**当我们谈论逃离北上广我们在想些什么？**

**——基于Python的数据分析可视化**

1. **背景介绍**

“逃离北上广”一种社会现象，即逃离北京、上海、广州这些一线城市。它是在大城市房价居高不下、生活压力持续增长的情况下，在白领中兴起的思潮。要不要逃离北上广，到二三线城市去安放青春，这是两种价值，选择哪种是性格和每个人具体情况所决定的。2016年网络营销案件“逃离北上广”，引发轰动的舆论反响。2017年再度炒冷饭，引发热议。我们将从北京空气质量，就业压力，以及生活压力三个方面，通过数据分析告诉大家是否逃离北上广。

1. **小组成员**

董文伟 201721198595 软件工程 技术支持

郑宝英 201721198147 语言学及应用语言学 文档撰写、整理

匡燕芳 201721198173 语言学及应用语言学 答辩及翻译

1. **技术路线**
2. **数据来源**

天气后报网 北京历史天气http://www.tianqihoubao.com/lishi/beijing.html

北京市教委《2017年度北京地区高校毕业生就业质量报告》

腾讯与企鹅智库《城市出行半径大数据报告》

诸葛找房数据研究中心《北京市套均租金和平米租金情况》

高德地图《2017年度中国主要城市交通分析报告》

中国产业信息网 《2017年全国50城市房租收入比报告》

1. **技术路线**

**获取方法：**通过python的scrapy包对网站按广度优先进行爬取，并使用正则对网页结构进行解析，获取有用信息。

**具体步骤：**

1、首先，爬取了北京2011-2017年每天的天气数据，通过折线图统计并展示这几年北京雾霾天数的变化，经过初步分析2015和2016年雾霾比较严重，然后以2016年为例观察北京雾霾各月的变化，将全国省会雾霾的严重度在地图上标明，用来可视化展示全国空气质量情况。

2、其次，从毕业生就业率、房租收入比、通勤时间、城市拥堵情况，综合分析北漂族的就业压力和生活压力。

3、最后，统计百度百科与雾霾、北漂相关联词的词频，使用jieba分词生成两个词云，最后通过flask+echarts在网页上进行展示。

1. **网页展示**

**（详见WebPython/templates/index.html）**

1. **空气质量（雾霾）**
2. **何为“雾霾”**

雾霾天气是一种大气污染状态，雾霾是对大气中各种悬浮颗粒物含量超标的笼统表述，尤其是PM2.5(粒径小于2.5微米的颗粒物)被认为是造成雾霾天气的“元凶”。

1. **雾霾成因**
2. **机动车尾气排放**

环保部采样调查显示，北京地区机动车尾气排放所产生的PM2.5占总量的20%~30%，且主要是二次无机气溶胶。

1. **燃煤烟尘排放**

据环保部调查分析，造成2015年12月北京市严重雾霾污染的最主要因素即为燃煤烟尘排放，环境总体恶化趋势尚不能根本遏制，污染源排放压力还在增大。

1. **工业废气排放**

据中国科学院研究表明，工业废气排放是产生PM2.5的又一主要因素。以京津冀区域为例，二次无机气溶胶、硫酸盐和硝酸盐占整个PM2.5产生量的26%，工业生产、燃煤和扬尘污染分别占25%、18%和15%，其余来自于周围重工业省份的飘散与垃圾焚烧。

1. **其他因素**

城镇化的快速发展、人口密度的不断增长、鳞次栉比的高楼大厦层叠交错，都对流经城市的季风产生了明显阻碍，不利于大气污染物的扩散与稀释，最终导致PM2.5细颗粒物在城区附近的长期沉积。

1. **雾霾危害**

**短期内对弱势群体的直接危害，日后也将体现在我们这一代年轻群体身上。生活在雾霾严重的一线城市，约等于慢性自杀。**

1. **儿童篇**

* 短时间：危害儿童肺部发育及防御系统，引起急性呼吸道感染，如感冒、扁桃体炎、咽喉炎；导致儿童哮喘的急性发作。
  + 长时间：患哮喘病，使儿童肺部功能发生变变化，造成肺部功能缺陷，增加患呼吸系统疾病的风险，以及成年时期患并发症和死亡的风险。
* 根据美国研究表明，一岁儿童长期居住在汽车尾气污染严重的区域，7岁时容易患多动症。

