



# Module 9 – The Endgame, or Putting it All Together

EMC<sup>2</sup> PROVEN PROFESSIONAL



Introduction



Analytics Lifecycle



Basic Methods



Adv. Methods



Tools



Lab

## Module 9 – The Endgame, or Putting it All Together

Upon completion of this module, you should be able to:

- Articulate three tasks needed to operationalize an analytics project
- Explain how the four common deliverables of an analytics lifecycle project meet the needs of key stakeholders
- Use a framework for creating final presentations for sponsors and analysts



Introduction



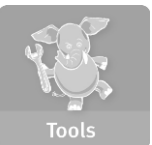
Analytics Lifecycle



Basic Methods



Adv. Methods



Tools



Lab

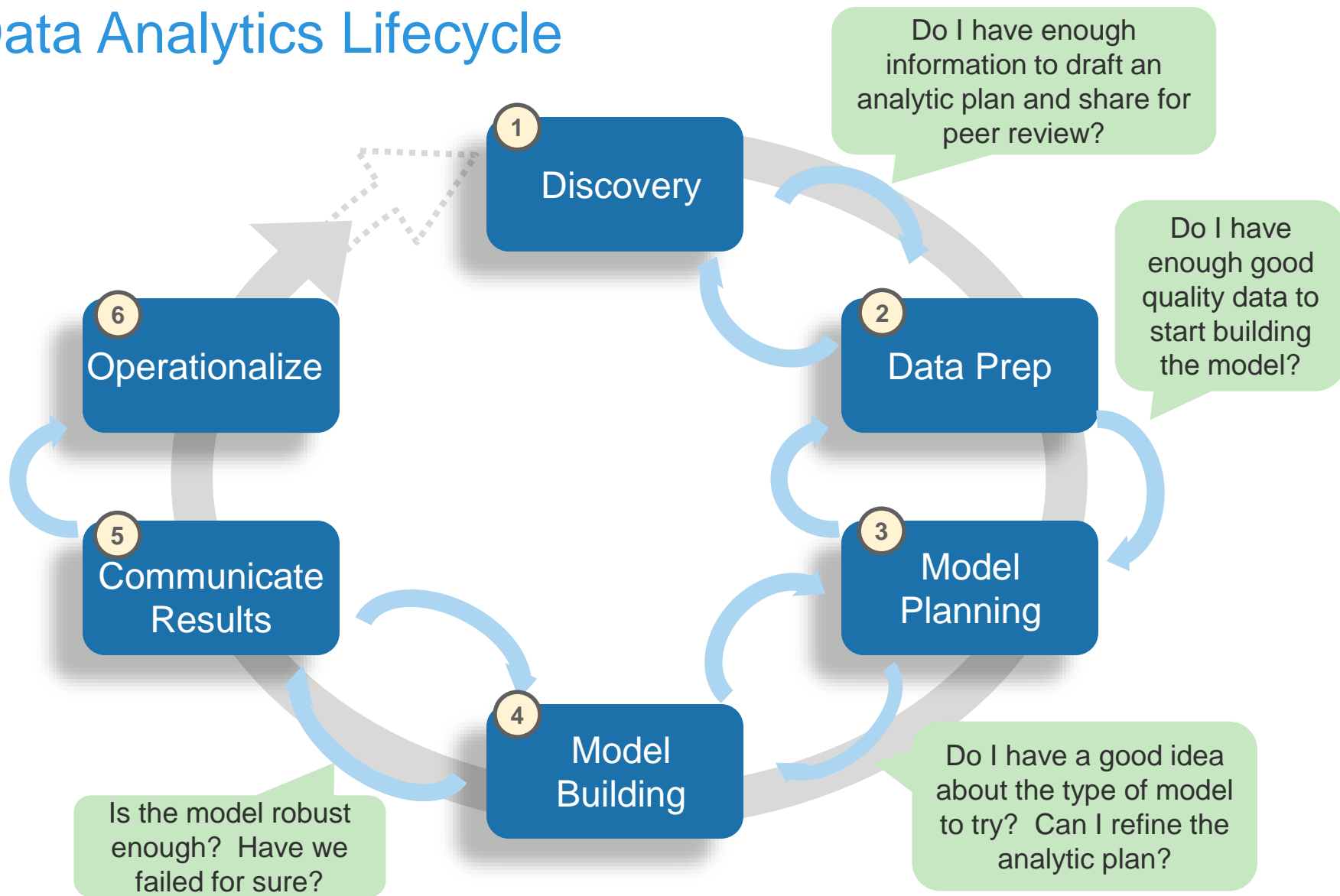
## Module 9 – The Endgame, or Putting it All Together

