**阜阳师范大学**

**计算机与信息工程学院**

**《****数据可视化分析》期末项目**

**（项目名称** **基于Django框架的数据分析岗位数据可视化分析）**

**姓 名： 周子苏**

**学 号：2019114054**

**班 级：数据科学与大数据技术**

**完成时间：2022年05月13日**

阜阳师范大学计算机与信息工程学院

**项目主题**

## 1 数据来源

### 1.1背景描述

这门课程是数据分析课程，当下数据分析是非常火热的领域，不论是什么研究基本上研究出来的结果都需要用数据分析的方法可视化出结果。因此我将数据分析师的岗位薪资待遇信息来分析一下当下数据分析师的主要薪资水平。一名大三的学生，面对即将结束的大学生活，大学毕业后是去考研还是去找工作是个值得深思的问题，为了给大家提供合理的建议，帮助大家更加了解当前行业的现状和要求，我做了关于Python相关岗位的数据分析，从各个维度带大家去了解当前的就业形势和行业现状，希望能够为大学毕业之后去找工作的同学提供帮助。

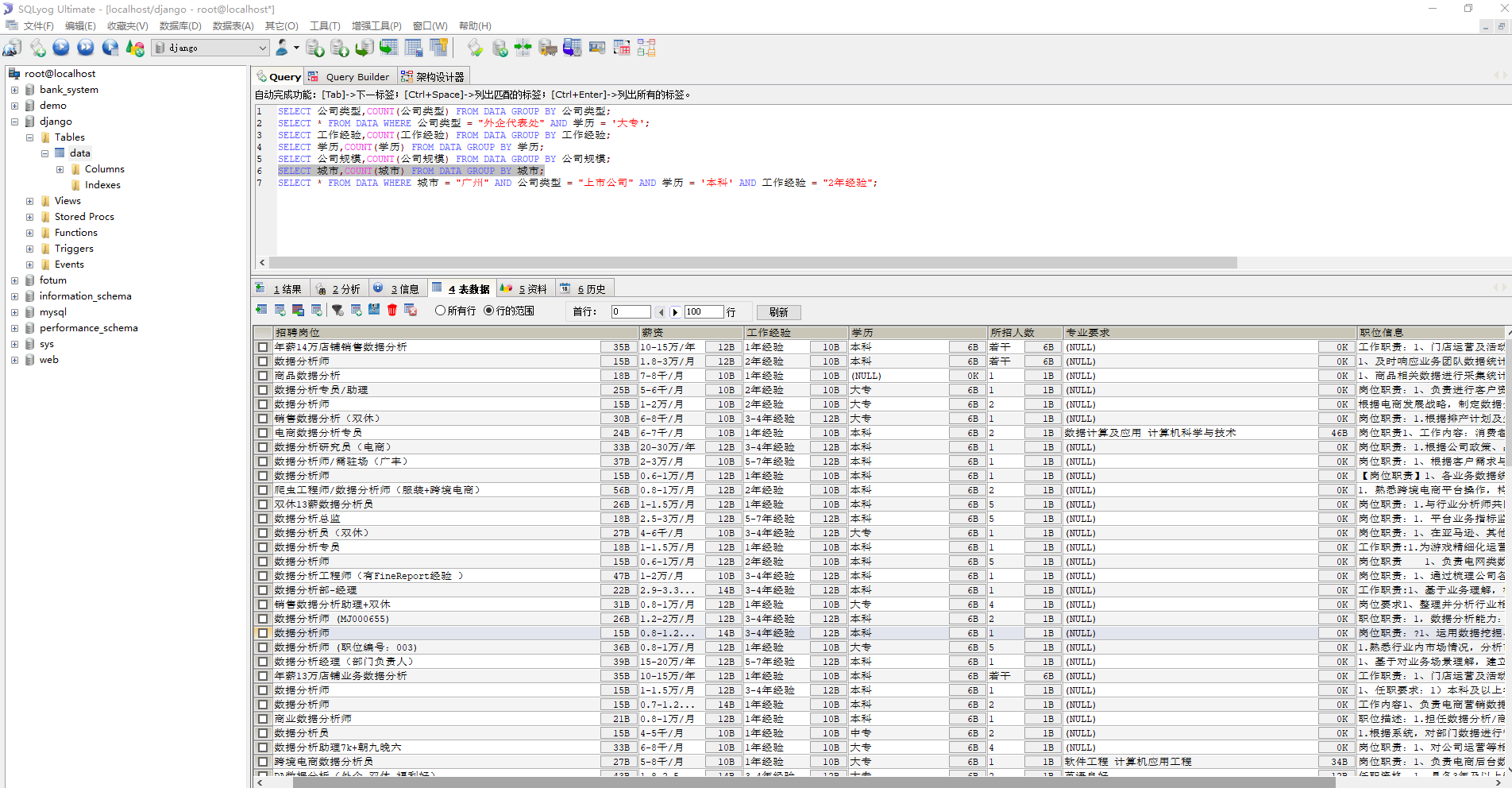
1.2 **数据获取来源**

网站链接：<https://www.heywhale.com/mw/dataset/62834c0dc08fe7829f219f1e/file>

## 2 数据预处理

### 2.1 数据库存放数据 （提取数据）

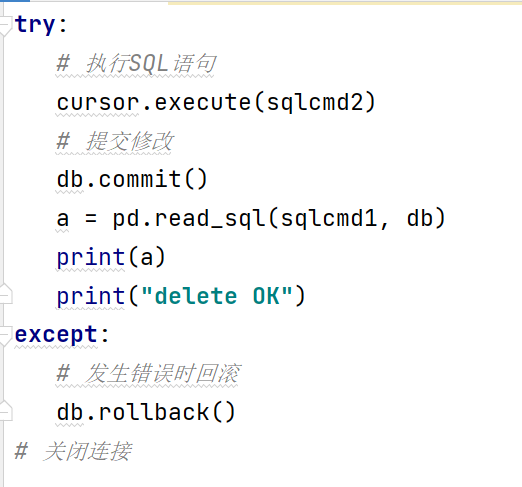
通过read\_csv、to\_sql来将数据转存到数据库当中



