

# Introductory Session

## PGP AIML Mentored Learning Session M1W1

# Agenda

1. Welcome & Introductions
2. Overview of AI and Data Science
3. Case Study – Example
4. Getting help from Python
5. QnA

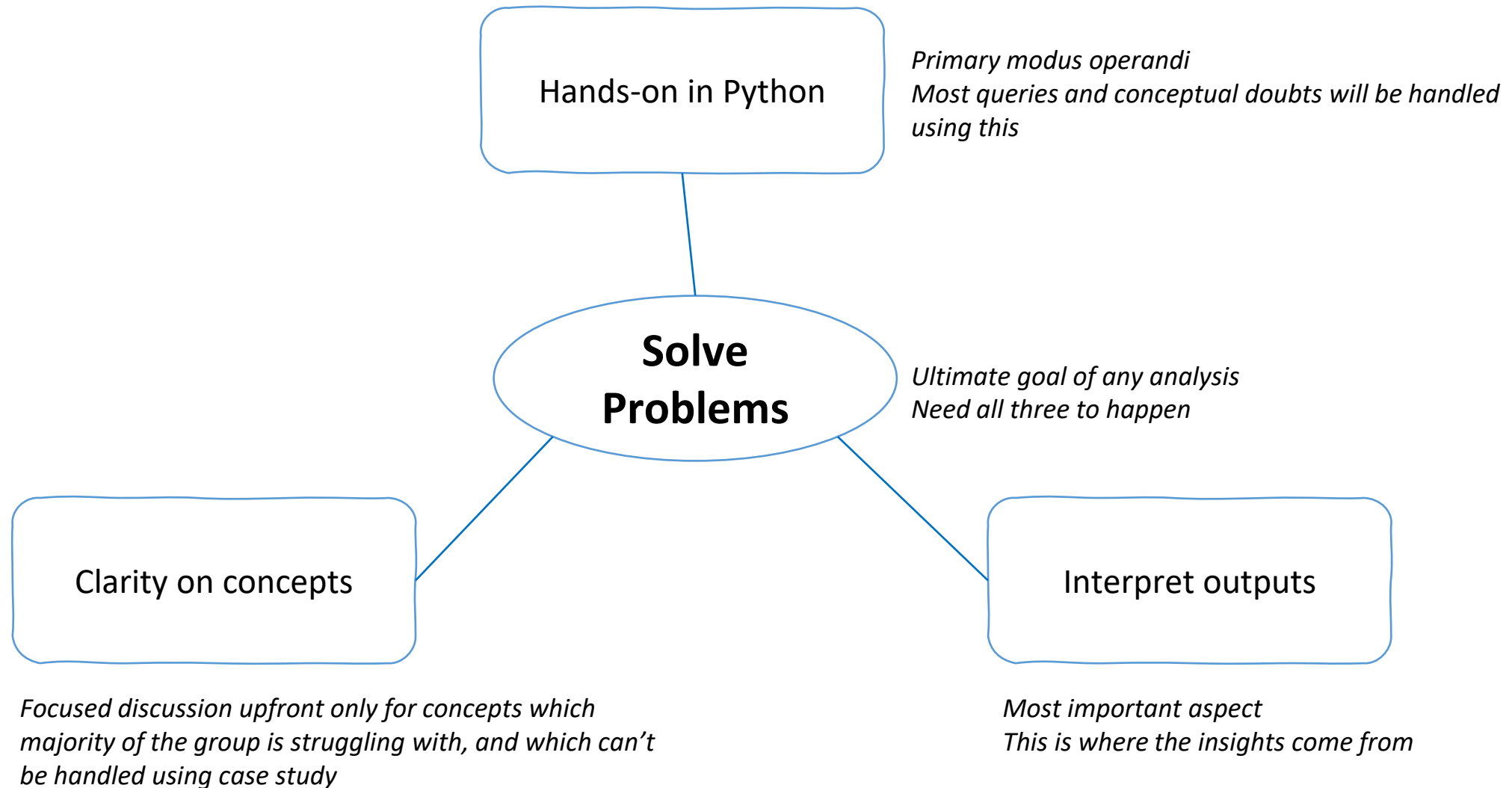
