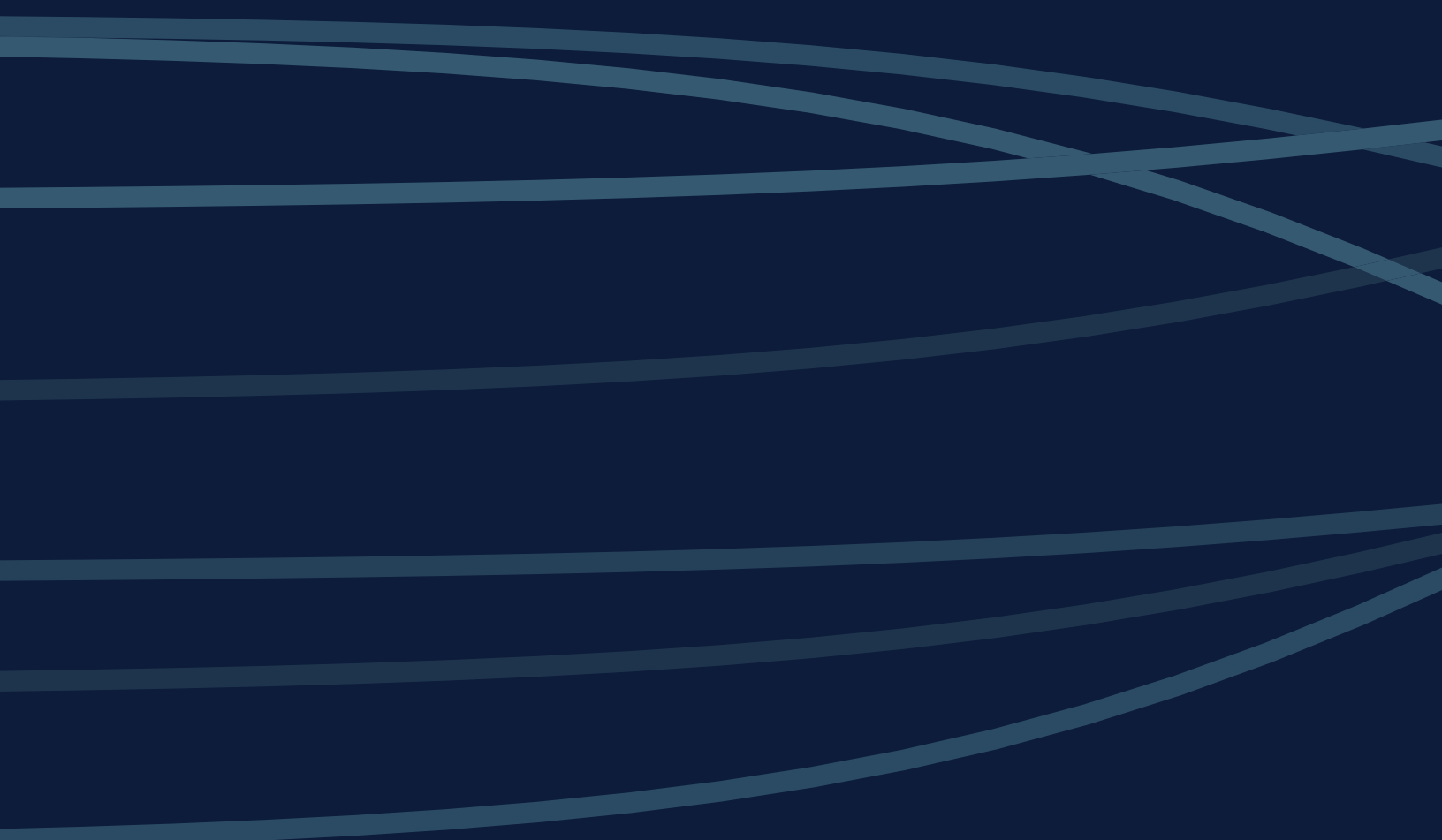# +++ +tableau®

Advanced Analytics
with Tableau

Tracy Rodgers, Product Marketing Manager

Introduction

We used to exist in a world of either-or. Either you knew how to program or advanced analytic techniques were out of your reach. Either you learned R, Python, and/or SAS, or you got someone else to do the heavy lifting. At Tableau, we believe that to truly augment human intelligence, we need to provide rich capabilities for users of all levels of technical ability. And that includes advanced analytics capabilities.

Tableau's analytical depth augments the workflows of data science groups at cutting-edge analytics companies like Netflix and Spotify. With a few clicks, you can create box plots, tree maps, predictive visuals, and cohort analyses. Connect to R, MATLAB, and Python, and use Tableau to extend your analysis, visualizing model results. Non-technical users can ask previously unapproachable questions, while data scientists can iterate and discover deeper insights faster, yielding more valuable findings.

In this paper we will explore how Tableau can help with all stages of an analytics project, but focusing specifically on a few advanced capabilities.

We will look at the following scenarios and the features that support them:

Segmentation and cohort analysis: Tableau promotes an intuitive investigative flow for rapid and flexible cohort analysis. Quickly move through different perspectives by slicing and dicing the data along as many dimensions as you want. If you want to augment your segmentation analysis with machine learning, automated clustering can help you discover patterns that would be difficult to spot in large data sets with high dimensionality.

Scenario and what-if analysis: By combining Tableau's flexible front-end with powerful input capabilities, you can quickly modify calculations and test different scenarios.

Sophisticated calculations and statistical functions: Tableau possesses a robust calculation language so you can augment your analysis with arbitrary calculations and perform complex data manipulations with concise expressions. Ranging from nesting basic aggregations to statistical calculations like standard deviation, mode, kurtosis, skewness, correlation, and covariance.

Time-series and predictive analysis: Since much of the world's data can be modeled by time series, Tableau natively supports rich time-series analysis, meaning you can explore seasonality and trends, sample your data, run predictive analysis like forecasting, and perform other common time-series operations within a robust UI.

External services integration: R, MATLAB, and Python integrations provide the power and ease of use of Tableau's front-end, while allowing experts to leverage prior work in other platforms and handle nuanced statistical and machine learning needs.

Contents

1. Segmentation and cohort analysis

Features: Clustering, Sets and Set Actions, Groups

To generate an initial hypothesis, business users and data experts often start the same way—by creating segments and/or conducting an informal cohort analysis.

Asking a series of basic questions about different segments helps analysts understand their data and validate their hypotheses like, "Do customers who pay with credit retain better than those who pay with check?" The ability to iterate rapidly can help drive model development and ensure projects stay on track.

Tableau possesses a rich set of capabilities to enable quick, iterative analysis, and comparison of segments. For example, with just a few calculated fields and some drag-and-drop operations, you can create a dashboard that breaks down patient volume by day and time, and further explore trends based on the department (Figure 1).
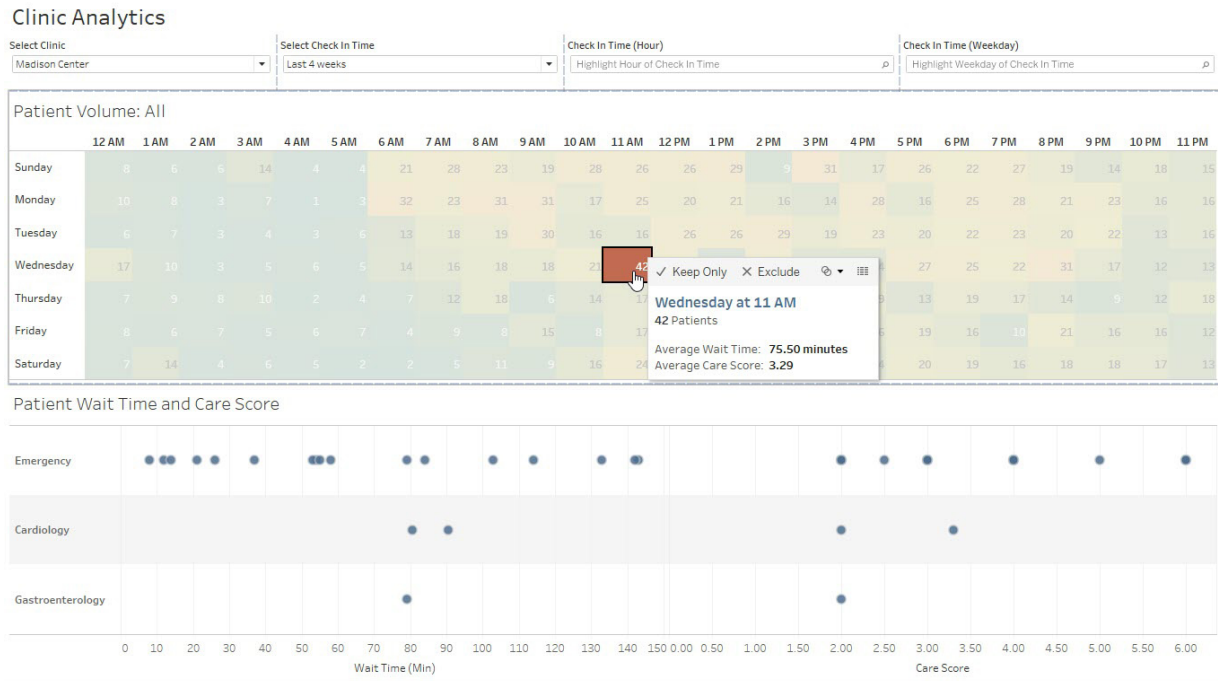