### Lesson 1: Operationalizing an Analytics Project

During this lesson the following topics are covered:

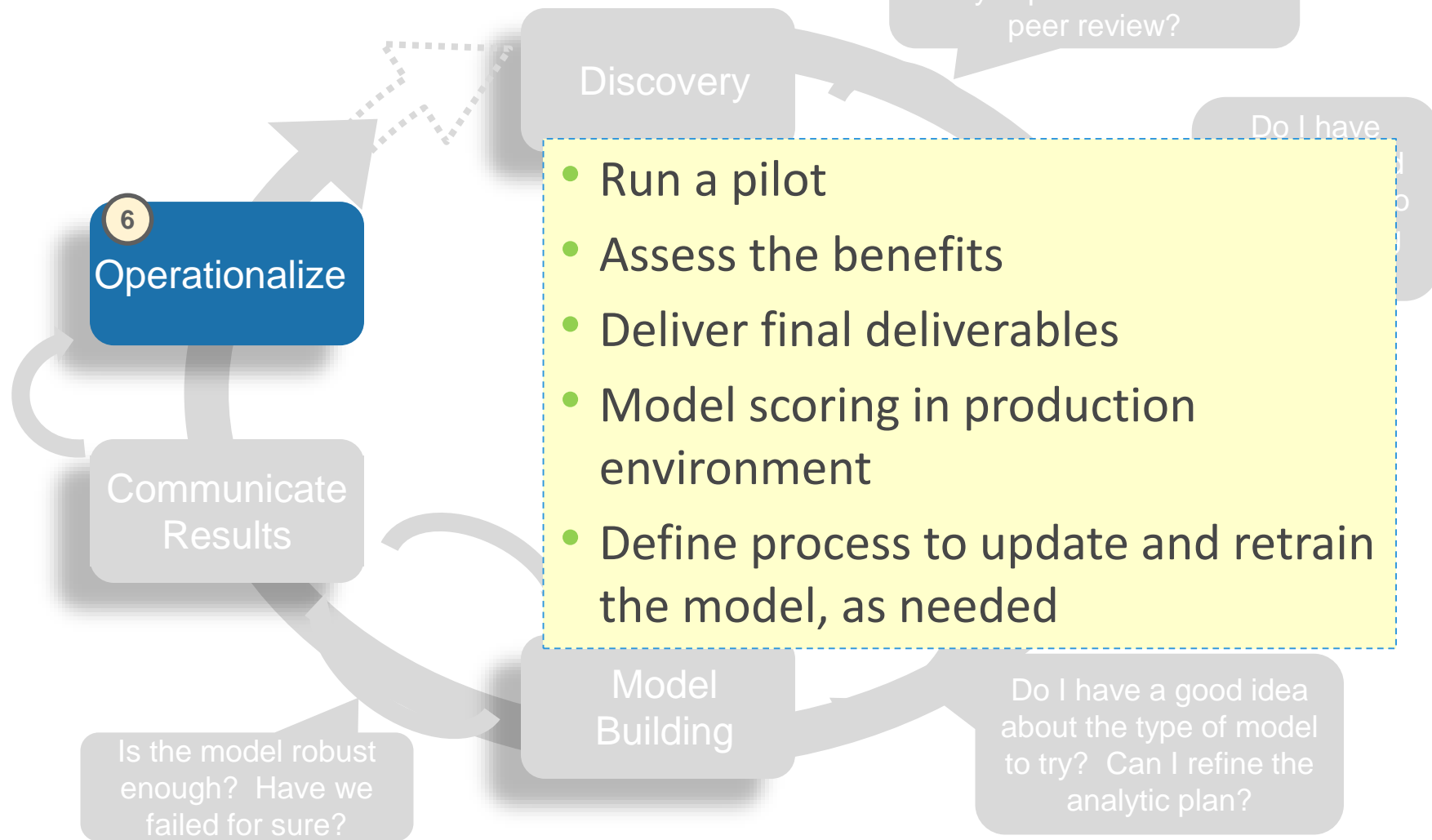
- Operationalizing a data analytics lifecycle project
- Key outputs needed for a successful analytic project, by stakeholder role
- 4 core deliverables to meet most stakeholder needs

