

This chapter focuses on the final phase of the Data Analytics Lifecycle: operationalize. In this phase, the project team delivers final reports, code, and technical documentation. At the conclusion of this phase, the team generally attempts to set up a pilot project and implement the developed models from Phase 4 in a production environment. As stated in Chapter 2, “Data Analytics Lifecycle,” teams can perform a technically accurate analysis, but if they cannot translate the results into a language that resonates with their audience, others will not see the value, and significant effort and resources will have been wasted. This chapter focuses on showing how to construct a clear narrative summary of the work and a framework for conveying the narrative to key stakeholders.

## 12.1 Communicating and Operationalizing an Analytics Project

As shown in Figure 12-1, the final phase in the Data Analytics Lifecycle focuses on operationalizing the project. In this phase, teams need to assess the benefits of the project work and set up a pilot to deploy the models in a controlled way before broadening the work and sharing it with a full enterprise or ecosystem of users. In this context, a pilot project can refer to a project prior to a full-scale rollout of the new algorithms or functionality. This pilot can be a project with a more limited scope and rollout to the lines of business, products, or services affected by these new models.

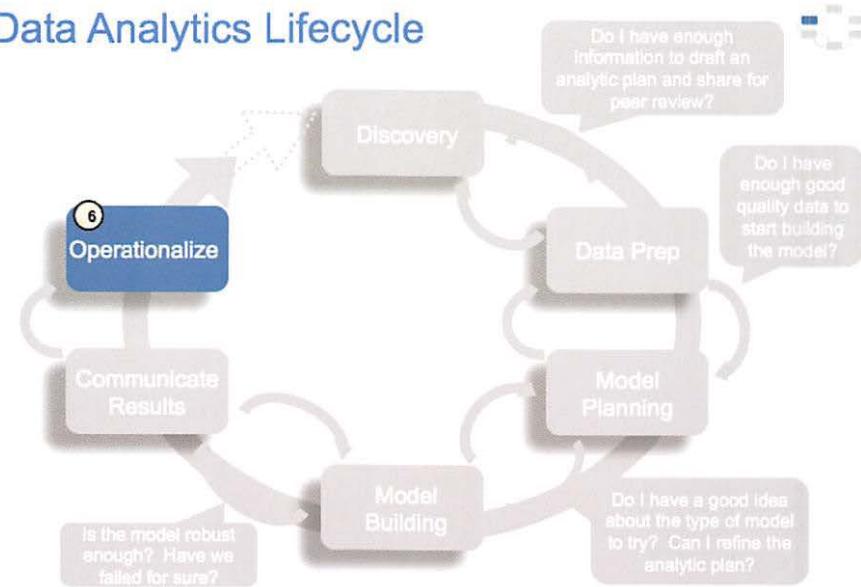


FIGURE 12-1 Data Analytics Lifecycle, Phase 6: operationalize

The team’s ability to quantify the benefits and share them in a compelling way with the stakeholders will determine if the work will move forward into a pilot project and ultimately be run in a production environment. Therefore, it is critical to identify the benefits and state them in a clear way in the final presentations.

As the team scopes the effort involved to deploy the analytical model as a pilot project, it also needs to consider running the model in a production environment for a discrete set of products or a single line of business, which tests the model in a live setting. This allows the team to learn from the deployment and make adjustments before deploying the application or code more broadly across the enterprise. This phase can bring in a new set of team members—namely, those engineers responsible for the production environment who have a new set of issues and concerns. This group is interested in ensuring that running the model fits smoothly into the production environment and the model can be integrated into downstream processes. While executing the model in the production environment, the team should aim to detect input anomalies before they are fed to the model, assess run times, and gauge competition for resources with other processes in the production environment.

Chapter 2 included an in-depth discussion of the Data Analytics Lifecycle, including an overview of the deliverables provided in its final phase, at which time it is advisable for the team to consider the needs of each of its main stakeholders and the deliverables, illustrated in Figure 12-2, to satisfy these needs.

