

# The doctor-patient behavior and medical resource allocation of Internet medical platform

廖家瑞 (Jiarui Liao) 1801212884

刘德明 (Deming Liu) 1801212889

曹阳 (Yang Cao) 1801212825

王如思 (Rusi Wang) 1801212938

常鑫磊 (Xinlei Chang) 1801212779

2020-01-13

## Project Introduction

Data Introduction and Preprocessing

Data Visualization

Data Modeling

Project Summary and Outlook

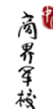
- China has made remarkable achievements in the development of medical and health services, but still faces some challenges, such as information asymmetry between doctors and patients, and uneven distribution of medical resources between regions, urban and rural areas.
- These problems lead to the contradiction between doctors and patients as well as the "difficulty and high cost" of high-quality hospitals.



# Project Introduction



**PHBS**  
北京大学汇丰商学院



- The convergence of Internet technology and medicine has brought about the emergence of Internet medical platforms, including medical service retrieval, medical knowledge sharing (DXY.com), online consultation and appointment website (haodf.com).
- Internet medical treatment will help patients to search for information, and also facilitate the reorganization and sharing of doctors' knowledge and skills and other resources in time and space. And this is conducive to the rational allocation of high-quality medical resources.

The screenshot shows the Haodf.com website interface. At the top, there is a navigation bar with the logo '好大夫在线' and the URL 'www.haodf.com'. Below the navigation bar, there is a search bar and a '搜索' button. The main content area features a large banner for '心脏支架术后到底能不能剧烈运动?' (Can I do intense exercise after heart stent surgery?). Below the banner, there is a section for '找好大夫' (Find Good Doctors) with a list of medical specialties and a table of doctors. The table lists various medical specialties such as Internal Medicine, Surgery, and Pediatrics, along with the names of doctors and their qualifications. On the right side of the interface, there is a '医生入口' (Doctor Entry) section with a '登录' (Login) button and a '注册' (Register) button. Below this, there is a '好大夫在线提示' (Haodf.com Online提示) section with a cartoon character and a message about the platform's purpose. At the bottom, there is a '最新出诊/停诊预报' (Latest Outpatient/Stopout Forecast) section with a table of hospital appointments and a '权威专家视点' (Authoritative Expert Viewpoint) section with a '更多' (More) link.

好大夫在线  
www.haodf.com

我的中心 我的订单 我的医生 手机好大夫 联系客服

首页 找大夫 按医院找 按疾病找 按专科找 专家团队 就诊 预约加号 海外就诊 品牌汇  
咨询 网络咨询 电话咨询 私人医生 家庭医生 指南 专家访谈 60秒问答 疾病专题 快速找药

心脏支架术后到底能不能剧烈运动?  
汪奇 副主任医师  
中国人民解放军总医院(301医院) 心血管内科

医生入口 登录 | 注册

好大夫在线提示  
好大夫在线是医患沟通平台, 医生基于患者自述病情所发表的言论仅供参考, 不能作为诊断和治疗的直接依据。

217003名权威专家在线解答  
立即免费咨询

最新出诊/停诊预报  
更多>>

权威专家视点  
更多>>

内科	外科	心脏	糖尿病	肺炎	哮喘	肾炎	脂肪肝	胃溃疡	类固醇	脑梗塞	更多>>
乙型肝炎	气胸	冠心病	糖尿病	肺炎	哮喘	肾炎	脂肪肝	胃溃疡	类固醇	脑梗塞	更多>>
贫血	肾结石	高血压	糖尿病	肺炎	哮喘	肾炎	脂肪肝	胃溃疡	类固醇	脑梗塞	更多>>
更多>>	更多>>	更多>>	更多>>	更多>>	更多>>	更多>>	更多>>	更多>>	更多>>	更多>>	更多>>

妇科	儿科	肿瘤	其他	不孕不育	月经失调	盆腔炎	子宫肌瘤	更多>>
不孕不育	小儿癫痫	肺癌	青光眼	不孕不育	月经失调	盆腔炎	子宫肌瘤	更多>>
多囊卵巢	小儿腹泻	肝癌	尖锐湿疣	多囊卵巢	卵巢囊肿	产前检查	更多>>	更多>>
更多>>	更多>>	更多>>	更多>>	更多>>	更多>>	更多>>	更多>>	更多>>

# Project Introduction



- Based on the doctor-patient data of haodf.com, this project focuses on the improvement effect of the Internet on the imbalance of medical resources by studying **the geographical distribution of patients and their willingness to consult.**

1

**Visualization of geographical distribution of medical interaction**

2

**Factors influencing the difference in medical interaction decision making (Regional factors, Willingness to consult)**

**咨询专家** 正规医院的 217003 位专家本人在线答疑

在此咨询, 21万名权威专家在线解答

我要免费咨询



卜丽萍 主任医师  
上海中山医院心内科  
今天

电话咨询

全部 精神科 妇科 肿瘤科 儿外科 儿内科 生殖中心 产科 内科 皮肤科

 <div>郝春生 主任医师 儿研所 小儿泌尿外科</div>	孩子小便正常, 也没有... 最近通话: 01-09 <a href="#">查看最新用户分享 &gt;&gt;</a>	200元/次(最长10分钟)	立刻预约
 <div>罗成华 主任医师 北京大学国际医院 腹膜后肿瘤外科</div>	主治医生说无法根治, ... 最近通话: 01-09 <a href="#">查看最新用户分享 &gt;&gt;</a>	500元/次(最长10分钟)	立刻预约
 <div>邱伟民 副主任医师 上海市第一妇婴保健院 妇科</div>	最近通话: 暂无 <a href="#">查看最新用户分享 &gt;&gt;</a>	180元/次(最长15分钟)	立刻预约

电话咨询仅提供建议, 不能诊断、处方、治疗。  
100%专家本人接听, 90%当天通话, 满意度高达99%!

