

Business Science Problem Framework



View
Business
as a Machine

Understand
the Drivers

Measure
the Drivers

Uncover
Problems &
Opportunities

Encode
Algorithms

Measure
Results

Report
Financial
Impact

Step 1: Isolate business unit being met
Step 2: Define objectives. Define machine in terms of people and processes
Step 3: Collect outcomes in terms of feedback. Feedback identifies problems.

Step 1: Investigate if objectives are being met
Step 2: Synthesize outcomes
Step 3: Hypothesize drivers

Step 1: Collect Data
Step 2: Develop KPIs

Step 1: Evaluate performance vs KPIs
Step 2: Highlight potential problem areas
Step 3: Review process and consider what could be missed or needed to answer questions

Step 1: Develop algorithms to predict and explain problem
Step 2: Tie financial value of individual decisions to optimize for profit
Step 3: Use recommendation algorithms to improve decisions

Step 1: Capture outcomes after decision making system is implemented
Step 2: Tie to financial benefits
Step 3: Report financial benefit of algorithms to key stakeholders

Step 1: Measure actual results.
Step 2: Tie to financial benefits
Step 3: Report financial benefit of algorithms to key stakeholders

Important: Show Cost of the Problem

CRISP Phase 1: Business Understanding

CRISP Phase 2: Data Understanding

CRISP Phase 3: Data Preparation

CRISP Phase 4: Modeling

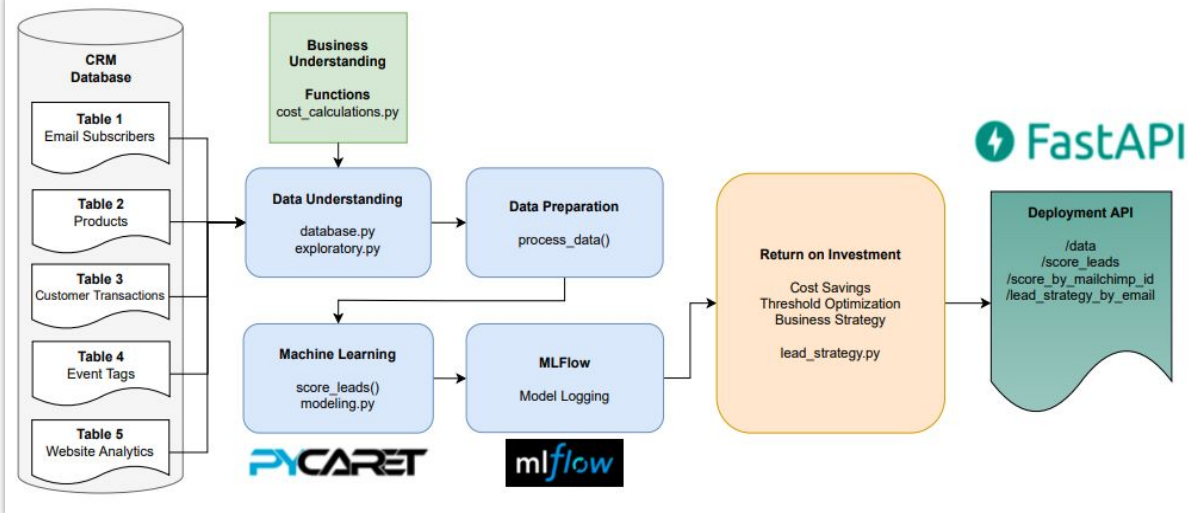
CRISP Phase 5: Evaluation

CRISP Phase 6: Deployment

CRISP-DM

Learn to apply the BSFP:
Data Science For Business With R Course (DS4B 201-R)
Machine Learning APIs with Python Course (DS4B 201-P)

Version 2.0



Student Orientation

Machine Learning API's with Python Course (DS4B 201-P)

Matt Dancho

Founder, Business Science





Agenda

1. Why learn from me?
2. My first BIG Project
3. How are my students succeeding with the BSPF? (Interview Secret)
4. Solving the Course Project with the BSPF
5. Course Workflow, Software, & Prerequisites
6. Let's go!



Business Science

Data Science Courses in R & Python
for Business

Why Learn from Me?





My CV

I love to **Train Data Scientists**





My CV

I love to **Build Software**

README.md

tidyquant

build passing codecov 45% CRAN 1.0.3 downloads 18K/month downloads 682K

Bringing financial and business analysis to the tidyverse

2-Minutes To Tidyquant

Our short introduction to `tidyquant` on YouTube.

```
graph LR; subgraph LeftBox [quantmod/xts packages]; L1[quantmod]; L2[xts]; L3[zoo]; L4[PerformanceAnalytics]; L5[TTR]; end; subgraph RightBox [tidyverse packages]; R1[dplyr]; R2[tidyr]; R3[ggplot2]; R4[purrr]; R5[and more!]; end; LeftBox <--> TQ[tidyquant]; TQ <--> RightBox;
```



My CV

I love to **train data scientists**.

This love has opened up doors to **consulting & corporate training events**.

They learned:

- **Same Framework**
- **Same Workflow**
- **Completely different Problem**



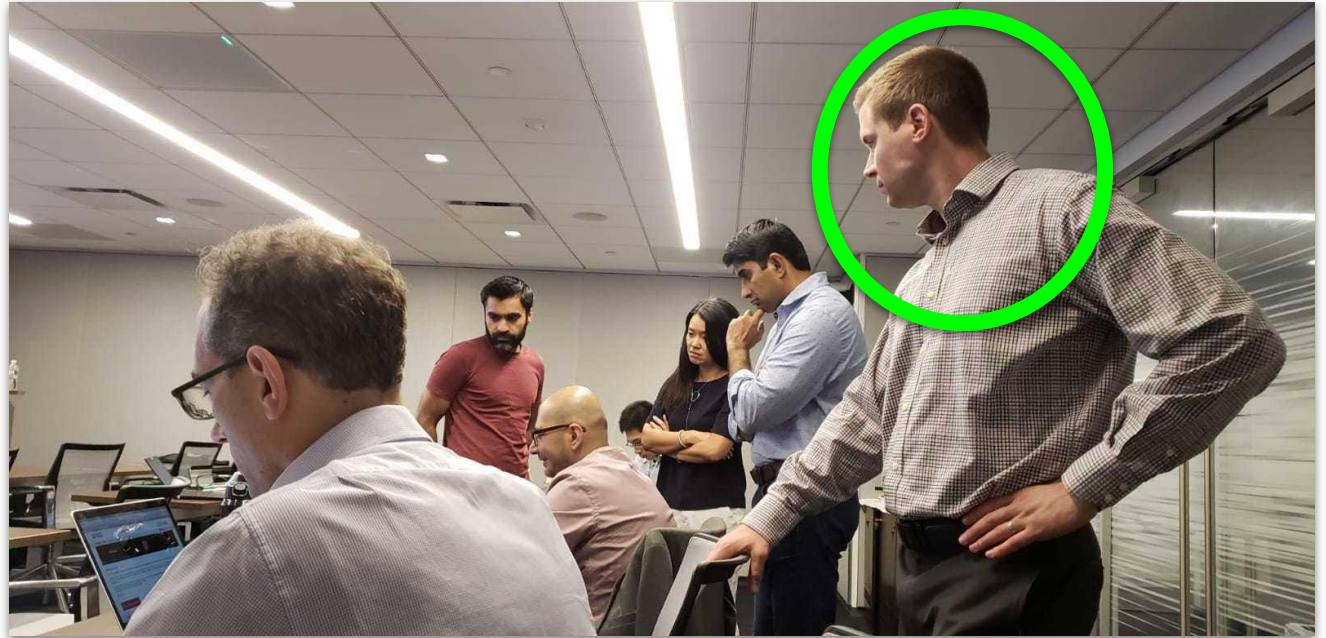
S&P Global



My CV

What you're learning
is what companies
paid **\$25,000 to
\$50,000.**

My methodologies
to solve business
problems with data
science.