# Agenda

1. Welcome & Introductions
2. Overview of AI and Data Science
3. Case Study – Example
4. Getting help from Python
5. QnA

# Introductions



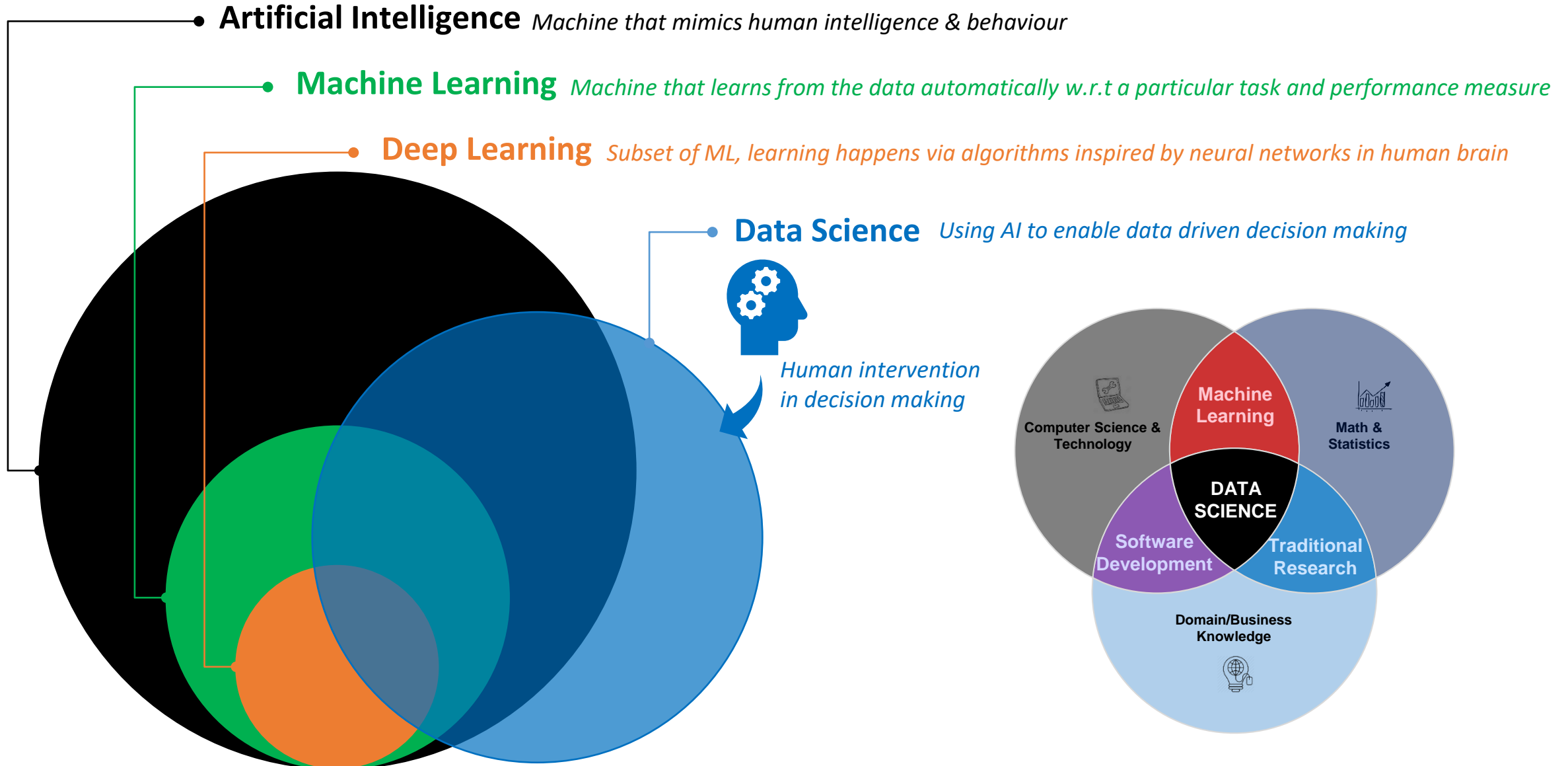
# Expectations from Mentored Learning Sessions



