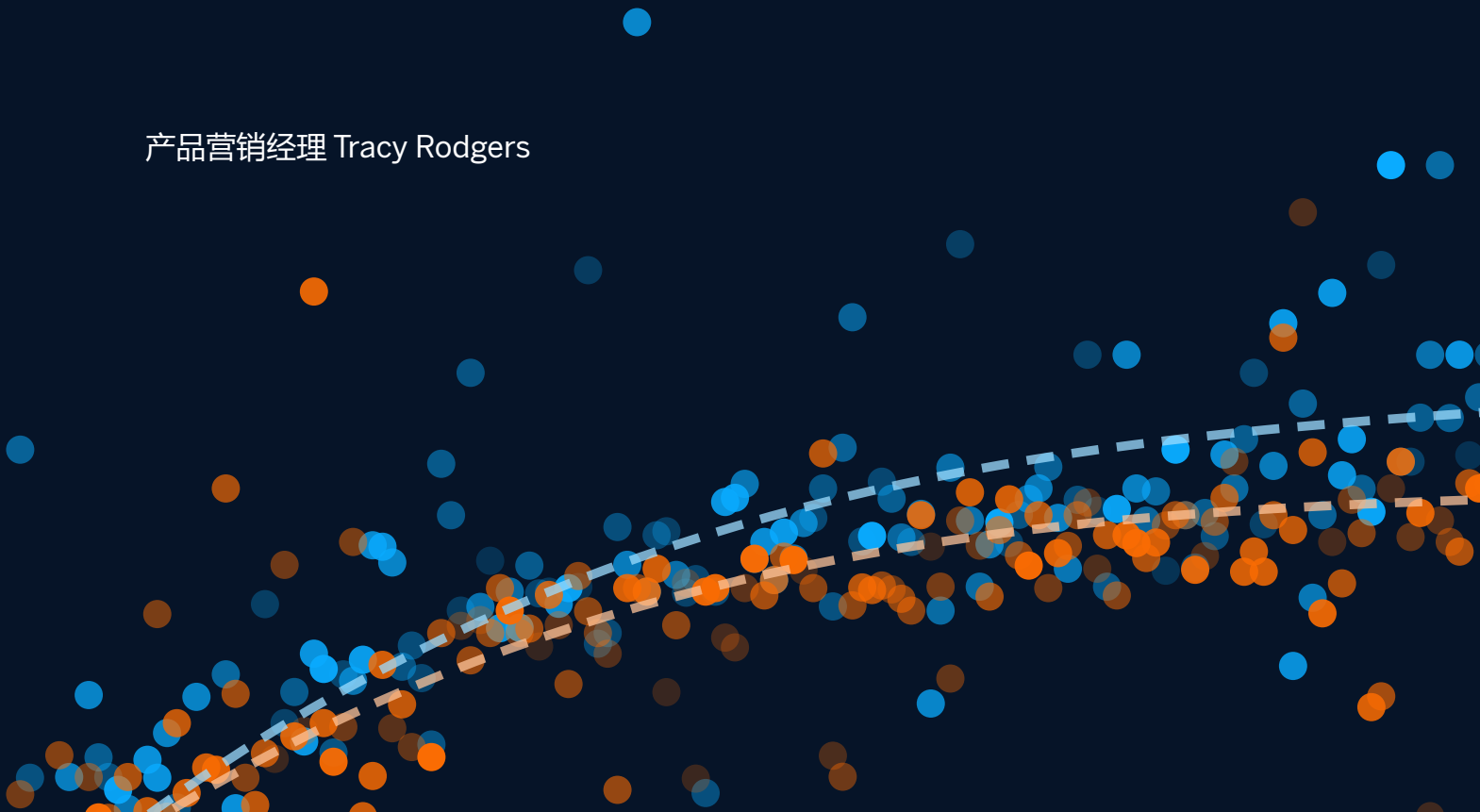




使用 Tableau 进行高级分析

产品营销经理 Tracy Rodgers



简介

我们曾经生活在一个非此即彼的世界。您要么懂得如何编程，要么与高级分析技术无缘。要么学习 R、Python 和/或 SAS，要么请人帮您攻克难关。Tableau 相信，为了真正地扩充人类智能，我们需要为技术能力千差万别的用户提供丰富的功能。这其中包括高级分析功能。

在 [Netflix](#) 和 [Spotify](#) 之类的前沿分析公司，Tableau 的分析深度又足以扩充数据科学团队的工作流。只需单击几下，您就可以创建盒须图、树状图、预测性可视化和同期群分析。还可以连接到 R、MATLAB 和 Python，使用 Tableau 扩展分析，对模型结果进行可视化。非技术用户可以提出此前难以企及的问题，而数据科学家可以更快地迭代和发现更深的见解，因此能够获得更有价值的发现。

该白皮书探索 Tableau 如何能够一方面为分析项目的各个阶段提供帮助，另一方面又专注于几项具体的高级功能。

我们将讨论以下方案以及为这些方案提供支持的功能：

分段和同期群分析：Tableau 提倡直观的调查流程，以实现快速灵活的同期群分析。通过根据所需维度切分和切割数据，转换不同的视角探索问题。如果您希望用机器学习来扩充自己的分段分析，可借助自动化聚类分析从高维度大数据集中发现原本很难察觉的模式。

方案和假设分析：通过将 Tableau 灵活的前端和强大的输入功能结合使用，您可以快速修改计算并测试不同的方案。

复杂计算和统计函数：Tableau 拥有强大的计算语言，因此您可以轻松地通过任意计算扩展分析，并利用简洁的表达式执行复杂的数据操作。范围涵盖基本聚合内嵌和统计计算（例如标准差、众数、峰度、偏度、相关性和协方差）。

时间序列和预测分析：世界上的大部分数据都可以按照时间序列进行建模，因此 Tableau 自身支持多种时间序列分析。也就是说，您可以在强大的 UI 中探索季节性变化和趋势，对数据进行抽样，运行预测性分析（例如预测）以及执行其他常见的时间序列运算。

外部服务集成：R、MATLAB 和 Python 集成为 Tableau 前端提供了易用的强大功能，同时使专家能够利用其他平台上现有的工作，并处理有细微差别的统计和机器学习需求。

目录

分段和同期群分析	4
聚类分析.....	5
集和集动作.....	6
群组.....	7
假设和方案分析	9
参数.....	9
Story points.....	10
复杂计算	10
详细级别表达式.....	10
表计算.....	11
时间序列和预测分析	12
时间序列分析.....	12
预测.....	14
外部服务集成	16
Python、R 和 MATLAB 集成.....	17
结语	19
关于 Tableau	20
相关白皮书	20
了解其他资源	20

1.分段和同期群分析

功能：聚类分析、集和集动作、群组

为了生成最初的假设，业务用户和数据专家经常以相同方式着手 - 创建分段和/或执行非正式同期群分析。

通过对不同的分段提出一系列基本问题，分析师可以更好地理解自己的数据、验证自己的假设。例如，“与使用支票付款的客户相比，使用信用卡付款的客户是否更易于保留？”快速迭代的能力有助于推动模型开发，确保项目始终在正确轨道上运行。

Tableau 具有丰富的功能，可以对分段实现快速迭代分析和比较。例如，只需通过一些计算字段和拖放操作，即可创建根据日期和时间划分的患者数量细分仪表盘，并进一步探索基于科室的趋势（图 1）。

