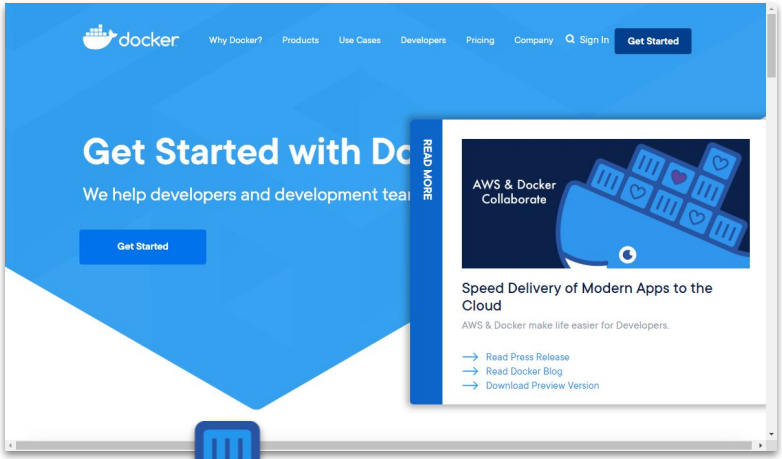
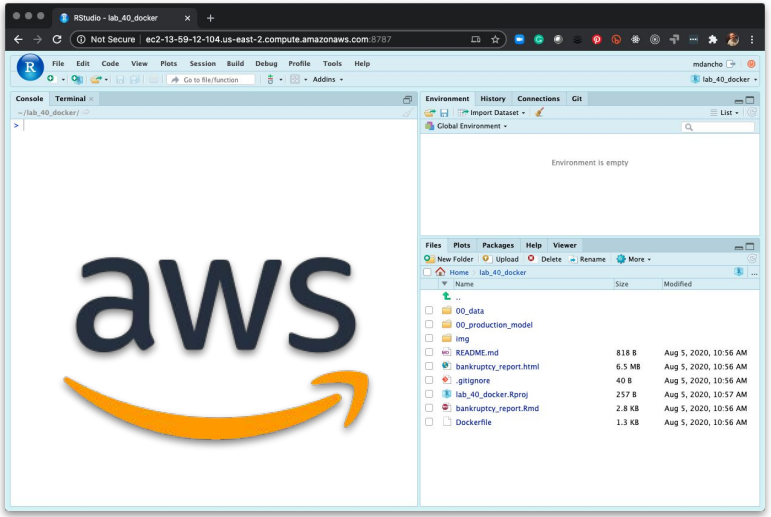
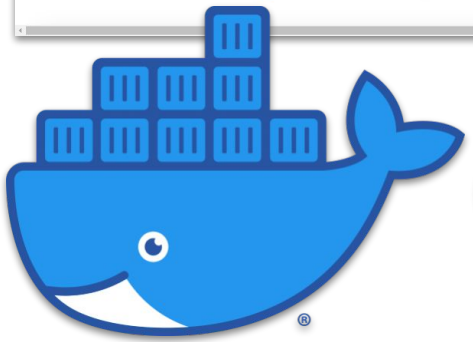






Docker for Data Science

Reproducible Machine Learning

Matt Dancho & David Curry

Business Science Learning Lab

Learning Lab Structure

- **Presentation**
  (20 min)
- **Demo's**
  (30 min)
- **Pro-Tips**
  (15 mins)

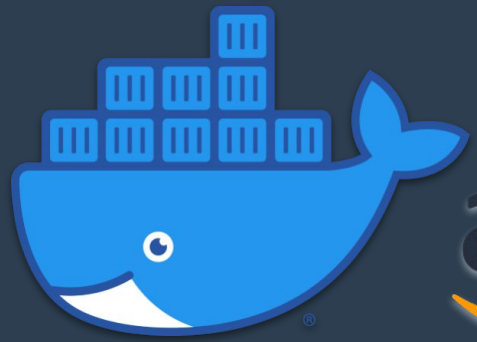
Matt Dancho

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

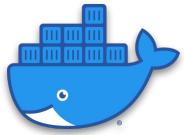
Founder of Sure Optimize, David works with businesses to help improve website performance and SEO using data science. His passion is **ethical Machine Learning initiatives**.

MLOps Series



- **Business Case**
  - Problem with H2O reproducibility
  - How Docker makes H2O reproducible
- **Docker**
  - Key Concepts
    - Docker Workflow
    - Working w/ GitHub
- **30-Min Demo**
  - Bankruptcy Classifier
    - MLFlow + H2O Grid Search
    -
  - **[Bonus] Run RStudio + H2O in Cloud w/ AWS**

- **Keys to Your Transformation**
  - **Reproducible Analysis**
  - **Deploying Applications**



MLOps Series

- **Lab 39 - MLFlow**
  - MLFLow Tracking API - H2O Machine Learning Artifacts
- **Lab 40 - Docker**
  - Reproducible Analysis
- **Lab 41 - Drake & Renv**
  - Workflow Dependencies, Caching Expensive Ops & Package Management
- **Lab 42 - Plumber APIs**
  - Building APIs for apps to connect to

