



model	engine	class	classprob	conf_int	numeric	pred_int	quantile	raw
boost_tree()	C5.0	✓	✓	✗	✗	✗	✗	✓
	spark	✓	✓	✗	✓	✗	✗	✗
	xgboost	✓	✓	✗	✓	✗	✗	✓
	C5.0	✓	✓	✗	✗	✗	✗	✓

# Machine Learning in

Hands-On Coding with  
New **parsnip** package



Matt Dancho & David Curry  
*Business Science Learning Lab*



# Learning Lab Structure

- **Presentation**  
(15 & 15 min)
- **Coding**  
(30 min)
- **Q&A**  
(10 mins)

## Your Hosts!



**Matt Dancho**

Founder of Business Science, Matt designs and executes educational courses and workshops that deliver immediate value to organizations. His passion is **up-leveling future data scientists** coming from **untraditional backgrounds**.



**David Curry**

Founder of Sure Optimize, David works with businesses to help improve website performance and SEO using data science. His passion is **ethical Machine Learning initiatives**.



- **Why Machine Learning?**
  - Key Benefits
  - High Impact Problems
- **Example - Applied Product Price Model**
  - Develop predictive pricing model
  - Real Product Data
  - **Parsnip** R Package - 3 Algorithms
- **Learning Recommendations**
  - Key algorithms
  - Apply advance ML to Business

# XGBoost

```
187 # XGBoost ----
188 set.seed(1)
189 model_04_xgboost <- boost_tree("regression",
190                               mtry = 18,
191                               trees = 1000,
192                               min_n = 2,
193                               tree_depth = 6,
194                               learn_rate = 0.29,
195                               loss_reduction = 60
196                               ) %>%
197   set_engine("xgboost") %>%
198   fit(Price_num ~ ., data = train_tbl %>% select(-row_id))
199
200 model_04_xgboost %>%
201   predict(new_data = test_tbl) %>%
202   bind_cols(test_tbl %>% select(Price_num)) %>%
203   yardstick::metrics(truth = Price_num, estimate = .pred)
204
```

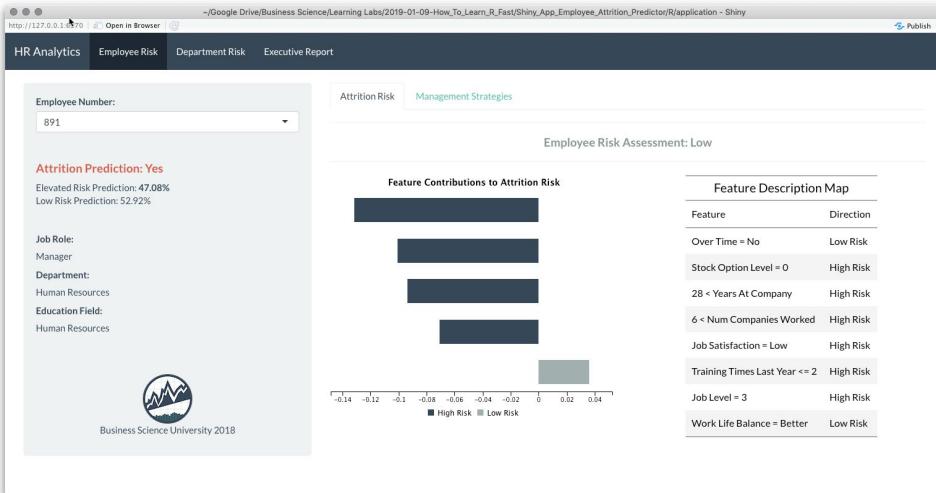


# Why Learn Machine Learning?

Most **in-demand** skill of the century

**Prediction** - Developing a model that learns a complex system, generalizes to new information

**Explanation** - Using the model to understand which features are important



# Machine Learning Solves High Impact Problems



## Sales

- Customer Churn
- Demand Forecasting
- Product Backorders
- Pricing Optimization