# Data Analytics Lifecycle



# Data Analytics Lifecycle

## Final Deliverables

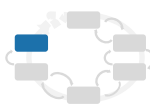


# Key Outputs from a Successful Analytic Project, by Role



Role	Description	What the Role Needs in the Final Deliverables
Business User	Someone who benefits from the end results and can consult and advise project team on value of end results and how these will be operationalized	<ul style="list-style-type: none"> <li>• <b>Sponsor Presentation</b> addressing: <ul style="list-style-type: none"> <li>• Are the results good for me?</li> <li>• What are the benefits of the findings?</li> <li>• What are the implications of this for me?</li> </ul> </li> </ul>
Project Sponsor	Person responsible for the genesis of the project, providing the impetus for the project and core business problem, generally provides the funding and will gauge the degree of value from the final outputs of the working team	<ul style="list-style-type: none"> <li>• <b>Sponsor Presentation</b> addressing: <ul style="list-style-type: none"> <li>• What's the business impact of doing this?</li> <li>• What are the risks? ROI?</li> <li>• How can this be evangelized within the organization (and beyond)?</li> </ul> </li> </ul>
Project Manager	Ensure key milestones and objectives are met on time and at expected quality.	
Business Intelligence Analyst	Business domain expertise with deep understanding of the data, KPIs, key metrics and business intelligence from a reporting perspective	<ul style="list-style-type: none"> <li>• Show the <b>analyst presentation</b></li> <li>• Determine if the reports will change</li> </ul>
Data Engineer	Deep technical skills to assist with tuning SQL queries for data management, extraction and support data ingest to analytic sandbox	<ul style="list-style-type: none"> <li>• Share the <b>code</b> from the analytical project</li> <li>• Create <b>technical document</b> on how to implement it.</li> </ul>
Database Administrator (DBA)	Database Administrator who provisions and configures database environment to support the analytical needs of the working team	<ul style="list-style-type: none"> <li>• Share the <b>code</b> from the analytical project</li> <li>• Create <b>technical document</b> on how to implement it.</li> </ul>
Data Scientist	Provide subject matter expertise for analytical techniques, data modeling, applying valid analytical techniques to given business problems and ensuring overall analytical objectives are met	<ul style="list-style-type: none"> <li>• Show the <b>analyst presentation</b></li> <li>• Share the <b>code</b></li> </ul>

# 4 Core Deliverables to Meet Most Stakeholder Needs



## 1. Presentation for Project Sponsors

- “Big picture” takeaways for executive level stakeholders.
- Determine key messages to aid their decision-making process.
- Focus on clean, easy visuals for the presenter to explain and for the viewer to grasp.

## 2. Presentation for Analysts

- Business process changes.
- Reporting changes.
- Fellow data scientists will want the details and are comfortable with technical graphs (such as ROC curves, density plots, histograms).

