

---

# Advanced Analytics with Tableau

Ian Coe, Product Manager

Bora Beran, Staff Product Manager

Piper Stull-Lane, Product Marketing Manager

We used to exist in a world of either-or. Either you knew how to program or advanced analytic techniques were out 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. We believe in enabling people of all skill levels to get insights and validation from their data.

Tableau is unique among analytics platforms in that it serves both business users and data scientists. Its intuitive interface and **visual query language** empowers non-programmers to conduct deep analysis without writing code. And its analytical depth augments the workflows of data science groups at cutting-edge analytics companies like Facebook and Amazon.

With a few clicks, you can create box plots, tree maps, predictive visuals, and cohort analyses. Connect to R 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 better, more valuable findings.

In this paper we will explore how Tableau can help with all stages of an analytics project, but focus specifically on a few advanced capabilities. Broadly, 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. Move through different perspectives quickly by slicing and dicing the data along as many dimensions as you want. If you would like 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.

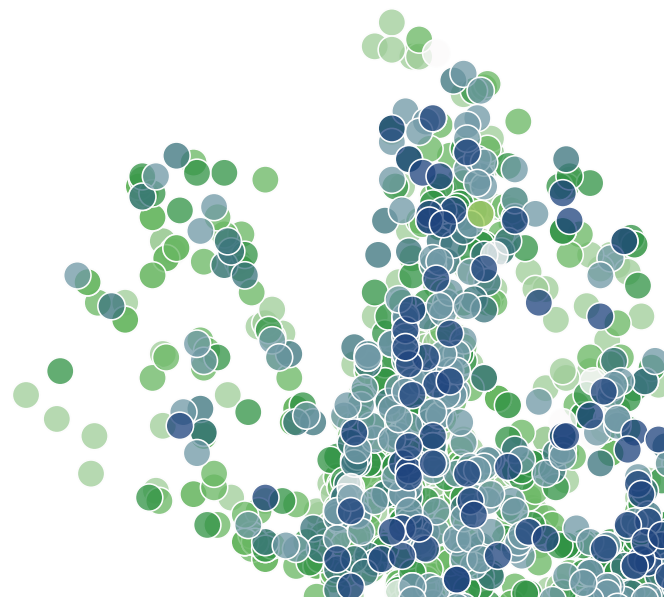
**Sophisticated calculations and statistical functions:** Tableau possesses a robust calculation language, which makes it easy to augment your analysis with arbitrary calculations and perform complex data manipulations with concise expressions. Ranging from nesting basic aggregations to statistical calculations such as 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, trends, sample your data, run predictive analysis such as forecasting, and perform other common time-series operations within a robust UI.

**External services integration:** R and Python plugins 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.

Jump to...

SCENARIO	FEATURES
1. <a href="#">Segmentation and Cohort Analysis</a>	Calculated Fields, Clustering, Sets, Groups
2. <a href="#">Scenario and What-If Analysis</a>	Parameters, Stories
3. <a href="#">Sophisticated Calculations</a>	Smart Authoring Language, LOD Expressions, Quick Table Calculations, Statistical Functions
4. <a href="#">Time-Series and Predictive Analysis</a>	Trend Analysis, Forecasting
5. <a href="#">External Services Integration</a>	R and Python Integration



# 1. Segmentation and Cohort Analysis

*Features: Calculated Fields, Clustering, Sets, Groups*

The ability to iterate rapidly can help drive model development and ensure projects stay on track.

---

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 (e.g. “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. The ideal platform for this phase should support the following:

---

Rapid ideation:

Provide an intuitive investigation canvas and near-instant feedback to questions asked as part of the analytical flow.

Simple set operations:

Create and combine cohorts using standard set operations or a simple UI.

Data issue handling:

Correct data errors and adjust cohorts without needing permissions to modify the underlying data source.

Seamless updates when data changes:

Propagate data updates through the analysis without running manual update scripts or refreshing caches.