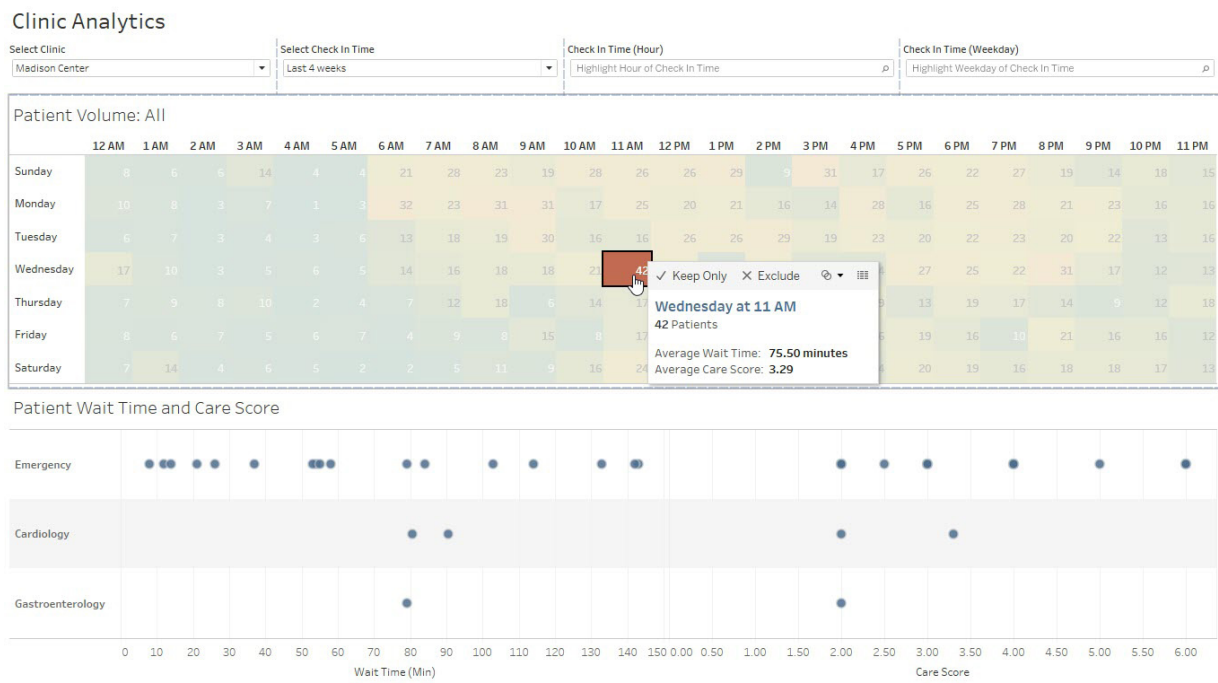


图 1：此交互式临床仪表盘显示特定日期和时间的患者数量、就诊原因以及等候时长。

聚类分析

Tableau 中的**聚类分析**是另一个使用无人监督的机器学习来分割数据的选项。聚类分析特别适合需要考虑很多变量的情况。

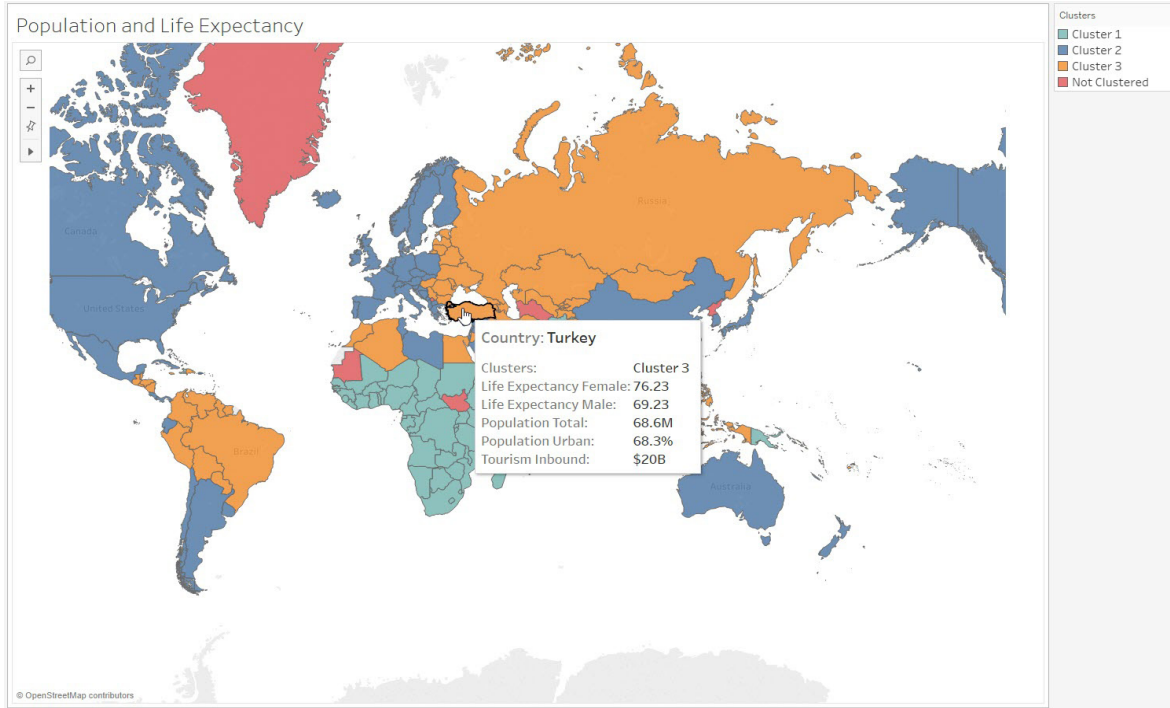


图 2：使用聚类分析自动分割数据。

在图 2 中，您可以看到使用联合国发展指标后，对各国/地区进行聚类分析获得的结果。该算法建议使用三群集解决方案，三个群集分别对应不发达国家/地区、发展中国家/地区和发达国家/地区。

借助 Tableau 的灵活界面，我们还可以轻松地测试不同的理论，探索同期群分布。借助可视化迭代功能，您无需像使用其他工具时一样调整脚本和重新运行模拟，从而可以节省时间。

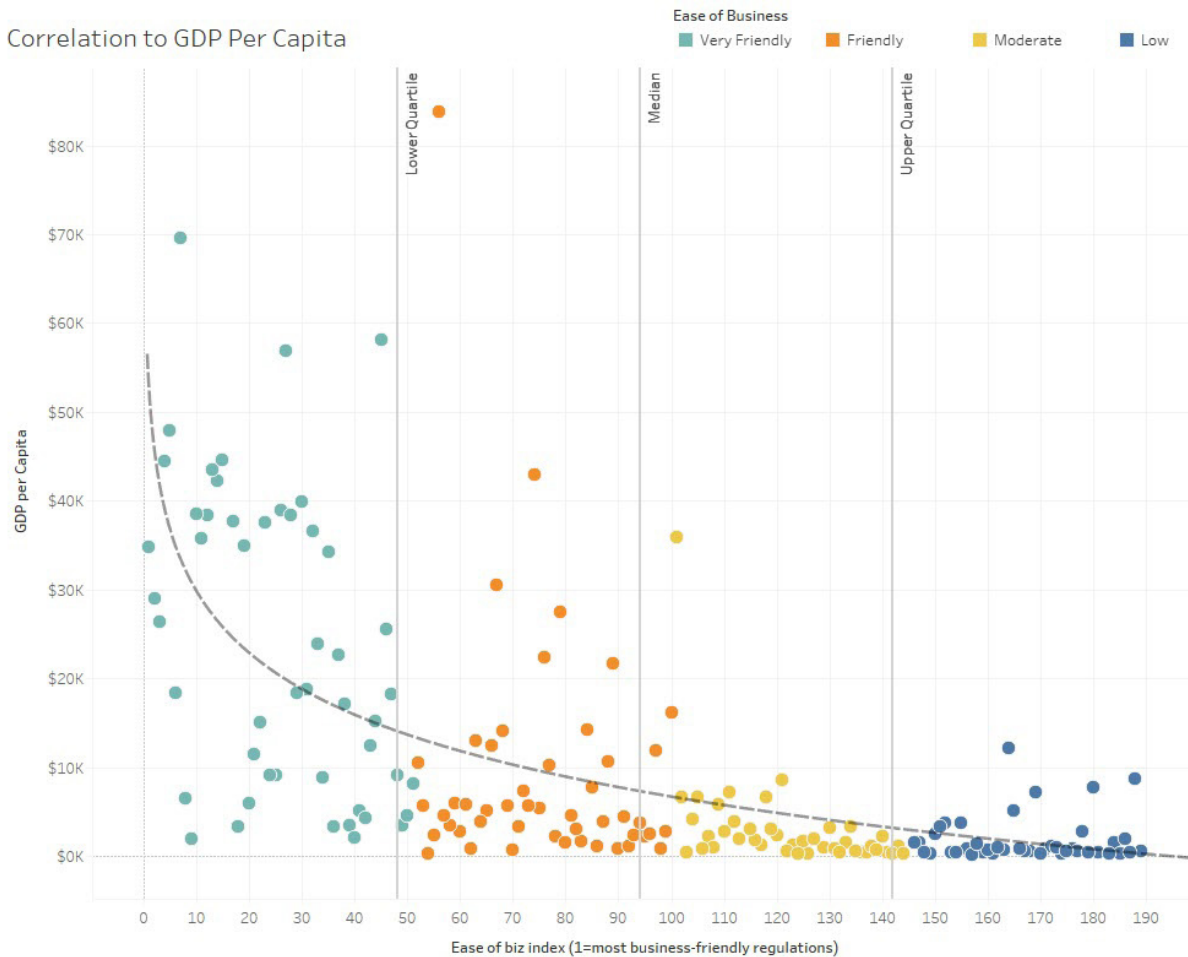


图 3：在短短几秒内分割和查看数据在此示例中，您可以看到根据“有利营商法规”和人均 GDP 划分的国家/地区。

如图 3 所示，只需将在聚类分析中使用的分段字段和度量拖至画布，即可生成查看数据的新方式，突出显示四个国家/地区群集之间的差异。

集和集动作

借助集，您可以通过手动选择或编程逻辑来定义数据对象集合。集可用于一系列方案中，包括筛选、突出显示、同期群计算以及异常值分析。您还可以通过合并多个集来测试不同的方案，或者通过创建多个同期群来进行模拟——例如，通过合并独立生成的不同客户组来进行客户保留分析，或者应用多个连续标准。

