Course3: Data science methodology

Module 1: From Problem to Approach and From Requirements to Collection

- Video: Course Introduction
- Reading: Helpful Tips for Course Completion
- Reading: Syllabus

Lesson 1: From Problem to Approach

- Video: Data Science Methodology Overview
- Video: Business Understanding
- Video: Analytic Approach
- Hands-on Lab: From Problem to Approach
- Reading: Lesson 1 Summary: From Problem to Approach
- Practice Quiz: From Problem to Approach
- Glossary: From Problem to Approach
- Graded Quiz: From Problem to Approach

Lesson 2: From Requirements to Collection

- Video: Data Requirements
- Video: Data Collection
- Hands-on Lab: From Requirements to Collection
- Reading: Lesson 2 Summary: From Requirements to Collection
- Practice Quiz: From Requirements to Collection
- Glossary: From Requirements to Collection
- Graded Quiz: From Requirements to Collection

Module 2: From Understanding to Preparation and from Modeling to Evaluation

Lesson 1: From Understanding to Preparation

- Video: Data Understanding
- Data Preparation - Concepts
- Data Preparation - Case Study
- Hands-on Lab: From Understanding to Preparation
- Reading: Lesson 1 Summary: From Understanding to Preparation
- Practice Quiz: From Understanding to Preparation
- Glossary: From Understanding to Preparation
- Graded Quiz: From Understanding to Preparation

Lesson 2: From Modeling to Evaluation

- Video: Modeling - Concepts
- Video: Modeling - Case Study
- Video: Evaluation
- Hands-on Lab: From Modeling to Evaluation
- Reading: Lesson 2 Summary:From Modeling to Evaluation
- Practice Quiz: From Modeling to Evaluation
- Glossary: From Modeling to Evaluation
- Graded Quiz: From Modeling to Evaluation

Module 3: From Deployment to Feedback

- Video: Deployment

- Video: Feedback
- Video: Storytelling
- Video: Course Summary
- Reading: Module 3 Summary:From Deployment to Feedback
- Practice Quiz: From Deployment to Feedback
- Glossary: From Deployment to Feedback
- Graded Quiz: From Deployment to Feedback

Module 4: Final Project and Assessment

Final Project

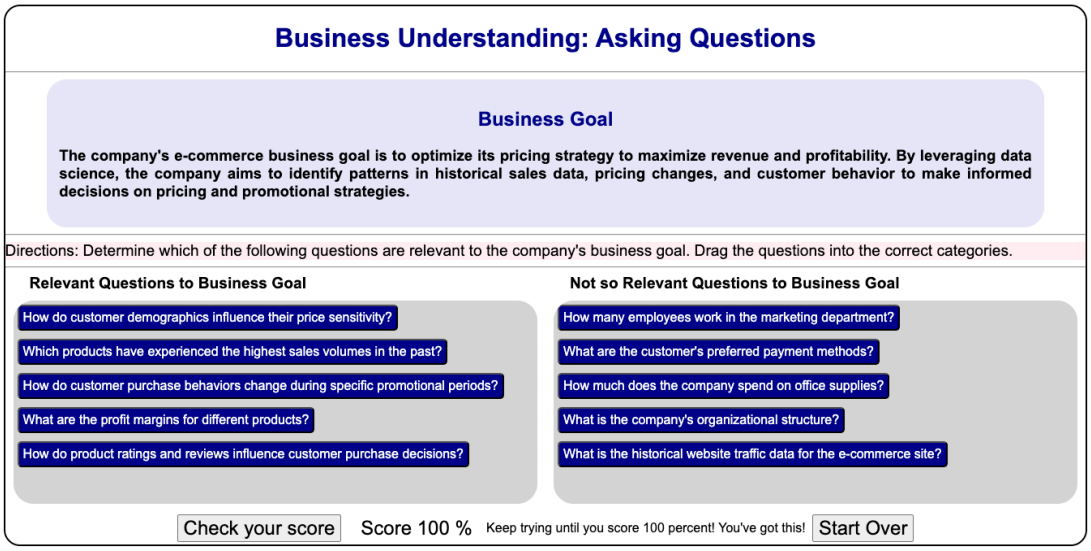
- Video: Introduction to CRISP-DM
- Reading: Final Assignment Overview
- Peer Review: Final Assignment

Course Summary and Final Quiz

- Reading: Review What You Learned
- Graded Quiz: Final Quiz

Course Wrap Up

- Reading: Congratulations and Next Steps
- Reading: Thanks from the Course Team
- Reading: IBM Digital Badge



Analytical Approach

Identifying the pattern to address the question

Business Goal

A transportation company aims to optimize its delivery routes and schedules to minimize costs and improve delivery efficiency. The company wants to use data science to identify the most optimal routes and delivery time windows based on historical delivery data and external factors such as traffic and weather conditions.