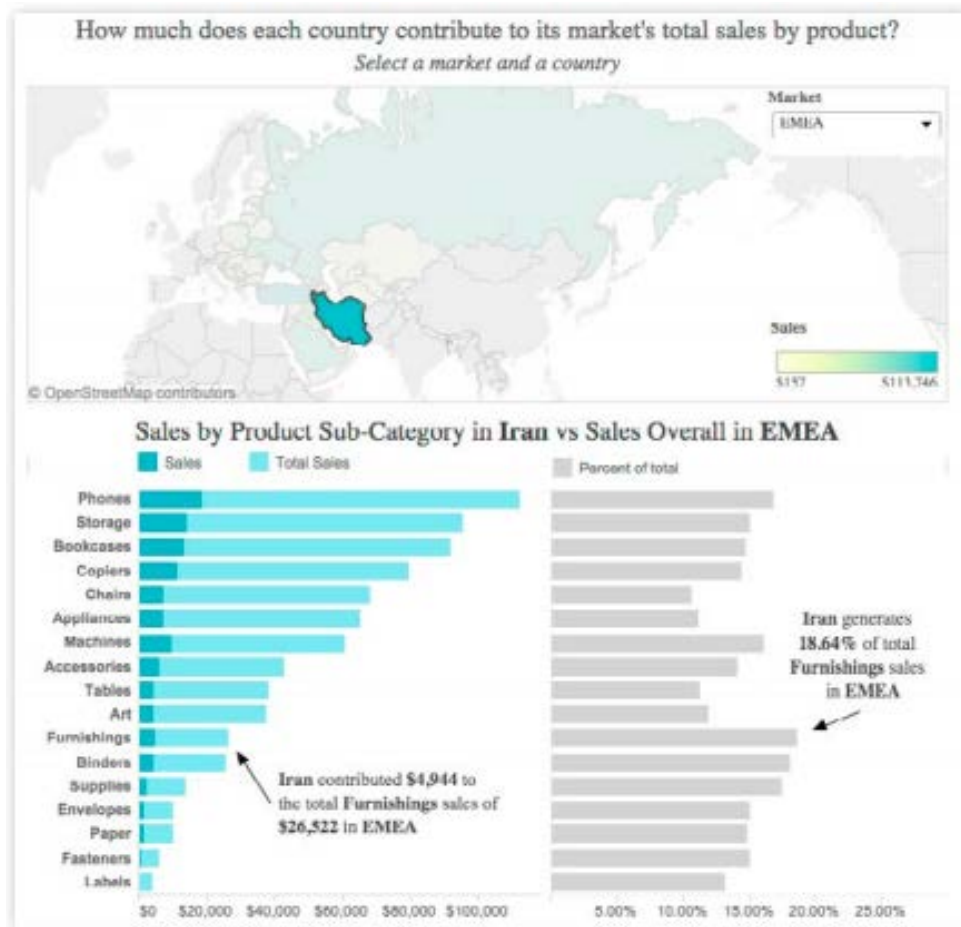


Figure 1: This interactive dashboard shows sales contribution by country and product.

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 a country's contribution to total sales across product categories (Figure 1).

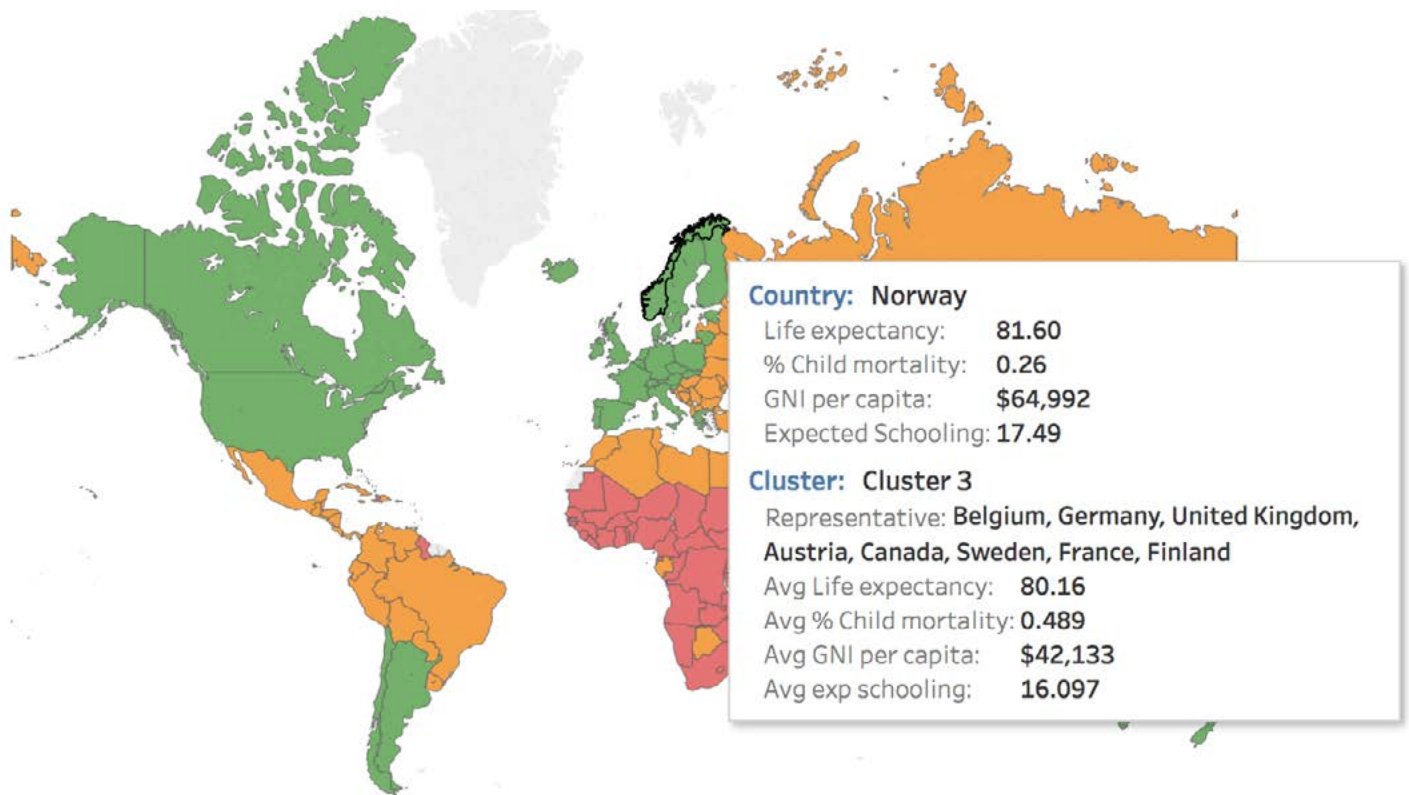


Figure 2: Automatically segment your data using clustering

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. In Figure 2, you can see results of clustering countries based on the United Nation’s development indicators. The algorithm recommends a 3-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. Tableau’s ability to iterate visually saves time tweaking scripts and re-running simulations in other tools.