## 3. Code for technical people

## 4. Technical specs of implementing the code



Introduction



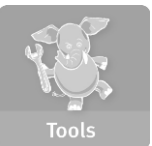
Analytics Lifecycle



Basic Methods



Adv. Methods



Tools



Lab

## Module 9: The Endgame, or Putting it All Together

### Lesson 1: Summary

During this lesson the following topics were covered:

- Operationalizing a data analytics lifecycle project
- Key outputs needed for a successful analytic project, by stakeholder role
- 4 core deliverables to meet most stakeholder needs





Introduction



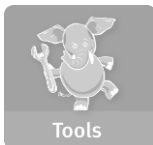
Analytics Lifecycle



Basic Methods



Adv. Methods



Tools



Lab

# Module 9: The Endgame, or Putting it All Together

## Lesson 2: Creating the Final Deliverables

During this lesson the following topics are covered:

- Brief review of YoyoDyne case study
- Using a core set of materials to deliver presentations for two different audiences
- Comparing the main focus areas for sponsors and analyst audiences
- Using a framework to organize the main pieces of your final presentations
- Tips for sharing your code and technical documentation

# YoyoDyne Churn Prediction Case Study



## *Mini Case Study: Churn Prediction for Yoyodyne Bank*

### Situation Synopsis

- Retail Bank, Yoyodyne Bank wants to improve the Net Present Value (NPV) and retention rate of the customers .
- They want to establish an effective marketing campaign targeting customers to reduce the churn rate by at least five percent.
- The bank wants to determine whether those customers are worth retaining. In addition, the bank also wants to analyze reasons for customer attrition and what they can do to keep them.
- The bank wants to build a data warehouse to support marketing and other related customer care groups.

# Use Analytic Plan to Guide Final Presentation

**Mini Case Study:**  
**Churn Prediction for**  
**Retail Banking**

Components of Analytic Plan	Retail Banking: Yoyodyne Bank
Discovery Business Problem Framed	How do we identify churn/no churn for a customer?
Initial Hypotheses	Transaction volume and type are key predictors of churn rates
Data & Scope	5 months of customer account history
Model Planning - Analytic Technique	Logistic regression to identify most influential factors predicting churn
Result & Key Findings	Once customers stop using their accounts for gas and groceries, they will soon erode their accounts and churn. If customers use their debit card fewer than 5 times per month, they will leave the bank within 60 days.
Business Impact	If we can target customers who are high-risk for churn, we can reduce customer attrition by 25%. This would save \$3 million in lost customer revenue and avoid \$1.5 million in new customer acquisition costs each year.

# Key Aspects of Final Presentation Material



- **Reflect on the project:**

- ▶ Consider the context of the problems you set out to solve.
- ▶ Identify observations about the model outputs, scoring, results.
- ▶ Identify Key Messages, and any unexpected insights.

- **Tailor outputs to the audience**

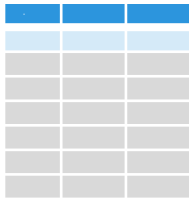
	Project Sponsor Presentation	Analyst Presentation
Focus	What	How
Objectives	<ul style="list-style-type: none"><li>• Show <u>that</u> you met the project goals</li><li>• Give your sponsor talking points to evangelize the work<ul style="list-style-type: none"><li>• Emphasize ROI and business value</li><li>• Mention if the models can be deployed within sponsor's SLA</li></ul></li></ul>	<ul style="list-style-type: none"><li>• Show <u>how</u> you met the project goals</li><li>• Share your methods so analysts can learn from it for future projects<ul style="list-style-type: none"><li>• Discuss methods, techniques, and technologies used.</li><li>• Provide specific model accuracy and speed (example: how well will it meet SLAs).</li></ul></li></ul>

# Develop Core Material you can use to Deliver Presentations to 2 Main Audiences

 = Same components for both presentations
  = Different components for Sponsor vs. Analyst presentation