My First BIG Project

Origin of the BSPF



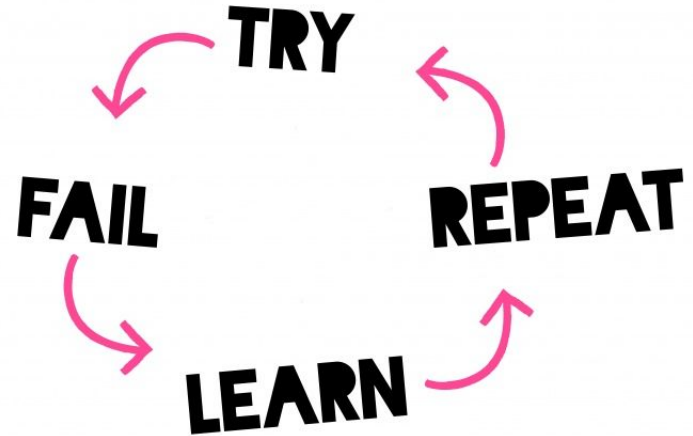


Failure comes before success

It's OK to Fail.

We all fail.

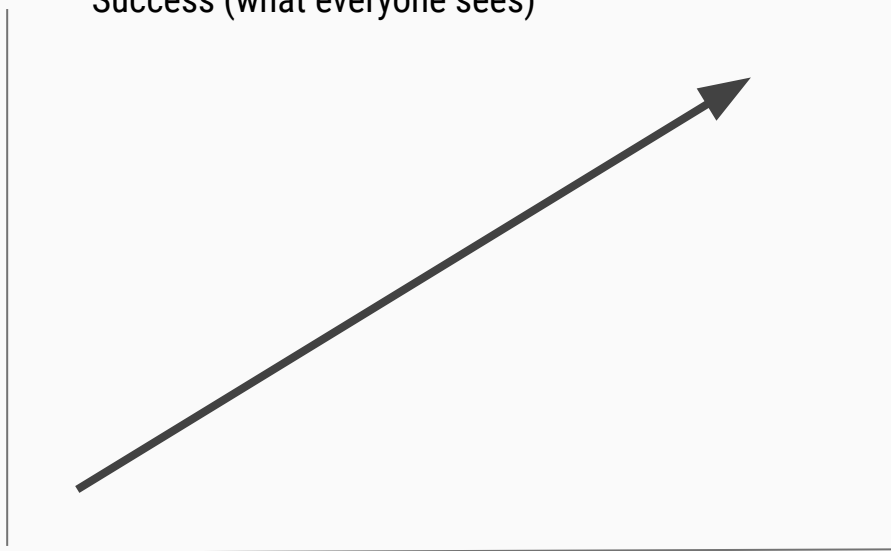
Those that are successful learn
from failure.





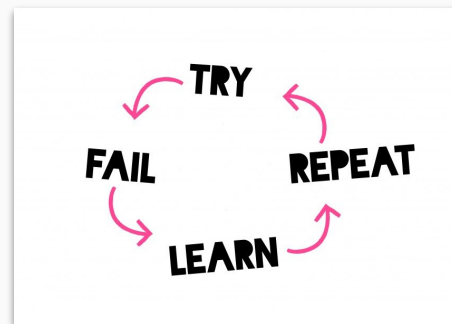
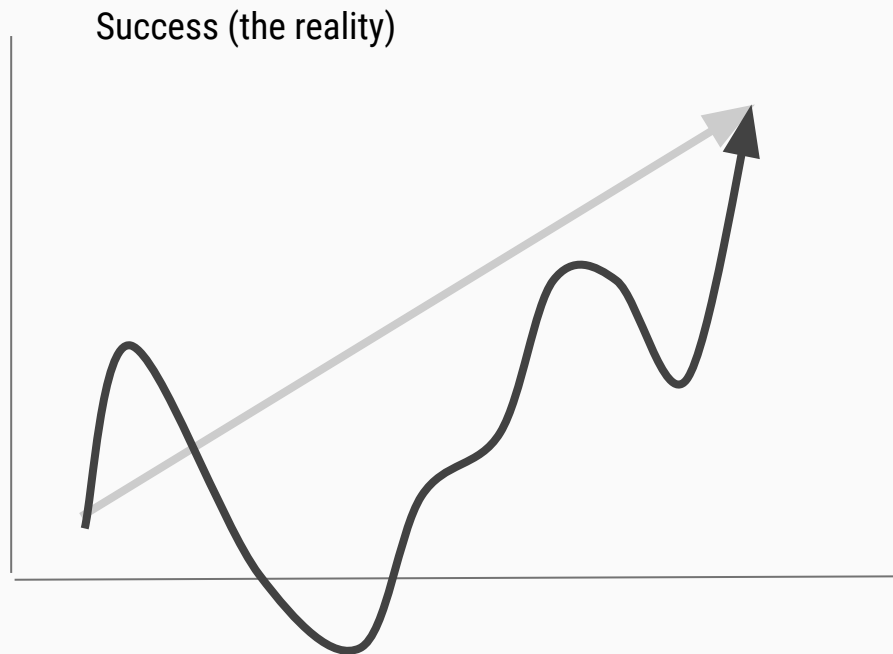
What **people think** success looks like

Success (what everyone sees)





