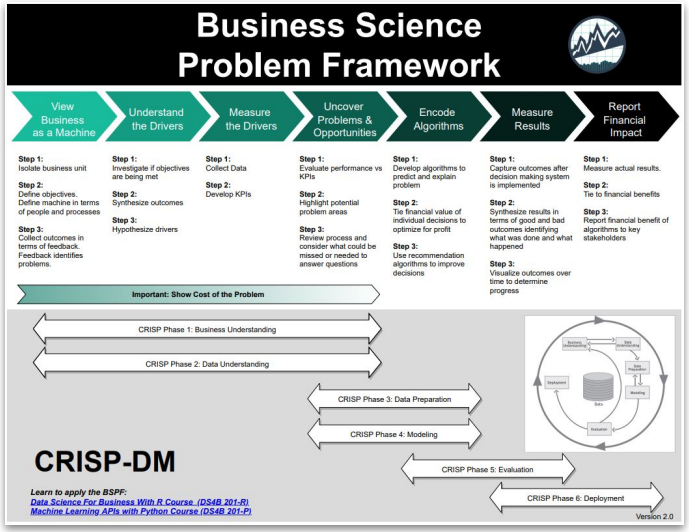
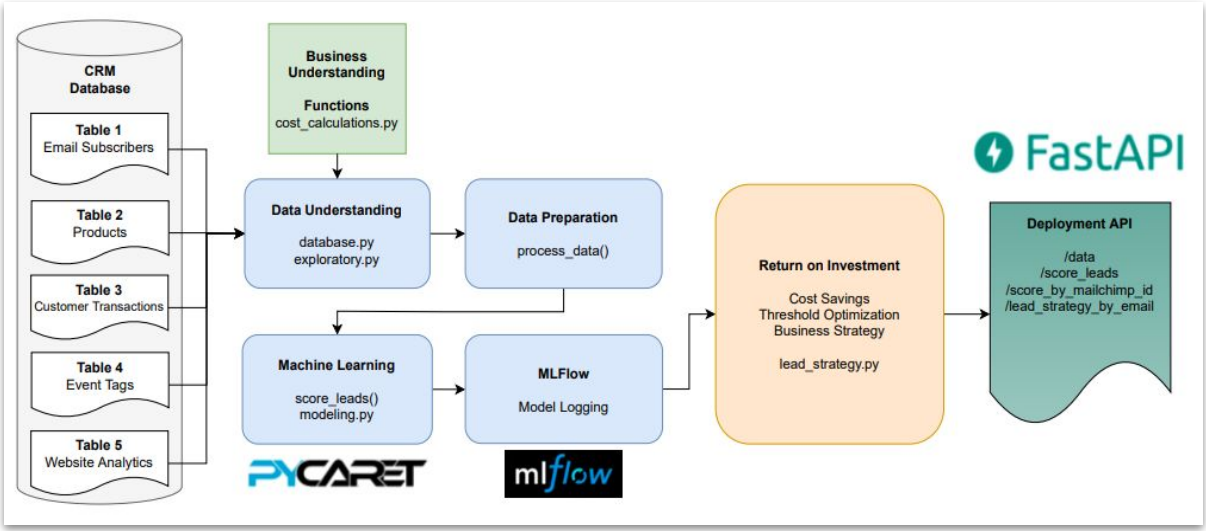




Student Orientation

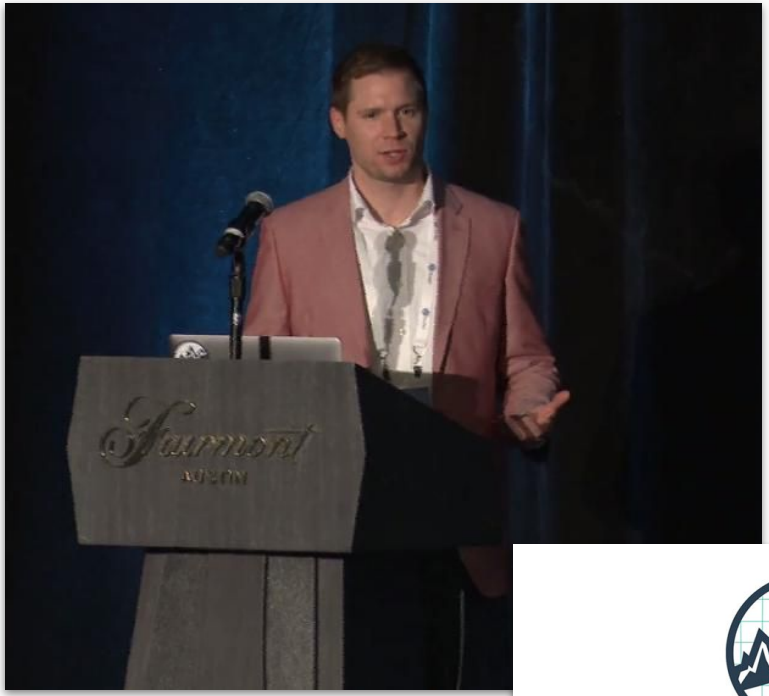
Machine Learning API's with Python Course (DS4B 201-P)

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

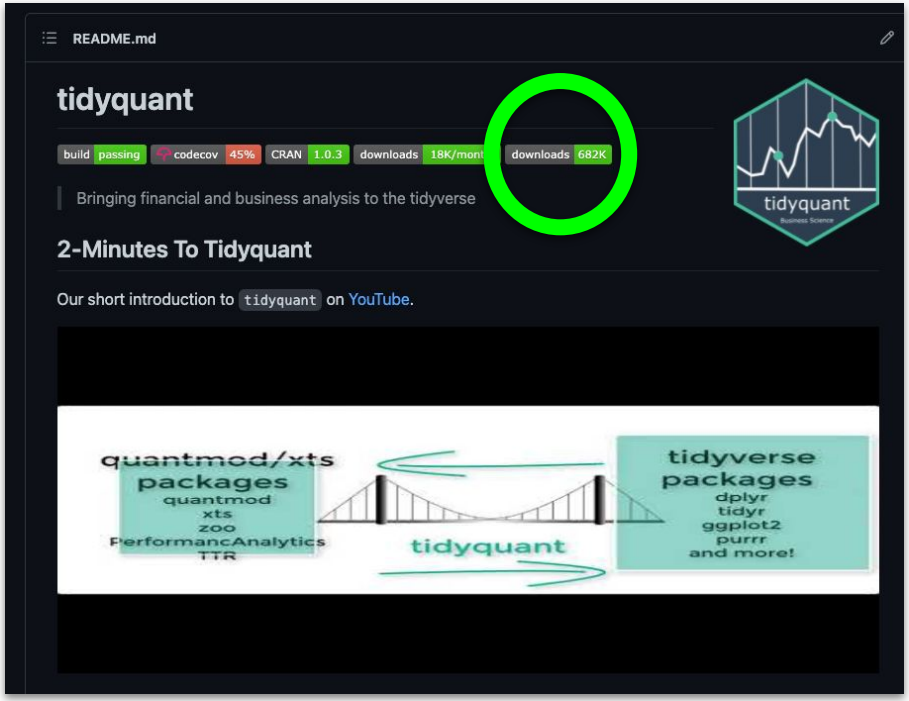
Why Learn from Me?