# Agenda

1. Welcome & Introductions
2. Overview of AI and Data Science
3. Case Study – Example
4. Getting help from Python
5. QnA

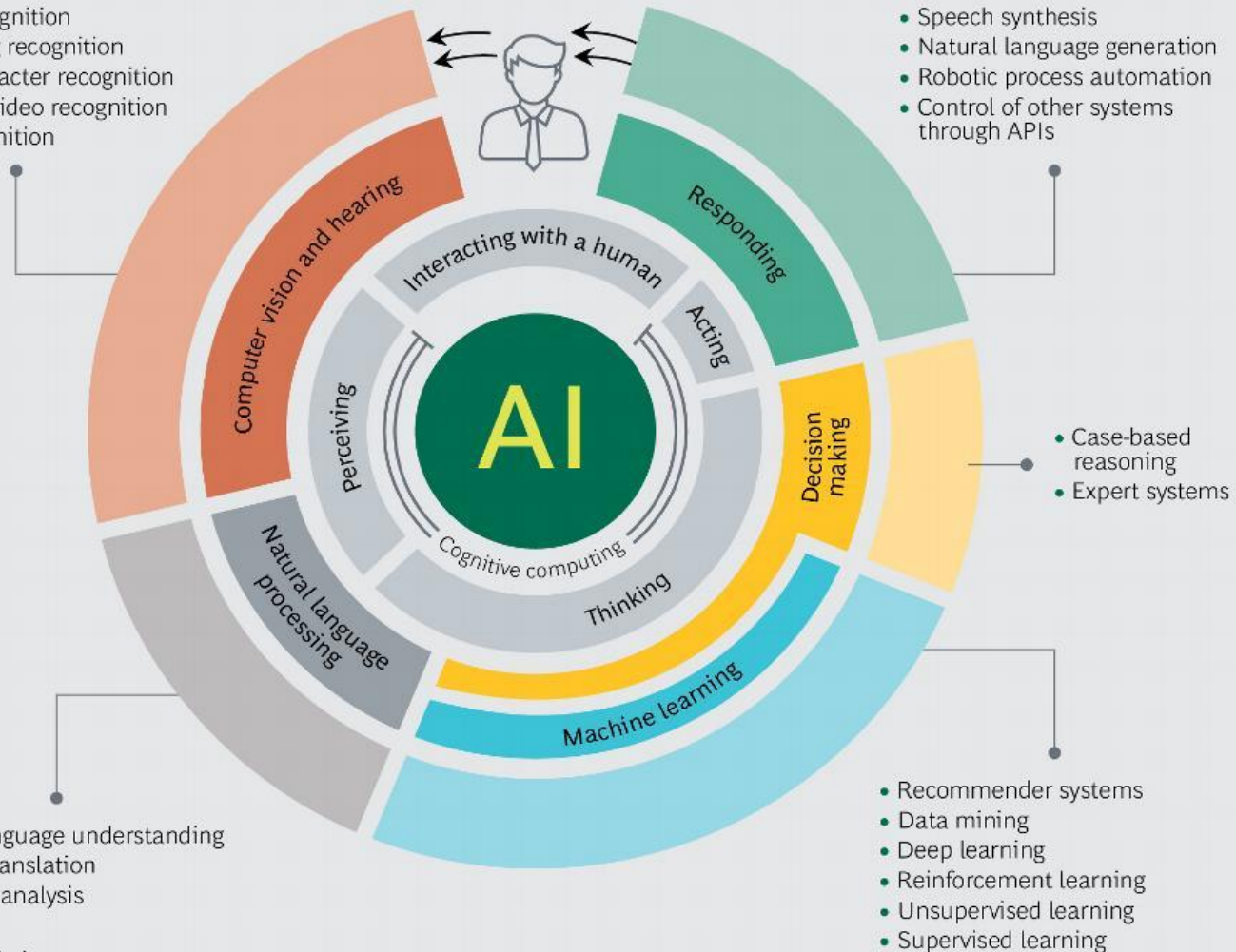
# AI or ML or DL or DS – Questions?



