

综合利用各种图表进行分析，可使用 Excel，Echarts，Tableau，Power BI，Processing，D3.js、Python、MySQL 等

挑战 1.1：伪基站常流动于人口密集的区域，以各种名义向一定范围内的手机发送垃圾短信，因此，了解掌握伪基站出行的时空模式，能够帮助执法人员尽早阻止和抓获不法分子，从而更好地维护社会秩序。然而仅仅从垃圾短信中很难确定其对应的伪基站，即无法确定来自同一台伪基站设备的垃圾短信，相同的垃圾短信有可能来自不同的伪基站，同一个伪基站可能不送不同的短信。请从宏观时空分析的角度出发，对垃圾短信数据进行可视分析，揭示伪基站的总体时空活动规律。

1、统计垃圾短信总量，发送号码总量，基站位置总量。

2、伪基站时间活动规律

- (1) 按接收时间统计每天的垃圾短信数量
- (2) 统计伪基站每天各个时段的短信数量
- (3) 统计出现频率较高的伪造的号码

3、伪基站空间活动规律

- (1) 将短信记录的位置投射在地图上
- (2) 统计北京各行政区垃圾短信数量
- (3) 接收时间连续，经纬度在一定范围内且 MD5 相同的短信可以被认为是由同一个基站发出，据此计算出各行政区每天的伪基站总数量，该数值将一天中不同时间点的同一个伪基站视为不同的伪基站，伪基站的工作时长一般为 9~12 小时，所以还需将各区每天的伪基站总数量除以各区伪基站工作时间，则可得到各行政区域内伪基站的平均数量。

3、结合北京的政治地理等知识进行综合分析

挑战 1.2：不法分子通过设置伪基站设备能够发送不同类型的垃圾短信，请尝试对垃圾短信的具体内容进行分类，分类标准不限，例如：按垃圾短信类型可以分为广告、诈骗等等，按垃圾短信对人们的人生经济危害程度可以分为一般、严重等等。请尝试在问题 1 的基础上进一步分析伪基站发送不同类型垃圾短信的时空分布规律。（建议参赛者回答此题文字不多于 1500 字，图片不多于 8 张）

1、按照垃圾短信的类型分类

- (1) 对短信内容进行分词和词频统计
- (2) 选择前 100 高频词汇，利用词云显示短信文本中出现频率高的关键词
- (3) 分析垃圾短信具体正文，将垃圾短信分成几大类，每个大类又细分为一些小类

2、将垃圾短信按照经济危害程度进行分类

3、根据问题一中得到的伪基站活动时间规律，分析常见类别的垃圾短信的时间规律

- (1) 统计不同类别的短信在 24 小时之内的变化趋势及其高峰时间

4、统计各行政区内不同类别及经济危害程度的短信的分布规律

5、结合时间和空间进行综合分析

挑战 1.3：伪基站不仅破坏正常电信秩序，危害公共安全，扰乱市场秩序，而且严重损害群众财产权益，侵犯公民个人隐私，社会危害严重。据《人民网》统计，每年通过“伪基站”设备发送诈骗、赌博、推销、中奖等短信近千亿条，

伪基站已成为社会一大公害。请结合以上两题中得到的伪基站行为模式，向执法人员提出打击整治伪基站的有效建议和方案，并结合数据分析结果进行说明。

- 1、执法人员应根据危害程度和数量多少进行准确打击
 - 2、根据冒充的号码，与对应的运营商、商家合作，找出有用的线索
 - 3、分析伪基站的移动路线，定点抓捕打击。
 - 4、分析作案规律，是否有团伙等
 - 5、与交通部门协作，利用交通摄像头等发现可疑车辆
 - 6、针对时间、空间规律对不同类别的垃圾短信采取针对性的措施
-