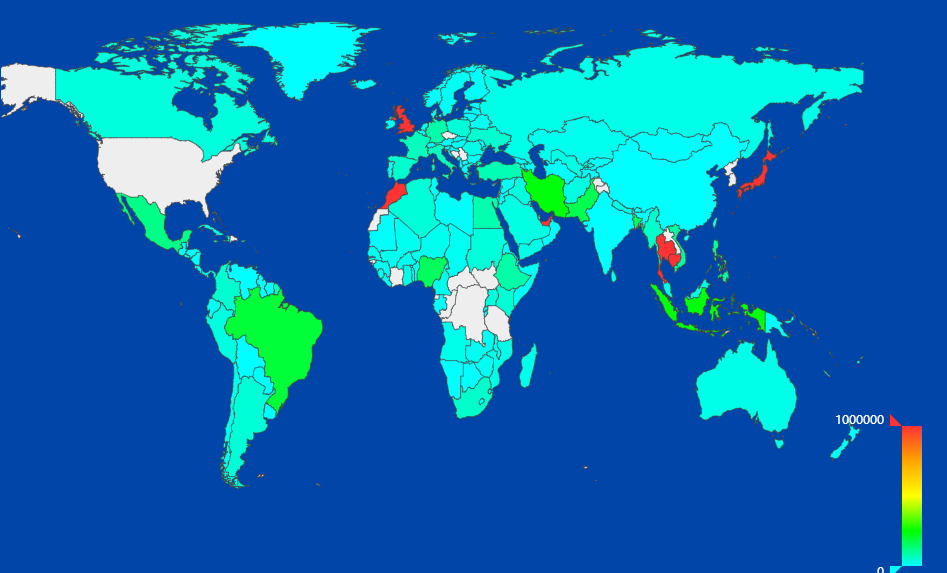
**基于TripAdvisor的旅游数据分析与可视化**

卢梦依 201721198599 朱述承 201721198327张 琪 201721198022

1. **主要内容**

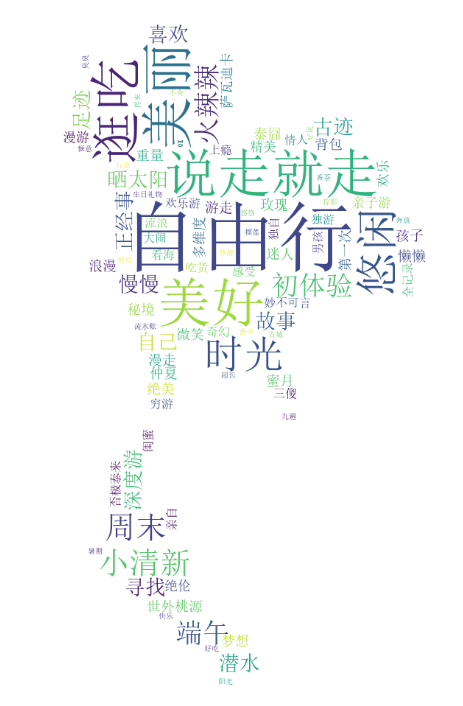
**1.1热门旅游点的分析与可视化**

通过统计TripAdvisor游记板块每个国家的游记数量，选出最受国人喜爱的5大出境旅游国家，并进行可视化。



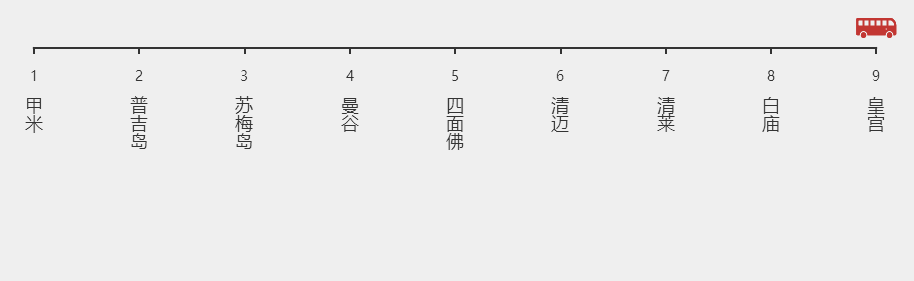
**1.2热门国家热词的分析与可视化**

统计每个热门国家下的游记标题出现的词，排除地名和虚词等，制作每个热门国家的旅游热词地图词云。



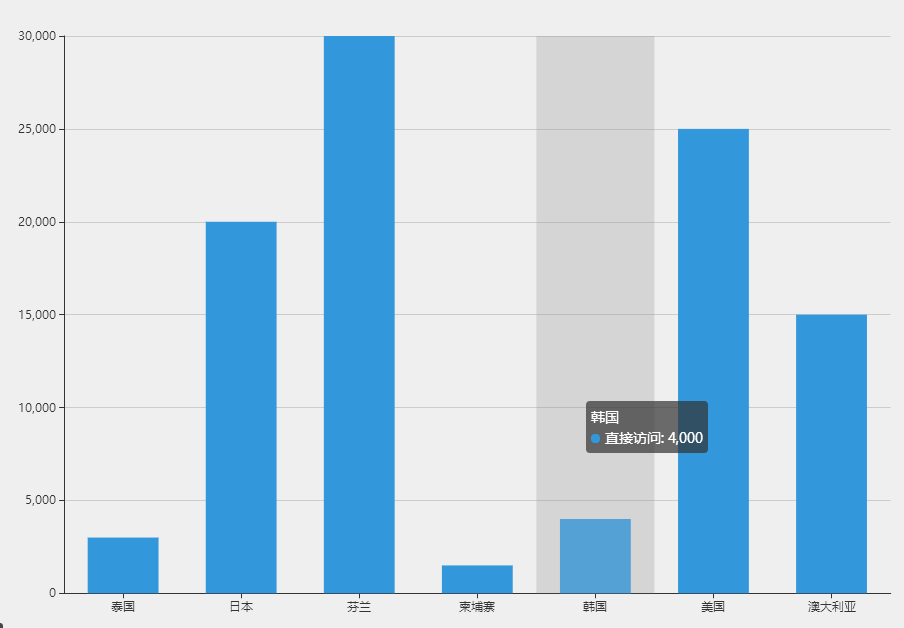
**1.3热门国家热点景点与城市的分析与可视化**

统计热门国家游记中出现的地名（城市和旅游景点），挑选频次较高的景点和城市，制作热点景点城市图。



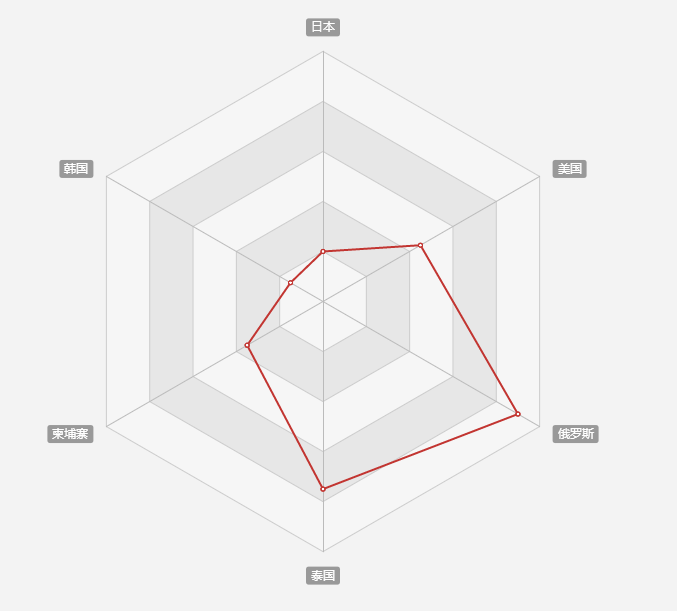
**1.4热门国家花销的分析与可视化**

统计热门国家平均旅行费用，制作热门国家花销直方图。



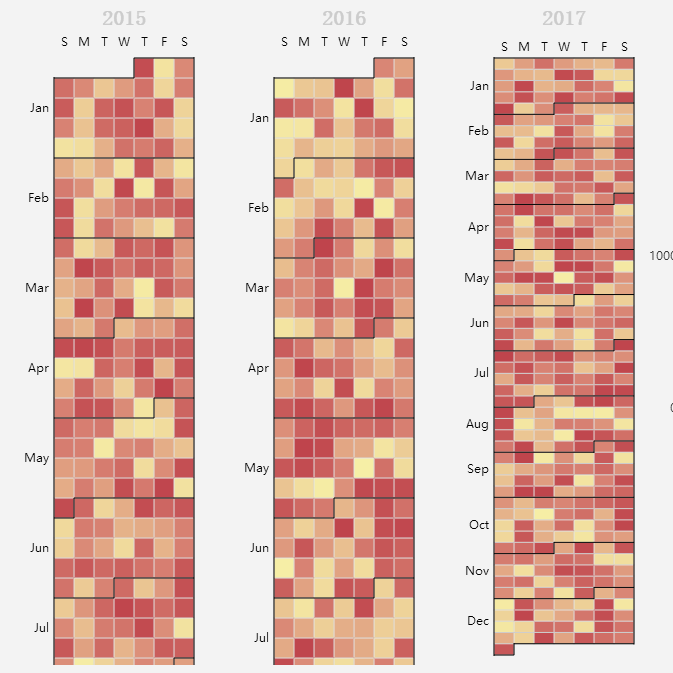
