Sharing-Bicycle System Prediction

Horace Wang

Mar.26 2019

Introduction & Problem

- Bicycle sharing service
- Customer rent and return bicycle at designed locations
- Availability at each station is not equal
- Shortage becomes a big issue for the company
- Need a method to predict the demand for each station

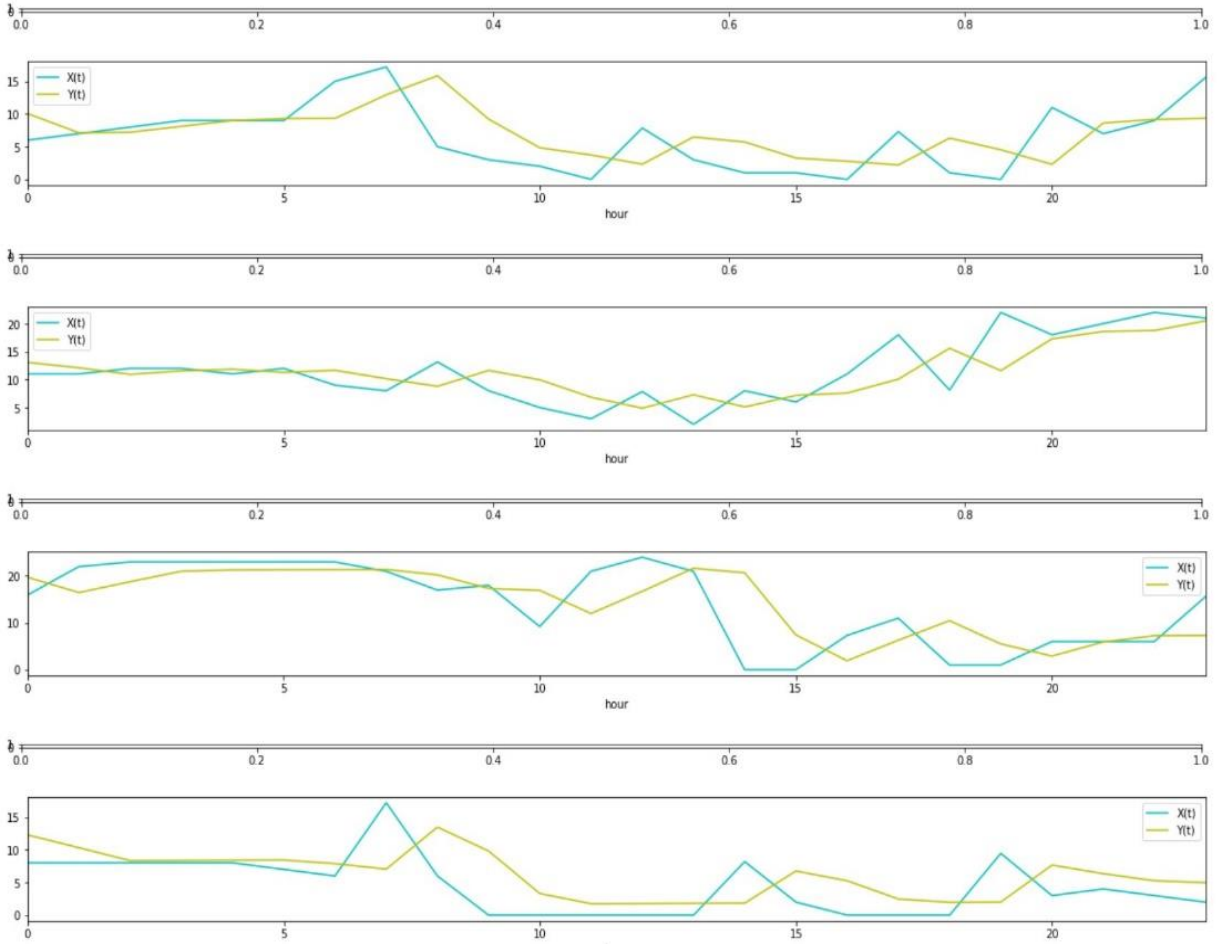
Data

- Citibike official website provides JSON file
  - https://feeds.citibikenyc.com/stations/stations.json
- Including following characteristics:
  - *executionTime*
  - *stationName*
  - *totalDocks*
  - *longitude*
  - *statusKey*
  - *stAddress1*
  - *city*
  - *location*
  - *testStation*
  - *landmark*
  - *stationBeanList*
  - *availableDocks*
  - *latitude*
  - *statusValue*
  - *availableBikes*
  - *stAddress2*
  - *postalCode*
  - *altitude*
  - *lastCommunicationTime*

