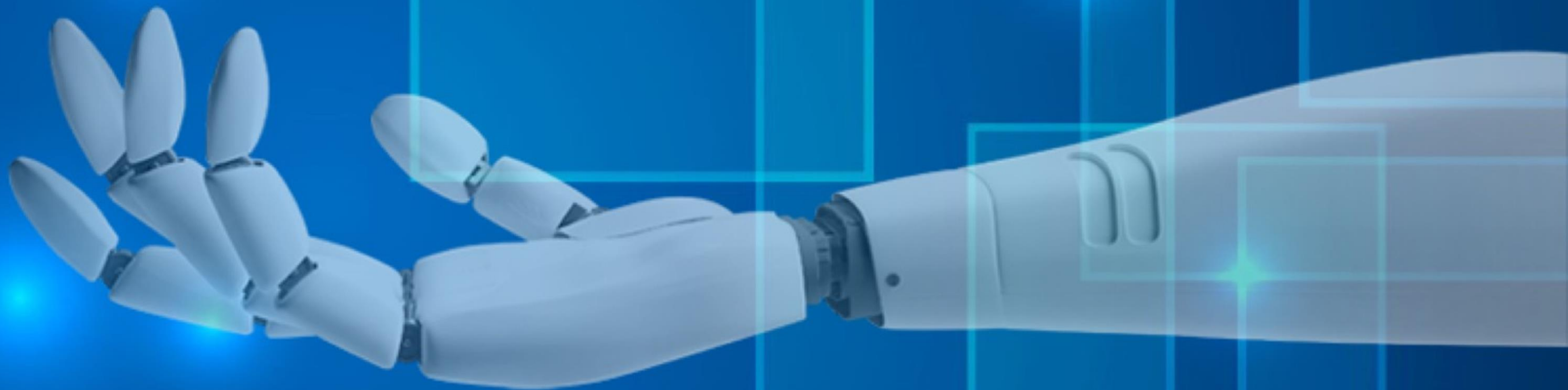


数据分析与可视化

数据分析概述



声音是数据吗？

音频设备可以采集声音，转化为音频数字信号，进而支撑很多有趣的应用，比如iphone的siri，搜狗的语音输入，微信的语音翻译，由此可见，在这个可以记录声音的时代，声音是一种数据，而且是一种具有强烈时代特征的数据

图像是数据吗？

数码成像技术的成熟可以让所有的图像都记录下来，而且分辨率很高，在此基础上，人们可以做进一步的分析与建模，进而支撑很多有趣的应用，例如：脸部识别，指纹识别，车牌号识别，在这个可以记录图像的时代，图像是一种数据，而且是一种具有强烈时代特征的数据

第一，凡是记录的都是数据；

第二，由于记录技术的进步，数据的定义具有强烈的时代特征

数据(data)是事实或观察的结果，是对客观事物的逻辑归纳，是用于表示客观事物的未经加工的原始素材。

1

理解数据

- 1、数据的类型
- 2、数据的呈现形式

达内教育研究院

数据的类型

数据类型的出现是为了把数据分成所需内存大小不同的数据，

在数据运算过程中，我们发现，数值、货币、日期与时间都可以进行加、减、乘、除等算术运算，所以统称为**数值型**；而文本只能进行简单的“计数”，不能进行算术运算，仍称**文本型**。

数据的呈现形式

1. 不同个体在同一标志上的不同取值某公司100名职工的月基本工资数据资料如图所示。

	A
1	基本工资
2	2390
3	2350
4	2380
5	2360
6	2430
7	2460
8	2440
9	2620
10	2170
11	2600

	A	B	C	D	E	F	G	H	I	J
1	100名职工的基本工资									
2	2390	2350	2380	2360	2430	2460	2440	2620	2170	2600
3	2650	2620	2640	2630	2320	2350	2380	2420	2490	2490
4	2510	2530	2540	2590	2620	2650	2700	2690	2370	2350
5	2380	2410	2440	2470	2500	2530	2560	2590	2620	2650
6	2680	2690	2320	2350	2380	2410	2440	2475	2500	2525
7	2560	2620	2630	2670	2690	2660	2320	2350	2380	2410
8	2440	2470	2500	2530	2560	2590	2620	2650	2680	2610
9	2320	2350	2380	2410	2440	2472	2500	2528	2560	2590
10	2620	2650	2680	2670	2320	2350	2380	2410	2440	2470
11	2500	2530	2560	2590	2620	2650	2680	2670	2370	2350

以理服人

数据的呈现形式

2. 不同个体在多个标志上的取值所组成的二维表格

在Excel中叫数据清单，如图所示

姓名	学校	年级	数学	物理	化学	总分
寸待杨	一中	高二	47	51	63	161
寸素香	二中	高三	36	41	63	140
寸静萍	四中	高一	98	80	73	251
寸德志	一中	高一	98	77	84	259
尹兴帅	一中	高二	97	78	100	275
尹兴松	二中	高一	69	73	64	206
尹丽蓉	二中	高三	84	75	66	225