**1.5热门国家游玩时长的分析与可视化**

统计热门国家的平均旅行时长，绘制热门国家游玩时长图。



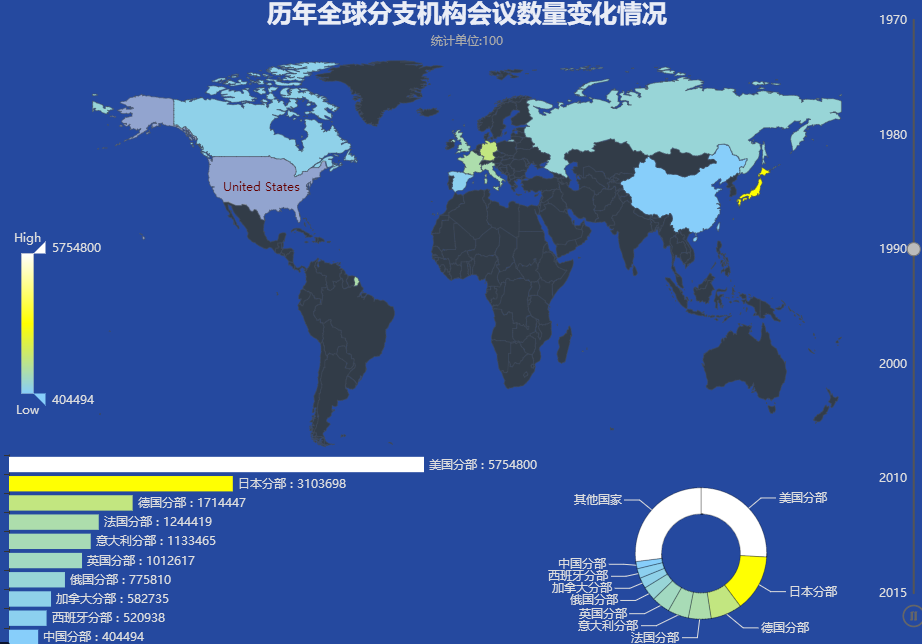
**1.6热门国家旅游季节的分析与可视化**

统计热门国家旅游的日期，绘制热门国家旅游季节图。



**1.7近几年热门国家变化趋势**

统计热门国家近几年的游记变化数量，制作近几年热门国家变化趋势图。



**1.8小结**

分析各项统计数据，给出热门旅游国家的历时变化和总体趋势，总结每个热门国家的特性词汇。根据用户需求，可以从游玩时间、游玩时长、预算支出、冷门与大热等等方面，给出国家景点城市推荐。

1. **技术路线**

**2.1数据来源与获取方法**

通过Python爬虫爬取TripAdvisor官网上的数据，如下图所示，包括游记题目、出行日期、旅游和时长等。



**2.2分析与统计**

统计内容包括每个国家总的游记数目、每个国家每年的游记数目、每个国家的高频热词、高频旅游景点和城市、每个国家的平均花销和平均旅游时长、每个国家的旅游时间等。

**2.3可视化与展示**

热门旅游点的展示可使用：

<http://gallery.echartsjs.com/editor.html?c=xSJagzcnLZ>

热门国家热词的展示可使用wordcloud词云。

热门国家热点景点与城市的展示可使用：

