* 求职竞争力：sql, python思路整理提升，简历提升
* Youtube数据集

Definition Check

* Type 1, type 2 error
* Hypothesis
* ABtest

进阶项目

* Google cloud
* AWS
  + 处理超大量数据
  + 云平台运用云计算做到分布计算
* \*LLM
* DA/BA: 偏得到general结论，从商业层面分析公司应该做什么决策，要说出商业洞察，只是为了说出opinion
* DS/MLE：更关注怎么用数据建出模型，更着重如何微调模型，选择更好的模型，对数据更好的处理来建出未来更合适的模型

拿到数据后

* Data exploration: 先看第一行，看包含哪些维度的信息
* Need to clarify:
  + Video id
  + Trending: 上榜日期
    - 2017年11月14日
  + 标题
  + Channel 名字
  + 视频属于哪一类，需要merge json文件
  + 发布时间
  + 标签
  + 浏览量
  + 点赞量
  + 点踩量
  + 评论数
  + 缩略图
  + 上榜的时候是不是不让评论：false：允许评论，true：不允许评论
  + 是不是允许打分
  + 统计数据时这个视频是不是还在，有没有删掉或出错
  + 总体描述
* 探索性分析