What **real success** looks like



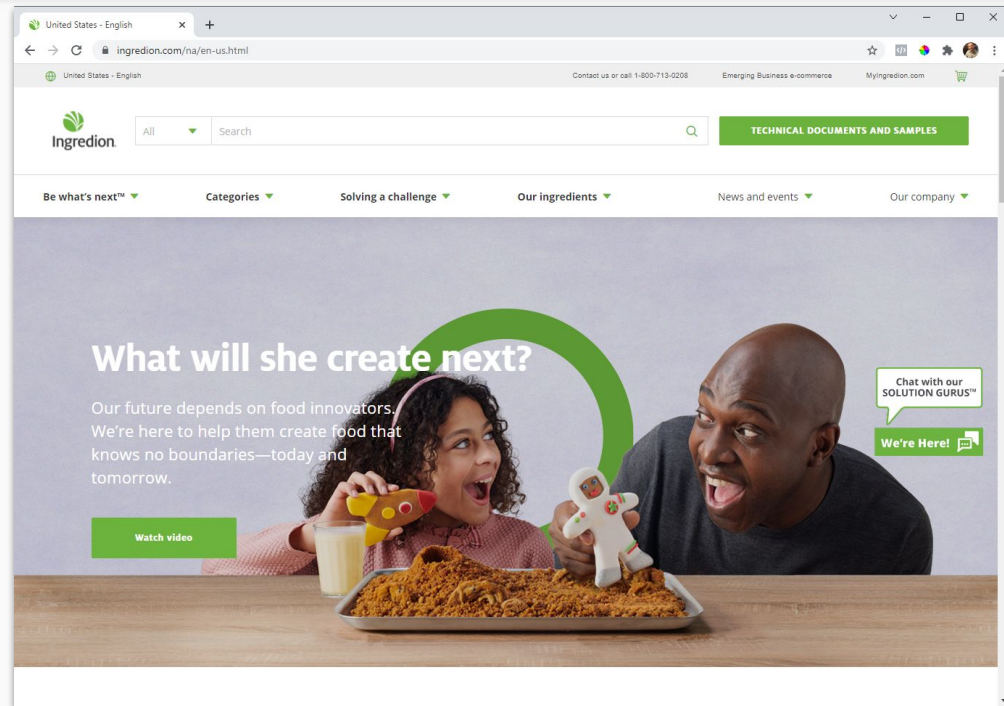


My first **BIG Opportunity**

Big Opportunity: Ingredion

\$6B in Revenue (2021)

Fortune 500 (#463 in 2021)



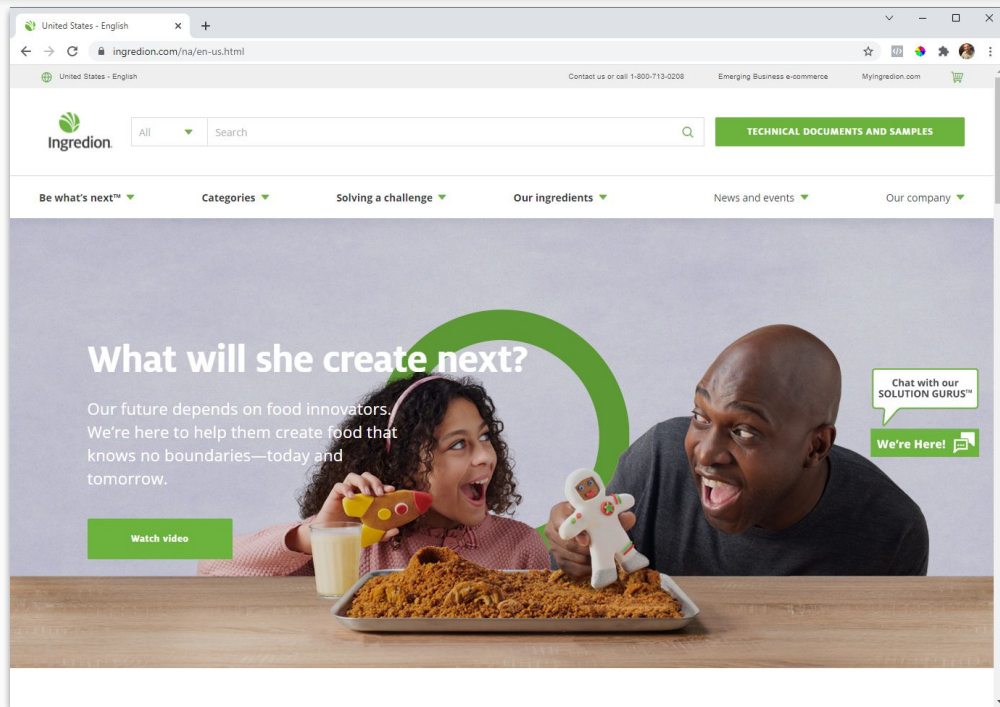


My first **BIG** Project

Human Resources Project

Goal: Apply predictive analytics

1. Got the Data on **Friday**.
2. Present a report on **Monday**.
3. Executives were to present to Board on **Wednesday**.





The Monday Report

We made an amazing report
(we thought)!

After presenting the report, I
heard... Silence.

Ingredion: Capability Review

Modeling Executive Potential

Business Science

September 25, 2017

Executive Summary

The objective is to identify the executive potential of candidates that are ready for the next level. We use





Lessons Learned

1. I **rushed** the process
2. I never understood clearly what the client **needed**
3. The client was **not involved**
4. We collectively never understood if there would be a clear **Return ON Investment**





Reviewing the Failure

Businesses **didn't need** predictive analytics for the sake of analytics.

They **needed a process** for converting data problems to business value.



