

## 一、分词与词频统计：

- 1.统计所有诗词的总词（字）频，获得乾隆全部词汇使用概况  
产出：TOP500 高频词表（可观察“朕”、“国”、“风”、“月”等词的地位）  
将每首诗的分词结果 txt 文件按“模糊顺序”编号  
创建文件，包含每首诗的[诗编号,分词内容]，结构化的待分析数据集

## 二、意象分析：

- 1.自然意象的偏好（自然类：山、水、云、松、梅、雪...），统计出现频率及共现词  
遍历所有诗，匹配意象词表，统计频次；对同一诗中出现的意象做共现计数  
发现乾隆最爱的自然景物（如“松”可能高频），观察“雪-梅”、“山-水”等固定搭配
  - 1.1 “乾隆的‘梅兰竹菊’排名”：统计四君子出现次数，可能发现乾隆最爱“竹”（因其象征正直，符合君主形象），而非文人更爱的“梅”。
2. “意象密度”变化计算每首诗的意象词数占总词数的比例  
按时间顺序（诗编号）绘制意象密度折线图，平滑处理（如每 50 首移动平均）
  - 2.1 发现乾隆作诗风格变化：早年意象密集 vs 晚年铺陈直白

\*意象密度，例：如某一个意象在其前后 250 首诗中出现了 n 次，即意象密度为 500 首诗中，意向密度为 $(n/500*100\%)$
- 3.私人趣味意象筛选特定小类收藏（鼎、玉、画）、饮食（茶、酒）  
统计这些词的出现频率，观察其集中出现的诗篇区域  
揭示乾隆的个人爱好在诗中的反映（如“茶”出现集中的时期可能对应其品茶活跃期）
  - 3.1“乾隆的诗有多‘俗’？”：统计“金”、“玉”、“宝”、“贵”等炫耀性词汇的出现频率。
  - 3.2“乾隆诗中的‘吃货’线索”：统计所有饮食相关词，看其偏好。
  - 3.3“深夜作诗爱好者”：统计“夜”、“月”、“灯”、“烛”等词，看其作诗时间偏好。

## 三、情感分析：

- 1.情感分析统计，创建积极、消极、中立三个情感词典，计算每首诗的情感倾向公式：  
 $(\text{积极词数}-\text{消极词数})/\text{总词数}$   
每首诗都有情感倾向的程度分数。

## 四、模糊时间分段与演变分析：

- 1.人工查询乾隆生平，划分关键时期（示例分段）：  
时期 1：继位初期（约诗编号 1-500）：1735-1745 年，年轻君主，锐意进取  
时期 2：统治中期（约 501-3000）：1746-1770 年，十全武功高峰，频繁巡幸  
时期 3：统治后期（3001-7000）：1771-1795 年，盛世隐忧，文字狱高峰  
时期 4：太上皇时期（7001-结尾）：1796-1799 年，退位后反思  
等等（示例分段，请自行分段）  
将时间分隔段上的诗词进行整理，将按时间顺序排列的诗词分为若干段。  
各时期分别进行意象 TOP10、情感均值、高频动词（巡、阅、忆、叹...）统计

- 1.1 “晚年词频泄露的真相”：统计晚年时期（最后 2000 首诗）中“倦”、“眠”、“旧”等 词的 出现频率，探究其身心状态。
- 1.2 “乾隆的旅行诗”：通过地名词（杭州、塞北、热河）识别“旅行诗”，看看这位皇帝爱往什么地方去旅游打卡
- 1.3“诗作长度与年龄关系”：统计每首诗字数，看是否随着年龄增长而变长（唠叨化）。

## 五、拓展生平经历与执政经历分析（待定，可以不做）：

### 1.乾隆形象建构

统计第一人称词汇（朕、吾、孤）的频率变化，分析其与“天”、“祖”、“臣”等词的共现  
统治权威的巩固过程、晚年自我认知变化

### 2.统治焦虑的流露

识别“忧”、“惕”、“艰”、“难”等词的出现段落，结合该时期历史事件（如白莲教起义萌芽）  
盛世光环下的压力与危机感

### 3.人际关系痕迹与关系图构建

人工筛选提及特定人物的诗（如“孝贤”（皇后）、“傅恒”（将领）），分析其情感倾向与出  
现时机

感情生活（悼亡诗）、君臣关系

### 4.执政重点转移

分时期统计政务相关词（农、河、赋、戍）的频率变化，绘制主题强度曲线  
水利工程、边疆政策、经济重心的演变