Various questions are targeted by data scientist to achieve this business goal

Directions: Identify the 'Question Pattern' relevant to each analytical approach. Drag each question into the relevant 'Analytical Approach' box.

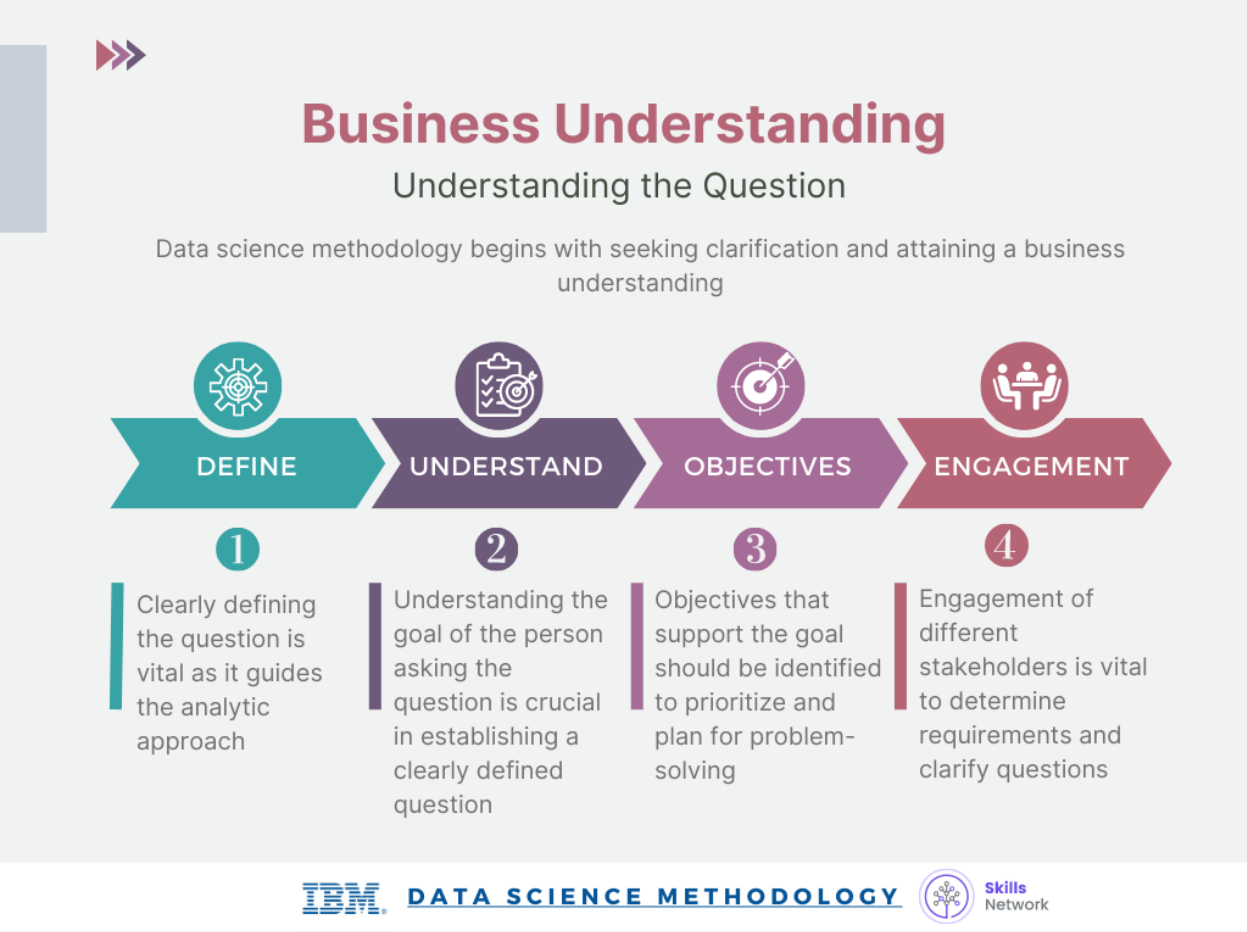
Predictive Model	Descriptive Model	Classification Model
How can we forecast the optimal number of delivery vehicles required for a specific day based on the expected order volume?	What historical data highlights the busiest delivery days and time intervals during the week based on past order data?	How can we cluster customer locations to create distinct groups for efficient delivery route planning, without explicitly making predictions based on past data?
What are the expected delivery time for each route considering historical traffic patterns and anticipated weather conditions?	What are the average delivery costs for different delivery routes, and how do they vary during different times of the day?	How can we group delivery regions based on customer density and order frequency to optimize delivery route planning?
How can we determine the most suitable delivery routes for perishable goods, ensuring timely deliveries without explicitly using past data to make predictions?	What are the most frequently used routes and their respective delivery time variations during peak and off-peak hours?	What are the various time slots in which delivery schedules can be classified to balance workload and minimize delivery delays?
How can we anticipate the potential impact of traffic incidents or road closures on delivery times to proactively adjust routes?	What insights can be gathered on the average delivery times for different vehicle types, how do these times vary based on the complexity of the delivery route?	How can we classify delivery routes into different categories based on the average delivery time and order volume?