My CV

I love to **Train Data Scientists**

My CV

I love to **Build Software**



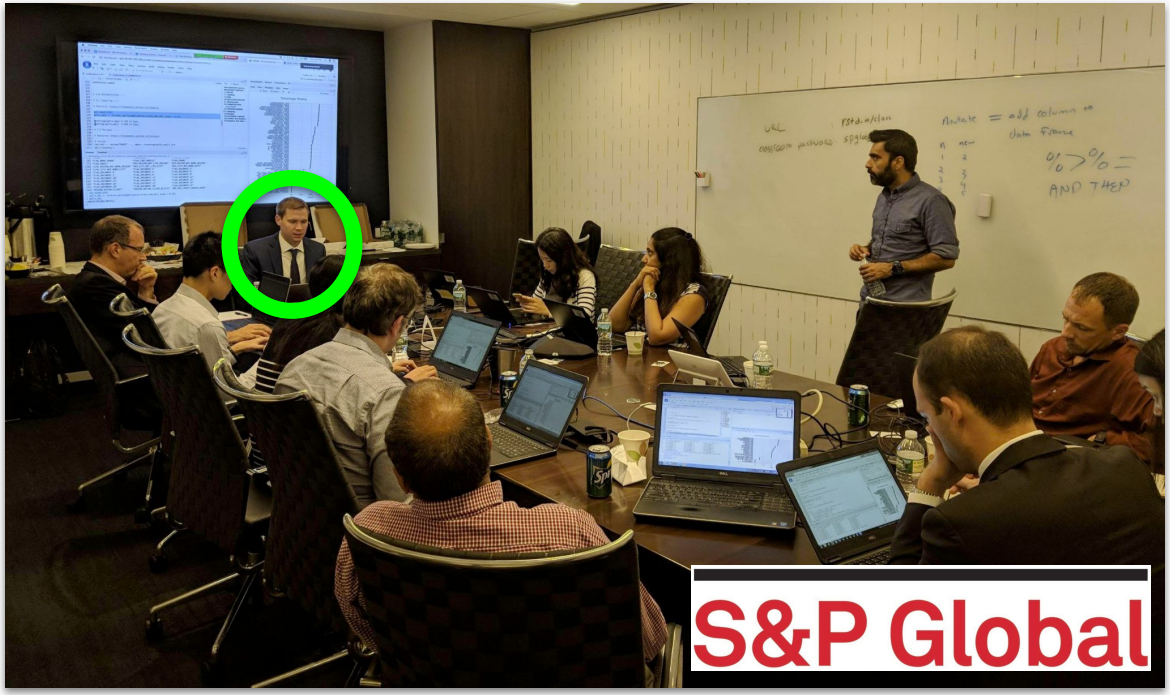
My CV

I love to **train data scientists.**

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

They learned:

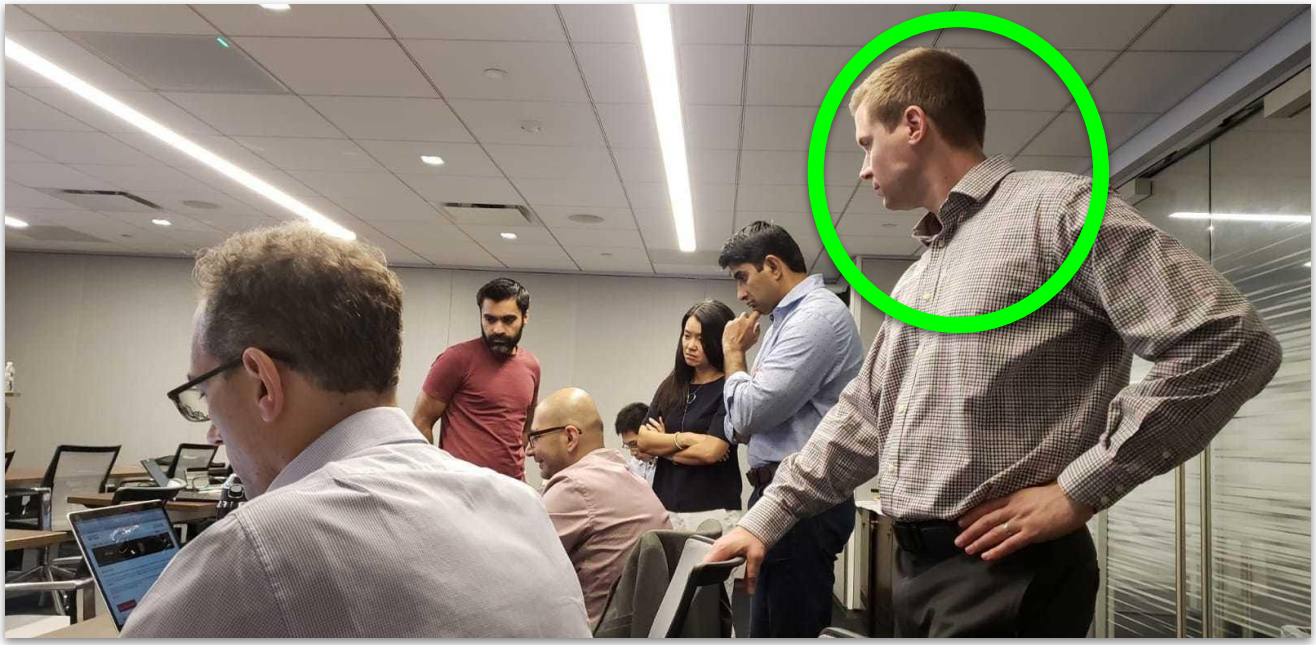
- **Same** Framework
- **Same** Workflow
- **Completely different** Problem



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

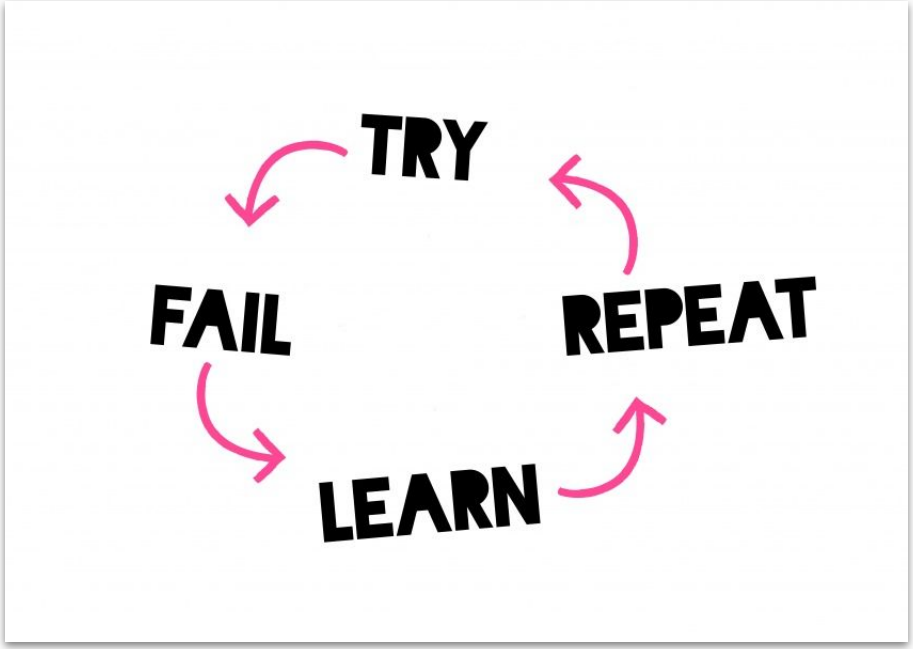
Origin of the BSPF

Failure comes before success

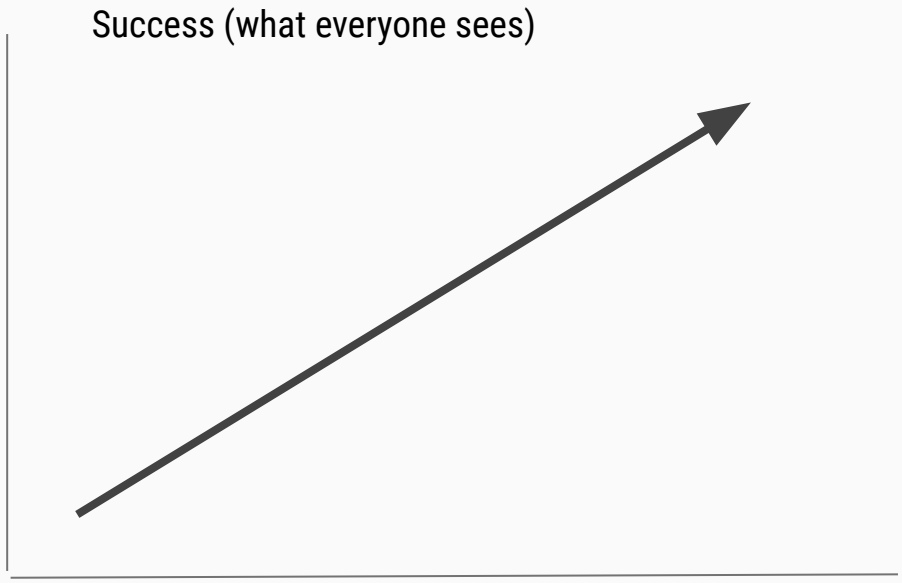
It’s OK to Fail.

