代价函数:

数据源：6个月的北京一卡通ic卡记录

数据预处理

（1）数据清洗：去除重复值，异常值，极端值

（2）数据填充：空值填充

（3）数据格式转换：

（4）特征抽取及缩放：由于不同特征之间采取不同的量纲，导致可能不同的特征数据之间量级相差很大，数值较小的数据被数值大的数据淹没，从而算法性能不佳；特征抽取及缩放是通过算法将不同特征的数据映射到[0,1]之间，或使其满足方差为1，均值为0的标准正态分布，从而提高算法性能；

模型设计：（1）基于kmeans的人物出行模式的聚类分析

1）k的确认：轮廓系数法

（2）基于时空分析的乘客个人偏好的分析

1）爬虫+时空分析

（3）乘客居住地与工作地的分析

（4）