## Finance

- Fraud Detection
- Credit Default Risk
- Loan Delinquency

## HR

- Employee Attrition
- Promotion Readiness

## Marketing

- Customer Journey
- Customer Segmentation
- Spend & Conversion Optimization
- Web Traffic Anomalies

## Accounting

- Cash Flow Forecasting
- Payment Anomaly Detection

# Machine Learning Solves High Impact Problems



## Sales

- Customer Churn
- Demand Forecasting
- Product Backorders
- Pricing Optimization

## Finance

Fraud Detection  
Credit Default Risk  
• Loan Delinquency

## HR

- Employee Attrition
- Promotion Readiness

## Marketing

- Will a customer leave - Yes/No?
- Customer Segmentation
- Spend & Conversion Optimization
- Web Traffic Anomalies

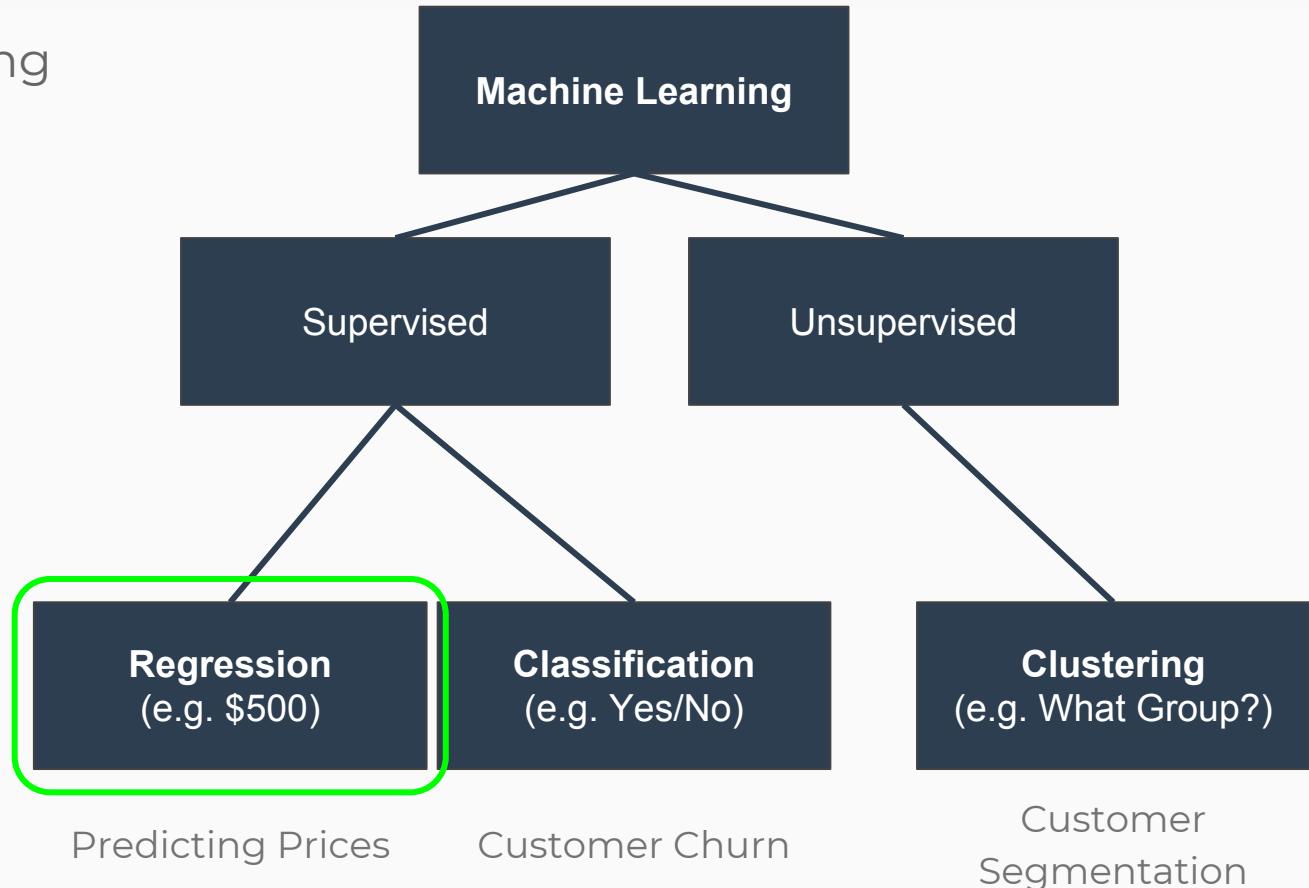
## Accounting

- Payment Anomaly Detection

# Machine Learning Primer

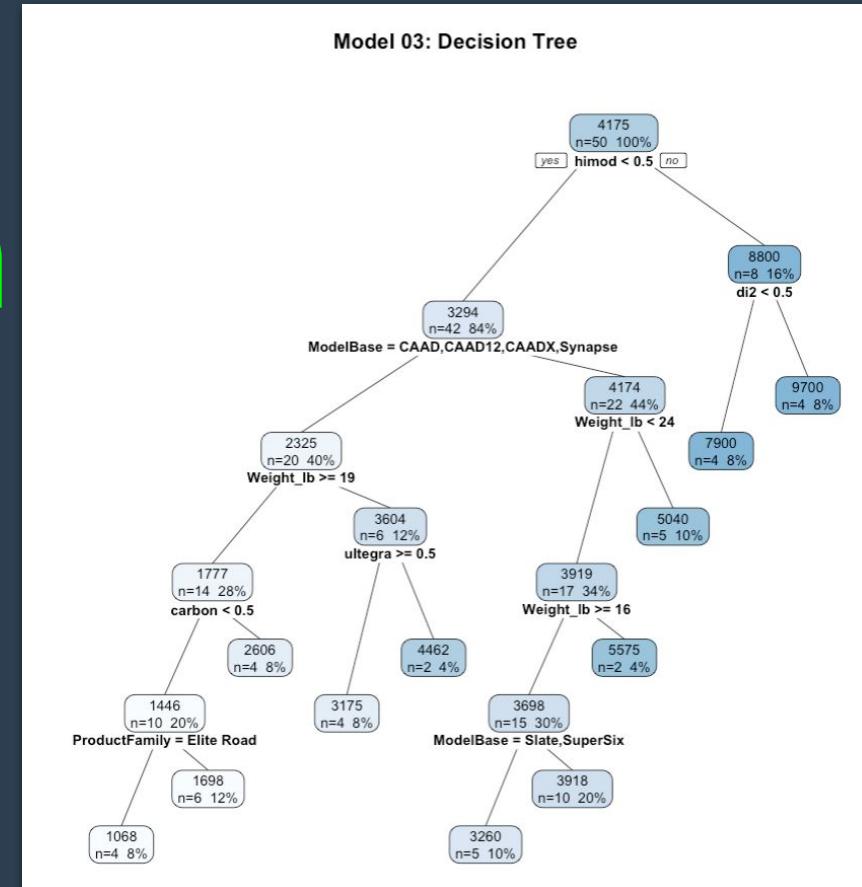


- We are performing **Regression**
- Stepping stone towards Classification
- Predicting a **Numeric Value (Price)**



