1

1.1 层次聚类 指标：’Balance','Qual_miles','Bonus_miles','Bonus_trans', 'Flight_miles_12mo','Flight_trans_12'

1.2 21 行：按 class 分类汇总每一类下 balance 列的长度，即每个 class 下 balance 记录条数; 22 行：按 class 分类汇总每一类下其它数据的均值

1.3 将分类减少到两类，原分类为 1 的一类，其他（2-20）为一类

1.4 分类减少到两类后，26：按 class 分类汇总每一类下 balance 列的长度，即每个 class 下 balance 记录条数; 27 行：按 class 分类汇总每一类下其它数据的均值

1.5 两大分类的 Balance VS Qual_miles、Bonus_trans VS Bonus_miles、Flight_trans_12 VS Flight_miles_12mo 散点图。信息：第一类乘客的这几个指标均较低

1.6 有无奖励的乘客柱状图，以及有无奖励乘客中 1 类 2 类的分布

1.7 ward 指标结果看起来更好，从三个散点图来看 1 类更加靠近（组内方差更小），与 2 类混杂更少（组间方差更大）

1.8

2 代码在 Cosmetics.R

2.1 Brushes, Concealer, Foundation

2.2 Eye.shadow, Concealer

2.3 如规则{Mascara=yes} => {Concealer=yes}，支持度 0.204 即同时购买这两件化妆品的记录占全部记录的 20.4%，信心水平 0.5714286 即所有购买了 Mascara 的记录中 57.1%也购买了 Concealer

2.4 按提升水平，{Nail.Polish=yes} => {Brushes=yes}和{Mascara=yes} => {Eye.shadow=yes}最有用

3

3.1 分类树法。结果每次按一个维度划分，得到面试人群和拒绝人群

3.2 聚类分析。结果是应聘简历中大概分为几个集群，以及每个集群特点

4 代码、画图见 Segmentation.R.

规则：TotalIntenCh2< 47260 为 PS；

TotalIntenCh2>= 47260, FiberWidthCh1>=11.2 为 WS；

TotalIntenCh2>= 47260, FiberWidthCh1<11.2, AvgIntenCh1< 165.1 为 PS；

TotalIntenCh2>= 47260, FiberWidthCh1<11.2, AvgIntenCh1>= 165.1, TotalIntenCh3< 55760 为 WS；

TotalIntenCh2>= 47260, FiberWidthCh1<11.2, AvgIntenCh1>= 165.1, TotalIntenCh3>= 55760 为 PS。