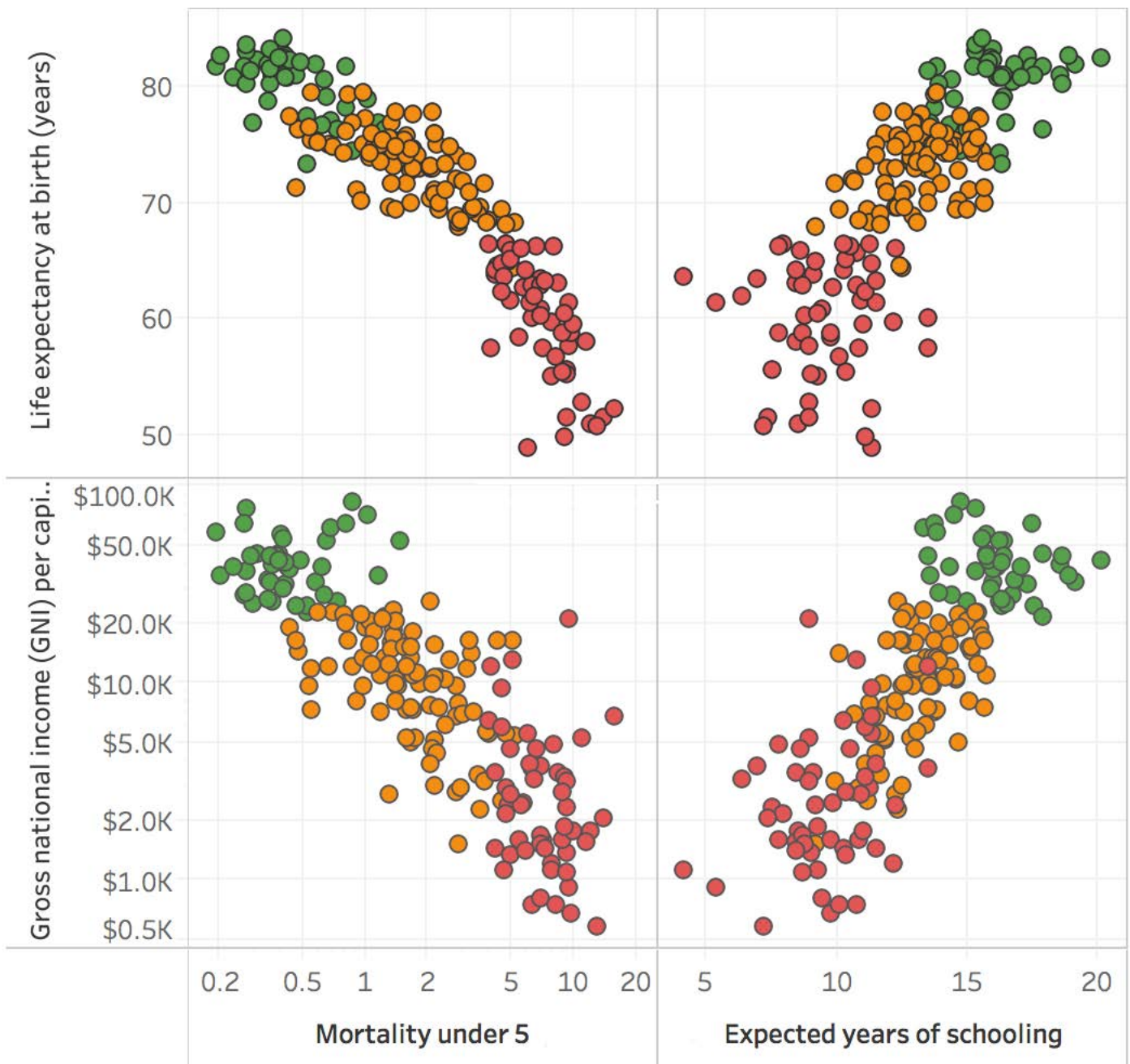


Figure 3: Segment and explore data in seconds

As seen in Figure 3, simply dragging the segmentation fields and measures used in clustering onto the canvas generates a small multiples view, highlighting differences between the three clusters of countries.

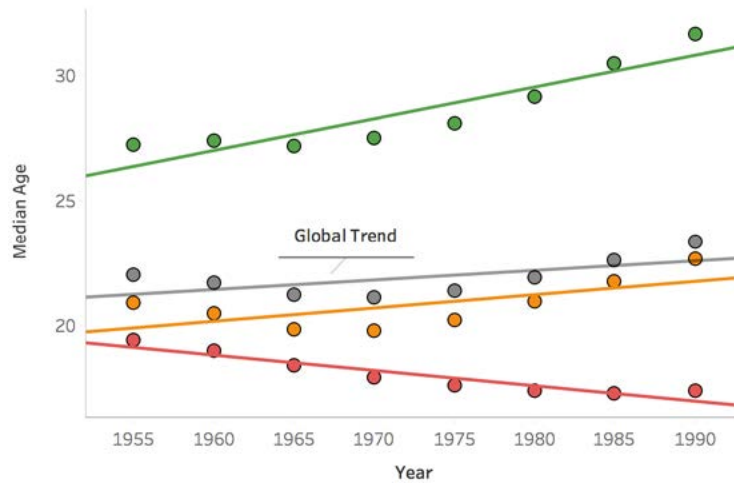


Figure 4: Explore and compare trends between segments

It is just as easy to look at how the average country in each group is trending against the average global age. Figure 4 clearly shows that while global values don't indicate any significant trends, the average age is increasing in developed countries, while underdeveloped countries' are decreasing due to large differences in birth rates. The trend line is automatically recomputed for each of the segments of interest without any additional work from the user.