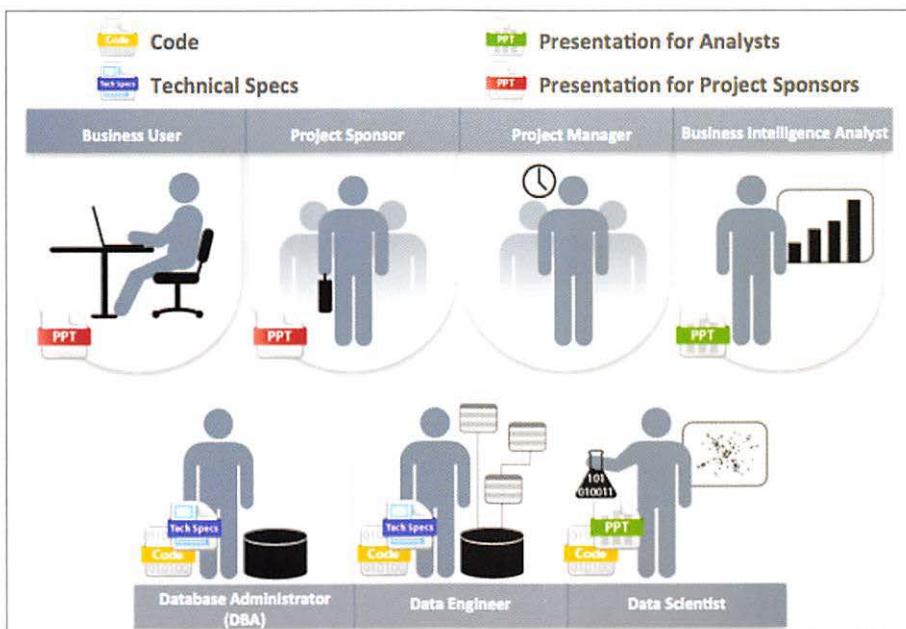


FIGURE 12-2 Key outputs from a successful analytic project

Following is a brief review of the key outputs for each of the main stakeholders of an analytics project and what they usually expect at the conclusion of a project:

- **Business User** typically tries to determine the benefits and implications of the findings to the business.
- **Project Sponsor** typically asks questions related to the business impact of the project, the risks and return on investment (ROI), and how the project can be evangelized within the organization and beyond.

- **Project Manager** needs to determine if the project was completed on time and within budget.
- **Business Intelligence Analyst** needs to know if the reports and dashboards he manages will be impacted and need to change.
- **Data Engineer and Database Administrator** (DBA) typically need to share the code from the analytical project and create technical documents that describe how to implement the code.
- **Data Scientists** need to share the code and explain the model to their peers, managers, and other stakeholders.

Although these seven roles represent many interests within a project, these interests usually overlap, and most of them can be met with four main deliverables:

- **Presentation for Project Sponsors** contains high-level takeaways for executive-level stakeholders, with a few key messages to aid their decision-making process. Focus on clean, easy visuals for the presenter to explain and for the viewer to grasp.
- **Presentation for Analysts**, which describes changes to business processes and reports. Data scientists reading this presentation are comfortable with technical graphs (such as Receiver Operating Characteristic [ROC] curves, density plots, and histograms) and will be interested in the details.
- **Code** for technical people, such as engineers and others managing the production environment
- **Technical specifications** for implementing the code

As a rule, the more executive the audience, the more succinct the presentation needs to be for project sponsors. Ensure that the presentation gets to the point quickly and frames the results in terms of value to the sponsor's organization. When presenting to other audiences with more quantitative backgrounds, focus more time on the methodology and findings. In these instances, the team can be more expansive in describing the outcomes, methodology, and analytical experiments with a peer group. This audience will be more interested in the techniques, especially if the team developed a new way of processing or analyzing data that can be reused in the future or applied to similar problems. In addition, use imagery or data visualization when possible. Although it may take more time to develop imagery, pictures are more appealing, easier to remember, and more effective to deliver key messages than long lists of bullets.

## 12.2 Creating the Final Deliverables

After reviewing the list of key stakeholders for data science projects and main deliverables, this section focuses on describing the deliverables in detail. To illustrate this approach, a fictional case study is used to make the examples more specific. Figure 12-3 describes a scenario of a fictional bank, YoyoDyne Bank, which would like to embark on a project to do churn prediction models of its customers. *Churn rate* in this context refers to the frequency with which customers sever their relationship as customers of YoyoDyne Bank or switch to a competing bank.

| Synopsis of YoyoDyne Bank Case Study                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>▪ YoyoDyne Bank is a retail bank that wants to improve its Net Present Value (NPV) and its customer retention rate.</li> <li>▪ It wants to establish an effective marketing campaign targeting customers to reduce the churn rate by at least five percent.</li> <li>▪ The bank wants to determine whether those customers are worth retaining. In addition, the bank wants to analyze reasons for customer attrition and what it can do to keep customers from leaving.</li> <li>▪ The bank wants to build a data warehouse to support marketing and other related customer care groups.</li> </ul> |

**FIGURE 12-3** Synopsis of YoyoDyne Bank case study example

Based on this information, the data science team may create an analytics plan similar to Figure 12-4 during the project.

| Components of Analytic Plan                | Retail Banking: YoyoDyne Bank                                                                                                                                                                                                                                                                                                                      |
|--------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Discovery Business Problem Framed</b>   | How can the bank identify customers with the highest likelihood for churn?                                                                                                                                                                                                                                                                         |
| <b>Initial Hypotheses</b>                  | Transaction volume and type are key predictors of churn rates                                                                                                                                                                                                                                                                                      |
| <b>Data and Scope</b>                      | 5 months of customer account history                                                                                                                                                                                                                                                                                                               |
| <b>Model Planning - Analytic Technique</b> | Logistic regression to identify most influential factors predicting churn                                                                                                                                                                                                                                                                          |
| <b>Result and Key Findings</b>             | <p>Key predictors of churn are:</p> <ol style="list-style-type: none"> <li>1. Once customers stop using their accounts for gas and groceries, their account holdings quickly diminish and the customers churn.</li> <li>2. If the customers use their debit card fewer than 5 times per month, they will leave the bank within 60 days.</li> </ol> |
| <b>Business Impact</b>                     | By targeting customers who are at high risk for churn, customer attrition can be reduced by 23%. This would save \$3 million in lost customer revenue and avoid \$1.5 million in new customer acquisition costs each year for the bank.                                                                                                            |

**FIGURE 12-4** Analytics plan for YoyoDyne Bank case study

In addition to guiding the model planning and methodology, the analytic plan contains components that can be used as inputs for writing about the scope, underlying assumptions, modeling techniques, initial hypotheses, and key findings in the final presentations. After spending substantial amounts of time in the modeling and performing in-depth data analysis, it is critical to reflect on the project work and consider

the context of the problems the team set out to solve. Review the work that was completed during the project, and identify observations about the model outputs, scoring, and results. Based on these observations, begin to identify the key messages and any unexpected insights.