Check your score Score 100 % Keep trying until you score 100 percent! You've got this! Start Over

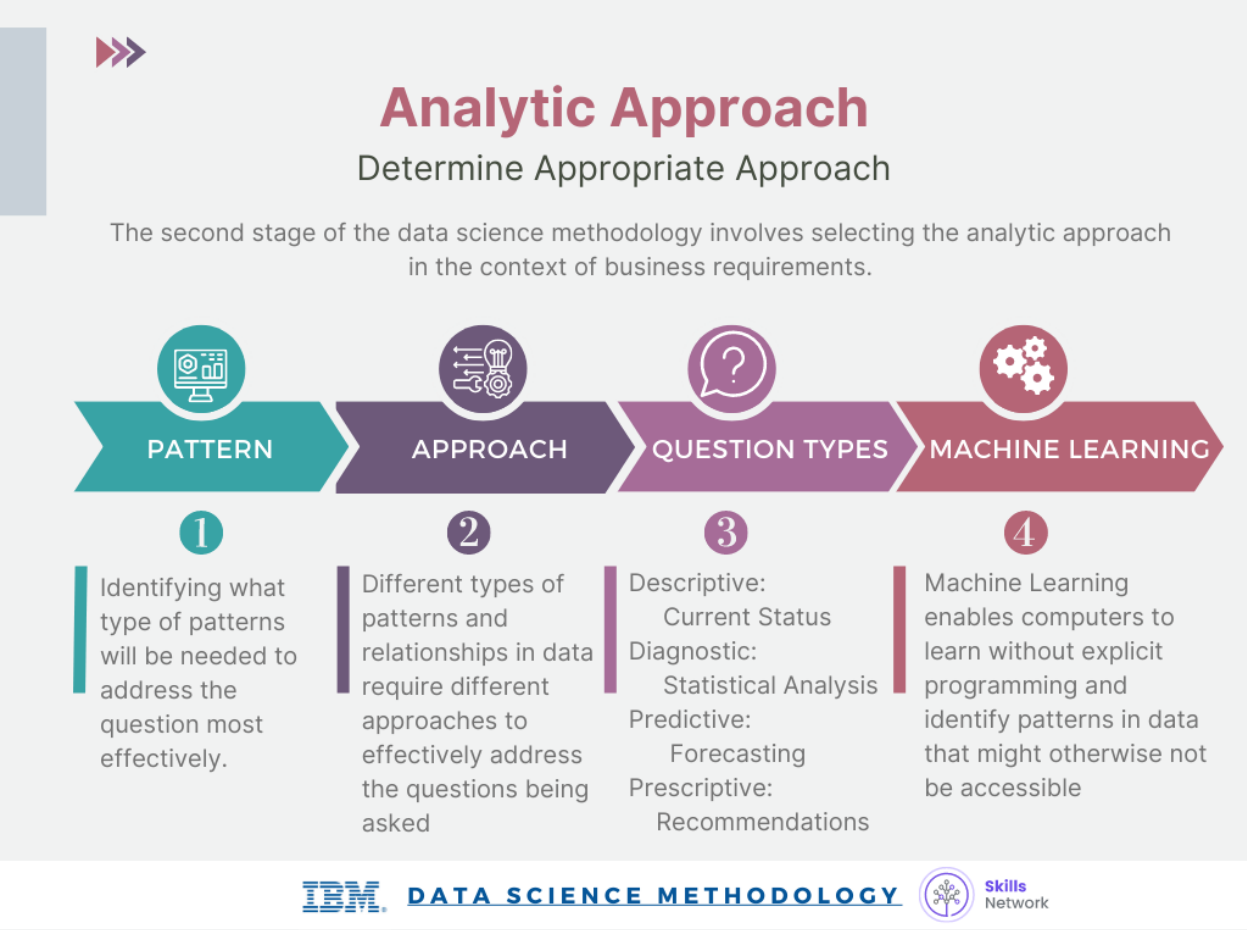
Back

Lesson summary
Module 1 Lesson 1: From Problem to Approach

- Foundational methodology, a cyclical, iterative data science methodology developed by John Rollins, consists of 10 stages, starting with Business Understanding and ending with Feedback.
- The primary goal of the Business Understanding stage is to understand the business problem and determine the data needed to answer the core business question.



- During the Analytic Approach stage, you can choose from descriptive, diagnostic, predictive, and prescriptive analytic approaches, whether to use machine learning with clustering associations.