# Regression Example

Applying ML  
Product Price Model





# Presentation Materials

[github.com/business-science/presentations](https://github.com/business-science/presentations)

 mdancho84	Learning Lab 05	Latest commit b959a21 6 minutes ago
 2017_05_19_R-Finance	-Add R User Group Panama Pres.	2 years ago
 2017_06_05_EARL_SF_2017	-Add R User Group Panama Pres.	2 years ago
 2017_09_19_R_UserGroup_Panama	EARL Boston 2017	a year ago
 2017_11_05_EARL_Boston_2017	EARL Boston 2017	a year ago
 2018_02_02_rstudio-conf-2018/presentation	presentation update	3 hours ago
 2018_04_19_SP_Global_Time_Series_Deep_Lea...	R/Finance 2018 Presentation	10 months ago
 2018_06_01_R-Finance-Time-Series-Platform...	R/Finance 2018 Presentation	10 months ago
 2018_10_13_DSGO18_Business_Science_Probl...	fix readme	5 months ago
 2019_01_09_How_To_Learn_R_Fast	how to learn R fast presentation	2 months ago
 2019_01_RStudio-Conf	Add RStudio Conference 2019 Presentation	13 days ago
 2019_02_13_Learning_Lab_Marketing_Analytics	update report	3 hours ago
 2019_03_13_Learning_Lab_05_Intro_to_Machin...	Learning Lab 05	6 minutes ago



# Data



- **Real Data:**  
Web-Scraped  
Product Data
- **Goal #1:** Build a  
prediction model for  
Bike Prices
- **Goal #2:** Understand  
what features are  
driving the model(s)

Synapse Carbon | The Most Complete Lineup of Bikes

https://www.cannondale.com/en/USA/Products/ProductCategory.aspx?nid=2cbf69ae-61b1-4002-acfa-8a1afaf61b49

cannondale

ROAD MOUNTAIN URBAN & FITNESS ELECTRIC WOMEN'S KIDS GEAR

## SYNAPSE CARBON

SYNAPSE HI-MOD DISC DURA-ACE DI2 \$10,500 [VIEW DETAILS](#)

SYNAPSE HI-MOD RED ETAP AXS \$9,500 [VIEW DETAILS](#)

SYNAPSE HI-MOD DISC RED ETAP \$8,400 [VIEW DETAILS](#)

# Machine Learning Package



- **Parsnip:**

- New
- Connects to modeling packages

- Part of **tidymodels** framework

- **Links**

- [tidymodels.github.io/parsnip](https://tidymodels.github.io/parsnip)
- [Model List](#)

Screenshot of a web browser displaying the "List of Models" page for the `parsnip` package. The URL is <https://tidymodels.github.io/parsnip/articles/articles/Models.html>. The page shows a comparison table of various modeling engines across different model types.

model	engine	class	classprob	conf_int	numeric	pred_int	quantile	...
boost_tree()	C5.0	✓	✓	✗	✗	✗	✗	✗
	spark	✓	✓	✗	✓	✗	✗	✗
	xgboost	✓	✓	✗	✓	✗	✗	✗
decision_tree()	C5.0	✓	✓	✗	✗	✗	✗	✗
	rpart	✓	✓	✗	✓	✗	✗	✗
	spark	✓	✓	✗	✓	✗	✗	✗
linear_reg()	glmnet	✗	✗	✗	✓	✗	✗	✗
	keras	✗	✗	✗	✓	✗	✗	✗
	lm	✗	✗	✓	✓	✓	✗	✗
logistic_reg()	spark	✗	✗	✗	✓	✗	✗	✗
	stan	✗	✗	✓	✓	✓	✗	✗
	glm	✓	✓	✓	✗	✗	✗	✗
logistic_reg()	glmnet	✓	✓	✗	✗	✗	✗	✗
	keras	✓	✓	✗	✗	✗	✗	✗

# Demo Time

Let's code!

# Learning Recommendations

- Learn Key Algorithms
- Apply to Advanced ML to Business
- Distribute Results

# Learn the Key Algorithms

*Linear Regression  
Regularized Regression  
Decision Trees  
Random Forests  
Gradient Boosted Machines  
Support Vector Machines*

## Business Analysis with R



- **7 Week System**
  - Data Science Fundamentals
- **Week 6 (Machine Learning)**
  - 5 Hours of Lessons to go from Zero to ML-Pro

**DS4B 101-R: Business Analysis With R**

Your Data Science Journey Starts Now! Learn the fundamentals of data science for business with the tidyverse.