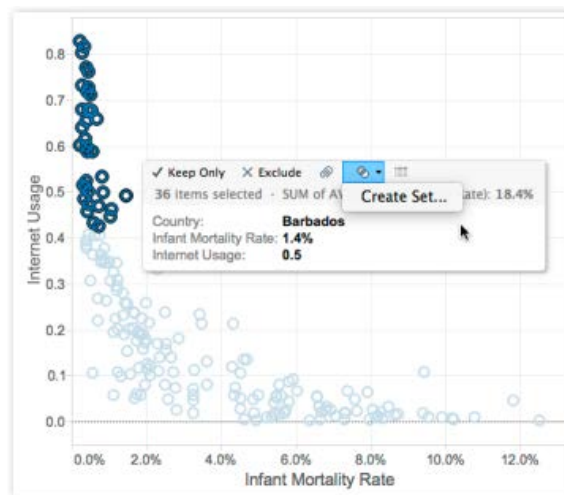


Figure 5: Define Sets graphically

Using sets, you can define collections of data objects either by manual selection (Figure 5) or 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 (Figure 6) 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.





Figure 6: Combine multiple Sets



Figure 7: Create a Group

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—for example, creating a group of the English-speaking countries as shown in Figure 7. This allows the analyst to customize the presentation and control the aggregation of data throughout the analysis.

In addition, Groups 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.

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 such as Sets and Groups will update as well. This means that cohort membership updates automatically without manually re-running 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, Stories*

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 by providing the following capabilities:

---

Simple controls:

A flexible set of input controls allows you to add text, numeric inputs, or even more complex controls such as sliders.

Full platform integration:

You can use the input values across Tableau to control thresholds in expressions, drive the cardinality of a report, filter data sets, or do any combination of these.

Snapshot interesting results:

Easily flag and share scenarios using Tableau's ability to store input values but keep analysis live and updating.

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 8, 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 the impact to key metrics.



Figure 8: With this parameter-driven sales report, the interactor can explore the effect of quotas, commissions, and salaries within the organization.

When combined with Stories (Tableau's way of building a narrative with data), Parameters allow you to take snapshots of interesting results and continue exploring. Stories allow you to construct a presentation that continues to update with data changes and viz modifications. However, Stories are smart enough to retain Parameter values, so you can flag scenarios and have confidence 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 insight, while simple generation of intuitive visuals allows 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: Smart Authoring Language, LOD Expressions, Quick Table Calculations, Statistical Functions*

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. To fully empower analysts, the language should have the following capabilities:

---

Expressibility:

Author calculations using a robust computational framework backed by a library of functions.

Flexible aggregations:

Support aggregation at multiple levels of detail within the same analysis component.

Result set computations:

Enable complex lags and iterative calculations dependent on the order of data.

Although Tableau is easy to use, we also provide a powerful language backed by a library that can express complex logic. With calculated fields, you can easily perform arithmetic operations, express conditional logic, or perform specialized operations on specific data types. Two key capabilities that enable advanced analysis are Level of Detail (LOD) Expressions and Table Calculations.

A relatively new addition to the calculation language, LOD Expressions have greatly augmented the power and expressibility of the calculation language. With this new capability, 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 9 shows the running sum of purchase history for cohorts of customers bucketed by the quarter of their first purchase. (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 the earliest customers placed the biggest initial orders and remained loyal with subsequent large purchases. LOD Expressions turn segmentation that would otherwise require complex group-by statements in SQL into simple, intuitive expressions that are manipulable in Tableau's front-end.

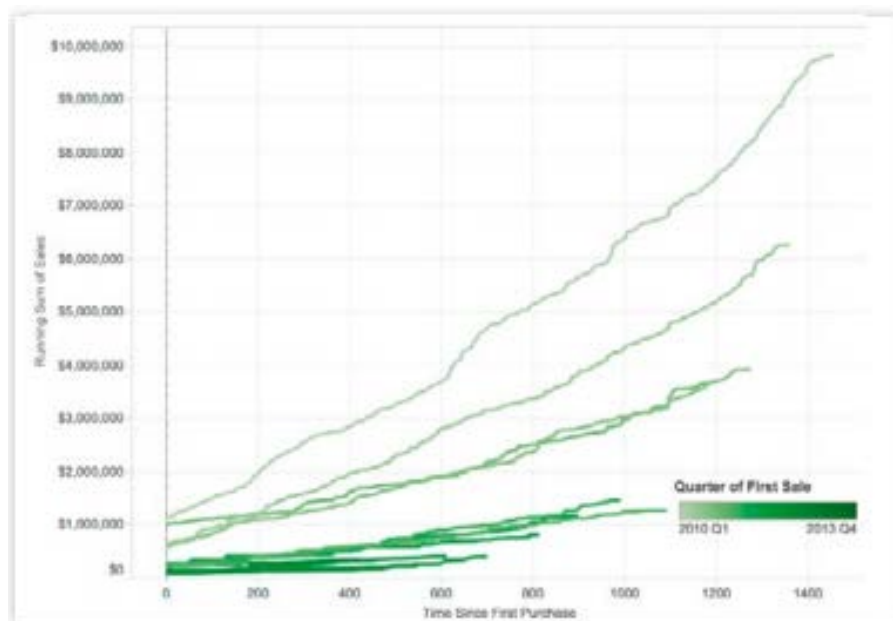


Figure 9: An LOD Expression is used to calculate the running sum of total sales by first quarter of purchase date

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 calculations 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. In fact, the running sum in Figure 9 was calculated using Quick Table Calculations. You can also create your own table calculations using the Table Calculation Functions in calculation language. These functions give workbook authors the power to precisely manipulate their result sets. Also, since all Table Calculation are expressible in the 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—such as manipulating aggregated data, and creating complex lags and 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 10: Down-sampling intraday data reveals possible insights about tipping patterns: drivers should consider working at night!

