# 資料分析

## 原始資料集

### 資料清洗

#### 僅留下需要的資料(公車、捷運)

本研究所取得的資料集包含了目前於MeNGo系統中的運輸業者….

#### 補齊捷運票證資料之旅客上車時間

本研究所使用的MeNGo使用者票證資料為高屏澎運輸研究發展中心提供，在初步資料處理階段，發現該資料集在欄位上有缺漏及紀錄資訊不完整的情形，主要缺漏在高雄捷運之刷卡資料上僅記錄該使用者當次刷卡之下車時間，其資料狀態紀錄為”定期票下車”(這裡應該可以補圖去說明他是怎樣缺漏的。)，因此為取得捷運使用者之上車時間，本研究透過TDX平台上所取得之’‘高雄捷運站間行駛時間’’，作為推估捷運使用者上車時間之基礎，用以補足捷運使用者在資料欄位中所缺少的上車時間資料。

具體流程….

#### 新增使用者特徵描述欄位(旅程時間、星期幾上車、上車時間(小時))

在完成資料補足的部分後，為在後續分析的方便進行，本研究於資料欄位中新增數個資料欄位，包括：旅程時間、星期幾上車、上車時間(小時)，此類使用者票證資料的旅行行為模式描述，

## 特徵資料集

## 資料降維

## 集群分析