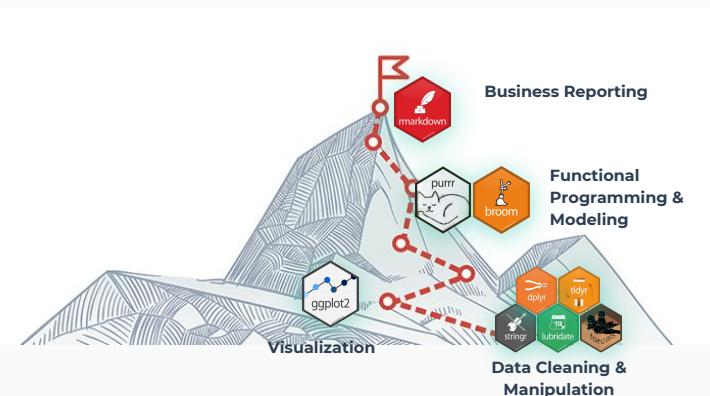
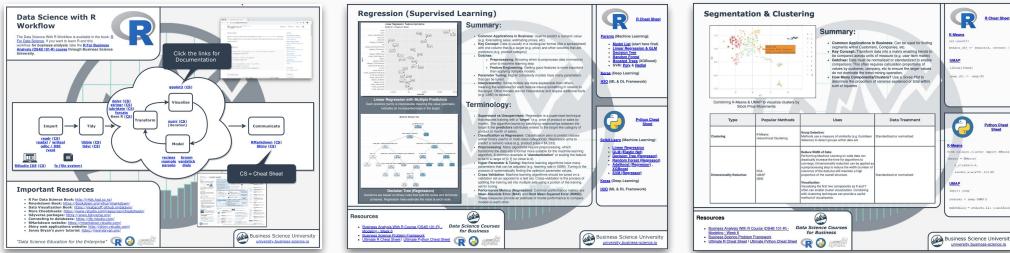
Matt Dancho

# Key Benefits

- Fundamentals - Weeks 1-5  
(25 hours of Video Lessons)
  - Data Manipulation (dplyr)
  - Time series (lubridate)
  - Text (stringr)
  - Categorical (forcats)
  - Visualization (ggplot2)
  - Programming & Iteration (purrr)
  - 3 Challenges
- **Machine Learning - Week 6**  
**(8 hours of Video Lessons)**
  - Clustering (3 hours)
  - Regression (5 hours)
  - 2 Challenges
- Learn Business Reporting - Week 7
  - RMarkdown
  - 2 Projects

# Business Analysis with R (DS4B 101-R)

Data Science Foundations  
**7 Weeks**



# Learn Binary Classification

*Most Critical Business Problems  
are Binary Classification (Yes/No)*

*Will this customer churn?*

*Is this employee going to leave?*

*Will this customer going to miss payment  
in next 90-days?*

## Data Science for Business with R

- **10 Week System**
  - Advanced Machine Learning
- **Churn with H2O AutoML**



The banner features the R logo, the H2O.ai logo, and various data visualization elements like a bar chart and a scatter plot.

**DS4B 201-R: Data Science For Business With R**

Solve a real-world churn problem with H2O AutoML (automated machine learning) & LIME black-box model explanations using R

