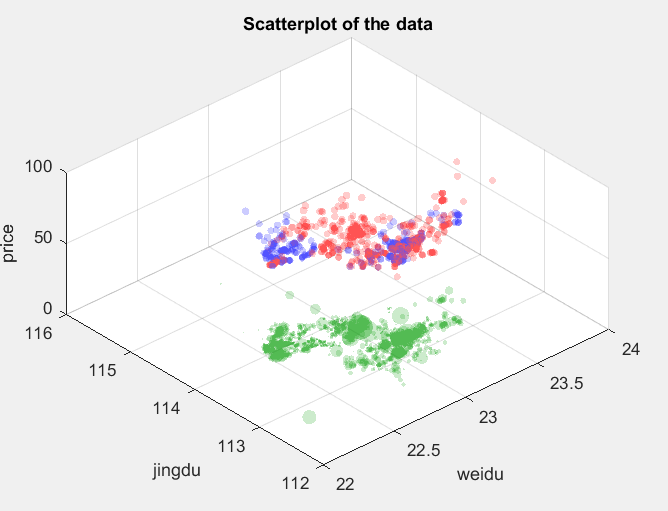
1 第一问

先将数据可视化，了解数据的信息。

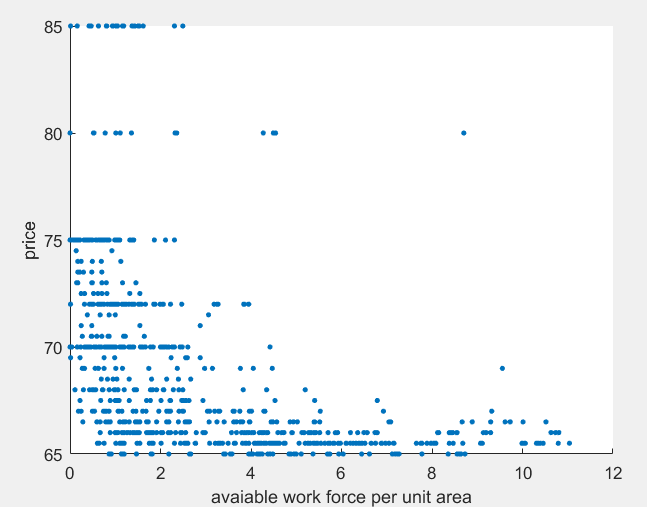
通过bubble plot画图



此图纵横坐标分别是经纬度，上面的小点是价格。红色是完成的任务，蓝色是未做的任务。颜色的深浅体现了事件发生的频率，举个例子，

Price=0所在平面是会员的信息，点的具体坐标表示了会员的所在位置，点的大小描述了会员的预定任务限额（也就是能够完成的任务数目）。

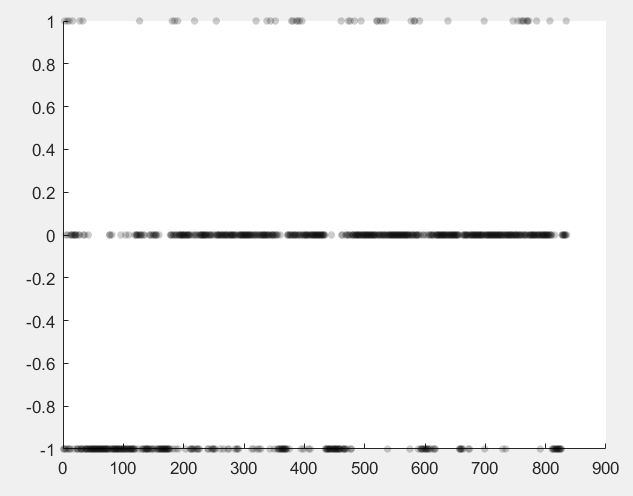
基本思路是任务的定价与任务周围的区域单位面积内任务完成能力相关（举个例子，某任务周围有10个人，把这10个人的任务限额全部相加，将所得到的和除以面积，就是单位面积内任务完成能力）。