Methodology

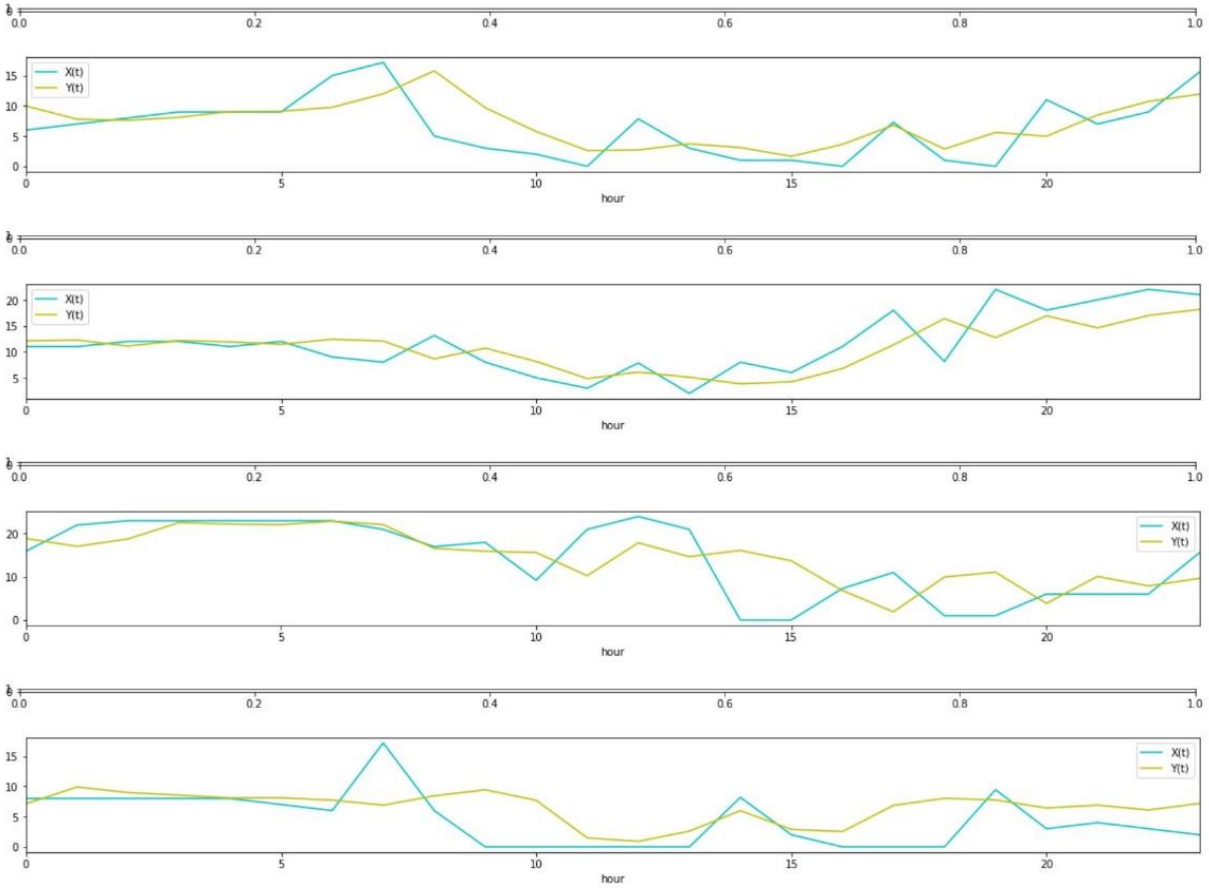
- Different Analyzing Method:
  - Linear Regression
  - SVR
  - Decision Tree
  - Natural Network
- Evaluation Standards:
  - R-square
  - Means Squared Error