集动作功能是指可在集中存储可视化的一系列数据点。任何工作表中的用户所选内容可以更新集，并修改经过协调的单个所选内容中的所有目标工作表。这可实现比例刷（图 4）等多种用例。

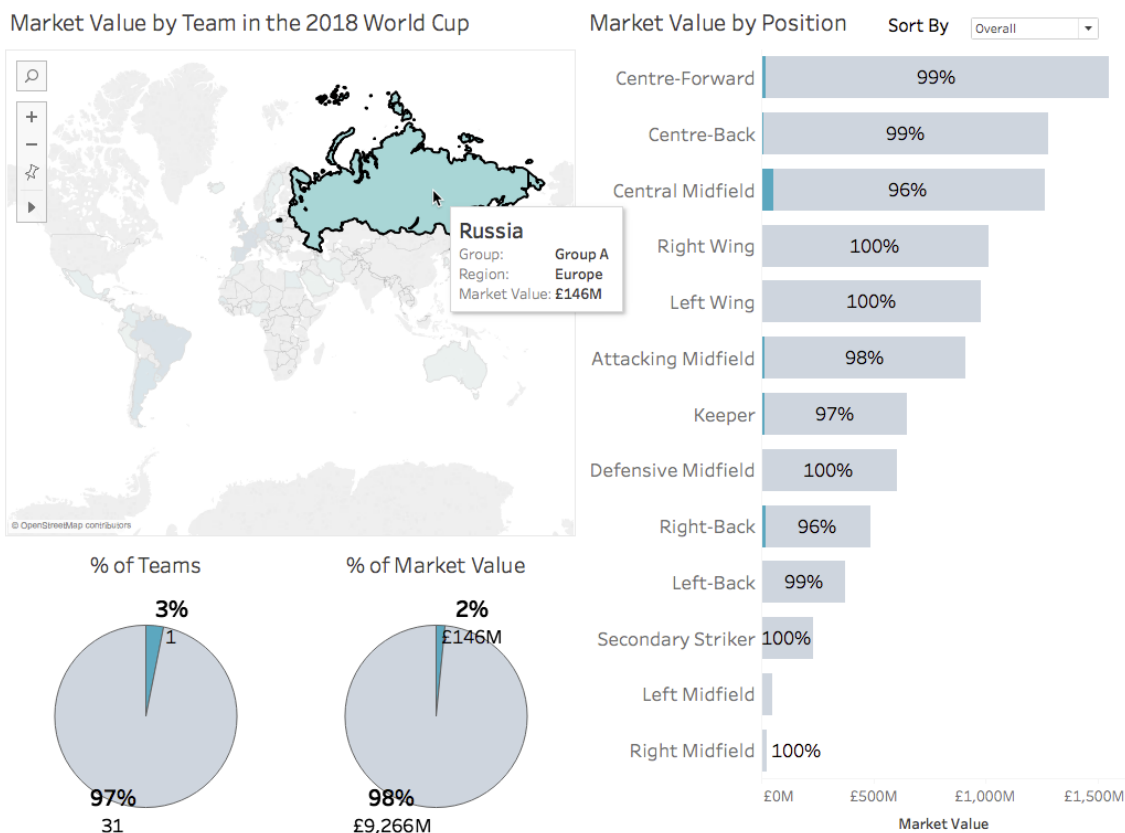


图 4：集动作可实现比例刷，比例刷是一种交互式分析技术，可以显示所选内容相对于总幅度的幅度。选择“俄罗斯”会显示世界杯中，按场上位置划分的该国球员市场价值。

群组

为了支持创建临时类别和建立分层结构的需求，Tableau 提供了名为“组”的功能。组还有助于满足基本的数据清理需求。

借助组，用户可以通过直观方式为现有的分析任务创建数据结构，将相关成员合并在一个字段中。例如，如果您要使用显示按专业划分的平均测验分数的视图，可能需要将某些专业组合在一起，以创建类别。英语和历史可以合并到名为“人文科学专业”的组，而生物和物理可以合并到名为“科学专业”的组。通过这种操作，分析师可以对演示进行自定义，并控制整个分析中的数据聚合。

此外，组还可以用于解决数据的一致性和质量问题。例如，加利福尼亚州可能用全称表示，也可能用 CA 或 Calif 来表示，由于没有相应权限，分析师和业务用户常常无法通过直接更改源系统来解决问题，这就意味着微小的数据错误可能对解释性分析造成极大阻碍。如果因为要提出数据更改请求而不得不停止提出问题，项目就会被延迟，创意的快速部署就会被中断。借助组，您可以快速定义一个包含所有别名的新分段，将其用于自己的分析，这样就可以继续提出问题，保持流程的连续性。

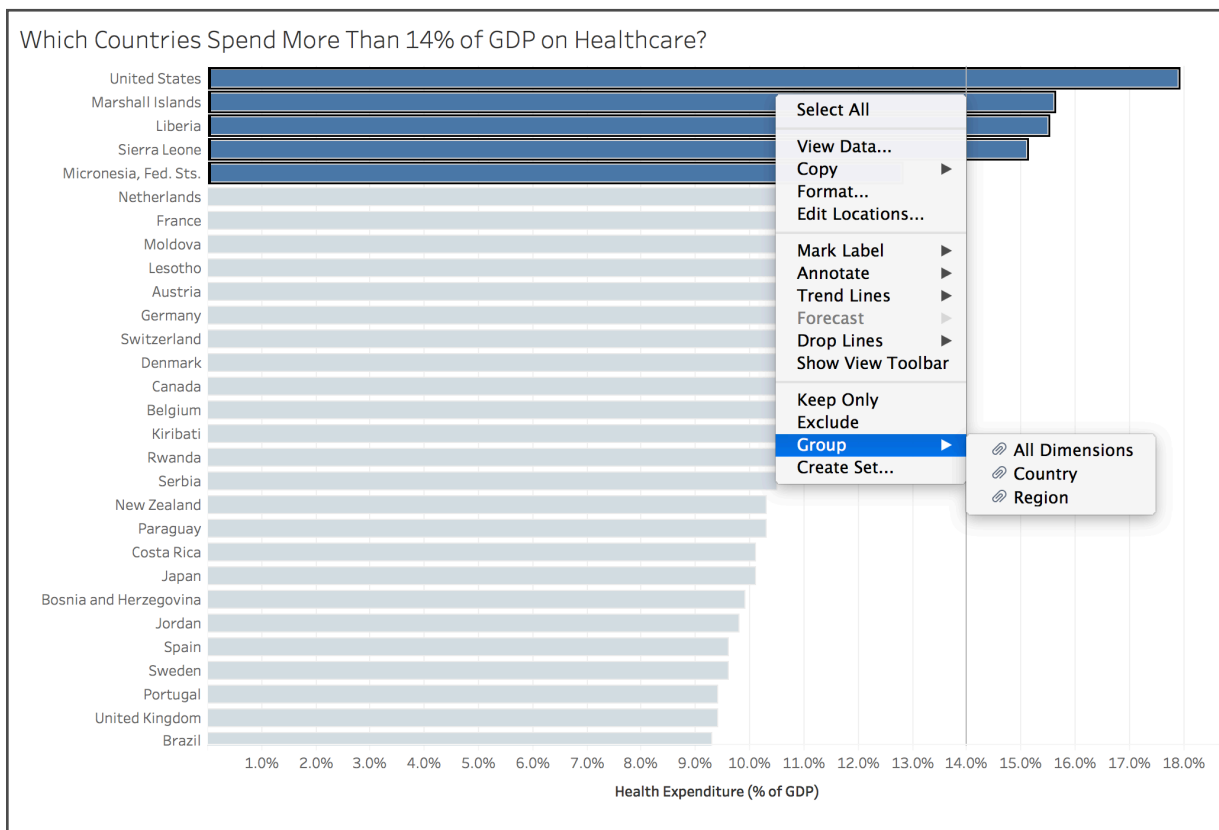


图 5：创建组，将相关成员合并到一个字段中。在此示例中，您可能需要将在医疗保健支出方面 GDP 占比最大的五个国家/地区分为一组。

所有这些功能都继承了简单的更新机制。在 Tableau 中，如果您选择实时连接并更新自己的数据，您的分析及所有基础组件（例如集和组）也会更新。也就是说，同期群成员会自动更新，无需以手动方式重新运行报告或其依赖的脚本。简单的更新机制有助于减轻报告负担，并且提供了另外一种方案测试方法。借助简单更新，您能够在无需更新分析堆栈的情况下，通过换出基础数据来探查对初始条件的敏感性。