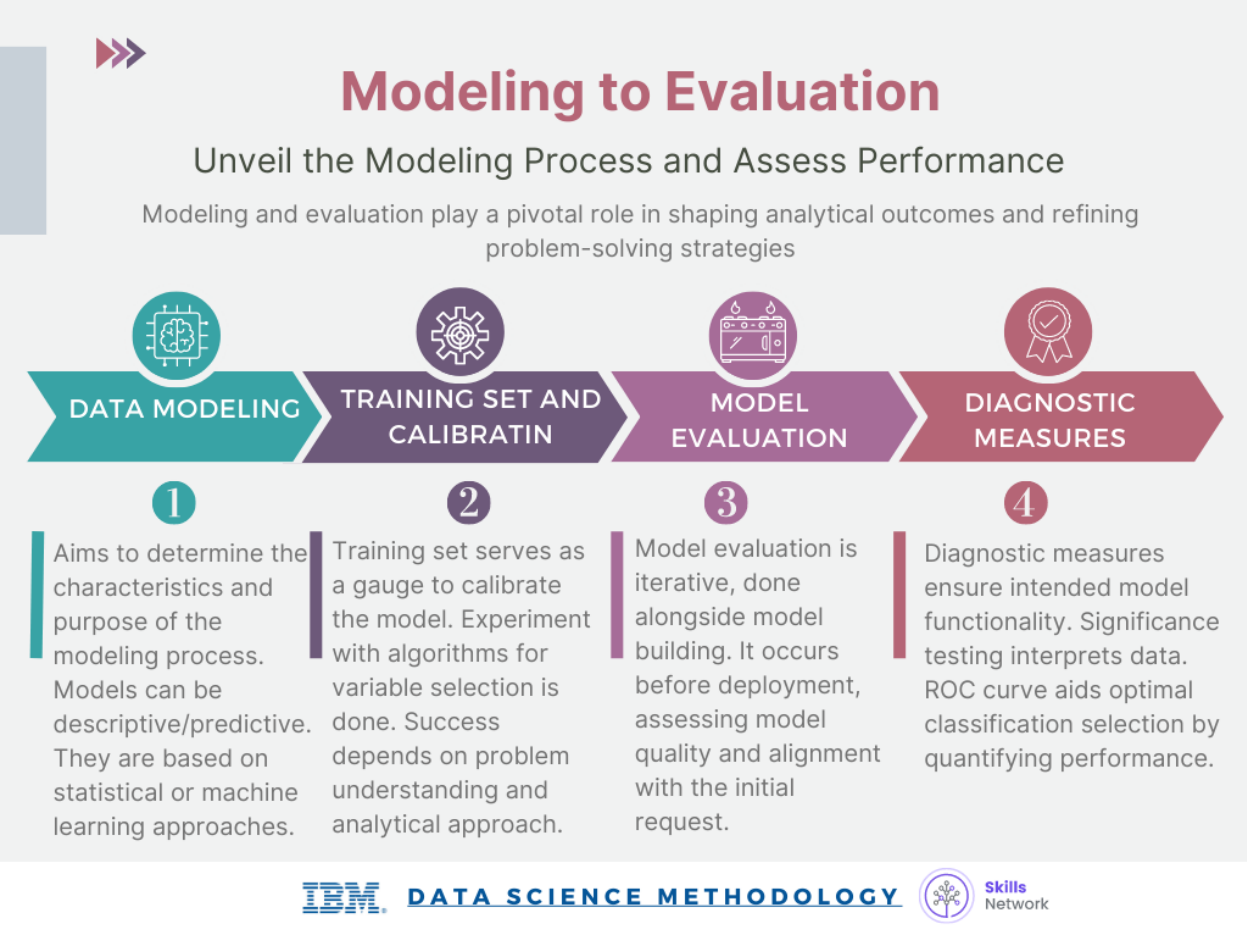


- Decision tree classification is a predictive analytics approach that's easy for non-data scientists to implement

Week 2:
Lesson summary
Module 2 Lesson 2: Modeling to Evaluation

Congratulations! You have completed this lesson. At this point in the course, you know:

- The end goal of the Modeling stage is that the data model answers the business question.
- The data modeling process uses a training data set. Data scientists test multiple algorithms on the training set data to determine whether the variables are required and whether the data supports answering the business question. The outcome of those models are either descriptive or predictive.



- The Evaluation phase consists of two stages, the diagnostic measures phase, and the statistical significance phase.
- During the Evaluation stage, data scientists and others assess the quality of the model and determine if the modelanswers the initial Business Understanding question or if the data model needs adjustment.
- The ROC curve, known as the receiver operating characteristic curve, is a useful diagnostic tool for determining the optimal classification model. This curve quantifies how well a binary classification model performs, declassifying the yes and no outcomes when some discrimination criterion is varied.