In addition, it is important to tailor the project outputs to the audience. For a project sponsor, show that the team met the project goals. Focus on what was done, what the team accomplished, what ROI can be anticipated, and what business value can be realized. Give the project sponsor talking points to evangelize the work. Remember that the sponsor needs to relay the story to others, so make this person's job easy, and help ensure the message is accurate by providing a few talking points. Find ways to emphasize ROI and business value, and mention whether the models can be deployed within performance constraints of the sponsor's production environment.

In some organizations, the data science team may not be expected to make a full business case for future projects and implementation of the models. Instead, it needs to be able to provide guidance about the impact of the models to enable the project sponsor, or someone designated by that person, to create a business case to advocate for the pilot and subsequent deployment of this functionality. In other words, the data science team can assist in this effort by putting the results of the modeling and data science work into context to help assess the actual value and cost of implementing this work more broadly.

When presenting to a technical audience such as data scientists and analysts, focus on how the work was done. Discuss how the team accomplished the goals and the choices it made in selecting models or analyzing the data. Share analytical methods and decision-making processes so other analysts can learn from them for future projects. Describe methods, techniques, and technologies used, as this technical audience will be interested in learning about these details and considering whether the approach makes sense in this case and whether it can be extended to other, similar projects. Plan to provide specifics related to model accuracy and speed, such as how well the model will perform in a production environment.

Ideally, the team should consider starting the development of the final presentation during the project rather than at the end of the project as commonly occurs. This approach ensures that the team always has a version of the presentation with working hypotheses to show stakeholders, in case there is a need to show a work-in-process version of the project progress on short notice. In fact, many analysts write the executive summary at the outset of a project and then continually refine it over time so that at the end of the project, portions of the final presentation are already completed. This approach also reduces the chance that the team members will forget key points or insights discovered during the project. Finally, it reduces the amount of work to be done on the presentation at the conclusion of the project.

### 12.2.1 Developing Core Material for Multiple Audiences

Because some of the components of the projects can be used for different audiences, it can be helpful to create a core set of materials regarding the project, which can be used to create presentations for either a technical audience or an executive sponsor.

Table 12-1 depicts the main components of the final presentations for the project sponsor and an analyst audience. Notice that teams can create a core set of materials in these seven areas, which can be used for the two presentation audiences. Three areas (Project Goals, Main Findings, and Model Description), can be used as is for both presentations. Other areas need additional elaboration, such as the Approach. Still other areas, such as the Key Points, require different levels of detail for the analysts and data scientists than for the project sponsor. Each of these main components of the final presentation is discussed in subsequent sections.

**TABLE 12-1** Comparison of Materials for Sponsor and Analyst Presentations

| Presentation Component         | Project Sponsor Presentation                                                                                                                   | Analyst Presentation                                                                                                                                                                                                                                                                                                                             |
|--------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Project Goals                  | List top 3–5 agreed-upon goals.                                                                                                                |                                                                                                                                                                                                                                                                                                                                                  |
| Main Findings                  | Emphasize key messages.                                                                                                                        |                                                                                                                                                                                                                                                                                                                                                  |
| Approach                       | High-level methodology                                                                                                                         | High-level methodology<br>Relevant details on modeling techniques and technology                                                                                                                                                                                                                                                                 |
| Model Description              | Overview of the modeling technique                                                                                                             |                                                                                                                                                                                                                                                                                                                                                  |
| Key Points Supported with Data | Support key points with simple charts and graphics (example: bar charts).                                                                      | Show details to support the key points.<br>Analyst-oriented charts and graphs, such as ROC curves and histograms<br>Visuals of key variables and significance of each                                                                                                                                                                            |
| Model Details                  | Omit this section, or discuss only at a high level.                                                                                            | Show the code or main logic of the model, and include model type, variables, and technology used to execute the model and score data.<br>Identify key variables and impact of each.<br>Describe expected model performance and any caveats.<br>Detailed description of the modeling technique<br>Discuss variables, scope, and predictive power. |
| Recommendations                | Focus on business impact, including risks and ROI.<br><br>Give the sponsor salient points to help her evangelize work within the organization. | Supplement recommendations with implications for the modeling or for deploying in a production environment.                                                                                                                                                                                                                                      |

## 12.2.2 Project Goals

The Project Goals portion of the final presentation is generally the same, or similar, for sponsors and for analysts. For each audience, the team needs to reiterate the goals of the project to lay the groundwork for

the solution and recommendations that are shared later in the presentation. In addition, the Goals slide serves to ensure there is a shared understanding between the project team and the sponsors and confirm they are aligned in moving forward in the project. Generally, the goals are agreed on early in the project. It is good practice to write them down and share them to ensure the goals and objectives are clearly understood by both the project team and the sponsors.

