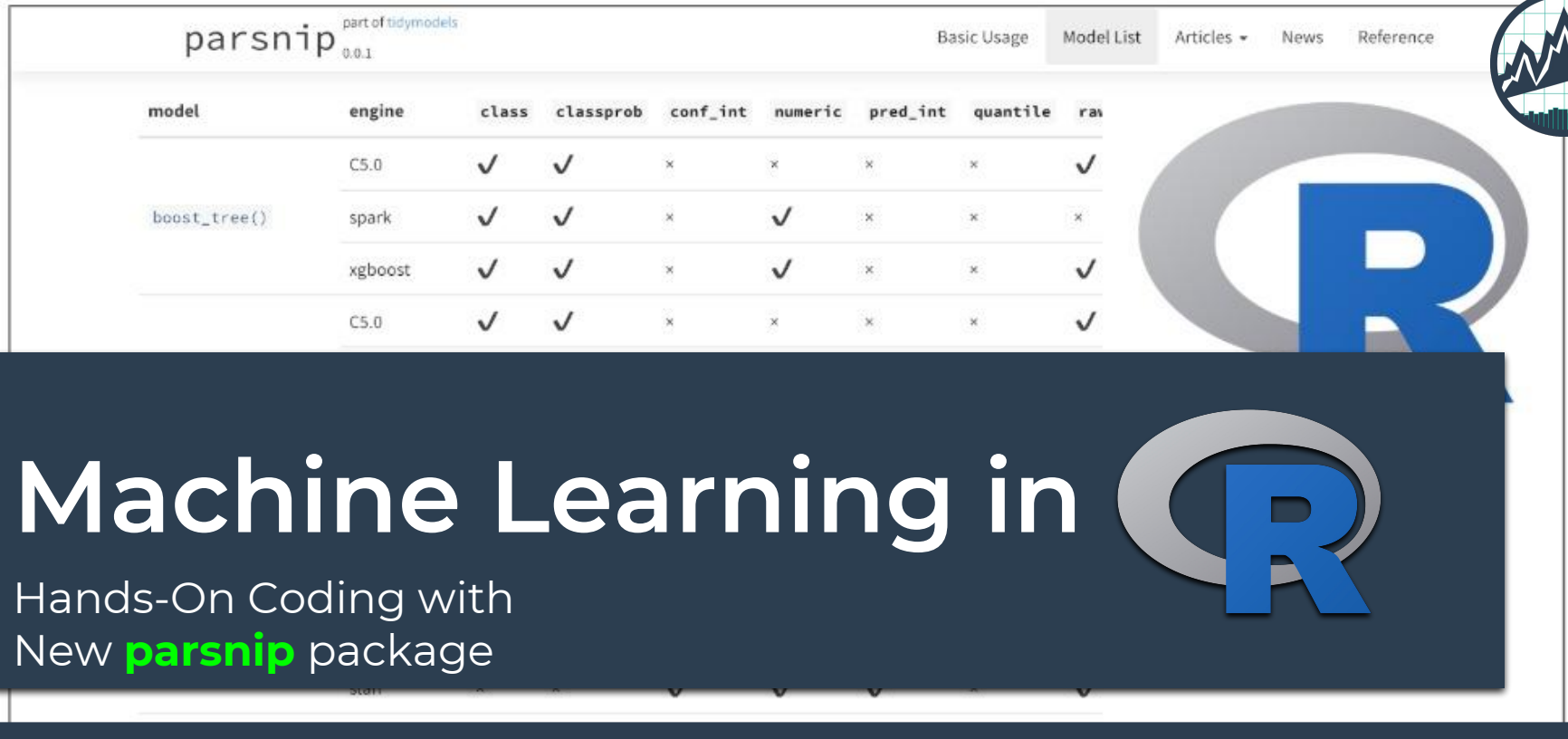


Machine Learning in R

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**.

Agenda

- **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

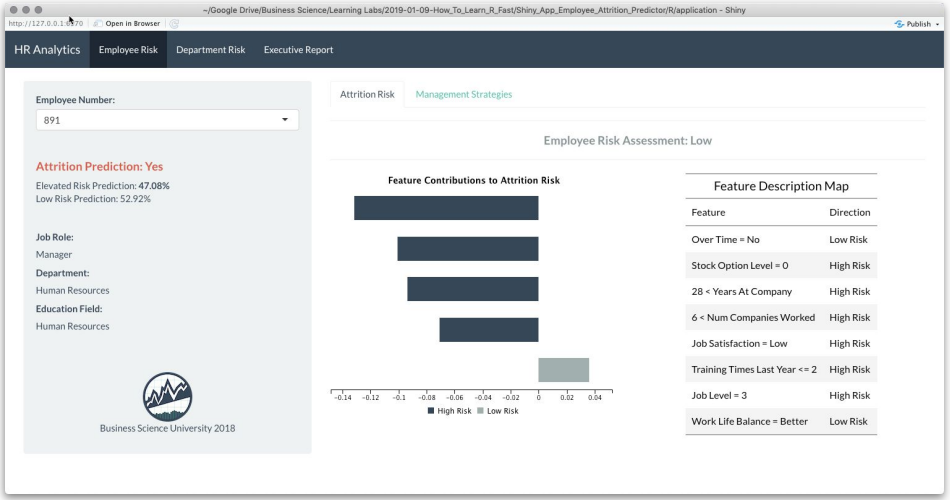
```
# XGBoost ----
set.seed(1)
model_04_xgboost <- boost_tree("regression",
                               mtry = 18,
                               trees = 1000,
                               min_n = 2,
                               tree_depth = 6,
                               learn_rate = 0.29,
                               loss_reduction = 60
                               ) %>%
    set_engine("xgboost") %>%
    fit(Price_num ~ ., data = train_tbl %>% select(-row_id))

model_04_xgboost %>%
    predict(new_data = test_tbl) %>%
    bind_cols(test_tbl %>% select(Price_num)) %>%
    yardstick::metrics(truth = Price_num, estimate = .pred)
```