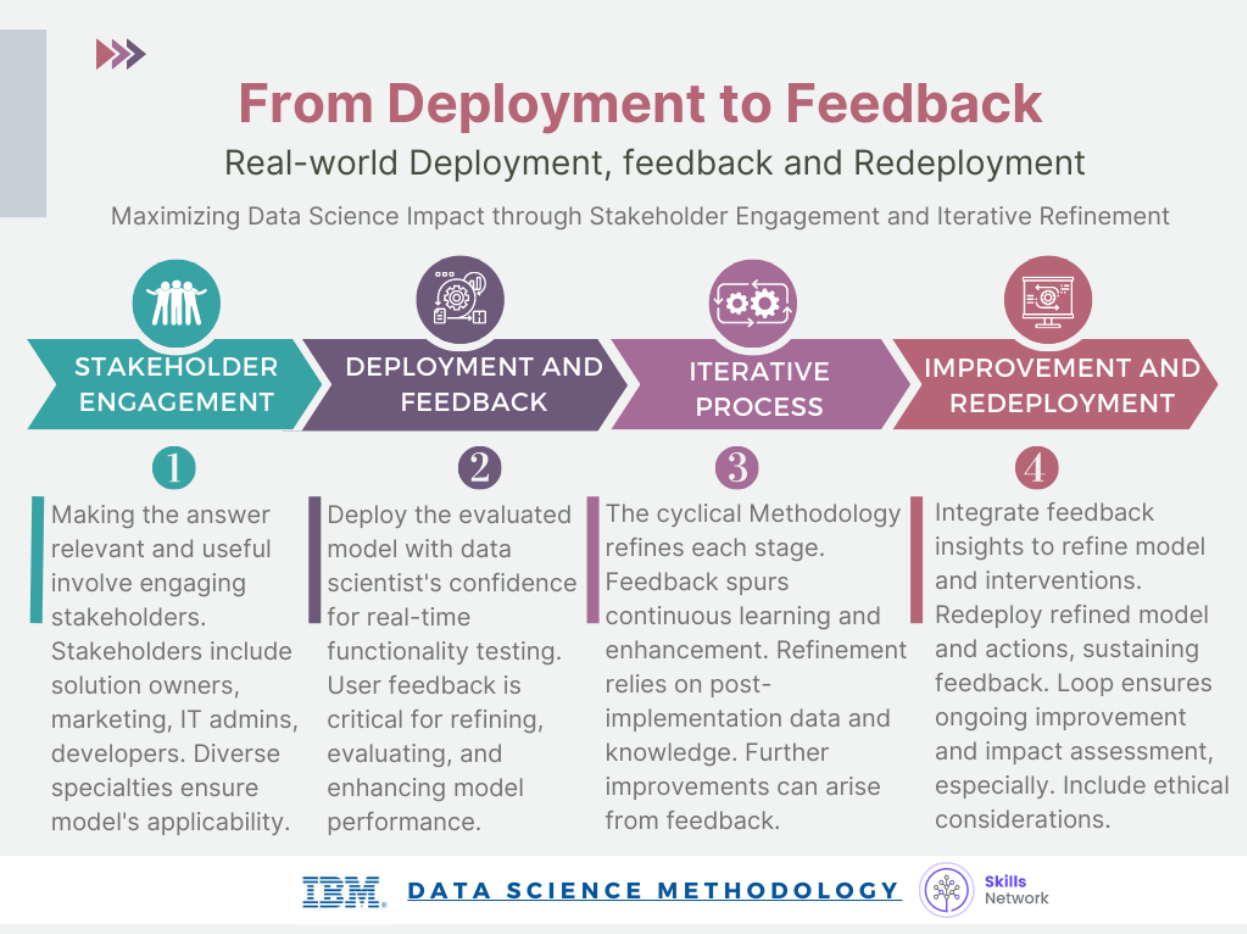
Lesson summary

Module 3 Lesson 1: Deployment to Feedback

Congratulations! You have completed this lesson. At this point in the course, you know:

- Stakeholders, including the solution owner, marketing staff, application developers, and IT administration evaluate the model and contribute feedback.
- During the Deployment stage, data scientists release the data model to a targeted group of stakeholders.
- Stakeholder and user feedback help assess the model's performance and impact during the Feedback stage.

- The model's value depends on iteration; that is, how successfully the data model incorporates user feedback.



Summary this course:

Review what you learned

After completing this course, you learned many facts about data science methodology. Here are 14 key, high-level takeaway facts you'll want to remember from this course.

- Foundational methodology, a cyclical, iterative data science methodology developed by John Rollins, consists of 10 stages, starting with Business Understanding and ending with Feedback.
- CRISP-DM, an open source data methodology, combines several data-related methodology stages into one stage and omits the Feedback stage resulting in a six-stage data methodology.
- The primary goal of the Business Understanding stage is to understand the business problem and determine the data needed to answer the core business question.
- During the Analytic Approach stage, you can choose from descriptive diagnostic, predictive, and prescriptive analytic approaches and whether to use machine learning techniques.
- During the Data Requirements stage, scientists identify the correct and necessary data content, formats, and sources needed for the specific analytical approach.
- During the Data Collection stage, expert data scientists revise data requirements and make critical decisions regarding the quantity and quality of data. Data scientists apply descriptive statistics and visualization techniques to thoroughly assess the content, quality, and initial insights gained from the collected data, identify gaps, and determine if new data is needed, or if they should substitute existing data.
