COMPUTER SCIENCE & ENGINEERING
TEXAS A&M UNIVERSITY

2018 TAMIDS Data Science Competition

Team Bulgogi

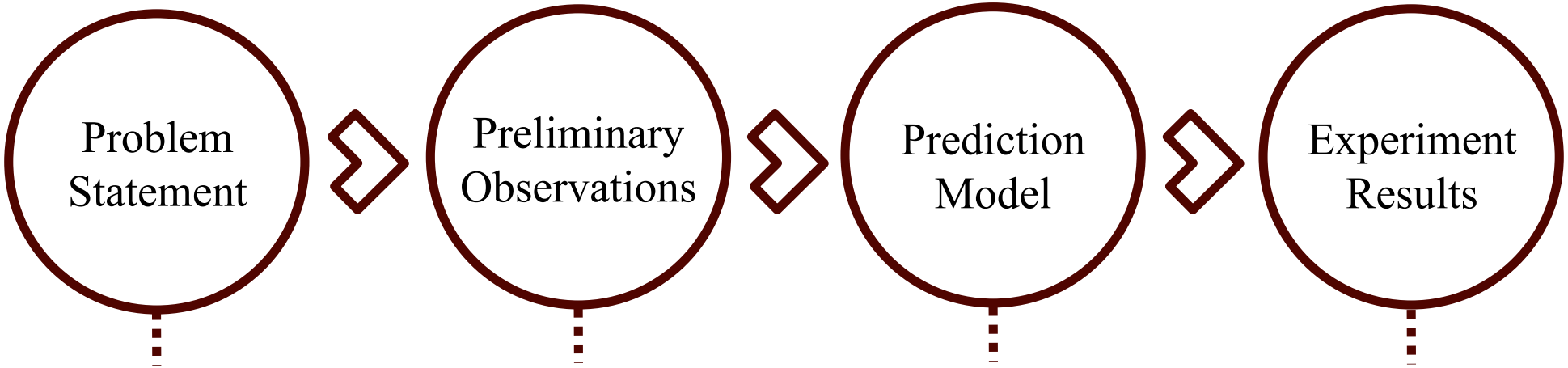
Donghwa Shin, Tae Jun Jeon, Ye Yuan[1]

Department of Computer Science & Engineering, Texas A&M University, College Station, TX, USA

Emails: {donghwa_shin, tjeon90, ye.yuan}@tamu.edu

1: TA for CSCE 633 (Machine Learning) - she does not have any contribution for the project

Overview



Dataset

Time Series

Location Based

Big Picture

Random Sampling

Neural Network

Final Prediction

Compare Baseline

Problem Statement

- Ride sharing services, *Uber* and *Lyft*, are challenging taxi business

Problem Statement

- Given the dataset containing taxi trips from 2013-2017

1) Build visualizations and predictive models explaining how the Chicago taxi business has changed over time

2) Predict the median weekly revenue for Chicago taxis

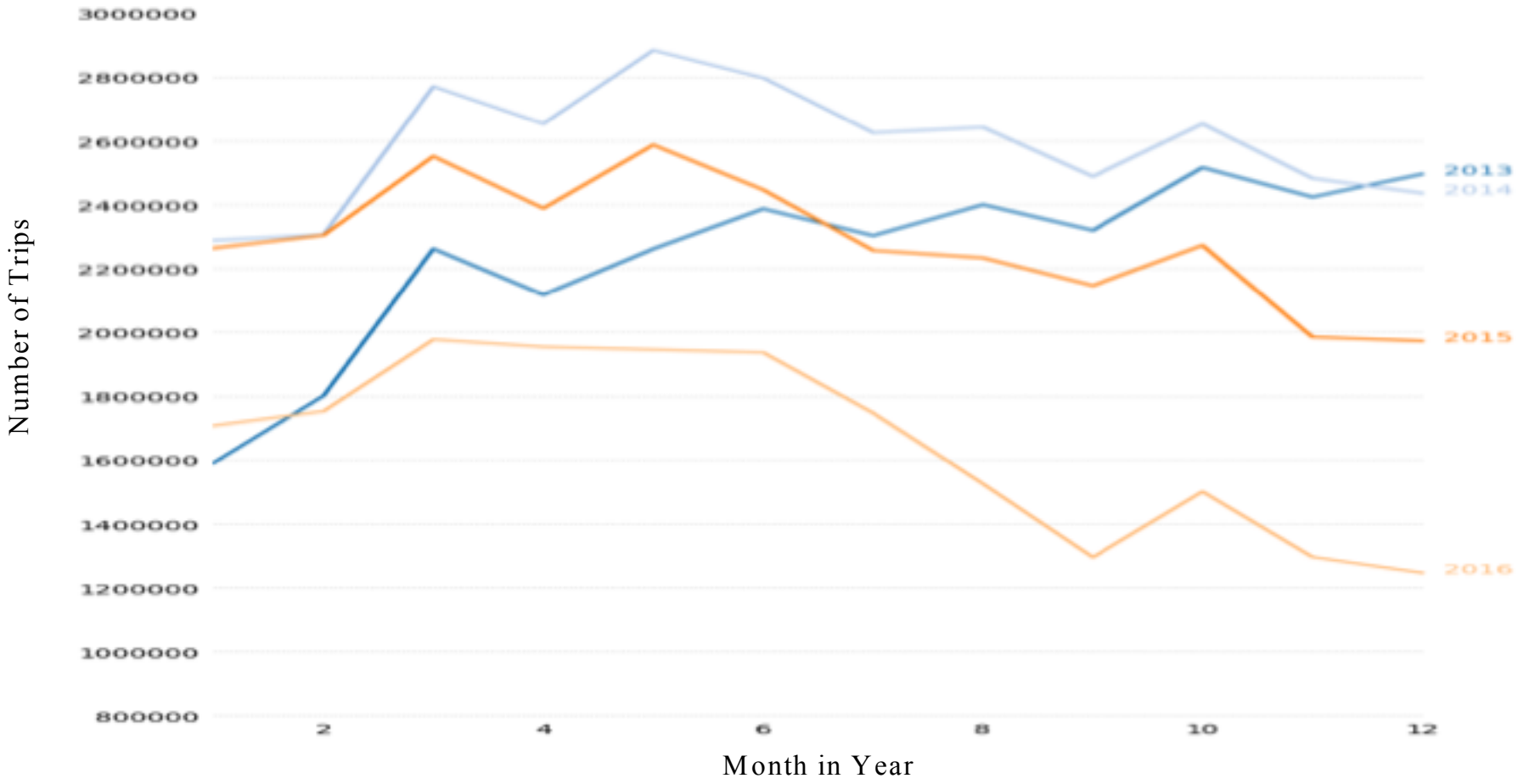
Dataset

- Dataset contains more than 110 million Chicago taxi rides from Jan. 1, 2013 to Jul. 31, 2017