Matt Dancho

# Data Science For Business with R

## (DS4B 201-R)



The screenshot displays the H2O AutoML interface running in a Jupyter Notebook environment. The left pane shows a code cell with R script for generating a legend and plotting ROC curves. The right pane shows four performance plots: ROC, Precision vs Recall, Gain, and Lift. Below the plots, a table lists various models with their names and AUC values.

```
5.13  # Legend using codelign
5.14  p_legend <- get_legend(p)
5.15  pd %>% theme(legend.position = "none")
5.16
5.17  p <- codelign_grid(pd, x_d, y_d, max_d = 2)
5.18
5.19  p[[1]] + ggtitle("H2O Model Metrics", vsize = 10, fontface = "bold",
5.20    color = palette$light[1][1])
5.21
5.22  p[[1]] + draw_grid(p[[1]] %>% filter_by(isupper(lower_box))%, size = 10,
5.23    color = palette$light[1][1])
5.24
5.25  ret <- plot_grid(p[[1]], p[[2]], p[[3]],
5.26    max_d = 1, rel_heights = c(0.6, 0.05, 1, 0.05 * max_model))
5.27
5.28  h2o.show_progress()
5.29
5.30  return(ret)
5.31 }
5.32
5.33 automl_models %>% head(5) %>% select(-model_id) %>% arrange(-auc)
5.34
5.35 plot_h2o_performance(automl_models %>% head(5), order_by = "auc_lgbm", n_models = 2)
```

Model	AUC
StackedEnsemble	0.84701
StackedEnsemble_0_AutoML_20180103_051516_0.84701	0.84701
GAM_0_AutoML_20180103_051516_0.84700	0.84700
GBM_0_AutoML_20180103_051516_0.84699	0.84699
COR_grid_R_AutoML_20180103_051516_model_4_0.85145	0.85145

# H2O AutoML

# **Weeks 5 & 6: H2O**

## Modeling & Performance

In **Week 5 (modeling)**, you learn generate 30+ models & visualize results using ggplot2

In **Week 6 (performance)**, you go in-depth learning ROC & AUC, Precision vs Recall, & Gain & Lift Plots

# Data Science For Business with R (DS4B 201-R)



## Week 7: Explaining Black-Box Models

*“The business won’t care how high your AUC is if you can’t explain your Machine Learning Model”*

In **Week 7 (Explaining Models)**, you use LIME to explain the “black-box” ensemble models & extract insights to answer **what is causing churn**

# Distribute Results

*Businesses need  
Web Applications*

*Interactive Visualizations*

*Dashboard Interface*

*Built-In Prediction & Explanation*

## Data Science for Business with R

- **6 Week System**
  - Integrate ML into
  - Web App

DS4B 301-R: Shiny Web Apps For Business

Build Distributed Web Applications using Machine Learning

Coming Soon!



# R Shiny Web Apps For Business

## (DS4B 301-R)

~/Google Drive/Business Science/Learning Labs/2019-01-09-How\_To\_Learn\_R\_Fast/Shiny\_App\_Employee\_Attrition\_Predictor/R/application - Shiny  
http://127.0.0.1:870 / Open in Browser

HR Analytics Employee Risk Department Risk Executive Report

Employee Number: 891

Attrition Prediction: Yes  
Elevated Risk Prediction: 47.08%  
Low Risk Prediction: 52.92%

Job Role: Manager  
Department: Human Resources  
Education Field: Human Resources

**H2O.ai**

Attrition Risk Management Strategies

Employee Risk Assessment: Low

Feature Contributions to Attrition Risk

Feature Description Map

Feature Direction

Over Time = No	Low Risk
Stock Option Level = 1	High Risk
28 < Years = No	High Risk
6 < Num of Children = No	High Risk
Job Satisfaction = High	High Risk
Training Times Less than 2	High Risk
Job Level = 3	High Risk
Work Life Balance = Better	Low Risk





# Business Science Learning System

## THE PLAN



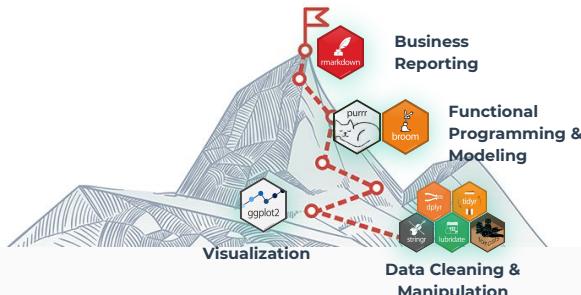
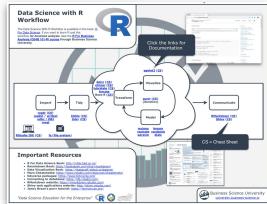
### Business Analysis with R (DS4B 101-R)