理解数据

字段。 字段是事物或现象的某种**特征**，如“姓名” “性别” “籍贯” 和“居住地址” 等都是字段。

记录。 **记录**是事物或现象某种特征的具体表现，如姓名可以是张三、李四和王五，性别可以是男性或者女性。

数据表。 将字段和记录组织在一起形成的关系结构就是数据表。从概念中可以看到，记录、字段和数据表是一个有机的整体，三者之间相互依赖才得以存在。

		字段		字段		数据表				
	A	B	C	D	E	F	G	H	I	J
1	编号	部门	员工姓名	办公应用	电脑操作	管理能力	礼仪素质	企业文化	企业制度	总分
2	BH002	人事部	张嘉利	82	75	73	80	83	80	473
3	BH003	财务部	何时韦	71	91	82	83	70	84	481
4	BH013	人事部	兰慧芳	83	75	82	93	77	73	483
5	BH008	行政部	王豆豆	77	79	94	82	82	72	486
6	BH001	人事部	董天宝	90	95	72	84	72	79	492
7	BH005	财务部	钟嘉惠	93	83	71	78	81	86	492
8	BH009	后勤部	刘星星	79	94	79	78	82	81	493
9	BH006	后勤部	何思佺	83	80	80	83	88	80	494
10	BH012	后勤部	杨天雄	77		91	77	81	76	494
11	BH007	行政部	高雅婷	81		93	83	80	83	501
12	BH011	财务部	邓丽梅	76	88	87	91	80	83	505
13	BH004	后勤部	马田东	94	87	85	84	86	71	507
14	BH010	人事部	赵大宝	86	73	93	93	70	94	509

记录

员工考核成绩表

以理服人

2

什么是数据分析？

- 1、数据分析的定义
- 2、数据分析流程
- 3、数据分析价值

达内教育研究院

最专业的数据分析：

有针对性的收集、加工、整理数据，并采用统计和挖掘技术分析和解释 数据的科学与艺术！

最简洁的数据分析：

简单的很，就是分析数据。

描述数据特征，预测数据趋势，展示分析结果，

从一大堆数据中提取到你想要的信息，就是数据分析。

数据分析是指用适当的**统计分析方法**，对收集来的大量数据进行分析，隐藏在数据背后的信息提炼出来，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

数据分析的过程主要包括6个既相对独立又相互联系的阶段，下面分别进行介绍。

1. 提出分析目标
2. 收集数据
3. 数据处理
4. 数据分析与建模
5. 数据可视化
6. 提出建议并推动落地

1、提出分析目标

只有明确目的，数据分析才不会偏离方向，否则得出的结果不仅没有指导意义。

现状统计：利用统计学知识进行数据统计，比如按月/产品统计销售额等

问题定位：波动解释型：某天的销售额突然下降了，某天的新用户留存突然降低了，这时候往往需要分析师去解释波动的**原因**，数据复盘型：类似于月报、季报，在互联网领域常见于app某某功能上线了一段时间后，数据分析师往往需要复盘一下这个功能的表现情况，看看有没有什么问题