Learning Labs PRO

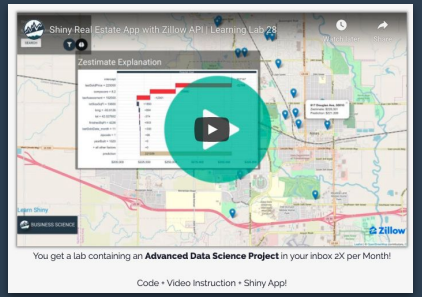
40 Labs

1.5-Hour Courses

2X Per Month

100's of Valuable Bonuses

$349/year or $39/month



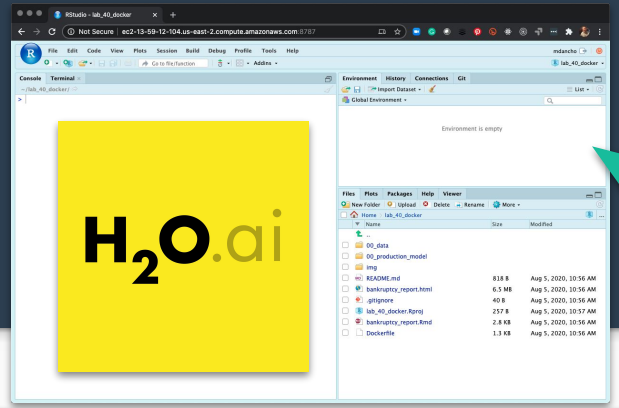
Learn Data Science Tools

Get Labs *2X per Month* on Advanced Topics

https://university.business-science.io/p/learning-labs-pro

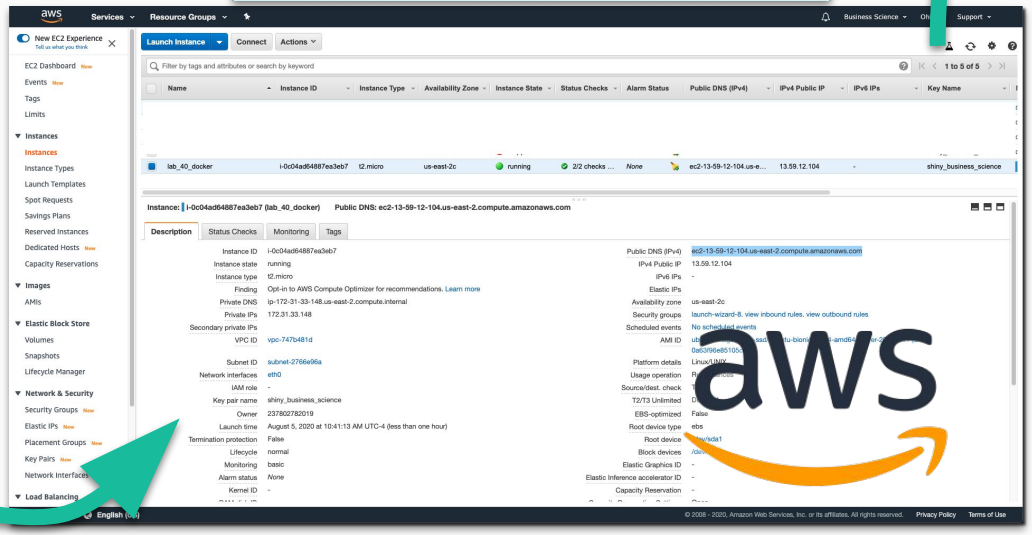
Bonuses

for LL PRO Members Today



Bonus #1

AWS Tutorial



Bonuses

for LL PRO Members Today

Bonus #2

MLFLOW Model Tracking Automation

```
28
29  # Iterate through Grid Search
30  run_h2o_grid_search_with_mlflow(
31
32      data       = data_prepared_tbl,
33      target     = "class",
34
35      # H2O Grid Search
36      h2o_init   = TRUE,
37      n_trees    = c(25, 50, 100),
38      max_depth  = c(5, 10),
39
40      # MLFLOW
41      launch_mlflow_ui        = TRUE,
        mlflow_tracking_uri     = "mflow_test_2",
        mlflow_experiment_name = "Bankruptcy Prediction API"
44  )
45
```

H2O.ai

mlflow

Problem

Reproducing Analysis

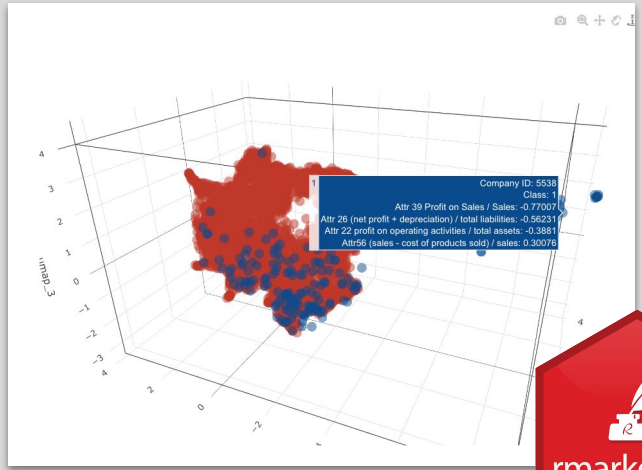
Version

Incompatibility

You can not load models built from prior versions of H2O

Problem

Your Machine



rmarkdown

Their Machine

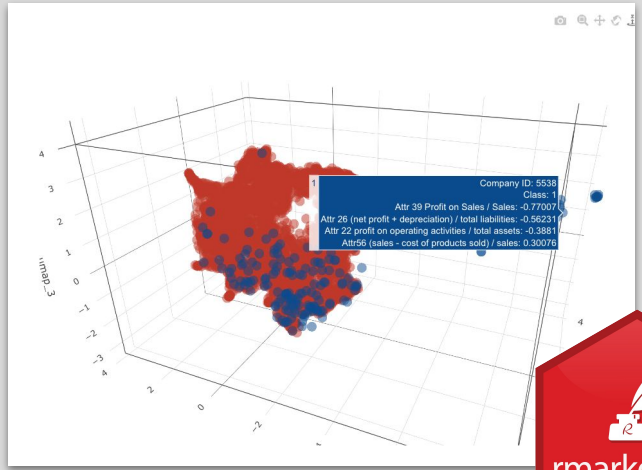
```
69
70  ### Load Production Model
71
72  ```{r}
73  path <- file.path(rprojroot::find_rstudio_root_file(),
74                    "00_production_model/PROD_H2O_MODEL1")
75
76  h2o_model <- h2o.loadModel(path)
77  ```
```

```
ERROR: Unexpected HTTP Status code: 412 Precondition Failed (url =
http://localhost:54321/99/Models.bin/)

water.exceptions.H2OIllegalArgumentException
 [1] "water.exceptions.H2OIllegalArgumentException: Illegal argument:
water.api.FSIOException: FS IO Failure: \n accessed path :
file:/Users/mdancho/Desktop/learning_labs/lab_40_docker/00_production_
File not found"
```