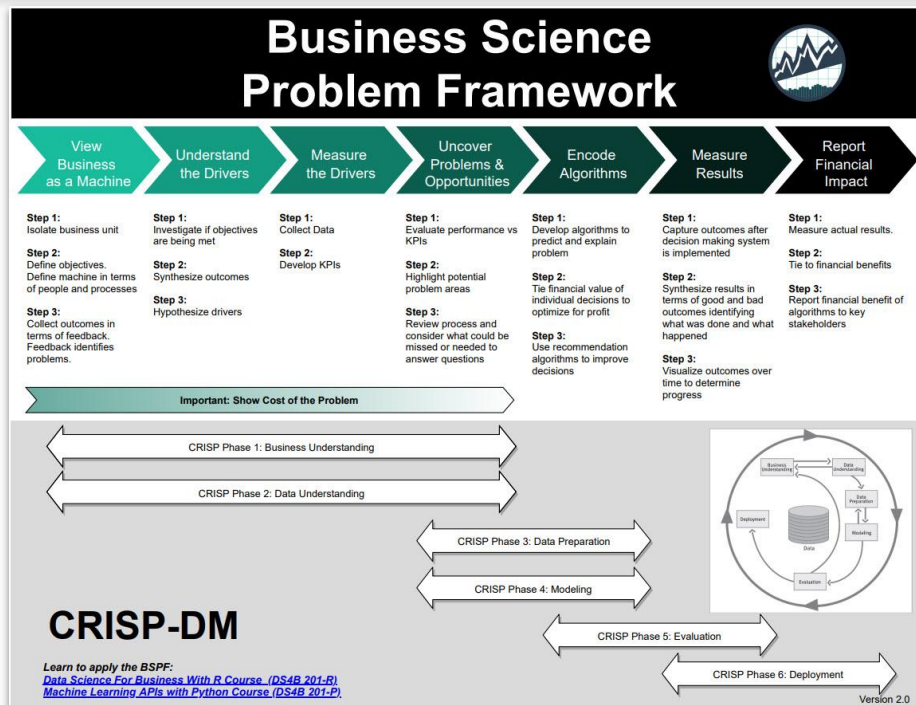


The BSPF was born

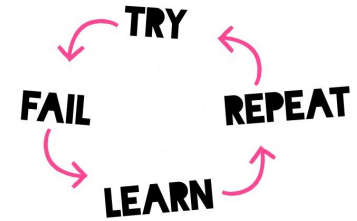
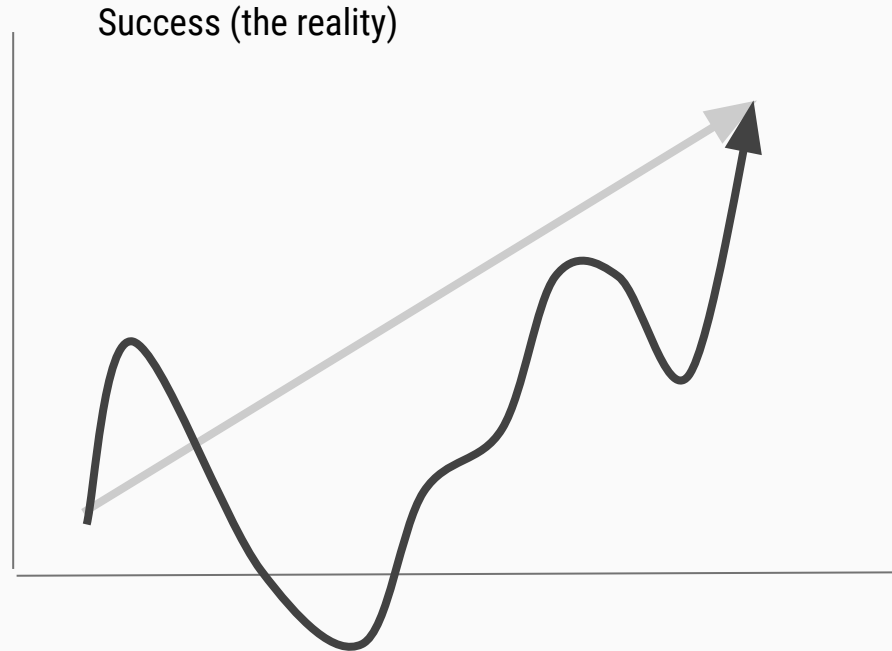
A problem solving framework.

Key Improvements:

1. Makes sure the **team is involved**.
2. Focus is on **cost of the problem** and **Return On Investment** of the solution.



What **happened** as a result of the BSPF



How are my students are succeeding with the BPSF?

The Interview Secret



Business Science Problem Framework



Step 1:
Isolate business unit

Step 2:
Define objectives.
Define machine in terms of people and processes

Step 3:
Collect outcomes in terms of feedback.
Feedback identifies problems.

Step 1:
Investigate if objectives are being met

Step 2:
Synthesize outcomes

Step 3:
Hypothesize drivers

Step 1:
Collect Data

Step 2:
Develop KPIs

Step 1:
Evaluate performance vs KPIs

Step 2:
Highlight potential problem areas

Step 3:
Review process and consider what could be missed or needed to answer questions

Step 1:
Develop algorithms to predict and explain problem

Step 2:
Tie financial value of individual decisions to optimize for profit

Step 3:
Use recommendation algorithms to improve decisions

Step 1:
Capture outcomes after decision making system is implemented

Step 2:
Synthesize results in terms of good and bad outcomes identifying what was done and what happened

Step 3:
Visualize outcomes over time to determine progress

Step 1:
Measure actual results.

Step 2:
Tie to financial benefits

Step 3:
Report financial benefit of algorithms to key stakeholders

Important: Show Cost of the Problem

CRISP Phase 1: Business Understanding

CRISP Phase 2: Data Understanding

CRISP Phase 3: Data Preparation

CRISP Phase 4: Modeling

CRISP Phase 5: Evaluation

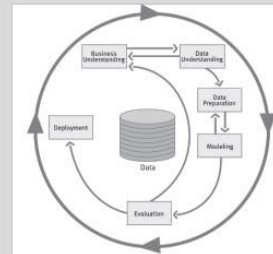
CRISP Phase 6: Deployment

CRISP-DM

Learn to apply the BSPF:

[Data Science For Business With R Course \(DS4B 201-R\)](#)

[Machine Learning APIs with Python Course \(DS4B 201-P\)](#)



Version 2.0

The Interview Secret.

Proof #1: Jennifer used the BSPF to land her dream job at JP Morgan



Thanks to Matt and what I've learned so far, *I was able to do an in-depth analysis of Consumer Financial Protection Bureau (CFPB) data, following his Business Science Process Framework and complete the project using RMarkdown. The polished, finished product impressed the hiring manager so much, he was willing to fast-track an offer.*" -Jennifer Cooper, VP Strategic Analytics

Jennifer Cooper has started her new role as a VP of Strategic Analytics at a major bank.

J.P.Morgan

Proof #2: Masatake used the BSPF in the verbal interview for Accenture





Masatake Hirono
Data Scientist at BCG GAMMA
[View full profile](#)

Masatake Hirono • 1st
Data Scientist at BCG GAMMA
2yr • 

After struggling to balance with my work for many months, I've finally completed the Business Science University DS4B 201-R: Data Science For Business With R, taught by [Matt Dancho](#). Unlike other MOOCs, this course showed me how to place data analytics in real business settings. **Without this course, I would have never attempted to pay attention to business/financial impacts, generated through my analysis. His instruction turned me a more advanced data scientist and helped me find a new career opportunity.** I will start to work at one of the most prestigious management consulting firms in October as a cognitive & analytics consultant. Highly recommended if you would like to use R as a professional business person!


accenture



Proof #2: Masatake used the BSPF in the verbal interview for Accenture




accenture



Solving the Course Project

Email Lead Scoring

(What you are solving with BSPF in this course)



Business Science Problem Framework



Step 1:
Isolate business unit

Step 2:
Define objectives.
Define machine in terms
of people and processes

Step 3:
Collect outcomes in
terms of feedback.
Feedback identifies
problems.

Step 1:
Investigate if objectives
are being met

Step 2:
Synthesize outcomes

Step 3:
Hypothesize drivers

Step 1:
Collect Data

Step 2:
Develop KPIs

Step 1:
Evaluate performance vs
KPIs

Step 2:
Highlight potential
problem areas

Step 3:
Review process and
consider what could be
missed or needed to
answer questions

Step 1:
Develop algorithms to
predict and explain
problem

Step 2:
Tie financial value of
individual decisions to
optimize for profit

Step 3:
Use recommendation
algorithms to improve
decisions

Step 1:
Capture outcomes after
decision making system
is implemented

Step 2:
Synthesize results in
terms of good and bad
outcomes identifying
what was done and what
happened

Step 3:
Visualize outcomes over
time to determine
progress

Step 1:
Measure actual results.

Step 2:
Tie to financial benefits

Step 3:
Report financial benefit of
algorithms to key
stakeholders

Important: Show Cost of the Problem

CRISP Phase 1: Business Understanding

CRISP Phase 2: Data Understanding

CRISP Phase 3: Data Preparation

CRISP Phase 4: Modeling

CRISP Phase 5: Evaluation

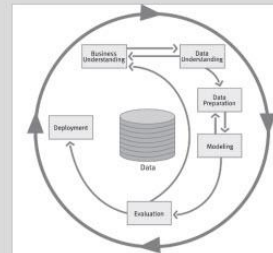
CRISP Phase 6: Deployment

CRISP-DM

Learn to apply the BSPF:

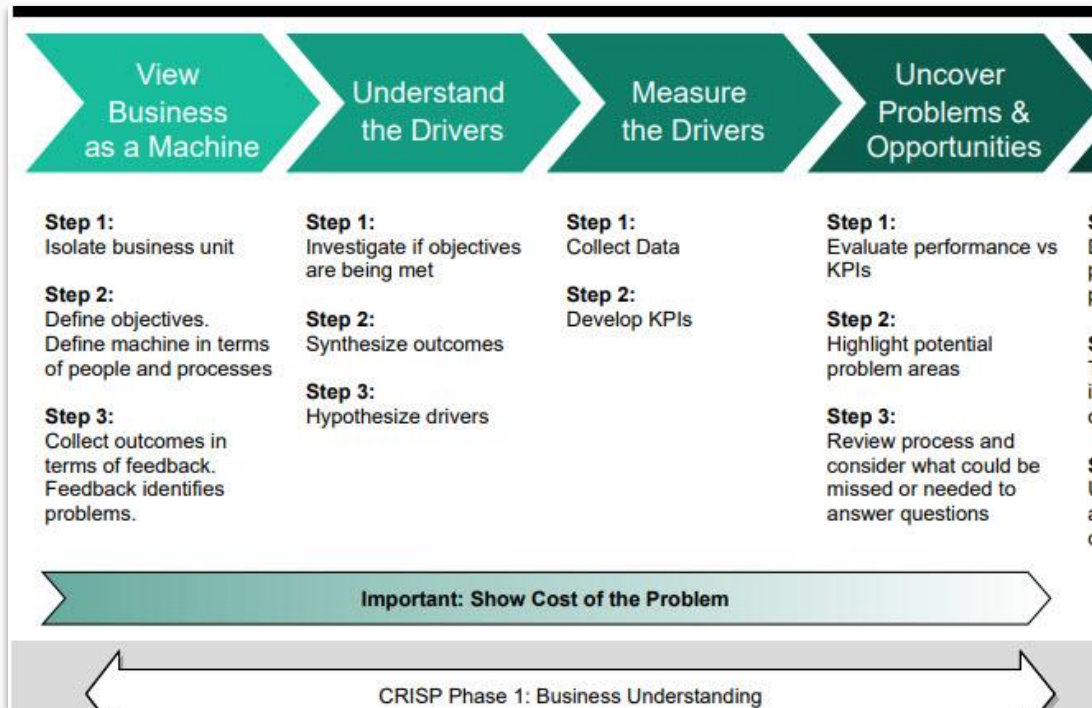
[Data Science For Business With R Course \(DS4B 201-R\)](#)

[Machine Learning APIs with Python Course \(DS4B 201-P\)](#)



Version 2.0

Can be used to
solve **any**
business problem.



Goal: To understand if improvement is needed.

Stages help:

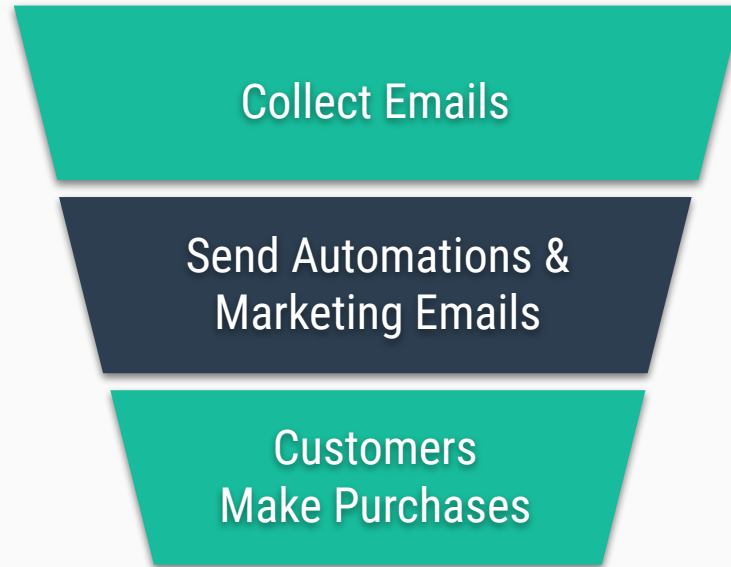
- **Define** the Objectives
- **Bring together** the team
- **Understand** the business process
- **Show** the cost
- **Develop** KPIs

Stage 1: View Business as a Machine



Business Process

Who is involved? **Marketing**
What does the process look like? 📌



1

View
Business
as a Machine

Step 1:
Isolate business unit

Step 2:
Define objectives.
Define machine in terms
of people and processes

Step 3:
Collect outcomes in
terms of feedback.
Feedback identifies
problems.

Problem: Revenue is slowing.

Objective: Make Purchase

Process: Companies send 2 Types
of Emails:

1. "Value Driven" (Nurture)
2. "Sales Driven" (Make Purchase)

Outcome: High Unsubscribe Rates

\$3,000,000

Per Year

2

Understand the Drivers

Step 1:
Investigate if objectives
are being met

Step 2:
Synthesize outcomes

Step 3:
Hypothesize drivers

Important: Show Cost of the Problem

Cost of losing unsubscribers due to email blasts.

Send 5 sales blasts per month. Lose 500 subscribers per email blast.

5% customer conversion rate & customer lifetime value of \$2000.

3

Measure the Drivers

Step 1:
Collect Data

Step 2:
Develop KPIs

Step 3:
Evaluate KPIs

Step 4:
High probability

Collect data with Marketing Team.

Together answer, **“What does a recent customer look like?”**

[Back to Business Science - Combined List](#)



Added via Landing Page: [FINAL R-Track Squeeze Page - Business Science Comb](#) on September 3rd, 2020 at 12:36 PM ET

Opted in on November 17th at 7:59 PM ET

Email Marketing Engagement: Rarely ⓘ



Subscribed

Actions ▾

Tags

+

[rtrack_freewebi...](#) × [r-tips](#) × [webinar](#) ×
[time-series-web...](#) × [time-series-wai...](#) ×

56.8%

Open rate

15.9%

Click rate

—
Total revenue

—
Average order value

[Connect your store](#) to get order data.

Profile Information

[Edit](#)

Email Address

First Name

Last Name

Birthday

Status

URL

Company

Title

Other

[Edit](#)

Language

English

Location ⓘ

Oakland, CA, USA

Favorite email client

Gmail

Preferred email for...

html

Last Updated

Wed, Nov 17, 2021 7:59 pm

Write a Note

1000 characters remaining

Write an internal note here

[Add Note](#)

Activity Feed

All Activity ▾

November 22nd



Opened the email [Monday before Black Friday - 2021](#).

11:43 AM ET



Clicked a link in the email [Monday before Black Friday - 2021](#).

11:43 AM ET



Opened the email [Monday before Black Friday - 2021](#).

11:43 AM ET



Was Sent the email [Monday before Black Friday - 2021](#).