## 4. Time-Series and Predictive Analysis

Features: Trend 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 garner further insight, you may want to leverage predictive

capabilities. Ideally, you should be able to add predictive analytics without a large effort so you can explore multiple scenarios quickly. To work well with time series and predictive analysis, an analytics platform should support the following:

---

#### Seasonality exploration:

Examine seasonal effects with simple, intuitive tools.

#### Elegant sampling:

Flexible control over time granularity and aggregations.

#### Windowed calculations:

Perform arbitrary computations over moving temporal windows

#### Relative date filters:

Quickly filter to relevant ranges based on current values.

#### Integrated analytics objects:

Analytics objects, such as trend lines and forecasts, should automatically update with the data and support cohort analysis.

#### Simple quality metrics:

Quality metrics should be readily accessible for any model.

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. In Figure 10, we are studying the tipping patterns from all the taxi rides in New York City. We can easily adjust our granularity to find interesting patterns within the data. With a single click, you can disaggregate the data or view daily, monthly or annual summaries or quickly view year-over-year or quarter-over-quarter sales growth.

Leveraging the dual axis feature and discretized aggregation, you can start looking at multiple time series. In this case, the chart indicates that on weekends, there may be an inverse relationship

between the average number of rides on a given day and the average tip amount (Figure 11). This certainly could be the result of random variation or driven by another latent variable, but perhaps the quality of service goes down as volume increases. Without the ability to quickly inspect time series at different levels of granularity and aggregation, you might not be able to generate the question.

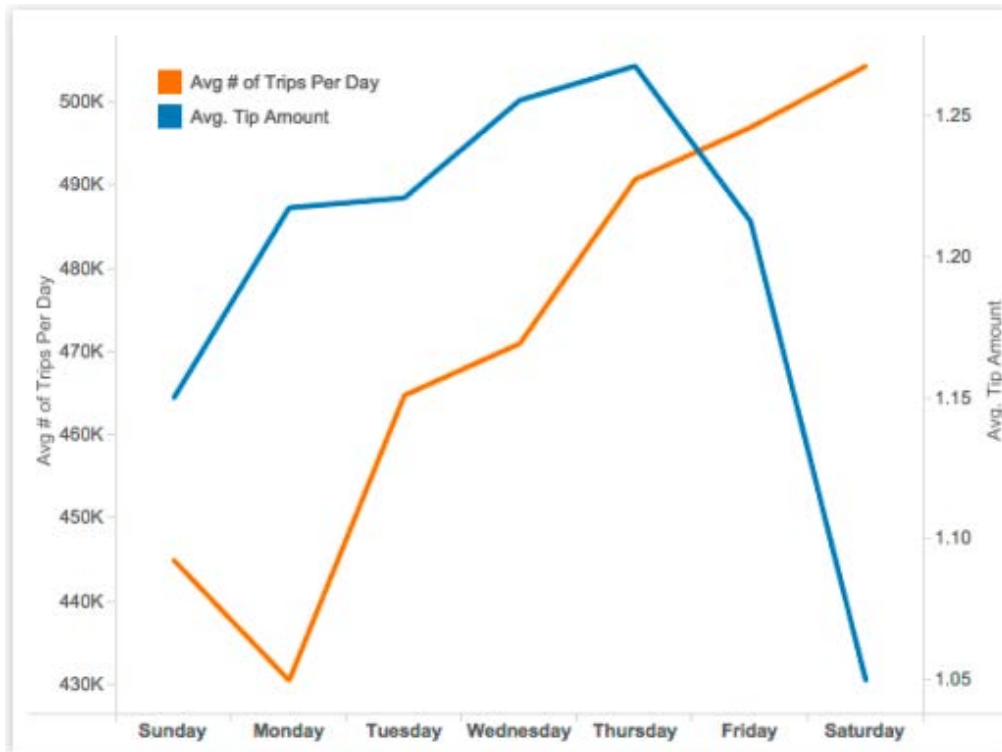


Figure 11: The dual axes plot shows an inverse relationship between rides and tip amount.

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, such as last week or last month. These periods are updated each time you open the view, 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 (Figure 12).

As previously discussed, Tableau’s Table Calculations feature lets you choose from a common set of time-series manipulations (Quick Table Calculations) or to use calculation language to write custom computations.





Figure 12: This time-series analysis shows the moving average of a stock price.

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.

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 modeling 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 13, 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.