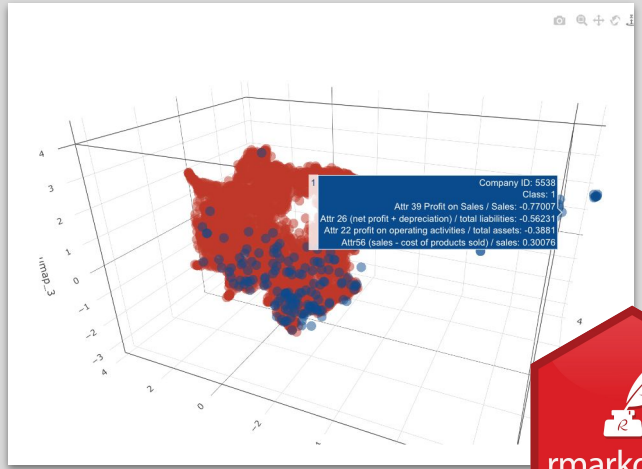
Solution

Your Machine



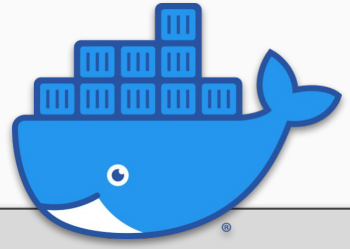
rmarkdown

Their Machine

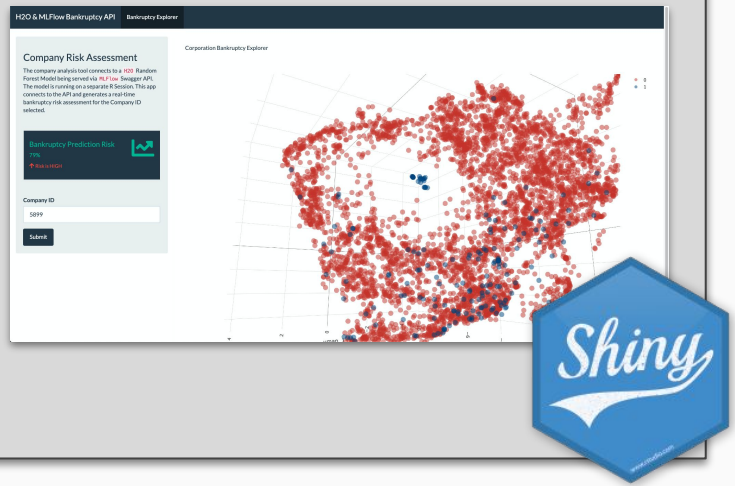


rmarkdown

Solution



Your Machine

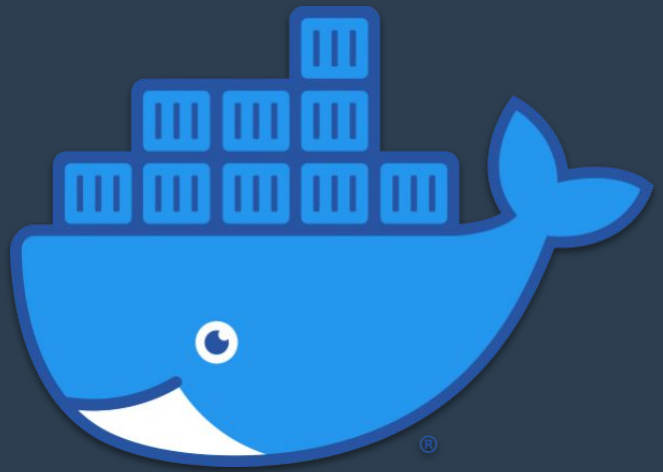


Their Server

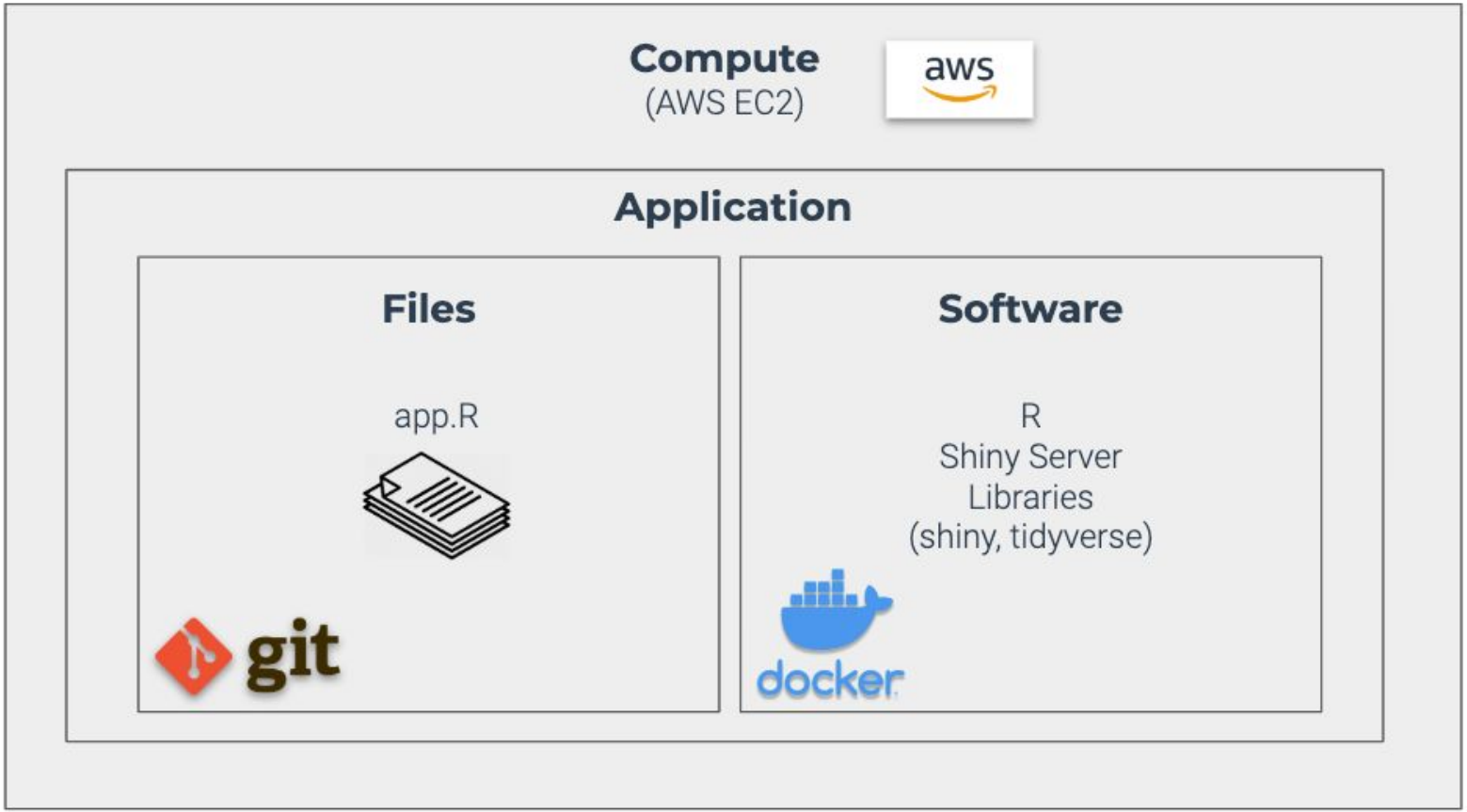
What about Applications?

What is Docker

80/20 Concepts



What is Docker?



What is Docker?

3 Key Concepts

1. Image & Dockerfile
2. Docker Hub
3. Containers

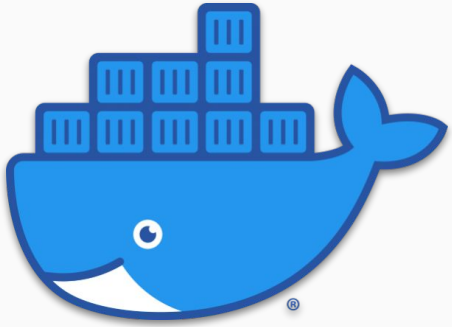


Image & Dockerfile

docker image build

00_lab_40_setup.R* | 01_docker_cli_scripts | Dockerfile | bankruptcy_report.Rmd*

```
1  FROM rocker/verse:latest
2
3  ## Install debian packages
4  RUN apt-get update -qq && apt-get -y --no-install-recommends install \
5
6      # usethis dependencies
7      build-essential \
8      libcurl4-gnutls-dev \
9      libxml2-dev \
10     libssl-dev \
11
12     # sf dependencies
13     lbzip2 \
14     libfftw3-dev \
15     libgdal-dev \
16     libgeos-dev \
17     libgsl0-dev \
18     libgl1-mesa-dev \
19     libglu1-mesa-dev \
20     libhdf4-alt-dev \
```

```
(base) matthews-mbp-2:lab_40_docker mdancho$ docker image ls
REPOSITORY                TAG                 IMAGE ID            CREATED             SIZE
mdancho/h2o-verse         3.30.0.1            65d265478042        5 hours ago         4.7GB
h2o-verse                 3.30.0.1            52a24e2dbbc4        16 hours ago        3.75GB
h2overse                  3.30.0.1            52a24e2dbbc4        16 hours ago        3.75GB
mdancho/h2o-verse         <none>              52a24e2dbbc4        16 hours ago        3.75GB
<none>                    <none>              0f3cf8c4bb04        17 hours ago        2.5GB
<none>                    <none>              1eedd05b7fe9        18 hours ago        3.17GB
<none>                    <none>              9bbbb0857a5c        18 hours ago        3.17GB
<none>                    <none>              406952c8aee4        19 hours ago        2.59GB
bsu-shiny-models          latest              2103a3a85290        5 months ago        4.59GB
mdancho/bsu-shiny-models  latest              2103a3a85290        5 months ago        4.59GB
```