1. **老人篇**

* 引起或加重急性呼吸道感染和炎症，如感冒、扁桃体炎、咽喉炎
  + 引起已有疾病的发作或恶化，包括慢性阻塞性肺病、哮喘、心脏病、动脉粥硬化、血栓等
* 增加心肌梗死和中风的危险。

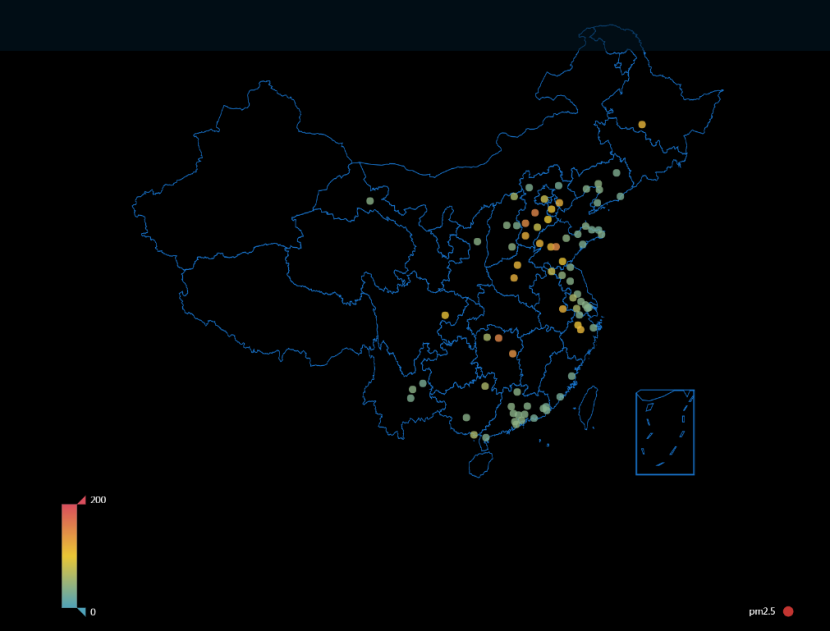
1. **孕妇篇**

* 新生儿体重过轻
  + 宫内发育迟缓
* 出生缺陷
  + 增加新生儿死亡的风险

**结合网站数据分析：**

从2011年开始，美国大使馆持续播报PM2.5的行为引发了社会关注，加之政府终于承认“雾霾”这一词，导致我们对雾霾更加关注。

统计数据显示，2016年雾霾较为严重，我们用相关数据制作全国主要城市空气质量图，直观展示2016年全国各省市全年雾霾的分布情况，颜色越红代表雾霾越重。



通过分析2011到2017年数据，发现北京雾霾天数从2011年开始增多，

到2015年达到最多。



以2016年为例，北京一年中1月到8月雾霾天数几乎差别不大，到9月开始雾霾天数直线上升，分析可能由于冬天供暖等原因。



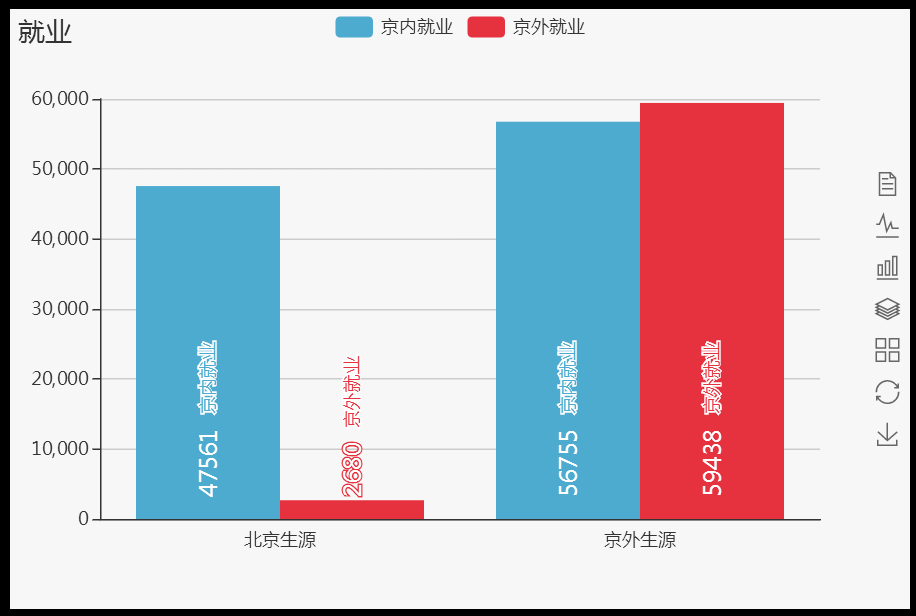
**原因分析：**从上表看到2016年2月雾霾天数骤降，分析可能的原因主要有：

* 2016年2月8日为春节，春节前后，不管是管理力度还是宣传力度都大大提升，“红火春节绿色过”更加深入人心。
* 2月为高校寒假时间，再加上春节前后，相当数量的人离京回乡过春节，在京人口大幅度下降。
* 此外，2016年2月份国内也举行了多次会议，对雾霾的减少可能也会稍有影响。

1. **就业压力（以2017年北京高校毕业生数据为例）**

北京市教委公布了2017年度北京地区高校毕业生就业质量报告，报告数据主要来源于 2017届北京地区 92所普通高校（不含科研院所）的毕业生就业信息库和 85所普通高校的毕业生就业创业状况调查数据。

首先，调查数据显示，2017北京地区高校毕业生23万人，就业率超过90%，这是一个令人振奋人心的消息。然而，在对毕业生的就业地区进行调查分析时，可以得出如下结论：



截至 2017年 10月 31日，92所普通高校已就业毕业生人数共计 167055人。其中，已就业北京生源毕业生中，47561人在北京就业，占 94.67%；2680人在京外就业，占 5.33%；已就业京外生源毕业生中，56755人在北京就业，占 48.85%；59438人在京外就业，占 51.15%。

