

深圳市数据分析岗位需求和薪资情况分析

项目简介

想要从事深圳市的数据分析岗位，就要先对这个岗位和城市需求有相应的了解，我能想到的比较直接的办法是从企业招聘需求中获取讯息来指导自己的学习方向和简历准备。

本次项目拟用集搜客爬取拉钩网上数据分析和商业数据分析岗位信息，然后进行数据清洗和分析，了解深圳市数据分析岗位需求和特点。

数据来源和数据集

本项目所使用的数据集来自拉勾网，是通过集搜客这一网络爬虫工具来爬取的。本次爬取得到的字段信息如下：

内容	字段名称
公司名称	company
岗位名称	title
所属行业	industry
月薪	month_salary
融资阶段	phase
公司规模	scale
所属区	district
经验要求	experience
学历要求	quanlification
全职/兼职	full_or_parttime
职位描述和任职要求	description

在 Excel 中将数据表去掉无关列（如 “_clueid”，_fullpath’ 等）并将所有字段名称改为英文后得到如下表格：

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	company	title	industry	month_salary	phase	scale	district	experience	qualification	full_or_parttime	description							
2	传易	贷中数据分析师(010279)	移动互联网 社交	10k-20k	不需要融资	150-500人	南山区	经验1-3年	本科及以上	全职	工作职责 1、负责风险数据收集、风险识别、设计预警、数据挖掘、搭建风险监测系统。2、监测贷前贷中							
3	KLOOK 客路旅行	数据分析师	旅游	15k-25k	0轮及以上	500-2000人	福田区	经验3-5年	本科及以上	全职	工作职责 Strategize and work with tech team on data collection to enable business decisions, such as operati							
4	赢时胜软件	数据分析师 (DA)	金融	10k-20k	上市公司	500-2000人	福田区	经验3-5年	大专及以上	全职	岗位职责: 1、负责数据仓库产品的整体架构、模型设计2、负责数据相关需求调研、数据模型设计3、							
5	梦享网络	数据分析师	移动互联网 广告营销	13k-25k	0轮	50-150人	南山区	经验3-5年	学历不限	全职	岗位职责: 1、负责产品海量数据的提取及分析挖掘,为相关部门提供数据支持; 2、深入理解业务,协助							
6	KLOOK 客路旅行	数据分析师	旅游	18k-35k	0轮及以上	500-2000人	福田区	经验1-3年	本科及以上	全职	【岗位职责】1、用户模型研究,根据不同业务场景提供用户相关数据分析2、挖掘业务数据需求,为产品、市							
7	平安产险	数据分析师	金融	11k-15k	上市公司	2000人以上	福田区	经验1-3年	本科及以上	全职	岗位职责: 负责数字化商业分析工作,根据业务经营目标及策略,从用户、产品、营销、竞争等多角度开展							
8	华云卓越	数据分析师 (数据零售和数据服务 移动互联网)	10k-15k	不需要融资	500-2000人	南山区	经验不限	本科及以上	全职	工作内容: 1、基于成熟的可视化数据处理及分析工具配置数据报表门户; 2、负责数据报表的数据清洗优化								
9	华云中盛	数据分析师	金融	10k-15k	不需要融资	500-2000人	南山区	经验3-5年	本科及以上	全职	工作职责1、基于成熟的可视化数据处理及分析工具配置数据报表门户; 2、负责数据报表的数据清洗优化							
10	汇合	数据分析	企业服务 金融	8k-12k	不需要融资	2000人以上	南山区	经验1-3年	本科及以上	全职	1、全栈熟练指本科、数学、统计学相关专业优先2、1年以上数据分析工作经验3、熟练掌握python编程语言而							
11	慧择网	数据分析师	移动互联网 电商	5k-9k	0轮	500-2000人	南山区	经验不限	本科及以上	全职	岗位职责: 1、负责数据事业部报表数据分散、管理数据分散问题; 2、根据业务需求,制作日常经营数据口							
12	苏州精正	数据分析	企业服务 金融	28k-56k	不需要融资	少于15人	南山区	经验5-10年	本科及以上	全职	岗位职责 (工作内容) 1.收集、整理、统计、记录、分析运营所有各项日常运营指标数据,提供建议和决策; 2.能							
13	苏州精正	数据分析	企业服务 金融	20k-40k	不需要融资	少于15人	南山区	经验5-10年	本科及以上	全职	岗位职责1) 收集、整理、统计、记录、分析运营所有各项日常运营指标数据,提供建议和决策支持; 2) 能							
14	贝壳	贝壳-数据分析师	房产家居	7k-14k	0轮及以上	2000人以上	福田区	经验不限	本科及以上	全职	工作职责1.提供数据支持,熟悉新房相关业务,建立有效的业务健康测评体系,设计完善的数据看板,为业							