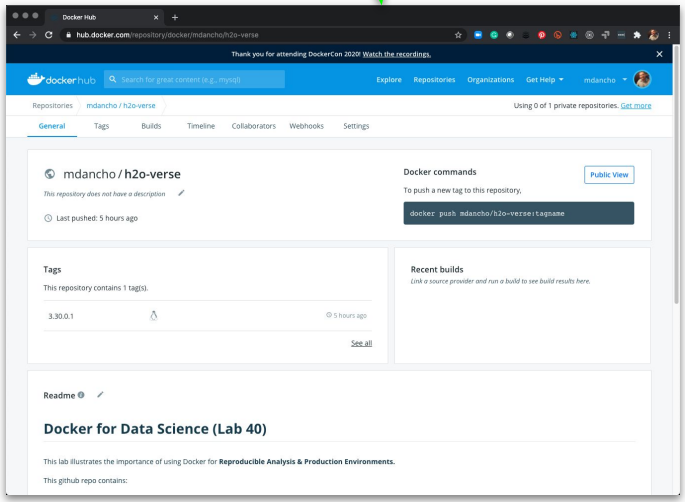
h2o-verse has a 1.5 hour build time.

Docker Hub

h2o-verse has a 1.5 hour build time.

```
(base) matthews-mbp-2:lab_40_docker mdancho$ docker image ls
REPOSITORY                  TAG                 IMAGE ID            CREATED             SIZE
mdancho/h2o-verse           3.30.0.1            65d265478042        5 hours ago         4.7GB
h2o-verse                   3.30.0.1            52a24e2dbbc4        16 hours ago        3.75GB
h2overse                    3.30.0.1            52a24e2dbbc4        16 hours ago        3.75GB
mdancho/h2o-verse           <none>              52a24e2dbbc4        16 hours ago        3.75GB
<none>                      <none>              0f3cf8c4bb04        17 hours ago        2.5GB
<none>                      <none>              1eedd05b7fe9        18 hours ago        3.17GB
<none>                      <none>              9bbbb0857a5c        18 hours ago        3.17GB
<none>                      <none>              [illegible]aee4     [illegible] hours ago   2.5[illegible]
bsu-shiny-models            latest              [illegible]5290     [illegible]hs ago
mdancho/bsu-shiny-models    latest              [illegible]         [illegible]go
```