查看全部大夫 >>

网上咨询

Project Introduction

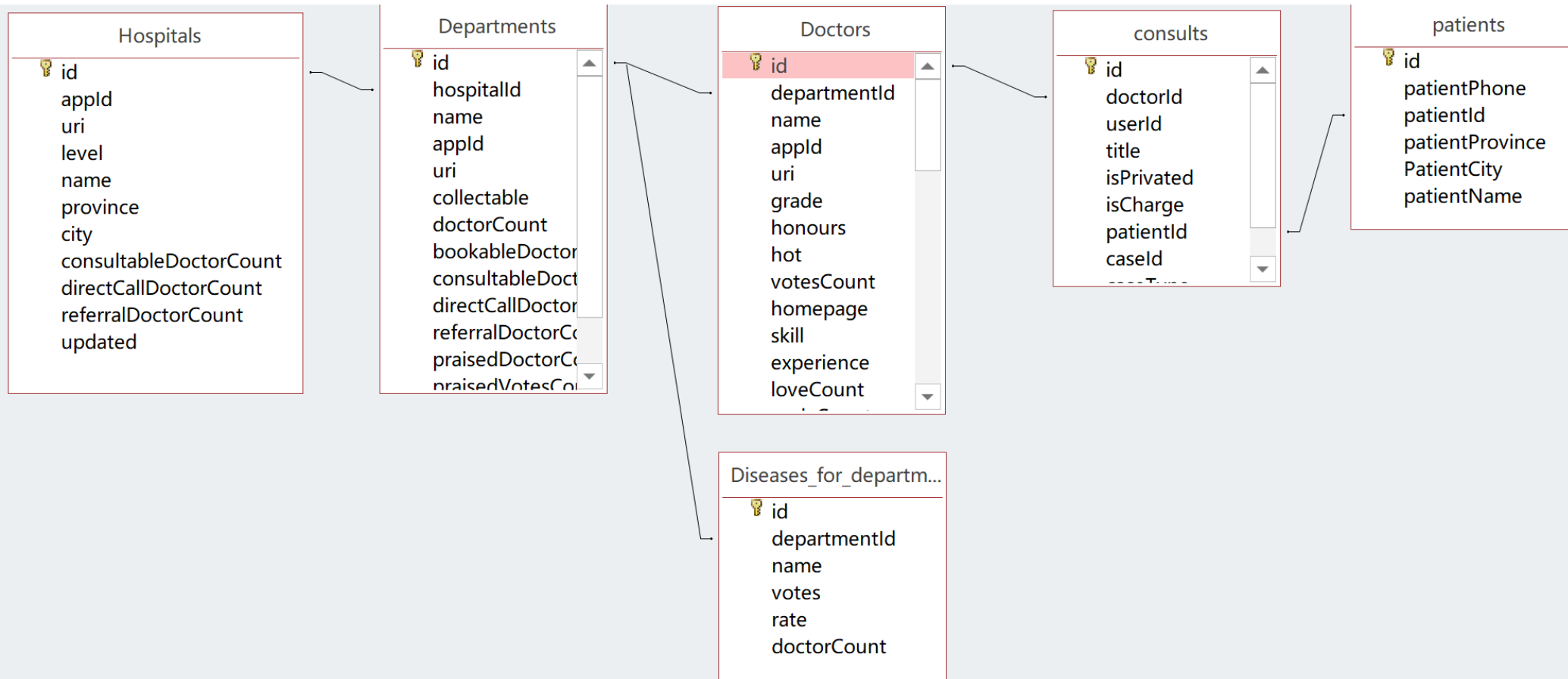
Data Introduction and Preprocessing

Data Visualization

Data Modeling

Project Summary and Outlook

- On the website of Haodf, we collected all relevant data from most departments of all hospitals in some regions (Shanghai, Anhui and Gansu were sampled accordingly)
- Data mainly includes information of hospitals, doctors, patients.





## ■ Location, departments, level...

当前位置: 好大夫在线-智慧互联网医院 > 医院 > 上海第一人民医院 > 首页


### 上海第一人民医院(三级甲等)

 医院介绍: 上海市第一人民医院建于1864年3月1日, 时称Shanghai Genera...


[查看详情>>](#)

 地 址: 北部地址:上海市虹口区海宁路100号;南部地址:上海市新松江路650号

[看地图>>](#)

 怎么 走: 北部乘车线路: 13、63、921、17、910、6、61、5...

[查看详情>>](#)

 电 话: 021-63240090(本部总机),021-63240090-5218(咨询台),021-37798103(南部门诊办公室),021-37798853(南部办公室)

 咨询大夫: 418位大夫在线


[咨询本院大夫](#)

邵怡大夫 泌尿外科 今天在线

李朋大夫 男科/盆底与尿... 今天在线

擅长: 输尿管镜钬激光碎石术治疗输尿管...

擅长: 男性不育症规范化诊疗: 无精子症...

 免费转诊: 本院开通转诊服务的大夫74名

 [申请本院大夫免费转诊](#)

已成功转诊患者8431人次

 电话咨询: 本院可直接通话大夫255名

 [预约与本院大夫通话](#)



## ■ Location, departments, level...

当前位置: 好大夫在线-智慧互联网医院 > 上海市第一人民医院 > 心内科 > 首页

### 基本信息

科室介绍: 一、上海第一人民医院心血管内科室概况: 上海第一人民医院心血管内科最早成立于1864年, 早期隶属于大内科, 90年代独立建科, 在乐文熙、颜和昌、许群、江智文等历任科主任的带领下, 不断创造着心内科的辉煌历史。1998年后引进留学美国Johns Hopkins大学的孙宝贵教授担任科主任, 科室又进入了一个全新的飞速发...[完整介绍>>](#)

网络咨询: 本科室可线上提问大夫14名

 [立即咨询本科室大夫](#)

免费转诊: 本科室开通转诊服务的大夫9名

 [预约本科室大夫免费转诊](#)

已成功转诊患者698人次

电话咨询: 本科室可直接通话大夫13名

 [预约与本科室大夫通话](#)



- Location, consulting text, related diseases, number of consulting, time of consulting...

首页 患者服务区 电话咨询 家庭医生 预约挂号 文章 患友会 心意礼物

已扫13,003次

在此简单描述病情，向陈松文大夫提问

咨询陈松文大夫

陈松文的咨询范围：心律失常（室上速、预激、房早、房速、房扑、房颤、室早、室速等；房室传导阻滞、病态窦房结综合征等）、扩张型心脏病、心力衰竭等各类心血管疾病

服务区 (全部患者: 3692)