数据集一共 451 行，11 列

拟分析的问题

主要是希望通过实际的数据为自己和求职深圳数据分析岗位的小伙伴提供参考性的建议，具体针对以下几个问题：

- 1、深圳各区对数据分析人才的需求对比；
- 2、深圳各行业对数据分析的需求对比；
- 3、数据分析师岗位在深圳总体薪酬情况；
- 4、深圳不同行业数据分析师平均薪酬对比；
- 5、深圳数据分析师岗位随着经验的积累薪资的变化情况；
- 6、数据分析师所需要掌握的技能 and 需要具备的核心能力。

技术和工具

本项目主要分为两大部分。第一部分是数据爬取，采用的是集搜客爬虫工具。第二部分是数据清洗和分析，分析工具为 jupyter notebook，以 python 编程语言为主。用 Padas 来进行数据整理和统计分析，用 matplotlib 作图形的可视化，用 seaborn 进行图形的美化。在进行技能需求分析的时候，用 jieba 作为分词工具包，并用 wordcloud 制作词云。

数据整理

```
1. import pandas as pd
2. import numpy as np
3. import matplotlib as mpl
4. import matplotlib.pyplot as plt
5.
6. #读取数据
7. data = pd.read_excel('/Users/wenying/Desktop/Data_Lagou.xlsx')
8. #清理重复行
9. jobs=data.drop_duplicates(["company","title","description"])
10. #显示数据集大小和前 5 行
11. print(jobs.info)
12. jobs.head(5)
```

	company	title	industry	month_salary	phase	scale	district	experience	quanlification	full_or_parttime	description
0	传易	贷中数据分析师(J10279)	移动互联网, 社交	10k-20k	不需要融资	150-500人	南山区	经验1-3年	本科及以上	全职	工作职责:\n1. 负责风险数据收集、风险识别、设计预警、数据挖掘, 搭建风险监测系...
1	KLOOK 客路旅行	数据分析师	旅游	15k-25k	D轮及以上	500-2000人	福田区	经验3-5年	本科及以上	全职	工作职责:\n\n Strategize and work with tech team on...
2	赢时胜软件	数据分析师 (DA)	金融	10k-20k	上市公司	500-2000人	福田区	经验3-5年	大专及以上	全职	岗位职责: \n1. 负责数据仓库产品的整体架构、模型设计\n2. 负责数...
3	梦享网络	数据分析师	移动互联网, 广告营销	13k-25k	B轮	50-150人	南山区	经验3-5年	学历不限	全职	岗位职责: \n1、负责产品海量数据的提取及分析挖掘, 为相关部门提供数据支持。 \n2、深...
4	KLOOK 客路旅行	数据分析	旅游	18k-35k	D轮及以上	500-2000人	福田区	经验1-3年	本科及以上	全职	【岗位职责】 \n1. 用户模型研究, 根据不同业务场景提供用户相关数据分析\n2. 挖掘业务数...