8:00 AM ET

3.3X

vs Industry Unsub Rate

3

Measure
the Drivers

Step 1:
Collect Data

Step 2:
Develop KPIs

Step
Evalu
KPIs

Step
High
prob

High Unsubscribe Rates

KPI (Industry Average): 0.0015

Unsubs per Sales Email of 500 & List size of 100,000 = 0.005

Stage 4: Uncover Problems & Opportunities



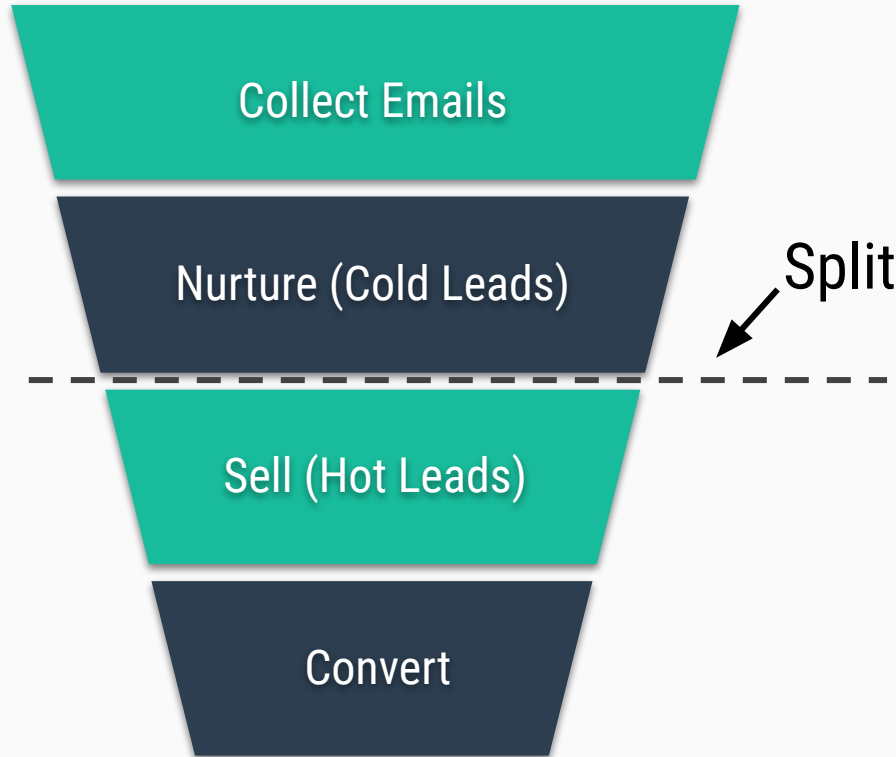
4

Uncover Problems & Opportunities

Step 1:
Evaluate performance vs KPIs

Step 2:
Highlight potential problem areas

Step 3:
Review process and consider what could be missed or needed to answer questions



High Unsub Rates:

- 3X Industry avg (KPI)
- \$3M Cost Per Year

Instead of Sending Blasts, **split the list:**

1. **Nurture “Cold Leads”:** Get them to take actions that make them more similar to customers.
2. **Sell “Hot Leads”:** Already have taken actions that customers do.

Leadership Goal: Keep Sales High (98%)

Stage 5: Step back and let the Data Science Team **Encode Algorithms!**



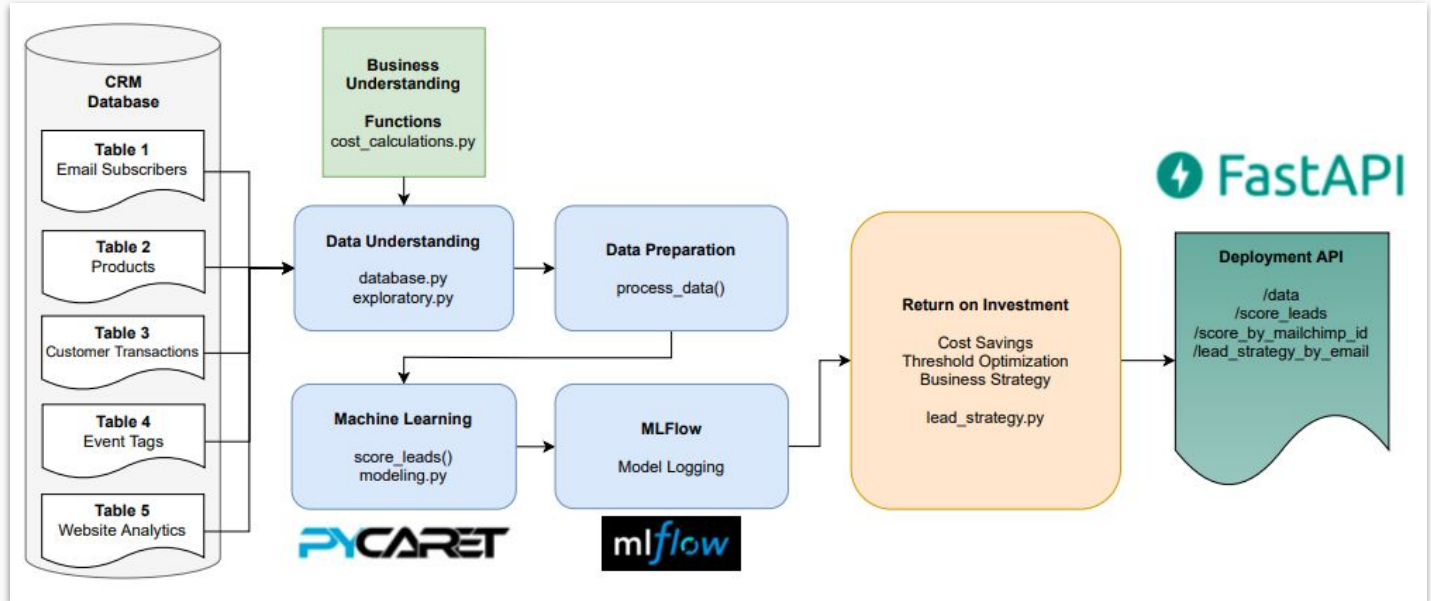
5

Encode Algorithms

Step 1:
Develop algorithms to predict and explain problem

Step 2:
Tie financial value of individual decisions to optimize for profit

Step 3:
Use recommendation algorithms to improve decisions



And we get a **Recommended Strategy** for Marketing.



5

Encode Algorithms

Step 1:

Develop algorithms to predict and explain problem

Step 2:

Tie financial value of individual decisions to optimize for profit

Step 3:

Use recommendation algorithms to improve decisions

```
import email_lead_scoring as els ...
```

```
leads_df = els.db_read_and_process_els_data() ...
```

```
... Transformation Pipeline and Model Successfully Loaded
```

```
Saved Customers Per Month: 12.124855665445061
```

```
Savings Per Month: 24249.71133089012
```

```
Cost Per Month: 5197.505197505197
```

```
Expected Value Per Month: 19052.206133384923
```

```
----
```

```
optimization_results['lead_strategy_df'] ...
```

	user_email	Score	made_purchase	rank	gain	category
19675	lilburn.beattyjones@gmail.com	0.4998	1	1	0.001040	Hot-Lead
6482	dr.cassie.turcotte.sr@yahoo.com	0.4994	1	2	0.002079	Hot-Lead
4043	gustavo.nolan@gmail.com	0.4985	1	3	0.003119	Hot-Lead
9002	porter.hermann@gmail.com	0.4979	1	4	0.004158	Hot-Lead
18342	zadie.kirlin@gmail.com	0.4978	1	5	0.005198	Hot-Lead
...
13403	treva.ortiz.dds@gmail.com	0.0273	0	19915	1.000000	Cold-Lead
10121	wanita.wyman@gmail.com	0.0273	0	19916	1.000000	Cold-Lead
10696	lyn.muller@gmail.com	0.0273	0	19917	1.000000	Cold-Lead
10114	tennie.murphy.dvm@gmail.com	0.0273	0	19918	1.000000	Cold-Lead
6475	pasquale.fahey@gmail.com	0.0273	0	19919	1.000000	Cold-Lead

19919 rows x 6 columns

Email Lead Scoring Strategy

Recommended Strategy Buy-In:

Give the organization a way to save money in the years to come.



