《表格数据处理与分析复习（一）》拓展资源

数据与数据分析报告

**一、数据与数据分析**

数据分析是指有针对性的收集、加工、整理数据，并采用统计、挖掘技术分析和解释数据。数据分析是为了提取有用信息和形成结论而对数据加以详细研究和概括总结的过程。

数据是客观事实,对于客观事物发生，发展的数字化记录。随着科学技术的发展，数据的概念内涵越来越广泛包括数值，文本，声音，图像，视频。

**二、数据分析大事记**

**1. 计算机时代崛起**

计算机的发展和计算技术的发展极大地增强了数据分析的过程。1880年，在使用计算机之前，美国人口普查局花了7年的时间来处理收集到的信息并完成最终报告。使用此设备，在18个月内完成了1890年的人口普查。

**2. 关系数据库诞生**

关系数据库由埃德加·科德（Edgar F. Codd）在1970年代发明，并在1980年代非常流行。关系数据库（RDBM）允许用户编写Sequel（SQL）并从其数据库中检索数据。关系数据库和SQL提供了能够按需分析数据的优势，并且仍在广泛使用。它们易于使用，对于维护准确的记录非常有用。

**3. 数据仓库诞生**

在1980年代后期，收集数据的数量继续显着增长，部分原因是硬盘驱动器的成本较低。在此期间，开发了数据仓库的体系结构，以帮助将来自操作系统的数据转换为决策支持系统。数据仓库通常是云的一部分，或者是组织的大型机服务器的一部分。与关系数据库不同，数据仓库通常经过优化，可快速响应查询。

**4. 商业智能(BI)崛起**

20世纪80年代-90年代，数十家BI厂商进入市场。数据仓库技术的发展大大推动了商业智能的发展，传统存储在各个地方的业务数据开始集中在一起。应运而生的技术还包括ETL（数据抽取、转换、加载）和OLAP（联机分析处理）。

**5. 数据挖掘崛起**

数据挖掘始于1990年代，是在当时多个学科发展的基础上发展起来的。随着数据库技术的发展应用，数据的积累不断膨胀，导致简单的查询和统计已经无法满足企业的商业需求，急需一些革命性的技术去挖掘数据背后的信息。

**6. 大数据来临**

2000年到2010年是大数据兴起和备受关注的时期，谷歌的“三驾马车”：谷歌文件系统、MapReduce和 BigTable。亚马逊也发表了一篇关于 Dynamo系统的论文。这几篇论文奠定了大数据时代的基础。随着大数据的到来，海量的数据以及新的技术发展，帮助公司将数据转化为洞察力。

**7. 数据科学家**

2012年9月， Tom Davenport和DJ Patil 在《哈佛商业评论》上发表了“数据科学家：21世纪最性感的工作” 。

**8. 增强分析**

2017年，Rita Sallman，Cindi Howson和Carlie Idonies在Gartner的研究论文中引入了增强分析的概念，并将其描述为一种新的数据分析方法，可使用机器学习和自然语言生成（NLG）自动化见解。增强型数据分析大大提升了数据分析效率，降低数据分析的门槛，人人都可以像数据科学家一样，对数据进行多维度的自动钻取，自动加载各种模型进行深度分析。

**三、数据分析报告的 6 个步骤**

在数据分析报告的 6 个步骤中，我认为最重要的是「**明确目标**」，花费时间最长的是「收集数据」和「处理数据」，最有价值的是「**结论建议**」。

**1、明确目标**

***正确地定义问题、合理地分解问题、抓住关键的问题。***

当明确目标之后，我们需要梳理分析思路，搭建分析框架，开始思考以下问题：

采用哪些分析指标？

运用哪些分析思维？

使用哪些分析工具？

明确目标，是确保数据分析过程有效进行的先决条件，可以为后续的步骤提供清晰的方向。

**2、收集数据**

收集数据是围绕数据分析目标，按照分析思路和框架，收集相关数据的过程，为后续的步骤提供素材和依据。

收集的数据包括原始数据和二手数据，其中原始数据包括组织机构内部的数据库、调查得到的数据等；二手数据包括统计局发布的数据、公开出版物中的数据等。

收集数据的基本要求是：***真实性、及时性、同质性、完整性、经济性和针对性***。

**3、处理数据**

处理数据，是从大量杂乱无章的原始数据中，抽取对解决问题有价值的数据，并进行加工整理，形成适合数据分析的样式，保证数据的一致性和有效性，这是数据分析之前必不可少的阶段。

数据的处理主要包括数据清洗、数据转化、数据抽取、数据合并、数据计算等过程，原始数据一般都需要经过一定的处理，才能用于后续的数据分析工作。

在处理数据的过程中，准确性尤为重要，如果数据本身存在错误，那么即使采用最先进的数据分析方法，得到的结果也是错误的，不具备任何参考价值，甚至还会误导决策。

**4、分析数据**

分析数据，是对客观真实的数据，运用恰当的方法和工具，进行科学有效的分析。

**5、展现数据**

通过数据分析，隐藏在数据背后有价值的信息逐渐浮现出现，此时需要通过合适的方式展现出来，让人一目了然，提高信息传递的效率。

通常情况下，展现数据的方式通常是用图表说话，即数据可视化，常用的数据可视化图表有很多。

**6、结论建议**

一份好的数据分析报告，需要有明确的结论建议。

如果换位思考，站在决策者的角度，更想知道的是可行的解决方案。

**如果数据分析报告没有明确的结论建议，那么也就失去了报告的灵魂。**

所以，要想制作出更有价值的数据分析报告，不仅要掌握数据分析的思维和工具，而且还要熟悉业务，这样才能提出更好的建议。

**小结**

数据分析报告的制作过程，通常可以分解为明确目标、收集数据、处理数据、分析数据、展现数据、结论建议等 6 个步骤，这是对整个数据分析过程的总结，为决策者提供科学、严谨的决策依据，从而降低企业的经营风险，提高企业的核心竞争力。

如果把数据分析报告比作一个产品，制作报告的人就是产品经理，看报告的读者就是用户。

作为「产品经理」，同理心很重要，通过自我体验来理解他人，乔布斯能瞬间把自己变成傻瓜，这是同理心的一种表现。数据分析的思维和工具也很重要，它们是数据分析的基础。想象力是广袤的天空，但不是天马行空，而是基于同理心的推演，运用数据分析的思维和工具，让推演更加科学有效。

在一份数据分析报告的背后，有许多枯燥的、基础的准备工作要做，例如数据采集、数据仓库、数据治理等等。

如果没有高质量的数据作为坚实的地基，那么数据分析报告的高楼大厦是不稳固的。**如果没有明确数据分析的目标，那么后面的工作可能就是胡拼乱凑，用一堆图表堆砌的花架子，并不能解决实际的问题。**

数据分析报告不要搞形式主义，而要有实质的内容，还要关注细节。