



日照职业技术学院
RIZHAO POLYTECHNIC

数据采集与治理

前导

数据采集与治理概述

主讲：赵娜





- 01** 数据分析的背景
- 02** 什么是数据分析
- 03** 数据分析的应用场景
- 04** 数据分析的流程
- 05** 为什么选择**Python**做数据分析
- 06** 常见的数据分析工具



数据分析的背景

随着计算机技术全面地融入社会生活，网络数据得到了爆发性地增长，驱使着人们进入了一个崭新的**大数据时代**。





思考：

数据库里面的数据这么多，怎么快速地拿到有价值的数据呢？

数据分析就可以从海量数据中获得潜藏的有价值的信息，帮助企业或个人预测未来的趋势和行为。



结论

不管你从事什么行业，掌握了**数据分析能力**，往往在其岗位上**更有竞争力**。



什么是数据分析？



思考：
什么是数据分析呢？

数据分析是使用适当的统计分析方法对收集来的大量数据进行分析，从中提取有用信息和形成结论，并加以详细研究和概括总结的过程。

- 数据分析的目的在于，将隐藏在一大批看似杂乱无章的数据信息集中提炼出来有用的数据，以找出所研究对象的内在规律。

在统计学领域中，数据分析可以划分为如下三类：

从一组数据中，可以摘要并且描述这份数据的集中和离散情形。

描述性数据分析

从海量数据中找出规律，并产生分析模型和研究假设。

探索性数据分析

验证科研假设测试所需的条件是否达到，以保证验证性分析的可靠性。

验证性数据分析



数据分析的应用场景

● 营销方面的应用

通过会员卡形式获得消费者的个人信息，以便对消费者的购买信息进一步研究其购买习惯，发现各有价值的目标群体。



● 零售方面的应用

在美国零售业曾经有这样
一个传奇故事，某家商店
将纸尿裤和啤酒并排放置在
一起销售，结果纸尿裤和
啤酒的销量双双增长！



● 网络安全方面的应用

新型的病毒防御系统可以使用数据分析技术，建立潜在攻击识别分析模型，监测大量网络活动数据和相应的访问行为，识别可能进行入侵的可疑模式。



● 交通物流方面的应用

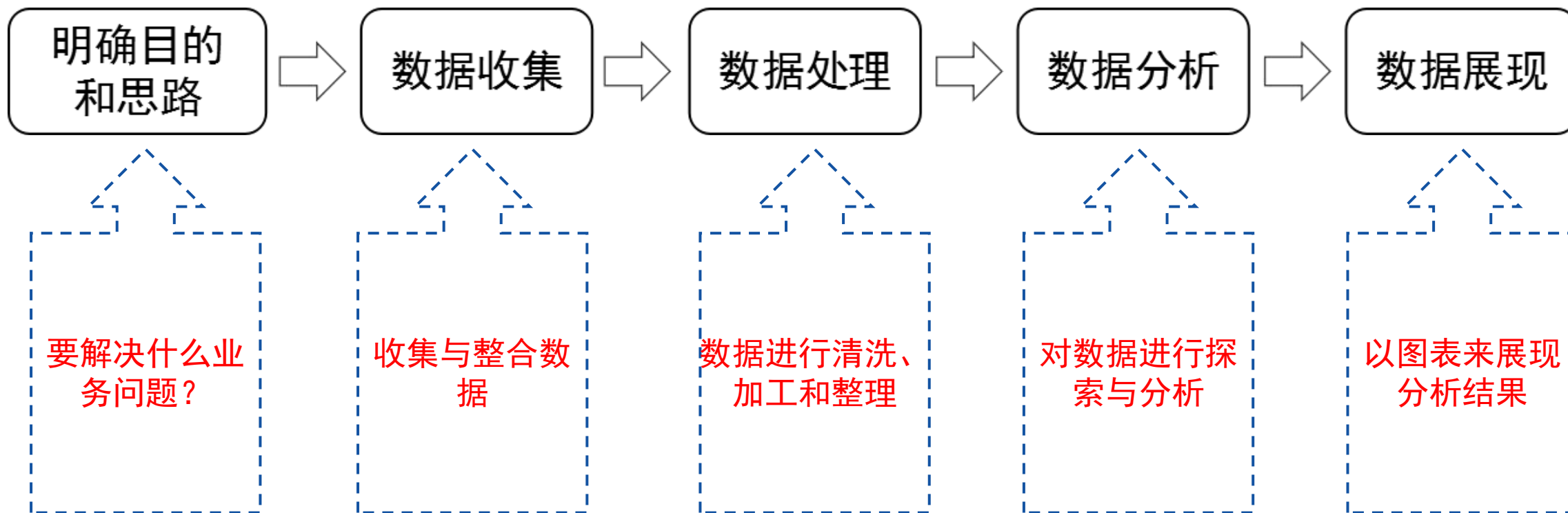
用户可以通过业务系统和GPS定位系统获得数据，使用数据构建交流状况预测分析模型，有效预测实时路况、物流状况、车流量、货物吞吐量，进而提前补货，制定库存管理策略。





数据分析的流程

数据分析大致可以分为以下五个阶段：



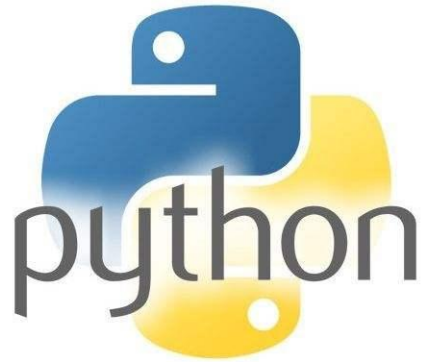


为什么选择Python做数据分析

为什么选择Python做数据分析



日照职业技术学院
RIZHAO POLYTECHNIC



思考?

为什么选择Python做数据分析?

选择Python做数据分析，主要考虑的是Python具有以下优势：

- 语法简单精炼，适合初学者入门
- 拥有一个巨大且活跃的科学计算社区
- 拥有强大的通用编程能力
- 人工智能时代的通用语言
- 方便对接其它语言



常见的数据分析工具

Python本身的数据分析功能并不强，需要安装一些第三方的扩展库来增强它的能力。

NumPy

Pandas

Matplo
tlib

Seaborn

NLTK



日照职业技术学院
RIZHAO POLYTECHNIC

感谢观看

主讲：赵娜