预测未来：预测销售额，预测房价，预测客户是否会流失，预测客户是否会违约等

2. 数据获取

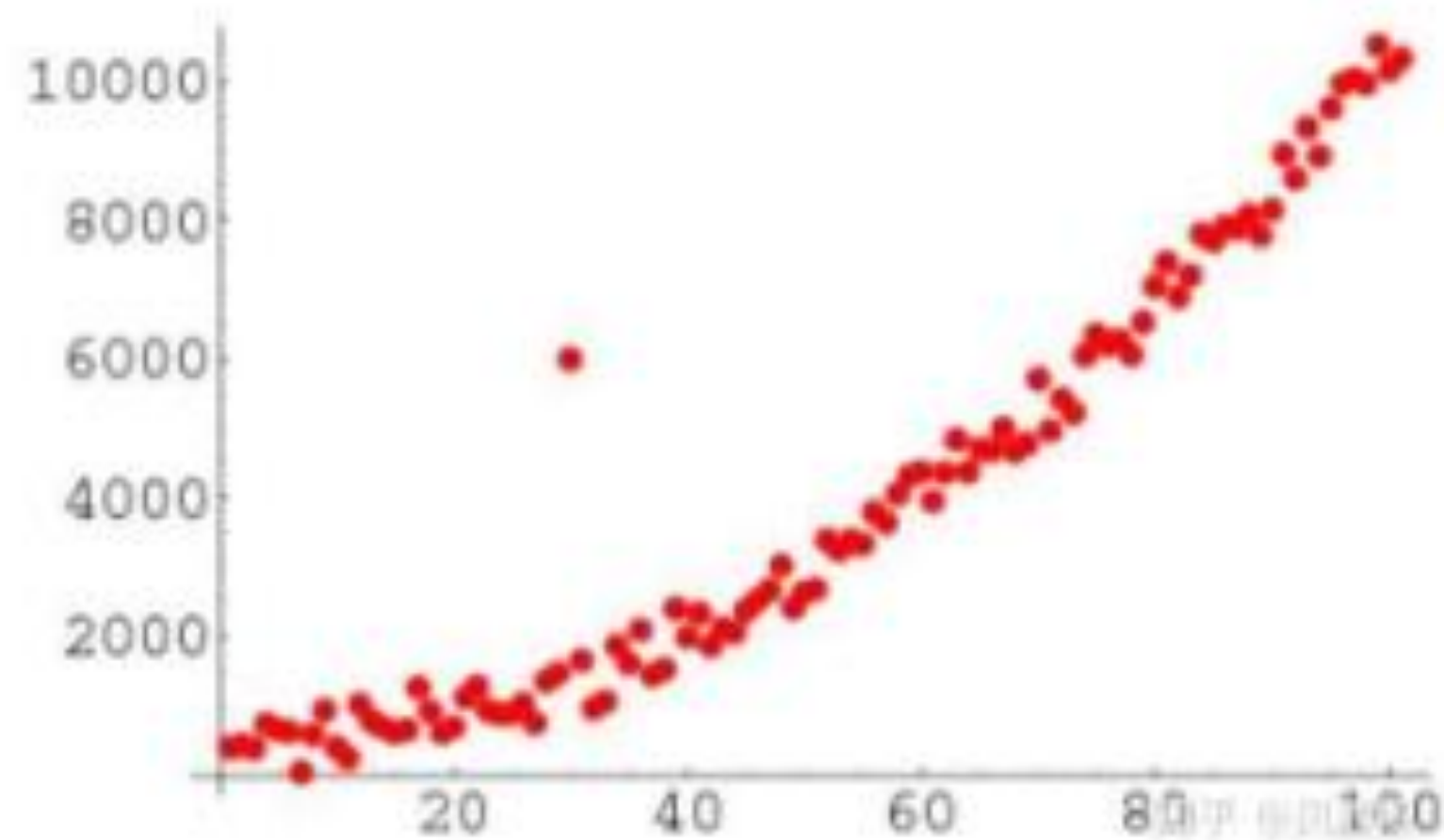
- (1) 内部数据：业务数据，客户资源，在职员工信息等
- (2) 网络数据：国家相关部门统计信息，各种信息查询平台及各种网站
- (3) 市场调查：运用科学的方法，有目的的系统搜集，记录和整理有关市场营销的信息和资料，分析市场情况，了解市场现状以及发展趋势，为市场预测和营销策略提供客观和正确的资料
- (4) 网络爬虫/第三方统计工具