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 Optimizatio

Finance

- Fraud Detection
- Credit Default Ris
- Loan Delinquency

HR

- Employee Attrition
- Promotion Readiness

Marketing

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

Accounting

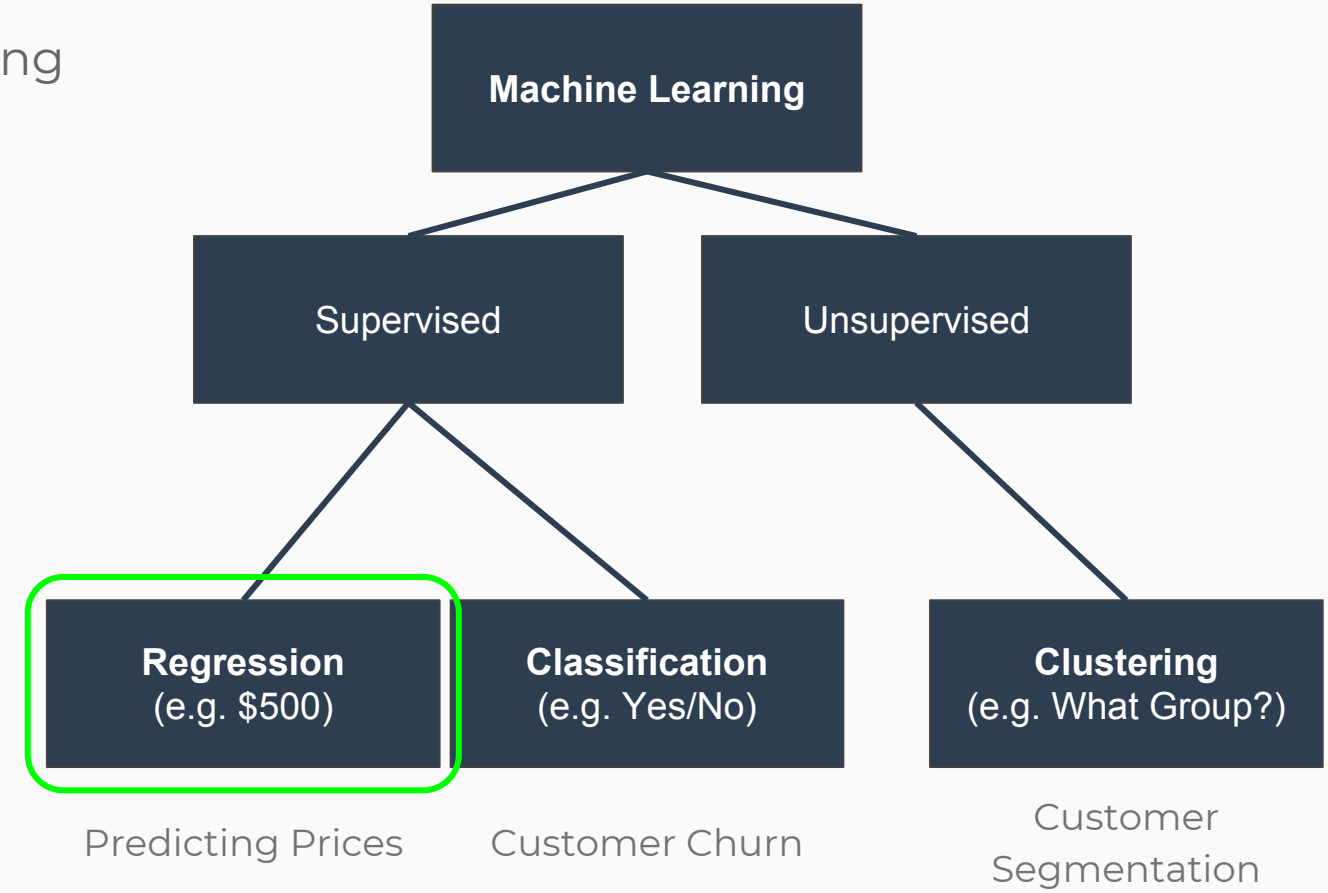
- Payment Anomaly Detection

Binary Classification

Will a customer leave - Yes/No?