# AI technologies come in many forms

## EXHIBIT 1 | AI and Robotics Technologies Come in Many Forms, Giving Rise to a Broad Variety of Applications

- Speech recognition
- Handwriting recognition
- Optical character recognition
- Image and video recognition
- Facial recognition



Source: BCG analysis.

Note: APIs = application programming interfaces.



## AI – Examples –Self Driving Car



- Constant data input stream through sensors and cameras
- Machine learning engine at play to analyze and predict what could happen in next 'X' seconds
- Decision engine uses predictions to decide what actions to take
- Robotics machinery executes those actions

# ML Example – Spam or No Spam?

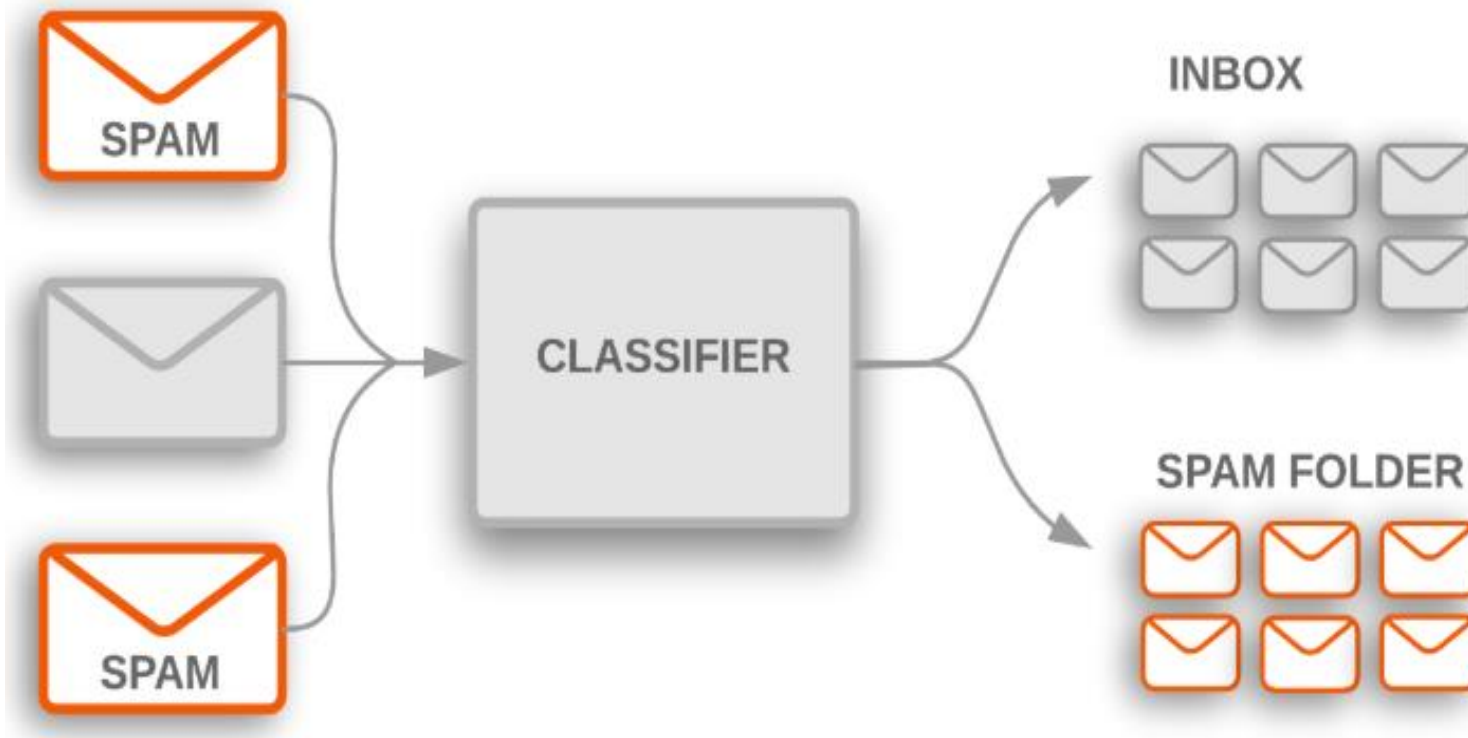
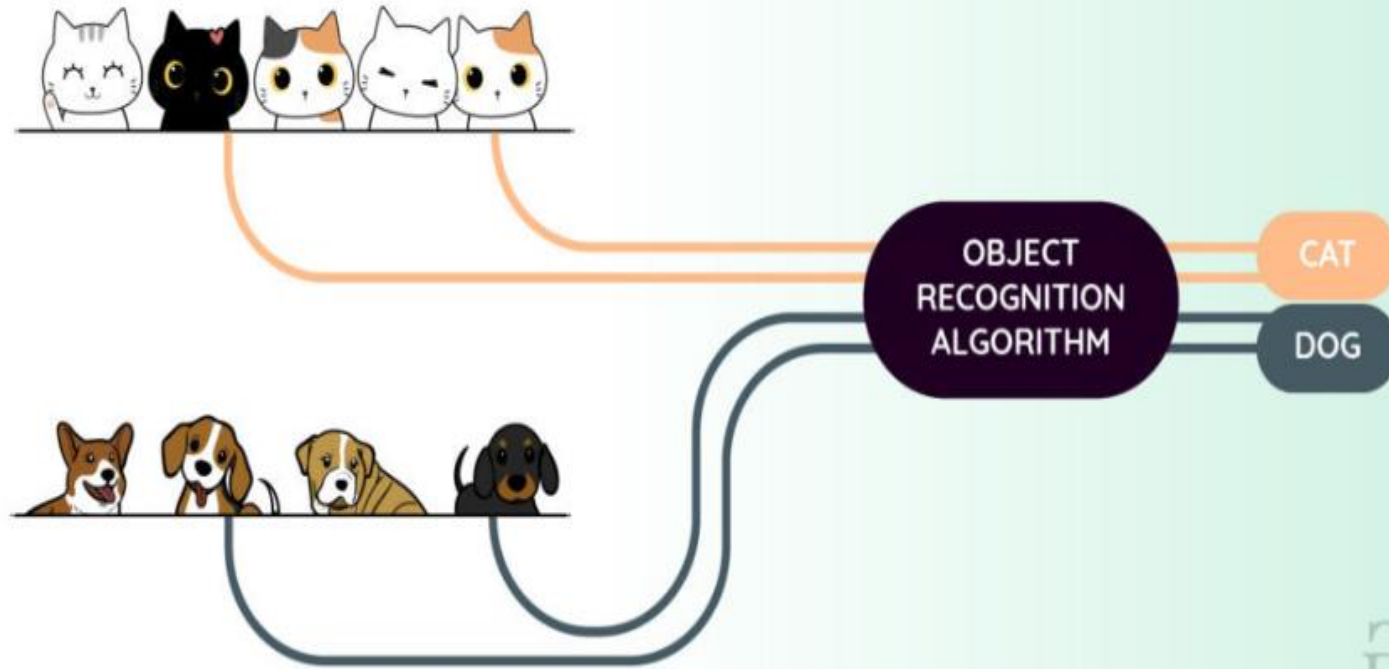