3. 数据处理

对数据进行数据合并，数据清洗，数据转换和数据标准化。

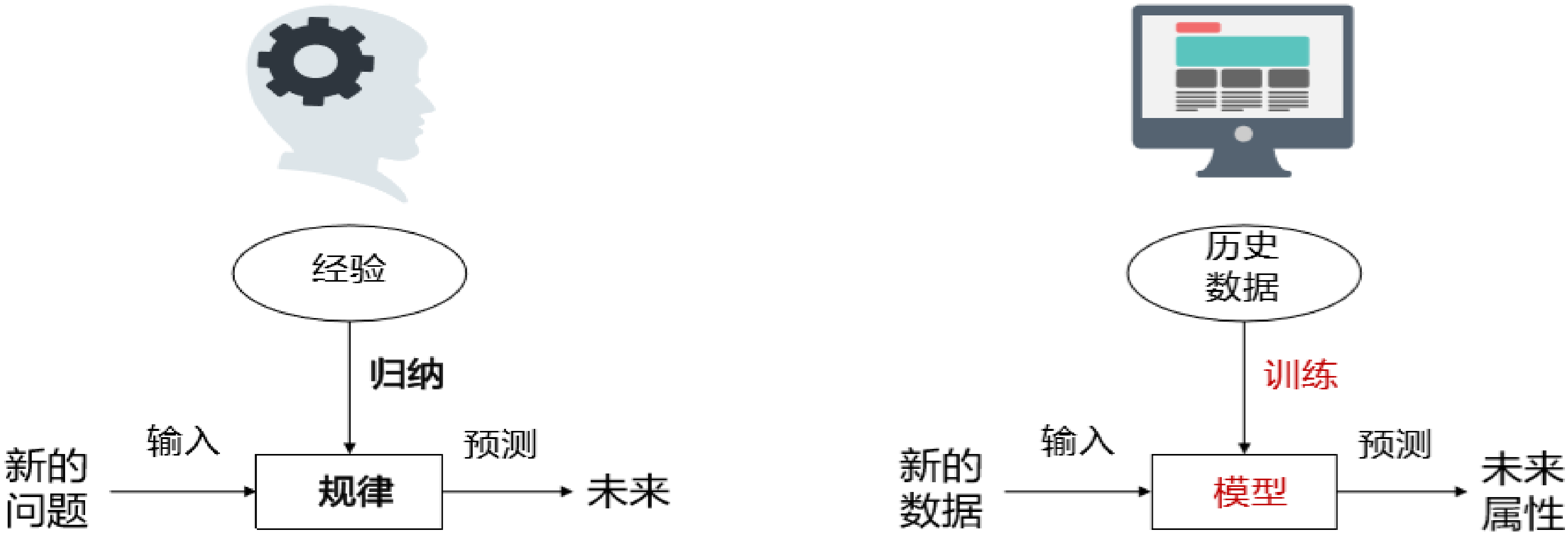
数据处理之后使得整体数据变为干净整齐，可以直接用于下一步的分析。举例如下：

- 数据缺失：数据无法获取或遗漏，若使用空值会对分析造成影响
- 异常数据：数据统计错误或者数据不合常理



4、数据分析与建模

- 简单分析方法：数据分析思维+业务指标+描述性统计分析+探索性数据分析
- 深层业务逻辑建模分析：数据挖掘，使用算法搭建模型，分析完成复杂的数据分析工作，重点挖掘数据价值，寻找模式与规律。

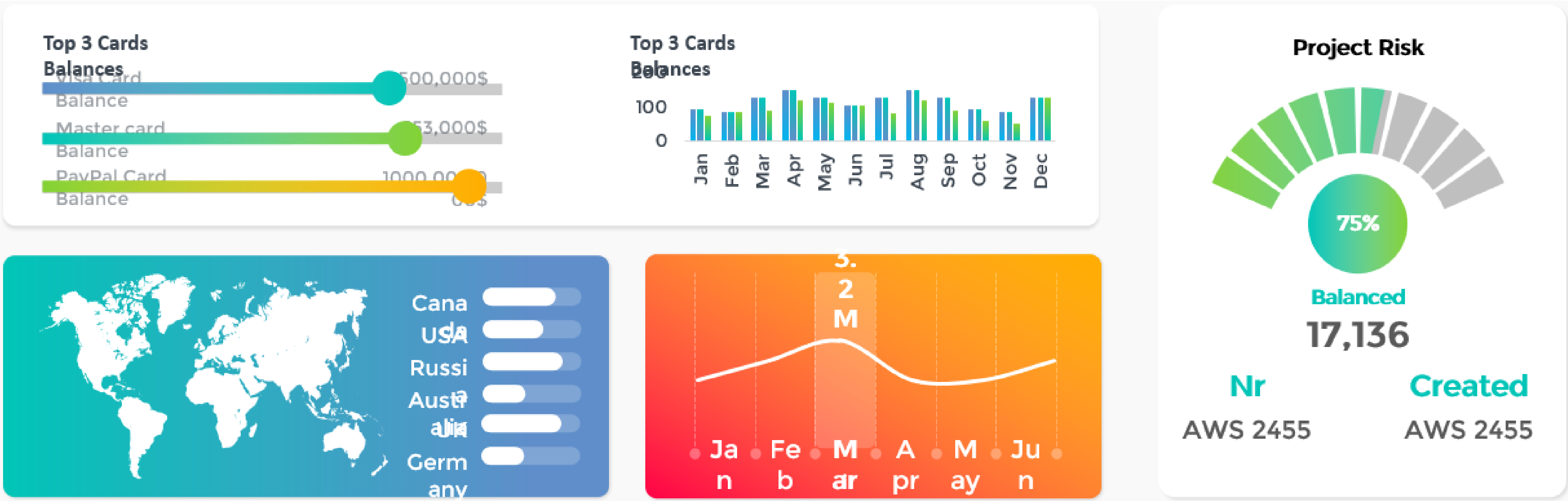


以理服人

5、数据可视化

通过数据分析得出结论后，还需要用图表展示出来，俗话说得好，“文不如表，表不如图”，用图表可以更清晰展现你的结论。

以理服人



6、提出建议推动落地

基于分析目标得出结论后，还应根据结论提出相对应的改进建议，并推动建议落地，这样才能完成一个完整的数据分析闭环。

数据分析的过程主要包括6个既相对独立又相互联系的阶段。



以理服人

商业价值

提高收入、降低成本、预测风险

以理服人