We can **decrease sales emails 28%** through targeting but still **achieve 98% of sales revenue**.

5

Encode Algorithms

Step 1:
Develop algorithms to predict and explain problem

Step 2:
Tie financial value of individual decisions to optimize for profit

Step 3:
Use recommendation algorithms to improve decisions

```
optimization_results['results_df'] ...
```

	count	sum_made_purchase
category		
Cold-Lead	5367	20
Hot-Lead	14552	942

```
942 / (942 + 20)
```

```
[18] ✓ 0.2s
```

```
... 0.9792099792099792
```

After nurturing takes effect, organization should then begin seeing **increased revenue** each month. (\$100,000/month)

```
optimization_results['expected_value'] ...
```

```
19052.206133384923
```

Stage 6: Measure Results



6

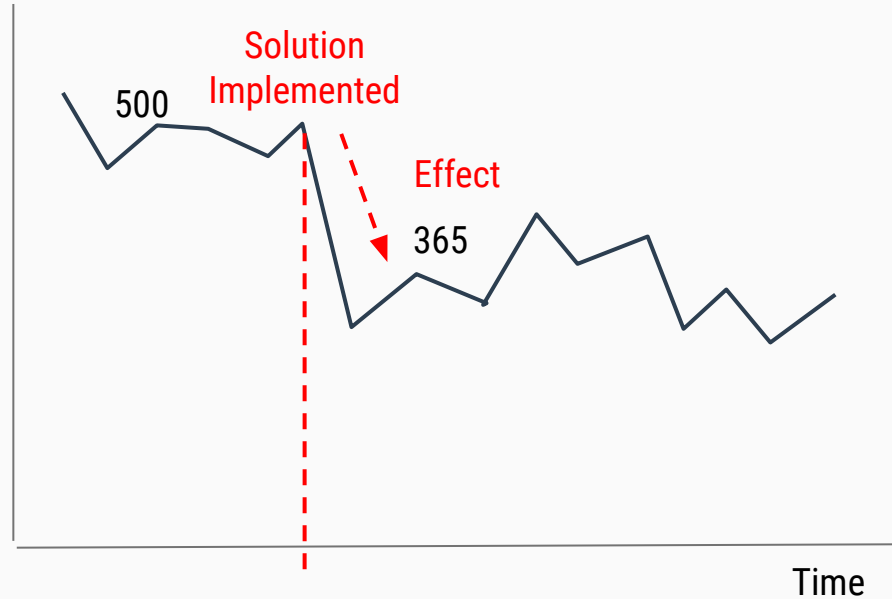
Measure Results

Step 1:
Capture outcomes after decision making system is implemented

Step 2:
Synthesize results in terms of good and bad outcomes identifying what was done and what happened

Step 3:
Visualize outcomes over time to determine progress

Email Unsubscribers Per Sales Email



Stage 7: Report Financial Impact



7

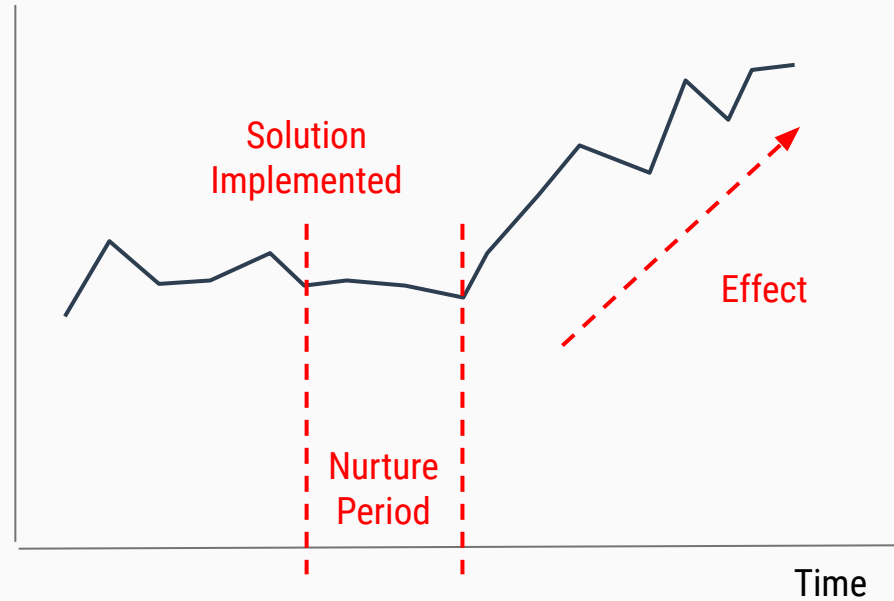
Report Financial Impact

Step 1:
Measure actual results.