Figures 12-5 and 12-6 show two examples of slides for Project Goals. Figure 12-5 shows three goals for creating a predictive model to anticipate customer churn. The points on this version of the Goals slide emphasize what needs to be done, but not why, which will be included in the alternative.

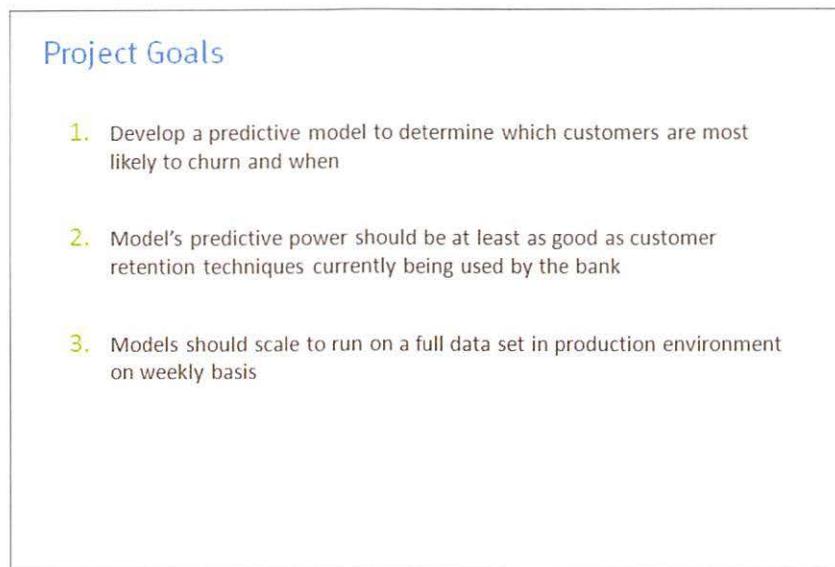


FIGURE 12-5 Example of Project Goals slide for YoyoDyne case study

Figure 12-6 shows a variation of the previous Project Goals slide in Figure 12-5. It is a summary of the situation prior to listing the goals. Keep in mind that when delivering final presentations, these deliverables are shared within organizations, and the original context can be lost, especially if the original sponsor leaves the group or changes roles. It is good practice to briefly recap the situation prior to showing the project goals. Keep in mind that adding a situation overview to the Goals slide does make it appear busier. The team needs to determine whether to split this into a separate slide or keep it together, depending on the audience and the team's style for delivering the final presentation.

One method for writing the situational overview in a succinct way is to summarize it in three bullets, as follows:

- **Situation:** Give a one-sentence overview of the situation that has led to the analytics project.
- **Complication:** Give a one-sentence overview of the need for addressing this now. Something has triggered the organization to decide to take action at this time. For instance, perhaps it lost 100

customers in the past two weeks and now has an executive mandate to address an issue, or perhaps it has lost five points of market share to its biggest competitor in the past three months. Usually, this sentence represents the driver for why a particular project is being initiated at this time, rather than in some vague time in the future.

- **Implication:** Give a one-sentence overview of the impact of the complication. For instance, if the bank fails to address its customer attrition problem, it stands to lose its dominant market position in three key markets. Focus on the business impact to illustrate the urgency of doing the project.

## Situation & Project Goals

**Situation**

1. YoyoDyne Bank wants to improve the Net Present Value (NPV) and retention rate of the customers
2. In the last 90 days, YoyoDyne has lost 6 of its top 100 customers and is seeing increased competition from its biggest competitor
3. Without a fast remediation plan, YoyoDyne risks losing its dominant position in three key markets

**Goals of YoyoDyne “Churn Project”**

1. Develop a predictive model to determine which customers are most likely to churn and when
2. Model's predictive power should be at least as good as customer retention techniques currently being used by the bank
3. Models should scale to run on a full data set in production environment on weekly basis

FIGURE 12-6 Example of Situation & Project Goals slide for YoyoDyne case study

### 12.2.3 Main Findings

Write a solid executive summary to portray the main findings of a project. In many cases, the summary may be the only portion of the presentation that hurried managers will read. For this reason, it is imperative to make the language clear, concise, and complete. Those reading the executive summary should be able to grasp the full story of the project and the key insights in a single slide. In addition, this is an opportunity to provide key talking points for the executive sponsor to use to evangelize the project work with others in the customer's organization. Be sure to frame the outcomes of the project in terms of both quantitative and qualitative business value. This is especially important if the presentation is for the project sponsor. The executive summary slide containing the main findings is generally the same for both sponsor and analyst audiences.

Figure 12-7 shows an example of an executive summary slide for the YoyoDyne case study. It is useful to take a closer look at the parts of the slide to make sure it is clear. Keep in mind this is not the only format for conveying the Executive Summary; it varies based on the author's style, although many of the key components are common themes in Executive Summaries.

