



餐廳客人數預測模型

報告人: 賴呈恩

訓練資料檔案說明 (皆為2016, 2017)



air_visit_data.csv

餐廳實際到場客人數

air_reserve.csv

Air 平台的預約資訊

hpg_reserve.csv

HPG 平台的預約資訊

air_store_info.csv

air平台上餐廳的資訊(類別、地區、經緯度)

hpg_store_info.csv

hpg平台上餐廳的資訊(類別、地區、經緯度)

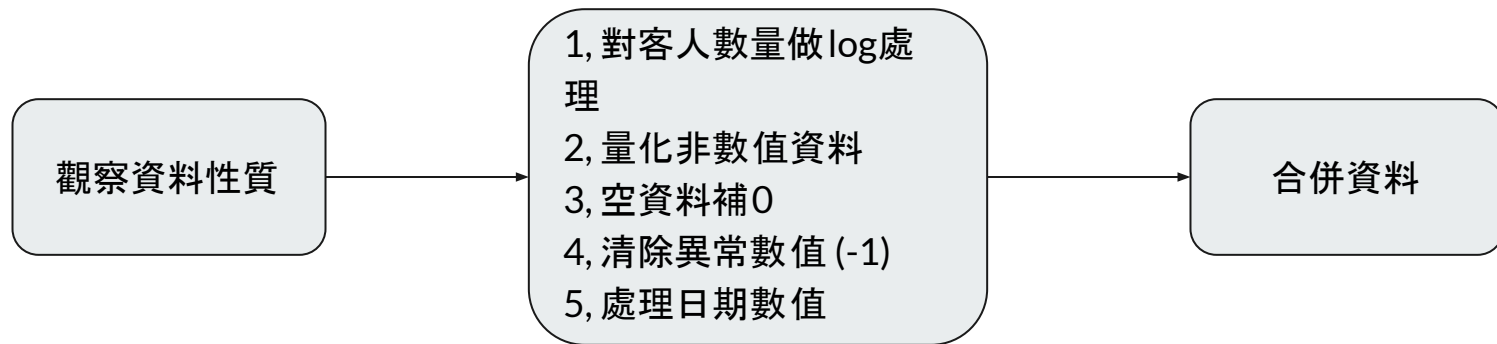
store_id_relation.csv

air 和 hpg 平台上的餐廳對應ID

date_info.csv

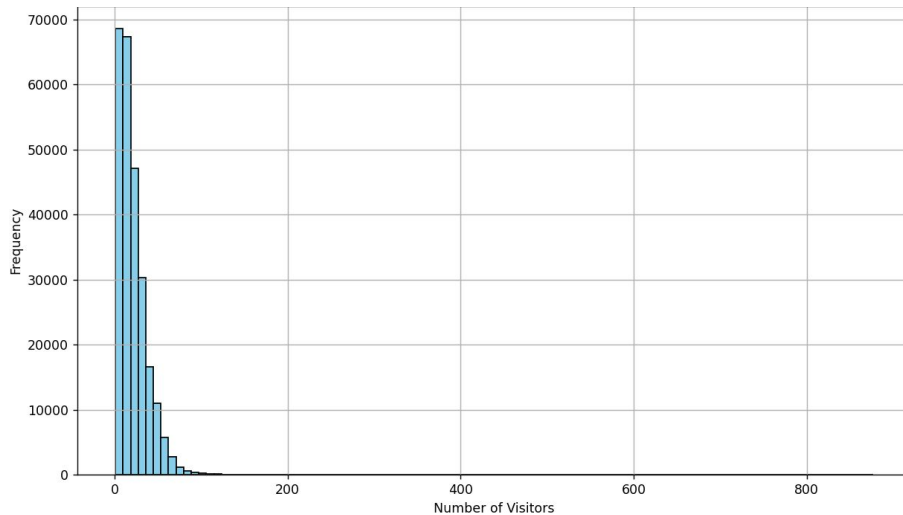
每天是否為假日、周幾等資訊

訓練資料處理流程



為何做log處理而不是標準化

項目	$\log(1+x)$	標準化
對極端值效果	有效壓縮極端值, 減少其影響	對極端值仍較敏感
數據分佈改變效果	使其更接近常態或對稱分佈	僅調整尺度, 不改變分佈形狀