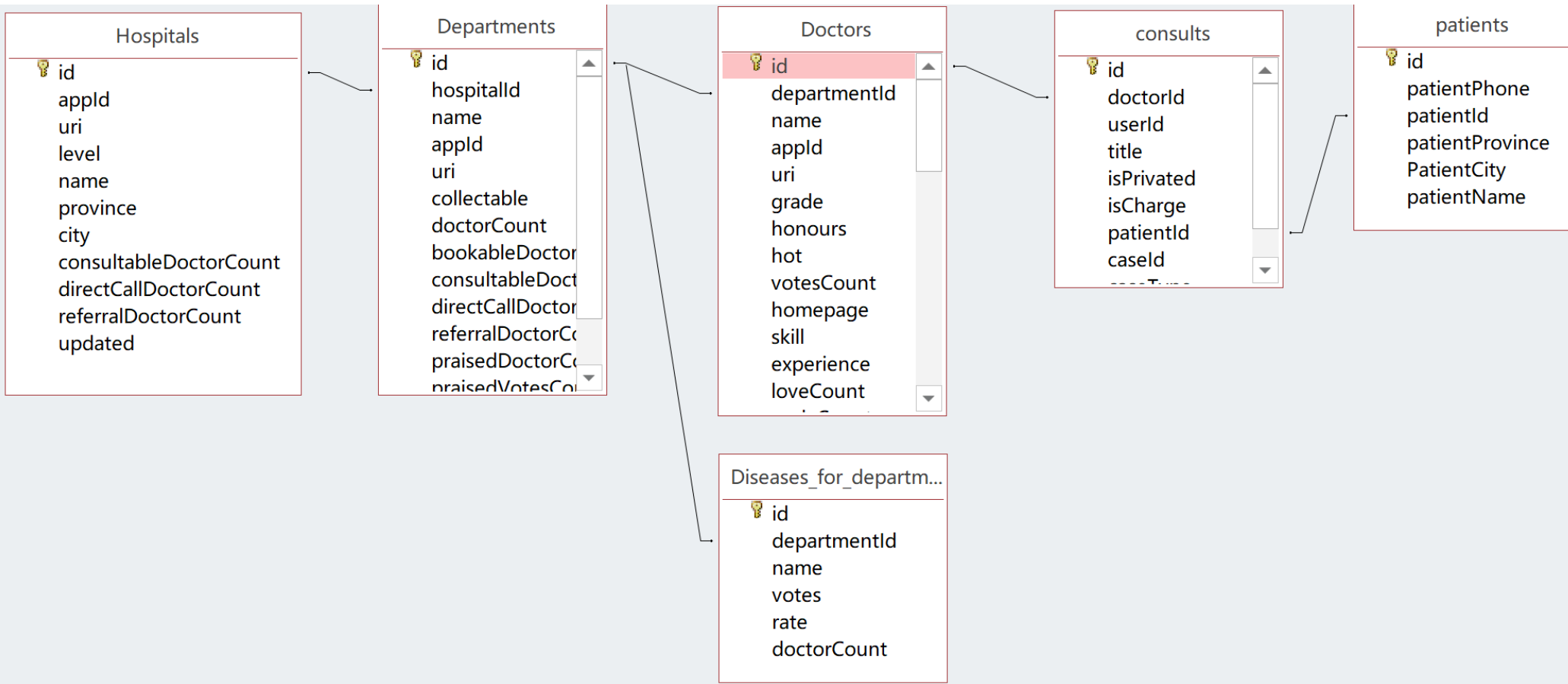
搜索

患者	标题	相关疾病	对话数 (医/患)	最后发表
h***	频发室早患者成功申请图文问诊 <span>已付费</span>	频发室早	13 ( 5/8 )	2017-03-08 by 陈松文
1***	室性早搏，检查：直接住院了 药物：倍他乐... <span>已付费</span>	室性早搏	56 ( 20/36 )	2017-03-07 by 1***
h***	射频消融手术费多少钱	盐城一院诊断房扑	7 ( 3/4 )	2017-03-07 by h***
h***	就是想了解这跟孩子的室速有关系吗	做射频消融术一年多，...	7 ( 3/4 )	2017-03-07 by 陈松文
1***	上海心律失常患者诊后咨询上海第一人民医...	心律失常	12 ( 5/7 )	2017-03-07 by 陈松文
B***	心律失常，房颤，心律失常，房颤；反复心悸...	心律失常，房颤	15 ( 6/9 )	2017-03-06 by 陈松文
h***	最低心率25，用不用装起搏器	房室传导阻滞	2 ( 1/1 )	2017-03-06 by 陈松文
h***	阵发性室上速患者成功申请图文问诊 <span>已付费</span>	阵发性室上速	43 ( 19/24 )	2017-03-05 by h***

# Data Overview — — Database



- We split previous data into several tables based on 2NF (considering the difficulty of query, we did not implement 3NF)



- Messy code
- Redundancy

```
select patients.patientcity, count(1) as counts from
(SELECT departments.id,
 departments.name, doctors.id,
 hospitals.name,
 consults.patientid,
 patients.patientprovince,
 patients.patientcity
FROM ((Departments
INNER JOIN hospitals
ON departments.hospitalid=hospitals.id)
INNER JOIN doctors
ON departments.id=doctors.departmentid)
INNER JOIN consults
ON doctors.id=consults.doctorid)
INNER JOIN patients
ON consults.patientid=patients.patientid
WHERE hospitals.province='上海')group by patients.patientcity
```

patientcity	counts
	408024
-----	1
&#65533;&#1011;&#65533;	1
&#65533;&#65533;&#65533;	4
&#65533;&#65533;&#65533;	1
&#65533;&#65533;&#8000;	1
&#65533; &#65533;	1
?	18
??	300
???	39
	1
14	1
2	6
231	1
5	1
6	2
ü?	42
wuxi	1
阿坝	222
阿克苏	564
阿克苏地区	1
阿拉善盟	146
阿勒泰	180
阿勒泰地区	4
阿里	22
安康	825

# Data Preprocessing



patientcity	counts
	408024
-----	1
&#65533;&#1011;&#65533;	1
&#65533;&#65533;&#65533;&#65533;	4
&#65533;&#65533;&#65533;&#65533;&#65533;	1
&#65533;&#65533;&#65533;&#65533;&#65533;&#65533;	1
&#65533;&#65533;&#65533;&#65533;&#65533;&#65533;&#65533;	1
?	18
??	300
???	39
□ □ □ □ □	1
14	1
2	6
231	1
5	1
6	2
ü?	42
wuxi	1
阿坝	222
阿克苏	564
阿克苏地区	1
阿拉善盟	146
阿勒泰	180
阿勒泰地区	4
阿里	22
安康	825



patientcity	counts
阿坝	222
阿克苏	565
阿拉善盟	146
阿勒泰	184
阿里	22
安康	825
安庆	5522
安庆市	27
安顺	476
安阳	1711
鞍山	1447
鞍山市	2
巴东	1
巴南	8
巴彦淖尔	855
巴音郭楞	631
巴中	722
白城	630
白沙	1
白山	627
白银	583
白云	2
百色	342
蚌埠	3826
包头	1708

Project Introduction

Data Introduction and Preprocessing

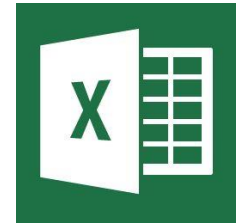
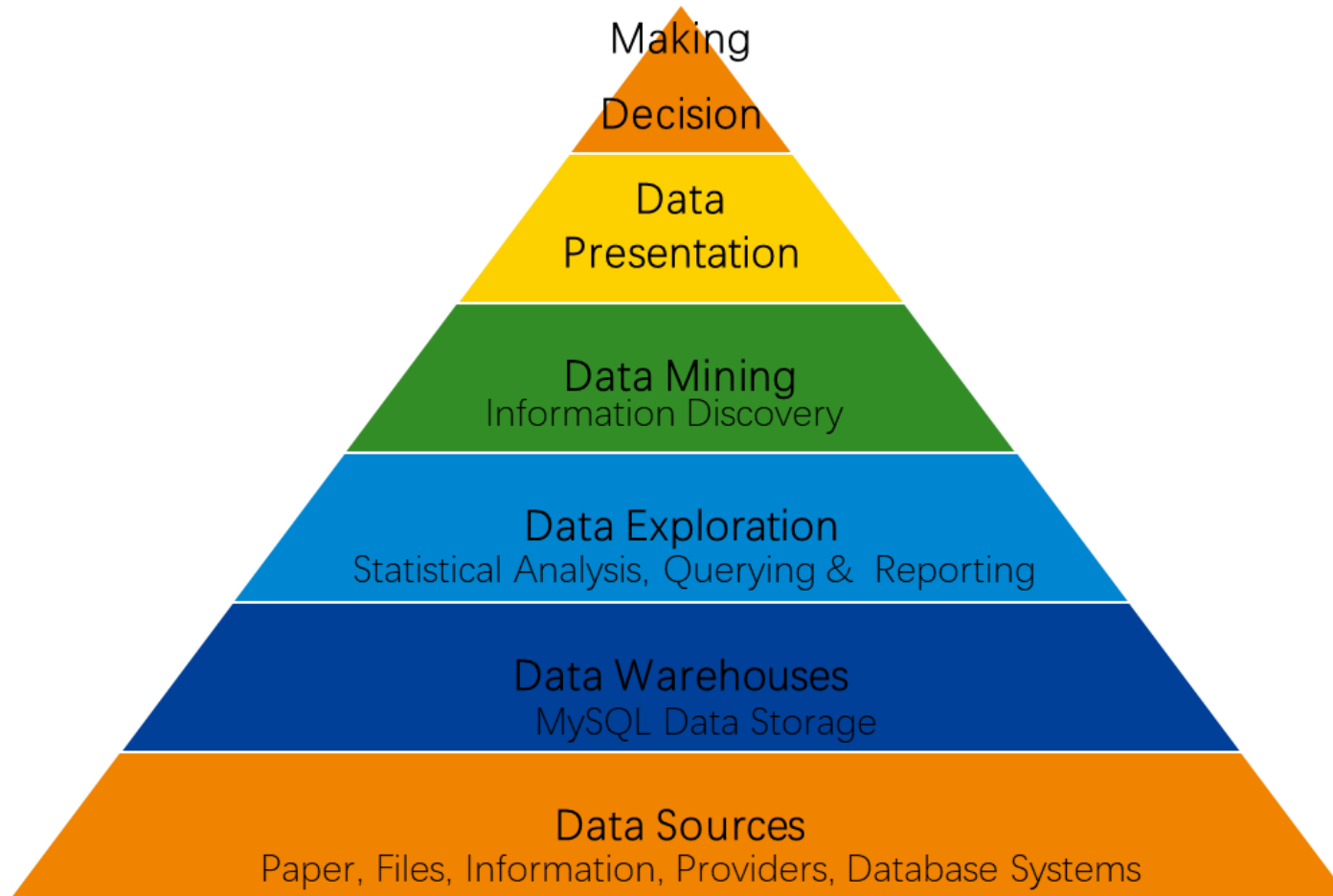
Data Visualization