Presentation Component	Project Sponsor Presentation	Analyst Presentation
<b>Project Goals</b>	<ul style="list-style-type: none"> <li>List top 3 agreed upon goals</li> </ul>	<ul style="list-style-type: none"> <li>List top 3 agreed upon goals</li> </ul>
<b>Main Findings</b>	<ul style="list-style-type: none"> <li>Emphasize key message</li> </ul>	<ul style="list-style-type: none"> <li>Emphasize key message</li> </ul>
<b>Approach</b>	<ul style="list-style-type: none"> <li>High Level Methodology</li> </ul>	<ul style="list-style-type: none"> <li>High Level Methodology</li> <li>Relevant details on modeling techniques and technology</li> </ul>
<b>Model Description</b>	<ul style="list-style-type: none"> <li>Overview of the modeling technique</li> </ul>	<ul style="list-style-type: none"> <li>Overview of the modeling technique</li> </ul>
<b>3 key points supported with data</b>	<ul style="list-style-type: none"> <li>Support key points with simple charts and graphics (example: bar charts)</li> </ul>	<ul style="list-style-type: none"> <li>Show details to support the key points</li> <li>Analyst-oriented charts and graphs (ROC curves, histograms)</li> <li>Visuals of key variables and significance of each</li> </ul>
<b>Model Details</b>	<ul style="list-style-type: none"> <li>Omit this section, or discuss only at a very high level</li> </ul>	<ul style="list-style-type: none"> <li>Show the code or main logic of the model, Include the model type, variables, technology used to execute it and score data.</li> <li>Identify key variables and impact of each</li> <li>Describe expected model performance and any caveats</li> <li>Detailed description of the modeling technique</li> <li>Discuss variables, scope, predictive power</li> </ul>
<b>Recommendations</b>	<ul style="list-style-type: none"> <li>Focus on business impact of doing this, including risks and ROI</li> <li>Give the sponsor salient points to help him or her evangelize the work within the organization</li> </ul>	<ul style="list-style-type: none"> <li>Supplement recommendations with any implications for the modeling, or for deploying in a production environment.</li> </ul>

# Project Goals



Presentation Component	Project Sponsor Presentation	Analyst Presentation
Project Goals	<ul style="list-style-type: none"> <li>List top 3 agreed upon goals</li> </ul>	<ul style="list-style-type: none"> <li>List top 3 agreed upon goals</li> </ul>

## Example 1 of Goals slide

Project Goals

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

## Example 2 of Goals slide, with Situation overview

Situation & Project Goals

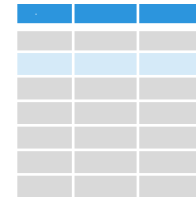
Situation

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

Goals of YoyoDyne “Churn Project”

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

# Main Findings (Executive Summary)



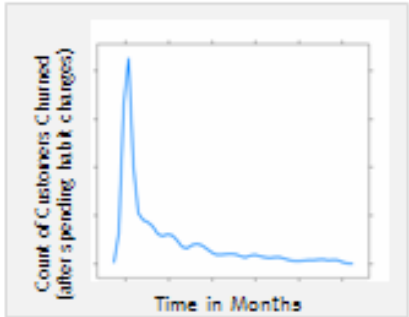
Presentation Component	Project Sponsor Presentation	Analyst Presentation
Main Findings (executive summary)	<ul style="list-style-type: none"> <li>Emphasize key message</li> </ul>	<ul style="list-style-type: none"> <li>Emphasize key message</li> </ul>

- Enable reader to grasp full synopsis in 1 slide
- Frame outcomes in terms of business value
- Generally same, or very similar, for both types of audiences

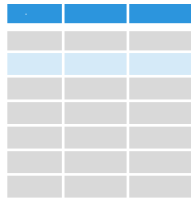
## Executive Summary

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

- **Customers churn within 60 days of changing their spending habits**
  - ▶ Most often after customers stop using bank cards for gas and grocery
  - ▶ 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 churner'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**