去除重复后，数据集显示有 448 行，11 列数据，数据完整度很好，便于后续的分析。

数据分析

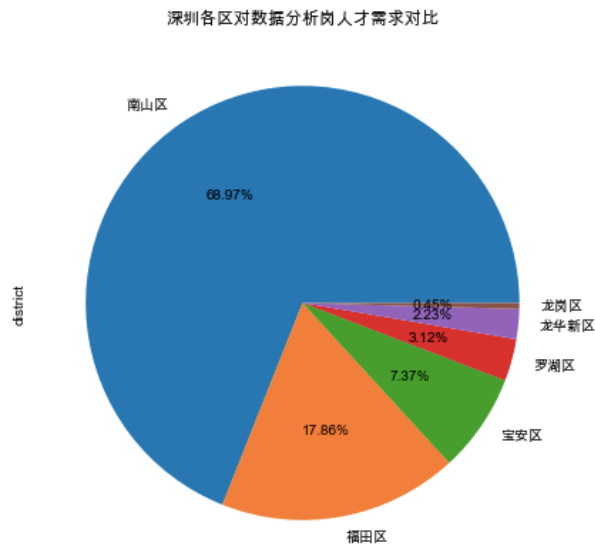
一、深圳市各区对数据分析岗人才需求对比

```
1. # 查看 Mac 电脑中的中文字体情况
2. import matplotlib
3. font_list=sorted([f.name for f in matplotlib.font_manager.fontManager.ttflis
t])
4. for i in font_list:
5.     print (i)
6. #选定一种中文字体，使图形标注能正常显示中文
```

```

7. plt.rcParams['font.family'] = ['Arial Unicode MS']
8. #绘制深圳各区数据人才需求饼状图
9. plt.rcParams["figure.figsize"]=7,7]
10. plt.title("深圳各区对数据分析岗人才需求对比")
11. jobs["district"].value_counts().plot(kind="pie",autopct="%1.2f%%")

```



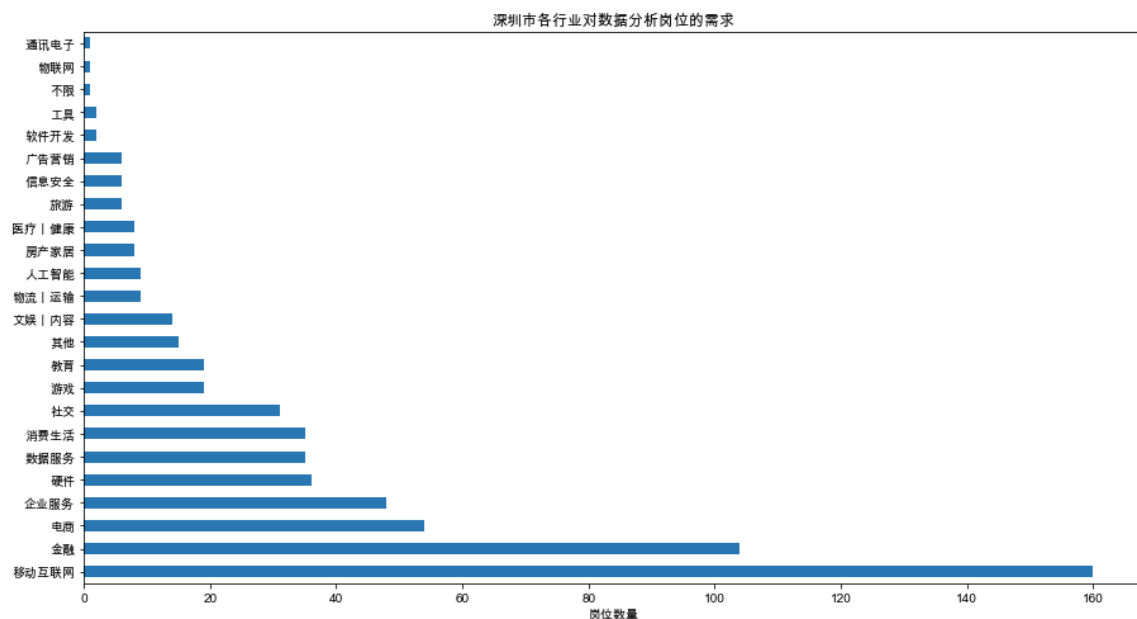
由上图可见，深圳市数据分析岗位相对集中在南山、福田、宝安这三个区，其中南山区需求占比最大，为 68.97%，想在深圳求职数据分析的话南山区是个很不错的选择。

二、深圳各行业对数据分析岗的需求对比

```

1. #由于每个公司可能会有不止一个行业背景，所以需要拆分“industry”这一列，拆分后的这一列重新命名为“industries”，形成的新表命名为“new_data”
2. new_data=jobs.drop("industry",axis=1).join(jobs["industry"].str.split(", ",expand=True).stack().reset_index(level=1,drop=True).rename("industries"))
3. print(new_data["industries"])
4. #绘制深圳市各行业对数据分析岗位需求的横向柱状图
5. fig=plt.figure(figsize=(15,8))
6. ax1=plt.subplot(111)
7. new_data["industries"].value_counts().plot(kind="barh")
8. plt.title("深圳市各行业对数据分析岗位的需求")
9. plt.xlabel("岗位数量")
10. plt.show()