Data Modeling

Project Summary and Outlook



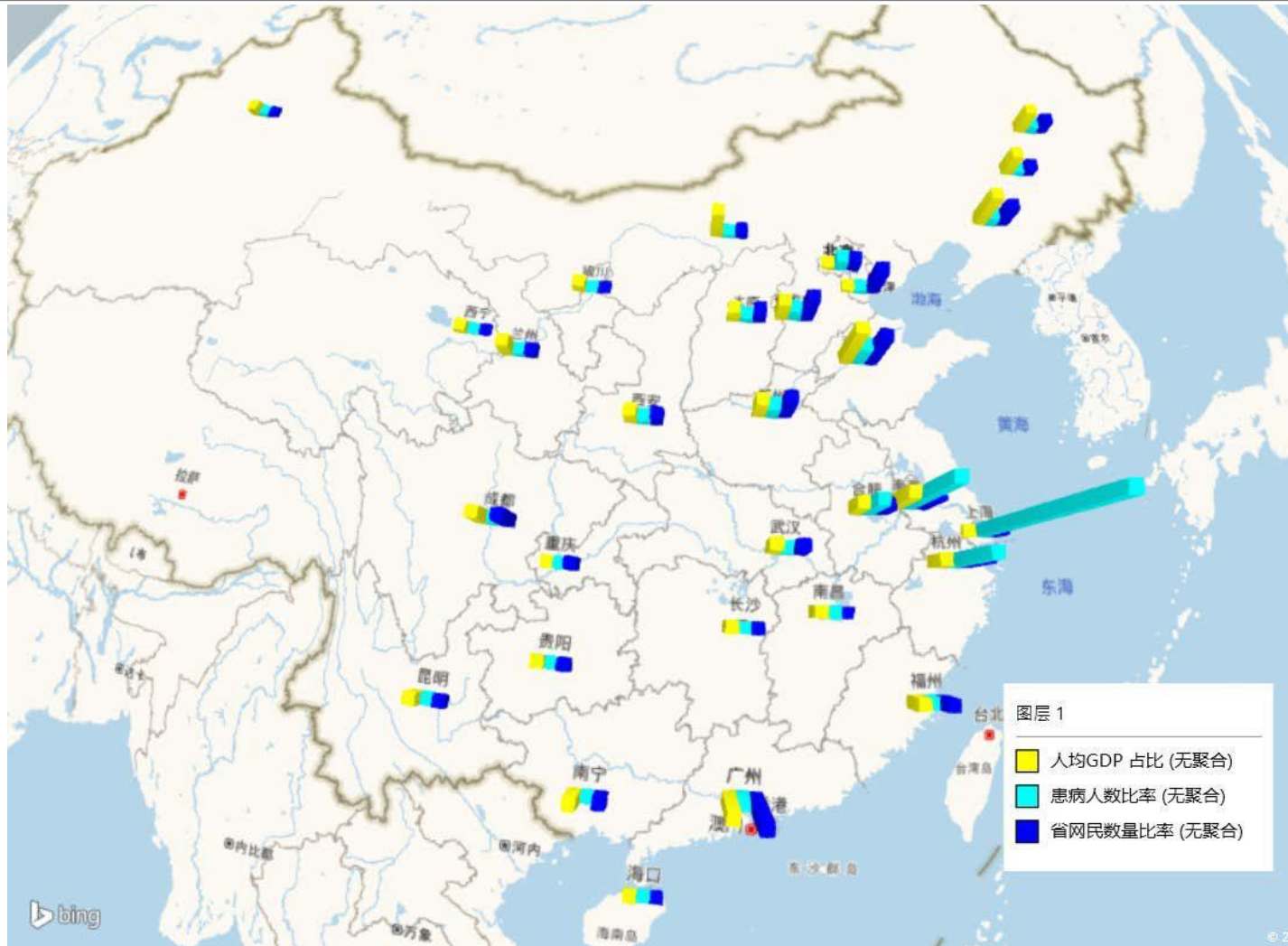
# The role of data visualization in this study



pyecharts

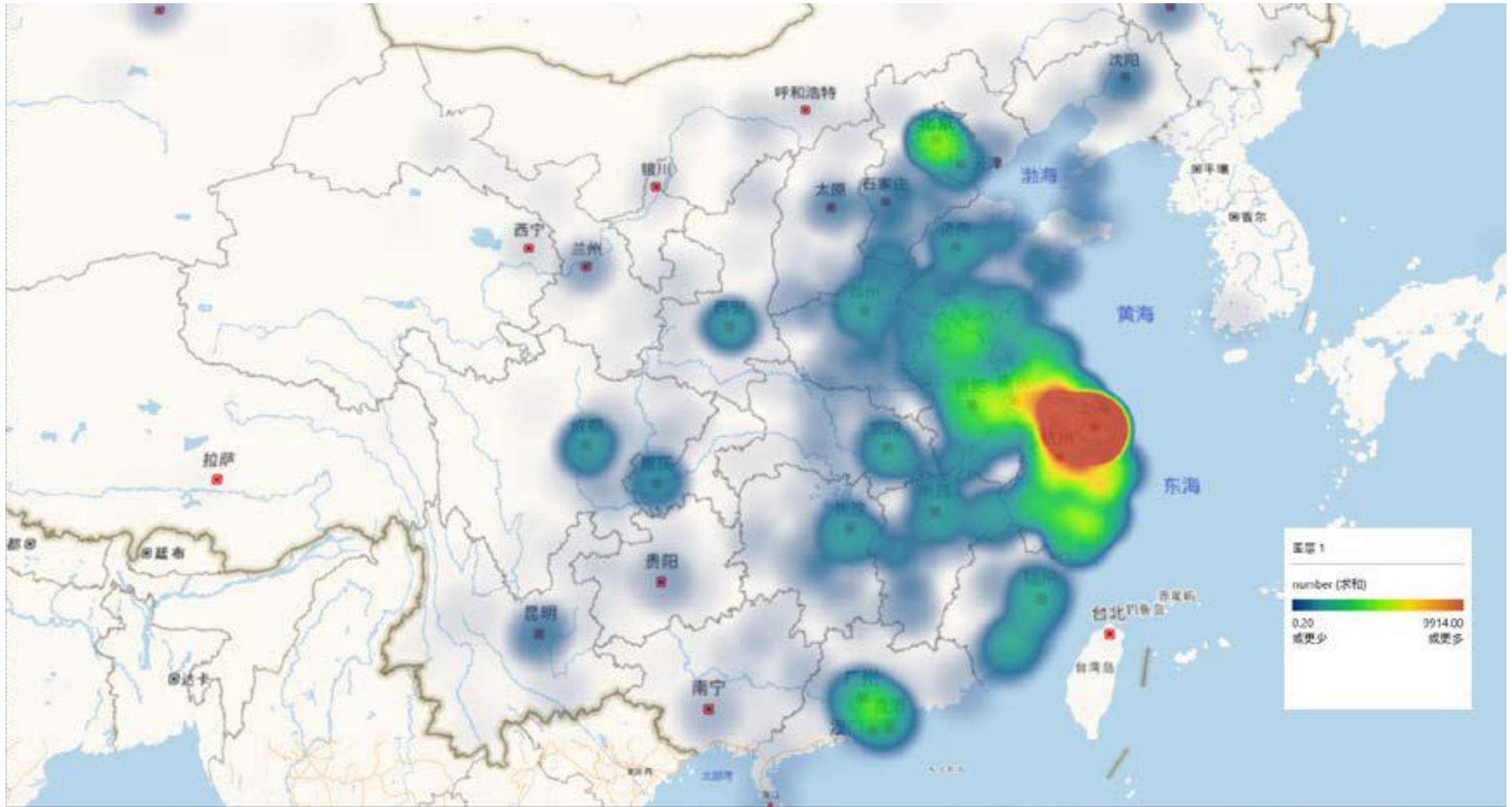
Python ❤ Echarts = pyecharts

# Cluster information graph by Power Map



Combine the per capita GDP, the number of patients, and the number of netizens, and then normalize them. The purpose is to intuitively test our hypothesis that the number of consultants is related to GDP and the number of Internet users