Figure 1: This interactive clinic dashboard shows the number of patients on a particular day and time, what they came in for, and how long they waited.

Clustering

Clustering in Tableau is another option that uses unsupervised machine learning for segmenting data, which is especially useful when a high number of variables need to be considered.
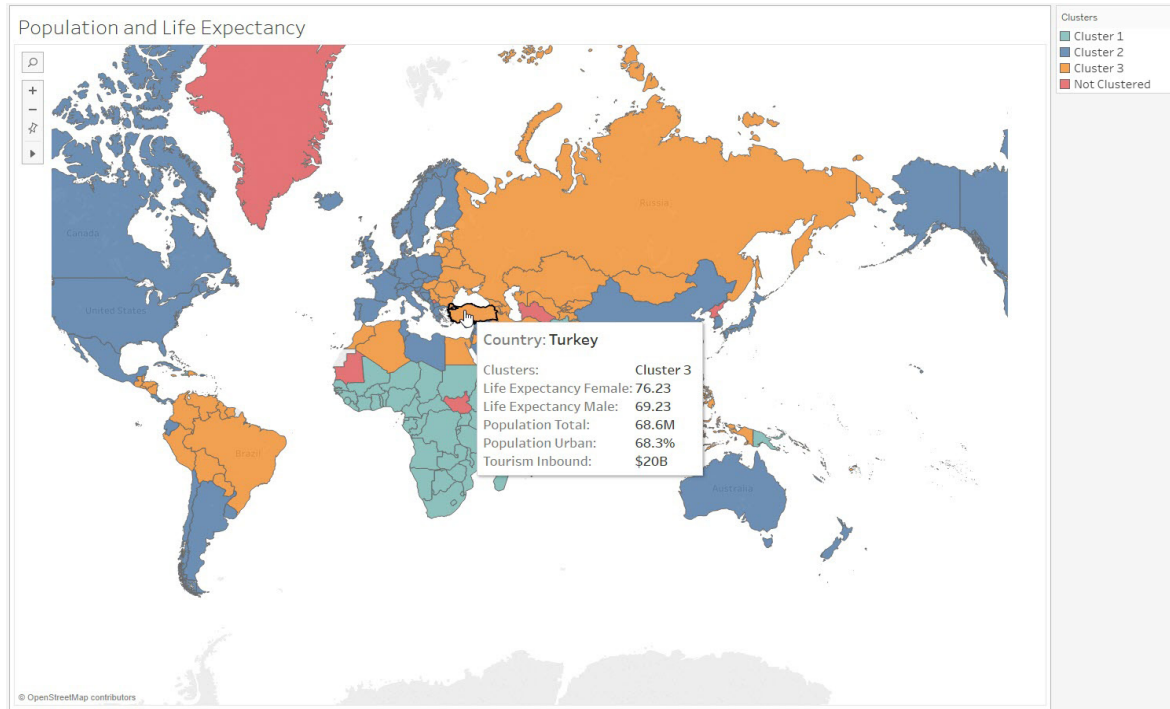

Population and Life Expectancy

Clusters
☐ Cluster 1
☐ Cluster 2
☐ Cluster 3
☐ Not Clustered

Country: **Turkey**

Clusters:	Cluster 3
Life Expectancy Female:	76.23
Life Expectancy Male:	69.23
Population Total:	68.6M
Population Urban:	68.3%
Tourism Inbound:	$20B

© OpenStreetMap contributors

Figure 2: Automatically segment your data using clustering.

In Figure 2, you can see results of clustering countries based on the United Nations' development indicators. The algorithm recommends a three-cluster solution that corresponds to underdeveloped, developing, and developed countries.

Tableau's flexible interface also makes it easy to test different theories and explore distributions across cohorts. The ability to iterate visually saves time tweaking scripts and rerunning simulations in other tools.
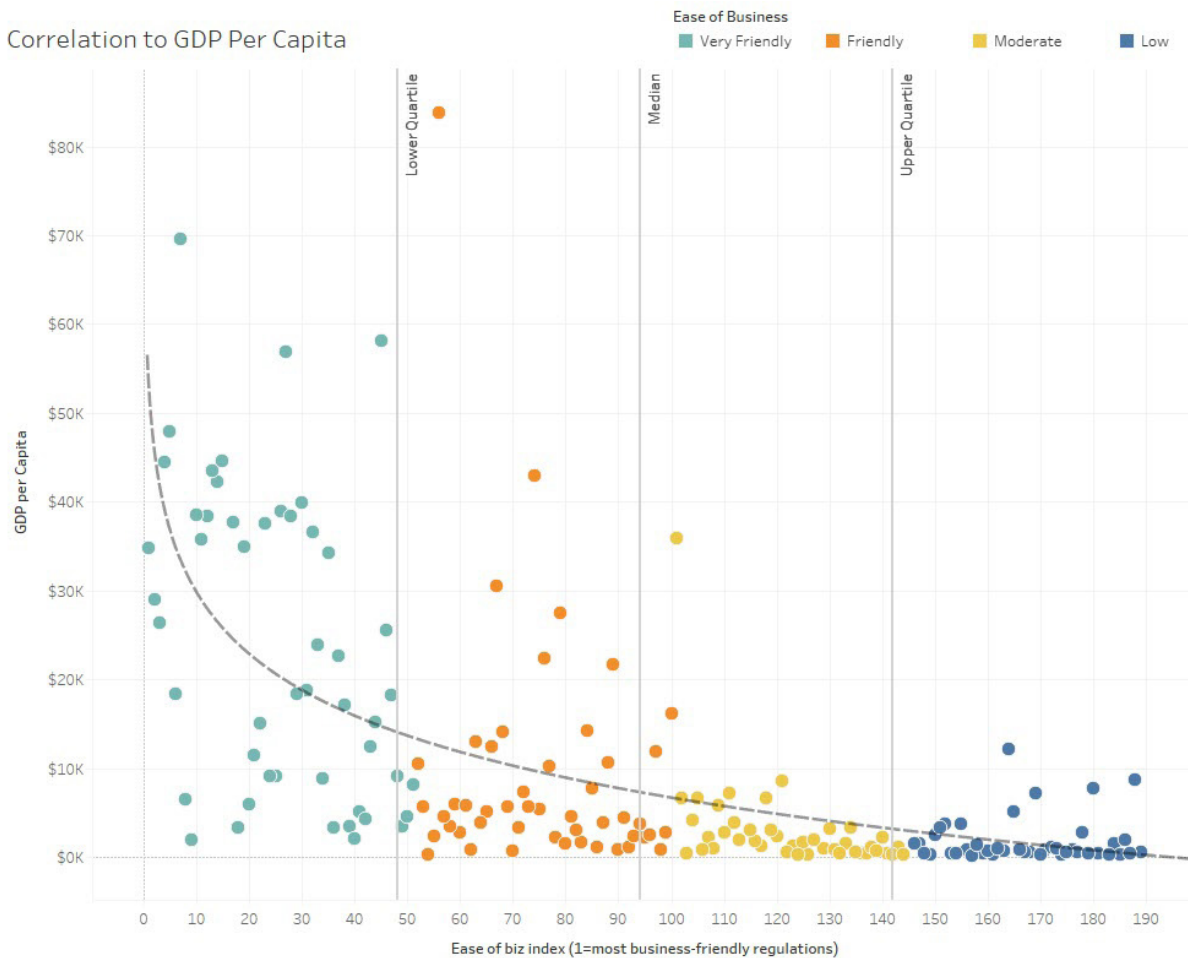
Figure 3: Segment and explore data in seconds. In this example, you can see countries segmented by "business-friendly regulations" and GDP per capita.