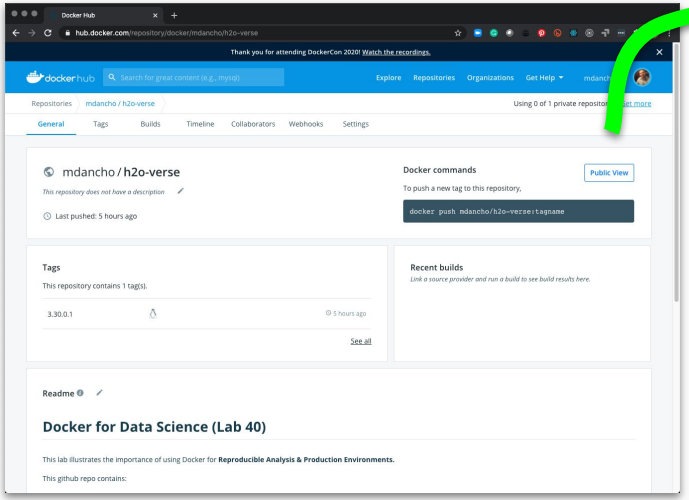
docker push image

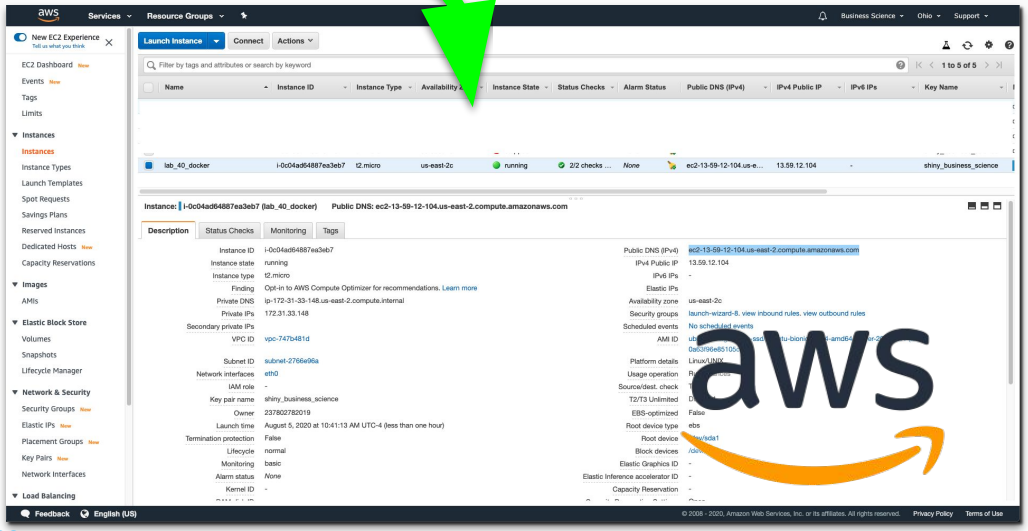


Source: https://hub.docker.com/repository/docker/mdancho/h2o-verse

Docker Image

```
docker pull mdancho/h2o-verse:3.30.0.1
```

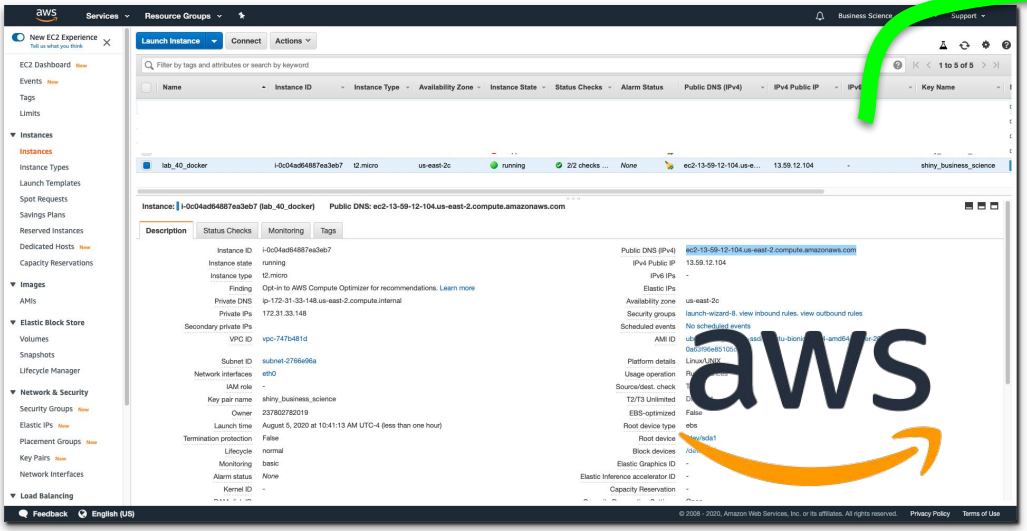


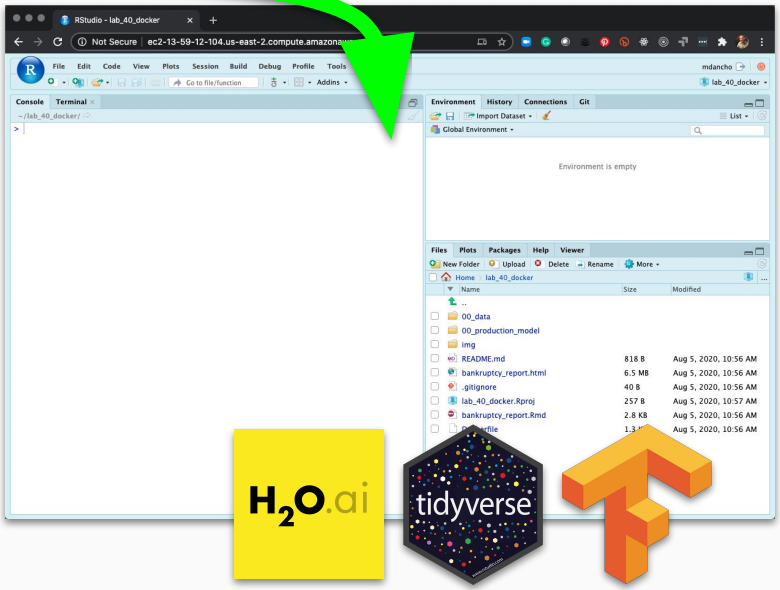


Source: https://hub.docker.com/repository/docker/mdancho/h2o-verse

Docker Container

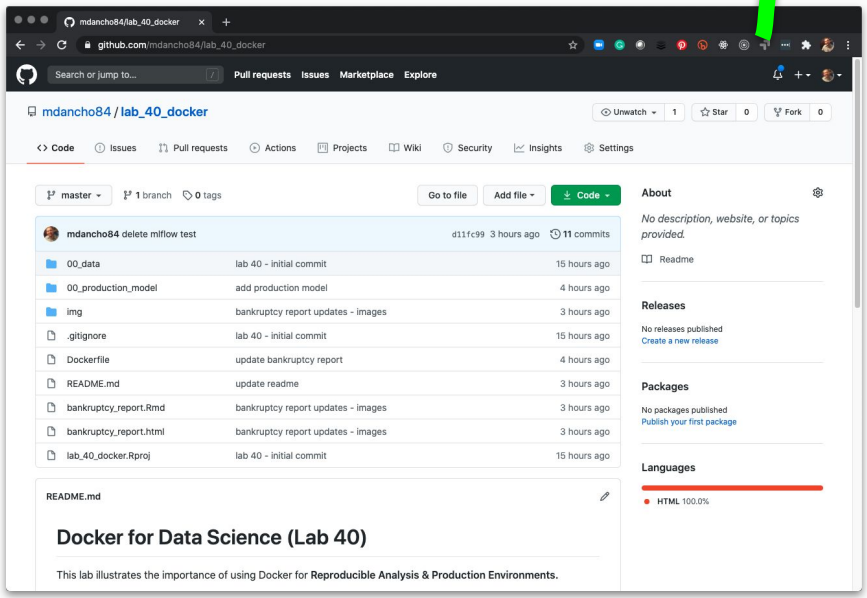
```
docker container run
```

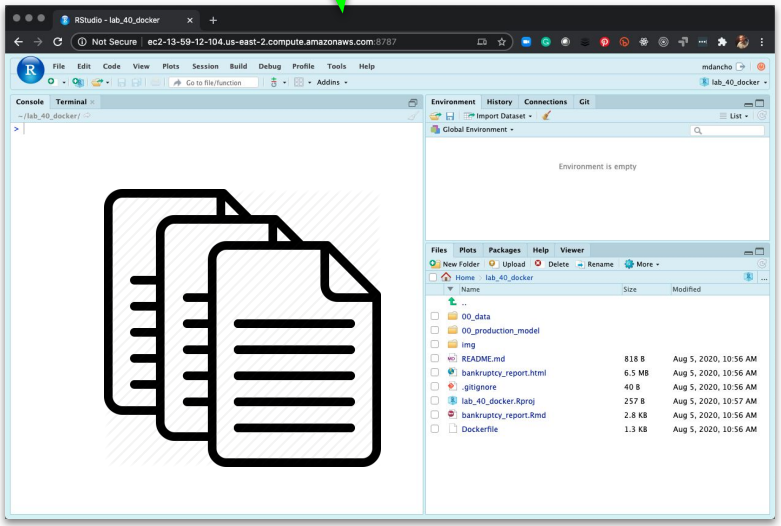




Source: https://hub.docker.com/repository/docker/mdancho/h2o-verse

GitHub





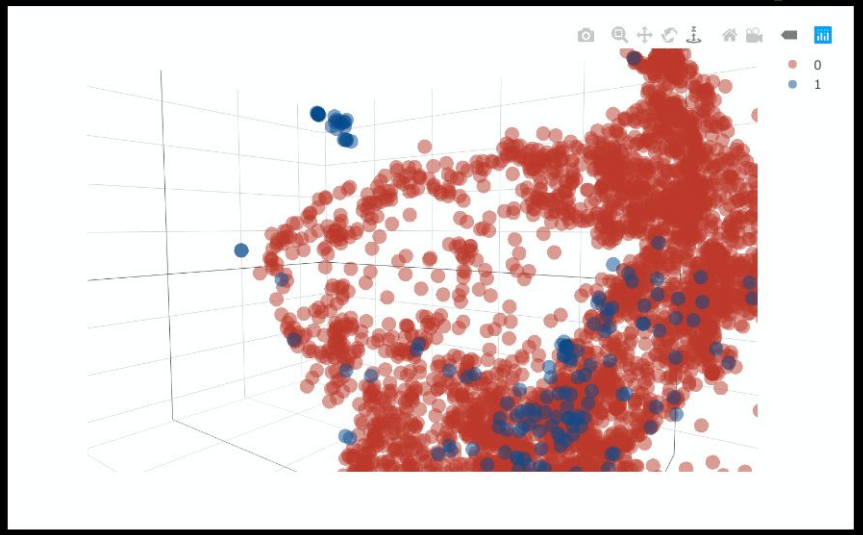
Source: https://github.com/mdancho84/lab_40_docker

