3.16

互动地图：<https://postal-codes.cybo.com/australia/sydney/>

EDA显著性不明显、pattern一无所获，因为数据量太大了，单一维度划分后还是会包括很多很多很多个subclasses，所以聚合起来看，成单率基本都是在0.6附近的，所以地图用作EDA可能不是一个report里很好的选择

地址信息处理：

用户侧暂无收获，聚类在地址、打车类型中或许有新发现，因为显著性较差，需要通过进一步分区分时提取有效规律

车程纬度：费用越高，车程距离越大，存在右偏取对数后为几何距离分段or Kmeans重新聚类变成一个分类变量，分短距离、中距离、长距离

时间维度：时间分箱后存在时间差异性，分时点（每小时）+分州 看成单率

热门地点独立分析：机场、农场、沙滩、大堡礁等热门地点，但容易有遗漏以及效率低

处理postcode、经纬度和地址：

五边形

px

Geopandas

sns.kde慎用hue参数，用了hue会让所有分类的kde下面积加总为1，如果两类数据总量差异很大的话结果非常不理想

—————————————————————————————————————————

Levia 3/17

目前字典存在将包含了cora的字段全部归类为“coral sea”珊瑚海的情况。而在实际地理情况下，应该不存在coral sea这个州。这可能是导致无法找到pattern的原因，目前正在修改。

PS：code已经上传，建议在excel里打开，看看有没有离谱的code。下面是参考