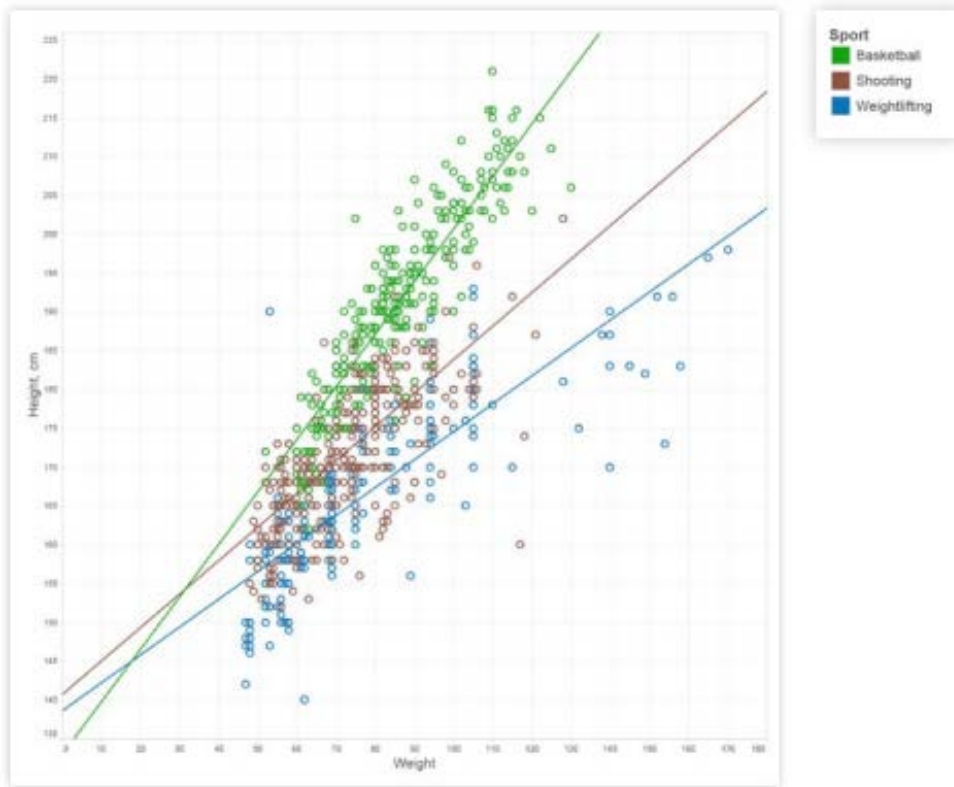


Figure 13: Trend lines highlight the relationship between height and weight by sport

As shown in Figure 14, Tableau contains a configurable forecasting ability for time-series data.

By default, Forecasting 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, from the length of the forecast to whether or not to account for seasonality, to the type of model used (additive or multiplicative).

The feature is also very easy to use, so a novice user can create a forecast with just a few clicks, while an advanced user can configure almost all aspects of the model. As with trend lines, details of the forecast quality are available with a single click. 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 UI (Figure 12).



Figure 14: Forecasting automatically predicts sales by region

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

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

Thus, 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. Supporting the integration with additional technologies enables the following:

---

Utilize virtually unlimited choice of methods:

Bring in algorithms and the latest advances from the broader community.

Leverage prior work:

Connect to preexisting logic and models to ensure best institutional practices and avoid replicating prior work.

Visualize and interrogate model results:

Use an intuitive front-end to help interpret, explore model results, and communicate to your colleagues.

Tableau integrates directly with Python and R 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.<sup>1</sup>

In Figures 15 and 16, you can see some examples in which R is used to compute descriptive statistics on a data set in Tableau, with Tableau used to visualize the results. Figure 15 is a graphical representation of the probability density for the data shown in the scatter plot and Figure 16 showcases significance testing.

<sup>1</sup> Tableau can also read R, SAS, and SPSS data files as a data source. While a complete discussion of data sources is beyond the scope of this paper, it's worth noting that Tableau can directly connect to the file outputs from several common stats programs

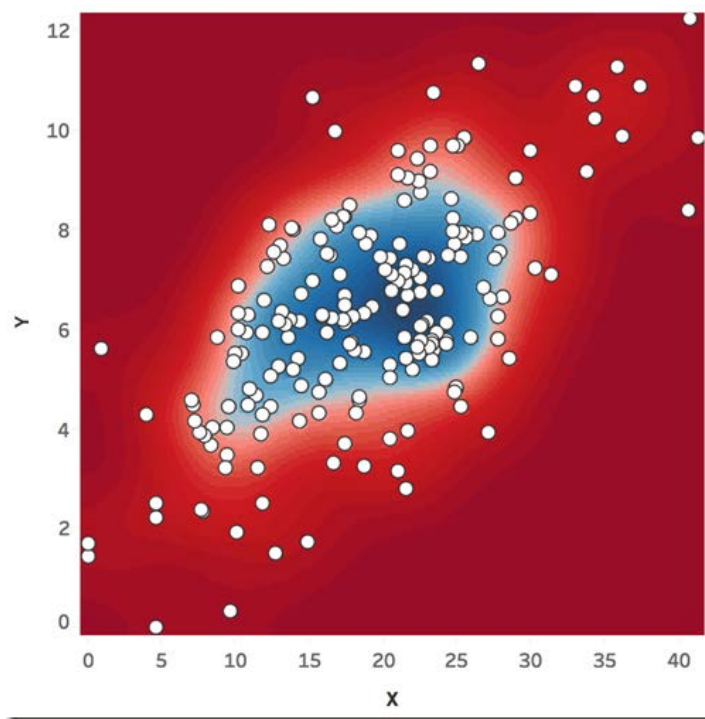


Figure 15: Bivariate kernel density computed in R, visualized in Tableau

Source: [boraberan.wordpress.com/](http://boraberan.wordpress.com/)

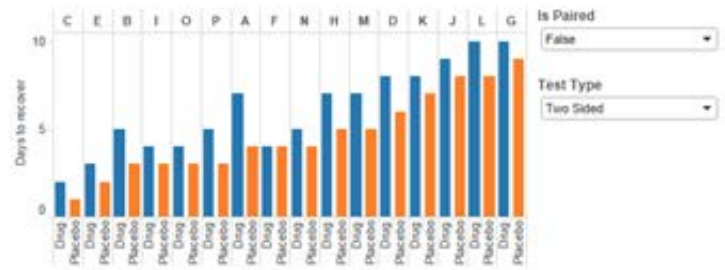
### Chi-square test of independence

**Contingency Table**

Order Priority	Delivery Truck	Express Air	Regular Air	
Critical	228	200	1,180	1,608
High	248	212	1,308	1,768
Medium	205	201	1,225	1,631
Low	250	190	1,280	1,720
Not Specified	215	180	1,277	1,672
	1,146	983	6,279	8,399

**P value**  
0.4205

### Welch Two Sample t-test



**P Value**  
0.1010

Figure 16: R and Tableau were used to calculate and visualize the results of significance testing

Source: [boraberan.wordpress.com/](http://boraberan.wordpress.com/)

The modeling can go much deeper than basic statistics. With R and Python integrations, you can analyze text to extract sentiment (Figure 17) or display results of prescriptive analytics such as supply chain or stock portfolio optimization (Figure 18).