We all fail.

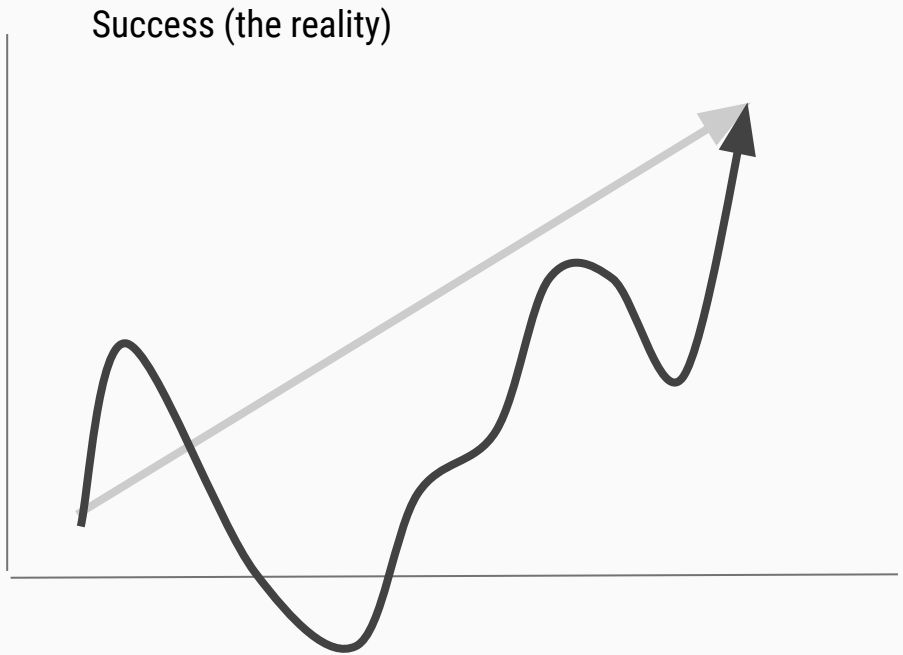
Those that are successful learn from failure.

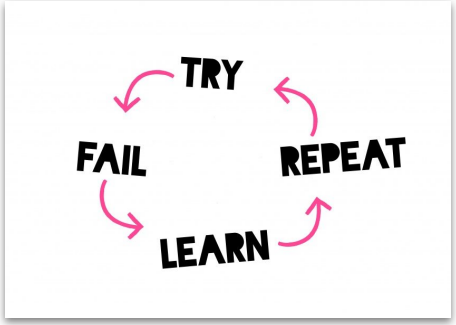


What people think success looks like



What real success looks like



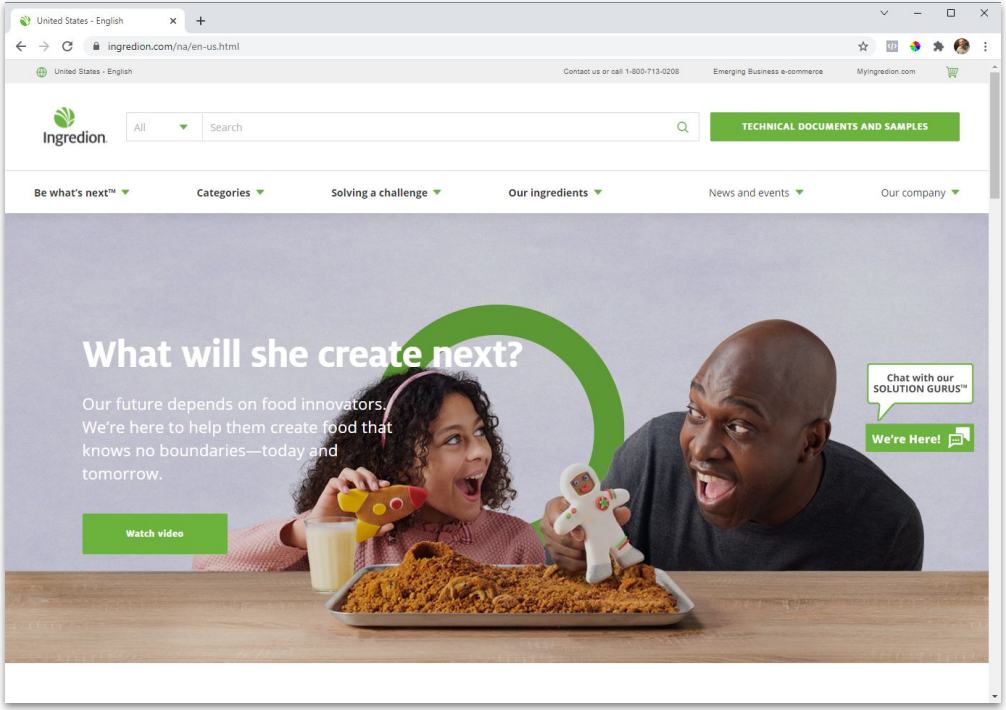


My first BIG Opportunity

Big Opportunity: Ingredion

$6B in Revenue (2021)

Fortune 500 (#463 in 2021)

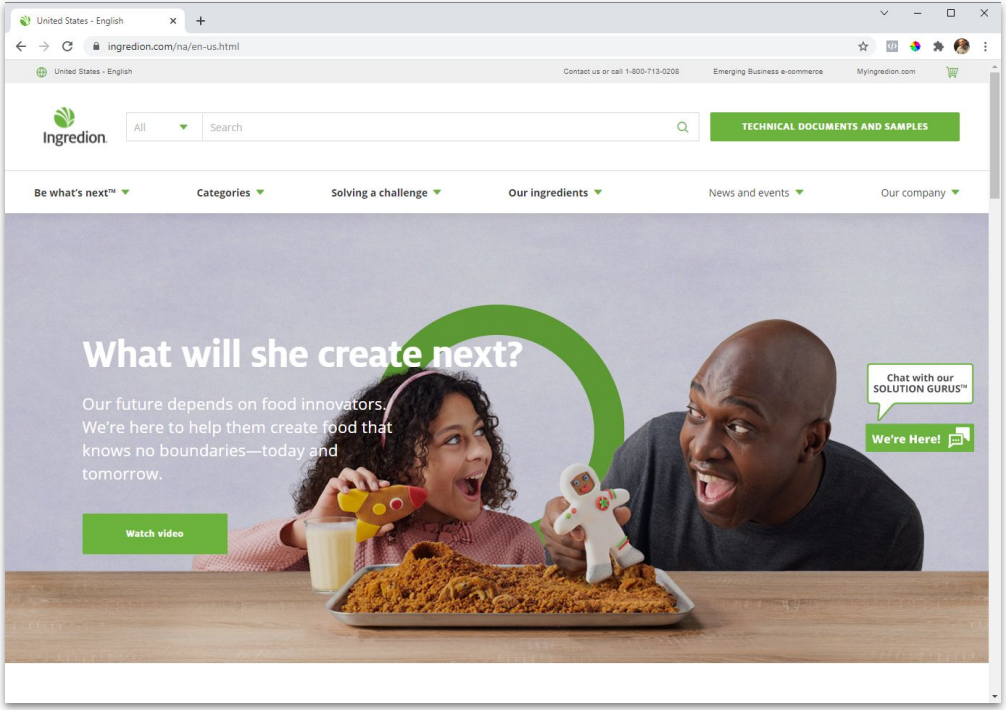


My first BIG Project

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.