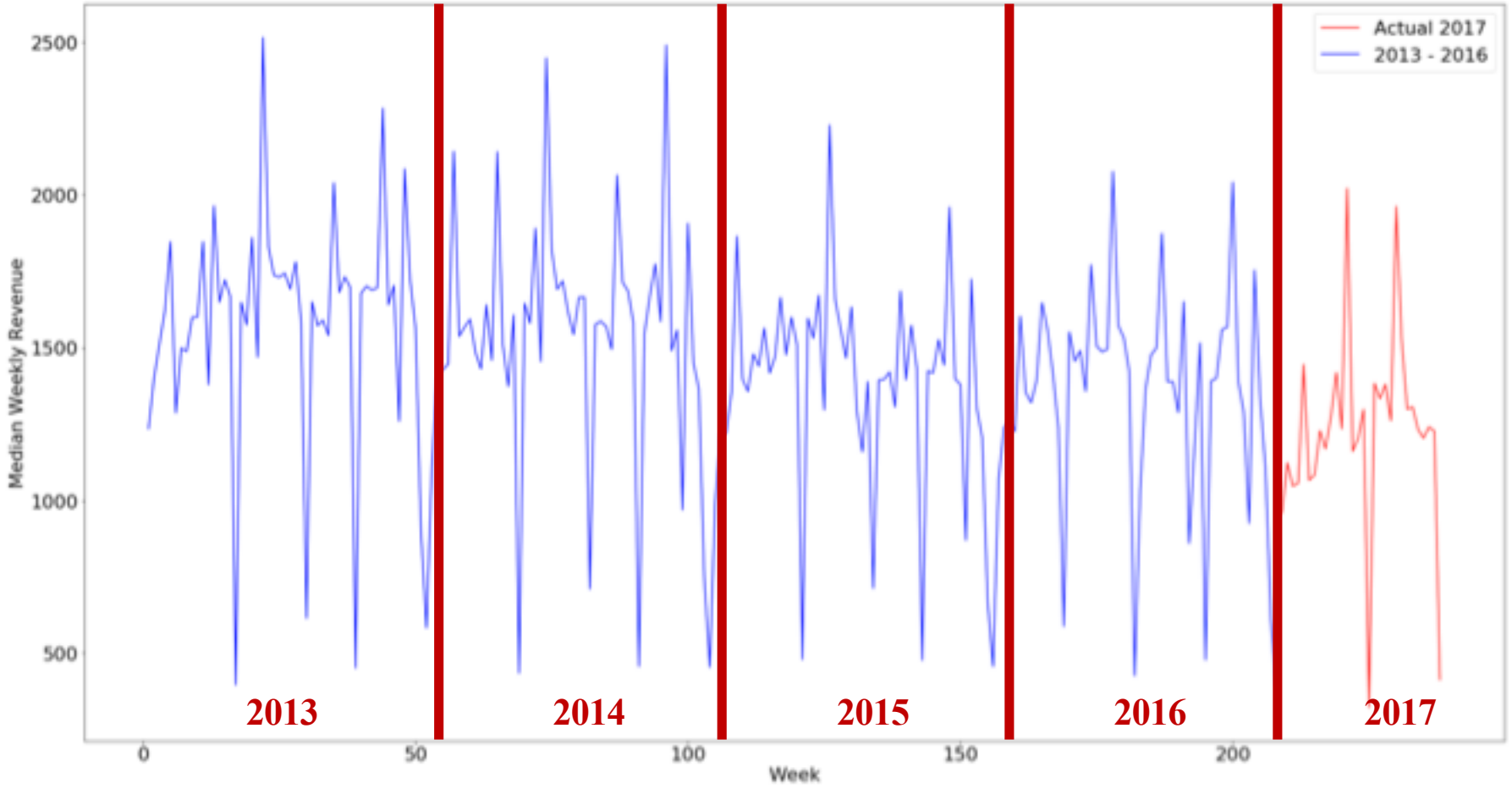
- 24 features including ‘Taxi ID’, ‘Trip Seconds’, ‘Trip miles’, ‘Pickup/Dropoff locations’, ‘Total Cost’, etc.