# Anatomy of an Executive Summary



## Business Impact

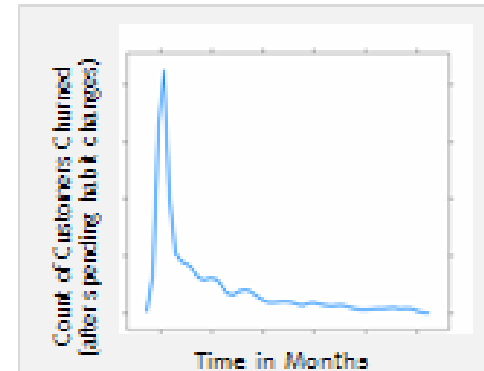
## Key Message

### Executive Summary

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

## Major Points

- **Customers churn within 60 days of changing their spending habits**
  - ▶ Most often after customers stop using bank cards for gas and grocery
  - ▶ 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 churner's contacts
  - ▶ With CRM data we can identify 20% of churners, adding social media data increases this to 30%

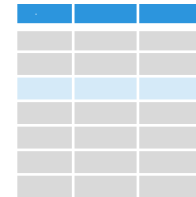


## SLA

**Models can run in minutes, rather than current process of monthly cycles**



# Approach



Presentation Component	Project Sponsor Presentation	Analyst Presentation
<b>Approach</b>	<ul style="list-style-type: none"> <li>High Level Methodology</li> </ul>	<ul style="list-style-type: none"> <li>High Level Methodology</li> <li>Relevant details on modeling techniques and technology</li> </ul>

## Example Approach slide, for Sponsors

## Example Approach slide, for Analysts

### Approach (for Sponsors)

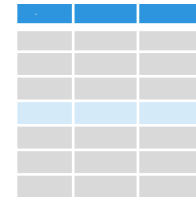
### Approach (for Analysts)

- 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 data sets, assess data quality and availability
- Developed churn model to identify customers most likely to leave the bank
  - Identify most influential factors
  - Provides 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

- 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 data sets, assess data quality and availability
- Developed churn model in R using a Generalized Additive Modeling technique
  - Minimizes variable transformations and binning
  - Provides greater explanatory power for analyzing impact of different factors on churn
- Impact of social network variables was examined and found to help identify more potential churners
- Worked 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

Note: Green boxes highlight differences between slides

# Model Description

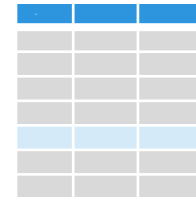


Presentation Component	Project Sponsor Presentation	Analyst Presentation
Model Description	<ul style="list-style-type: none"><li>Overview of the modeling technique</li></ul>	<ul style="list-style-type: none"><li>Overview of the modeling technique</li></ul>

## Model Description

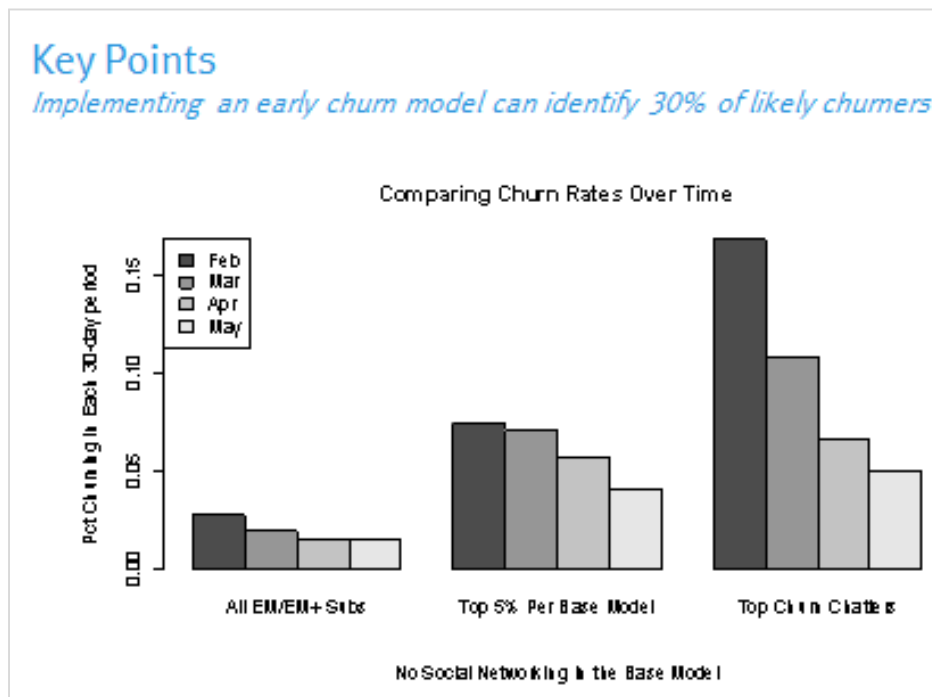
- **Overview of Basic Methodology:** predict the likelihood of churn for each customer. Identify customers with a greater probability for churn then compare with actual churn outcomes to train the algorithm and enable predictions for existing customers.
- **Model:** Logistic regression model
- **Dependent 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