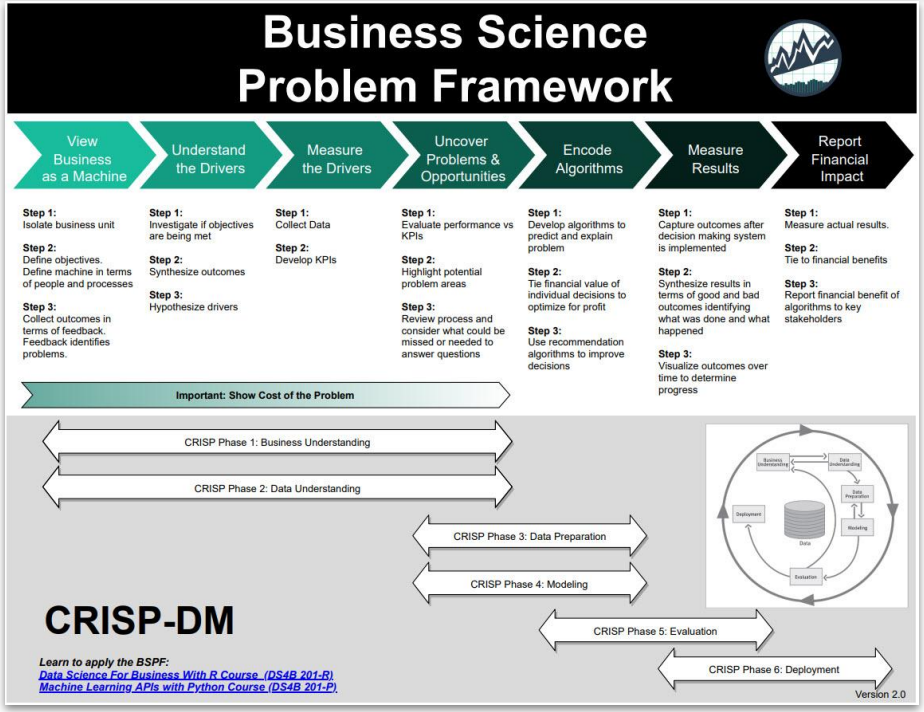
They **needed a process** for converting data problems to business value.

The BSPF was born

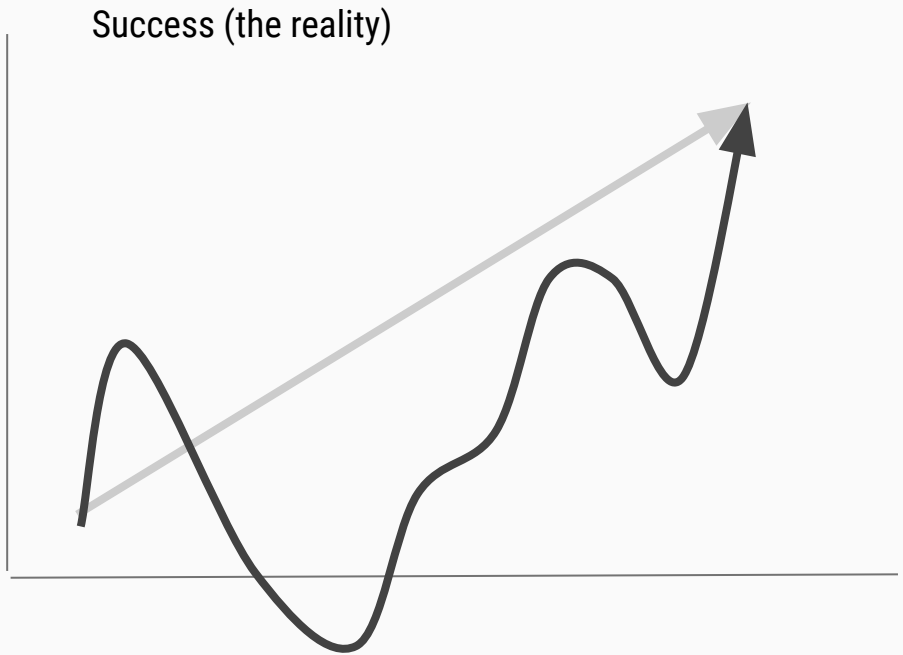
A **problem solving** framework.

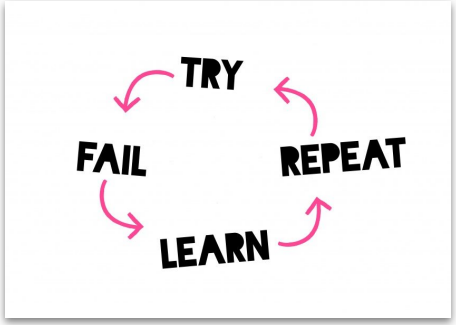
Key Improvements:

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

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 **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

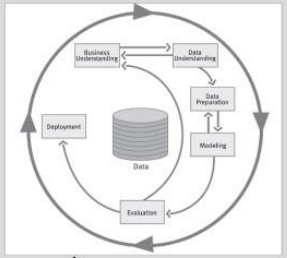
CRISP Phase 2: Data Understanding

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.

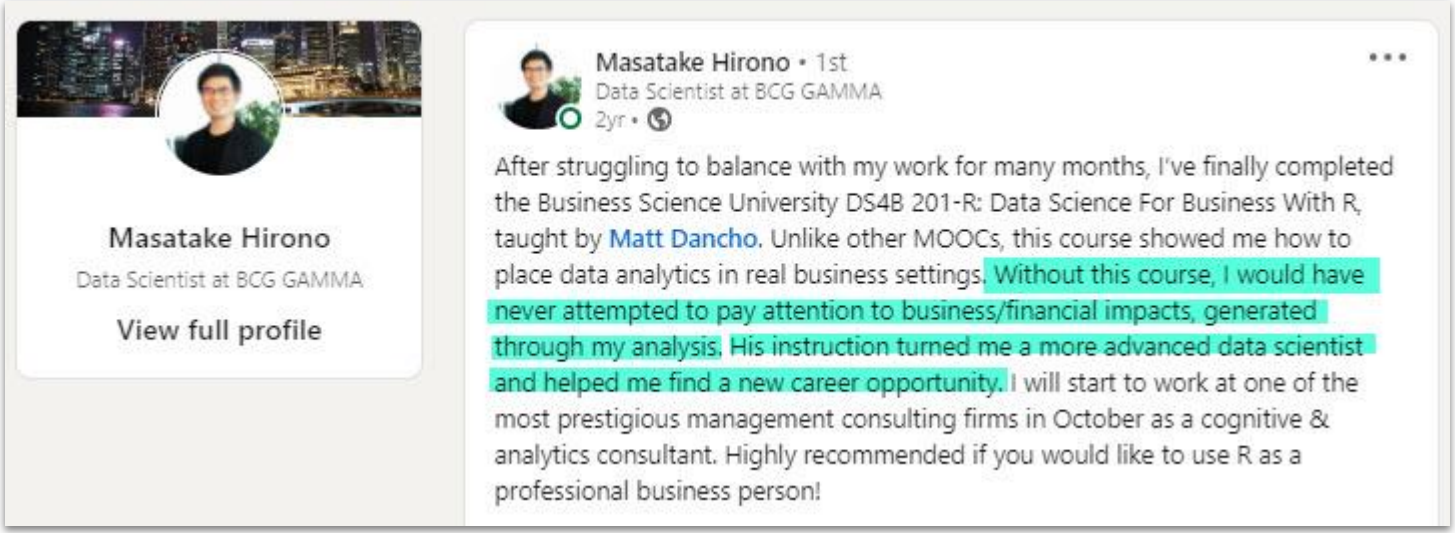
Proof #1: Jennifer used the BSPF to land her dream job at JP Morgan

Thanks to Matt and what I've learned so far, I was able do 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

BCG GAMMA

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

accenture

BCG GAMMA

Solving the Course Project

Email Lead Scoring

(What you are solving with BSPF in this course)

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 **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