## Executive Summary

*Running an early churn warning test each day using social media can reduce annual churn by 30 % and save \$4.5M annually*

- Customers churn within 60 days of changing their spending habits
  - Once customers stop using their accounts for gas and groceries, their account holdings quickly diminish and the customers churn
  - If customers use their debit card fewer than 5 times per month, they will leave the bank within 60 days
- Combining social networking data and existing CRM data increases the model's predictive power to identify churners
  - We can pinpoint social media chatter from bank customers and influence of chunner's contacts
  - With CRM data, we can identify 20% of churners, adding social media data increases this to 30%
- Models can run in minutes, rather than current process of monthly cycles

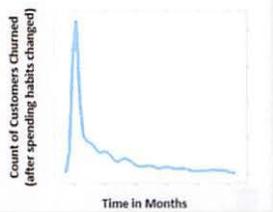


FIGURE 12-7 Example of Executive Summary slide for YoyoDyne case study

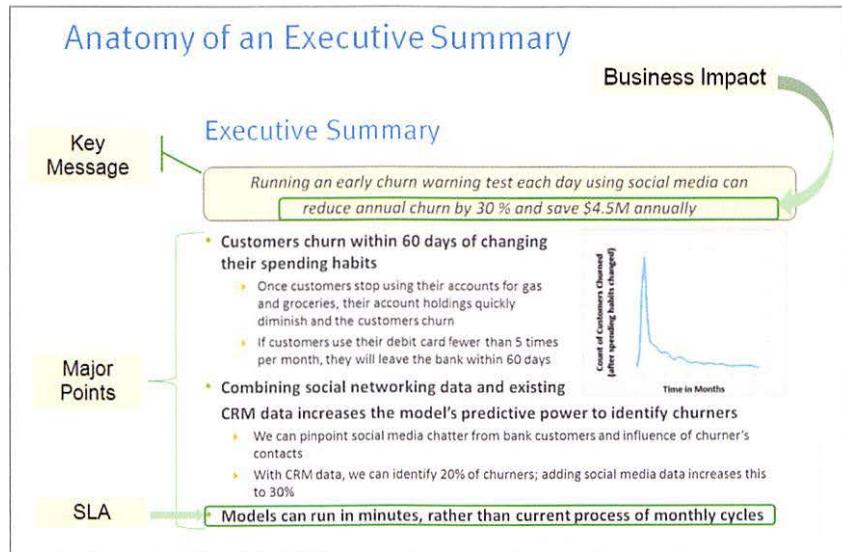
The key message should be clear and conspicuous at the front of the slide. It can be set apart with color or shading, as shown in Figure 12-8; other techniques can also be used to draw attention to it. The key message may become the single talking point that executives or the project sponsor take away from the project and use to support the team's recommendation for a pilot project, so it needs to be succinct and compelling. To make this message as strong as possible, measure the value of the work and quantify the cost savings, revenue, time savings, or other benefits to make the business impact concrete.

Follow the key message with three major supporting points. Although Executive Summary slides can have more than three major points, going beyond three ideas makes it difficult for people to recall the main points, so it is important to ensure that the ideas remain clear and limited to the few most impactful ideas the team wants the audience to take away from the work that was done. If the author lists ten key points, messages become diluted, and the audience may remember only one or two main points.

In addition, because this is an analytics project, be sure to make one of the key points related to if, and how well, the work will meet the sponsor's service level agreement (SLA) or expectations. Traditionally, the SLA refers to an arrangement between someone providing services, such as an information technology (IT) department or a consulting firm, and an end user or customer. In this case, the SLA refers to system performance, expected uptime of a system, and other constraints that govern an agreement. This term has become less formal and many times conveys system performance or expectations more generally related to performance or timeliness. It is in this sense that SLA is being used here. Namely, in this context, SLA

refers to the expected performance of a system and the intent that the models developed will not adversely impact the expected performance of the system into which they are integrated.

Finally, although it's not required, it is often a good idea to support the main points with a visual or graph. Visual imagery serves to make a visceral connection and helps retain the main message with the reader.



**FIGURE 12-8** Anatomy of an Executive Summary slide

#### 12.2.4 Approach

In the Approach portion of the presentation, the team needs to explain the methodology pursued on the project. This can include interviews with domain experts, the groups collaborating within the organization, and a few statements about the solution developed. The objective of this slide is to ensure the audience understands the course of action that was pursued well enough to explain it to others within the organization. The team should also include any additional comments related to working assumptions the team followed as it performed the work, because this can be critical in defending why they followed a specific course of action.

When explaining the solution, the discussion should remain at a high level for the project sponsors. If presenting to analysts or data scientists, provide additional detail about the type of model used, including the technology and the actual performance of the model during the tests. Finally, as part of the description of the approach, the team may want to mention constraints from systems, tools, or existing processes and any implications for how these things may need to change with this project.

Figure 12-9 shows an example of how to describe the methodology followed during a data science project to a sponsor audience.

### Approach (for Sponsors)

- Interviewed 14 members of retail lending team to understand YoyoDyne's lending policies and marketing practices for customer retention
- Collaborated with IT to identify relevant datasets and assess data quality and availability
- Developed churn model to identify customers most likely to leave the bank
  - ▶ Identify most influential factors
  - ▶ Provide greater explanatory power for analyzing impact of different factors on churn
- Mined and added social media data to the model to improve predictive power
- Worked with IT to simulate model performance within YoyoDyne's production environment

FIGURE 12-9 Example describing the project methodology for project sponsors

Note that the third bullet describes the churn model in general terms. Furthermore, the subbullets provide additional details in nontechnical terms. Compare this approach to the variation shown in Figure 12-10.

