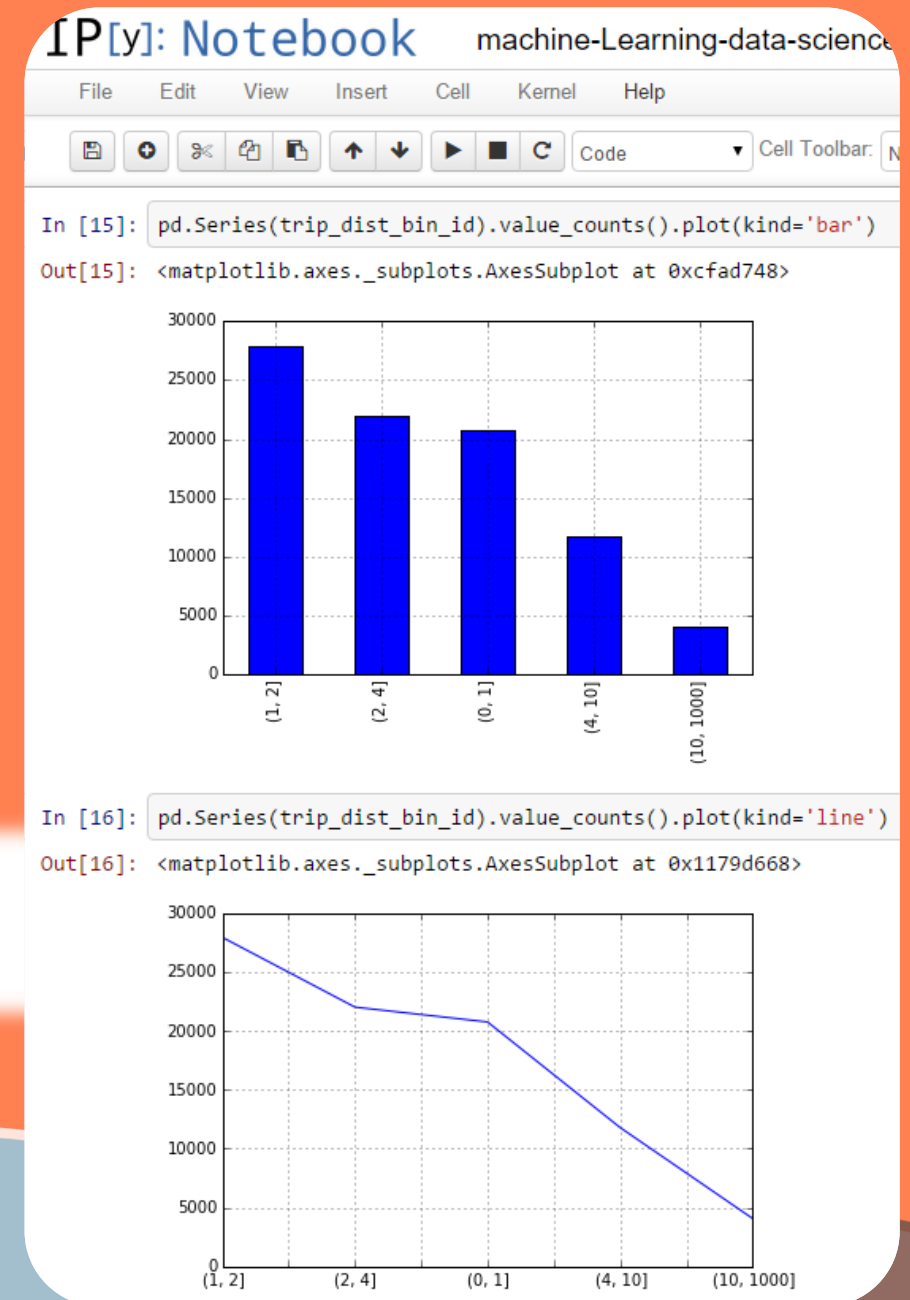


# Demo

Let's analyze some data!

## Getting ready for analytics

<http://aka.ms/cap/sqlwalkthrough>





How do I operationalize?

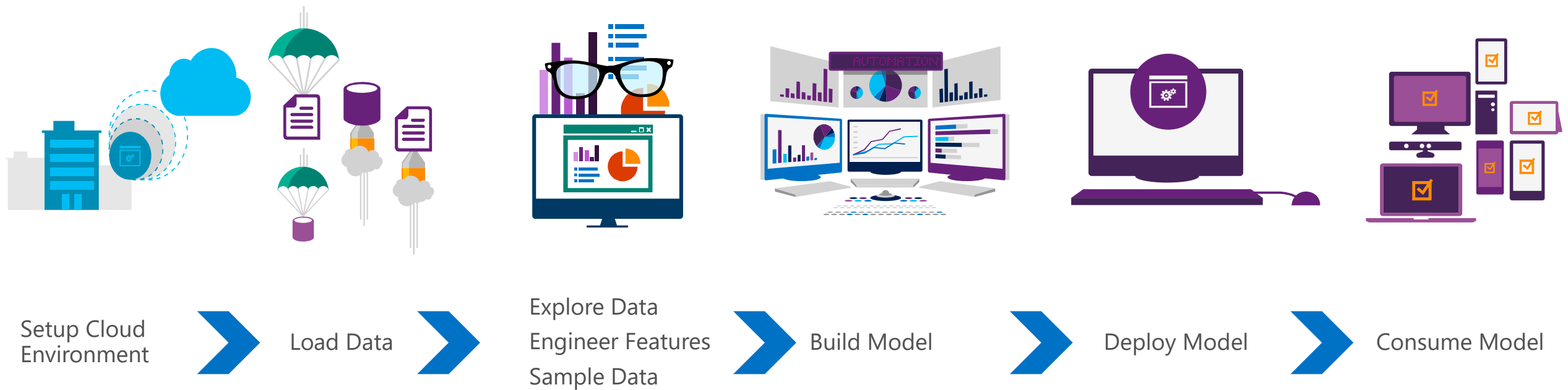
# Demo

Let's dive deeper

Operationalizing end-to-end  
analytics solutions



# Cortana Analytics Process



# Links

## Cortana Analytics Process

<http://aka.ms/CortanaAnalyticsProcess>

<http://aka.ms/cap/sqlwalkthrough>

## Questions or Feedback

[AzureDataScienceProc@microsoft.com](mailto:AzureDataScienceProc@microsoft.com)

## NYC Taxi Data

[http://chriswhong.com/open-data/foil\\_nyc\\_taxi/](http://chriswhong.com/open-data/foil_nyc_taxi/)

<http://gawker.com/the-public-nyc-taxicab-database-that-accidentally-track-1646724546>