Discussion (1/5)



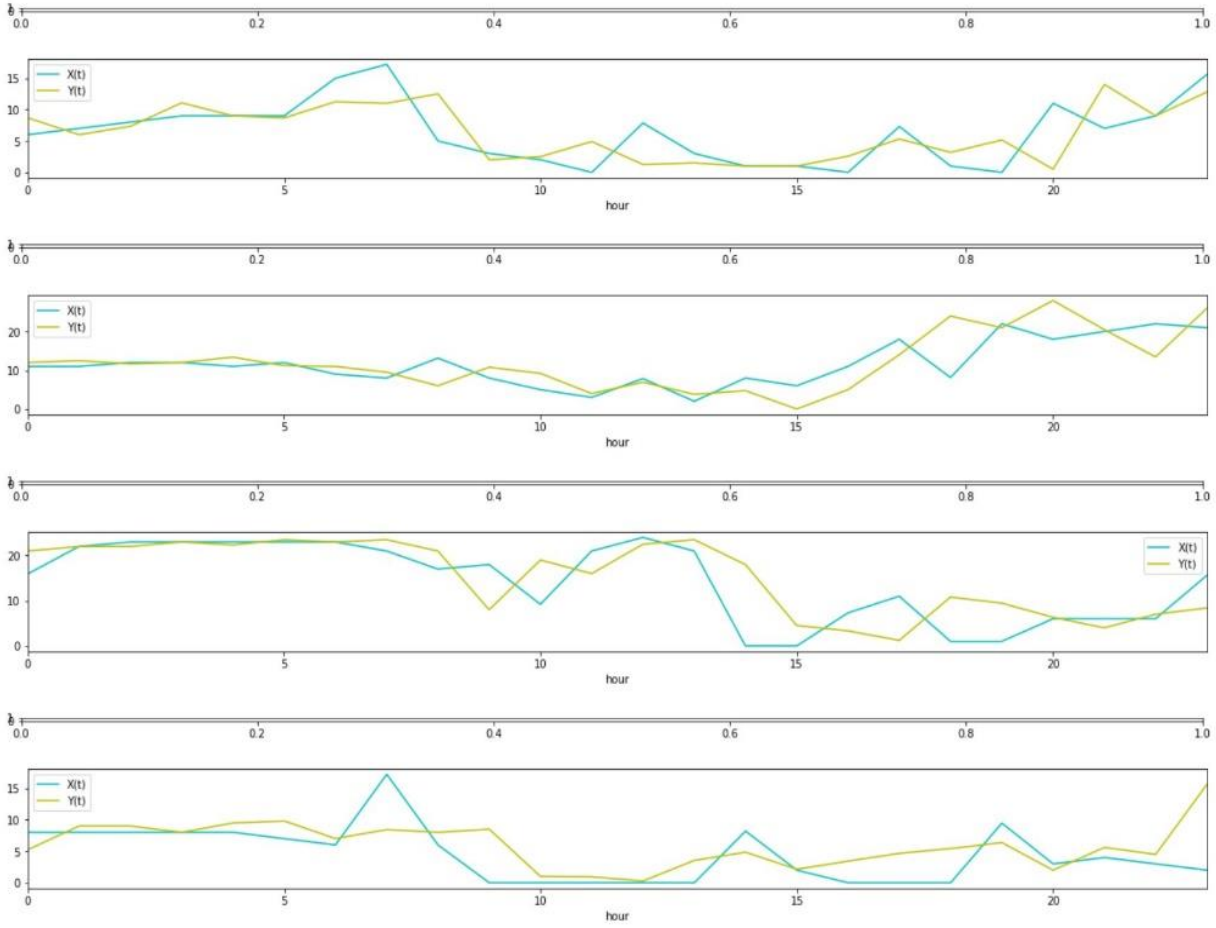
- Linear Regression

Discussion (2/5)



