



A New York Minute

Viktoriya Andonova
Linkedin: linkedin.com/vandonova
Github: github.com/vandonova

The Data

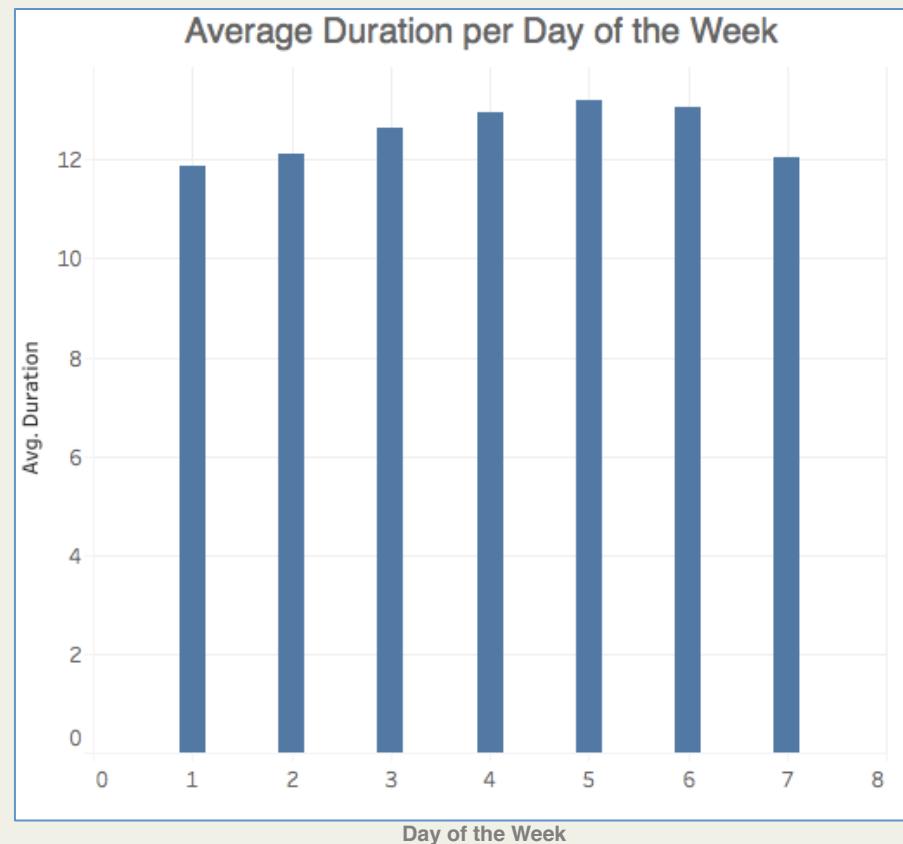
NYC Taxi Trips

6 years (2009-2015)

