# 需求

根据小组成员采用爬虫获得的数据，获得了广大网民在网络求医过程中，人员年龄，性别，咨询科室，访问时间等一系列信息。但是我们更关注，在这个过程中，网络求医这种新型医疗保健方式所隐藏的背后的规律。究竟是什么样的一群人带动了这个现象，他们以怎样的方式参与其中？这些都是我们所关心并且希望了解的。

# 实现思路

通过SQL和Excel对原始数据进行初步清理，并导入数据库中，采用SQL语言进行相关信息的提取

# 代码结构及部分结果

## （1）网络求医时间分布

代码结构：

#用HOUR()函数将取时间的小时，按小时分组判断出不同时间段网络求医人数

select

HOUR(d\_time) as 看病时间 ,count(d\_name) as 相同时间看病人数量

from d\_statistics where d\_name !=''

group by 看病时间

查询结果：网络求医时间分布.txt

0 6

1 16

2 34

3 32

4 39

5 80

6 79

7 176

8 236

9 153

10 16

11 7

12 18

13 21

14 32

15 21

16 29

17 65

18 105

19 183

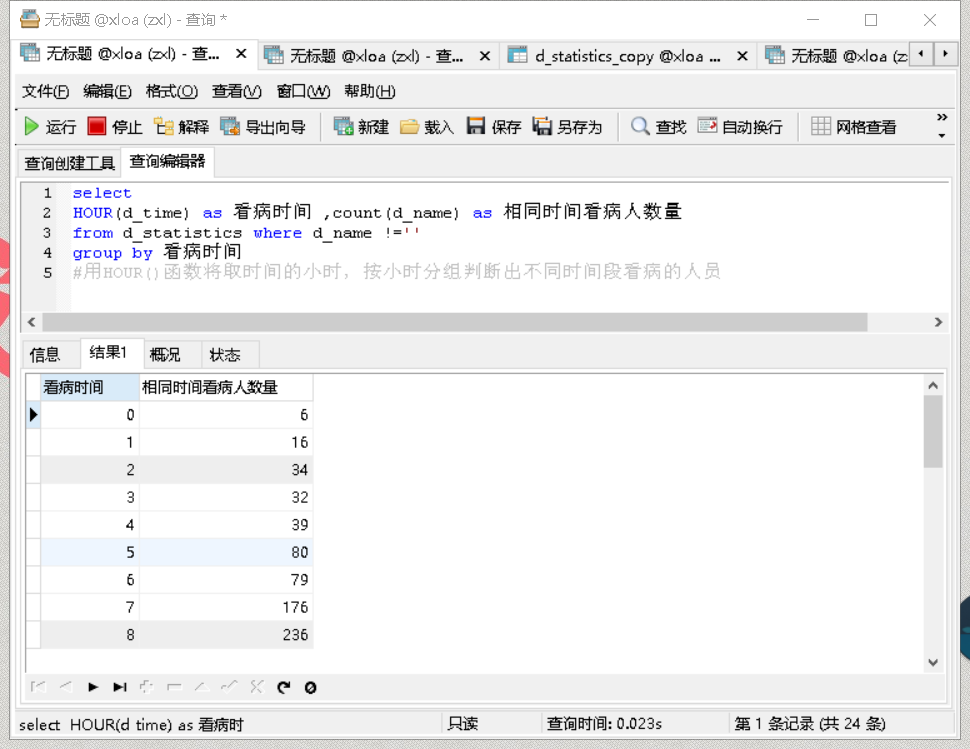
20 600

21 463

22 276

23 18501

注：其中第一列为时间，第二列为对应人数



## （2）网络求医人员年龄分析

代码结构：

#用LENGTH()筛选出，正确的年龄看病记录，分组查看不同病情，不同性别的人员

select d\_age as 年龄,count(d\_age) as 数量

from d\_statistics\_copy where

d\_name !='' and LENGTH(d\_age) <10 and LENGTH(d\_age)>1 and d\_age!=' 0个月'

GROUP BY d\_age

查询结果：网络求医年龄分布.txt

1 209

2 76

3 63

4 59

5 64

6 58

7 50

8 33

9 28

10 29

11 36

12 38

13 45

14 61

15 84

16 139

17 178

18 255

19 219

20 491

21 360

22 414

23 464

24 418

25 501

26 437

27 313

28 327

29 248

30 297

31 136

32 216

33 127

34 118

35 145

36 130

37 74

38 78

39 65

40 113

41 52

42 70

43 71

44 51

45 98

46 68

47 47

48 71

49 45

50 70

51 23

52 58

53 47

54 25

55 37

56 52

57 27

58 31

59 20

60 46

61 22

62 19

63 13

64 18

65 33

66 17

67 14

68 13

69 9

70 31

71 6

72 16

73 13

74 6

75 11

76 9

77 4

78 10

79 3

80 6

81 3

82 2

83 2

84 3

85 3

86 1

87 1

89 1

90 2

91 2

92 1

95 1

106 1

注：第一列为年龄，第二列为人数

## （3）网络求医人员性别分布

代码结构：

select d\_six as 性别,count(d\_six) as 数量

from d\_statistics\_copy

GROUP BY d\_six

查询结果：网络求医性别分布.txt

10934

女 6855

男 3566

注：空白项为求医者未提供性别信息