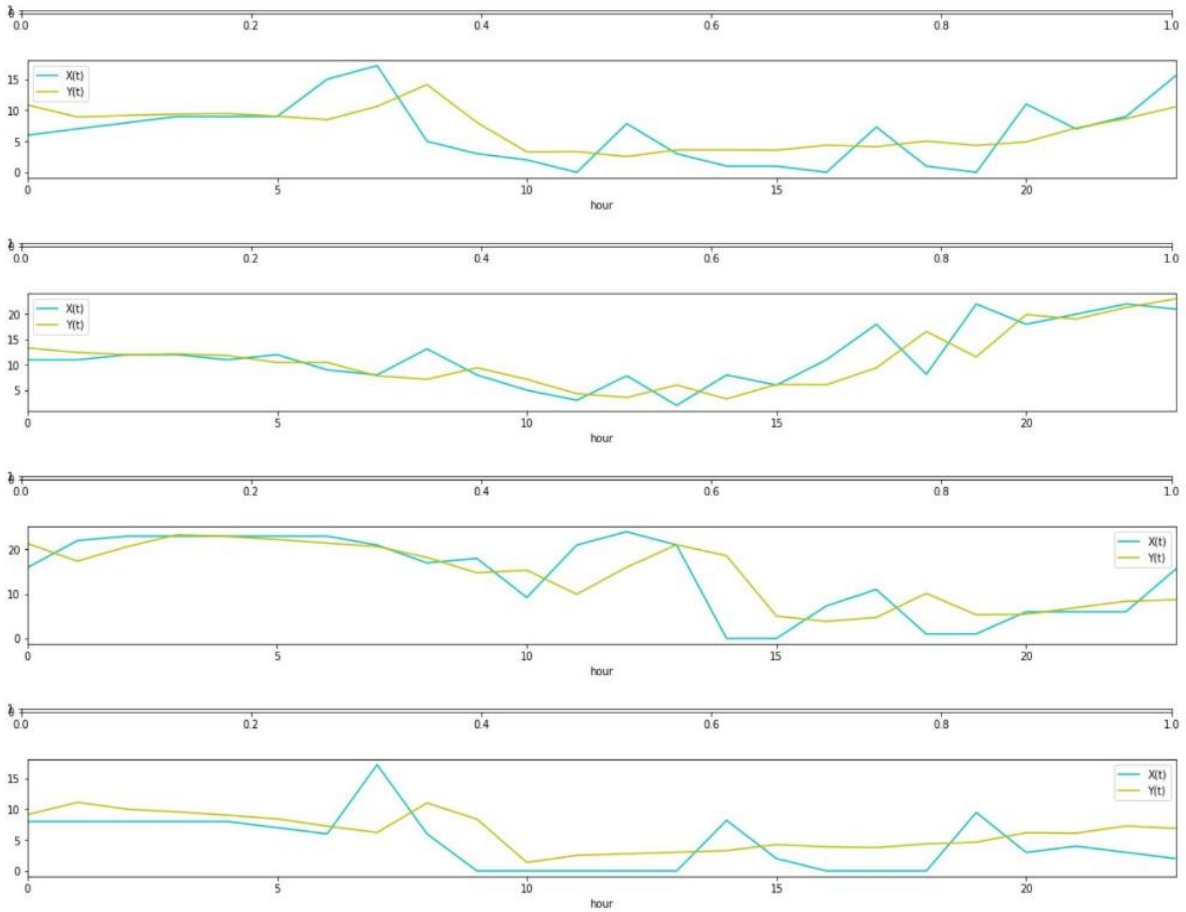
- SVR

Discussion (3/5)



- Decision Tree

Discussion (4/5)



- Natural Network

Discussion (5/5)

	Linear Regression	SVR	Decision Tree	Neural Network
R-square	0.547	0.556	0.485	0.622
MSR	18.453	18.063	20.960	15.387

- Neural Network has the highest R-square value and lowest Means Square Error. Thus, Neural Network is the best estimation model for Citibike's business for each station.

Conclusion

- Neural Network is the best model to estimate the demand of each station currently
- Future Improvement:
  - More data needed to have more accurate result:
    - Customer’s behavior
    - The relation between different stations
    - The differences between each week, seasons or years need to be considered

Thanks