Image Source – [developers.google.com](https://developers.google.com)

- Historical emails are categorized as spam or not
- Algorithm is constantly identifying the features that characterize a spam email, like large mailing list, or from a suspicious email id etc.
- When a new email comes in, the algorithm classifies it based on historical learning
- Every time a decision is overridden by the user, the algorithm adjusts itself







# DL Example – Is it a Cat or Dog?



- Machine is fed thousands of images of cats and dogs
- Algorithm is developed to identify different features in the images
- It learns overtime how a cat image differs from a dog image
- When fed a new image, it can now identify if it's cat or a dog

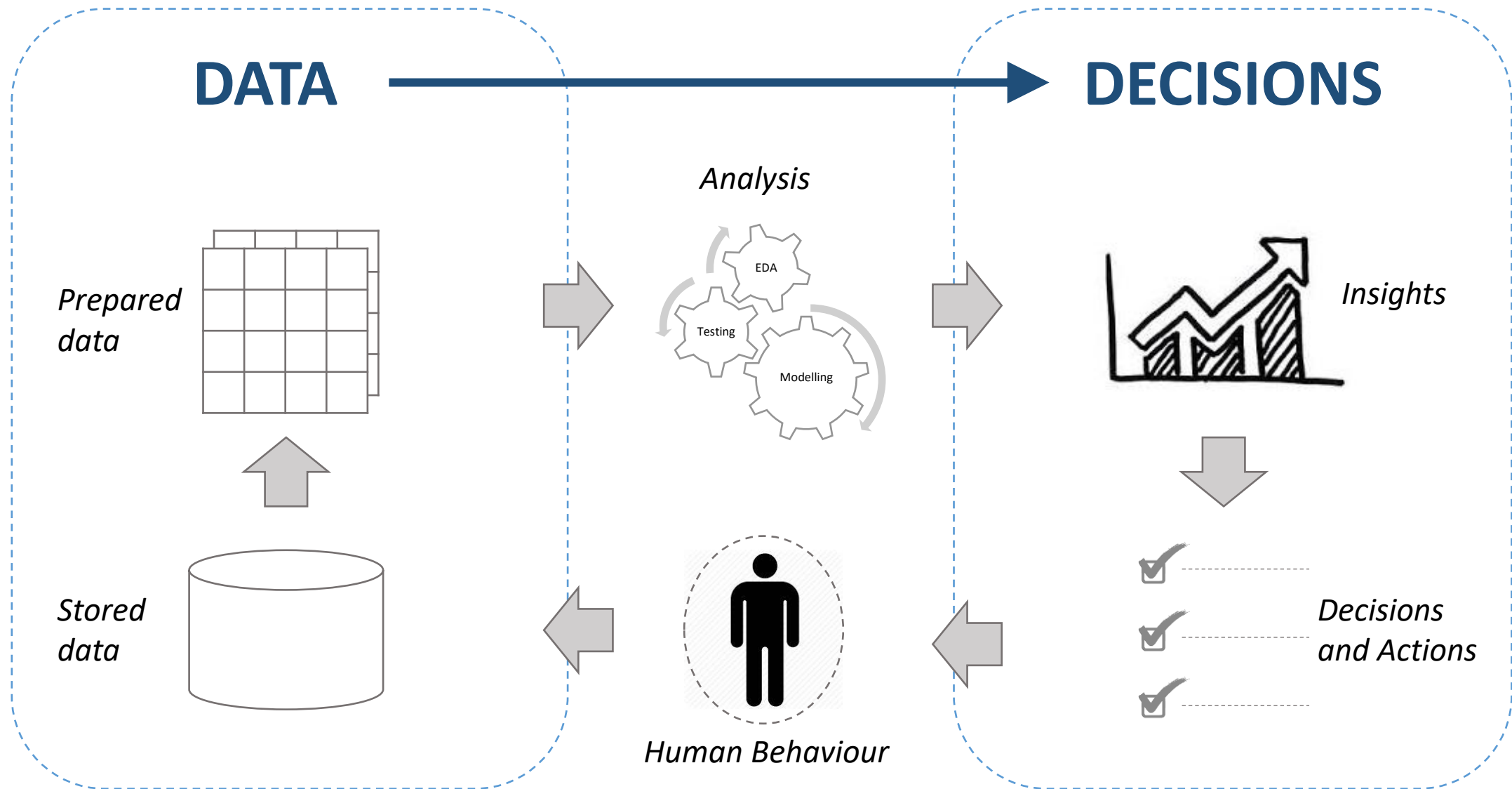
Image Source – [Geek for Geeks](#)

# DL Example – Is Twitter sentiment positive or negative?

Sentiment		Emotion	
Positive			
Negative			

- Meaningful words are extracted from tweets
- Words and their combinations are characterized by a certain 'emotion'
- Algorithms developed to identify the majority 'sentiment' and classify them