Tableau 允许用户快速分割数据并对其进行分类，从而让业务用户和分析师等都可以更为轻松地执行强大的同期群分析。这些功能还可帮助数据科学家调查初步假设和测试方案。

2.假设和方案分析

功能：参数、Story Points

用户有时希望探索某一个值或某一组值的改变，看其如何影响分析的输出。这可以用于测试不同的理论，为同事突出显示重要方案，或者调查新业务的可能性。借助 Tableau，您可以对分析输入进行尝试。通过使用简单控件，您可以轻松添加标记并共享方案，同时保持数据的最新状态。

参数

进行假设分析时，您可能需要更改计算的基准值、重新定义配额，或者设置初始条件。Tableau 中的参数让您可以轻松实现上述目的。您可以通过定义参数，提供一种更改模型或仪表板输入值的方法。参数可用于驱动计算、更改筛选器阈值，甚至选择要引入仪表板的数据。借助参数，非技术用户可以对不同输入进行实验，探索复杂模型的可能输出。

Tableau 的参数功能不但可以帮助您测试假设，还让您能够在交互式报告中展示假设分析的结果。在图 6 中，参数驱动关于销售佣金的假设分析。销售经理可以使用不同的佣金率、基本工资和配额进行实验，并在此过程中获得有关这些更改对关键指标影响的实时反馈。

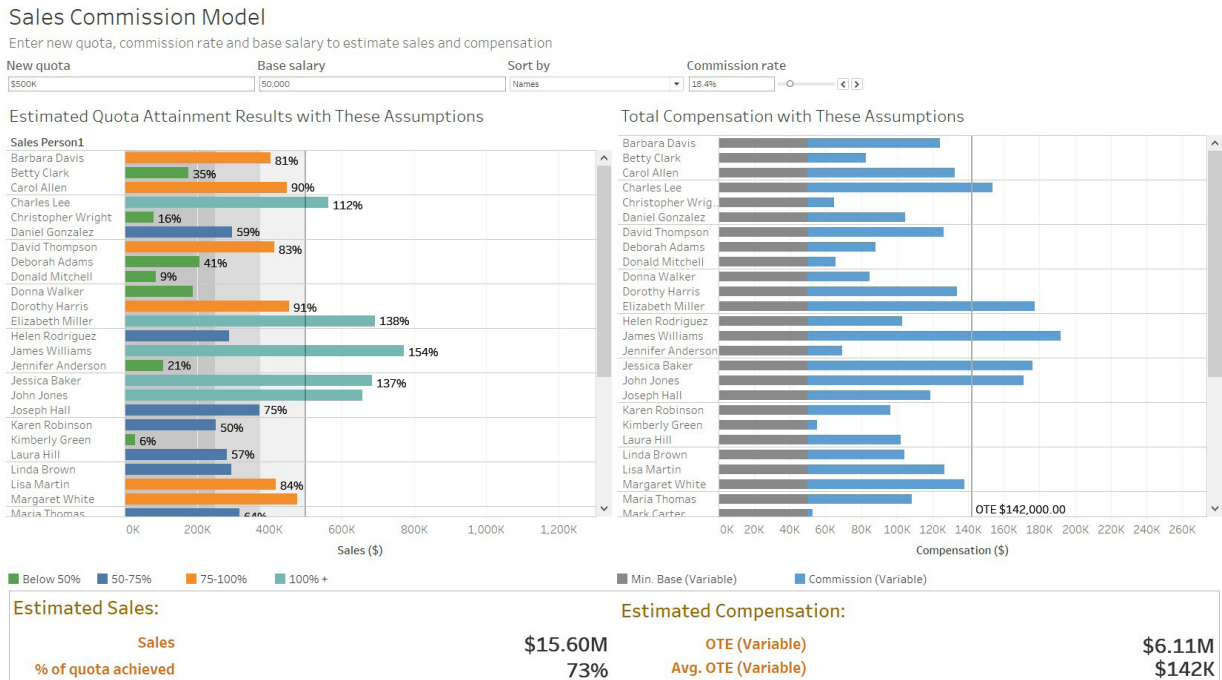


图 6：使用此参数驱动型销售报表，交互者可以掌握配额、佣金和工资在组织内产生的影响。

Story Points

参数可以和 **Story Points** (Tableau 使用数据构建叙述内容的方法) 配合使用, 帮助您获取有趣结果的快照并持续进行探索。借助 Story Points, 您可以构建一个随着数据变化和可视化修改不断更新的演示。然而, Story Points 具有足够的智能, 能够保留参数值, 因此您可以标记方案, 并确信自己可以在不中断分析流程的情况下返回这些方案。您还可以比较多组不同输入的结果, 无需担心过时的屏幕截图, 也无需重新运行模拟。

凭借集、组、拖放式分段和参数, Tableau 实现了从理论和问题到专业外观仪表板的转换, 甚至允许非专业人员提出问题并测试自己的方案。假设分析的简化让数据专业人员能够专注于分析中更加复杂的层面, 提供质量更高的见解, 而直观可视化让最终用户能够使用数据来提高参与度。更高的参与度有助于在整个组织中促进变革, 改进决策过程。

3. 复杂计算

功能: 计算字段、详细级别表达式、表计算

源数据通常并不包含全面分析所需的所有字段。分析师需要使用一种简单但功能强大的语言来转换数据和定义复杂逻辑。

虽然 Tableau 易于使用, 但它也具有支持复杂逻辑的强大功能。借助 Tableau 中的 **计算字段**, 您可以根据数据源中已存在的数据创建新数据。借助计算字段, 您可以轻松执行算术运算、表达条件逻辑, 或对特定数据类型进行特殊化操作。两种计算字段可以实现高级分析, 它们分别是详细级别 (LOD) 表达式和表计算。

详细级别表达式

LOD 表达式是 Tableau 计算语言的扩展, 是回答涉及单个可视化中多粒度级别问题的强大方法。借助 LOD 表达式, 我们可以通过非常简洁的表达式处理许多以前无法或很难处理的方案。

详细级别表达式大大简化了同期群分析 (如上文所述) 和多遍聚合。图 7 显示针对按第一个订单日期分桶为“新”和“现有”的客户同期群, 按区域市场划分的客户获取运行总和。在关于时间序列分析的下一部分中, 我们将了解这种计算语言通过哪些其他层面的功能来实现这种分析。该图表显示从一月开始直到全年剩余月份, 北美客户获取的下滑情况。通过详细级别表达式, 原本需要使用 SQL 复杂分组依据语句的分段, 现已转化为简单直观, 并能够在 Tableau 前端编辑和调整的表达式。