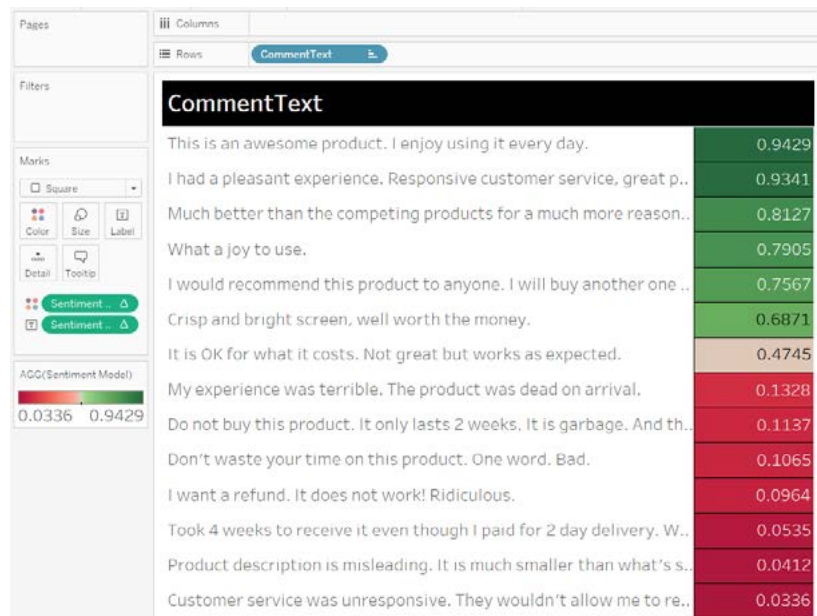


Figure 17: Sentiment Analysis

Source: [boraberan.wordpress.com/](http://boraberan.wordpress.com/)

In Figure 18, you can see an optimized portfolio computed and simulated in R, but visualized in Tableau.

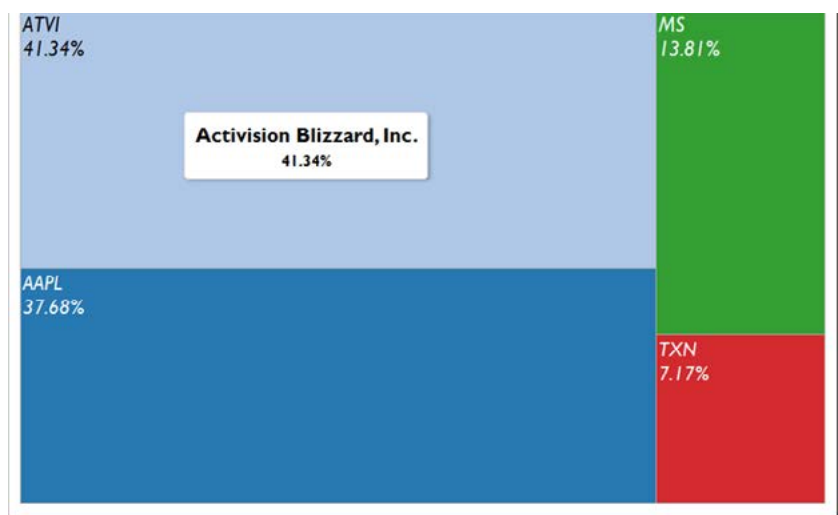


Figure 18: Stock portfolio optimization

Source: [boraberan.wordpress.com](http://boraberan.wordpress.com)

The integration supports running R or Python code directly inside Tableau, as well as models and scripts defined outside Tableau. Figure 19 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 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 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.