30-Min Demo

lab_40_docker - master - RStudio

00_lab_40_setup.R | Dockerfile | 01_docker_cli_scripts | README.md | bankruptcy_report.Rmd

```
    add_markers(opacity = 0.5)
```



Reproducible Bankruptcy Cl...
Libraries
Data
Bankruptcy Classification F...
Bankruptcy Classification F...
H2O Prediction Analysis
Load Production Model
Make Predictions
Bankruptcy UMAP Visualiza...
Apply UMAP
Plotly Visualization

5:11 | Lab 40: Docker for Data Science | R Markdown

Console | Terminal | R Markdown | Jobs

~/Desktop/learning_labs/lab_40_docker/

```
> # Plotly Visualization ----
> plot_data_tbl %>%
+     plot_ly(x = ~ umap_1, y = ~ umap_2, z = ~ umap_3,
+             color = ~ class, colors = c('#BF382A', '#0C4B8E'),
+             hovertemplate = ~ tooltip) %>%
+     add_markers(opacity = 0.5)
>
```

Environment | History | Connections | Git | Tutorial

Files | Plots | Packages | Help | Viewer

Publish

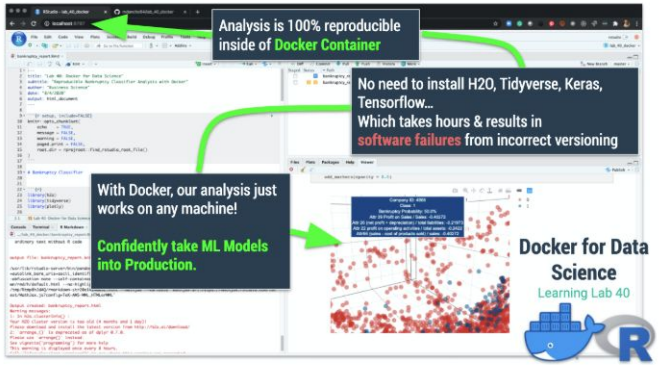
Lab 40: Docker for Data Science

Reproducible Bankruptcy Classifier Analysis with Docker

Business Science

8/5/2020

Reproducible Bankruptcy Classifier Analysis



Libraries

```
library(h2o)
library(tidyverse)
library(plotly)

# Not in container
```

Your transformation

- Build & deploy apps with AWS & Docker

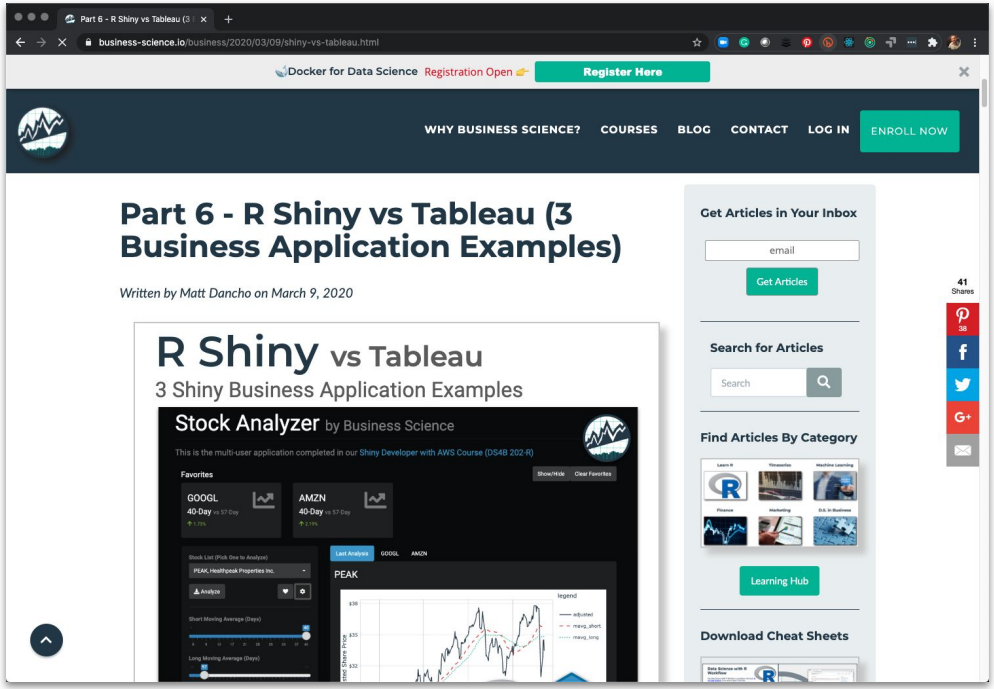
Shiny vs Tableau

5-Year Trend going on.