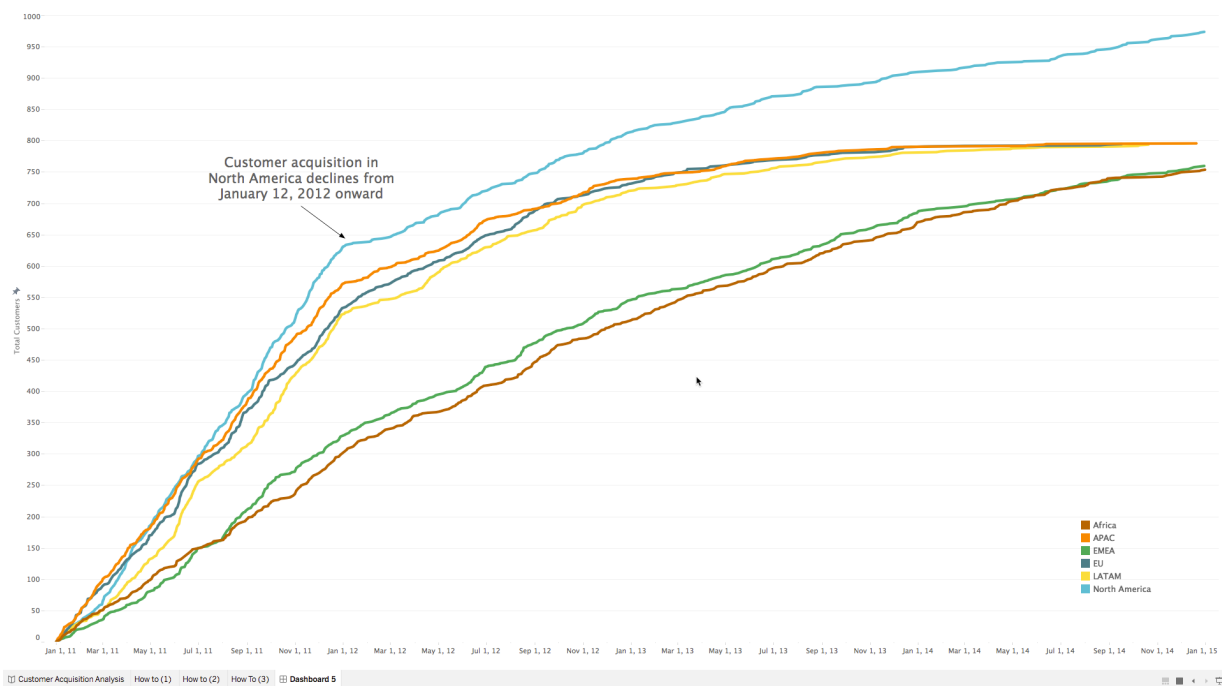


图 7：此可视化使用 LOD 表达式查看各市场总体客户获取的日常趋势。LOD 表达式可确保我们不会将同一客户错误地算作新客户。

表计算

Tableau **表计算**可以实现具有相对性质的计算。更具体地讲，表计算是应用于表中所有值的计算，常常依赖于表结构本身。这种类型的计算包括许多时间序列运算，例如滞后或运行总和，但也包括排序和加权平均之类的计算。

在 Tableau 中，可以通过两种方法使用表计算。第一种方法涉及一组常用的表计算，称为**快速表计算**。它们让您能够通过一键式操作定义表计算，因此是很好的入手点。您还可以借助**表计算函数**创建自己的表计算。借助这些函数，工作簿作者可以对自己的结果集进行精确操控。另外，由于所有表计算都可以使用同一计算语言表达，您可以选择一个快速表计算作为起点，并根据需要对其进行手动编辑，增加其复杂度。

借助表计算，只需点击几次或使用简单的表达式，即可完成复杂的数据库工作，例如操纵聚合数据和创建依赖于数据结构的聚合。两者都允许非技术用户自行操作、为专家们节省了大量时间，还可避免编写耗时的 SQL 代码。

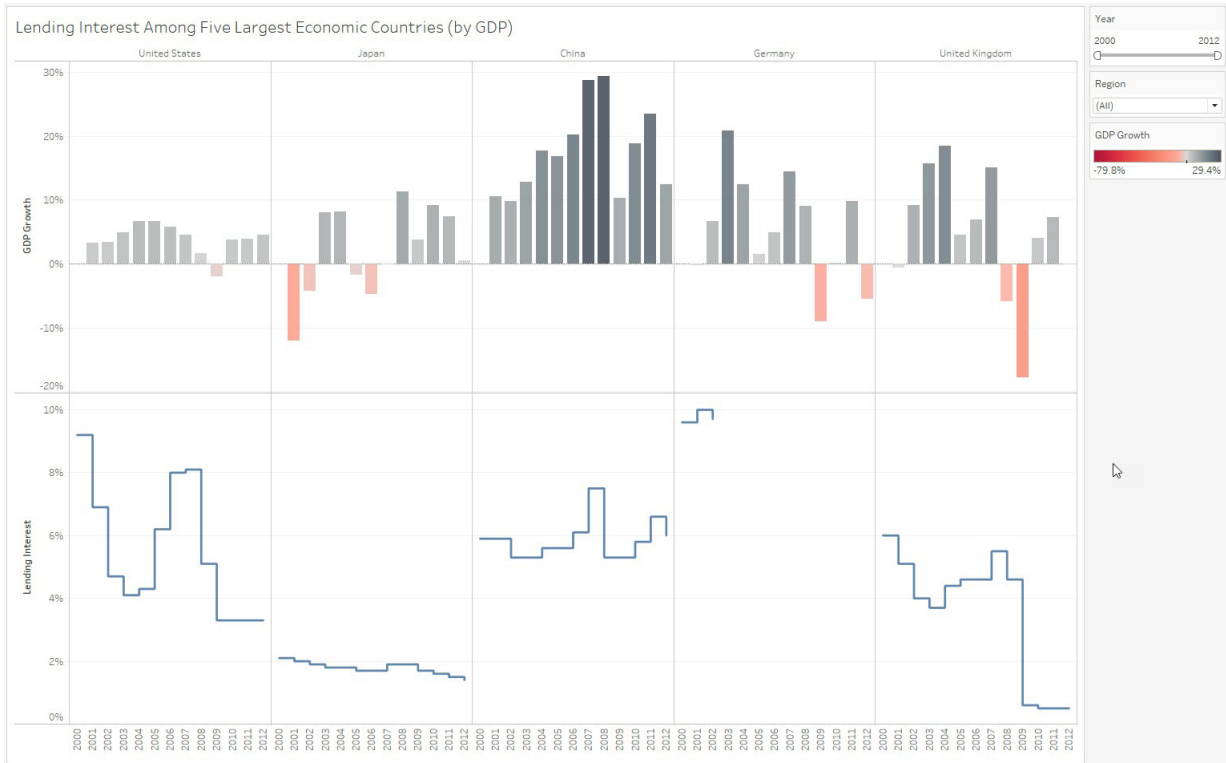


图 8：通过使用 Tableau 内置的快速表计算（百分比差异）之一，可以快速查看利率上升对一些国家/地区 GDP 的影响。

4. 时间序列和预测分析

功能：时间序列分析、预测

从传感器读数到股票价格和毕业率，世界上的大部分数据都能够以时间序列的形式进行有效建模。因此，时间是分析项目中最常用的独立变量之一。为了获取更多见解，您可能需要利用预测性功能。借助 Tableau，您无需大费周折就能够添加预测性分析，这样您便可以快速探索多个方案。