此图反映了任务周边单位面积可使用人力与任务定价的关系，当任务周边可用人力较多时，由于竞争激烈，可以将价格订的较低，因此在5~12的范围内基本都是低价。

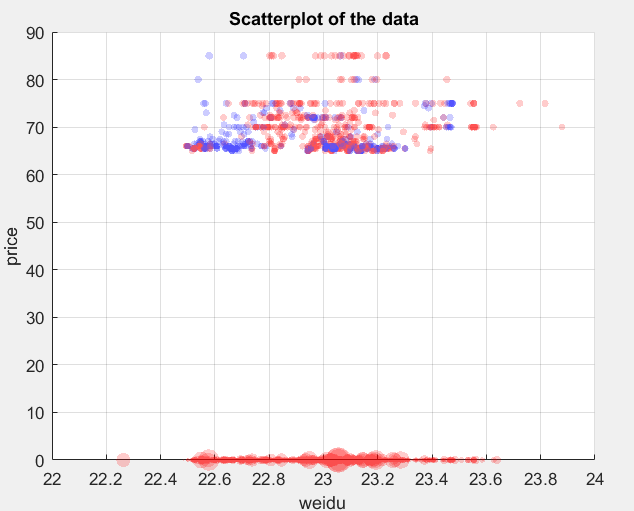
当可用人力较少时，需要通过控制价格来吸引周边不多的人力来完成任务，因此最高价主要出现在可使用人力较少的地区。

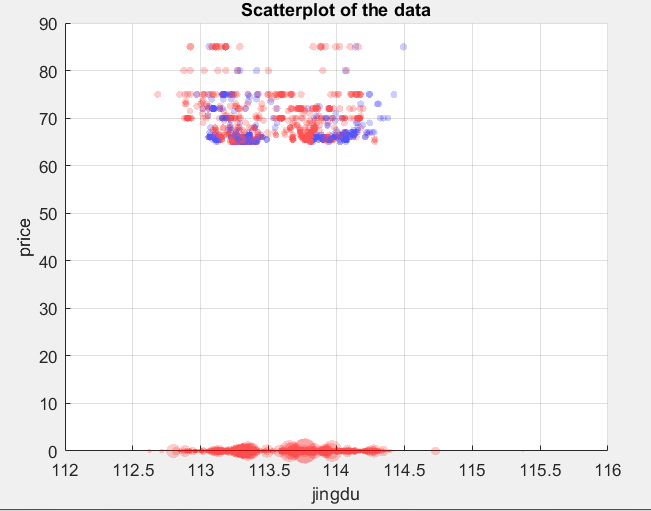
以可使用人力的密度、任务的价格为自变量，对任务完成情况做logistic回归，



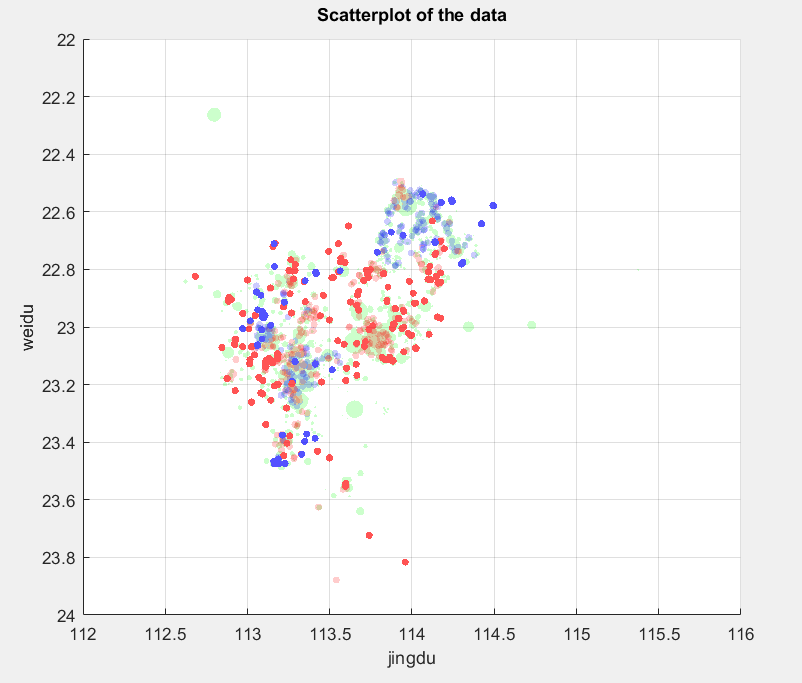
回归的情况，位于0的区域为成功预测结果的情况。

解释：任务未完成的原因





看横轴时，大的红圈表示可使用劳动力较多处，根据定价策略，应当定价较低。观察蓝色区域（即未完成任务），发现价格较低处（65总有）未完成任务出现的频次较高，区域颜色明显较深。一个可能的解释是在对这些任务定价定得过低，对会员们没有足够的吸引力。



同时，值得注意的是蓝色在空间位置上也有明显的聚集现象，在精度114~114.5和维度22.6~22.8范围内，大量出现未完成任务。一个可能的解释是在该区域定价策略失效，无法对该区域的会员产生足够吸引力。