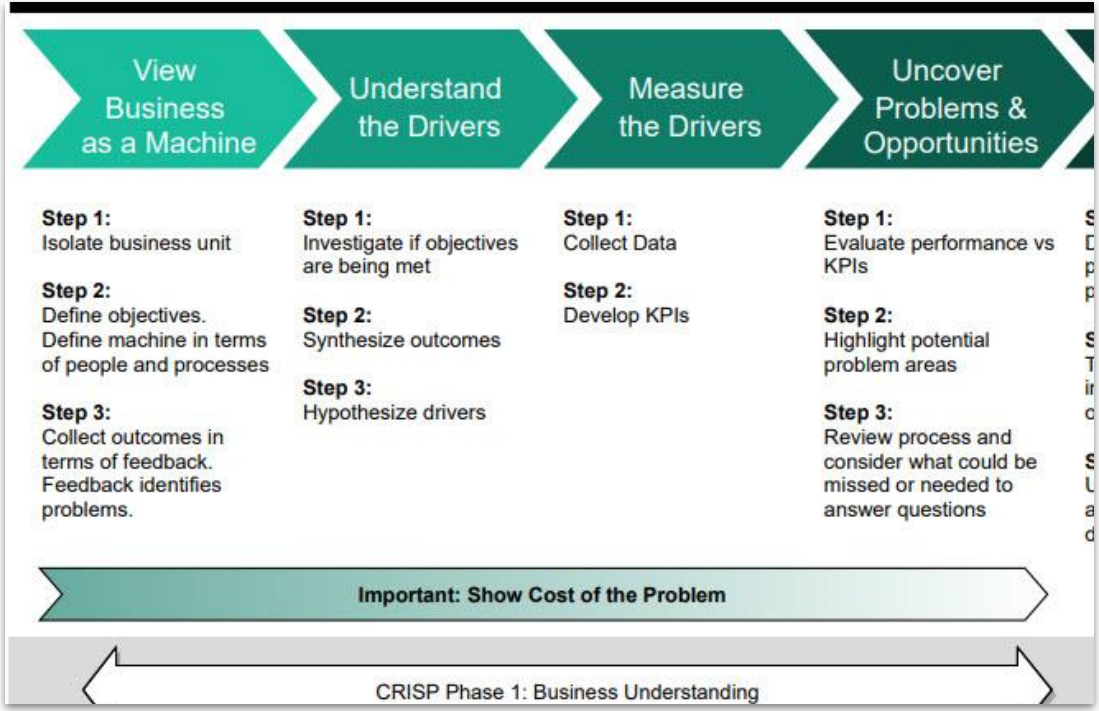
Learn to apply the BSPF:

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.

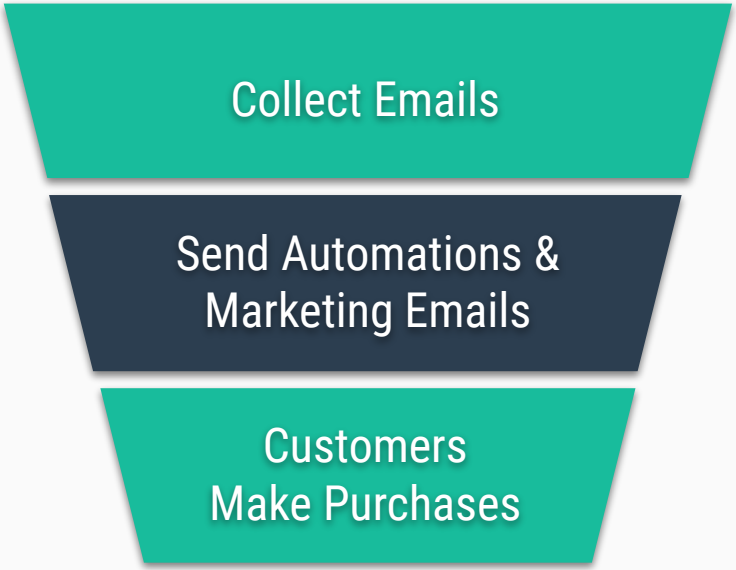
Stages help:

- **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? 👇



Problem: Revenue is slowing.

Objective: Make Purchase

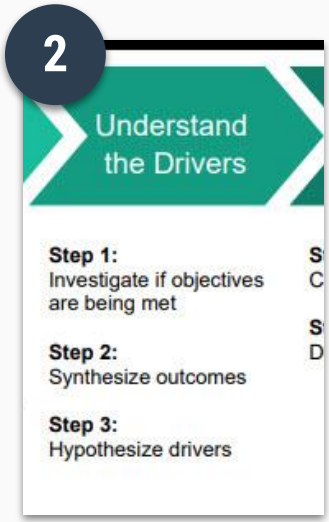
Process: Companies send 2 Types of Emails:

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

Outcome: High Unsubscribe Rates

$3,000,000

Per Year



Important: Show Cost of the Problem

Cost of losing unsubscribers due to email blasts.

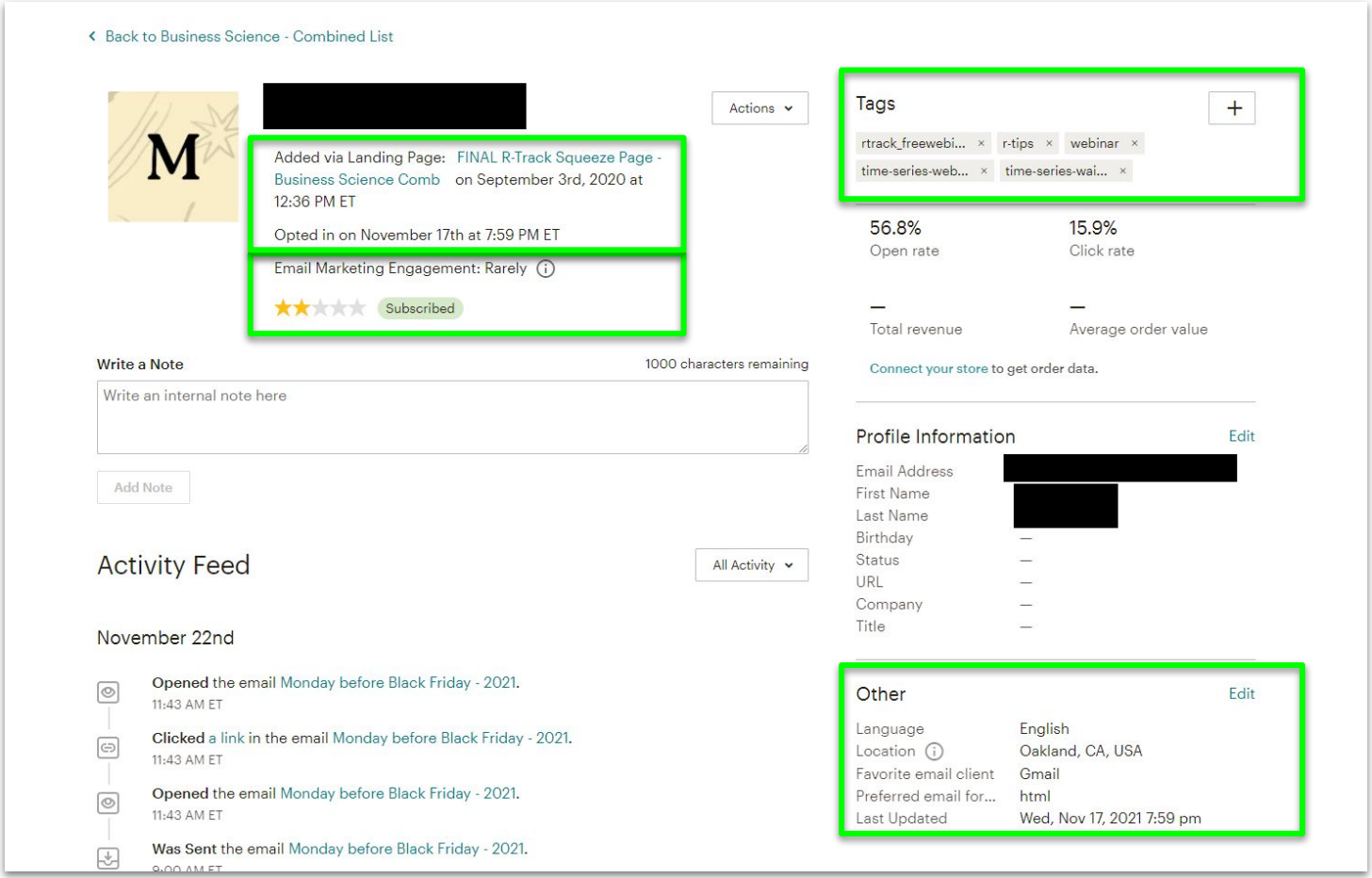
Send 5 sales blasts per month. Lose 500 subscribers per email blast.