<http://gallery.echartsjs.com/editor.html?c=xByj7eOSMx>

热门国家花销的展示可使用：

<http://gallery.echartsjs.com/editor.html?c=bar-tick-align>

热门国家游玩时长的展示可使用：

<http://echarts.baidu.com/demo.html#radar>

热门国家旅游季节的展示可使用：

<http://echarts.baidu.com/demo.html#calendar-vertical>

近几年热门国家变化趋势的展示可使用：

<http://gallery.echartsjs.com/editor.html?c=xBkzbFl3WM&v=2>

**2.4可能会遇到的问题与解决办法**

**2.4.1地图可视化工具受限**

由于最近颁布的地图管理条例严格要求了地图出版物的发行，故Echarts上有关世界地图的可视化工具十分有限。故考虑采用其他形式的可视化工具（如柱状图、饼状图、关系图等）进行可视化。

**2.4.2网站的反爬机制**

TripAdvisor的官方网站具有反爬特性，频繁访问会使ip被封，使之禁止访问，影响数据爬取的效率。故考虑用代理IP访问网站的新方法。

1. **进度安排**

12.11-12.18数据爬取、数据处理，12.18-12.25统计与分析，12.25-1.1数据可视化，1.1-1.8搭建可视化网站，完成项目报告并翻译。