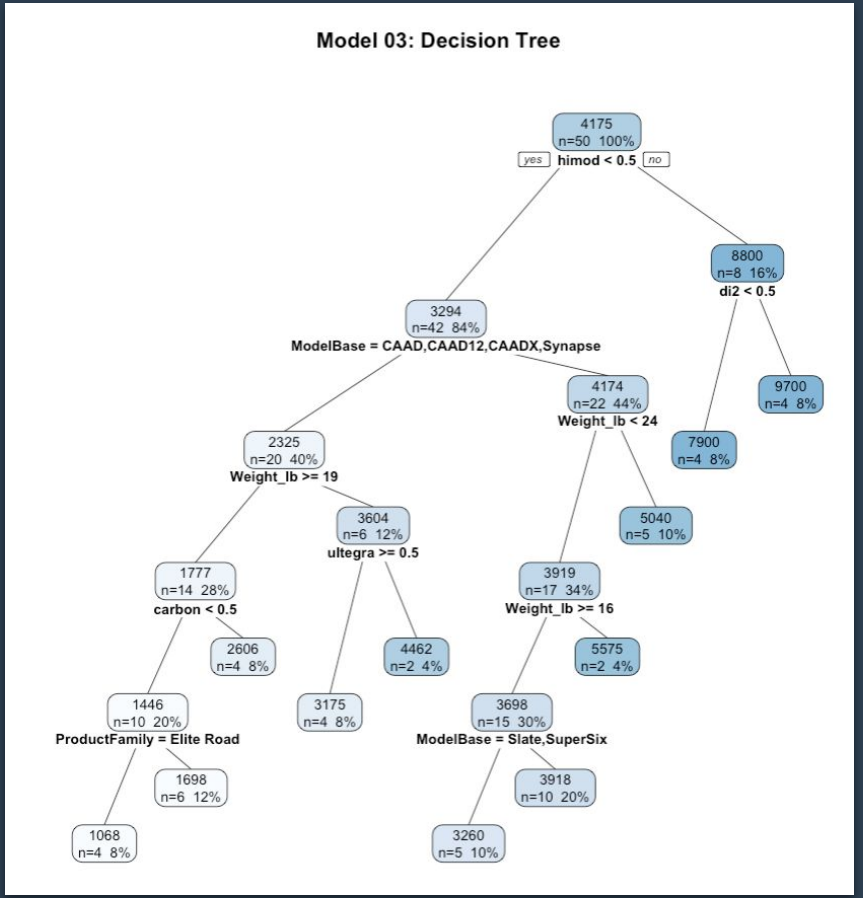
Machine Learning Primer

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



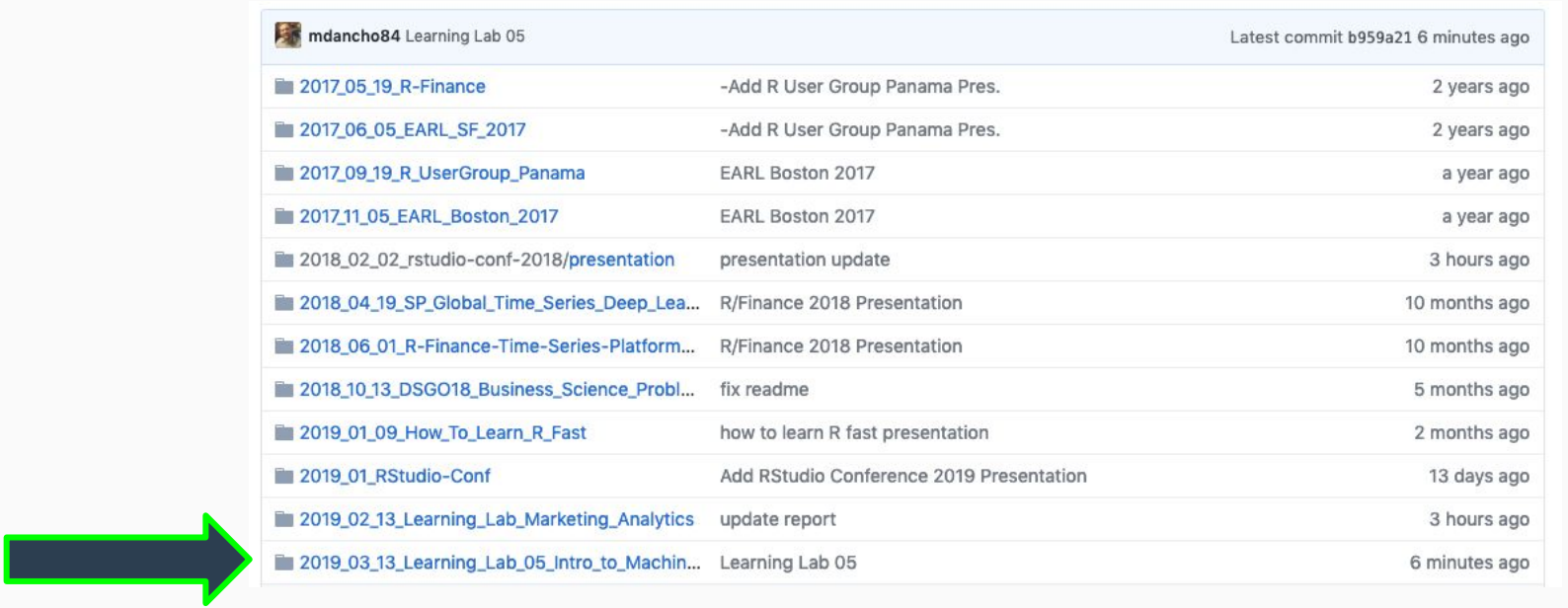
Regression Example

Applying ML Product Price Model



Presentation Materials