Analytics is a process of going from data to decision and it is a cycle





# What is Data Science and what is it not?

```
class BigFile:
    def __init__(self, datadir, ndims):
        self.datadir = datadir
        self.names = [x.strip() for x in str.split(open(self.datadir).read()) if x.strip()]
        self.name2index = dict(zip(self.names, range(len(self.names))))
        self.ndims = ndims
        self.featurefile = os.path.join(datadir, "feature.bin")
        print "[BigFile] ndims: %d" % self.ndims
        print "[BigFile] binary: %s" % self.featurefile
        print "txt: %s" % self.datadir

    def read(self, requested, isname=True):
        if isname:
            index_name_array = [(self.name2index[x], x) for x in requested if x in self.names]
            return index_name_array
        else:
            return [self.names[x] for x in requested]
```

{Python}

Not about code, but the solution

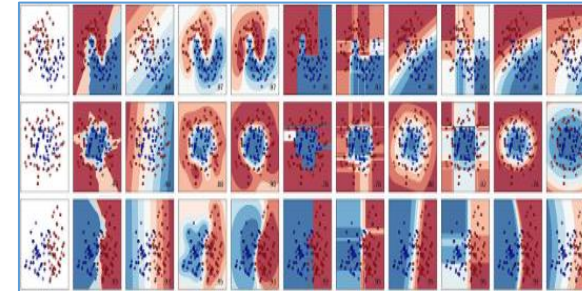
$$c = S_0 N(d_1) - K e^{-rT} N(d_2)$$

$$p = K e^{-rT} N(-d_2) - S_0 N(-d_1)$$

where  $d_1 = \frac{\ln(S_0 / K) + (r + \sigma^2 / 2)T}{\sigma\sqrt{T}}$

$$d_2 = \frac{\ln(S_0 / K) + (r - \sigma^2 / 2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}$$

Not about just building a model, but about interpreting what the output means (insights)