As seen in Figure 3, simply dragging the segmentation fields and measures used in clustering onto the canvas generates a new way to view the data, highlighting differences between the four clusters of countries.

Sets and Set Actions

Using Sets, you can define collections of data objects either by manual selection or by using programmatic logic. Sets can be useful in a number of scenarios including filtering, highlighting, cohort calculations, and outlier analysis. You can also combine multiple Sets in order to test different scenarios or create multiple cohorts for simulations—for example, combining different, independently generated customer groups for a retention analysis, or applying multiple successive criteria.

The Set Actions feature means that a selection of data points within a visualization can be stored within a set. A user selection in any of the sheets can update the set and modify all target sheets in a single, coordinated selection. This allows for a variety of use cases like proportional brushing (Figure 4).
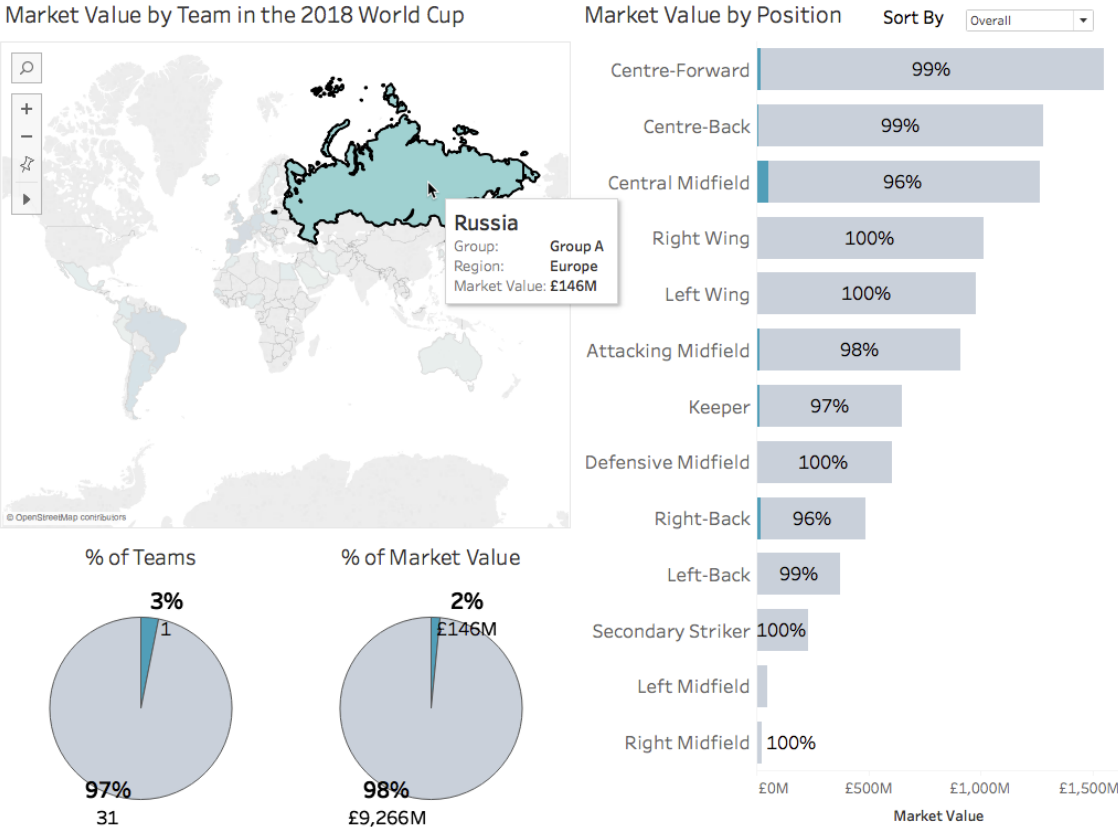

Figure 4: Set Actions allow for proportional brushing, an interactive analytics technique that displays the magnitude of a selection relative to the total magnitude. Selecting "Russia" shows the country's market value by the position of players in the World Cup.

Groups

To support the need for creating ad-hoc categories and establishing hierarchies, Tableau has a feature called Groups. Groups can also help with basic data cleaning needs.

Groups let users structure data in an intuitive way for the analysis task at hand, combining related members in a field. For example, if you are working with a view that shows average test scores by major, you might want to group certain majors together to create categories. English and History might be combined into a group called Liberal Arts Majors, while Biology and Physics might be grouped as Science Majors. This allows the analyst to customize the presentation and control the aggregation of data throughout the analysis.

Groups can also help when data has consistency and quality issues. For example, California may be called by its full name, but may also be referred to as CA or Calif. Analysts and business users often do not have permissions to change source systems directly to clean up issues, meaning small data errors can greatly encumber exploratory analysis. Having to stop asking questions in order to request data changes delays projects and disrupts the rapid development of ideas. With Groups, you can quickly define a new segment that includes all of the alternate names for the purposes of your analysis and continue to ask questions without disrupting your flow.
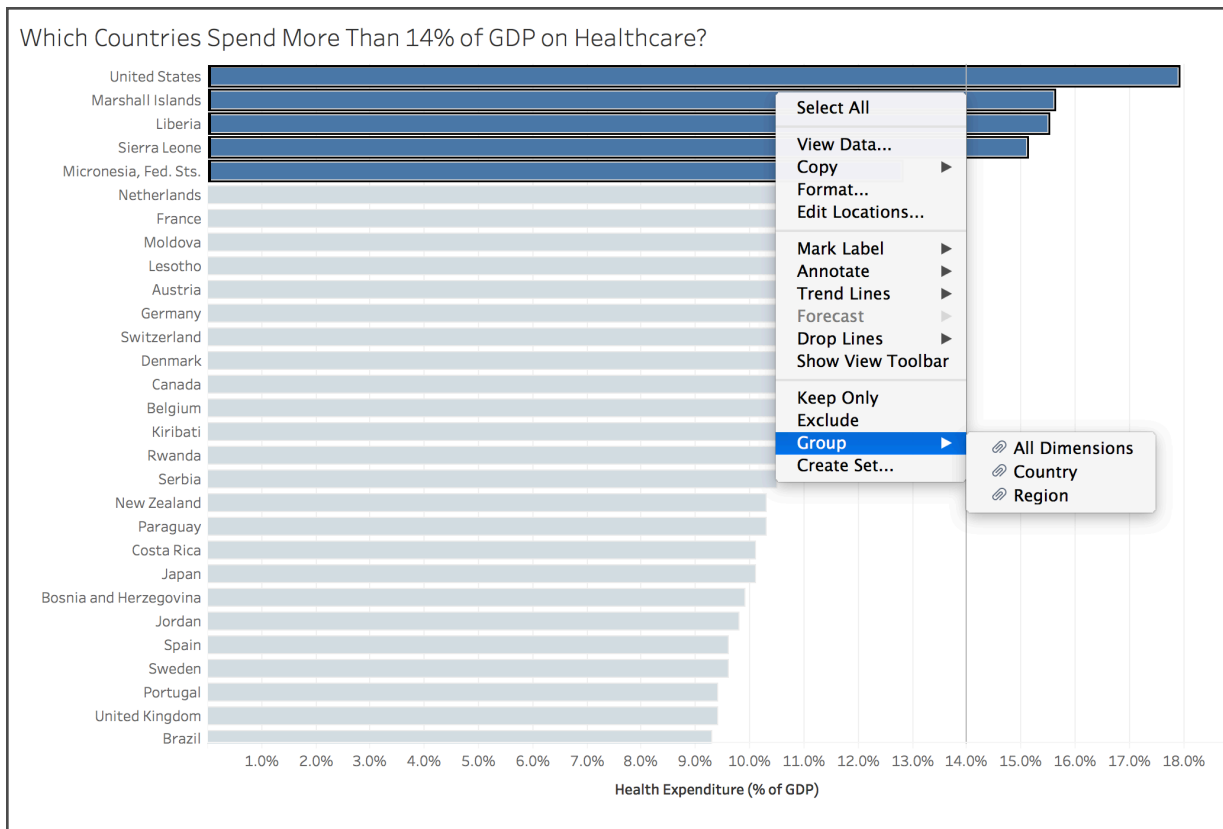

Figure 5: Create a group to combine related members in a field. In this example, you may want to group the top five countries that spend the largest percentage of their GDP on healthcare.