# Key Points Supported With Data

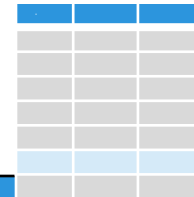


Presentation Component	Project Sponsor Presentation	Analyst Presentation
<b>Key Points Supported With Data</b>	<ul style="list-style-type: none"> <li>Support key points with simple charts and graphics (such as bar charts)</li> </ul>	<ul style="list-style-type: none"> <li>Show details to support the key points</li> <li>Analyst-oriented charts and graphs (ROC curves, histograms)</li> <li>Visuals of key variables and significance of each</li> </ul>

- Identify key points based on your insights and observations resulting from the data and model scoring results
- Illustrate your key points with charts and visualizations
- Use simpler charts for Sponsors



# Model Details



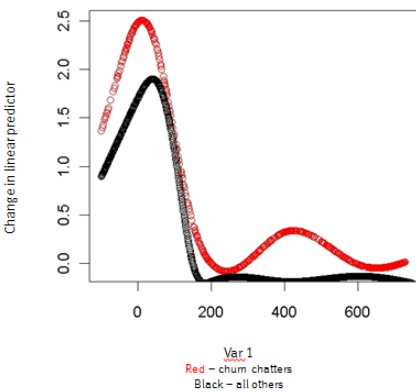
Presentation Component	Project Sponsor Presentation	Analyst Presentation
Model Details	<ul style="list-style-type: none"> <li>Omit this section, or discuss only at a very high level</li> </ul>	<ul style="list-style-type: none"> <li>Show the code or main logic of the model, Include the model type, variables, technology used to execute it and score data.</li> <li>Describe expected model performance and any caveats</li> <li>Detailed description of the modeling technique, variables, and model performance</li> </ul>

## Model Details

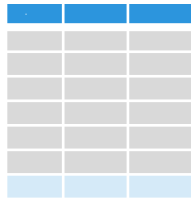
- Candidate variables: 22 from CRM, 154 from call history, and 12 social networking variables
- Through PCA and discussion with domain experts, we reduced ~190 variables to the 9 most predictive of customer churn
- General Additive Model (GAM) model built in R :

```
gam.wsn.by2 <- 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(tvar5, 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)
```

Var 1 has a larger and earlier impact on churn-chatters



# Recommendations



Presentation Component	Project Sponsor Presentation	Analyst Presentation
<b>Recommendations</b>	<ul style="list-style-type: none"> <li>Focus on business impact of the project, including risks and ROI</li> <li>Give the sponsor salient points to help him or her evangelize the work within the organization</li> </ul>	<ul style="list-style-type: none"> <li>Supplement recommendations with any implications for the modeling, or for deploying in a production environment.</li> </ul>

## 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.

# Quick Summary of Final Presentation Components

## Situation & Project Goals

### Situation

1. Yoyodyne Bank wants to improve the Net Present Value (NPV) and retention rate of the customers.
2. In last 90 days, Yoyodyne has lost 6 of its top 100 customers, and is seeing increased competition from their 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

