

评论分析：

所有区域总的

情感分析。共得到多少条有效句子，其中，“正面”的句子有多少条，占比多少，中性情感倾向的句子有多少条，占比多少，负面情感句子多少，占比多少。



顾客 IP 属地分析

疑问：顾客 IP 属地分布分析有的是空白，没有 IP 属地，有缺失项，还能否做。

V	W	
点	评论ip	论出
2	浙江	情侣
1	辽宁	情侣
0	甘肃	情侣
0	江苏	情侣
2	四川	情侣
0	浙江	情侣
1	广东	情侣
3	四川	情侣
4	上海	情侣
0	上海	其他

在提取顾客体验情感评论数据时，计算出不同 IP 属地评论占总评论比例，从而得到顾客不同地区评论数，可以得出顾客的地区分布，绘制一个中国地图。例如：



供应商回复时间和评论时间做差，看携程回复是否及时。

AA	AB	AC	AD	
程安排	评论时间	共应商回	供应商回	图片
程安排	2020-02-02	尊敬的游	2020-03-18	http
程安排	2020-01-25	尊敬的游	2020-03-18	http
程安排	2020-02-12	亲爱的携	2020-03-26	http
程安排	2020-01-26	11:39:44		http
程安排	2020-01-26	17:39:56		http
程安排	2020-01-24	尊敬的游	2020-03-18	http
程安排	2020-01-23	13:30:33		http
程安排	2020-02-09	尊敬的游	2020-03-18	http

下面分三大区域做：

东南亚，日韩，欧美每个区域再用 Stata 等统计软件汇总东南亚区域铺了多少条线路，也就是 E 列有多少个不同的产品，相同价格对应一个产品线路，（H 列）票价格均价是多少，（K 列）东南亚这么多条线路出游人数一共多少。（P 列）计算出东南亚区域整体满意度多少。

情感分析共得到多少条有效句子，其中，“正面”的句子有多少条，占比多少，负面情感句子多少，占比多少。做成三线表。

E		H		O		P		K	
产品名		票价	评	企	产品满意度	评		出游人	品
印度尼西亚				日 (935)	92.7	PRC		7516	携程
印度尼西亚		5752		日 (935)	92.7	PRC		7516	携程
印度尼西亚		5752		日 (935)	92.7	PRC		7516	携程
印度尼西亚		5752		日 (935)	92.7	PRC		7516	携程
印度尼西亚		5752		日 (935)	92.7	PRC		7516	携程
印度尼西亚		5752		日 (935)	92.7	PRC		7516	携程
印度尼西亚		5752		日 (935)	92.7	PRC		7516	携程

产品线路	出游总人数	平均单价	整体好评率	好评占比	差评占比
东南亚					
日韩					
欧美					

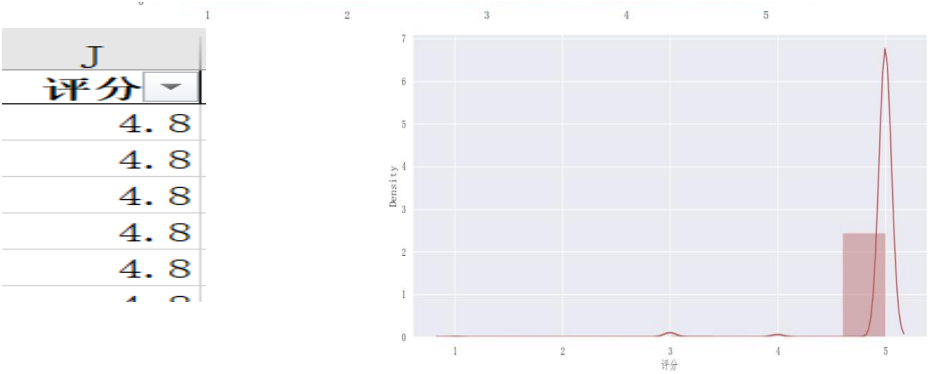
评论出行类型。做一个三线表，各种类型占比

X	
论出行	评
情侣夫妻	多
情侣夫妻	早
情侣夫妻	望
情侣夫妻	送
情侣夫妻	主

例如：

情侣夫妻	单独旅行
东南亚			
日韩			
欧美			

计算出不同评分的比例（J 列）从而得到顾客评分频率，横坐标为评分三大区域分开评分趋势图。做成 Excel 统计图表折线图。



三大区域分别 LDA 分析

LDA 虽然可以直接对文本进行主题分析，但是文本的正面评价和负面评价混淆在一起，因此，将文本分为正面评价和负面评价两个文本，再分别进行 LDA 主题分析。将文本分为正面评价和负面评价两个文本主要采用的是（）软件进行，该软件可以自动生成“正面情感结果”“负面情感结果”“中性情感结果”三大类，本文仅对“正面情感结果”“负面情感结果”文本进行 LDA 主题分析，分别挖掘三大区域产品的正面评价和负面评价，挖掘优势和不足。