Inherent to all of these capabilities are simple updates. In Tableau, if you choose a live connection and update your data, your analysis and all the underlying components like Sets and Groups will update as well. This means that cohort membership updates automatically without manually rerunning reports or dependent scripts. Simple updates help ease the reporting burden and are yet another way to test scenarios. They make it possible to swap out the underlying data in order to probe the sensitivity to initial conditions without any need to update the analysis stack.

By letting users quickly segment and categorize their data, Tableau enables business users and analysts alike to perform powerful cohort analysis with relative ease. These capabilities also help data scientists investigate initial hypotheses and test scenarios.

2. What-if and scenario analysis

Features: Parameters, Story Points

Sometimes users want to explore how changing a particular value or set of values affects the output of their analysis. This could be used to test different theories, highlight important scenarios for colleagues, or investigate new business possibilities. With Tableau, you can experiment with the inputs of your analysis. And simple controls allow you to easily flag and share scenarios while keeping data fresh.

Parameters

When performing a what-if analysis, you may want to change the base value of a calculation, redefine a quota, or set initial conditions. Parameters in Tableau make this an easy task. By defining a parameter, you provide a way to change the input values into your model or dashboard. Parameters can drive calculations, alter filter thresholds, and even select what data goes into the dashboard. Non-technical users can leverage parameters to experiment with different inputs and explore possible outputs from complex models.

In addition to helping you test hypotheses, Tableau's Parameter feature lets you showcase results from a what-if analysis in an interactive report. In Figure 6, parameters drive a what-if analysis around sales commissions. The sales manager can experiment with commission rates, base salaries, and quotas, all while getting real-time feedback on how these changes impact key metrics.
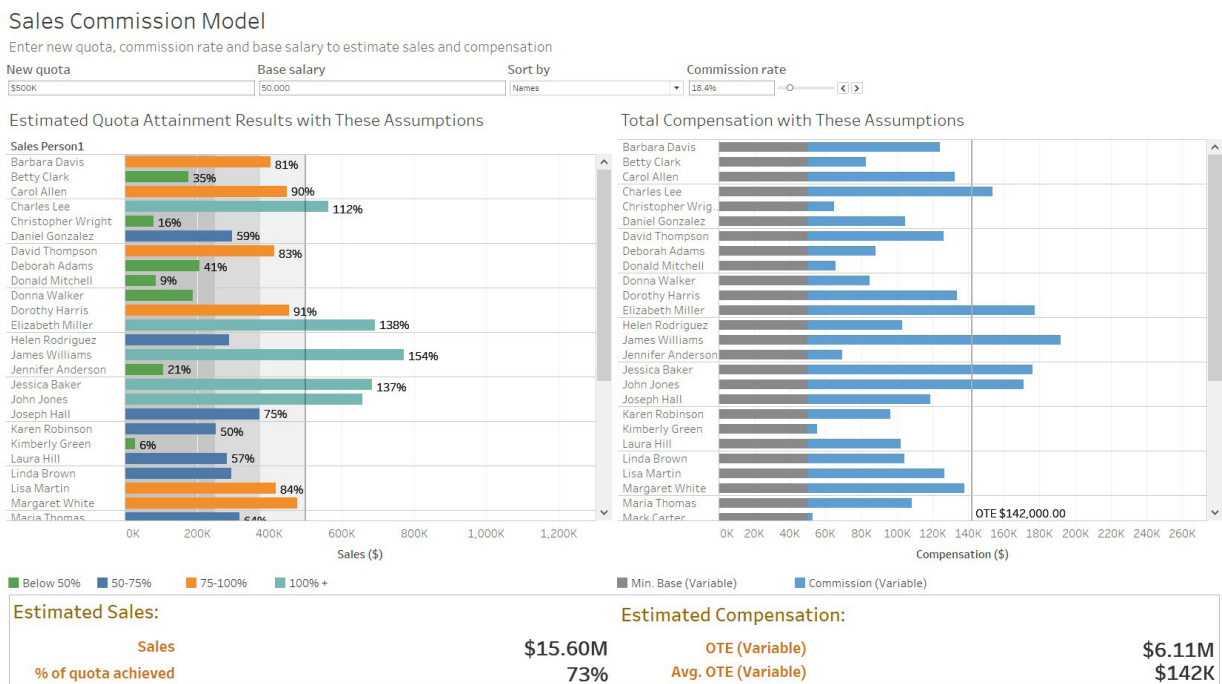

Figure 6: With this parameter-driven sales report, the interactor can explore the effect of quotas, commissions, and salaries within the organization.

Story Points

When combined with Story Points (Tableau's way of building a narrative with data), Parameters allow you to take snapshots of interesting results and continue exploring. Story Points allow you to construct a presentation that continues to update with data changes and viz modifications. However, Story Points are smart enough to retain Parameter values, so you can flag scenarios and have the confidence that you can return to them without interrupting your analytical flow. You can also compare the results from several different sets of inputs without worrying about stale screenshots or rerunning simulations.

With Sets, Groups, drag-and-drop segmentation, and Parameters, Tableau makes it possible to move from theories and questions to a professional-looking dashboard that allows even non-experts to ask questions and test their own scenarios. Streamlining what-if analysis empowers data professionals to focus on the more complex aspects of the analysis and deliver greater insights, while intuitive visuals allow end users to engage with the data. This increased engagement helps drive change and empower better decision-making throughout an organization.

3. Sophisticated calculations

Features: Calculated fields, Level of Detail Expressions, Table Calculations

Typically, source data does not contain all the fields necessary for a comprehensive analysis. Analysts need a simple, yet powerful language to transform data and define intricate logic.

Although Tableau is easy to use, it also has powerful capabilities that support complex logic. Calculated fields in Tableau allow you to create new data from data that already exists in your data source. With calculated fields, you can easily perform arithmetic operations, express conditional logic, or perform specialized operations on specific data types. Two types of calculated fields that enable advanced analysis are Level of Detail (LOD) Expressions and Table Calculations.

Level of Detail Expressions

LOD Expressions are an extension of Tableau's calculation language and they are a powerful way to answer questions involving multiple levels of granularity in a single visualization. With LOD Expressions, many previously impossible or challenging scenarios can now be handled with a very simple, concise expression.