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

3

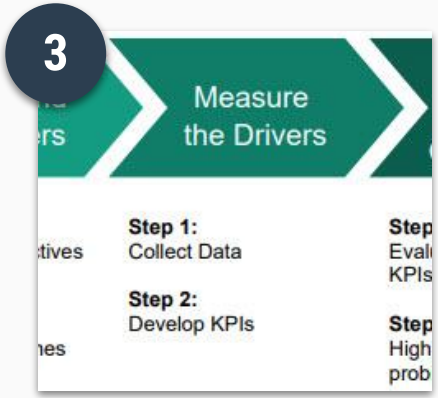
Collect data with Marketing Team.

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



3.3X

vs Industry Unsub Rate

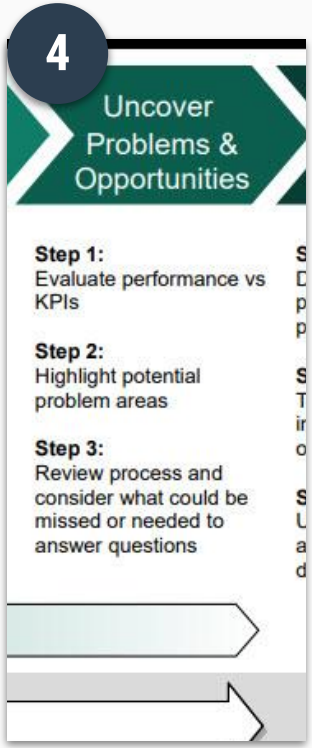


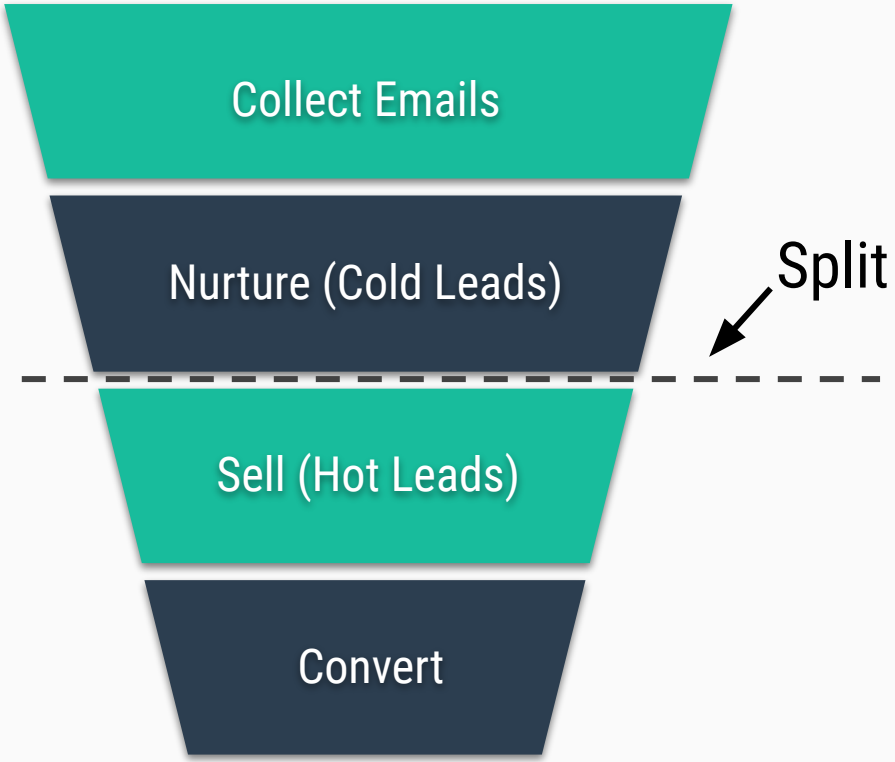
High Unsubscribe Rates

KPI (Industry Average): 0.0015

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

Stage 4: Uncover Problems & Opportunities





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!

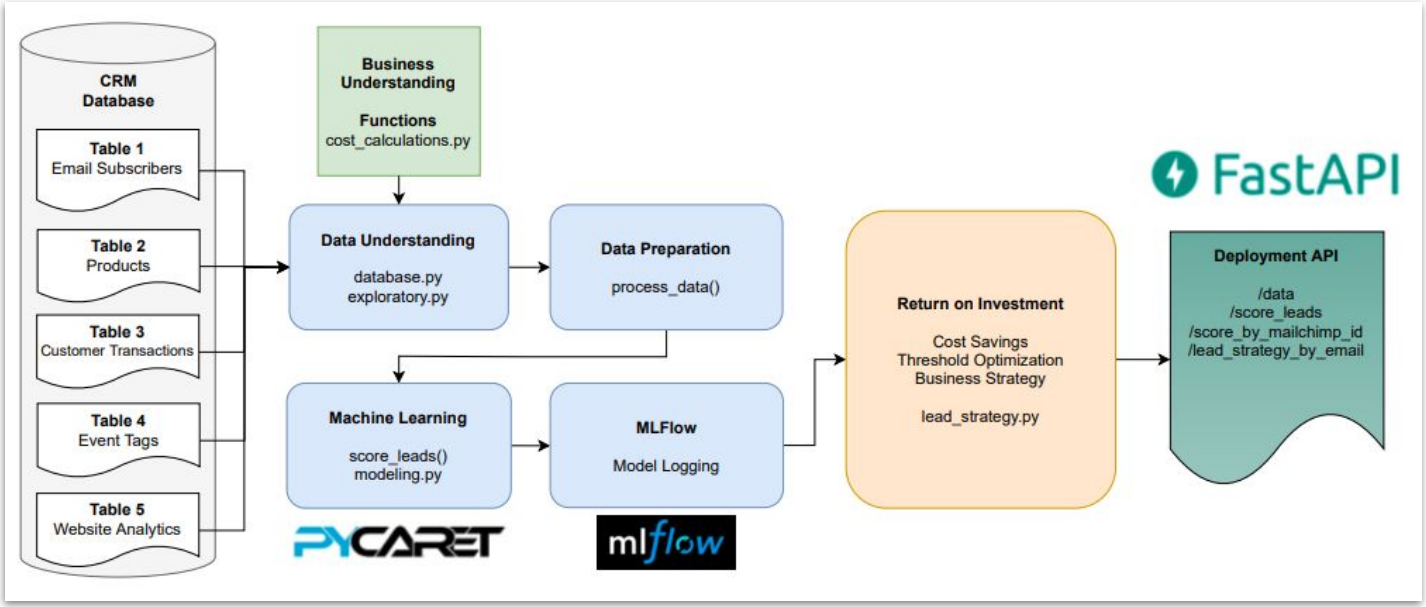
5

Encode Algorithms

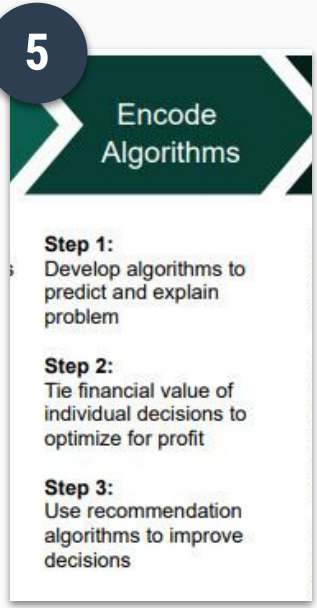
Step 1:
Develop algorithms to predict and explain problem