### Data Science For Business with R (DS4B 201-R)

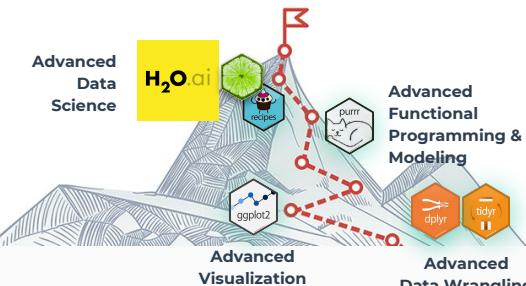
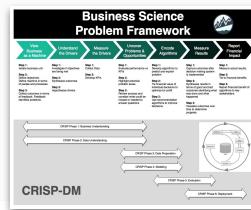
### R Shiny Web Apps For Business (DS4B 301-R)

#### Project-Based Courses with Business Application

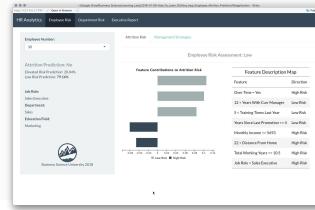
Data Science Foundations  
**7 Weeks**



Machine Learning & Business Consulting  
**10 Weeks**



Web Application Development  
**6 Weeks**





# THE RESULT



Business Analysis with R

(DS4B 101-R)

7 WEEKS



Data Science For Business with R

(DS4B 201-R)

10 WEEKS



R Shiny Web Apps For Business

(DS4B 301-R)

6 WEEKS

**23 WEEKS!**  
**(OR LESS)**



# 15% OFF PROMO Code: **learninglabs**



**DS4B 101-R: Business Analysis With R**

Your Data Science Journey Starts Now! Learn the fundamentals of data science for business with the tidyverse.

Matt Dancho

MSRP: ~~\$349~~

Your Price! **\$297**

Save: \$52

**DS4B 201-R: Data Science For Business With R**

Solve a real-world churn problem with H2O AutoML (automated machine learning) & LIME black-box model explanations using R.

Matt Dancho

MSRP: ~~\$499~~

Your Price! **\$421**

Save: \$78

Courses Included with Purchase

**DS4B 201-R: Data Science For Business With R**  
Solve a real-world churn problem with H2O AutoML (automated machine learning) & LIME black-box model explanations using R.  
 Matt Dancho **\$495**

**DS4B 101-R: Business Analysis With R**  
Your Data Science Journey Starts Now! Learn the fundamentals of data science for business with the tidyverse.  
 Matt Dancho **\$349**

Original Price: **\$844**



MSRP: ~~\$844~~

Your Price! **\$637**

Save: \$207

# THE BONUS



Start your journey with **15% OFF**  
**PROMO Code: learninglabs**  
[university.business-science.io](http://university.business-science.io)

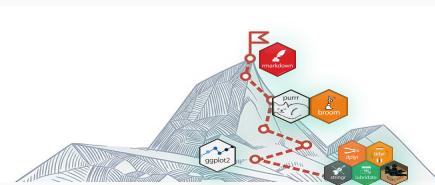
Business Analysis with R  
(DS4B 101-R)

Data Science For Business with R  
(DS4B 201-R)

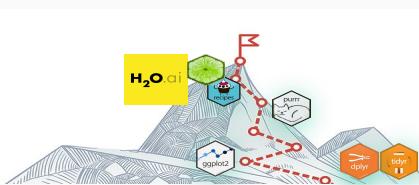
R Shiny Web Apps For Business  
(DS4B 301-R)

## Project-Based Courses with Business Application

Data Science Foundations  
**7 Weeks**



Machine Learning &  
Business Consulting  
**10 Weeks**



Web Application Development  
**6 Weeks**