### 2.2 数据清洗

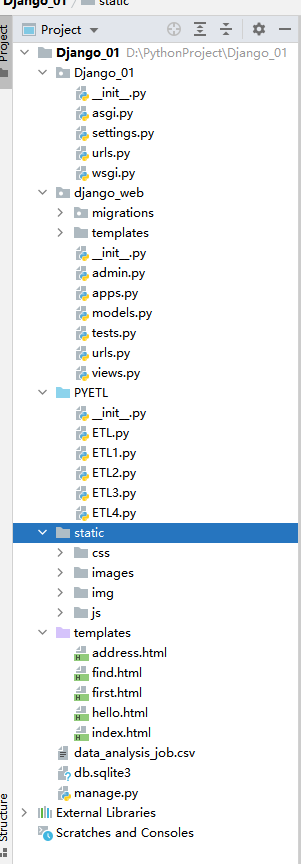
data.csv文件中有一些数据存在null值

DELETE FROM data WHERE XXX IS NULL



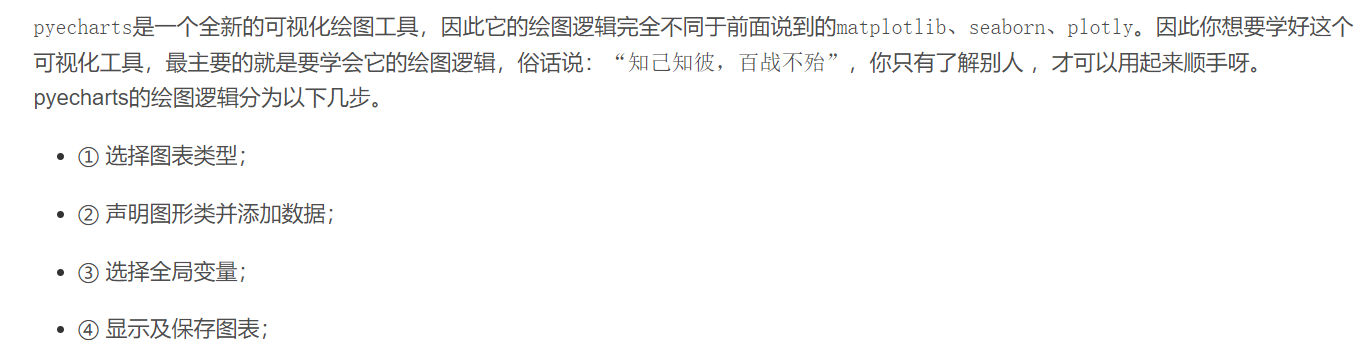
## 3数据分析

### 3.1 整体框架



### 3.2 相关技术

Django、Pyecharts



### 3.3 前后端交互

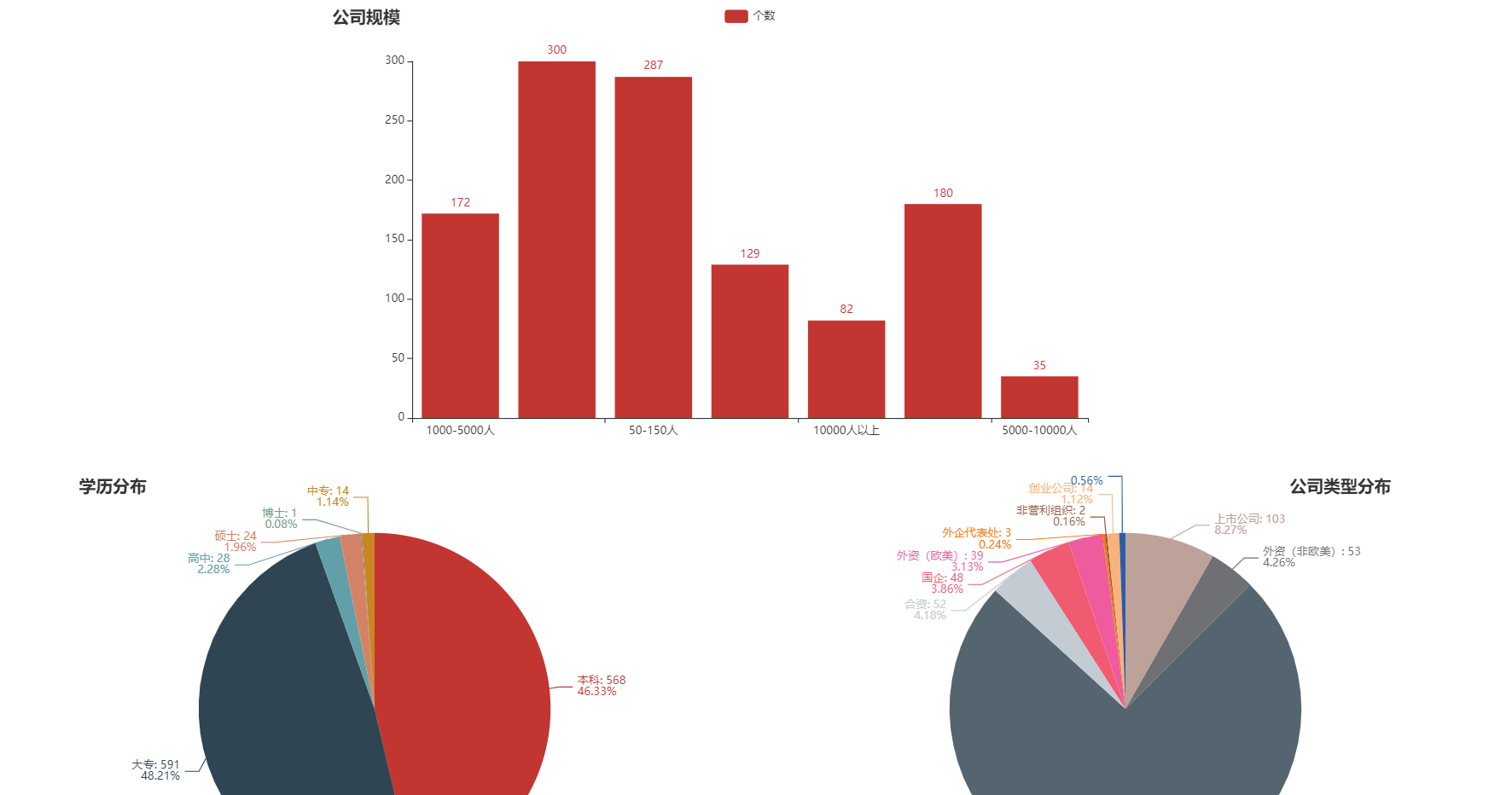
1.ETL3

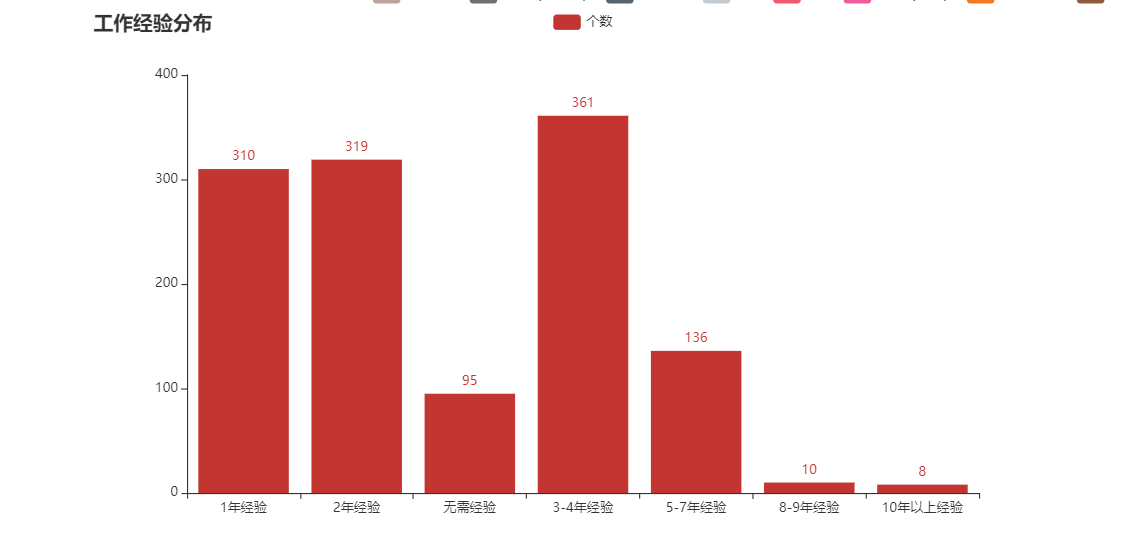
**import** pandas **as** pd  
**import** pymysql  
  
  
db = pymysql.connect(  
 host=**"localhost"**,  
 database=**"django"**,  
 user=**"root"**,  
 password=**"111"**,  
 port=3306,  
 charset=**'gbk'** )  
  
**def** find\_data(x,y,z,k):  
 sql1 = **"SELECT \* FROM DATA WHERE 城市 = '%s' AND 公司类型 = '%s' AND 学历 = '%s' AND 工作经验 = '%s'"** % (x,y,z,k)  
 a = pd.read\_sql(sql1, db)  
 list1 = a.values.tolist()  
 **return** list1