## Executive Summary

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

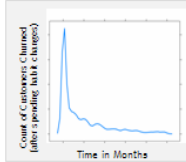
### Customers churn within 60 days of changing their spending habits

- ▶ Most often after customers stop using bank cards for gas and grocery
- ▶ 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 churner'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

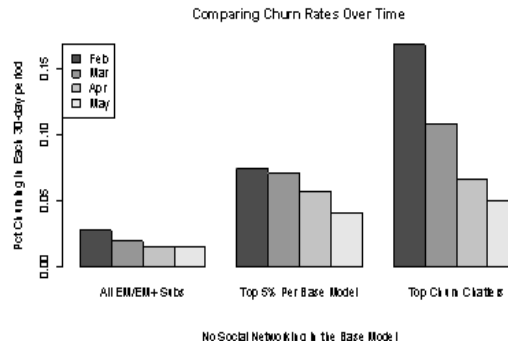


## Model Description

- **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

## Key Points

*Implementing an early churn model can identify 30% of likely churners*



## 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 data sets, assess data quality and availability
- Developed churn model to identify customers most likely to leave the bank
  - ▶ Identify most influential factors
  - ▶ Provides 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

## 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
- Generated

### 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.

# Top 10 Tips, Tricks, & Pitfalls to Avoid for the Final Presentation

1. Be visual. Generally, the more visual the better. Up to a point.
2. Be MECE (Mutually Exclusive and Collectively Exhaustive).
3. Tie your ideas together....don't force people to tie your ideas together, guide people and help them draw logical connections.
4. Don't forget that not everyone has gone through the Discovery phase like you have.
5. Context is key. Orient people to the project itself, as well as the graphics you use, the terminology and jargon (spell out acronyms).
6. Don't assume people see the obvious benefits.
7. Measure and quantify the benefits. Be specific. “\$8.5M in annual cost savings” is much stronger than “Great Value”.
8. Be patient. You may have to tell your story more than once...consider these sessions opportunities to refine your message and share good work that was done.
9. Let the intended audience guide you in shaping the right message and level of detail.
10. Avoid long bulleted lists 😊

# Overview of Code & Technical Documentation

- **Consider the interests of your technical audience:**
  - ▶ How will the project affect them?
  - ▶ In what ways will it change their day-to-day roles, or existing processes?
  - ▶ Be aware of the implications of your work on their roles as you create these technical deliverables.
- **2 Technical deliverables:**
  - ▶ Code
  - ▶ Technical specifications and documentation.



# Considerations for Technical Specifications & Documentation

*Approach the documentation as if it's for an API (application programming interface)*

- **Inputs & Pre-processing:**

- ▶ Discuss the expected pre-processing steps before data goes to the model code.
- ▶ Document expected input, data format, source tables, and units.
- ▶ Describe the processing script are you using.
- ▶ Explain how the outputs are created.

- **Exception handling :**

- ▶ Explain how to deal with exceptions to the model.
- ▶ Provide guidance for making decisions on the exceptions.

- **Post-processing:**

- ▶ After you create the output, discuss any post-processing before going to the next step.
- ▶ Interpreting a threshold as opposed to a simple yes/no.

# Providing Your Code

- Test for accuracy in the production environment
- Ensure the code will run quickly and meet SLAs
- Include comment lines in the code
- Hold a briefing with the engineers who will implement the code



Introduction



Analytics Lifecycle



Basic Methods



Adv. Methods



Tools



Lab

# Module 9: The Endgame, or Putting it All Together

## Lesson 2: Summary

During this lesson the following topics were covered:

- Brief review of YoyoDyne case study
- Using a core set of materials to deliver presentations for two different audiences
- Comparing the main focus areas for sponsors and analyst audiences
- Using a framework to organize the main pieces of your final presentations
- Tips for sharing your code and technical documentation