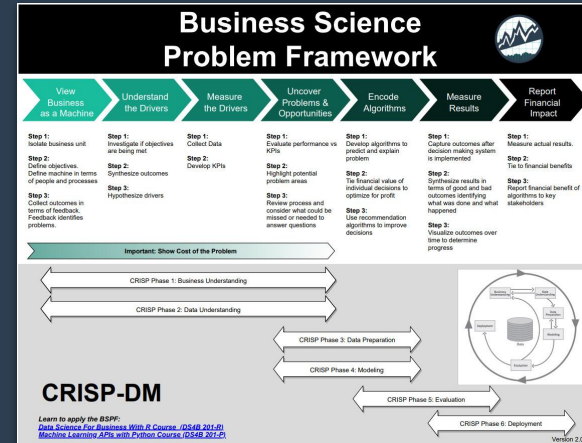
Step 2:
Tie to financial benefits

Step 3:
Report financial benefit of
algorithms to key
stakeholders

Revenue



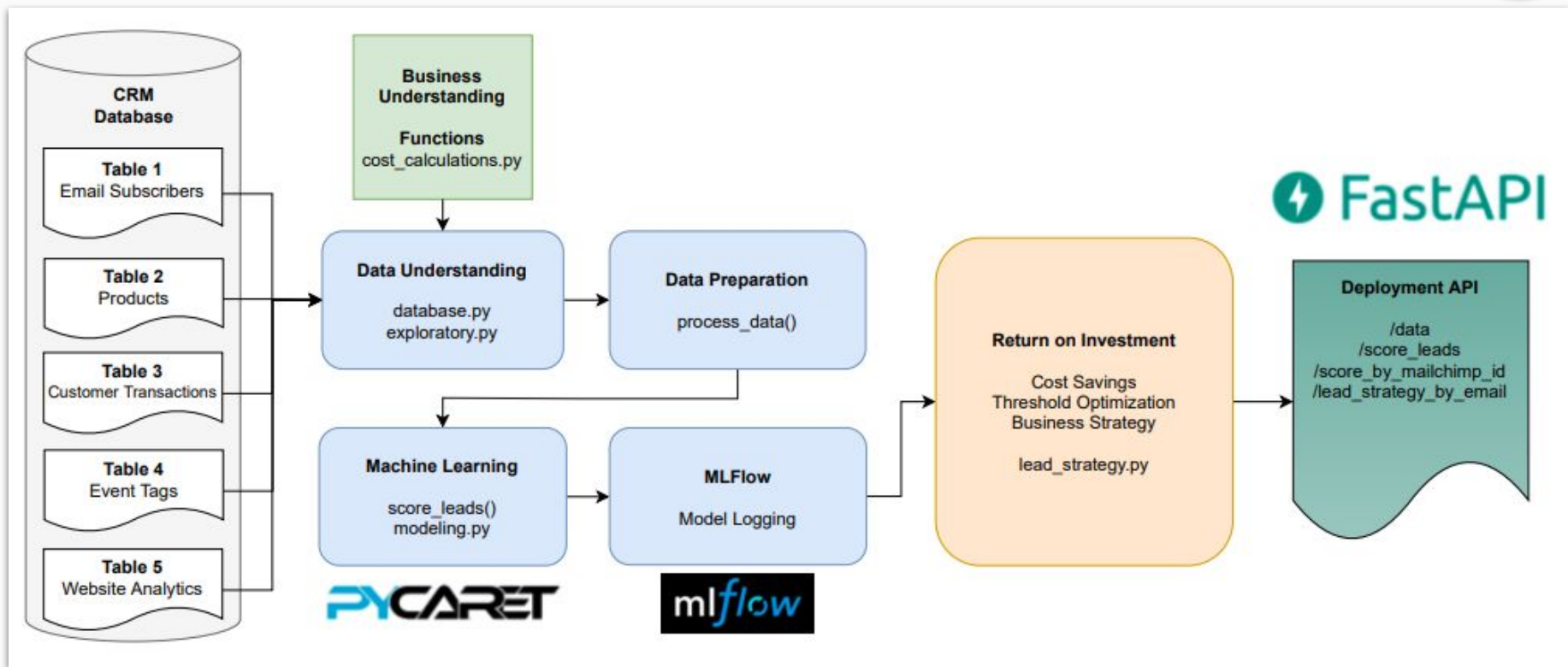
You've just given the organization a way to save millions in the years to come.



Course Workflow

Software & Prerequisites





Create a Python Package

Named:
email_lead_scoring

Process:

1. Create Analysis
2. Convert to Python Functions
3. Add functions to the Python package





Python Software Taught In-Depth

The logo for PyCaret, featuring the word "PYCARET" in a stylized font. "PY" is in blue, "CARET" is in black, and there is a small blue triangle above the "A".The logo for mlflow, featuring the text "mlflow" in a lowercase, sans-serif font. "ml" is white and "flow" is blue.The logo for FastAPI, featuring a green circular icon with a white lightning bolt inside, followed by the text "FastAPI" in a green, sans-serif font.


We learn how to use these production tools in this course.

Python Software NOT Taught In-Depth



You must be proficient with Pandas to be successful in this course.
We dive right into Pandas in Module 1. So get ready!

If Pandas is difficult, then **take the Python for Data Science Automation Course.**



```
O2_model_scoring_function.py x
O4_machine_learning > O2_model_scoring_function.py > ...

You, 2 days ago | 1 author (You)
1 | You, a week ago * machine learning - mlflow
2 # BUSINESS SCIENCE UNIVERSITY
3 # COURSE: DS4B_201-P PYTHON MACHINE LEARNING
4 # MODULE 4: MACHINE LEARNING | MODEL LEAD SCORING FUNCTION
5 # -----
6
7 import pandas as pd
8 import pycaret.classification as clf
9 import email_lead_scoring as els
10
11 leads_df = els.read_and_process_els_data()
12
13 def model_score_leads(data, model_path = "models/blended_models_final_1")
14     mod = clf.load_model(model_path)
15     predictions_df = clf.predict_model(mod, data = data)
16     leads_scored_df = pd.concat([1-predictions_df['Score'],data], axis = 1)
17     return leads_scored_df
18
19 model_score_leads(leads_df)
20
21 # TEST OUT
22
23 import email_lead_scoring as els
24
25 leads_df = els.db_read_and_process_els_data()
26
27 leads_df
28
29 els.model_score_leads(leads_df, model_path="models/blended_models_final_1")
```

VS Code interface showing a Python file named `O2_model_scoring_function.py` with the following code:

```
1 | You, a week ago * machine learning - mlflow
2 # BUSINESS SCIENCE UNIVERSITY
3 # COURSE: DS4B_201-P PYTHON MACHINE LEARNING
4 # MODULE 4: MACHINE LEARNING | MODEL LEAD SCORING FUNCTION
5 # -----
6
7 import pandas as pd
8 import pycaret.classification as clf
9 import email_lead_scoring as els
10
11 leads_df = els.read_and_process_els_data()
12
13 def model_score_leads(data, model_path = "models/blended_models_final_1")
14     mod = clf.load_model(model_path)
15     predictions_df = clf.predict_model(mod, data = data)
16     leads_scored_df = pd.concat([1-predictions_df['Score'],data], axis = 1)
17     return leads_scored_df
18
19 model_score_leads(leads_df)
20
21 # TEST OUT
22
23 import email_lead_scoring as els
24
25 leads_df = els.db_read_and_process_els_data()
26
27 leads_df
28
29 els.model_score_leads(leads_df, model_path="models/blended_models_final_1")
```

The status bar at the bottom indicates: Python 3.7.1 64-bit ('ds4b_201_p': conda), 0 errors, 11 warnings, Server not selected, You, a week ago, Ln 1, Col 1, Spaces: 4, UTF-8, LF, Python, Indents: 0, Kite: indexing.



```
1 # DS4B 201-P: MACHINE LEARNING WITH PYTHON ---- You, a month ago
2 # Environment Setup ----
3
4 # Conda Environment Setup Instructions ----
5 # - Using an environment.yml file with conda
6 # - Important Terminal Commands:
7 #   conda env create -f environment.yml
8 #   conda env update -f environment.yml
9 #   conda env export -n ds4b_201_p > envname.yml
10 #   conda env remove -n ds4b_201_p
11 #   conda env list
12
13 name: ds4b_201_p
14 channels:
15   - anaconda
16   - conda-forge
17   - defaults
18 dependencies:
19   - python=3.7.1
20   - pip
21   - pip:
22     # CORE
23     - numpy==1.20.2
24     - pandas==1.3.4
25     - matplotlib==3.3.4
26     - statsmodels== 0.13.0
27
28     # VISUALIZATION
29     - plotly==5.3.1
30     - plotnine==0.8.0
31
32     # EDA
33     - pyjanitor==0.21.2
34     - sweetviz==2.1.3
```

Interactive-1 X

X Clear All Restart Interrupt Variables Save ds4b_201_p (Python 3.7.1)

Started 'Python 3.7.1 64-bit ('ds4b_201_p': conda)' kernel
Python 3.7.1 (default, Dec 10 2018, 22:54:23) [MSC v.1915 64 bit (AMD64)]
Type 'copyright', 'credits' or 'license' for more information
IPython 7.29.0 -- An enhanced Interactive Python. Type '?' for help.

```
import pandas as pd
import pycaret.classification as clf
import email_lead_scoring as els
```

[1] ✓ 3.3s

environment.yml M 32 # EDA

logs.log 33 - pyjanit

requirements.txt U 34 - sweetvi

35

36

37 # DATABAS

- sqlalch

master* Python 3.7.1 64-bit ('ds4b_201_p': conda) 0 6

Conda Environment: Needed to Replicate Our Analysis & Push Code to Production

Let's go!