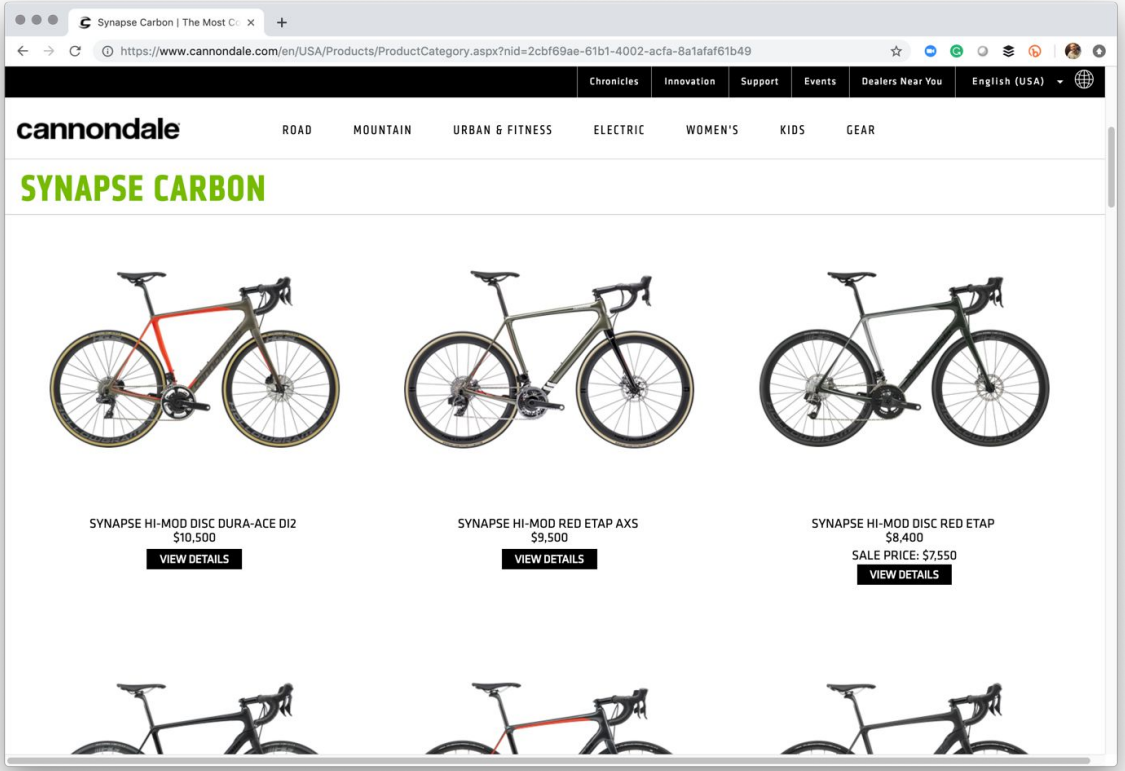
github.com/business-science/presentations



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

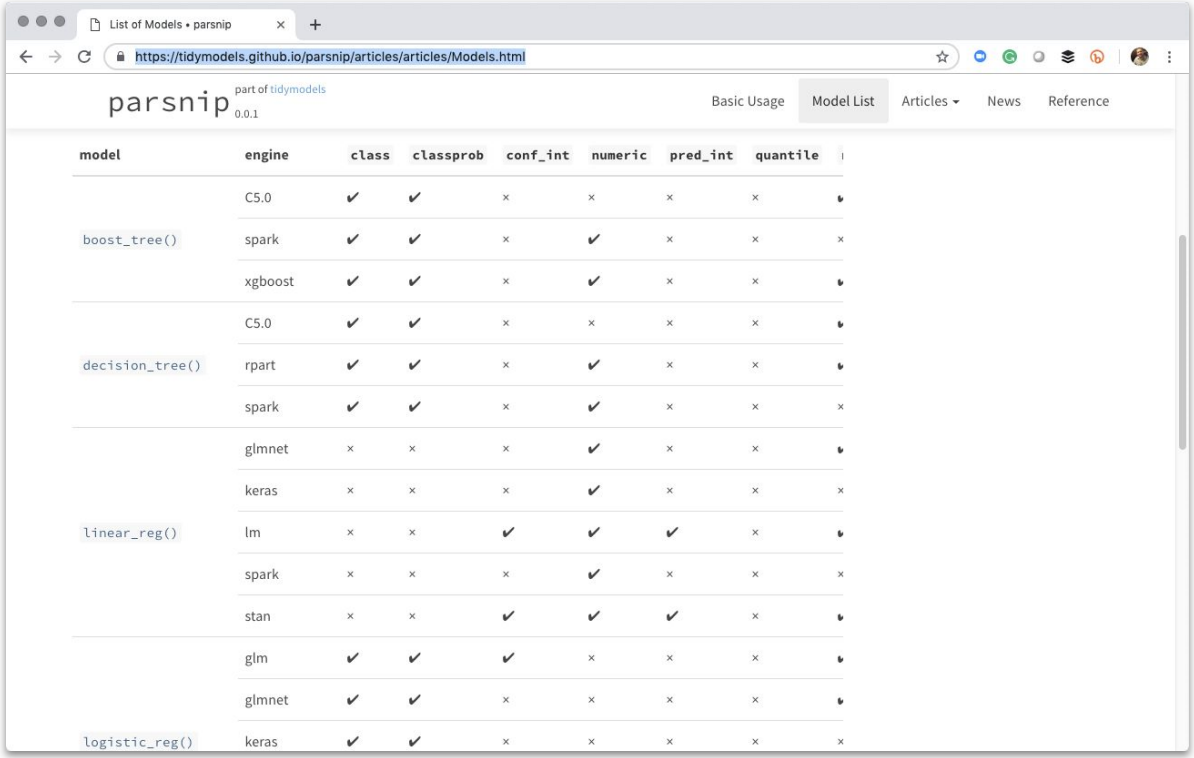
Data

- **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)