```



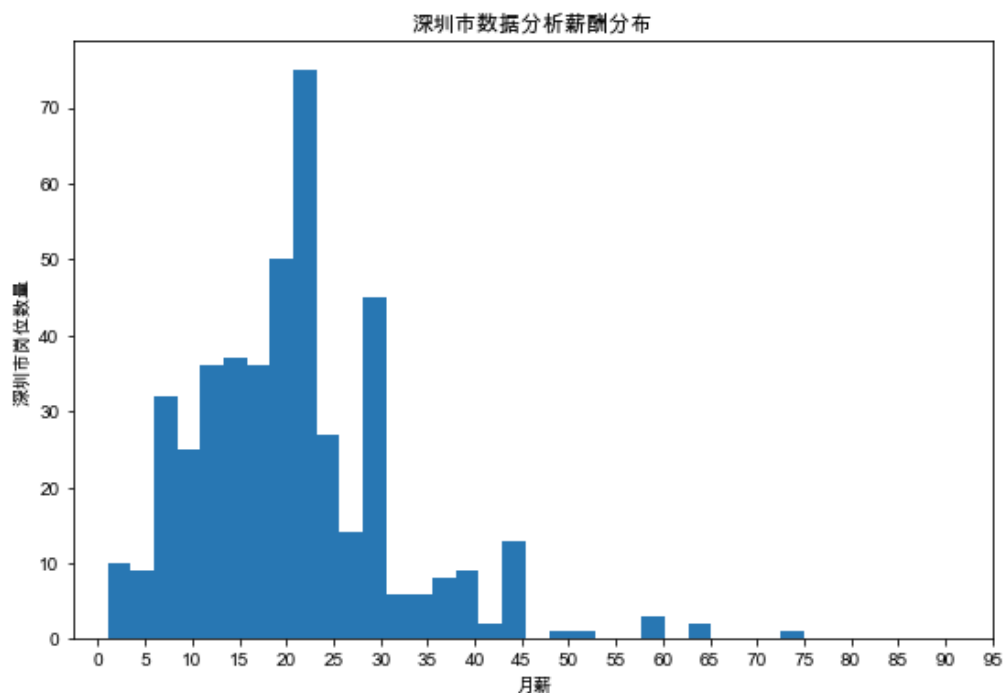
由上图可以看到，深圳市移动互联网和金融领域对数据分析岗位需求最大，其次为电商和企业服务领域。对于想转行数据分析领域的小伙伴来说，互联网、金融、电商这几个领域是不错的选择：可以分析的数据量大，行业规模成熟，对数据分析能力的提升无疑会更快。

三、数据分析岗位在深圳总体薪酬情况

```

1. #因为薪酬那一列显示是一个区间，并且相互间有重叠不方便进行后续数据分析，所以考虑定义一个函数，取区间两端中值作为月薪资替代值
2. def avg_salary(salary):
3.     try:
4.         s_list = salary.split('-')
5.         s_min = int(s_list[0][: -1])
6.         s_max = int(s_list[1][: -1])
7.         s_avg = float(s_min + s_max)/2
8.     except ValueError:
9.         s_list = salary.split('k')
10.        s_avg = float(int(s_list[0]))
11.    return s_avg
12. jobs['avg_month_salary'] = jobs['month_salary'].apply(avg_salary)
13. #绘制深圳市数据分析薪酬分布条形图
14. fig = plt.figure(figsize = (9,6))
15. ax2 = plt.subplot(111)
16. rect = ax2.hist(jobs['avg_month_salary'], bins = 30)
17. ax2.set_title(u'深圳市数据分析薪酬分布')
18. ax2.set_xlabel(u'K/月')
19. plt.xticks(range(0,100,5))
20. plt.xlabel("月薪")
21. plt.ylabel("深圳市岗位数量")

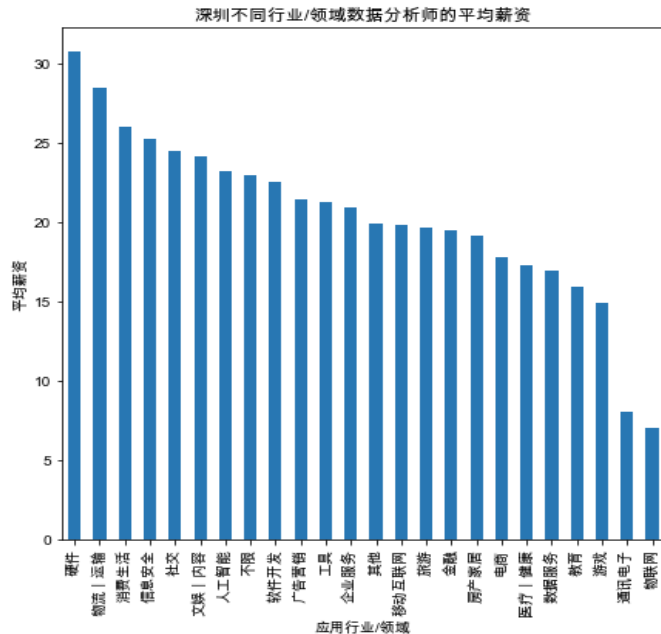
```