LOD Expressions greatly simplify cohort analysis (as described in a previous section) and multi-pass aggregations. Figure 7 shows the running sum of customer acquisitions by regional market for cohorts of customers bucketed into "new" and "existing" by their first order date. In the next section on time-series analysis, we'll look at some of the other aspects of the calculation language that make this analysis possible. The chart reveals that customer acquisition in North America declined from January on through the rest of the year. LOD Expressions turn segmentation that would otherwise require complex group-by statements in SQL into simple, intuitive expressions that you can edit and adapt in Tableau's front-end.
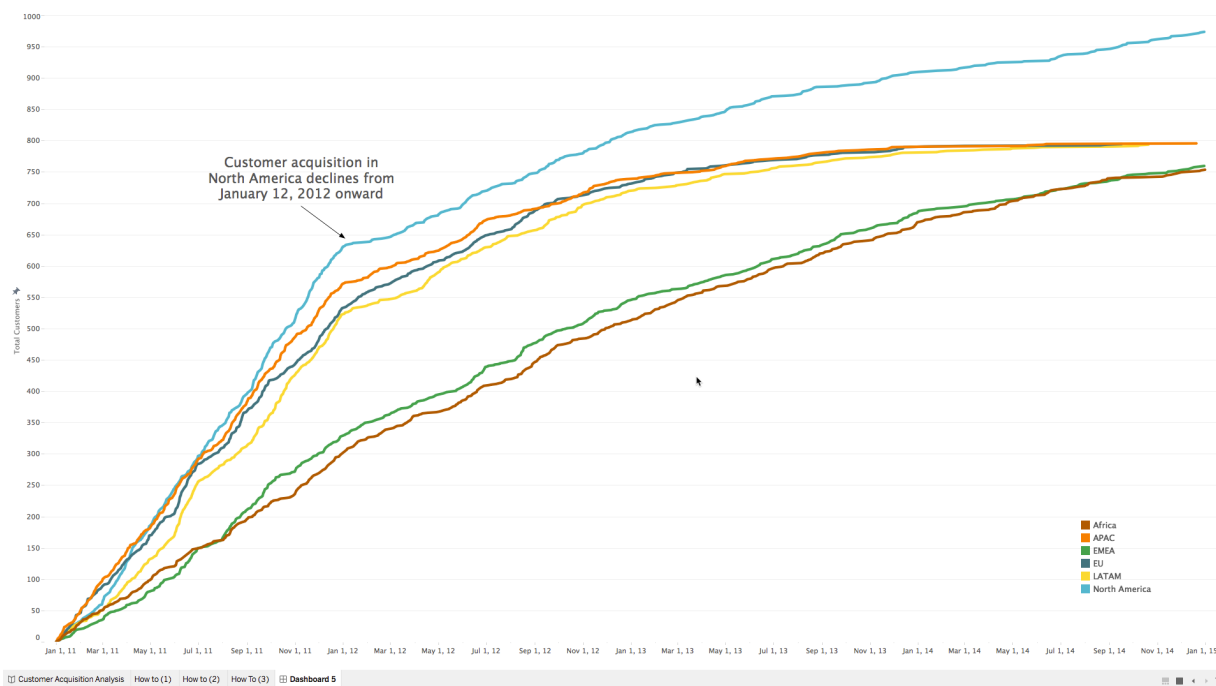
Figure 7: This visualization uses an LOD Expression to see daily trends of total customer acquisition by market. An LOD Expression ensures that we don't miscount repeat customers as new customers.

Table Calculations

Tableau Table Calculations enable computations that are relative in nature. More specifically, Table Calculations are computations that are applied to all values in a table and are often dependent on the table structure itself. This type of calculation includes many time-series operations such as lags or running sums, but also computations like ranking and weighted averages.

In Tableau, there are two ways to work with Table Calculations. The first is a collection of commonly-used Table Calculations called Quick Table Calculations. These let you define a Table Calculation with one click and are a great place to start. You can also create your own Table Calculations using the Table Calculation Functions. These functions give workbook authors the power to precisely manipulate their result sets. Also, since all Table Calculations are expressible in the same calculation language, you can use one of the Quick Table Calculations as a starting point and edit it manually if you need additional complexity.

With Table Calculations, challenging database work—like manipulating aggregated data, and creating data structure-dependent aggregations—requires just a few clicks or a simple expression. This both empowers non-technical users and saves experts countless hours and laborious SQL code.
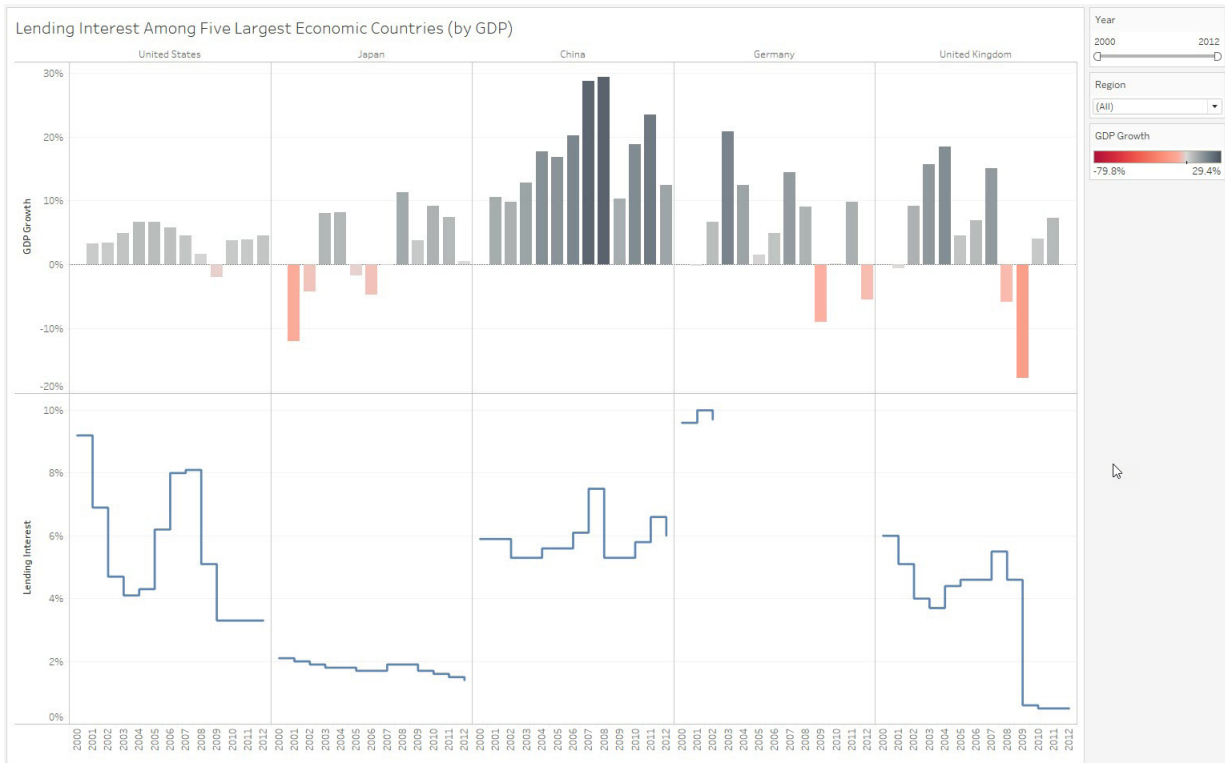
Figure 8: Using one of Tableau's built-in Quick Table Calculations (percent difference), you can quickly see how a rise in interest rates affects GDP in some countries.

4. Time-series and predictive analysis

Features: Time-series analysis, Forecasting

From sensor readings to stock market prices to graduation rates, much of the world's data can be effectively modeled as time series. As such, time is one of the most common independent variables used in analytics projects. And to find even more insights, you may want to leverage predictive capabilities. With Tableau, you can add predictive analytics without a large effort so you can explore multiple scenarios quickly.

Time-series analysis

In Tableau, a flexible front end and powerful back end makes time-series analysis a simple matter of asking the right questions. Analysis starts by just dragging the fields of interest into the view and beginning the interrogation process.