Machine Learning Package

- **Parsnip:**
  - New
  - Connects to modeling packages
- Part of **tidymodels framework**
- **Links**
  - tidymodels.github.io/parsnip
  - Model List



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	×	×	✔	✔	✔	×
	spark	×	×	×	✔	×	×
	stan	×	×	✔	✔	✔	×
logistic_reg()	glm	✔	✔	✔	×	×	×
	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

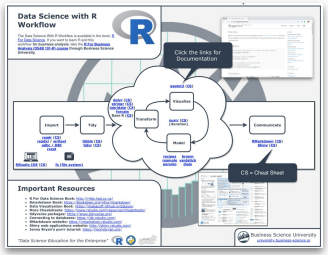
Key Benefits

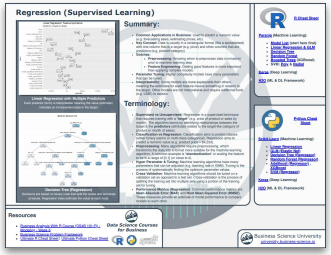
- Fundamentals - Weeks 1-5 (25 hours of Video Lessons)
  - Data Manipulation (dplyr)
  - Time series (lubridate)
  - Text (stringr)
  - Categorical (forcats)
  - Visualization (ggplot2)
  - Programming & Iteration (purrr)
  - 3 Challenges
- **Machine Learning - Week 6 (8 hours of Video Lessons)**
  -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

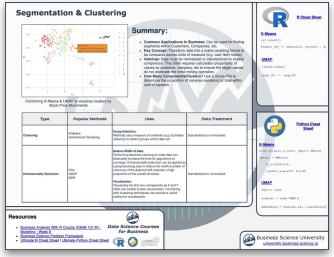
Business Analysis with R (DS4B 101-R)

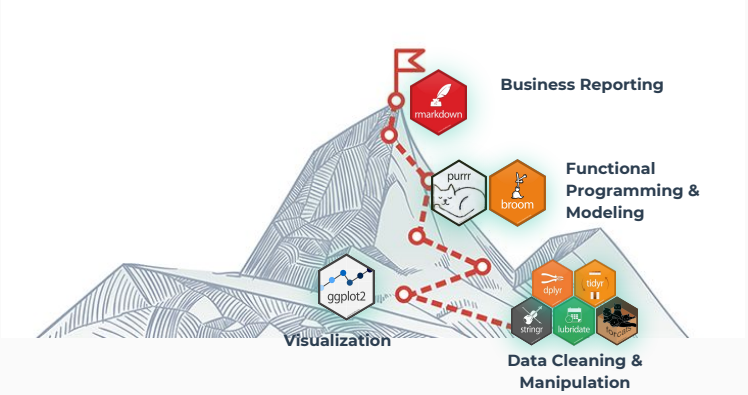
Data Science Foundations

7 Weeks









Learn Binary Classification

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

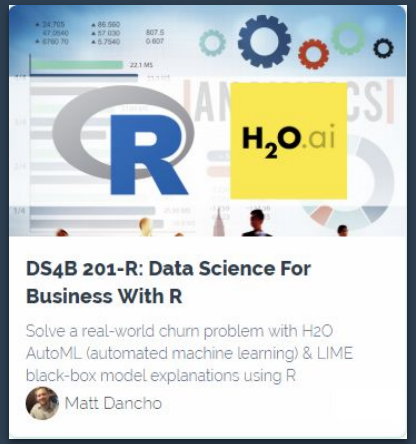
Will this customer churn?

Is this employee going to leave?

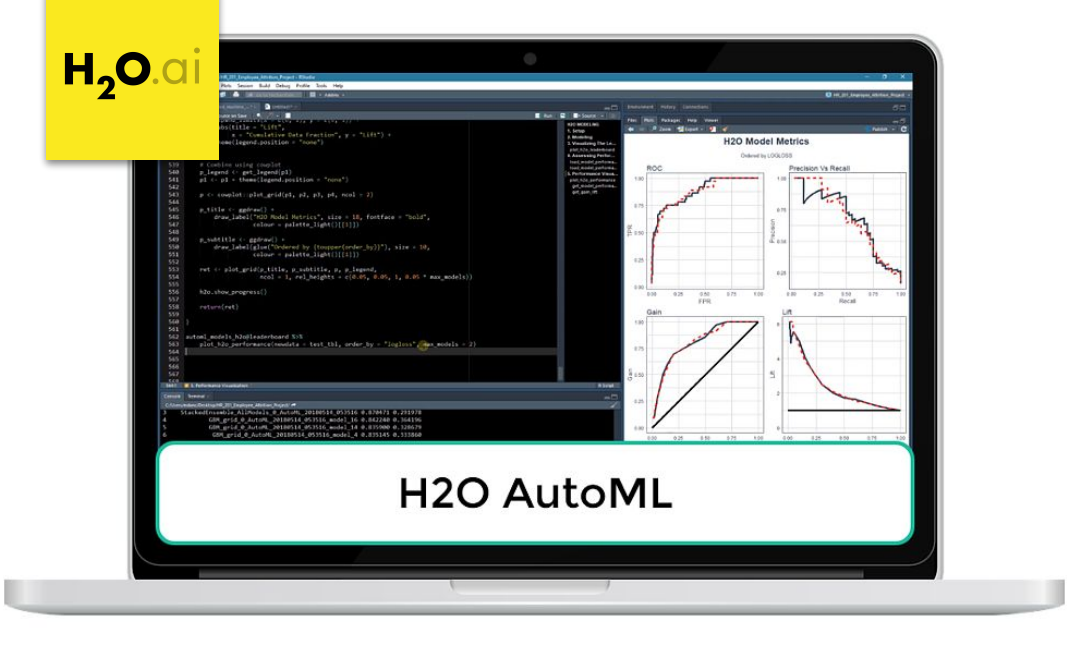
Will this customer going to miss payment in next 90-days?