从上图可以看到，深圳大多数数据分析师的收入集中在 **5k-30k** 每月，少部分人能够获得更高的薪酬(**30-45k/月**)，极少数人薪酬达到 **50-75k/月**，让人充满期待。需要说明的是，因为岗位薪酬值我取的是区间的中值作为代表值进行的分析，因此，实际的薪酬分布情况可能会比图中的情况更好一些。综合来看，深圳数据分析师的整体薪酬收入和其他行业相比还是有较大的优势。

四、深圳不同行业数据分析师平均薪酬对比

```
1. #将 avg_month_salary 这一列数据加到上文中对行业分列后的数据集 new_data 中去
2. new_data['avg_month_salary'] = new_data['month_salary'].apply(avg_salary)
3. new_data.head(5)
4. #绘制深圳不同行业或领域的数据分析师的平均薪资柱状图
5. new_data.groupby(['industries'])['avg_month_salary'].mean().sort_values(ascending=False).plot(kind="bar")
6. plt.xlabel("应用行业/领域")
7. plt.ylabel('平均薪资')
8. plt.title('深圳不同行业/领域数据分析师的平均薪资')
```



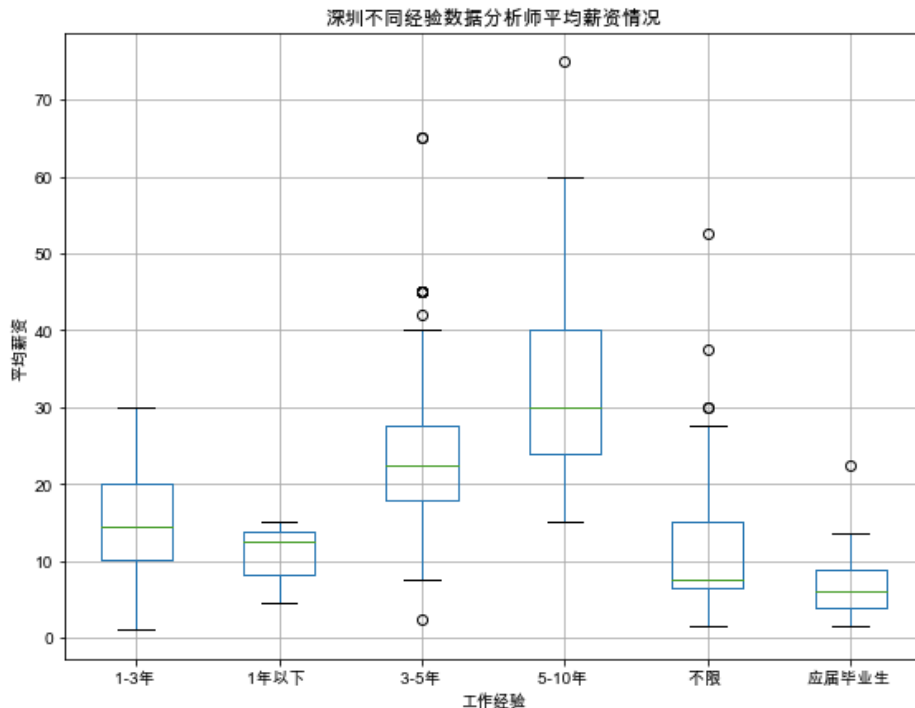
由上图看出，在深圳硬件细分领域的数据分析师平均薪资最高，除通讯电子和物联网领域数据分析师的平均薪资偏低，其他领域的数据分析师平均薪资都在 14k+

