Abstract

1. Introduction
   1. Introduce the issue and its importance
   2. Introduce the other methods to optimize the public transits (literature review)
   3. Introduce machine learning to do the optimization
2. Data

2.1. The composition of various data (mov\_gps, mov\_sensing, trip\_detail)

2.2. How the data is collected

2.3. How the data is organize into different frame

3. Methodology

3.1. How to use kmeans clustering to separate noises and passengers

3.2. Explain Sil score

3.3 Self….

3.3. Explain the model that I use to predict the number of passenger each stops( or live)

1. Result

4.1. validate the result from clustering (separation between noise and passenger)

4.2. validate the result from prediction

1. Conclusion & discussion
   1. Interpretation
   2. Application?
   3. Limitation
   4. Possible future directions?