Data Science for Business with R

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



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

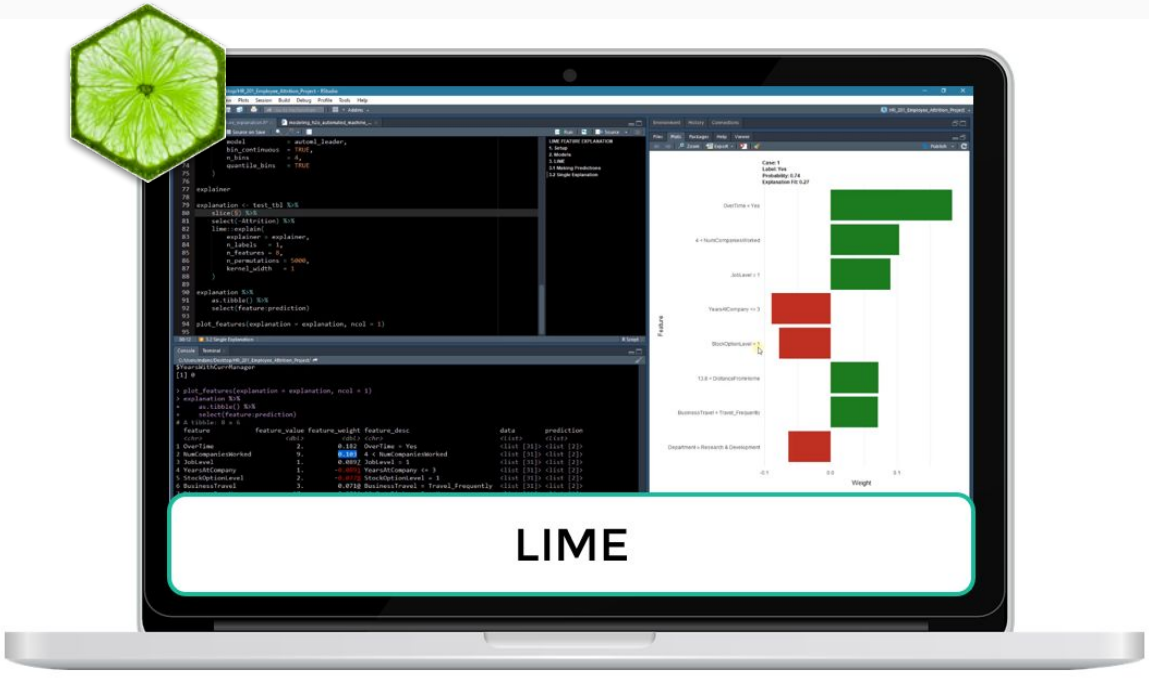


Weeks 5 & 6: H2O Modeling & Performance

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

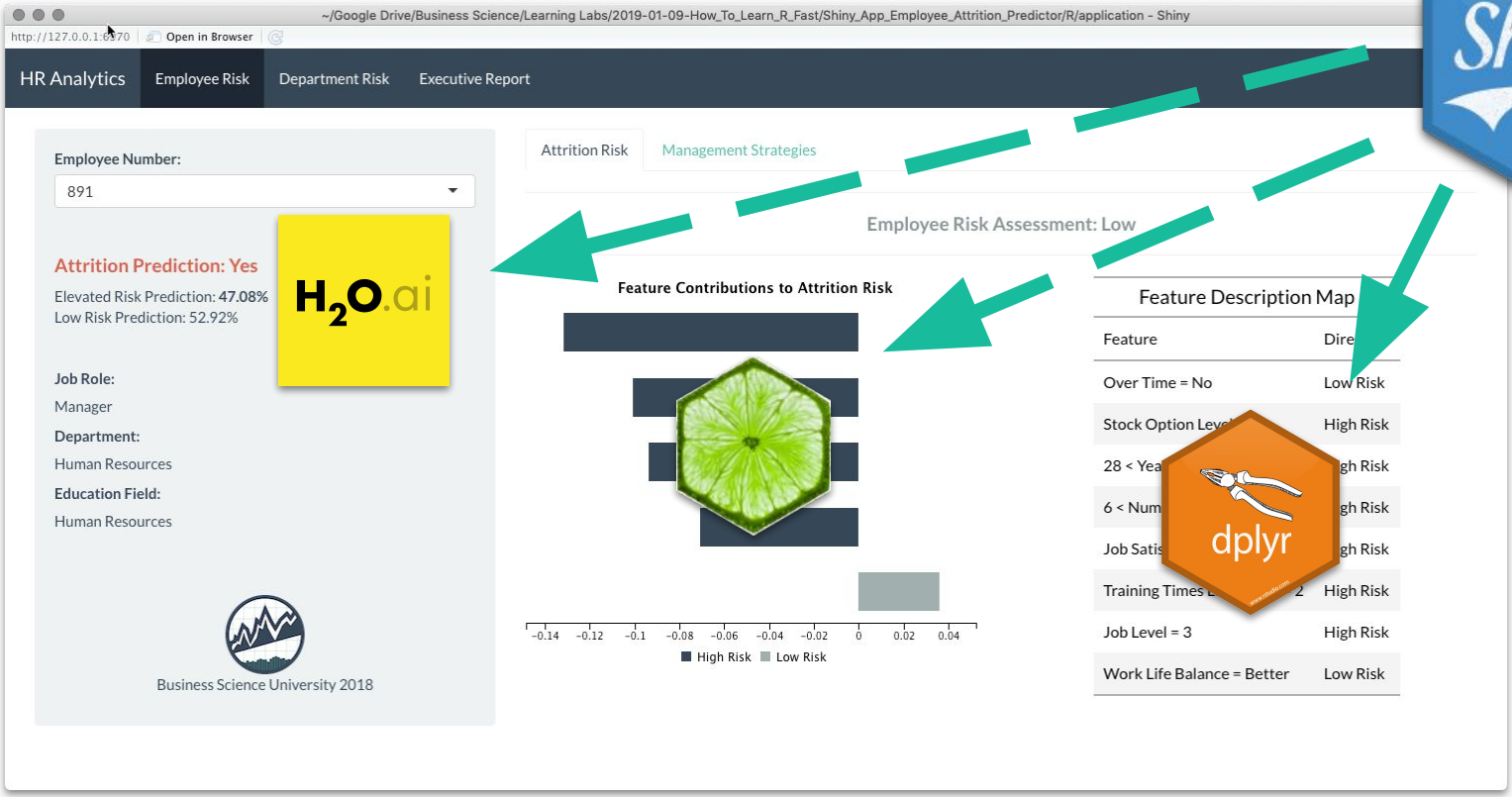
Dashboard Interface

Built-In Prediction & Explanation

Data Science for Business with R

- 6 Week System
  - Integrate ML into
  - Web App

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



Business Science Learning System