# Thermal map of online consult to Shanghai

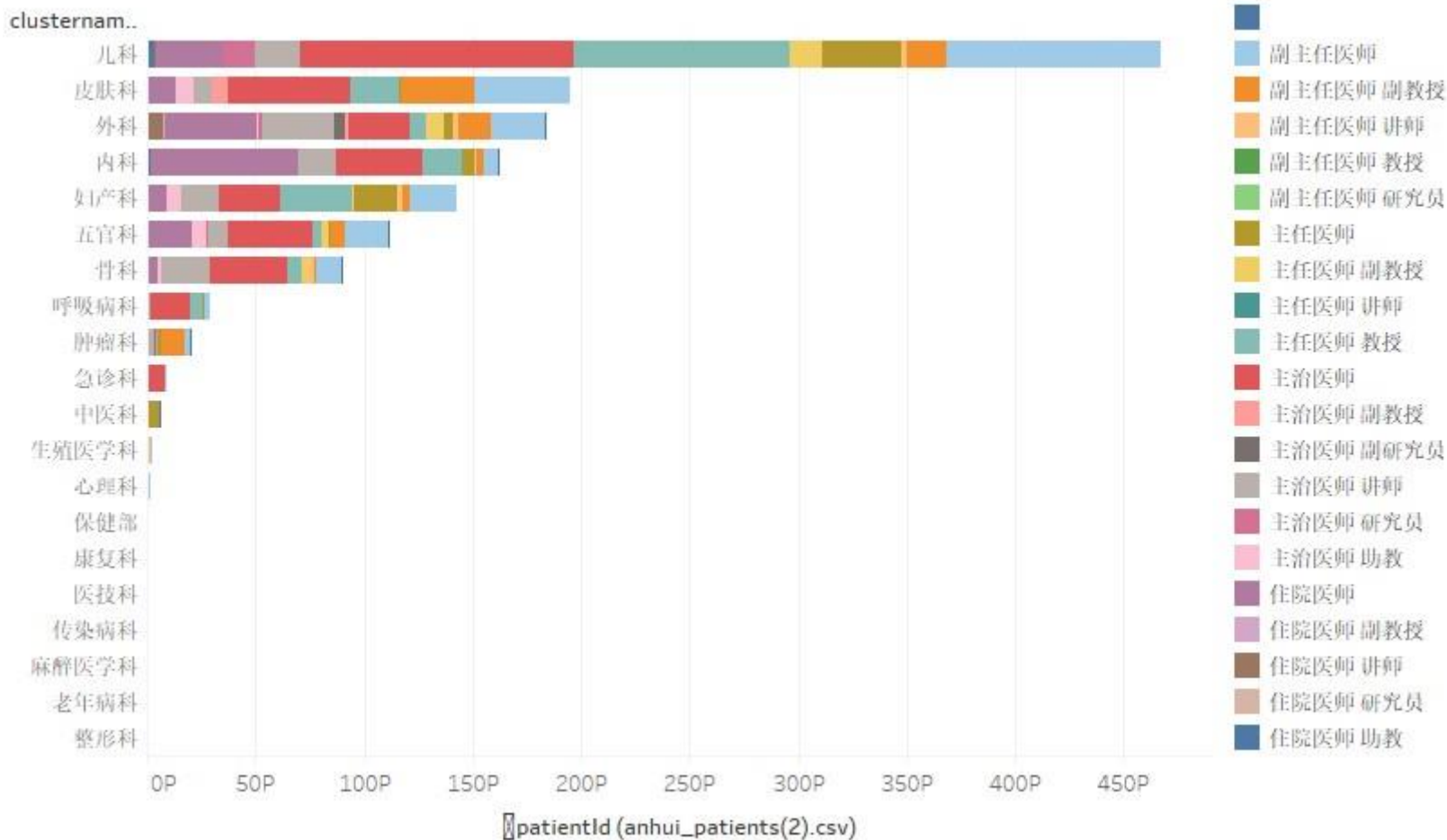


In order to better show the impact of regional differences, we have drawn a heat map for consulting Shanghai in cities across the country.

# Doctor level statistics of Anhui in consult list



安徽咨询人数（按科室）\_省

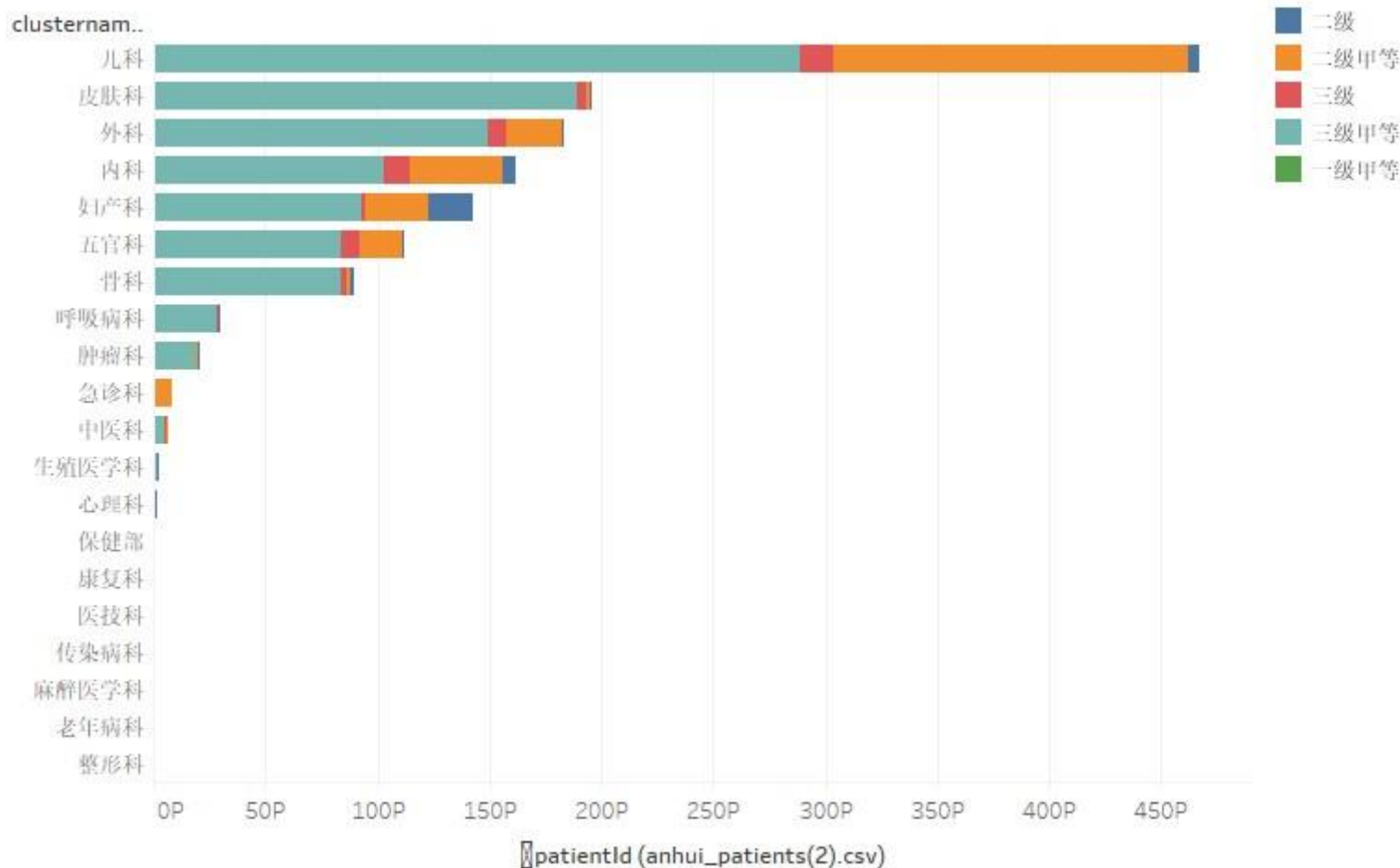


Chief, Associate Chief, Attending Doctor are the doctors most frequently consulted