Leveraging the dual axis feature and discretized aggregation, you can start looking at multiple time series. Without the ability to quickly inspect time series at different levels of granularity and aggregation, you might not be able to generate the question.
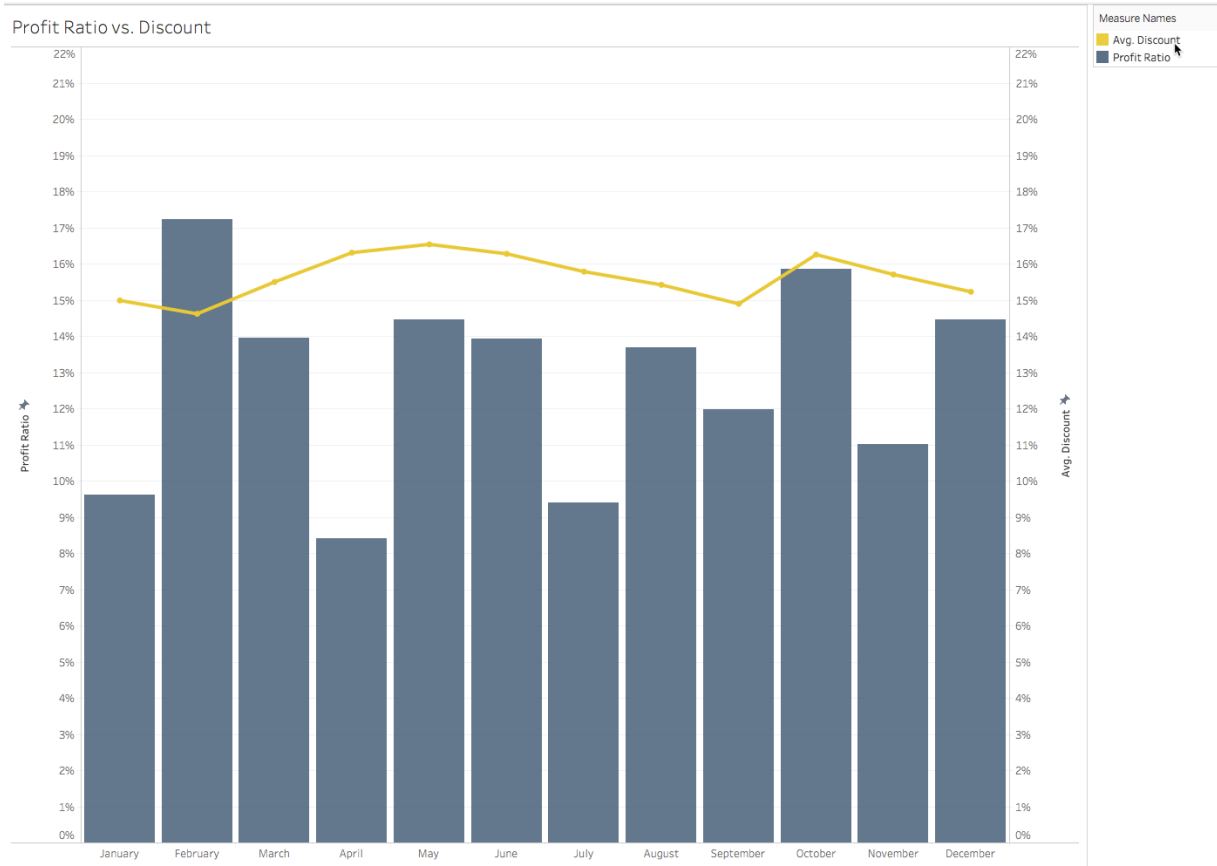
Figure 9: The dual axis plot shows the relationship between profit ratio and average discount.

To look at a specific time period, you can filter your data to a set of exact dates or take advantage of Tableau's relative date filters. With relative date filters, you can look at relative periods, like "last week" or "last month." These periods update each time you open the view to incorporate new data, making them a powerful tool for reporting.

When working with time series, it's often necessary to smooth or perform other temporal calculations. Tableau possesses a rich feature set designed to simplify common time-series operations such as moving averages, year-over-year calculations, and running totals.

As previously discussed, Tableau's Table Calculations feature lets you choose from a common set of time-series manipulations (Quick Table Calculations) or write custom computations.
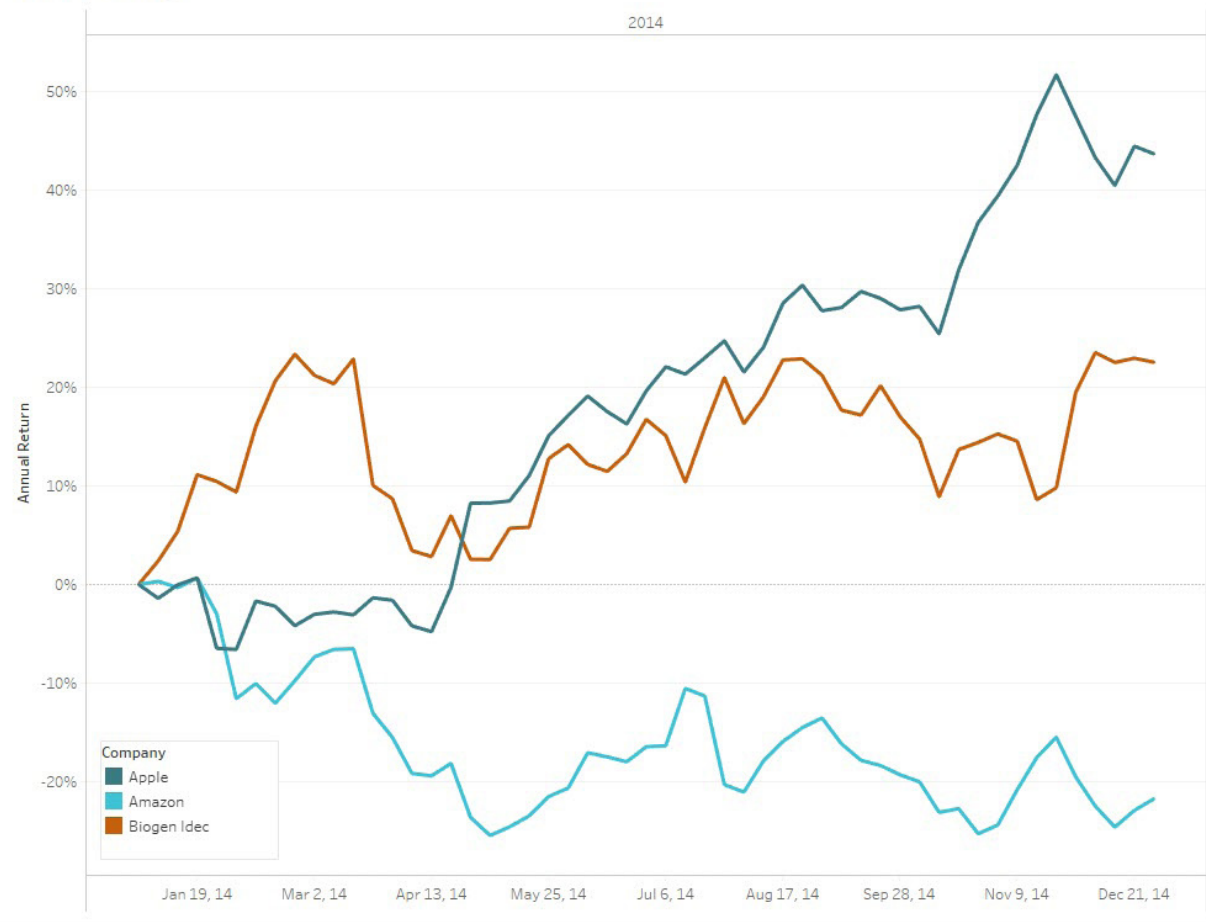
Figure 10: This time-series analysis shows monthly stock prices of three major companies.

Since time-series analysis is so common, Tableau's functionality helps finish projects faster and deliver more value to the organization. The intuitive functionality helps both data experts and business analysts to ask more and better questions of their data.

Forecasting

If you need to look toward the future, Tableau possesses several native modeling capabilities, including trending and forecasting.

You can quickly add a trendline to any chart and view details describing the fit (e.g. p-values and R-squared) simply by right-clicking on the line. Using Tableau's drag-and-drop functionality, you can model different groups with a single click as trend lines are fully integrated into the front end and can be easily segmented. As seen in Figure 11, Tableau automatically creates three trend lines for the different segments without any code. Tableau also supports several other types of fits, including logarithmic, polynomial, and exponential.
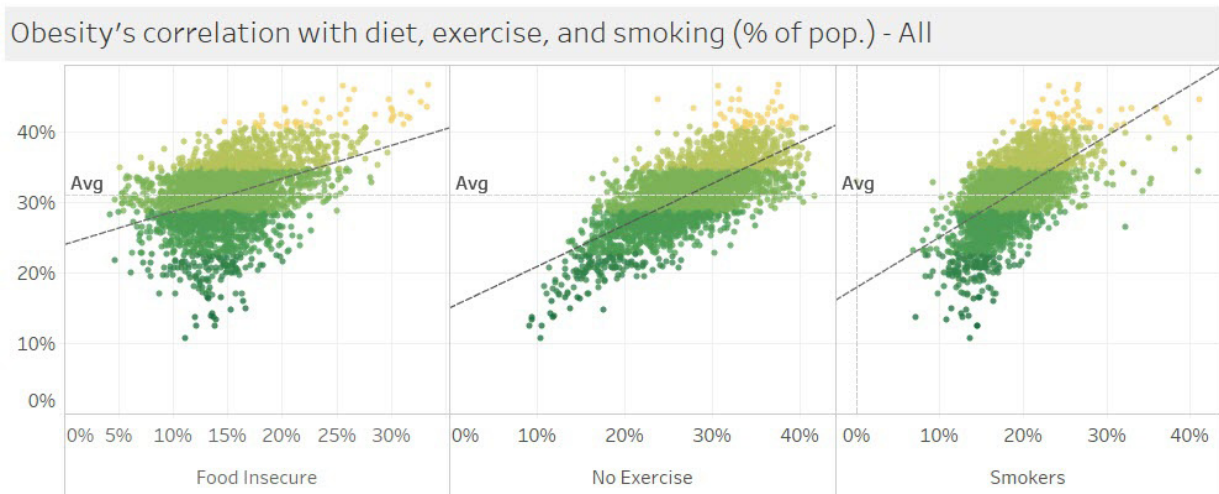
Obesity's correlation with diet, exercise, and smoking (% of pop.) - All

Figure 11: Trend lines highlight the relationship between obesity and food insecurity, exercise, and smokers.