- 2013-2016 for training and 2017 for testing

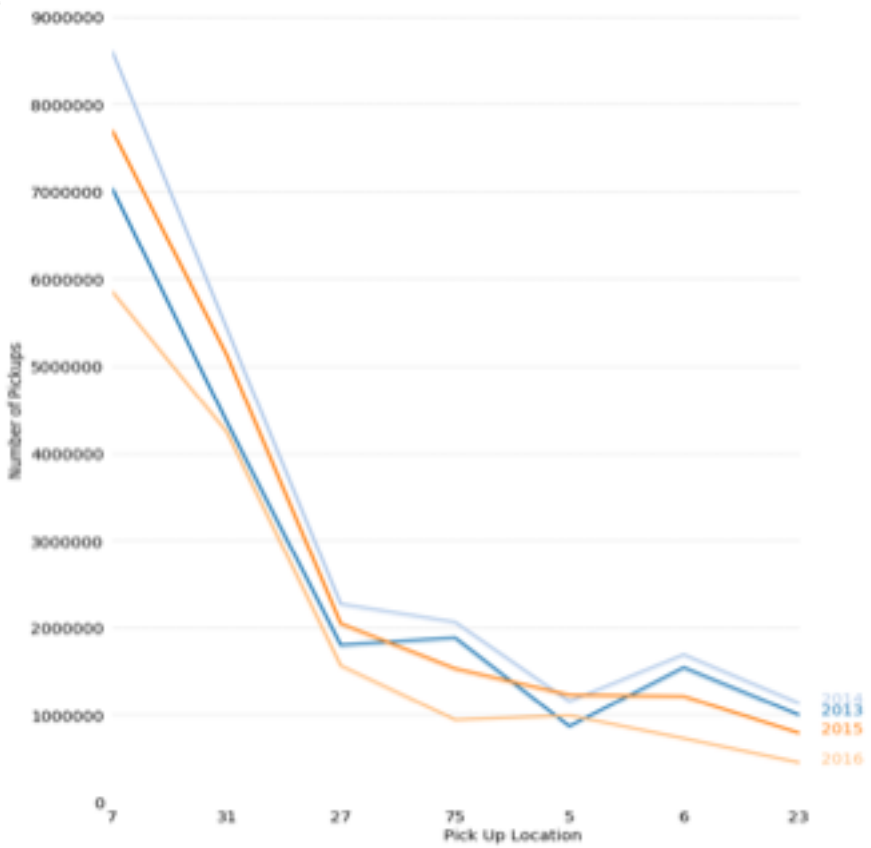
Observations

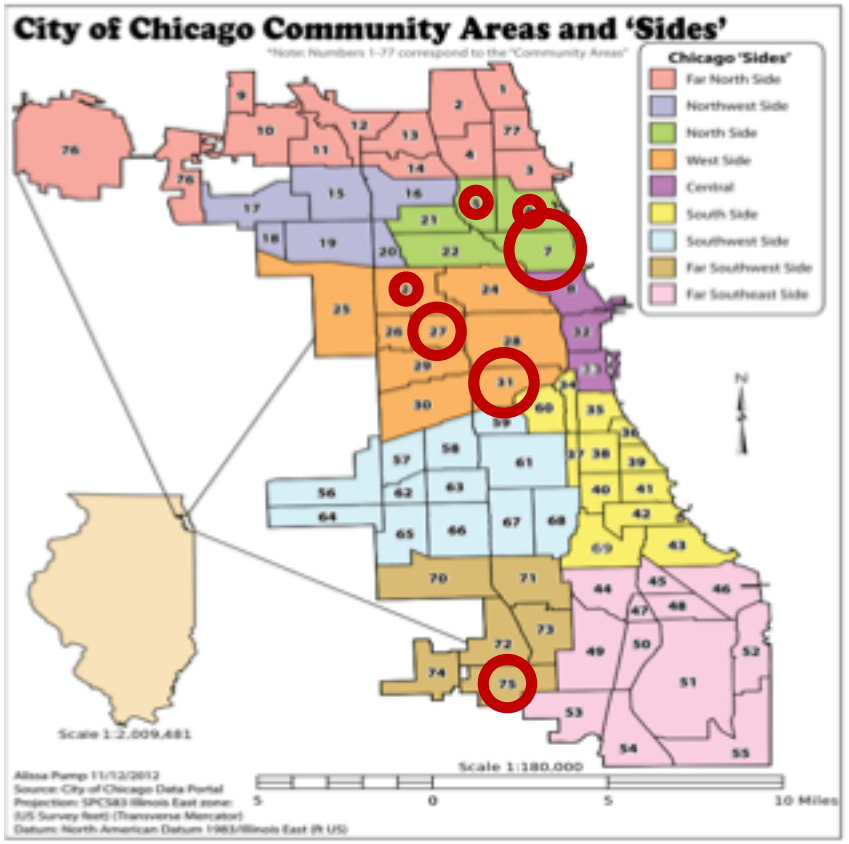


Observations

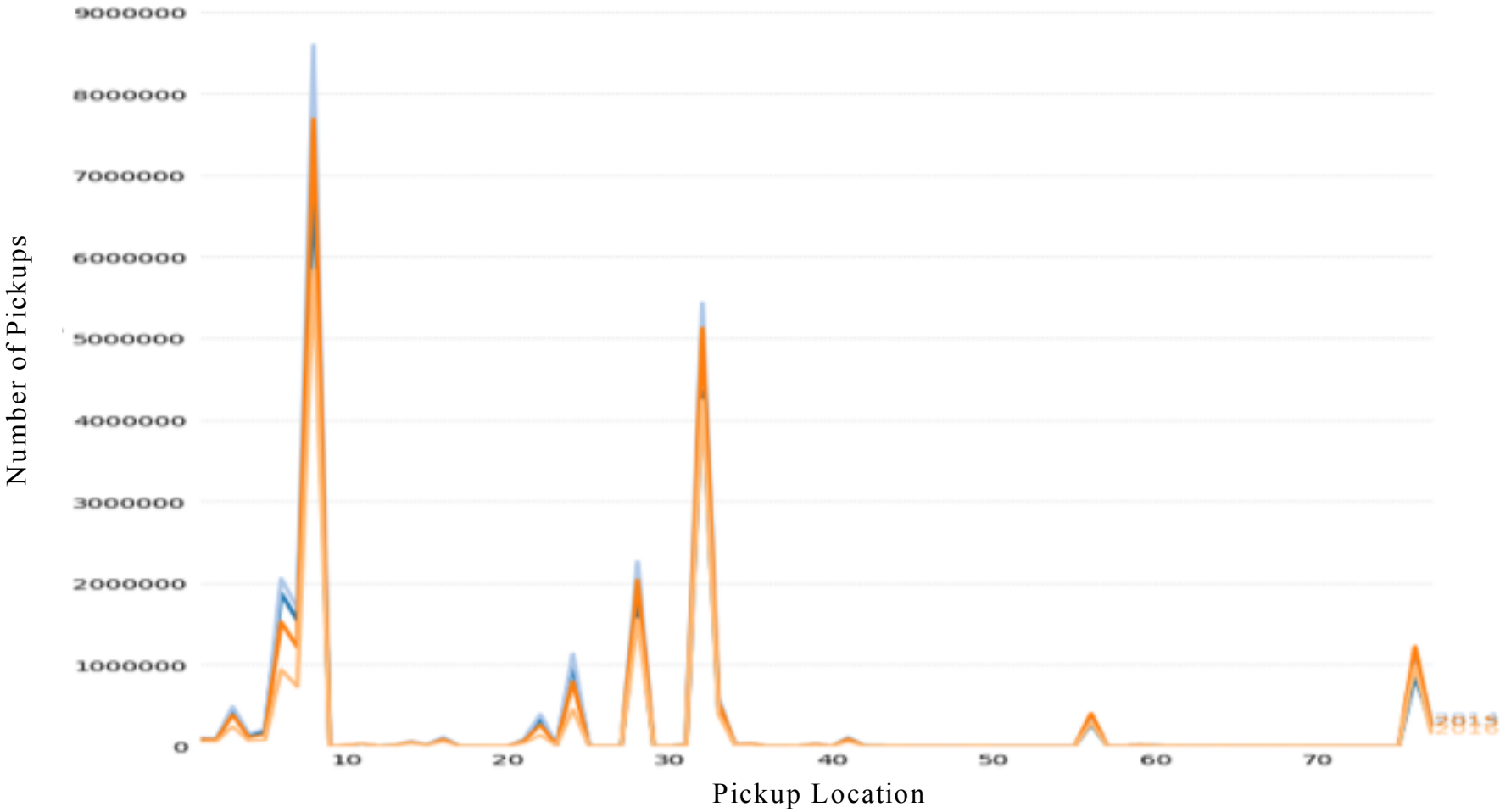


Observations