THE PLAN

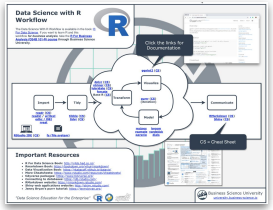
Business Analysis with R
(DS4B 101-R)

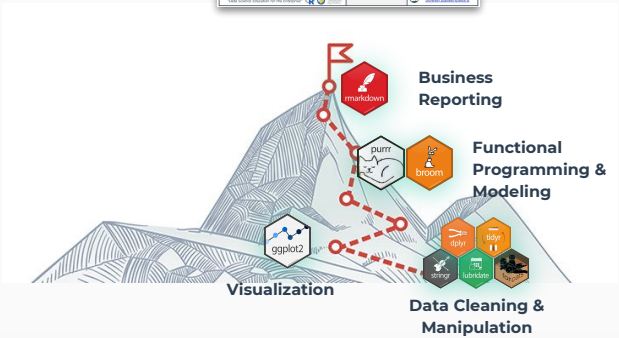
Data Science For Business with R
(DS4B 201-R)

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

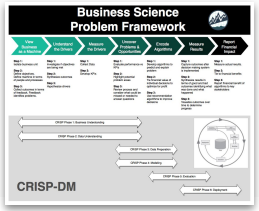
Project-Based Courses with Business Application

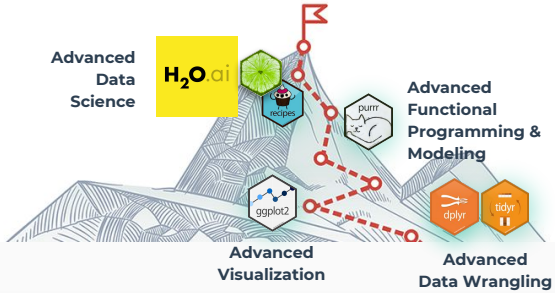
Data Science Foundations
7 Weeks



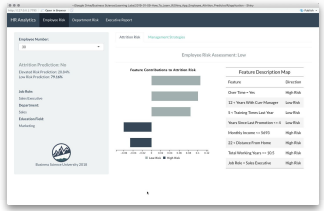


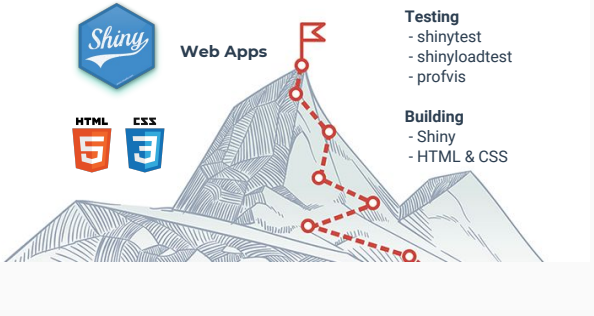
Machine Learning & Business Consulting
10 Weeks