- The Data Understanding stage encompasses all activities related to constructing the data set. This stage answers the question of whether the collected data represents the data needed to solve the business problem. Data scientists might use descriptive statistics, predictive statistics, or both.
- Data scientists commonly apply Hurst, univariates, and statistics such as mean, median, minimum, maximum, standard deviation, pairwise correlation, and histograms.
- During the Data Preparation stage, data scientists must address missing or invalid values, remove duplicates, and validate that the data is properly formatted. Feature engineering and text analysis are key techniques data scientists apply to validate and analyze data during the Data Preparation stage.
- The end goal of the Modeling stage is that the data model answers the business question. During the Modeling stage, data scientists use a training data set. Data scientists test multiple algorithms on the training set data to determine whether the variables are required and whether the data supports answering the business question. The outcome of those models is either descriptive or predictive.
- The Evaluation stage consists of two phases, the diagnostic measures phase, and the statistical significance phase. Data scientists and others assess the quality of the model and determine if the model answers the initial Business Understanding question or if the data model needs adjustment.

- During the Deployment stage, data scientists release the data model to a targeted group of stakeholders, including solution owners, marketing staff, application developers, and IT administration.,
- During the Feedback stage, stakeholders and users evaluate the model and contribute feedback to assess the model's performance.
- The data model's value depends on its ability to iterate; that is, how successfully the data model incorporates user feedback.

Mark as completed
Like
Dislike
Report an issue