# Hospital level statistics of Anhui in consult list



安徽咨询人数（按科室）\_省



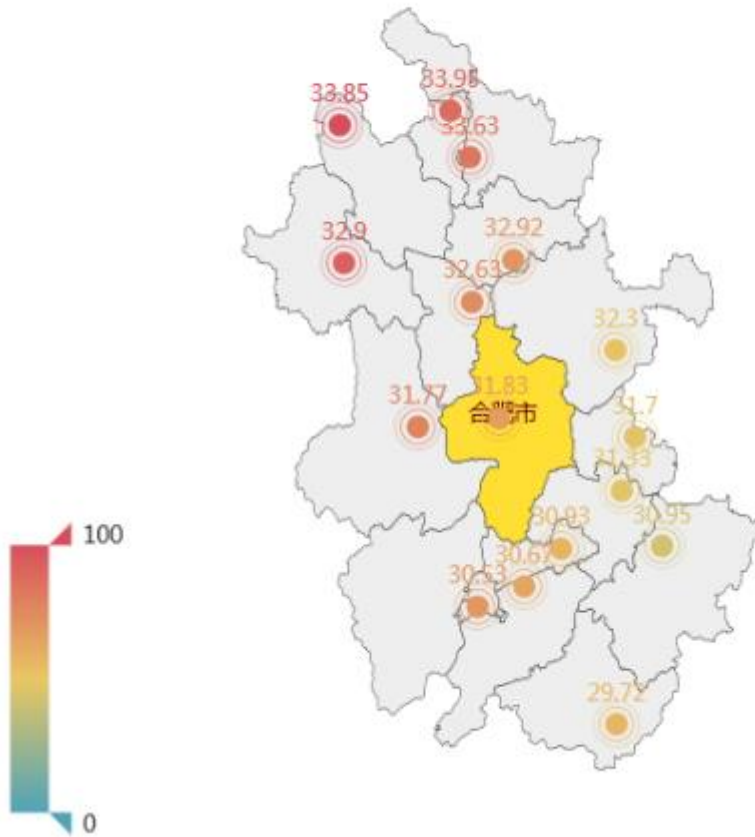
Horizontal comparison: The majority of patients in the same department choose: Grade-A Tertiary Hospital.



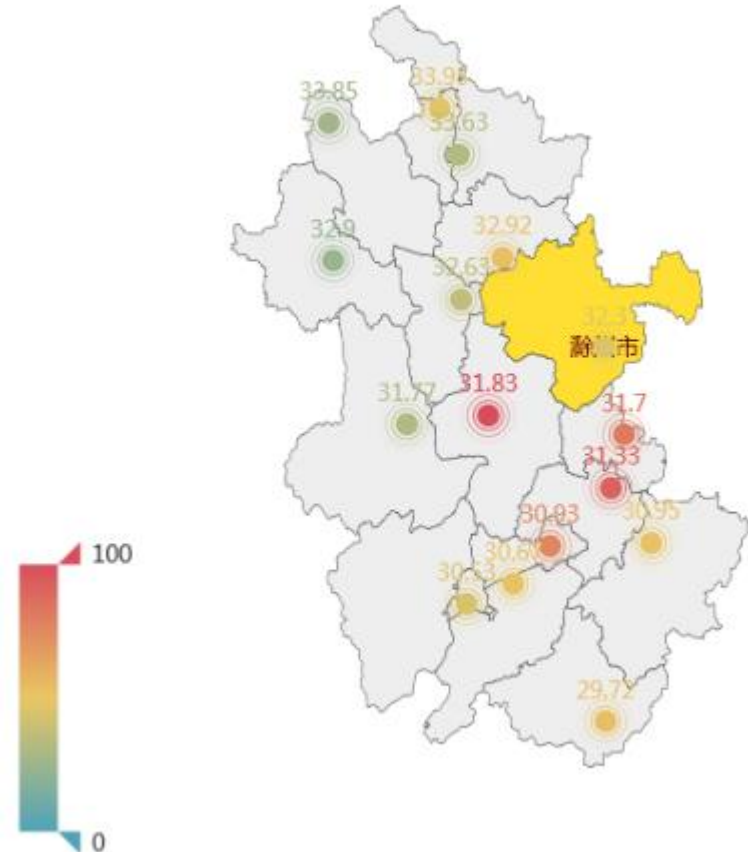
# The region feature of Anhui Province



Distance between Shanghai and cities in Anhui Province



Medical level score of cities in Anhui Province

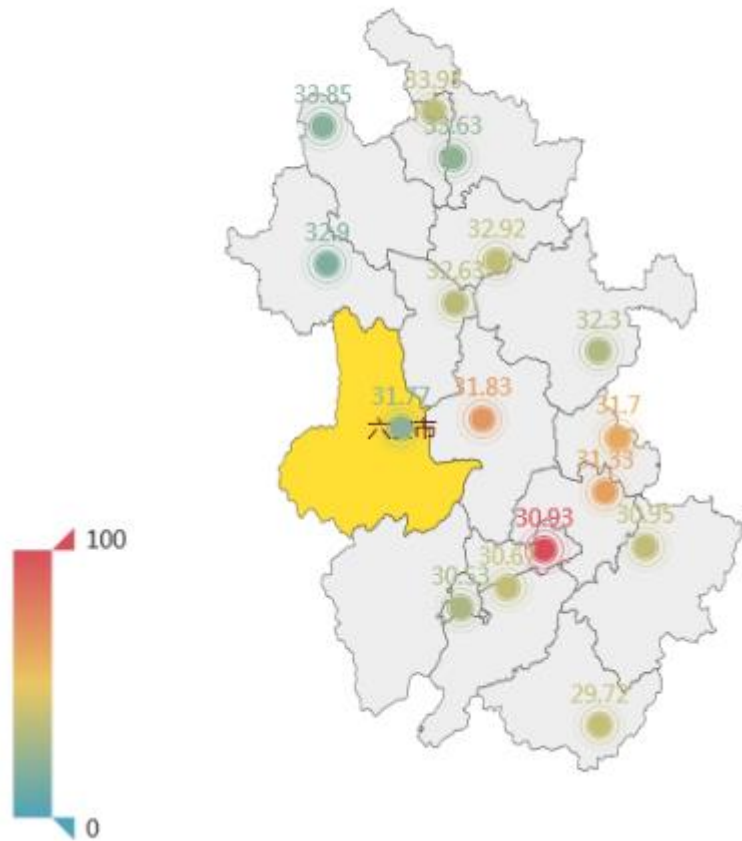


Through data processing and data visualization, we can have a unified understanding of the medical level and GDP of each city in Anhui Province.