By default, Tableau's forecasting functionality will run several different models in the background and select the best one, automatically accounting for data issues such as seasonality. Forecasting in Tableau uses a technique known as exponential smoothing. Exponential smoothing iteratively forecasts future values of a time series from weighted averages of past values. As mentioned previously, almost everything about the forecast is configurable. For example, you can select the length of the forecast, the type of model (additive or multiplicative), and whether or not to account for seasonality.

The feature is also very easy to use, so a novice user can create a forecast with just a few clicks using the default settings, while an advanced user can configure almost all aspects of the model. In addition to the statistical elements, Tableau provides novice users an estimate of the forecast quality by displaying confidence intervals. Forecasting also fits in seamlessly with the rest of Tableau, so you can easily segment and manipulate the forecast as you would any other analytic object in the user interface.
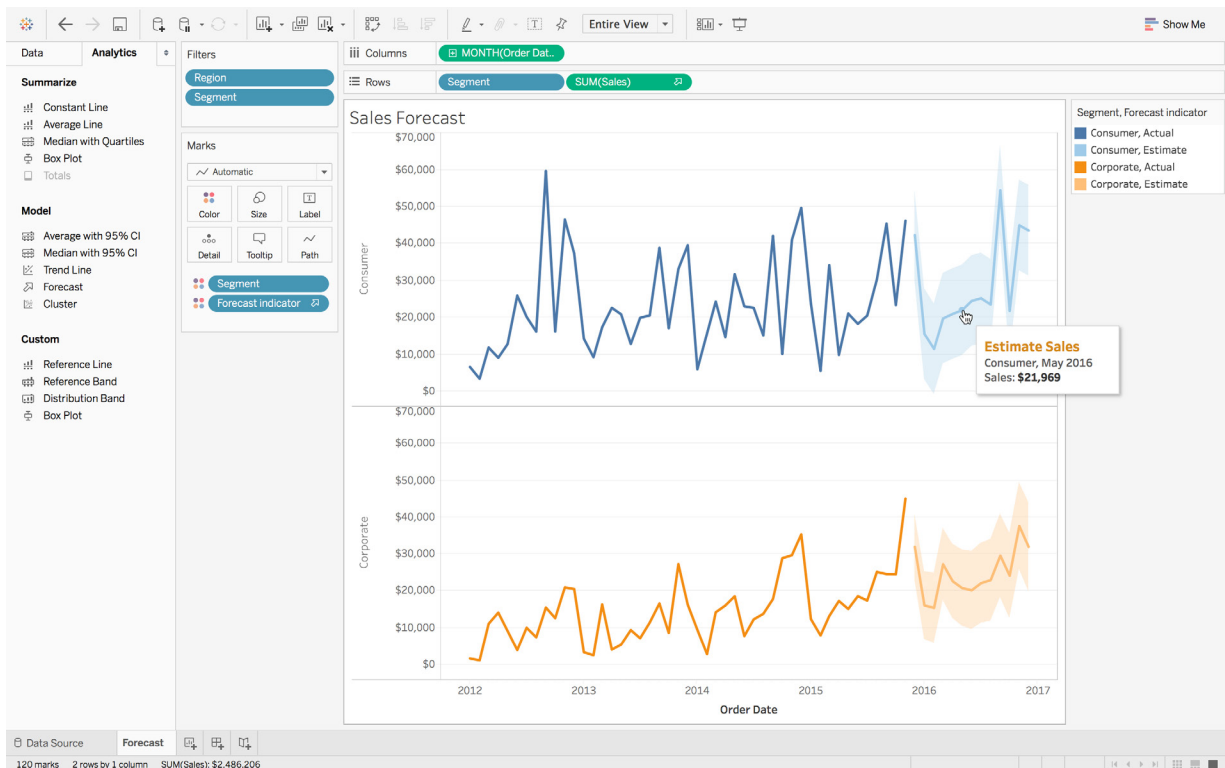
Figure 12: Forecasting automatically predicts sales by segment.

Easy predictive analytics adds tremendous value to any data project. By supporting both complex configuration and simple interactive modeling, a platform can serve both the data scientist and the end user.

5. External services integration

Features: Python, R, and MATLAB integrations

Many organizations make investments in analytic platforms and institutional knowledge; therefore, you may have very specific needs and a valuable corpus of existing work.

A comprehensive analytics platform must support the ability to integrate with other advanced analytics technologies, allowing you to expand the possible functionality and leverage existing investments in other solutions. With Tableau, you can integrate with external services like Python, R, and MATLAB—and because of Tableau's ease of use, you can democratize data science by making advanced algorithms accessible to everyone in your organization with a simple drag and drop.

Tableau integrates directly with Python, R, and MATLAB to support users with existing models and the leverage the worldwide statistics community. Tableau sends scripts and data to external predictive services such as RServe and TabPy. The results are then returned to Tableau for use by the Tableau visualization engine. This allows a Tableau user to call any function available in R or Python on data in Tableau and to manipulate models created in these environments using Tableau.
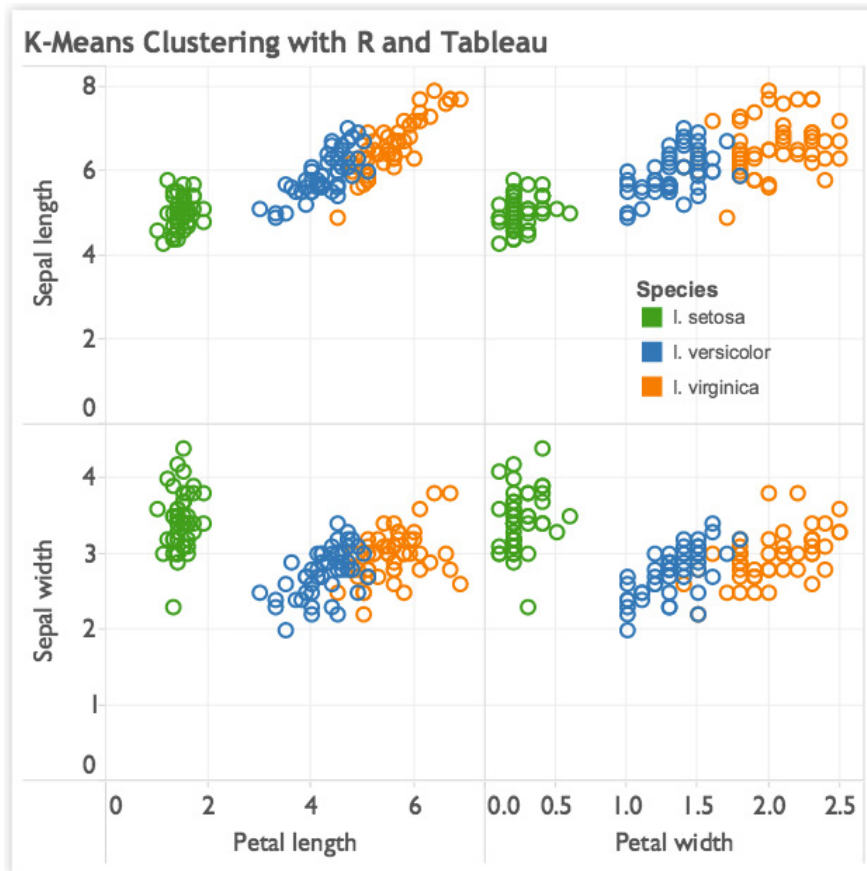

Figure 13: Example of k-means clustering algorithm with R, visualized in Tableau.
Clusters show differences in variables (sepal and petal length/width) for three types of Iris flowers.

