# 基本分析思路

确定业务问题：如何让房东在airbnb中获得更多的收入，并分解为2个子问题：1，什么影响airbnb的价格，2，如何提高airbnb的价格

对数据进行处理，将不具有分析意义的变量，如“ID\_airbnb”，“title\_airbnb”，在进入编程语言之前删去。对于非数值型，进行类别转换，即将非数值型转换为数值型。将缺失值超过百分之80的变量删去，将超过百分之八十相同的变量删去。最终得到的数据共有33个变量【变量的具体选取和转换方法见最后：变量转换】

对这两个问题，采用线性回归，决策树和聚类方法分别研究，算法运行后根据作业要求进行算法诊断，最终算法的性能和显著性均满足要求。

算法得到的主要结果：影响airbnb价格的因素有三个层次。首先是这个airbnb的基本条件，如房间类型，房间面积，房间数量等，第二是这个airbnb的附加设施情况，如有没有洗衣机，烘干机，等。最后是这个airbnb的服务情况，如回复时间，退改规则等。三种因素的影响依次递减。

最后提出建议，先改善基本条件，再改善附加设施，最后提高服务水平。

【可能用到的知识内容及其参考资料】

多元线性回归：<https://www.cnblogs.com/zhiranok/p/xianxinghuigui.html>

决策树：<https://blog.csdn.net/RH_Wang/article/details/81703151>

主成分分析：<http://www.360doc.com/content/19/0818/16/57890290_855677956.shtml>

聚类方法：<https://www.jianshu.com/p/e4d5a0fbcefe>

# 变量转换

ID&title

对分析无用，直接删除

# refund\_rule

* 매우 3 60일：4
* 엄격：3
* 일반：2
* 유연：1

转化为数字，从而能参与各种分析：

# response\_time

* 하루 이내:24
* 몇 시간 이내:6
* 1시간 이내:1
* 며칠 이상:72

# type\_airbnb

* 아파트의 개인실：1
* 아파트의 다인실：2
* 아파트 전체：3
* 주택의 개인실：4
* 집전체：5
* 로프트의 다인실：6
* 호스텔의 개인실：7
* 공동4：8
* 개인실：9
* 호텔의 9：10
* 게스트용 별채의 9：11
* 로프트 전체：12
* 로프트의 9：13
* 타운하우스 전체：14
* 부티크 10：15
* 게스트 스위트 전체：16
* B&b의 9：17
* 레지던스의 9：18
* 레지던스 전체：19
* 게스트용 별채 전체：20
* 타운하우스의 9：21
* 전원주택 전체：22
* 콘도(아파트) 전체：23
* 공동주택의 다인실：24
* 도미토리의 9：25
* 도미토리의 다인실：26
* 방갈로의 9：27
* 게스트 스위트의 9：28
* 저택 전체：29
* 통나무5：30
* 리조트 이용권 전체：31
* 타운하우스의 다인실：32
* 캠핑카(RV)：33
* 텐트의 9：34
* 방갈로 전체：35
* 산장의 9：36

# registration\_date

原格式：“4월 2013”，指2014年4月注册，为让该变量有实际意义，我们计算这个日期与2019年1月的差值，如4월 2013转化为：（2018-2013）\*12+（13-4）=69，即注册时间为69个月前。

# time\_stamp

可能是爬取数据的时间戳信息，对分析无用，删去

列名处理：

将【무선 인터넷，세탁기，건조기，다리미，노트북 작업 공간，필수품목，난방，에어컨，주방，아침식사，옷걸이，헤어드라이어，샴푸，침실문 잠금장치，소화기，일산화탄소 감지기，화재 감지기，구급 상자，TV，엘리베이터，온수，외부 주차(유료)，열쇠 보관함，케이블 TV，프런트데스크/경비원，해수욕 필수품，커피메이커，기본 조리 도구，식기류，전자레인지，냉장고，오븐，가스레인지，호스트가 게스트를 맞이함，여행 가방 보관 가능，침구，여분의 베개와 담요，도로 주차(무료)，기타，장기 숙박 가능，헬스장，식기세척기，게임기，체크아웃 전 청소 필요，호수나 강과 인접，실내 벽난로，이더넷(Ethernet) 연결 지원，전원 콘센트 덮개，암막 커튼，창문 안전장치 설치，어린이용 책과 장난감，욕조，베이비시터 추천 가능，아기 모니터，어린이용 식기，단층 주택，건물 내 무료 주차，바비큐 그릴，파티오 또는 발코니，건물 내 유료 주차，단독 사용하는 거실，스마트락，수영장，다기능/여행용 아기 침대，게스트 전용 출입문，자쿠지 욕조，정원 또는 뒷마당，포켓 와이파이(Pocket wifi)，아기 침대，유아용 식탁의자，디지털 도어록，벽난로 안전 장치，계단 차단문，테이블 모서리 보호대，기저귀 교환대，아기 욕조，전기차(EV) 충전시설，해변과 인접，호수로 연결됨，스키를 탄 채로 출입 가능】

直接翻译为英文，作为列名，单词间空格用下划线代替

经过基本调整的数据存储为AdjustedData.csv