1. **生活压力**
2. **通勤时间**

腾讯与企鹅智库联合发布的《城市出行半径大数据报告》数据分析：

* **工作日内部出行：北京平均半径9.3公里，居四城之首**

随着城市化的快速推进，人口向城市特别是一线城市涌入，随之产生了一线城市的通勤压力问题。对比北上广深这些一线城市的工作日通勤距离，北京市工作日平均出行半径为9.3公里，上班族通勤时间压力居四城之首。

* **假日外部出行：上海热情偏低，京穗深三市较高**

上海居民最喜欢市内出行，对上海市外出行的热情偏低。

在境外出行路线上，北京、上海居民的出行目的地已经比较多元化，分布在多个大洲，呈现向欧美发达国家的辐射分布趋势。

* **春节跨区出行：深圳空城率最高，京沪半数人离开**

大数据指出在春节期间至少有一半的人离开所工作居住的城市，其中深圳的空城率达到69%，为四城最高，堪称名副其实的“空城”。京沪也有半数以上的人离开。

* **城市交通规划：减少工作区和居住地出行半径**

通过大数据研究发现，职住分离使得一线城市的交通负重运行，解决职住分离的问题已成为提升幸福指数、降低城市运行成本的重要举措。

对高收入者来说，一方面能够通过改善交通工具而缩短通勤时间，另一方面愿意以更长的通勤距离来换得更高收入的职业或更好质量的住房。而低收入者因为住房支付能力低，迁居并改善住房条件的机会少。

北京的远距离居住密集区最为集中，燕郊、廊坊市区与工作区仅有公路交通相连，交通压力更为严峻。发展北京五环外及燕郊的轨道交通，提升北京出行效率较为重要。

上海的居住聚集地则相对分散，交通压力相对较小，但青浦、奉贤等居住区的出行时间仍然较长，发展连接郊环至中心城区的交通，有利于提升出行效率，缓解上海进城压力。

1. **房租成本**

诸葛找房数据研究中心对当前我爱我家、链家、中原[地产](http://house.china.com.cn/" \t "http://house.china.com.cn/newscenter/view/_blank" \o "地产)、21世纪不动产、丽兹行等中介公司挂牌租房房源及58同城、赶集网等互联网租房平台的数据进行整合分析。