Web Application Development
6 Weeks





THE RESULT

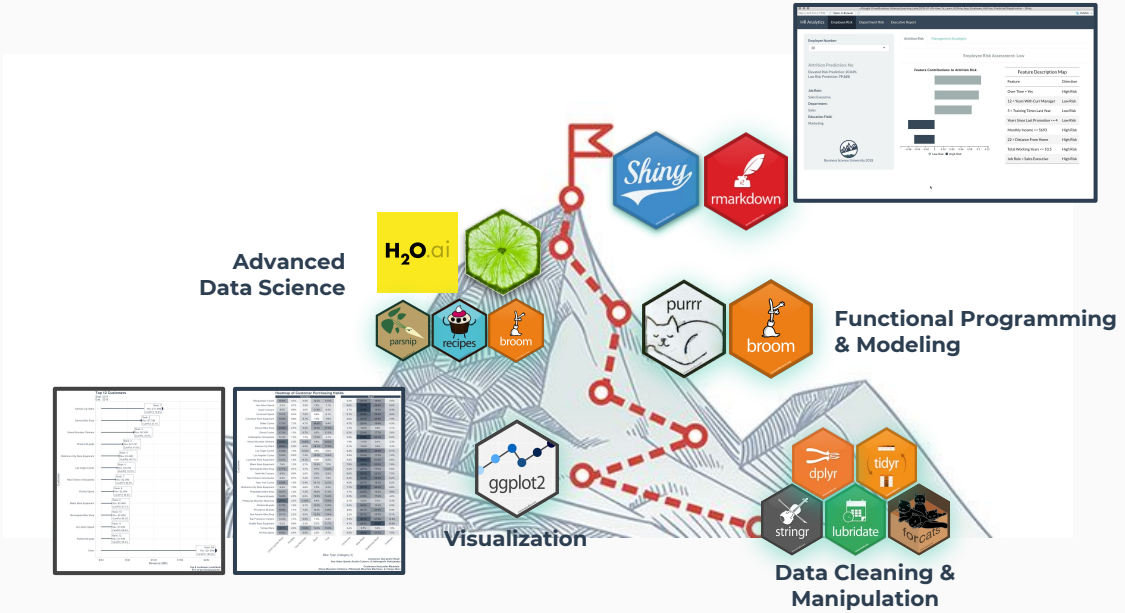
Business Analysis with R
(DS4B 101-R)
7 WEEKS

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)

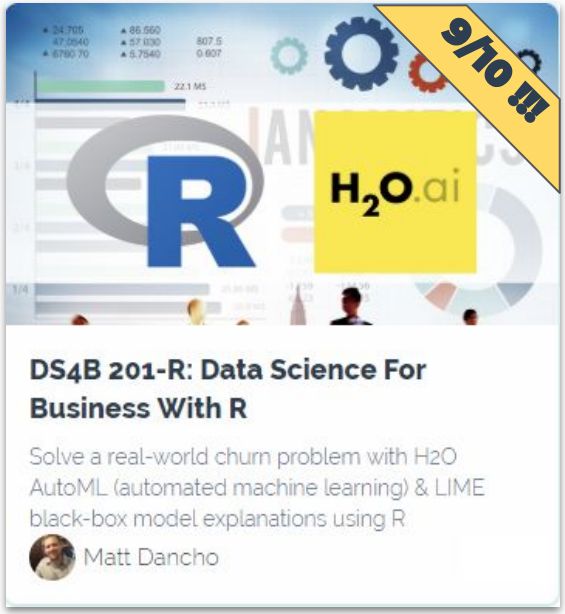


15% OFF PROMO Code: learninglabs

MSRP: ~~$349~~

Your Price! **$297**

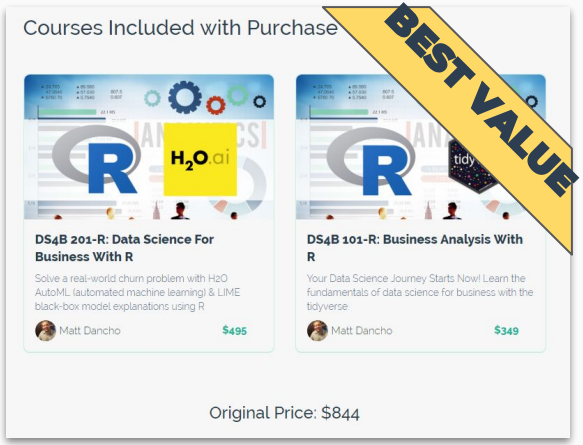
Save: $52



MSRP: ~~$499~~

Your Price! **$421**

Save: $78



R-TRACK BUNDLE

MSRP: ~~$844~~

Your Price! **$637**

Save: $207

THE BONUS

Start your journey with 15% OFF
PROMO Code: learninglabs

university.business-science.io

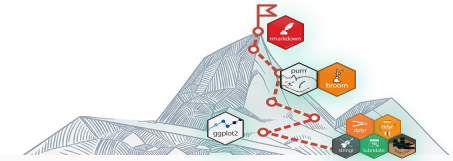
Business Analysis with R
(DS4B 101-R)

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

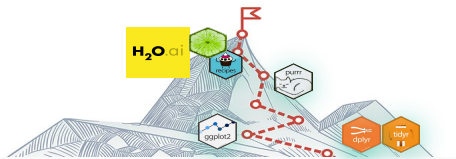
R Shiny Web Apps For Business
(DS4B 301-R)

Project-Based Courses with Business Application

Data Science Foundations
7 Weeks



Machine Learning &
Business Consulting
10 Weeks



Web Application Development
6 Weeks