

项目问题1提问回答汇总

2018年5月25日星期五

14:06

注：黑色字体为学员提问，红色字体为助教回答。问题

背景：

问题 1：至少写下两个你感兴趣的问题，请确保这些问题能够由现有的数据进行回答。

答案：

1.2012年-2015年成都市PM 2.5的数据有什么变化趋势

2.PM 2.5的含量与季节是否有关系

如同题目所说，同学你的提问一定是数据集可以回答的才有意义，这样后面的分析步骤才可以围绕自己提出的问题展开，然后希望没有编程基础的同学控制变量在两个到3个之间，过多的变量会导致后面可视化代码非常困难（当然学有余力的同学是建议多多拓展，研究多个变量之间的关系），比如问题：‘2012年-2015年成都市PM 2.5的数据有什么变化趋势’就是研究成都（固定研究城市，消除了城市变量）从2012-2015每年的年平均（自变量）PM2.5（因变量）的变化趋势。同学们提问题的时候可以自己分析一下哦。

学员1：

问题一就是写两个问题吗

怎么发给你啊

助教：

对哒

微信这里发给我就可以啦

研究的问题尽量具体，后面的步骤才好实施

1. 2010-2015年广州的PM2.5浓度与湿度的关系？

2. 2013年广州降水量对PM2.5浓度的影响。

因为年平均会多一各变量

总平均年份的变量就消失了

那第一个问题出两个可视化的图哦，图1广州每年pm2.5变化趋势，图2，上海

要是吧两图放在一起也行，但是项目中的可视化代码就不能套用了，需要改代码，所以还是分两个图会轻松一点

好的，这个提出的问题是我不需要之后去自行解决的是吗？

恩恩这就是你后面研究的问题

所以尽量先从简单的来

好，那我还是先研究一年的好了...

学员2：

1. 2012年-2015年成都市PM 2.5的数据在不同的月份有什么变化趋势？2. 哪个城市的PM2.5的含量较高？

我的问题就是这两个

哇超棒

基本不用做什么修改

就是第二个问题吧年份限定一下就更好了，比如年平均

学员3：

问题 1：至少写下两个你感兴趣的问题，请确保这些问题能够由现有的数据进行回答

（问题示例：1. 2012年-2015年上海市PM 2.5的数据在不同的月份有什么变化趋势？

答案：

第一个问题：PM2.5的数据在一天内有什么变化趋势？

第二个问题：PM2.5的数据与温度、湿度、风速和降雨量是否存在相关性？

看到xx的问题啦~

很棒~就是有点小问题

第一个是想研究哪个城市的PM2.5数据呢

要是能把城市写出来就更好了

这样啊，那就选北京吧

第二个问题想研究pm和哪个因素的相关性这很好，但是有一个问题就是东涛这边有5个因素，你就要出5张可视化的图（4张散点图+一张热力图）哦

xx这边没有编程基础，我担心代码量有些大

哦，这样啊，其实散点图改一下横坐标就一下子出来啦

优达项目这边的建议是控制一个变量（温度，湿度，风速，降雨量中的一个）与某个城市的PM2.5数据是否有相关性

嗯嗯道理是这样的~

那第二个我先做湿度的相关性分析吧

记得控制一下时间哦，比如北京2014年的PM2.5在一天内有什么变化趋势？

年份也是需要控制哒

学员4：

早哇，xx今天有空看一下项目哦~

有机会吧项目的问题1的答案给到我

有几个原则哈，尽量保证变量少哦，因为这个数据分析最好是由简到繁，一开始变量比较多后面可视化可能不能套用它给的栗子，可视化会比较复杂，其次记得给出年分和城市哈

1.PM2.5浓度与温度的关系（通过不同城市的比较，分析是否线性相关）2.2010-2015年 上海的PM2.5浓度变化（是否随季节呈周期性变化，同期相比是否有所变化）

是这样吗。。。

第一个问题城市最好确定为一个，然后年份也定下

第二个最好是上海的年平均pm2.5与季节是否有相关性

这样研究起来方向比较明确

我的理解是 pm2.5是否和温度有线性关系 应该是多个城市进行比较验证才能得出结论？

一个城市应该没有办法得出结论？存在偶然性？可能有其他因素影响？

理论上是需要研究多个城市，这样才接近真实

因为我们问题研究的就是一个城市，所以说问题中我们并没有把结论这个推广到所有城市-是否能适用于所有城市这样子。如果你的问题是研究所有的城市的话，你就要多研究几个城市，但是你如果你城市也多，你就会发现你做可视化需要出很多的图，这个在以后我们学完入门正式课程了以后是没问题的，但是因为这个项目比较简单，我们想要套用后面的可视化代码，尽量把变动控制在一个到两个之间，然后我们研究研究起来就会比较方便。

好滴好滴理解 我改一下 马上发你

嗯嗯对哒，越学到后面会越来越严谨

其实主要是怕打击大家在项目0的积极性

学员5:

2012年和2015年相比，pm2.5的数据有什么异同点？

上海市的pm2.5值与月份有相关吗？

这两个问题可以吗

第一个问题最好加上那个城市

因为你的变量是年份了，所以城市最好用一个

第二个加上哪年的上海市数据就完美啦

其他的没有，问题非常棒，小改一下就ok了

学员6：

问题一：2010年-2015年北京市和上海市的PM2.5浓度与湿度的关系？

问题二：2010年-2015年广州市和沈阳市的PM2.5浓度与温度的关系？

嗯呢很棒，控制了年份，城市和另一个变量

就是变量稍微有点多，后面可视化出图的时候可能代码量有点大

我建议第一个问题研究北京伤害pm2.5之间的差异

嗯嗯，我是感觉这两个问题其实有点重复

第二个问题取一个城市研究pm2.5与温度的关系

嗯嗯这样变量控制在两个就好了

还要考虑一下年分定在2010-2015汉章是想研究每年的年平均还是总得平均

因为年平均会多一个变量

总平均年份的变量就消失了

每年的吧，没想到年平均这个

好的，那第一个问题出两个可视化的图哦，图1北京每年pm2.5变化趋势，图2，上海

要是吧两图放在一起也行，但是项目中的可视化代码就不能套用了，需要改代码，所以还是分两个图会轻松一点

学员7：

问题一：北京和成都PM2.5在不同月份变化趋势对比是怎么样的。

问题二：成都PM2.5一天之内的变化趋势在不同季节中是怎么样的。

超棒，但是有个小问题，xx这个问题二稍微有点复杂哦，你可以handle嘛

因为既要研究每天的pm变化，还要对比不同月份～出图的话可能要出春夏秋冬4个图哦

如果想要吧四个图合在一个图中，用不同图例表示

这样子可视化的代码会有点复杂

那如果只看一个季节比如夏季会不会稍微简单些？

或者某一个月份？

我：

嗯控制了季节的变量会简单点

你可以控制两月份做一个对比

最冷的和最热的怎么样

这样对比下来也许会有想不到的结果

好那就1月 和8月

学员8：

第一个问题：2010-2015年，不同季节对五个城市的PM2.5值各自影响如何 第二个问题：2010-2015年，各个城市的空气质量呈现怎样的变化趋势

问题很棒呢，就是可能需要稍微做点调整，我们推荐是变量控制在两个以内，不然后面可视化的代码会比较多，比如第一个问题，研究不同的季节是个好想法，城市控制一个就好了，不然光第一个问题就要出5张可视化的图哦，第二个问题不用修改了，就是把空气质量改成PM2.5如何变化可能会严谨一点。

学员9：

1.北京在2010-2015年间有最好的空气质量是在那个地区？

2.从空气质量的变化趋势，选择那个城市能享受最好的空气？

区域这个问题我们研究不了哦

因为我们只能用到PM_US Post这一个数据，其他测绘点的数据因为缺失量太大

不适合做研究哦

对，我看了也是缺失量大。。我再改改第一个

嗯嗯所以最好从城市来对比，然后空气质量最好表达成pm2.5哦，因为数据集只提供了PM2.5，而空气质量包含很多

1.2015年间各个地区空气质量最好的时间段都是什么时候？会有一个时间是PM2.5最低的时候吗

2.从PM2.5的变化趋势，选择那个城市能享受最好的空气？

各个地区要是能换成一个固定城市就好了，比如，2015年间北京空气质量最好的时间段都是什么时候？会有一个时间是PM2.5最低的时候吗

不能做成多条线吗

当然可以哒，但是优达建议p0项目的变量控制少一点，要是xx学有余力变量多也没问题的哦

可视化的代码可能不能套用项目提供的模板，可视化的代码就需要重新编写

1.2015年间广州地区空气质量最好的时间段都是什么时候？会有一个时间是PM2.5最低的时候吗

2.从PM2.5的变化趋势，选择那个城市能享受最好的空气？

第二个问题要是解释想从哪个时间入手呢？

是每年年均还是固定研究一年年均，然后5个城市比较？

一年年均的五个城市比较

ok，那在第二个问题前面加上哪一年问题可能看起来更清晰哦

多谢耐心指导

我知道大神可以编更难代码~

233333不过数据分析还是由浅入深的好一点

不是不是，只是想的时候没想太多

学员10：

各城市PM2.5有效数据时间内，数值变化曲线与其它一项或几项数据变化曲线有无明显相关性。

五大城市PM2.5平均值大小是否与数据集内某个或某几个特征值有关联性。

HelloXX，问题很棒，就是可以再细化一下

比如你要研究哪年的数据

有效数据时间有点太宽泛了，提问的时候最好直接定义好

然后第一个问题的后半部分也是一样的，没有具体说与哪个数据有无相关性

城市也最好确定一个城市这样最好了

比如第一个问题可以问：2014年北京pm2.5与温度之间的变化规律是什么？

xx能不能根据我给你的提示重新吧第二个问题也改一下呢？

好的，在修改...

海永这边没基础，没关系的，这个问题我帮你往简单的方向引导

到时候你做项目会比较轻松的

第一个问题：2014年北京pm2.5与温度之间的变化规律是什么？

第二个问题：2014年上海与北京PM2.5平均值大小与两城市年平均气温的关系？

第二个我觉得比较两个城市之间的pm2.5就好了

要是加上气温可能要出4个图

可能稍微有点难

学员11：

第一个问题：2012年-2015年北京市PM 2.5的数据在不同的月份有什么变化趋势？

第二个问题：2012年-2015年北京市PM 2.5的数据和DEWP:露点(摄氏温度)、TEMP:Temperature(摄氏温度)、HUMI:湿度(%)、PRES:气压 lws: 累计风速(m/s)、precipitation 降水量/时(mm)、lprec: 累计降水量(mm)这几个气象数据有线性相关性吗？

第一个问题非常棒，考虑到了固定时间和城市

第二个问题想研究pm和哪个因素的相关性这很好，但是有一个问题就是东涛这边有5个因素，你就要出5张可视化的图（4张散点图+一张热力图）哦

可能可视化代码量有点大哦

优达项目这边的建议是控制一个变量（温度，湿度，风速，降雨量中的一个）与北京的PM2.5数据是否有相关性

好哒我改改。

主要是感受这个分析的过程和结论的撰写，后面学习正式的入门课程会很详细的讲几个相关变量之间用热力图输出相关关系图

要是孝天现在有兴趣我也可以提前发给你看一下哒~