**2020年全国高校计算机能力挑战赛**

**大数据应用赛答题卡第五部分**

**队长姓名： 印体锐 队长学校： 合肥工业大学**

# 第五部分 数据分析文档（5分）

主要数据处理的处理方式，数据分析的思路，数据可视化部分的呈现思路。

文档描述：

数据可视化的思路：

数据处理方面通过引入pandas，对大的数据表gongyu\_utf8.csv通过groupby

()函数分条件拆分成不同的小的Dataframe，条件例如区域，房价，户型等等。

图形显示方面选择了matplotlib和seaborn。题目中的饼图、直方图、柱状图分别对应 plot.pie(),seaborn.displot(),seaborn.barplot()函数，可以教直观的看出数据分布特性。

结果截图：