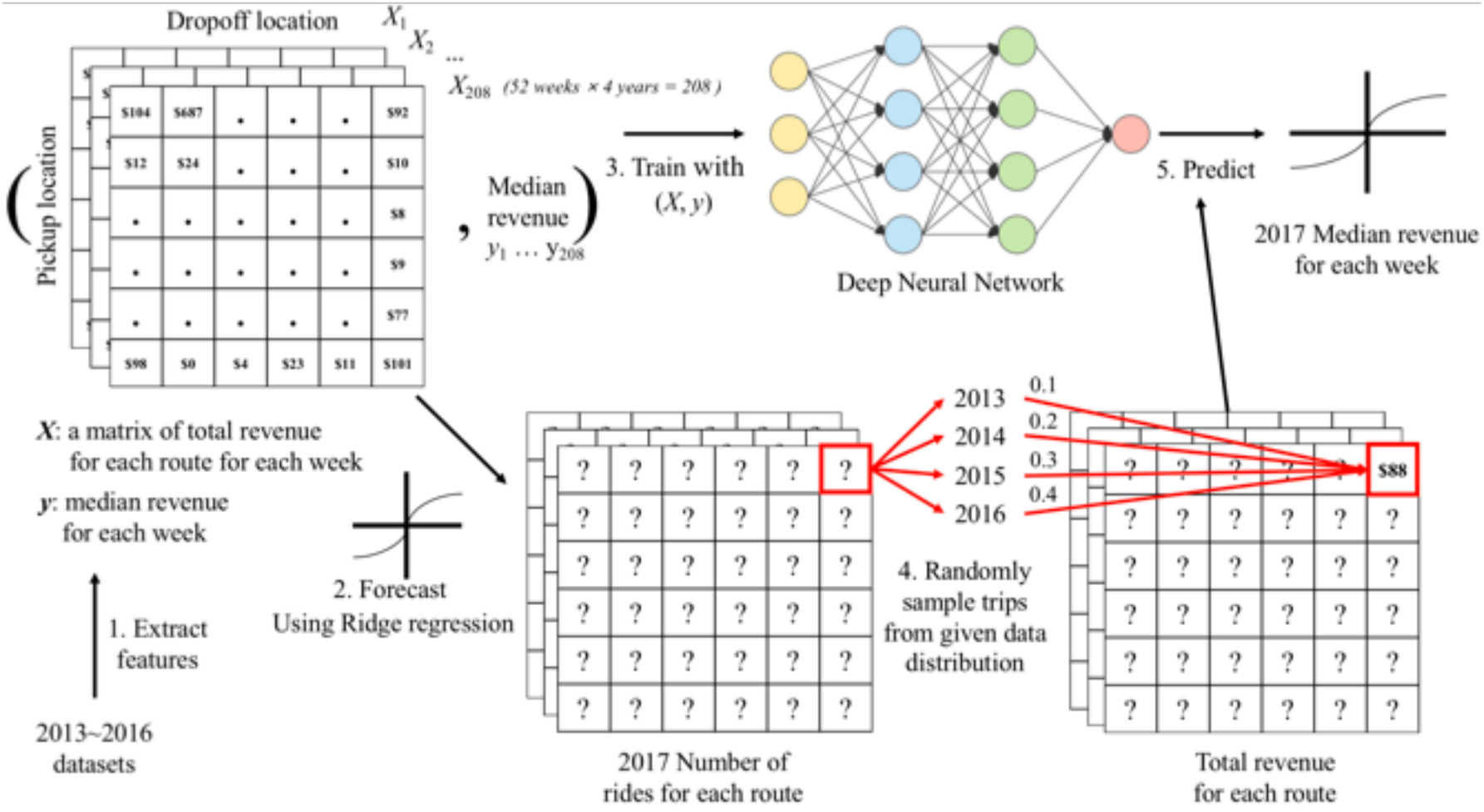




Observations

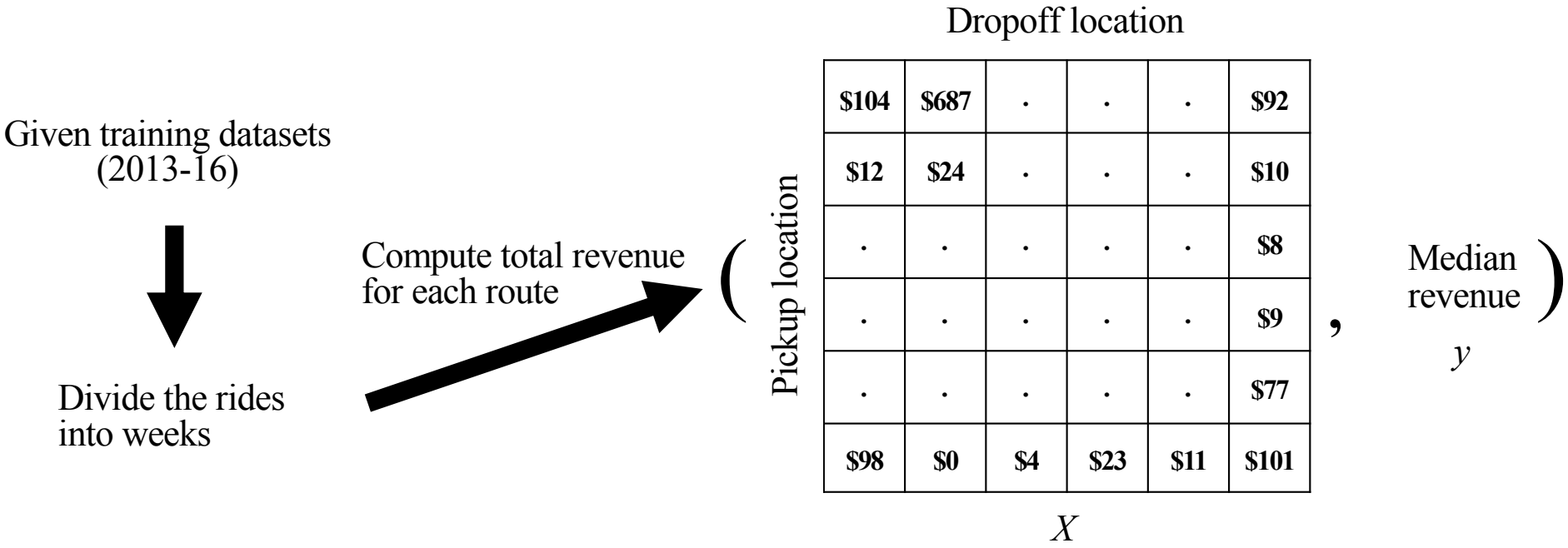


Prediction model



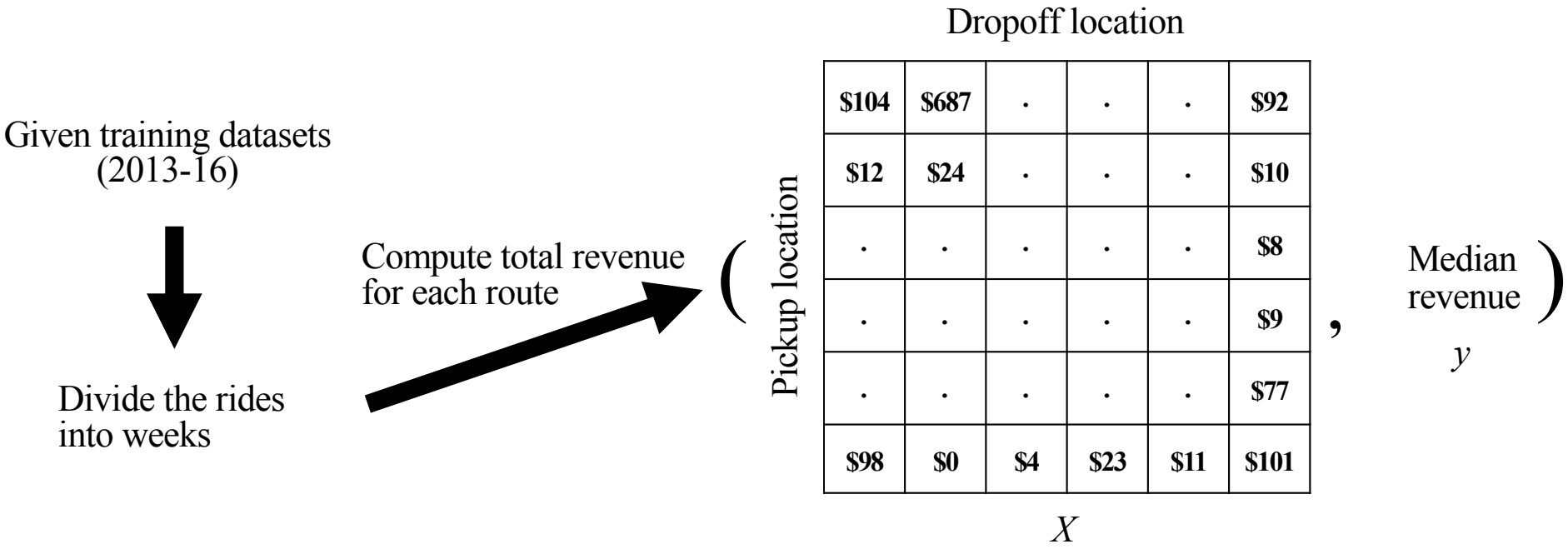
Prediction model

1. Extract features



Prediction model

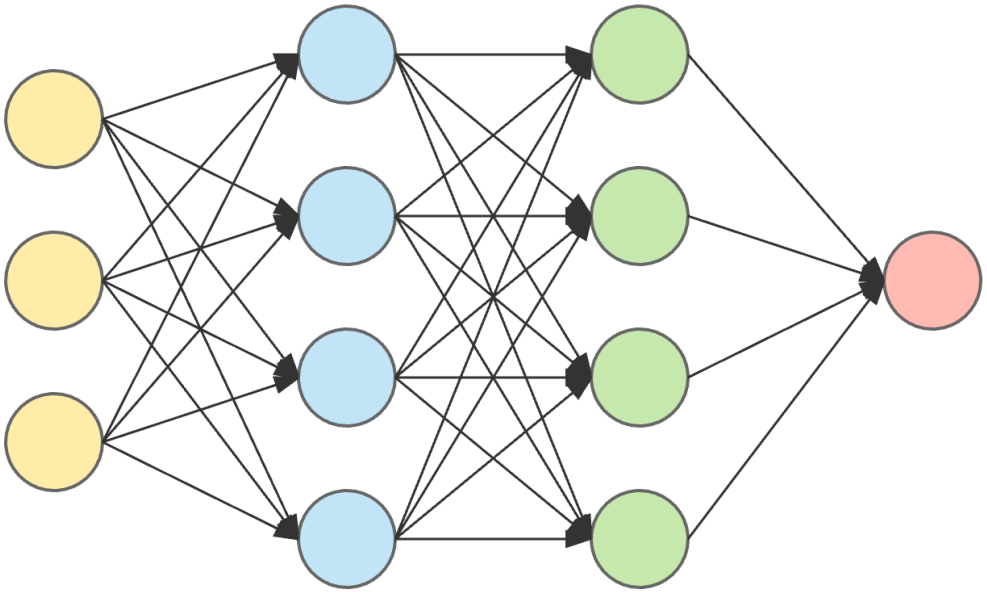