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.




```
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 × 6 columns

Email Lead Scoring Strategy

Recommended Strategy Buy-In:
Give the organization a way to save money in the years to come.

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

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

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

category	count	sum_made_purchase
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

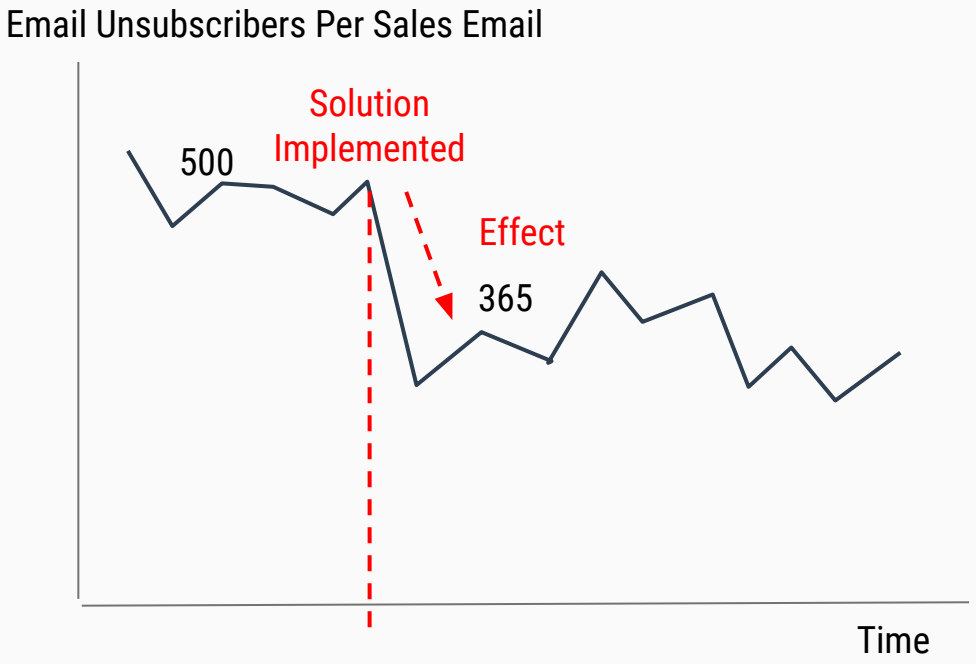
6

Measure Results

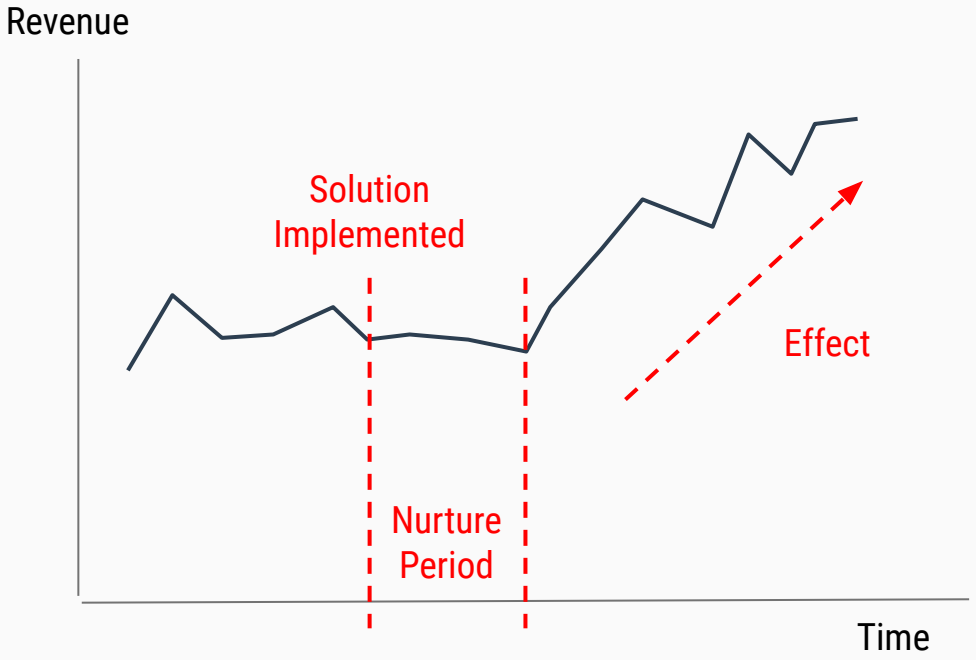
Step 1:
Capture outcomes after decision making system is implemented

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

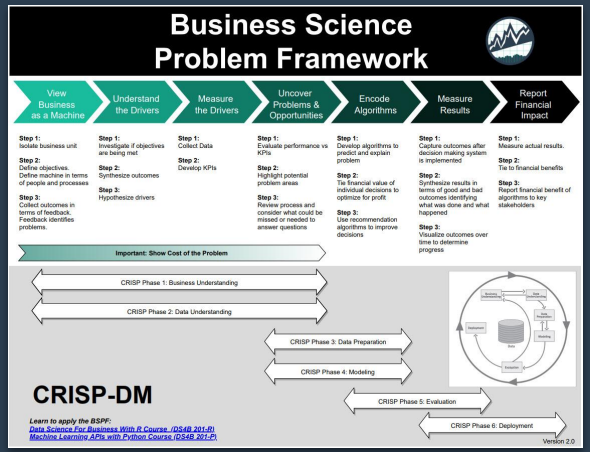
Step 3:
Visualize outcomes over time to determine progress



Stage 7: Report Financial Impact

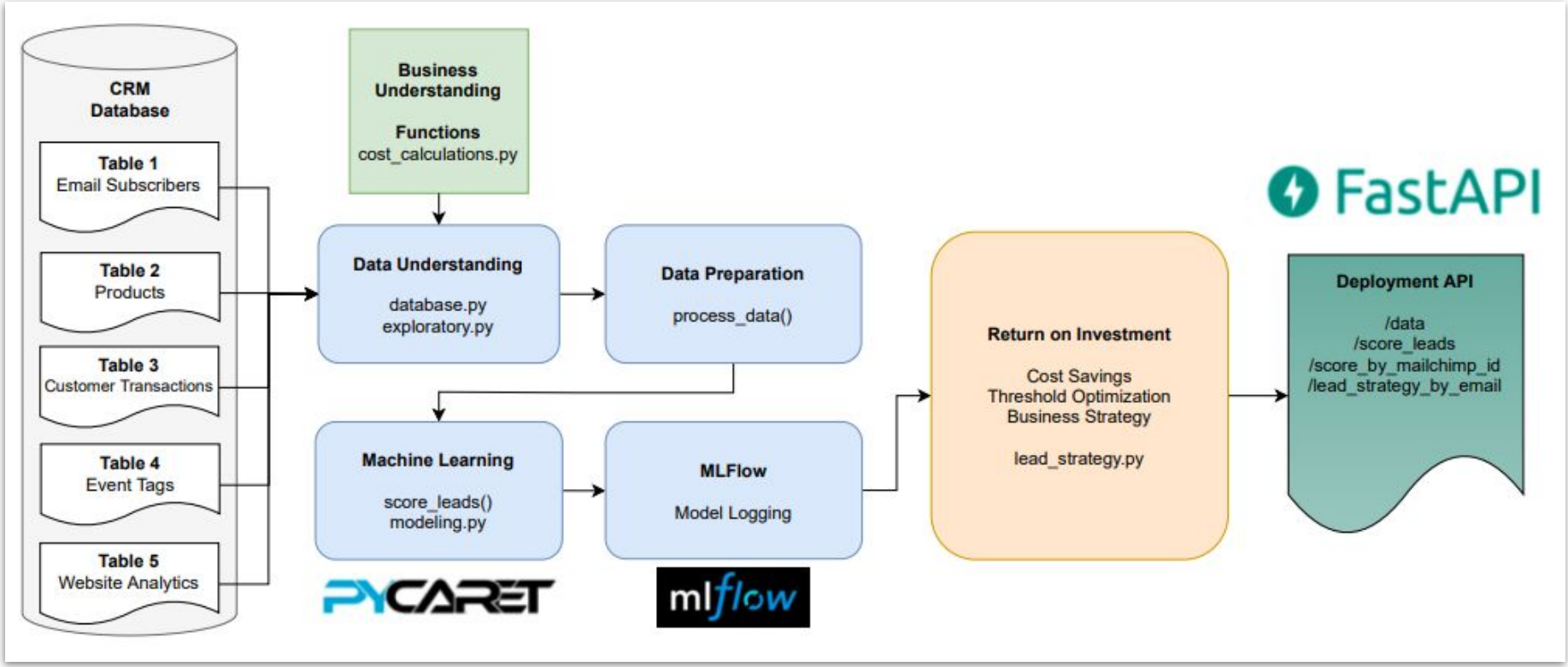


You've just given the organization a way to save millions in the years to come.



Course Workflow

Software & Prerequisites


CRM Database
Table 1 Email Subscribers
Table 2 Products
Table 3 Customer Transactions
Table 4 Event Tags
Table 5 Website Analytics
Business Understanding
Functions
cost_calculations.py
Data Understanding
database.py
exploratory.py
Data Preparation
process_data()
Machine Learning
score_leads()
modeling.py
MLFlow
Model Logging
PYCARET
mlflow
Return on Investment
Cost Savings
Threshold Optimization
Business Strategy
lead_strategy.py
FastAPI
Deployment API
/data
/score_leads
/score_by_mailchimp_id
/lead_strategy_by_email

Create a Python Package

Named:
`email_lead_scoring`

Process:

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