The modeling can go much deeper than basic statistics. With R and Python integrations, you can analyze text to extract sentiment (Figure 14) or display results of prescriptive analytics such as supply chain or stock portfolio optimization. The integration supports running R, MATLAB, or Python code directly inside Tableau, as well as models and scripts defined outside Tableau.
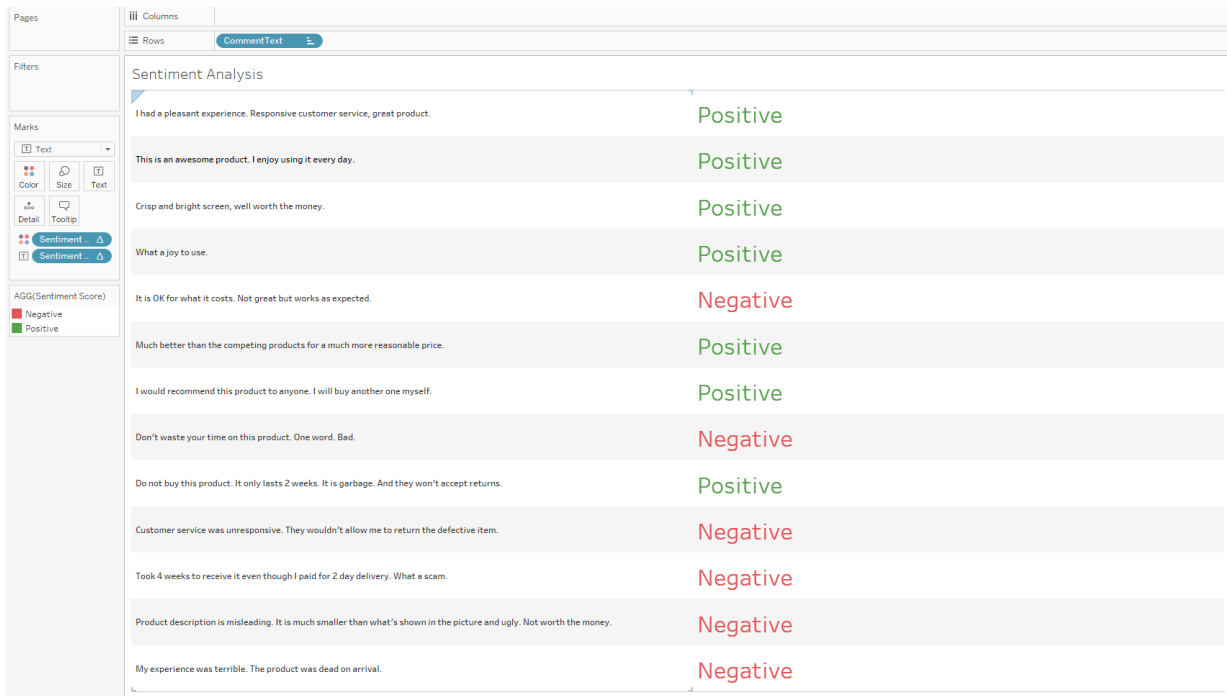

Figure 14: Example of a sentiment analysis built with the Tableau R integration.

Figure 15 shows a dashboard that uses gradient boosting, an ensemble machine learning technique to provide diagnosis suggestions for breast tumors based on historical cancer cases. This model is trained and statistically evaluated for accuracy by a data scientist in a Python programming environment and then published to TabPy (Tableau Python Server). The dashboard provides form entry fields to allow its users to type in values for different medical indicators, which are then passed to this hosted machine learning model and turned into diagnosis suggestions in real-time. This allows multiple analysts and dashboards to build on shared machine learning assets that data scientists can easily manage in a single location using their tool of choice.
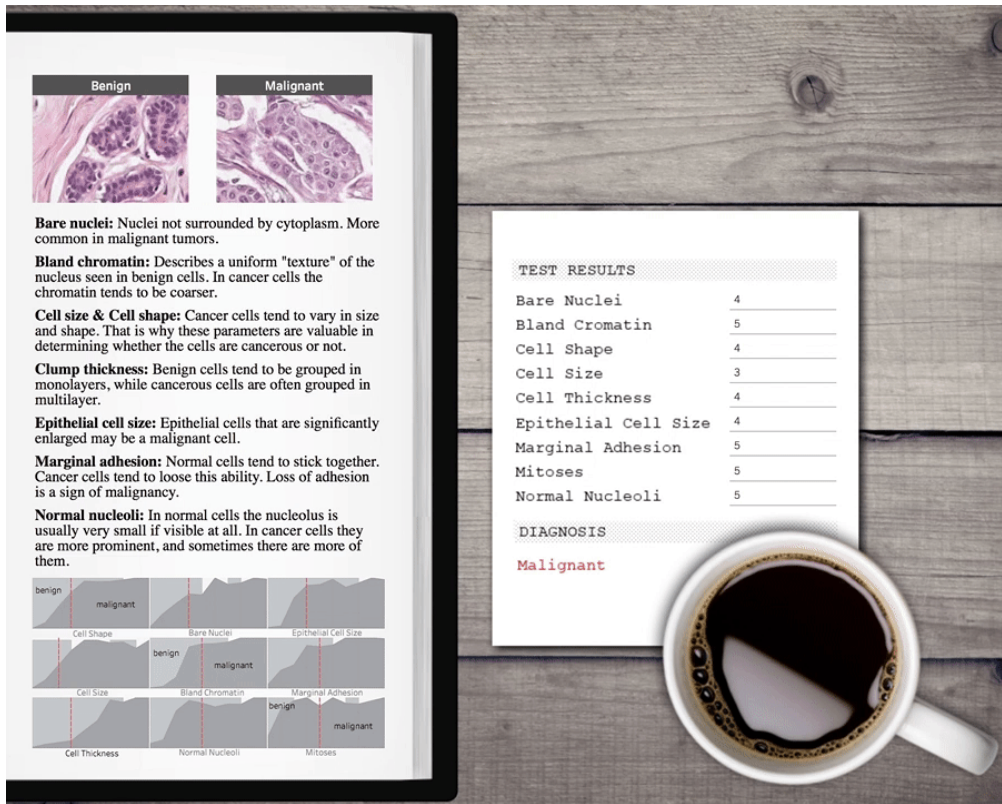
Figure 15: Ensemble machine learning model with input parameters.

The combination of Tableau with Python, R, or MATLAB is extremely powerful. It allows complex modeling results to be expressed as beautiful and engaging visuals for consumption by users of any skill level and lets them ask what-if questions and evaluate hypothetical scenarios by interacting with controls embedded into the dashboards.

Conclusion

In many ways, Tableau stands alone among analytics platforms. Because of our mission to augment human intelligence, we designed Tableau with both the business user and data scientist in mind. By staying focused on our mission to empower users to ask interesting questions of their data as quickly as possible, we built a platform that has valuable functionality for users of all levels.

Tableau's flexible front end allows business users to ask questions without needing to code or understand databases. Tableau also has the necessary analytical depth to be a powerful weapon in a data scientist's arsenal. By leveraging sophisticated calculations, R and Python integrations, rapid cohort analysis, and predictive capabilities, data scientists can complete complex analyses in Tableau and easily share the visual results. Whether you use Tableau for data exploration and quality control, or model design and testing, the interactive nature of the platform saves countless hours across the lifetime of a project. By making analysis more accessible and faster to complete at all levels, Tableau drives critical collaboration and better decision-making throughout the enterprise.

About Tableau

Tableau helps extract meaning from information. It's an analytics platform that supports the cycle of analytics, offers visual feedback, and helps you answer questions, regardless of their evolving complexity. If you want to innovate with data, you want an application that encourag- es you to keep exploring—to ask new questions and change your perspective. If you're ready to make your data make an impact, download a free trial of Tableau Desktop today.

Related whitepapers

Using R and Tableau

Understanding Level of Detail Expressions

Define Analytics

See all whitepapers

Explore other resources

Product Demo

Training & Tutorials

Community & Support

Customer stories

Solutions