1. Extract features



52 pairs of (*X*, *y*) for each year
208 pairs in total

Prediction model

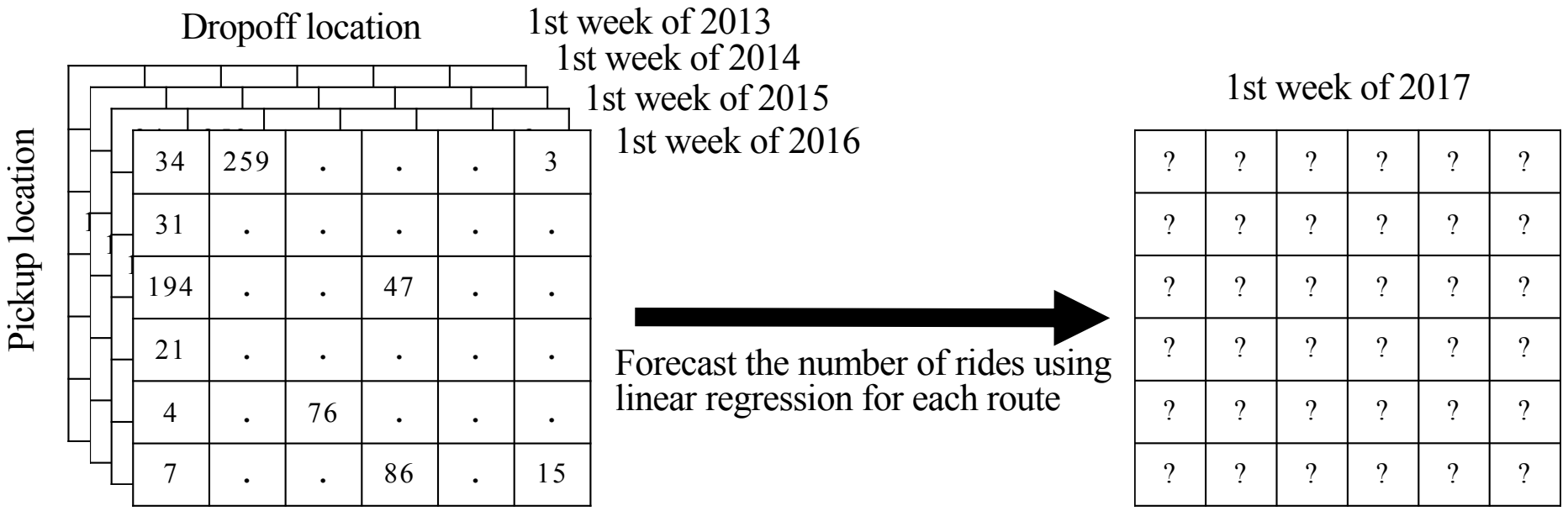
2. Train a deep neural network for regression



Train with 208 pairs of (X, y)