### Approach (for Analysts)

- Interviewed 14 members of retail lending team to understand YoyoDyne's lending policies and marketing practices for customer retention
- Collaborated with IT to identify relevant datasets and assess data quality and availability
- Developed churn model in R using a Generalized Addictive Modeling technique
  - ▶ Minimizes variable transformations and binning
  - ▶ Provide greater explanatory power for analyzing impact of different factors on churn
- Examined impact of social network variables and found that it helped identify more potential churners
- Work with IT to simulate model performance within YoyoDyne's production environment
- The model can be rapidly scored in the database over large datasets using a SQL code generator for the purpose

FIGURE 12-10 Example describing the project methodology for analysts and data scientists

Figure 12-10 shows a variation on the approach and methodology used in the data science project. In this case, most of the language and description are the same as in the example for project sponsors.

The main difference is that this version contains additional detail regarding the kind of model used and the way the model will score data quickly to meet the SLA. These differences are highlighted in the boxes shown in Figure 12-10.

### 12.2.5 Model Description

After describing the project approach, teams generally include a description of the model that was used. Figure 12-11 provides the model description for the Yoyodyne Bank example. Although the Model Description slide can be the same for both audiences, the interests and objectives differ for each. For the sponsor, the general methodology needs to be articulated without getting into excessive detail. Convey the basic methodology followed in the team's work to allow the sponsor to communicate this to others within the organization and provide talking points.

Mentioning the scope of the data used is critical. The purpose is to illustrate thoroughness and exude confidence that the team used an approach that accurately portrays its problem and is as free from bias as possible. A key trait of a good data scientist is the ability to be skeptical of one's own work. This is an opportunity to view the work and the deliverable critically and consider how the audience will receive the work. Try to ensure it is an unbiased view of the project and the results.

Assuming that the model will meet the agreed-upon SLAs, mention that the model will meet the SLAs based on the performance of the model within the testing or staging environment. For instance, one may want to indicate that the model processed 500,000 records in 5 minutes to give stakeholders an idea of the speed of the model during run time. Analysts will want to understand the details of the model, including the decisions made in constructing the model and the scope of the data extracts for testing and training. Be prepared to explain the team's thought process on this, as well as the speed of running the model within the test environment.

### Model Description

- **Overview of Basic Methodology:** predict the likelihood of churn for each customer. Identify customers with a greater probability for churn then compare with actual churn outcomes to train the algorithm and enable predictions for existing customers.
- **Model:** Logistic regression model
- **Dependent variable:** Binary variable, of churn/no churn
- **Scope:**
  - 500,000 Yoyodyne bank customers, based on churn within a 150 day period after 1/31/2011
  - 500,000 Customers with all churners through 6/30/11, plus a random sample of 45,000 accounts
  - All selected customers were Active, Suspended or Pending as of 2011-01-31
  - Call History detail data extracted from Call Data Record Warehouse for customers from 1/31/11 to 6/30/11
- **Sampling**
  - Training sample: 50,000 subscribers
  - Testing sample: 100,000 subscribers
- **The model developed has predictive power at least as good as the bank's current churn model**
  - We created a baseline model without social networking variables and the bank's marketing analytics team verified that the predictive power was at least as good as the current model
  - Social networking variables were added to the model and that further increased its predictive power

FIGURE 12-11 Example of a model description for a data science project

## 12.2.6 Key Points Supported with Data

The next step is to identify key points based on insights and observations resulting from the data and model scoring results. Find ways to illustrate the key points with charts and visualization techniques, using simpler charts for sponsors and more technical data visualization for analysts and data scientists.

Figure 12-12 shows an example of providing supporting detail regarding the rate of bank customers who would churn in various months. When developing the key points, consider the insights that will drive the biggest business impact and can be defended with data. For project sponsors, use simple charts such as bar charts, which illustrate data clearly and enable the audience to understand the value of the insights. This is also a good point to foreshadow some of the team's recommendations and begin tying together ideas to demonstrate what led to the recommendations and why. In other words, this section supplies the data and foundation for the recommendations that come later in the presentation. Creating clear, compelling slides to show the key points makes the recommendations more credible and more likely to be acted upon by the customer or sponsor.

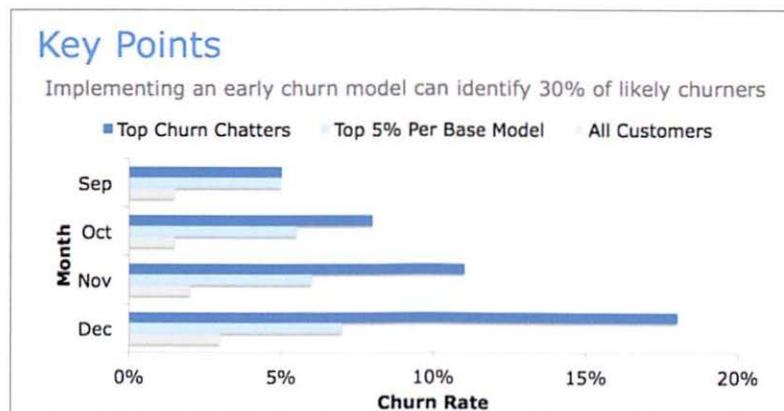


FIGURE 12-12 Example of a presentation of key points of a data science project shown as a bar chart

For analyst presentations, use more granular or technical charts and graphs. In this case, appropriate visualization techniques include dot charts, density plots, ROC curves, or histograms of a data distribution to support decisions made in the modeling techniques. Basic concepts of data visualization are discussed later in the chapter.