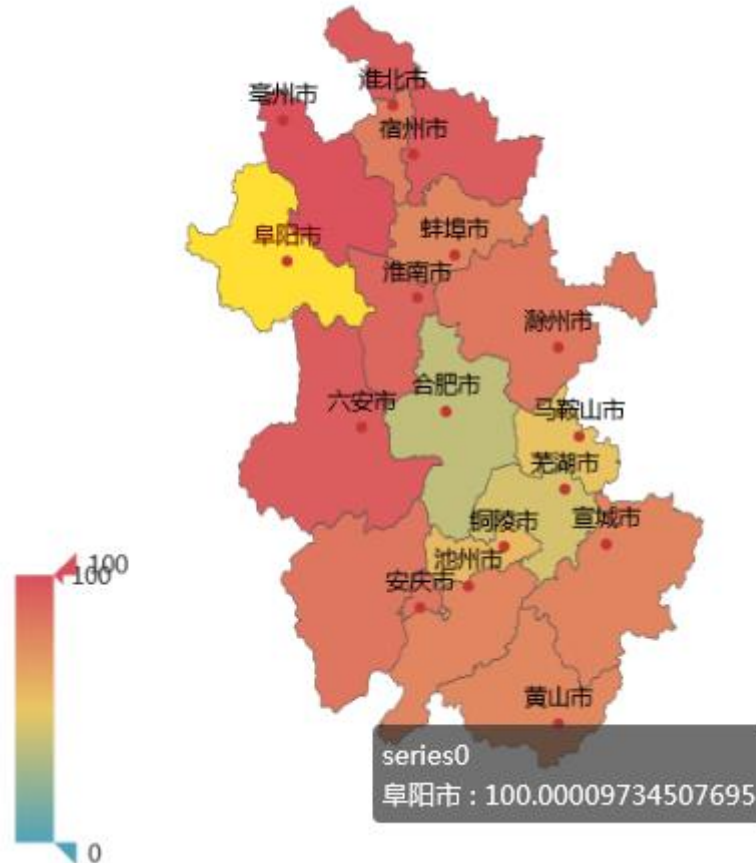
# The medical level gap between Shanghai and Anhui



GDP per capita of cities in Anhui Province



Difference of medical level between Shanghai and cities in Anhui



Those information could help us in the model construction.



Project Introduction

Data Introduction and Preprocessing

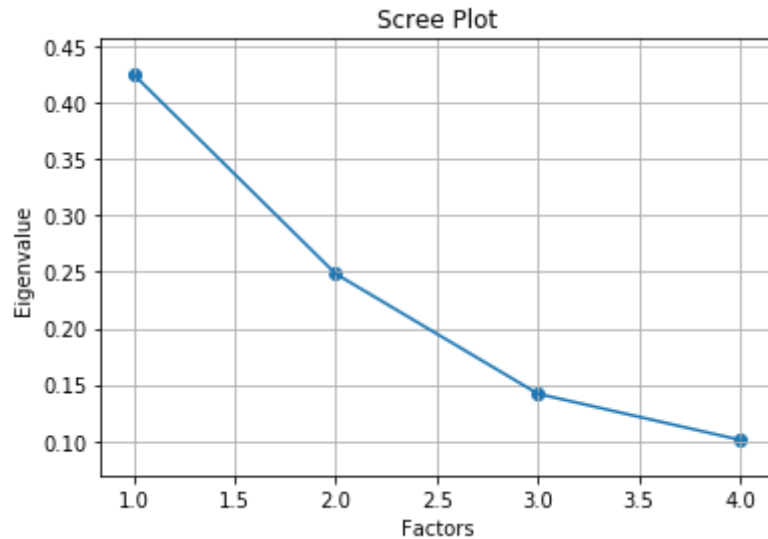
Data Visualization

Data Modeling

Project Summary and Outlook

- ☒ Construct Medical level indicator using PCA
- ☐ Logistic regression

- Method: Principal Component Analysis
- Tool: Python
- Features(indicators):
  - The elderly dependency ratio
  - Number of medical institutions
  - Number of hospitals
  - Number of medical beds per 10000 people
  - Number of medical personnel per 10000 people
  - Number of practicing physician per 10000 people
  - Average hospitalization day
  - GDP



	Components_1	Components_2	Components_3
Feature			
elderly dependency ratio	-0.421	-0.129	0.799
Number of medical institutions	0.520	0.582	0.499
Number of hospitals	0.349	0.758	-0.276
Number of medical beds	0.861	-0.405	0.054
Number of medical personnel	0.939	-0.207	0.125
Number of practicing physician	0.936	-0.238	0.157
Average hospitalization day	0.750	-0.426	-0.238
GDP	0.787	0.559	0.062

Three components can explain the 81.5% of the total variance

- The first component reflects the economic and medical resource level
- The second principal component reflects the degree of medical competition
- The third principal component reflects population and aging factors

Construct the Medical level using three components, the weight of each components can be calculated by following formula

$$Weight_i = \frac{Var(x_i)}{\sum_{i=1}^3 Var(x_i)} = \frac{\lambda_i}{\sum_{i=1}^3 \lambda_i}$$

	City	elderly dependency ratio	Number of medical institutions	Number of hospitals	Number of medical beds	Number of medical personnel	Number of practicing physician	Average hospitalization day	GDP	Medical level
0	Hefei	15.83	2217	146	56	61	23	11	5660.27	13276.51751
1	HuaiBei	13.51	718	73	55	50	19	10	760.39	6328.77547
2	Haozhou	14.60	1705	55	33	30	11	7	942.61	3576.32820
3	Suzhou	17.24	1880	73	35	39	15	8	1235.83	4248.65928
4	Bangfu	16.74	1413	83	54	53	19	10	1253.05	6995.21678
5	Fuyang	16.78	2644	97	40	38	14	8	1267.45	3242.62066
6	Huainan	17.39	1211	67	40	41	15	11	901.08	4861.05483
7	Chuzhou	15.85	1655	61	40	39	15	9	1305.70	6026.79145
8	Liuan	15.45	2701	37	43	44	19	8	1016.49	4209.93965
9	Maanshan	20.35	1005	55	37	50	19	9	1365.30	10923.09195
10	Wuhu	16.40	1436	78	51	54	21	9	2457.32	12188.90280
11	Xuancheng	19.02	1330	40	43	49	19	8	971.46	6864.70094
12	Tongling	17.43	306	18	38	37	14	14	911.60	10218.18171
13	Chizhou	15.54	982	32	44	49	20	9	544.74	6887.73138
14	Anqing	16.64	2497	72	43	49	19	9	1417.43	5877.07508
15	Huangshan	20.90	1153	31	49	62	23	10	530.90	7053.73530

- Construct Medical level indicator using PCA
- Logistic regression

- We add the weighted value of each component to form a new variable called Medical\_Level It means the comprehensive score of the city's medical level.
- Combine into the patient's data.
- However, every patient now has the same city attributes

地区	到上海的距离 (公里)	人均地区生产总值 (全市) /元	医疗水平综合得分	与上海医疗水平的差异
合肥市	408.1	67689	13276.51751	5827.21859
淮北市	532.5	35324	6328.77547	12774.96063
亳州市	607.8	17769	3576.3282	15527.4079
宿州市	503.9	20895	4248.65928	14855.07682
蚌埠市	429.7	35542	6995.21678	12108.51932
阜阳市	563.5	15303	3242.62066	15861.11544
淮南市	451.5	33361	4861.05483	14242.68127
滁州市	315.5	30562	6026.79145	13076.94465
六安市	472.4	19211	4209.93965	14893.79645
马鞍山市	286.1	60091	10923.09195	8180.64415
芜湖市	289.6	64039	12188.9028	6914.8333
宣城市	259.3	35726	6864.70094	12239.03516
铜陵市	350.9	97193	10218.18171	8885.55439
池州市	387.0	36267	6887.73138	12216.00472
安庆市	424.6	28808	5877.07508	13226.66102
黄山市	346.1	37306	7053.7353	12050.0008