2.ETL4

**from** pyecharts **import** options **as** opts  
**from** pyecharts.charts **import** Bar, Grid, Pie, Page  
**def** visualization(list1):  
 work\_experience, education, company\_type, company\_size = {}, {}, {}, {}  
 **for** i **in** list1:  
 **if** i[3] **not in** work\_experience:  
 work\_experience[i[3]] = 1  
 **else**:  
 work\_experience[i[3]] += 1  
 **if** i[4] **not in** education:  
 education[i[4]] = 1  
 **else**:  
 education[i[4]] += 1  
 **if** i[8] **not in** company\_type:  
 company\_type[i[8]] = 1  
 **else**:  
 company\_type[i[8]] += 1  
 **if** i[9] **not in** company\_size:  
 company\_size[i[9]] = 1  
 **else**:  
 company\_size[i[9]] += 1  
 **if None in** list(work\_experience.keys()):  
 work\_experience.pop(**None**)  
 **if None in** list(education.keys()):  
 education.pop(**None**)  
 **if None in** list(company\_size.keys()):  
 company\_size.pop(**None**)  
 **if None in** list(company\_type.keys()):  
 company\_type.pop(**None**)  
  
 bar1 = Bar()  
 bar1.add\_xaxis(list(company\_size.keys()))  
 bar1.add\_yaxis(**"个数"**, list(company\_size.values()))  
 bar1.set\_global\_opts(title\_opts=opts.TitleOpts(title=**"公司规模"**))  
  
 pie1 = (Pie()  
 .add(**""**, list(zip(list(education.keys()), list(education.values()))), center=[**"25%"**, **"50%"**])  
 .set\_global\_opts(title\_opts=opts.TitleOpts(title=**"学历分布"**, pos\_left=**'5%'**),  
 legend\_opts=opts.LegendOpts(pos\_left=**"5%"**, pos\_bottom=**"5%"**))  
 .set\_series\_opts(label\_opts=opts.LabelOpts(formatter=**"{b}: {c}\n{d}%"**))  
 )  
  
 bar2 = Bar()  
 bar2.add\_xaxis(list(work\_experience.keys()))  
 bar2.add\_yaxis(**"个数"**, list(work\_experience.values()))  
 bar2.set\_global\_opts(title\_opts=opts.TitleOpts(title=**"工作经验分布"**))  
  
 pie2 = (Pie()  
 .add(**""**, list(zip(list(company\_type.keys()), list(company\_type.values()))), center=[**"75%"**, **"50%"**])  
 .set\_global\_opts(title\_opts=opts.TitleOpts(title=**"公司类型分布"**, pos\_right=**'7%'**),  
 legend\_opts=opts.LegendOpts(pos\_right=**"5%"**, pos\_bottom=**"0%"**))  
 .set\_series\_opts(label\_opts=opts.LabelOpts(formatter=**"{b}: {c}\n{d}%"**))  
 )  
 grid1\_1 = Grid(init\_opts=opts.InitOpts(width=**'1600px'**))  
 grid1\_1.add(pie1, grid\_opts=opts.GridOpts(pos\_right=**"55%"**))  
 grid1\_1.add(pie2, grid\_opts=opts.GridOpts(pos\_left=**"55%"**))  
 page\_1 = Page(layout=Page.SimplePageLayout)  
 page\_1.add(bar1)  
 page\_1.add(grid1\_1)  
 page\_1.add(bar2)  
  
 **return** page\_1

## 4 可视化结果







## 结论

**从公司规模来看，各地区具有差异性：一线城市大公司较多**

**但普遍分布在50-150、150-500这两个阶段；**

**从学历分布来看，本科生依旧是就业主力军，其次为大专生；**

**从公司类型来看，民营企业占大比重，数据分析在国企和事业单位**

**也具有一定市场；**

**从工作经验来看，企业对工作经验比较重视，数据分析岗具有一定**

**经验比较吃香，3-4年的经验可以作为跳槽，涨薪的支撑。**