气象数据的来源是Reliable Prognosis（rp5.ru）网站，实际天气数据由地面气象数据国际自由交换系统提供，网站的天气预报一天两次完全更新，最新数据一天八次每隔三小时补充到网站数据库中，这里我们只给出太原和上海最近十年的气象数据，如有其它需要，自行到网页下载。

对于上海和太原2009年1月1日到2018年12月31日整十年的数据，年份比较早的数据大多数都是一天四次每隔6个小时补充到网站数据库中，最新数据是隔三小时记录一次气象数据，网站收集的信息包括：T（地面以上以上2米处的大气温度）、Po（气象站水平的大气压）、P（平均海平面的大气压）、U（地面高度2米处的相对湿度）、N（总云量）、VV（水平能见度）、Td（地面高度2米处的露点温度）、RRR（降水量）等。为了尽可能准确预测天气情况，需要综合考虑气象数据中的一些相关指标，这其中就会有许多潜在问题，比如对于所需数据中的缺失值该如何解决；对于不同维度的数据量纲之间存在差异，如何进行标准化能减小量纲对结果产生的影响；预测天气情况，气象数据各维度之间通常具有极强的依赖性及关联性，如何去进行维度约减对预测结果的影响最小；在预测天气状况建模的过程中，在运用各类数据进行分析处理时，如何能建立最佳模型，使预测结果更加准确等等。请运用数据挖据的相关知识进行分析。