- Therefore, we investigate the other information and invent a new variable called Utility Score.
- It measures the urgency or utility of a patient.
- More words to describe his illness, more phone calls, higher score
- A comprehensive score of Length of Description, Numbers of Phone Calls, Total Time of Consulting.

	patientId	score	city	distance	gdp	medical_level	medical_dif	patientId	prov	doctorId	name	doctorgrade
0	1013185098	8	阜阳	563.5	15303.0	3242.62066	15861.115440000000	1013185098	安徽	21587	杜永强	主任医师 教授
1	1013185098	8	阜阳	563.5	15303.0	3242.62066	15861.115440000000	1013185098	安徽	21587	杜永强	主任医师 教授
2	1013185098	8	阜阳	563.5	15303.0	3242.62066	15861.115440000000	1013185098	安徽	21587	杜永强	主任医师 教授
3	1013185098	8	阜阳	563.5	15303.0	3242.62066	15861.115440000000	1013185098	安徽	21587	杜永强	主任医师 教授
4	1013185098	8	阜阳	563.5	15303.0	3242.62066	15861.115440000000	1013185098	安徽	21587	杜永强	主任医师 教授
5	1013185098	8	阜阳	563.5	15303.0	3242.62066	15861.115440000000	1013185098	安徽	21587	杜永强	主任医师 教授
6	2255034500	5	亳州	607.8	17769.0	3576.3282	15527.4079	2255034500	安徽	16464	江山	副主任医师
7	2255034500	5	亳州	607.8	17769.0	3576.3282	15527.4079	2255034500	安徽	16464	江山	副主任医师
8	2255034500	5	亳州	607.8	17769.0	3576.3282	15527.4079	2255034500	安徽	16464	江山	副主任医师
9	2255034500	5	亳州	607.8	17769.0	3576.3282	15527.4079	2255034500	安徽	16464	江山	副主任医师
10	2255034500	5	亳州	607.8	17769.0	3576.3282	15527.4079	2255034500	安徽	16464	江山	副主任医师
11	2255034500	5	亳州	607.8	17769.0	3576.3282	15527.4079	2255034500	安徽	16464	江山	副主任医师
12	2255034500	5	亳州	607.8	17769.0	3576.3282	15527.4079	2255034500	安徽	16938	项平	主治医师
13	2255034500	5	亳州	607.8	17769.0	3576.3282	15527.4079	2255034500	安徽	16938	项平	主治医师
14	2255034500	5	亳州	607.8	17769.0	3576.3282	15527.4079	2255034500	安徽	16938	项平	主治医师
15	2255034500	5	亳州	607.8	17769.0	3576.3282	15527.4079	2255034500	安徽	16938	项平	主治医师
16	2255034500	5	亳州	607.8	17769.0	3576.3282	15527.4079	2255034500	安徽	16938	项平	主治医师
17	2255034500	5	亳州	607.8	17769.0	3576.3282	15527.4079	2255034500	安徽	16938	项平	主治医师
18	2255034500	5	亳州	607.8	17769.0	3576.3282	15527.4079	2255034500	安徽	24671	王静	主治医师

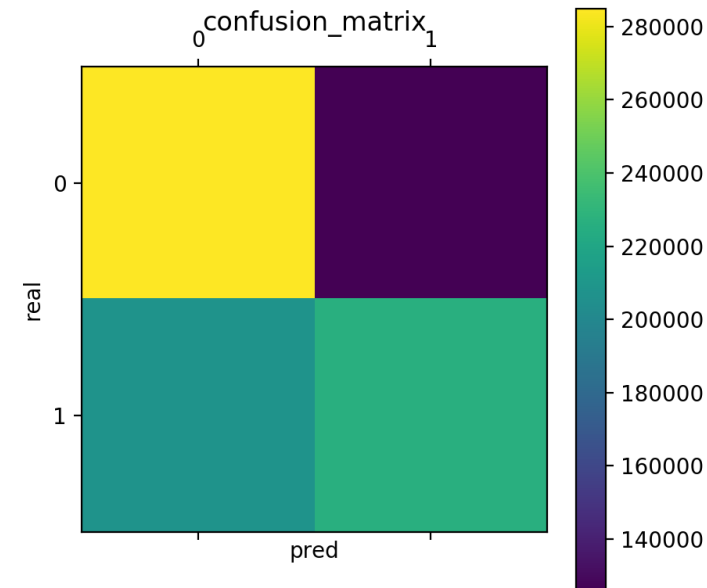
- Then we use Logistic Regression to process our data.
- The result shows that the accuracy, precision and recall are all about 60%, and the coefficients lead to our final results.
- The coefficients: (1) Distance < 0 (2) GDP > 0 (3) Medical Level < 0 (4) Utility Score > 0

```
precision    recall  f1-score   support

0           0.58        0.69        0.63       410909
1           0.64        0.52        0.58       433791

micro avg    0.61        0.61        0.61      844700
macro avg    0.61        0.61        0.60      844700
weighted avg 0.61        0.61        0.60      844700

[[-1.29484253e-03  1.76739044e-05 -1.20214732e-04  7.61322211e-02]]
('accuracy_score ', 0.6051177932993962)
('precision_score ', 0.6422969903463941)
('recall_score ', 0.5214884587278206)
[Finished in 33.0s]
```



Project Introduction

Data Introduction and Preprocessing

Data Visualization

Data Modeling

Project Summary and Outlook

## Whether Internet medical platform is conducive to the allocation of medical resources ?

- **Yes!**
- Taking Shanghai as an example, it can be seen from the visualization results that the consulting patients are all over the country, and people all over the country can enjoy the high-quality medical resources of Shanghai (**especially Grade-A tertiary hospital**) to a certain extent.
- Regional factors and willingness to consult affect patients' decision-making. **(1) The closer the distance is, (2) the higher the GDP is, and (3) the poorer the medical level is, the more patients choose Shanghai, and (4) the stronger their willingness to consult, the more they choose Shanghai.**

## What else can we do in the future ?

### DATA



- The data can be combined with the patient's personal information (family income, duration of illness, age) to provide a more detailed analysis of the factors influencing the patient's decision

### MODEL



- Time variables can be added into the model to study the changes of patients' decision-making in different periods.

1. 王霜奉.国外掀起移动医疗“风潮” [J] .上海信息化, 2014 (3) : 84.
2. 曹常义.移动医疗商业路径选择及移动运营商践行经验探讨 [J] .移动通信, 2014 (19) : 48-54.
3. 梁怡凡.移动医疗APP的盈利模式探索 [J] .中国市场, 2015 (25) : 137-140.
4. 唐超.重构中国互联网医疗模式 [J] .中国医院院长, 2015 (15) : 66-72.
5. 张浩辰.互联网与中国医疗模式变革研究 [J] .中国物价, 2016 (4) : 62-65.
6. 朱莉华, 曹乾, & 王健. (2009). 居民健康与卫生保健及医疗服务的可及性关系——基于chns 2006年数据的实证研究. 经济研究导刊(13), 211-213.
7. 宋正娜, 陈雯, 张桂香, & 张蕾. (2010). 公共服务设施空间可达性及其度量方法. 地理科学进展(10), 69-76.
8. Murray A T. Site placement uncertainty in location analysis [J] . Computers, Environment & Urban Systems, 2003, 27( 2) : 205—221.
9. Luo W, Wang F. Measures of spatial accessibility to health care in a GIS environment: synthesis and a case study in the Chicago region [J] . Environment & Planning B, 2003, 30( 6) : 865—884.
10. Tsou K W, Hung Y T, Chang Y L. An accessibility—based integrated measure of relative spatial equity in urban public facilities [J] . Cities, 2005, 22( 6) : 424—435.
11. Naess P. Accessibility, activity participation and location of activities: exploring the links between residential location and travel behavior [J] . Urban Studies, 2006, 43( 3) : 627—652.