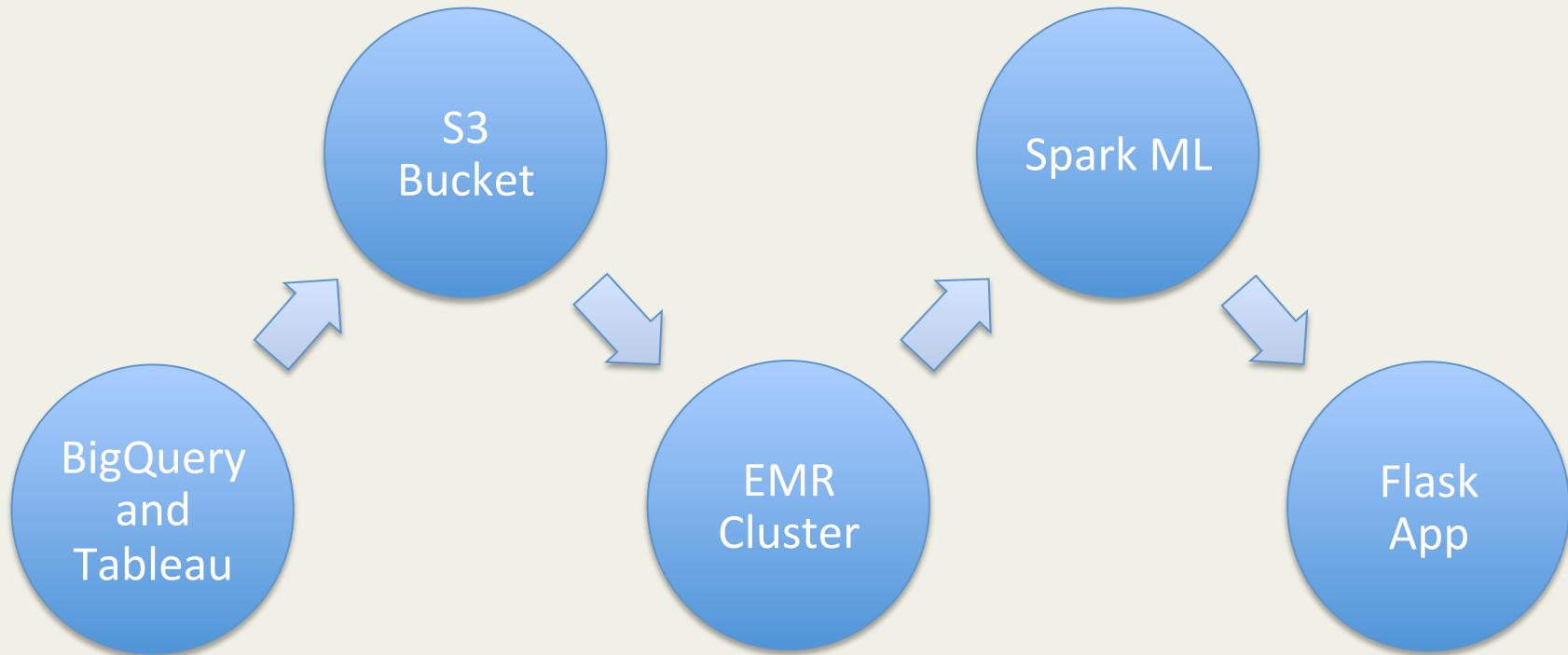
1.1 billion trips

5.6 billion miles

\$14.6 billion



The Workflow



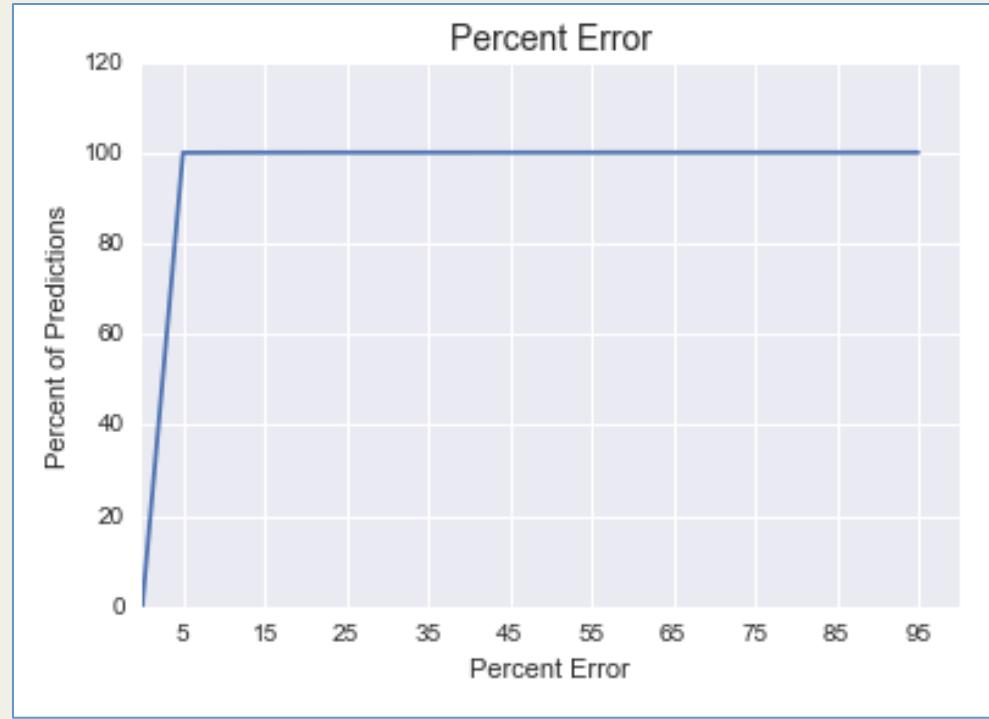
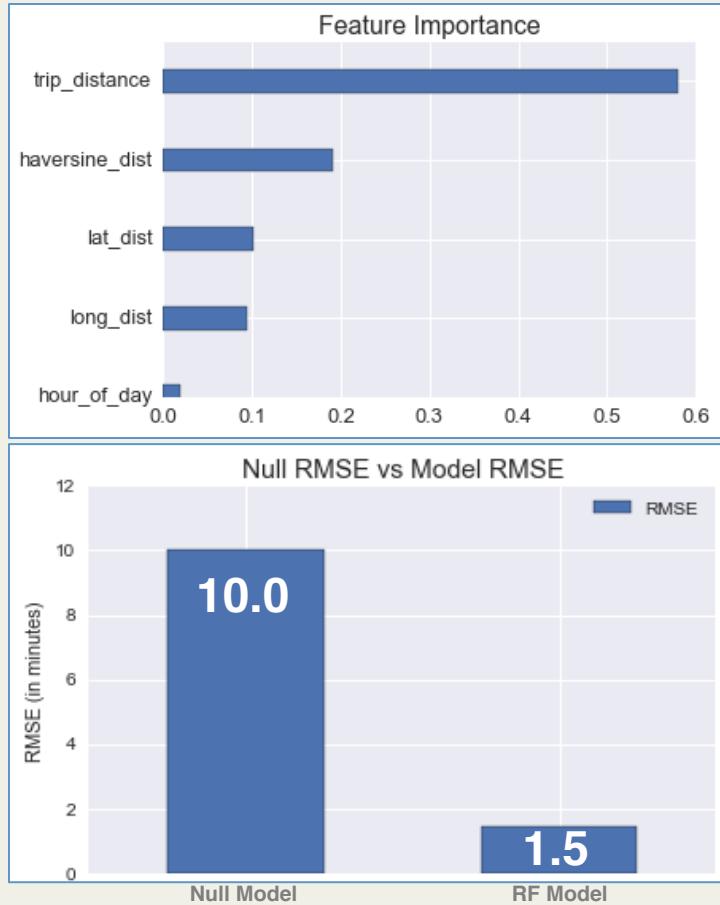
Model Building

Model: Random Forest Regressor

Response: log(duration)

	Space	Time
Features	Pickup lat/long Dropoff lat/long Trip distance Crow flies distance Latitude distance Longitude distance	Day of week Hour
Response		Duration