时间序列分析

在 Tableau 中，得益于灵活的前端和功能强大的后端，只需提出正确的问题即可进行时间序列分析。要开始分析，您只需将关注的字段拖至视图，然后开始进行询问。

您可以利用双轴功能和离散化聚合，查看多个时间序列。如果不能快速检查不同粒度和聚合级别的时间序列，您可能无法提出这个问题。

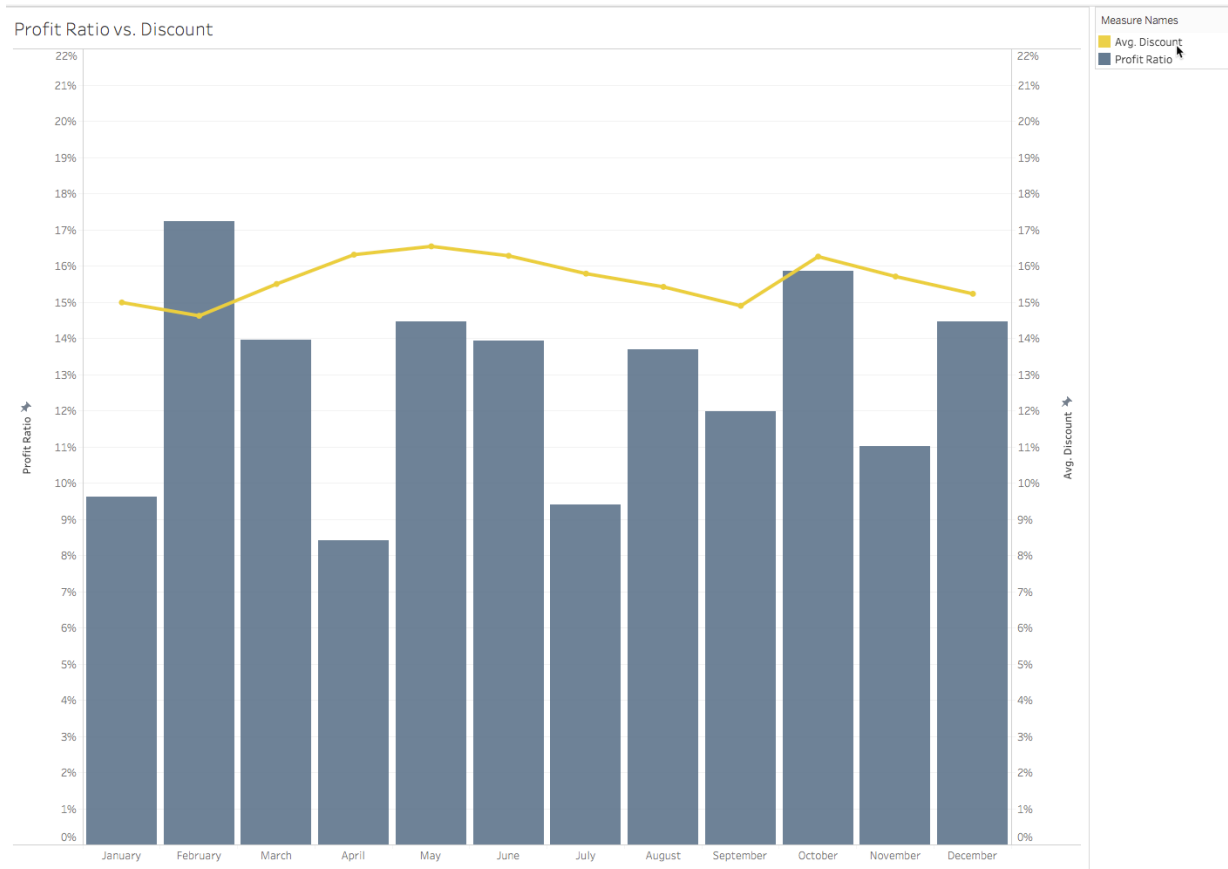


图 9：双轴图显示利润率和平均折扣之间的关系。

要查看特定时间段，您可以使用一组精确日期来筛选数据，也可以利用 Tableau 的相对日期筛选器。借助相对日期筛选器，您可以查看相对时间段，例如上周或上月。这些时间段会在您每次打开视图合并新数据时更新，因此是功能强大的报告工具。

使用时间序列时，我们常常需要进行平滑处理，或者执行其他时序计算。Tableau 提供了一组丰富的功能来简化常见的时序运算，例如移动平均、年同比计算以及汇总。

如前文所述，Tableau 的表计算功能让您既可以从一组常用的时序运算（快速表计算）中进行选择，也可以编写自定义计算。

Stock Tracking

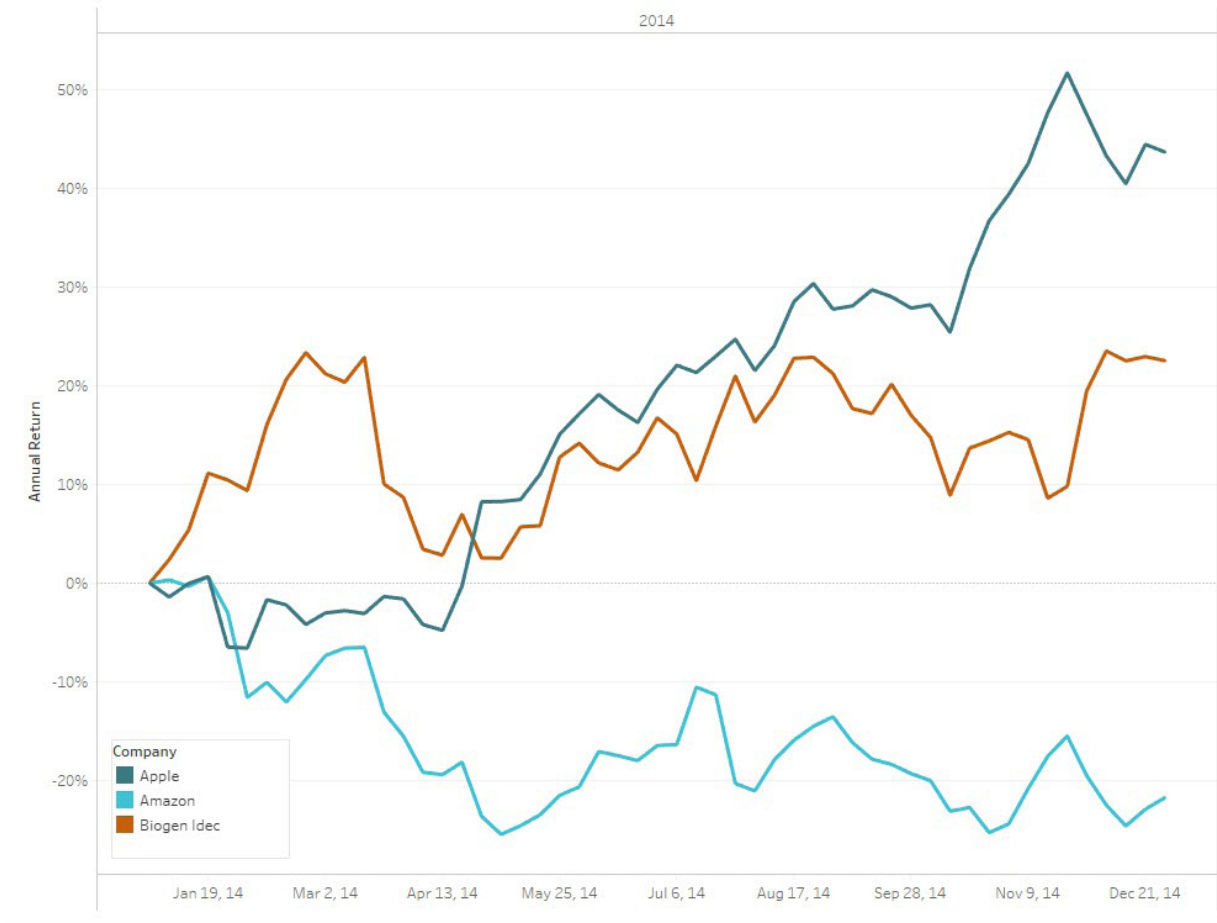


图 10：此时间序列分析显示三家大型公司的每月股票价格。

时间序列分析极为常用，因此 Tableau 的功能可以帮助您更快完成项目，为组织提供更多价值。借助直观的功能，数据专家和业务分析师都可以针对自己的数据提出更多、更好的问题。

预测

如果想展望未来，您可以利用 Tableau 提供的几种本机建模功能，包括趋势分析和预测。

只需右键单击，您就可以向任何图表快速添加趋势线，并查看描述拟合情况的详细信息（例如 p 值和 R 平方值）。您可以使用 Tableau 的拖放功能，通过单键操作对不同的组进行建模，趋势线已经完全集成到前端，可以轻松进行分段。如图 11 所示，Tableau 无需任何代码即可自动为不同的分段创建三条趋势线。Tableau 还支持几种其他类型的拟合，包括对数拟合、多项式拟合和指数拟合。