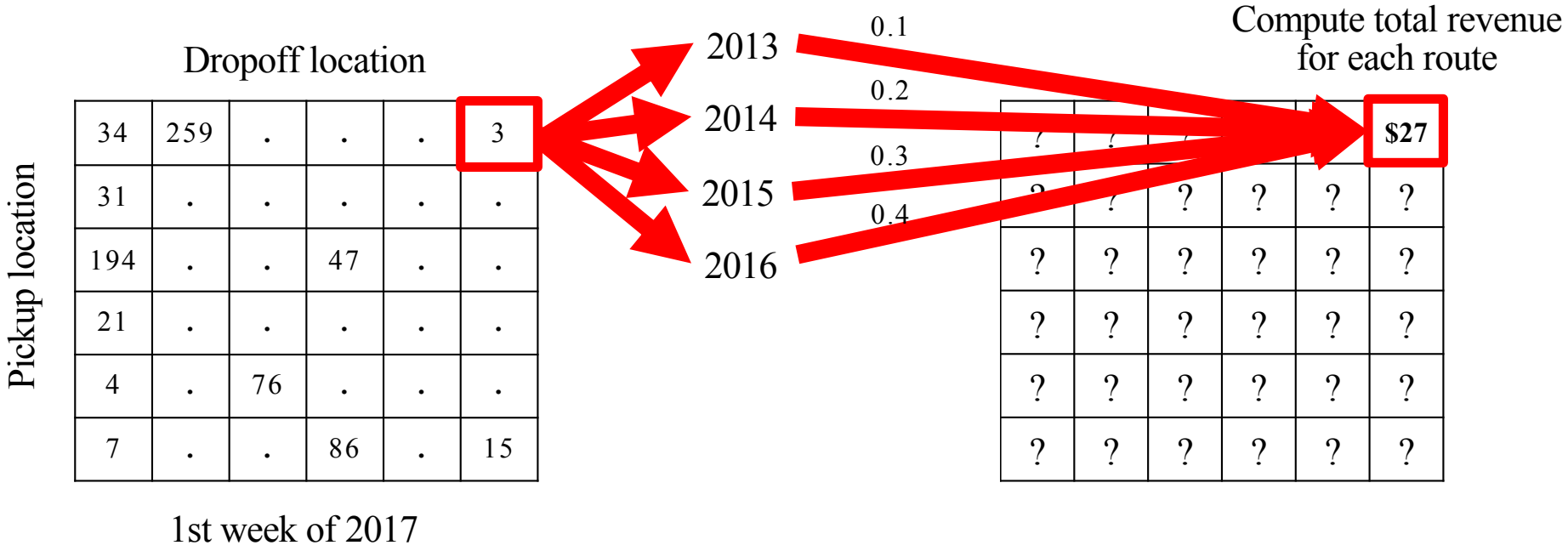
Prediction model

3. Forecast the number of rides using linear regression



Prediction model

4. Randomly sample rides from the training datasets

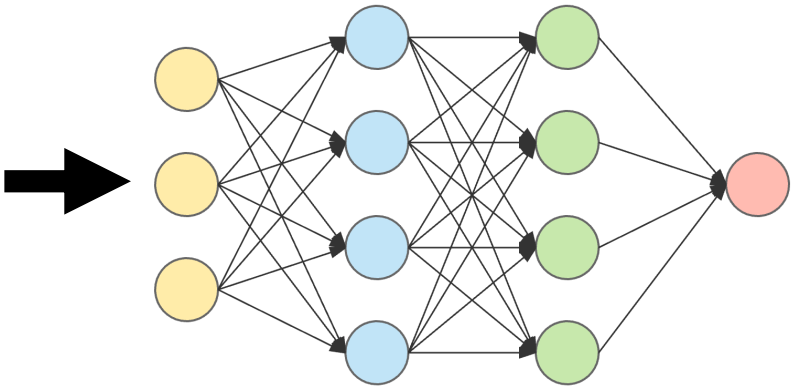


Prediction model

5. Predict median weekly revenue

$104	$687	.	.	.	$92
$12	$24	.	.	.	$10
.	.	.	.	.	$8
.	.	.	.	.	$9
.	.	.	.	.	$77
$98	$0	$4	$23	$11	$101