訓練結果

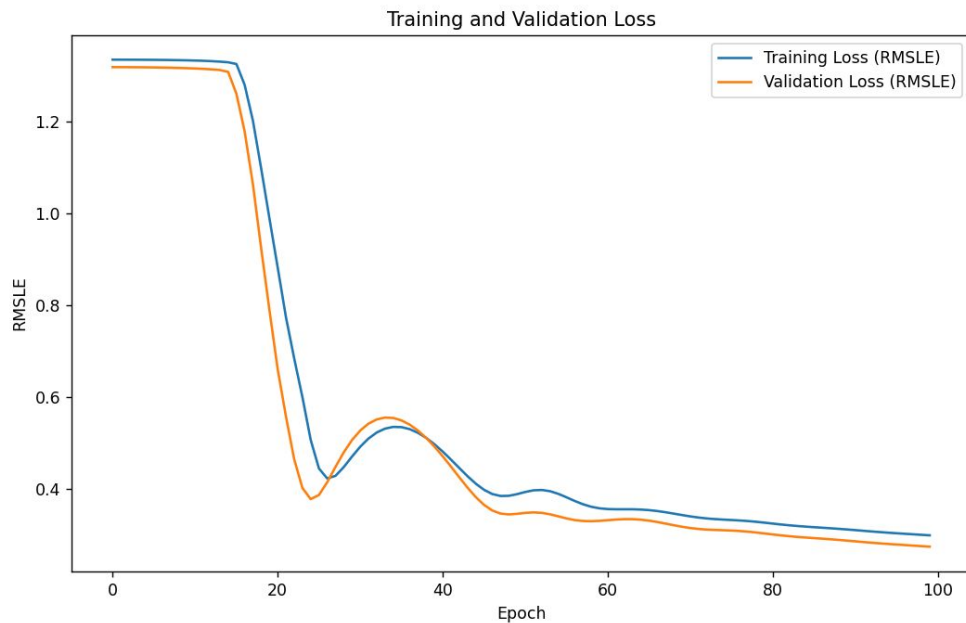
optimizer : adam

learning rate : 0.0001

train data size : 約140000筆

test data size : 約80000筆

full batch



整批資料平均化更新

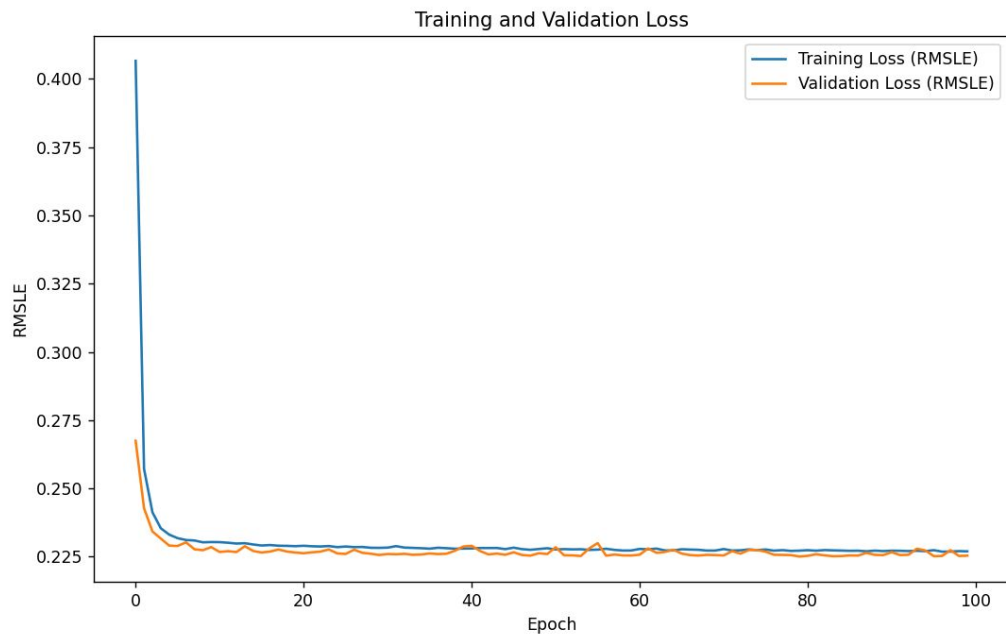
loss (rmsle)

訓練集:0.2994

驗證集:0.2747

測試集:0.3326

1024 batch



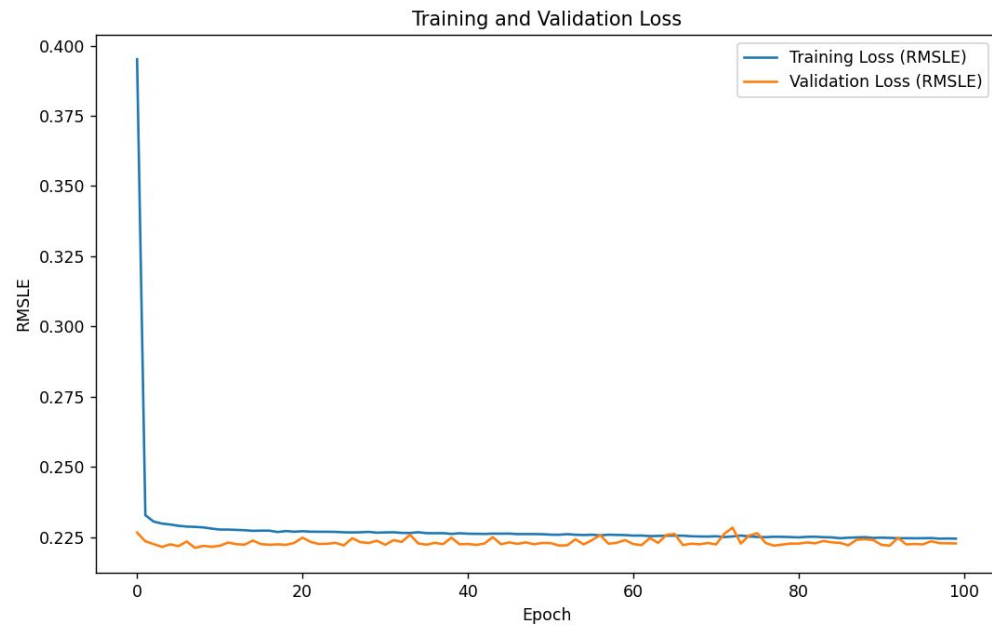
loss (rmsle)

訓練集:0.2273

驗證集:0.2254

測試集:0.2237

256 batch



loss (rmsle)

訓練集:0.2241

驗證集:0.2227

測試集:0.2204

還原各rmsle誤差值，並比較特定人數下的誤差百分比