****

1. **城区租金情况**

从每平米租金上来看，城六区与郊区差距较大，最大差距接近6倍。

其中西城以123元/平米的租金占据榜首，郊区方面，昌平由于占据西二旗等科技园区及昌平高教园等优质教育资源，以57.4元每平米领衔郊区房租。

从套均价来看，东城区以9577元/套占据榜首，朝阳以9565元/套微小差距紧随其后，海淀排名第三，套均价超9000元。

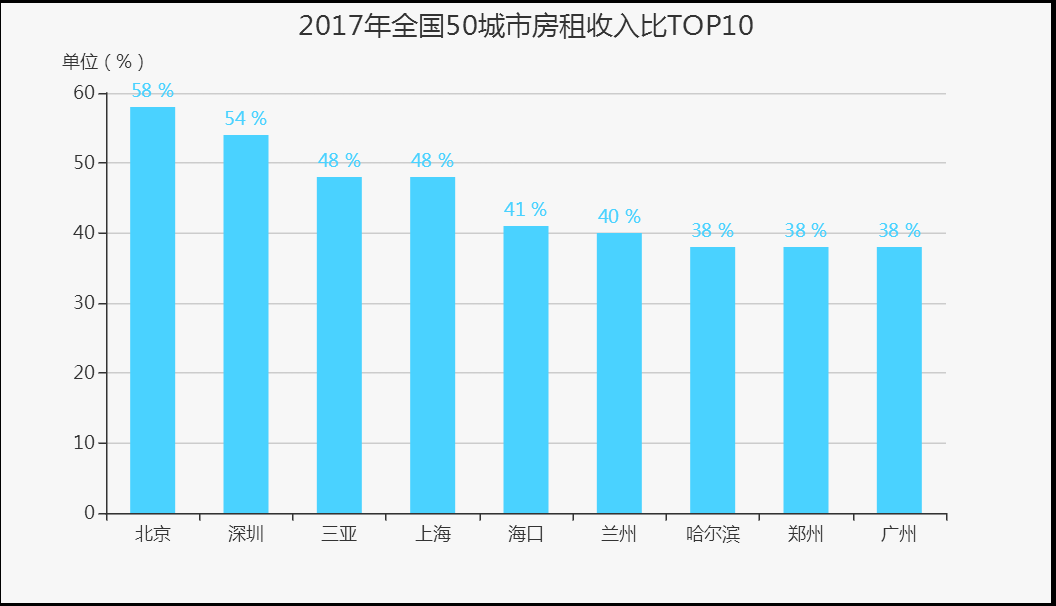
1. **商圈租金情况**

商圈综合排名方面，地处中心城区西城区的金融街凭借金融、文化、消费三大优势，以185.7元每月每平米的租金单价毋庸置疑的拔得头筹。例如，在官园租一个50平米的一居室，平均也要8500元。

1. **地铁线租金情况**

作为北京市唯一的两条内环线——2号线和10号线，占据地铁换乘优势，每平米遥遥领先其他地铁沿线，2号线更是以129.5元每平米每月占据榜首，均价是房山线的2.67倍。而在良乡大学城租一套50平的一居室，却只需要2000元不到。

根据中国产业信息网的统计数据，**北京的房租占总收入的比例高达58%**，也就是说，工薪族每个月将大部分的薪资都用在了房租上，上海、广州比例略低，但仍然数值较大，分别为48%和38%。



从图中可以看出，北上广等大城市，大家把大部分收入都用来租房子，所以北上广多为月光族。

1. **交通拥堵**

2018年1月18日，高德地图联合交通运输部科学研究院等机构发布了《2017年度中国主要城市交通分析报告》，报告显示全国26%的城市通勤处于拥堵状态，55%的城市高峰时处于缓行状态，其中北京、上海、广州拥堵情况均位列全国前十，人均年拥堵成本分别为268h、249h、247h。

根据数据来看，北上广地区高峰拥堵延时指数较高，高峰拥堵延时指数指在高峰拥堵时期所花费的时间与畅通时期所花费时间的比值，从数值来看，北京高峰拥堵时期花费的时间是平常的两倍还多，可见工薪族的时间成本很高。



众所周知，北上广这样的一线城市带来很多的机遇与挑战，尤其对于年轻人来说，能在这些城市工作是很可贵的机会。但是，综上所述，在具体分析北京高校毕业生就业率、房租收入比、通勤时间、城市拥堵等情况之后，会发现这些城市的就业压力和生活压力远远高于其他一线城市，需要付出的代价比我们想象的更大。因此，为了不把生命浪费在还房贷、通勤上下班的路上，很多人会选择逃离北上广，退居其他一线城市选择就业。