working progress update

1.  1.2 miilion 数据需要清洗， state name 不规范（大小写）， mileage outlier  很多，需要检查一下其他列是否有异常。 1.2 million 数据读取的时候编码有问题，我把有state 改动的传到drive 上了（见car1.csv）， 徐艺萌和jackie 可以用这个数据。

2.  我们昨天商量想要把新车数据的车辆features match 到1.2 million 的数据上，需要手动清理统一一下mode 名称。 建议选量最多的前20款车（已有结果），或者不同品牌同级车（need discuss）。 这部分Jackie 的Q3 会用到，可以开始做。

3.  目前结果中和后续问题相关的结果:

 compare  trade volume  in all states —>  结果与新车销量在各区销量一致

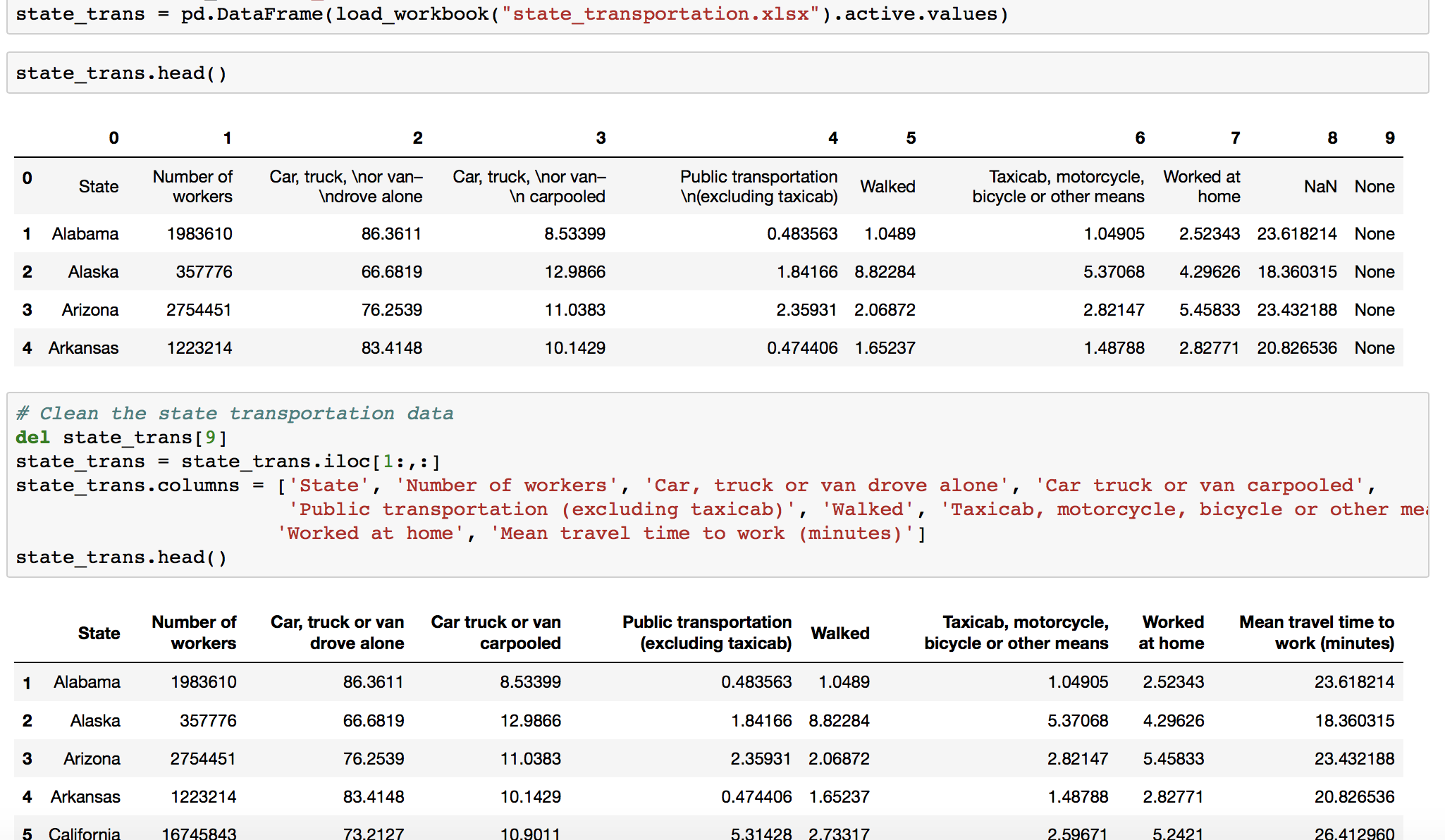
                               compare  trade volume in all makes

                               compare trade  volume in all make\_model  —> choose top 20 for feature analysis?

                              compare the price of typical models(median) in different state

4. 用Tableau 做辅助分析提高效率

Yimeng:



清理各州不同类型交通工具的使用状况