Obesity's correlation with diet, exercise, and smoking (% of pop.) - All

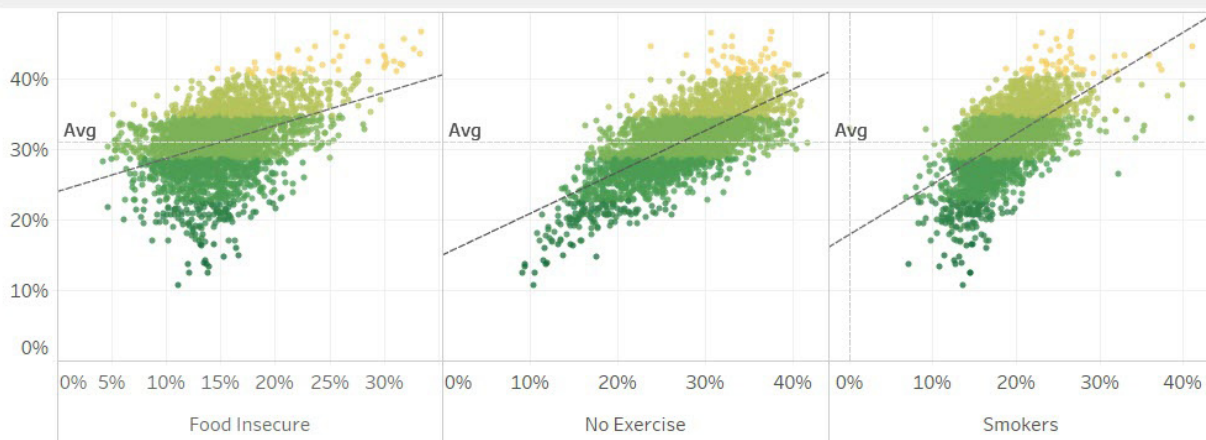


图 11：趋势线突出显示肥胖与食品不安全、锻炼和吸烟者之间的关系。

默认情况下，Tableau 的预测功能会在后台运行多个不同的模型，然后选择效果最好的模型，并且会自动考虑季节性等数据问题。Tableau 使用名为“指数平滑”的技术进行预测。“指数平滑”根据时间序列过去的值的加权平均数，以迭代方式预测将来的值。如前所述，预测功能的几乎所有设置都是可配置的。例如，您可以选择预测长度、模型类型（累加或累乘）以及是否考虑季节性。

该功能还具有极高的易用性，初学者只需使用默认设置单击几下即可创建预测，而高级用户可以在几乎所有层面对模型进行配置。在统计元素之外，Tableau 还通过显示置信区间，为初学者提供预测质量估计。预测功能还可以和 Tableau 的其他功能无缝配合，因此您可以像操作用户界面中的任何其他分析对象一样，对预测进行轻松分段和处理。

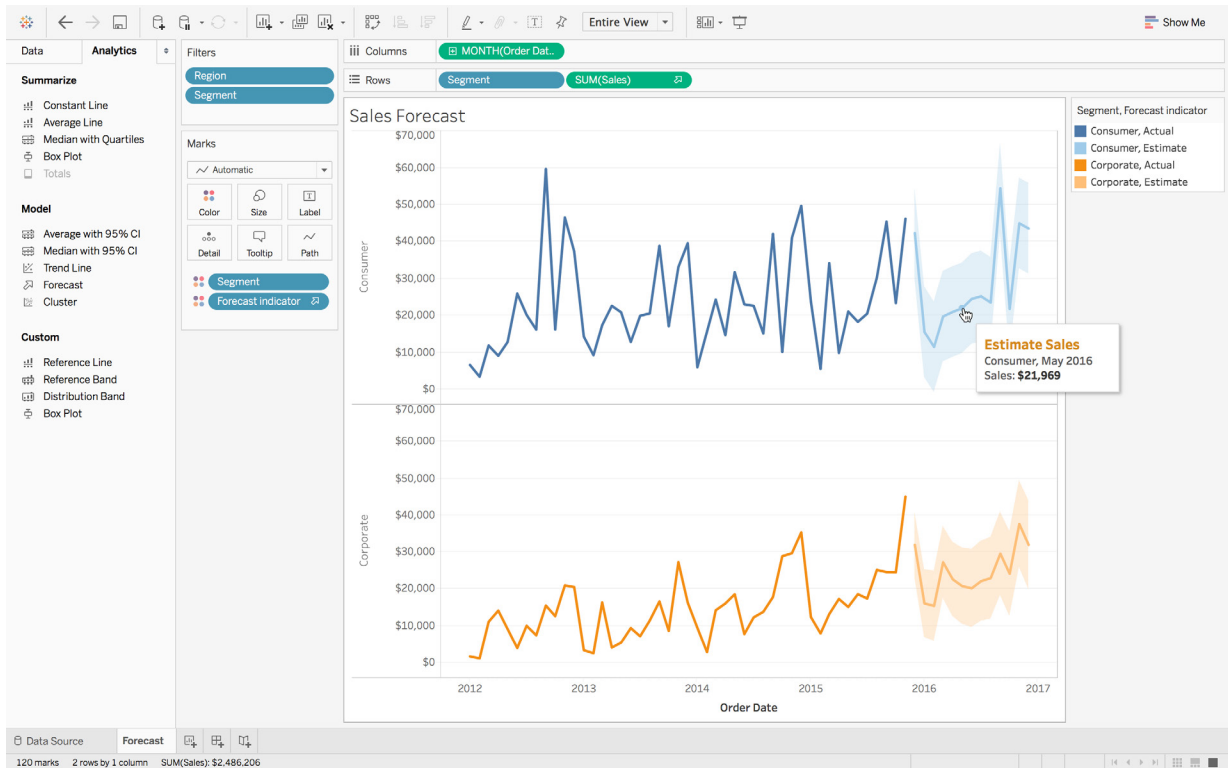


图 12：预测会按照分段自动预测销售额。

简单的预测分析可为任何数据项目带来巨大的价值。通过同时支持复杂的配置和简便的交互式建模，平台可同时为数据科学家和最终用户提供服务。

5.外部服务集成

功能：Python、R 和 MATLAB 集成

许多组织对分析平台和机构内知识进行投资。因此，您可能有非常具体的需求，以及很有价值的现有工作集合。

完善的分析平台必须支持与其他高级分析技术集成的功能，从而让您能够扩展可用的功能，充分利用您对其他解决方案的现有投资。借助 Tableau，您可以与 Python、R 和 MATLAB 等外部服务集成，由于 Tableau 易于使用，您可以让组织中的每个人通过简单拖放操作使用高级算法，从而普及数据科学。

Tableau 与 Python、R 和 MATLAB 直接集成，从而让用户可以使用现有模型，并从全球统计社区获得帮助。Tableau 将脚本和数据发送给 RServe 和 TabPy 之类的外部预测服务机构。随后，结果会返回到 Tableau，供 Tableau 可视化引擎使用。因此，Tableau 用户可以针对 Tableau 中的数据调用任何可用的 R 或 Python 函数，并对使用 Tableau 在这些环境中创建的模型实施操作。

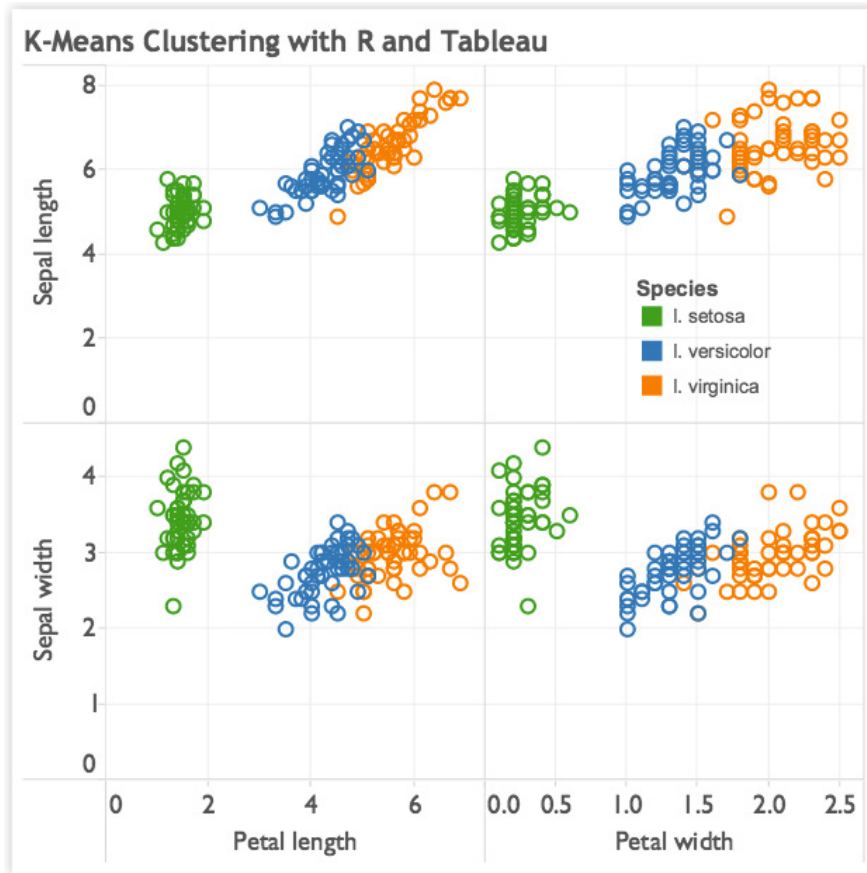


图 13：使用 R 的 K-均值聚类分析算法的示例，已在 Tableau 中可视化。
群集显示三种鸢尾花的变量（花萼和花瓣长度/宽度）差异。

建模深度可以远远超出基本统计。借助 R 和 Python 集成，您可以通过分析文本来提取情绪（图 14），也可以展示规范分析的结果，例如供应链或股票组合优化。此集成功能支持在 Tableau 中直接运行 R、MATLAB 或 Python 代码，以及在 Tableau 外部定义的模型和脚本。