Results



99.9 % of the observations have < 5% error

A New York Minute

Please enter your trip information

Origin: Columbia University

Destination: Times Square

Flexible

SUBMIT

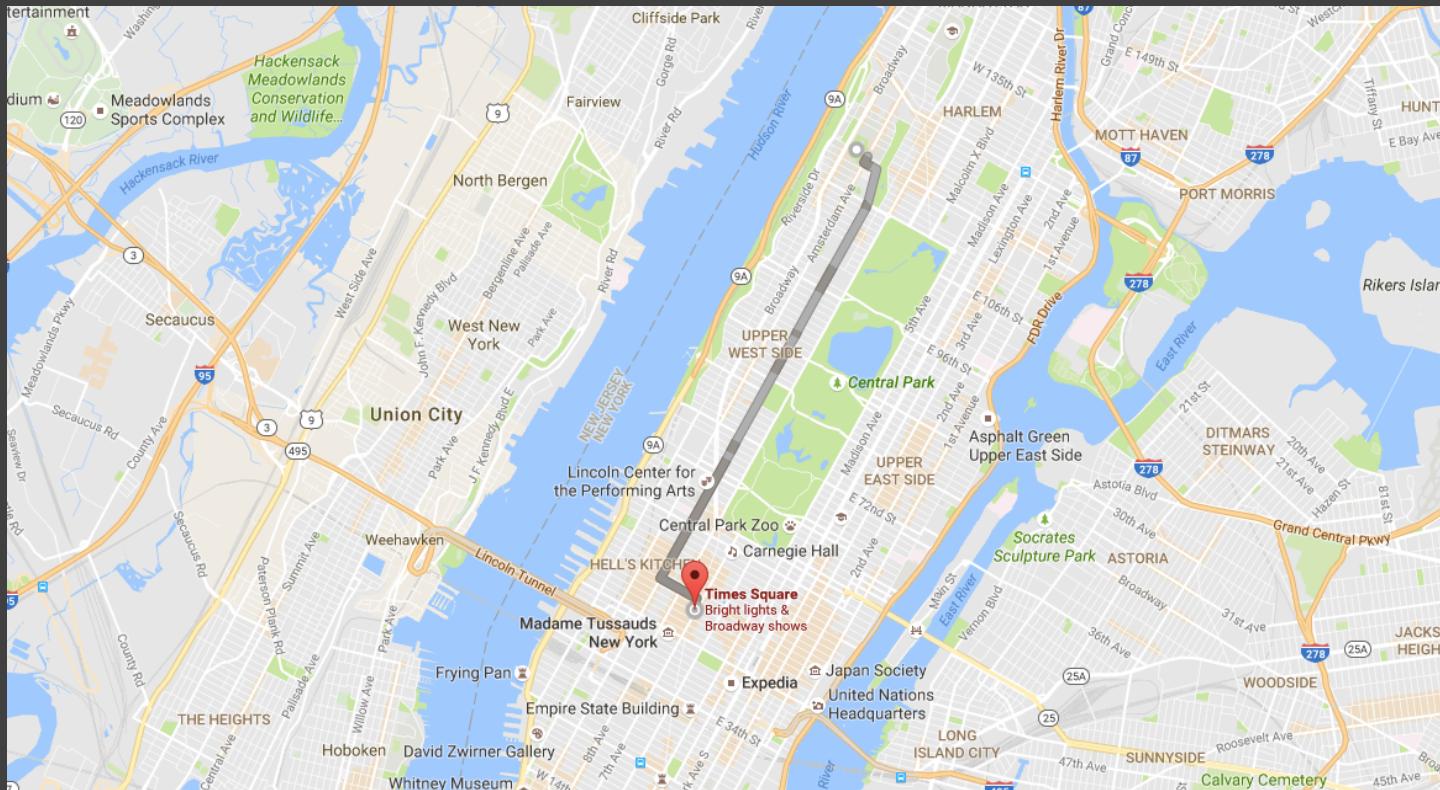
Trip Summary

Prediction: 20 min

99.9% confidence interval: 1 min

You will arrive within 21 minutes!

Sorry, no quicker trips within 2 hours.



Next Steps

Include **weather** and **traffic accident** data in the model

Use **geocoding** to identify routes that cross a bridge

Use the model and traffic data to predict the **price** of a taxi trip

Compare the performance to **Google maps**

Viktoriya Andonova

vba2102@columbia.edu

Thank you

Linkedin: linkedin.com/vandonova

Github: github.com/vandonova