- Docker +4000%
- AWS +400%

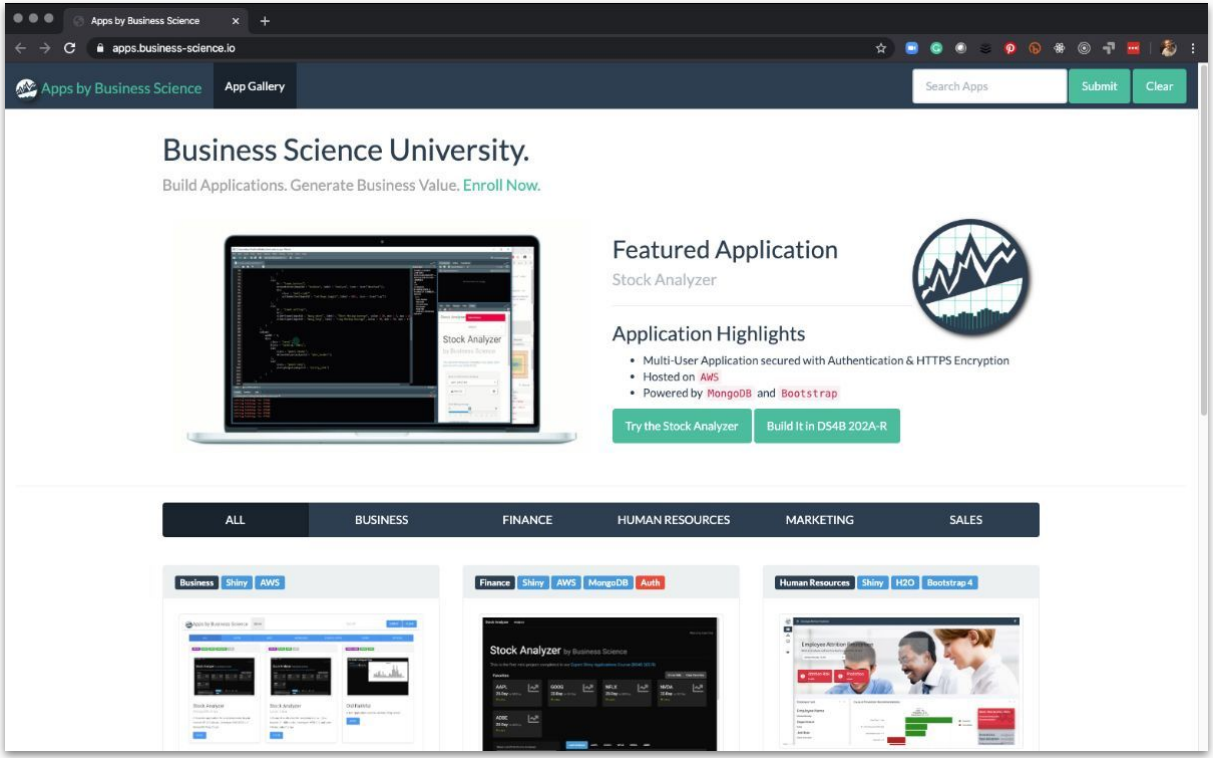
3 Business Examples

1. Application Library
2. Full Stack Web App
3. ML-Powered Dashboard



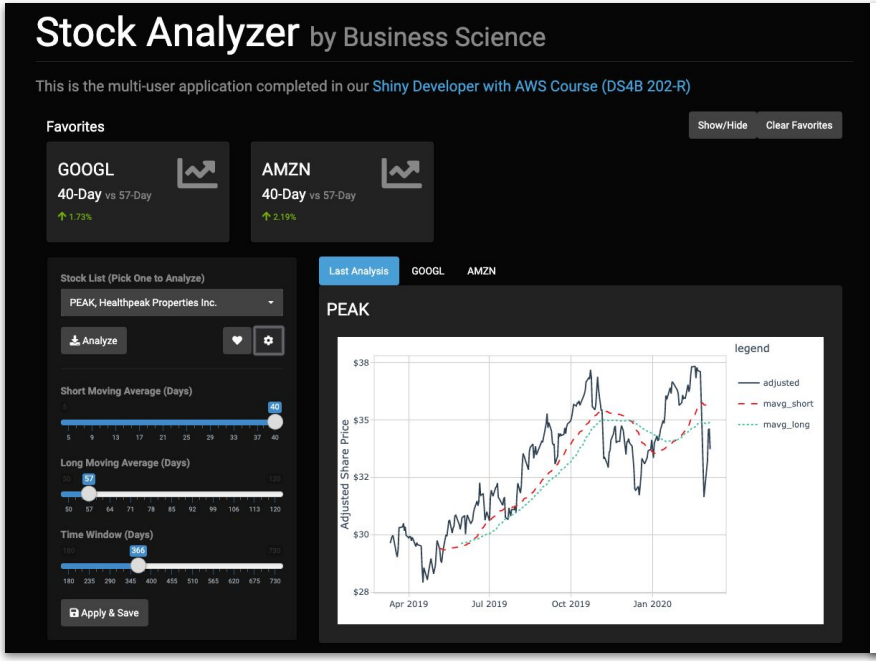
Souce: https://www.business-science.io/business/2020/03/09/shiny-vs-tableau.html