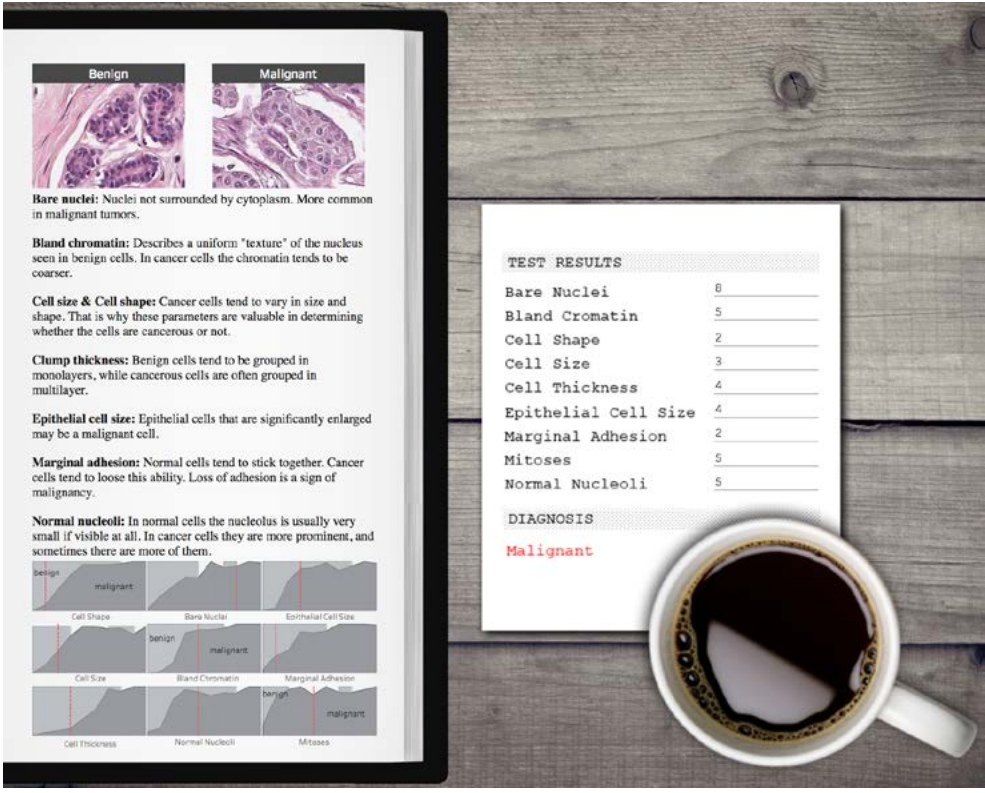


Figure 19: Ensemble machine learning model with input parameters

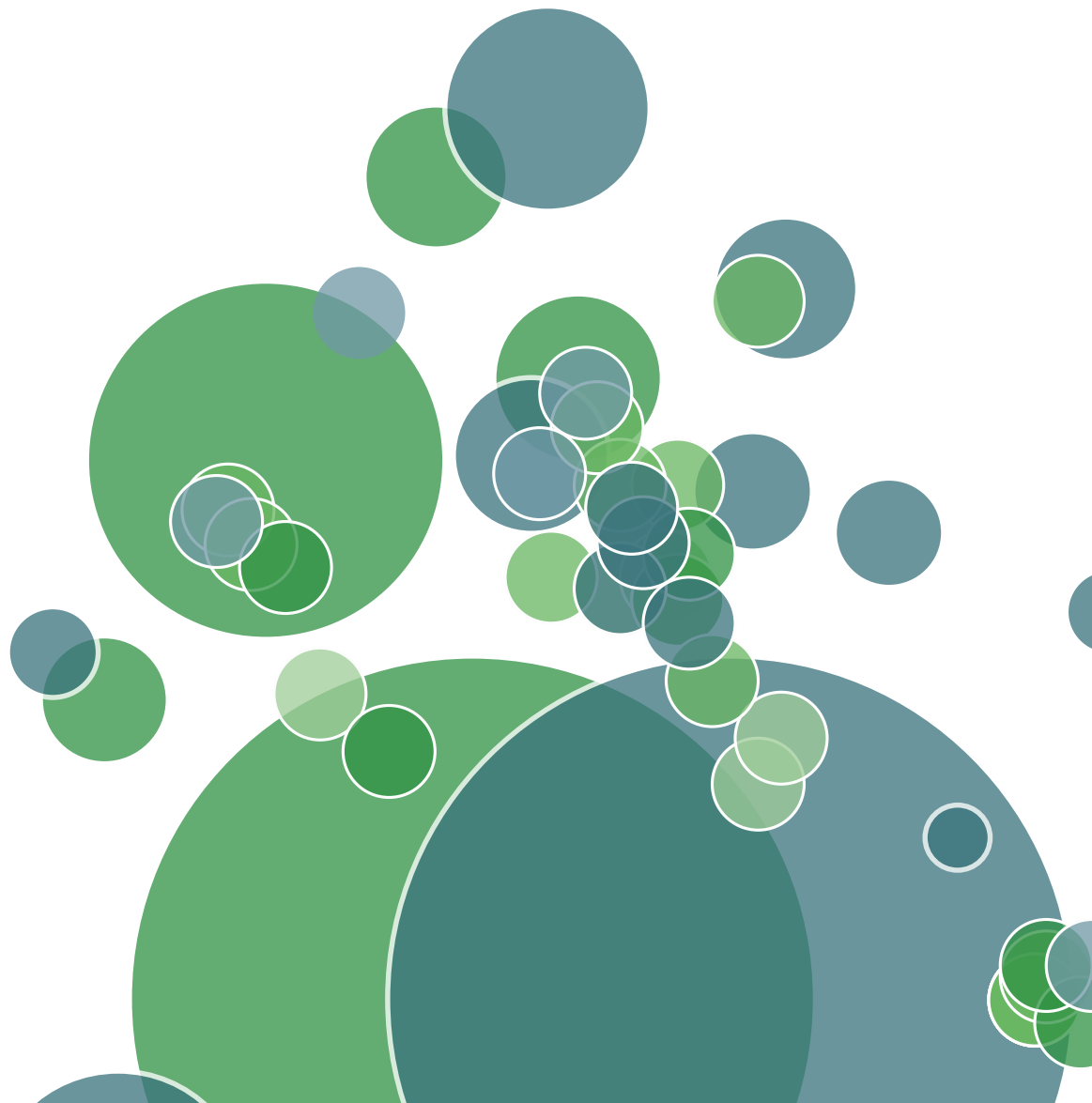
Source: [Building advanced analytics applications with TabPy on the Tableau Blog](#)

The combination of Tableau with Python and R 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 let's 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.

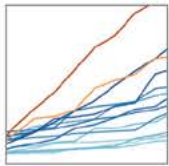
Tableau's flexible front-end allows business users to ask questions without needing to code or understand databases. Tableau also has the necessary analytical depth to be a powerful weapon in a data scientist's arsenal. By leveraging sophisticated calculations, R and Python integration, rapid cohort 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.





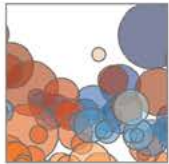
# About Tableau

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 encourages 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](#).



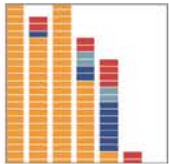
## Additional Resources

- [Download Free Trial](#)
- [Choose Your Own Adventure](#)



## 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](#)