## 12.2.7 Model Details

Model details are typically needed by people who have a more technical understanding than the sponsors, such as those who will implement the code, or colleagues on the analytics team. Project sponsors are typically less interested in the model details; they are usually more focused on the business implications of the work rather than the details of the model. This portion of the presentation needs to show the code or main logic of the model, including the model type, variables, and technology used to execute

the model and score the data. The model details segment of the presentation should focus on describing expected model performance and any caveats related to the model performance. In addition, this portion of the presentation should provide a detailed description of the modeling technique, variables, scope, and expected effectiveness of the model.

This is where the team can provide discussion or written details related to the variables used in the model and explain how or why these variables were selected. In addition, the team should share the actual code (or at least an excerpt) developed to explain what was created and how it operates. This also serves to foster discussion related to any additional constraints or implications related to the main logic of the code. In addition, the team can use this section to illustrate details of the key variables and the predictive power of the model, using analyst-oriented charts and graphs, such as histograms, dot charts, density plots, and ROC curves.

Figure 12-13 provides a sample slide describing the data variables, and Figure 12-14 shows a sample slide with a technical graph to support the work.

## Model Details

- Candidate variables: 22 from CRM, 154 from call history, and 12 social networking variables
- Through PCA and discussion with domain experts, we reduced ~190 variables to the 9 most predictive of customer churn
- General Additive Model (GAM) model built in R :

```
gam.wsn.b2 <- bam(volchurn ~ 120, p~  
  s(var1, bs="cs", by=c30, k=length(custom.knots))  
  +s(var2, bs="cs", by=c30)  
  +s(var3, bs="cs", k=5)  
  +s(var4, bs="cs", k=5, by=c30)  
  +s(var5, bs="cs", k=5)  
  +var6  
  +var7  
  +s(var8)  
  +s(var9),  
  knots=list(var1=custom.knots),  
  data=train.df, family=binomial, weight=weight, gamma=1.4)
```

FIGURE 12-13 Example of model details showing model type and variables

As part of the model detail description, guidance should be provided regarding the speed with which the model can run in the test environment; the expected performance in a live, production environment; and the technology needed. This kind of discussion addresses how well the model can meet the organization's SLA.

This section of the presentation needs to include additional caveats, assumptions, or constraints of the model and model performance, such as systems or data the model needs to interact with, performance

issues, and ways to feed the outputs of the model into existing business processes. The author of this section needs to describe the relationships of the main variables on the project objectives, such as the effects of key variables on predicting churn, and the relationship of key variables to other variables. The team may even want to make suggestions to improve the model, highlight any risks to introducing bias into the modeling technique, or describe certain segments of the data that may skew the overall predictive power of the methodology.

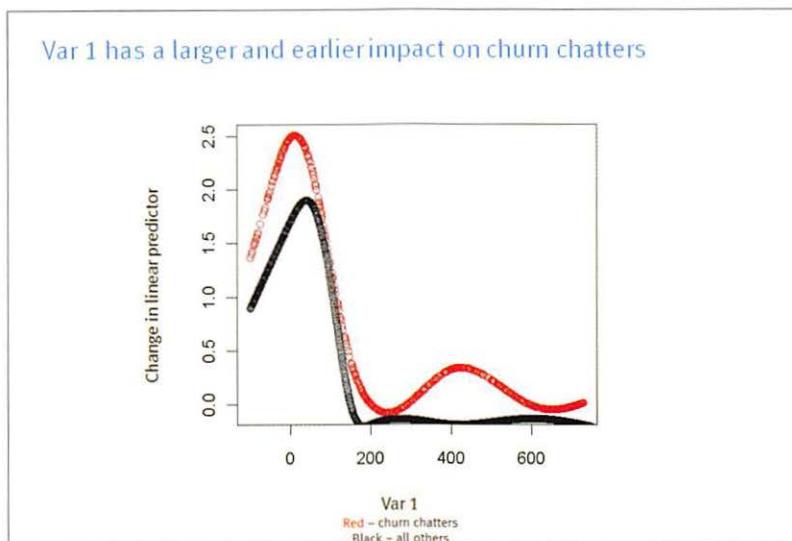


FIGURE 12-14 Model details comparing two data variables

### 12.2.8 Recommendations

The final main component of the presentation involves creating a set of recommendations that include how to deploy the model from a business perspective within the organization and any other suggestions on the rollout of the model's logic. For the Yoyodyne Bank example, Figure 12-15 provides possible recommendations from the project. In this section of the presentation, measuring the impact of the improvements and stating how to leverage that impact within the recommendations are key. For instance, the presentation might mention that every customer retained represents a time savings of six hours for one of the bank's account managers or \$50,000 in savings of new account acquisitions, due to marketing costs, sales, and system-related costs.