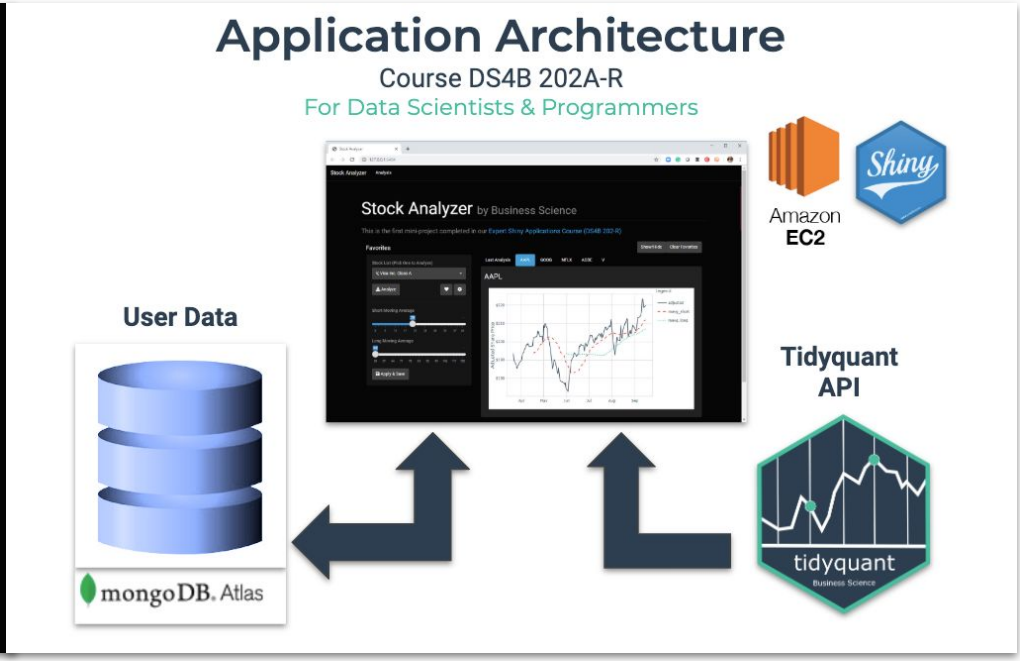
1. Application Library



Application Library

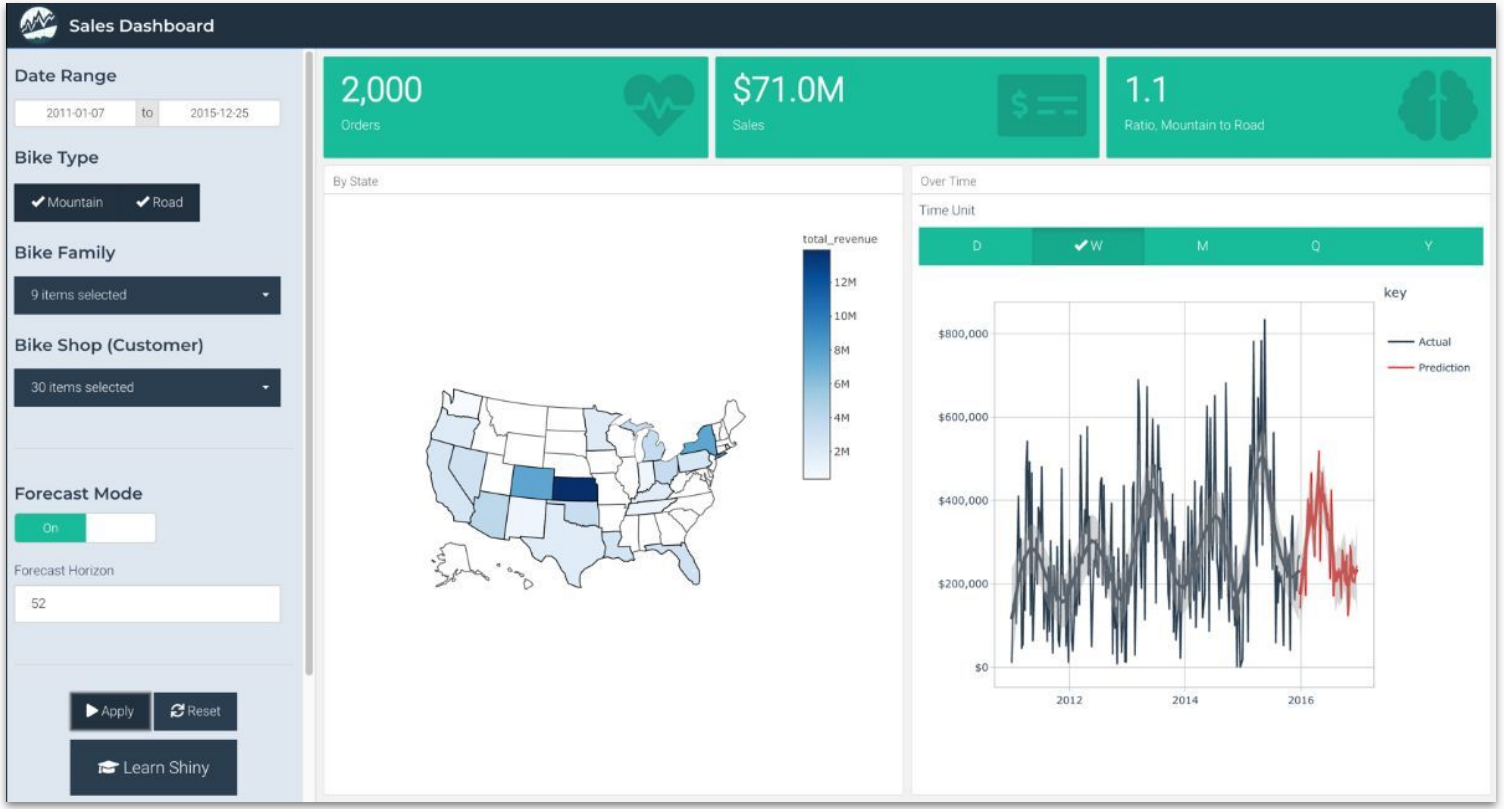
2. Full-Stack Web App





[Stock Analyzer]

3. ML-Powered Dashboard



[ML-Powered Dashboard]

4-Course R-Track System

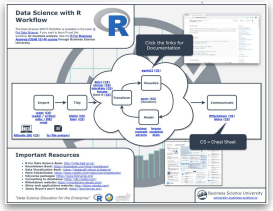
Business Analysis with R
(DS4B 101-R)

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

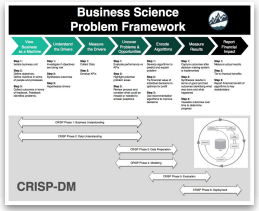
Web Apps & Shiny Developer
(DS4B 102-R + DS4B 202A-R)

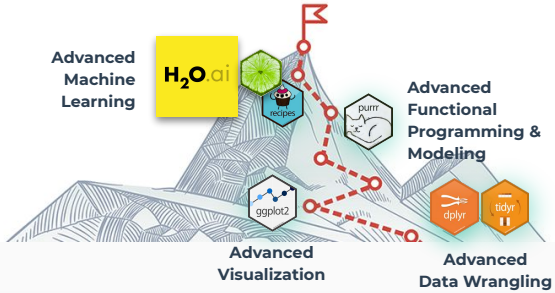
Project-Based Courses with Business Application

Data Science Foundations
7 Weeks

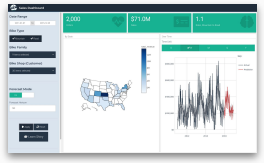


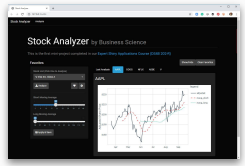
Machine Learning & Business Consulting
10 Weeks





Web Application Development
12 Weeks





Key Benefits

Frontend + Backend + Production Deployment

Frontend for Shiny
- Bootstrap

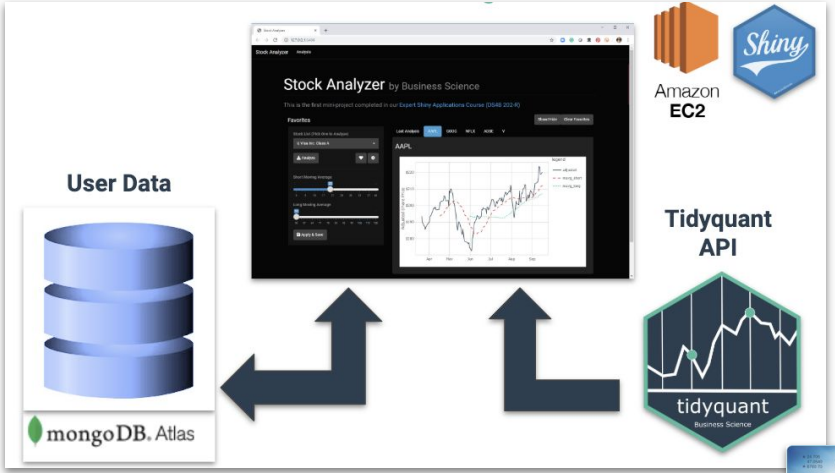
Backend for Shiny
- MongoDB Atlas Cloud
- Dynamic UI
- User Authentication
- Security

Production Deployment
- AWS
- EC2 Server
- SSL & HTTPS Encryption

Shiny Apps for Business (DS4B 202A-R)

Web Application Development

6 Weeks



aws

docker git Bootstrap mongoDB.

15% OFF PROMO Code: learninglabs

R-TRACK BUNDLE

4-Course Bundle - Machine Learning + Expert Web Applications (R-Track)

Go from Beginner to Expert Data Scientist & Shiny Developer in Under 6-Months

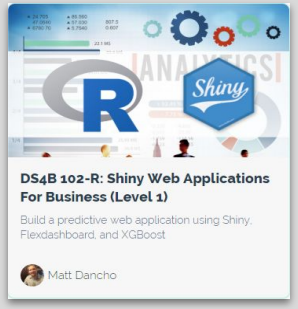
4 Course Bundle ~~$1,596~~

$127/mo
Limited Time

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



DS4B 102-R: Shiny Web Applications For Business (Level 1)

Build a predictive web application using Shiny, Flexdashboard, and XGBoost

Matt Dancho

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

DS4B 202A-R: Expert Shiny Developer with AWS

Learn how to build Scalable Data Science Applications using R, Shiny, and AWS Cloud Technology.

Matt Dancho

○	**Paid Course** 15% COUPON DISCOUNT	~~$1,596~~ **$1,356.60**
◉	**12 Low Monthly Payments** 15% COUPON DISCOUNT 12X Payment Plan	~~12 payments of $149/m~~ **12 payments of $126.65/m**

Career acceleration awaits

university.business-science.io