

--

資料集名稱	建立日期
divvy bike	2023/8/10

日期	描述
2023/8/13	增加欄: ride distance
2023/8/13	增加欄: ride_length
2023/8/13	增加欄: day_of_week
2023/8/14	增加欄: work_day
2023/8/13	增加欄: is_holiday
2023/8/14	增加欄: speed
2023/8/14	增加欄: rush_hour
步驟	日期
1	2023/8/10
2	2023/8/10

3	2023/8/10
4	2023/8/11-12

日期	策略
2023/8/11-12	用當前匹配值填補缺失值

日期	轉換
2023/8/14	將ride_length轉換為四類： 極短程、短程、中程、長程

更改日誌

資料簡述

上次更新

2023/8/14

更動

細節

根據GPS計算起點站和終點站之間的距離

計算騎車時長

星期幾

確定是不是上班日

1: 上班日 0: 假日

確定是不是假日

1: 假日 0: 上班日

計算騎車速率

確定是不是上班尖峰時段

1: 尖峰時段 0: 非尖峰時段

尖峰時段: 上班日7-9點 或 16-18點

資料清洗步

描述

整合過去12個月資料

基本基料驗證

進階資料驗證

處理缺失值

缺失值處理

細節

而缺失的緯度和經度，我找尋匹配的站點名稱填補缺失值
對於缺失的站點名稱，我找尋相同GPS位置的站點填補缺失值

資料轉換

須轉換原因

分配給 4
個類別之後會更容易分析用戶行為

志

資料來源
該數據由芝加哥交通部 (CDOT) 提供
https://divvy-tripdata.s3.amazonaws.com/index.html

更動原因
為未來分析做準備
為未來分析做準備
為未來分析做準備
為未來分析做準備
為未來分析做準備
為未來分析做準備
為未來分析做準備

驟
細節
檢查各欄位型態

<p>檢查是否有兩個站具有相同的 id 但不同的名稱， 如果它們確實是不同的站(距離 > 0.5公里)， 則更改 id</p>
<p>請見缺失值處理</p>

里
備註

備註