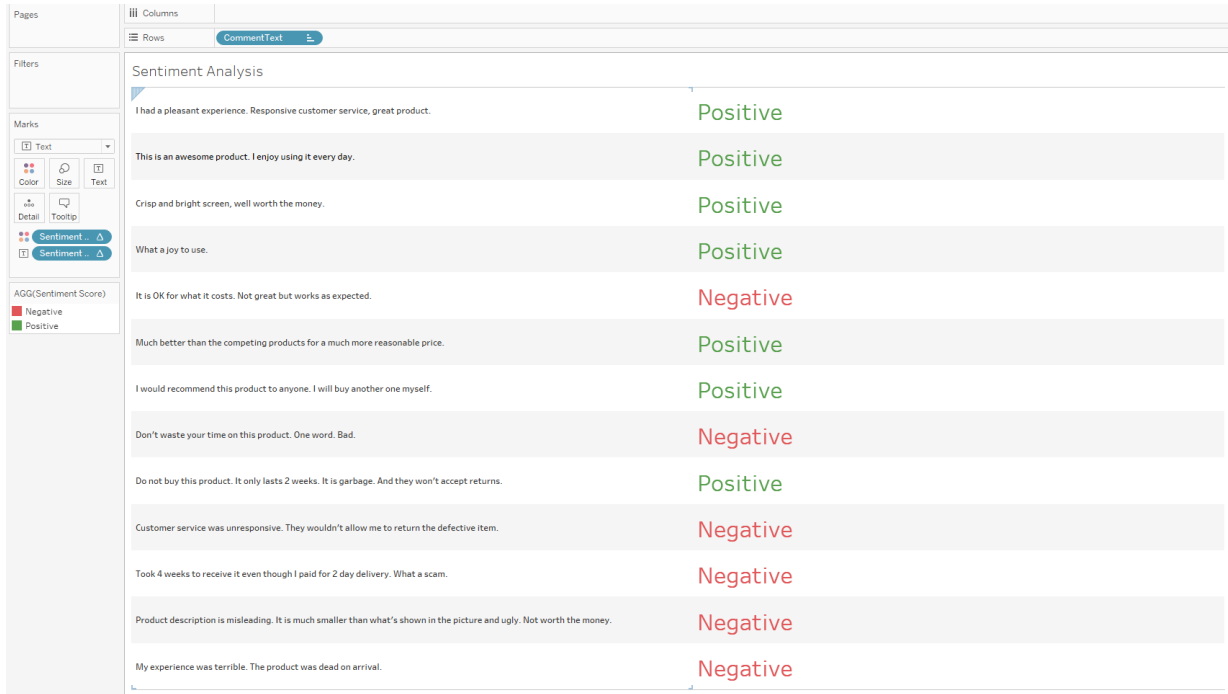


图 14：使用 Tableau R 集成构建情绪分析的示例。

图 15 中的仪表板使用名为“梯度提升”的集成式机器学习技术，根据历史癌症病例，针对乳腺肿瘤提供诊断建议。此模型经过训练，并由数据科学家在 Python 编程环境中对其准确性进行统计学评估，随后在 TabPy (Tableau Python Server) 上发布。此仪表板提供表格输入字段，使用户可以为不同的医学指标键入值，这些值被传送到托管的机器学习模型并实时转化为诊断建议。这样，众多分析师和仪表板可以扩充共享的机器学习资产，而数据科学家可以使用自己喜欢的工具，在同一个地点轻松管理这些资产。

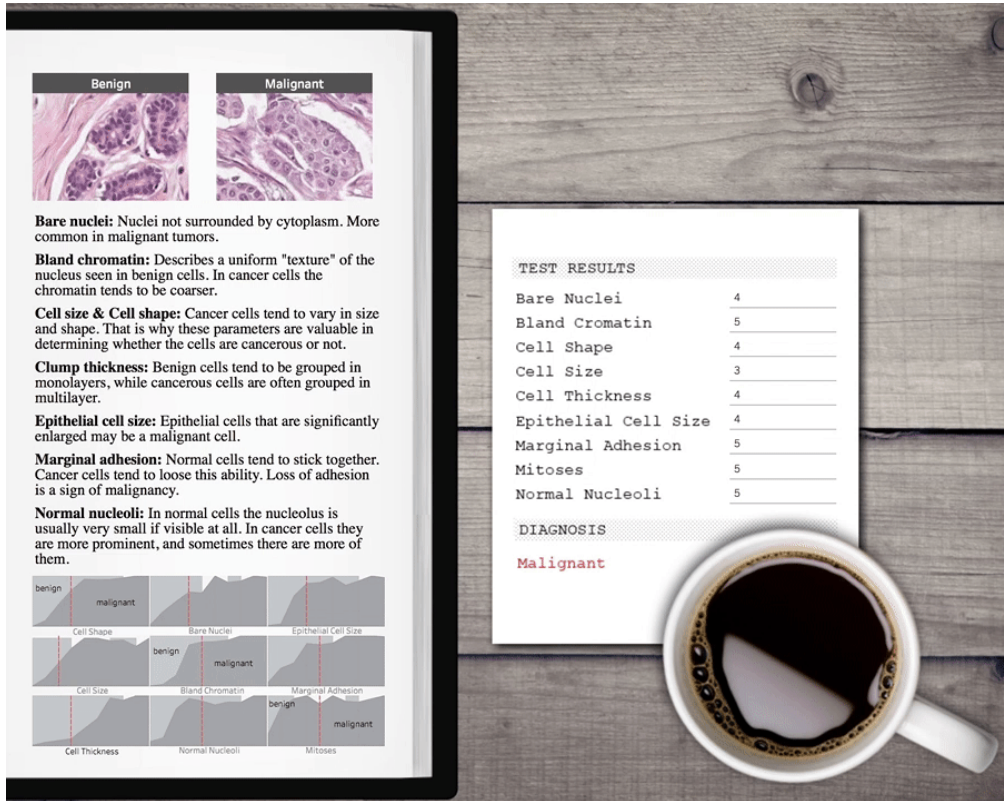


图 15：集成式机器学习模型及其输入参数。

Tableau 和 Python、R 或 MATLAB 的组合具有极其强大的功能。它可以将复杂的建模结果表达为美观而有趣的可视化，让各种技能水平的用户都能使用这些结果，并通过与仪表板中内嵌的控件进行交互，提出假设问题并评估假想方案。

结语

从很多方面来说，Tableau 都在分析平台中独树一帜。我们以扩充人类智能为使命，从业务用户和数据科学家的需求出发设计了 Tableau。我们始终专注于自己的使命 — 让用户能够以最快的速度针对自己的数据提出有趣的问题，构建出了能够为不同水平的用户提供有用功能的平台。

借助 Tableau 的灵活前端，业务用户可以在无需进行编码，也无需理解数据库的情况下提出问题。Tableau 还可以凭借其分析深度，成为数据科学家手中的一件强大武器。利用复杂计算、R 和 Python 集成、快速同期群分析以及预测功能，数据科学家可以在 Tableau 中完成复杂分析并轻松共享可视化结果。无论您使用 Tableau 的目的是进行数据探索和质量控制，还是进行模型设计和测试，该平台的灵活性都可以让您在整个项目周期节省大量时间。Tableau 可以让各级别的分析更加易于访问并缩短分析所需时间，从而在整个企业推动关键协作、提高决策质量。

关于 Tableau

Tableau 帮助您提取信息的意义。作为一个分析平台，它支持分析周期，提供可视化反馈，帮助您回答问题，即便这些问题的复杂程度与日俱增。如果您希望使用数据进行创新，您的应用程序必须能够鼓励您不断探索 — 提出新问题并改变视角。如果您已经准备好让自己的数据发挥重大作用，请立即下载 Tableau Desktop [免费试用版](#)。

相关白皮书

[使用 R 和 Tableau](#)

[了解详细级别表达式](#)

[定义分析](#)

[查看所有白皮书](#)

查看其他资源

[产品演示](#)

[培训与教程](#)

[社区与支持](#)

[客户故事](#)

[解决方案](#)