1st week of 2017

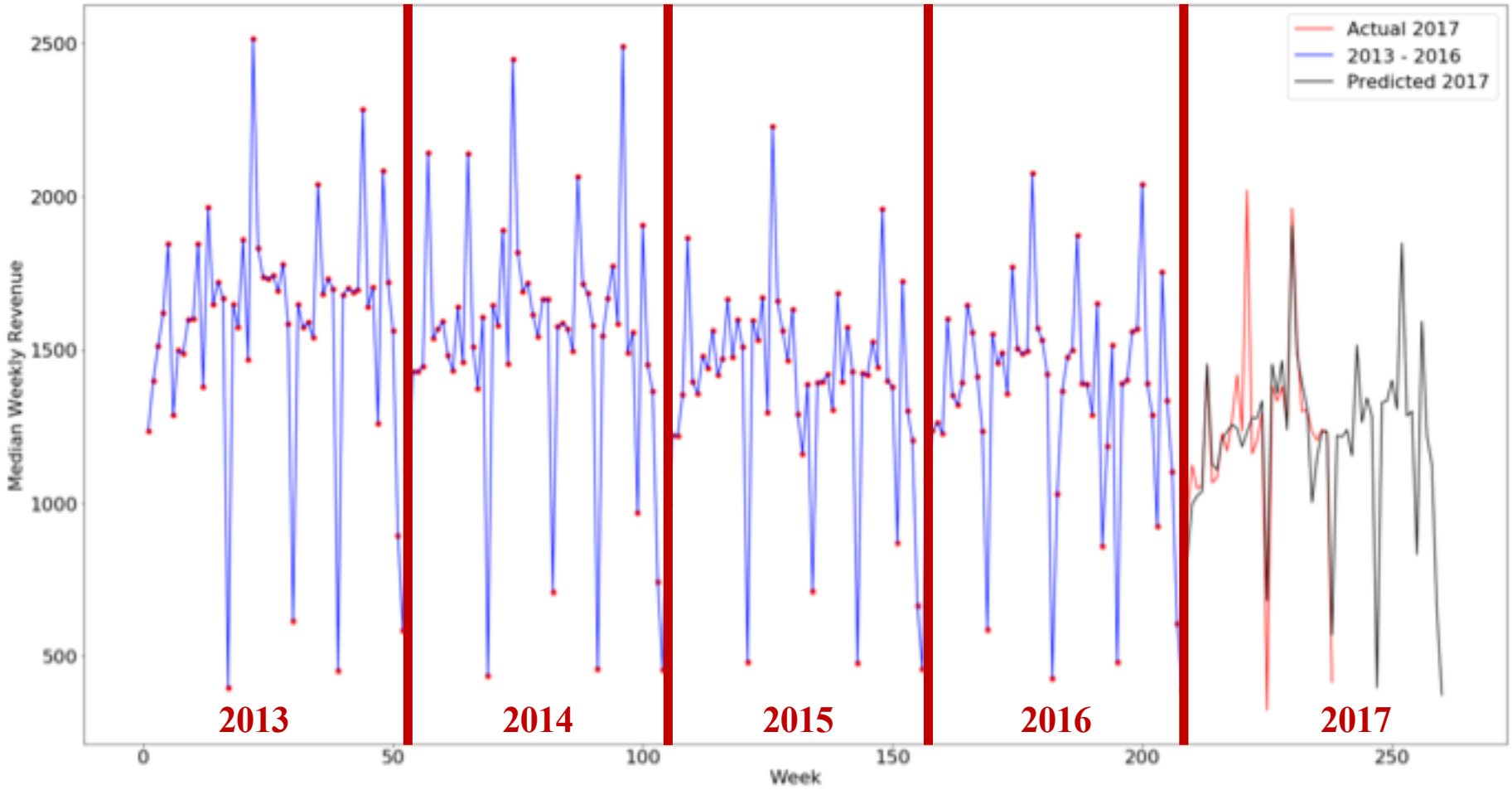


Neural network regressor

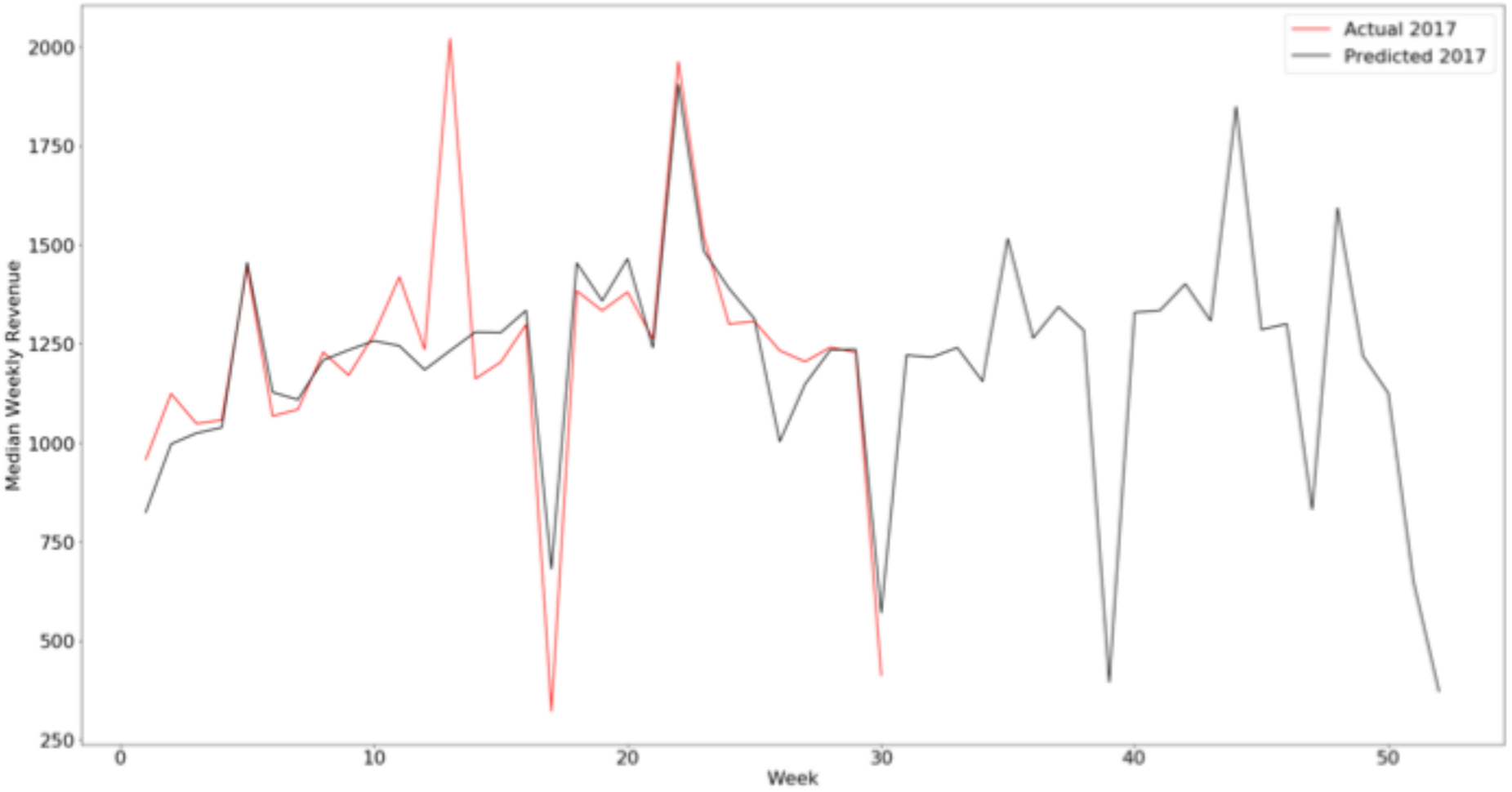
Predict

Median revenue for 1st week of 2017

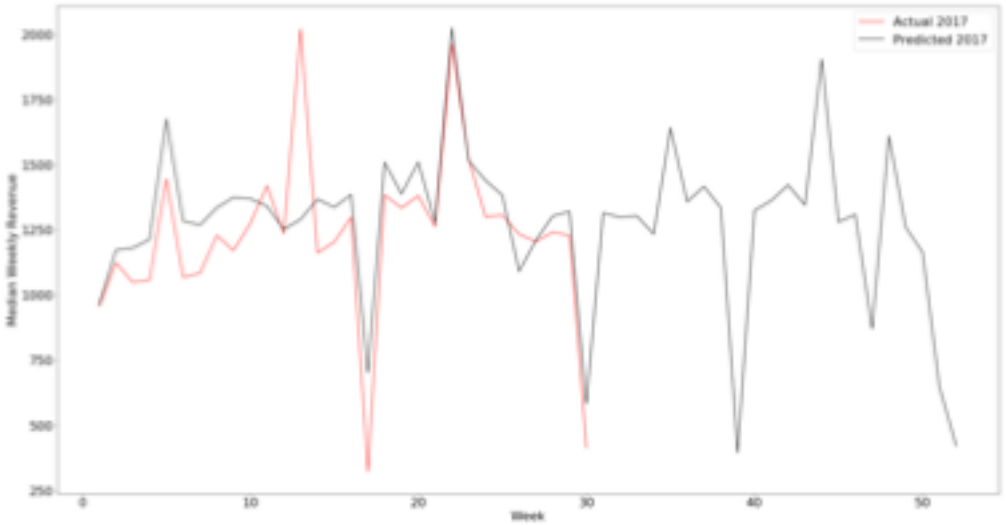
