PREDICTING TRANSPORTATION PRICES

Implementing AWS to Leverage NYC Taxi Data

TASK

How can taxi companies compete with rideshare companies like Uber and Lyft?



Use NYC Taxi data and AWS to build a model to predict the price of a trip



Implement into an app allowing the user to see how much they would need to pay prior to ordering a taxi

TOOLS





amazon S3



amazon SageMaker





Demonstrate using AWS with big data



Illustrate big data predictive modeling process



Address a real-world problem



Provide predictive analytics techniques



1.5 GB Data

12 files 12 months of data 1.37 mil rows of data

18 Columns

PROCESS

K<mark>ey</mark> Takeaways

NYC Taxi NYC TAXI **Data** Data Data **Data** warehouse for multiple **Storage** files Data **AWS** Create the table structure Redshift **Preparation** for data to live Use Python or **AWS** Model R to build and SageMaker **Building** test predictive model Model to be **Prediction** used for Model predicting prices

Accurately predicted taxi prices with use of multiple models

Taxi companies should consider using a method such as the one demonstrated here to better compete with ride share services

Opportunities for even more accurate results with better AWS tiers

GROUP

Adam Baker | bake0835@umn.edu Emily Harvey | harve424@umn.edu Ryan McKiernan | mckie023@umn.edu

Robin Peng | peng0363@umn.edu Nikhilesh Murali | naray242@umn.edu