```
import email_lead_scoring as els
els.cost_simulate_unsub_costs(
    email_list_monthly_growth_rate=[0, 0.03],
    customer_conversion_rate=[0.03, 0.06]
) \
    .pipe(els.cost_plot_simulated_unsub_costs)
```

✓ 4.2s

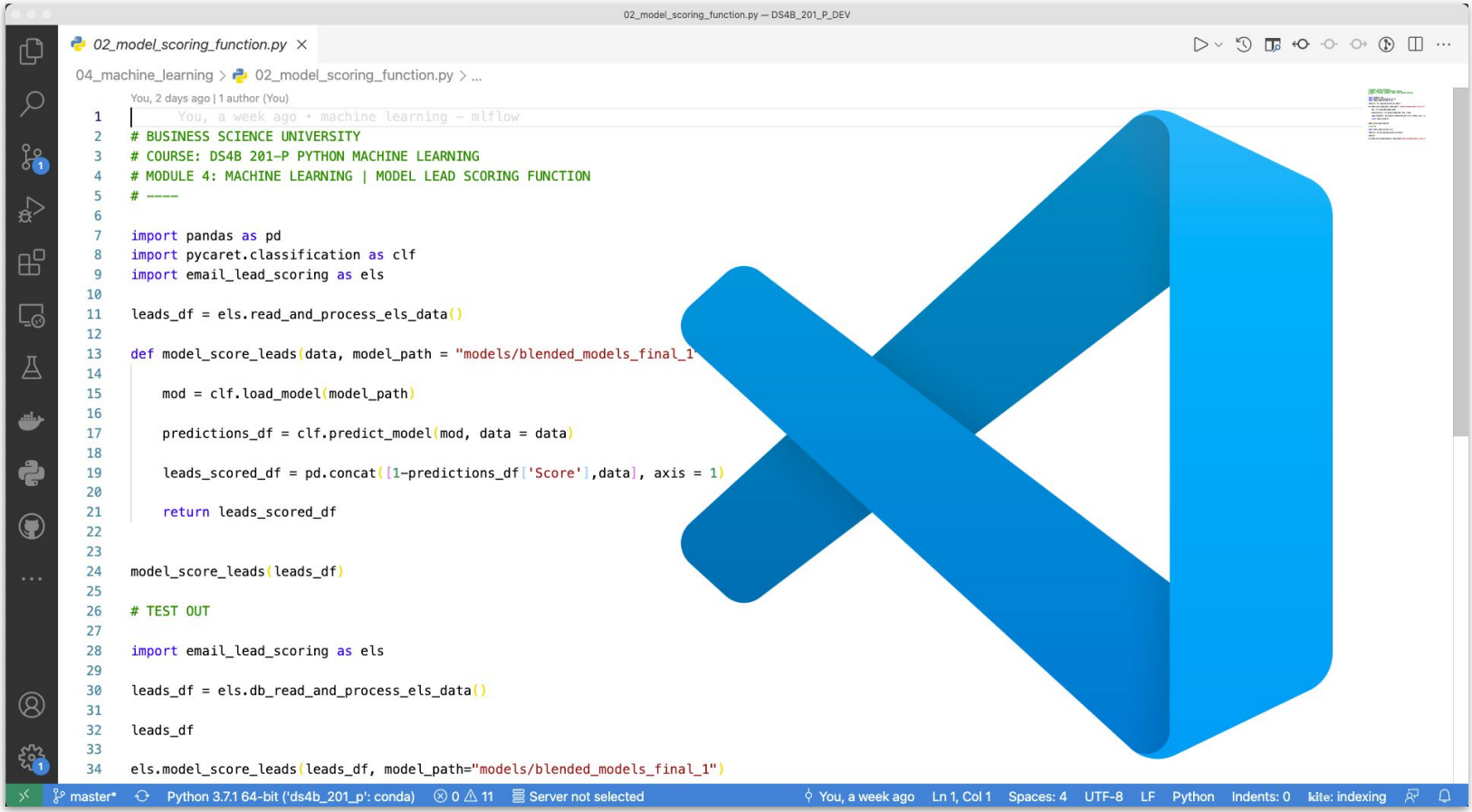
Python Software Taught In-Depth

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

Python Software NOT Taught In-Depth

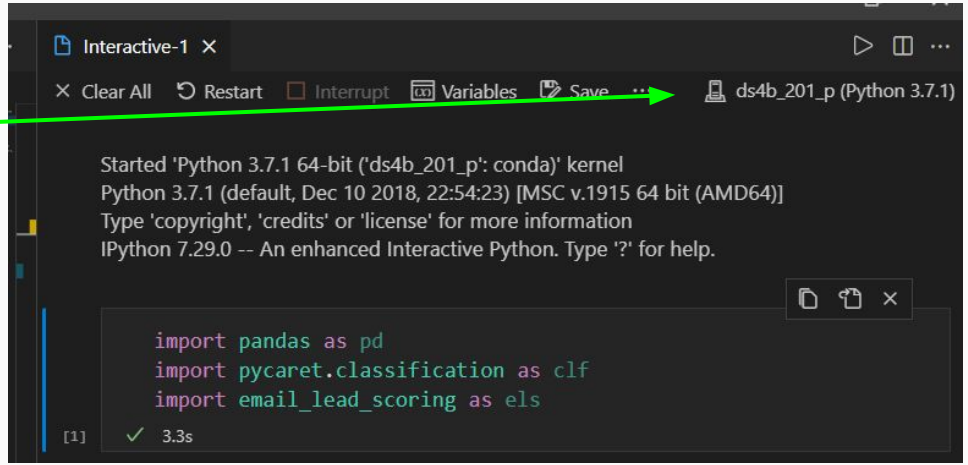
You must be proficient with Pandas to be successful in this course.
We dive right into Pandas in Module 1. So get ready!

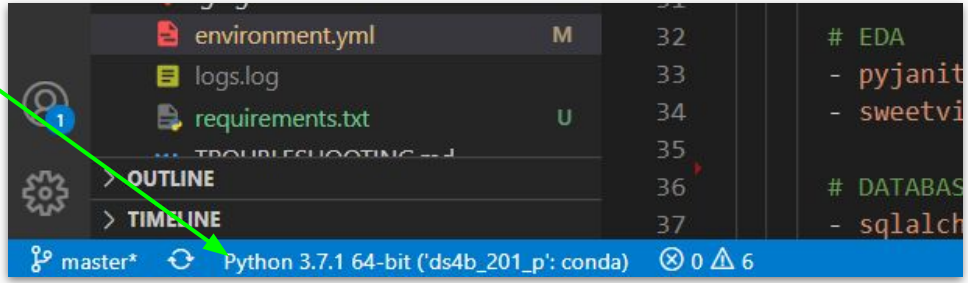
If Pandas is difficult, then **take the Python for Data Science Automation Course**.



VSCODE: Integrated Development Environment used in the Course

```yaml
# DS4B 201-P: MACHINE LEARNING WITH PYTHON ----
# Environment Setup ----

# Conda Environment Setup Instructions ----
#  - Using an environment.yml file with conda
#  - Important Terminal Commands:
#    conda env create -f environment.yml
#    conda env update -f environment.yml
#    conda env export -n ds4b_201_p > envname.yml
#    conda env remove -n ds4b_201_p
#    conda env list

name: ds4b_201_p
channels:
  - anaconda
  - conda-forge
  - defaults
dependencies:
  - python=3.7.1
  - pip
  - pip:
      # CORE
      - numpy==1.20.2
      - pandas==1.3.4
      - matplotlib==3.3.4
      - statsmodels== 0.13.0

      # VISUALIZATION
      - plotly==5.3.1
      - plotnine==0.8.0

      # EDA
      - pyjanitor==0.21.2
      - sweetviz==2.1.3
```





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

Let’s go!