Experiment results



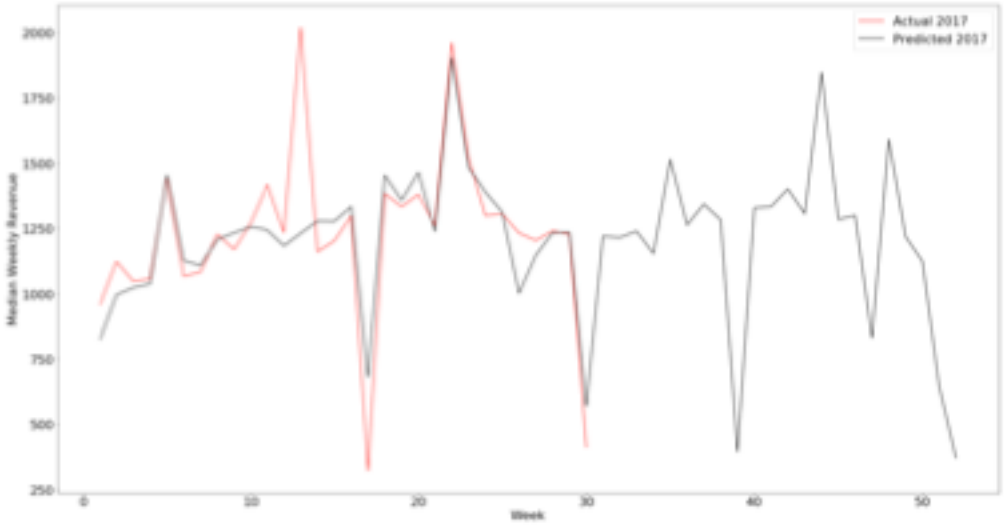
Experiment results



Experiment results



Ridge



Ours

RMSE	Lasso	Ridge	Our model
2013-2016	193.0228	193.0110	**178.4015**

Thank you!