For a presentation to a project sponsor audience, focus on the business impact of the project, including risks and ROI. Because project sponsors will be most interested in the business impact of the project, the presentation should also provide the sponsor with salient points to help evangelize the work within the organization. When preparing a presentation for analysts, supplement the main set of recommendations with any implications for the modeling or for deployment in a production environment. In either case, the

team should focus on recommending actions to operationalize the work and the benefits the customer will receive because of implementing these recommendations.

## Recommendations

- **Implement the model as a pilot, before more wide-scale rollout – test and learn from initial pilot on performance and precision**
  - ▶ Addressing these promptly can potentially save more customers from churning over time and also prevent more networking that seems to drive additional churn
  - ▶ An early churn warning trigger can be set up based on this model
- **Run the predictive model daily or weekly to be proactive on customer churn**
  - ▶ In-database scorer can score large datasets in a matter of minutes and can be run daily
  - ▶ Each customer retained via early warning trigger saves 4 hours of account retention efforts & 50k in new account acquisition costs
- **Develop targeted customer surveys to investigate the causes of churn, which will make the collection of data for investigation into the causes of churn easier**

**FIGURE 12-15** Sample recommendations for a data science project

### 12.2.9 Additional Tips on the Final Presentation

As a team completes a project and strives to move on to the next one, it must remember to invest adequate time in developing the final presentations. Orienting the audience to the project and providing context is important. On occasion, a team is so immersed in the project that it fails to provide sufficient context for its recommendations and the outputs of the models. A team needs to remember to spell out terminology and acronyms and avoid excessive use of jargon. It should also keep in mind that presentations may be shared extensively; therefore, recipients may not be familiar with the context and the journey the team has gone through over the course of the project.

The story may need to be told multiple times to different audiences, so the team must remain patient in repeating some of the key messages. These presentations should be viewed as opportunities to refine the key messages and evangelize the good work that was done. By this point in the process, the team has invested many hours of work and uncovered insights for the business. These presentations are an opportunity to communicate these projects and build support for future projects. As with most presentations, it is important to gauge the audience to guide shaping the message and the level of detail. Here are several more tips on developing the presentations.

- **Use imagery and visual representations:** Visuals tend to make the presentation more compelling. Also, people recall imagery better than words, because images can have a more visceral impact. These visual representations can be static and interactive data.

- Make sure the text is mutually exclusive and collectively exhaustive (MECE): This means having an economy of words in the presentation and making sure the key points are covered but not repeated unnecessarily.
- Measure and quantify the benefits of the project: This can be challenging and requires time and effort to do well. This kind of measurement should attempt to quantify benefits that have financial and other benefits in a specific way. Making the statement that a project provided “\$8.5M in annual cost savings” is much more compelling than saying it has “great value.”
- Make the benefits of the project clear and conspicuous: After calculating the benefits of the project, make sure to articulate them clearly in the presentation.

### 12.2.10 Providing Technical Specifications and Code

In addition to authoring the final presentations, the team needs to deliver the actual code that was developed and the technical documentation needed to support it. The team should consider how the project will affect the end users and the technical people who will need to implement the code. It is recommended that the team think through the implications of its work on the recipients of the code, the kinds of questions they will have, and their interests. For instance, indicating that the model will need to perform real-time monitoring may require extensive changes to an IT runtime environment, so the team may need to consider a compromise of nightly batch jobs to process the data. In addition, the team may need to get the technical team talking with the project sponsor to ensure the implementation and SLA will meet the business needs during the technical deployment.

The team should anticipate questions from IT related to how computationally expensive it will be to run the model in the production environment. If possible, indicate how well the model ran in the test scenarios and whether there are opportunities to tune the model or environment to optimize performance in the production environment.

Teams should approach writing technical documentation for their code as if it were an application programming interface (API). Many times, the models become encapsulated as functions that read a set of inputs in the production environment, possibly perform preprocessing on data, and create an output, including a set of post-processing results.

Consider the inputs, outputs, and other system constraints to enable a technical person to implement the analytical model, even if this person has not had a connection to the data science project up to this point. Think about the documentation as a way to introduce the data that the model needs, the logic it is using, and how other related systems need to interact with it in a production environment for it to operate well. The specifications detail the inputs the code needs and the data format and structures. For instance, it may be useful to specify whether structured data is needed or whether the expected data needs to be numeric or string formats. Describe any transformations that need to be made on the input data before the code can use it, and if scripting was created to perform these tasks. These kinds of details are important when other engineers must modify the code or utilize a different dataset or table, if and when the environment changes.

Regarding exception handling, the team must consider how the code should handle data that is outside the expected data ranges of the model parameters and how it will handle missing data values (Chapter 3, “Review of Basic Data Analytic Methods Using R”), null values, zeros, NAs, or data that is in an unexpected format or type. The technical documentation describes how to treat these exceptions and what implications may emerge on downstream processes. For the model outputs, the team must explain to what extent to post-process the output. For example, if the model returns a value representing