五、深圳数据分析师岗位随着经验积累薪资的变化情况

```

1. #整理工作经验这一列，去掉每个单元格前面的‘经验’二字
2. def remove_first_2_word(word):
3.     return word[2:]
4. jobs['experience'] = jobs['experience'].apply(remove_first_2_word)
5. #深圳市各工作经验下对应的岗位数量
6. experience_count=jobs.groupby(["experience"])[ "title"].count()
7. print(experience_count)
8. #找出”10 年以上“工作经验对应的那一行信息
9. print(jobs.loc[jobs['experience'].str.contains('10 年以上')])
10. # 因为工作经验为 10 年以上的岗位招聘需求只有一个，其薪资不能代表市场上平均水准，故删除此行后进行后续经验与平均薪资的统计
11. new_jobs=jobs.drop([431])
12. #绘制工作经验与平均薪资关系的箱线图
13. new_jobs.boxplot(column="avg_month_salary",by="experience",figsize=(9,7))
14. plt.title("深圳不同经验数据分析师平均薪资情况")
15. plt.xlabel("工作经验")
16. plt.ylabel("平均薪资")

```



从箱线图中可以看出，在10年内，工作经验为5-10年的数据分析师岗位最高，平均月薪为30k，个别达到70k；工作经验在3-5年的数据分析师平均月薪大于2k，工作经验为1-3年的数据分析师平均月薪在10-20k之间，看来数据分析师这一岗位随着经验的积累，薪资上升空间还是比较大的。

六、数据分析师所需掌握的技能 and 具备的相关能力

```

1. #加载分词和词云制作相关包
2. import jieba
3. import jieba.analyse
4. from wordcloud import WordCloud
5. #抽取 description 字段关键词,放在新建的列"key_words"下
6. def key_words(text):
7.     key_words = jieba.analyse.extract_tags(text, topK=20, withWeight=False,
8. allowPOS=())
9.     return key_words
10. jobs['key_words'] = jobs['description'].apply(key_words)
11. jobs.head(1)
12. # 对 key_word 这一列使用精确模式划分,得到"cut_words"列表
13. cut_words=jieba.lcut(str(jobs["key_words"].values),cut_all=False)
14. cut_words
15. #将分词结果中不相关的词去掉,其他放在列表"word_list"中

```


分析结论

一、需求情况：就区域需求来说，在深圳，数据分析岗位大量的招聘需求集中在南山和福田这两个区，其中南山区相应工作机会占比最大，为 68.97%，是数据分析求职 不错的选择；就行业需求来说，深圳市移动互联网和金融领域对数据分析岗的需求量最大，其次为电商和企业服务领域；就工作经验的需求来说，经验在 3-5 年的数据分析师在深圳比较吃香，其次为 1-3 年，可以看出这个职业是一个比较年轻化的职业。

二、薪资待遇情况：就总体薪资水平来说，深圳大多数数据分析师的收入集中在 5k-30k 每月，少部分人能够获得 30-45k 的月薪酬，极少数人能达到 50-75k/月；就行业/细分领域对薪资影响来说，除通讯电子和物联网领域的数据分析师的平均薪资偏低，其他领域的数据分析师在深圳的平均薪资都在 14k+；就工作经验对数据分析师薪资影响来说，工作经验为 1-3 年的数据分析师深圳平均月薪在 10-20k 之间，3-5 年的数据分析师在深圳的平均月薪大于 20k，5-10 年的平均月薪在 30k 左右，可见这一职业随着经验积累，薪资上升空间还是比较大的。

三、所需技能和核心能力：在深圳，数据分析师需求频率排在前列的技能包括 SQL、Python、Excel、数据挖掘和数据建模、Hive、Hadoop，另外，对业务的理解和熟悉程度、沟通能力也是招聘企业比较看中的。