Not about Accuracy, but about "Fit-for-use" & stability



Not about Complex Algorithms, but about Impact

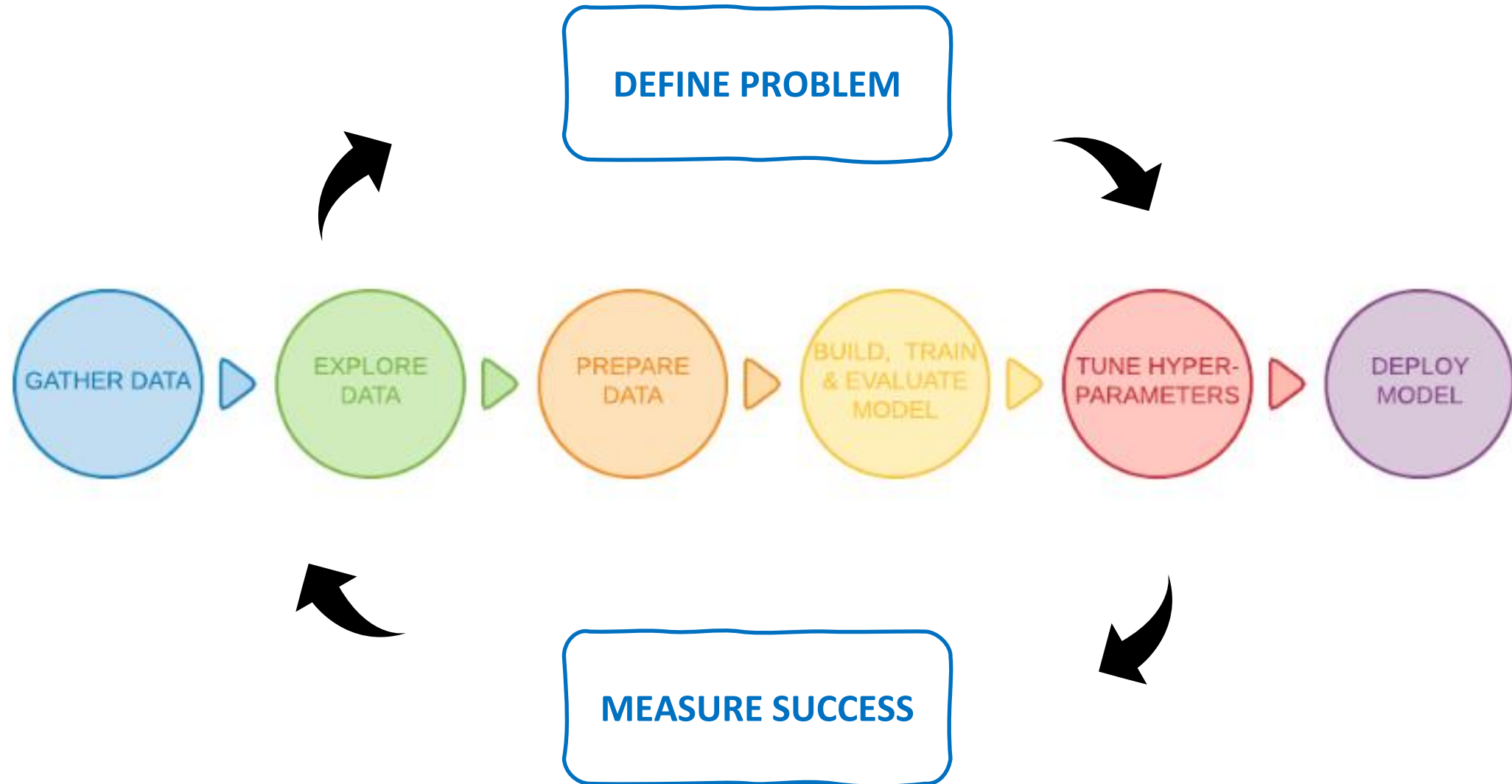


Not about train once, but about continuous learning & leveraging intelligence gained



More than accuracy, contextual delivery of insights & Help in decision making

# Implementing a machine learning model is a multi-step process

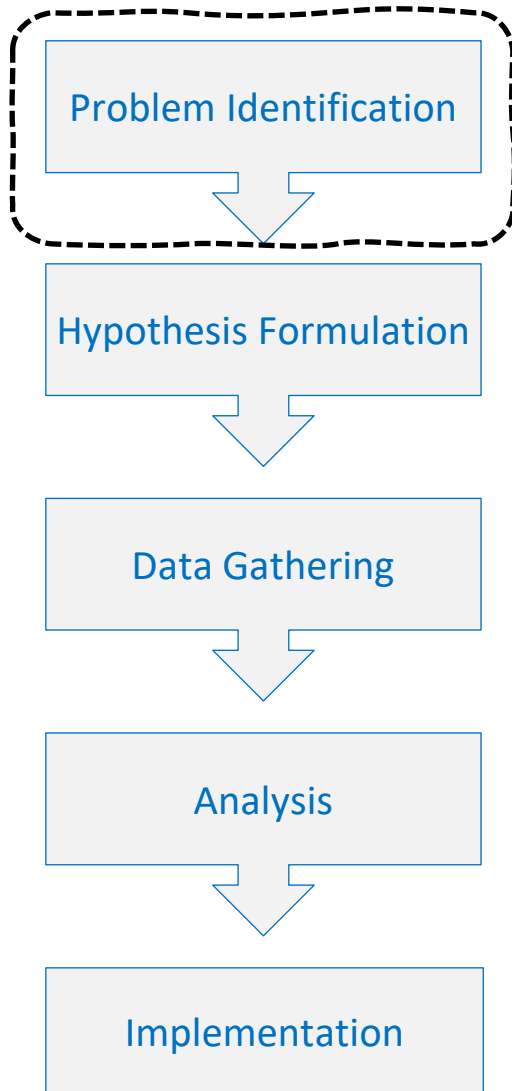


# Agenda

1. Welcome & Introductions
2. Overview of AI and Data Science
3. Case Study – Example
4. Getting help from Python
5. QnA



# Case Study – Example – Xcite Apparels



## Background

- Xcite Apparels is a chain of apparel stores across US with presence in 150+ cities and over 200+ stores
- They have a Loyalty card based membership system in place which helps them gather a lot of data around their customer and their purchase behavior
- Customers are generally spoilt for choice with a variety of options (online & offline) in the market
- Multiple retailers are fighting over the customers 'share of wallet'
- Customers can switch preferences easily
- There is Limited shelf space in the store

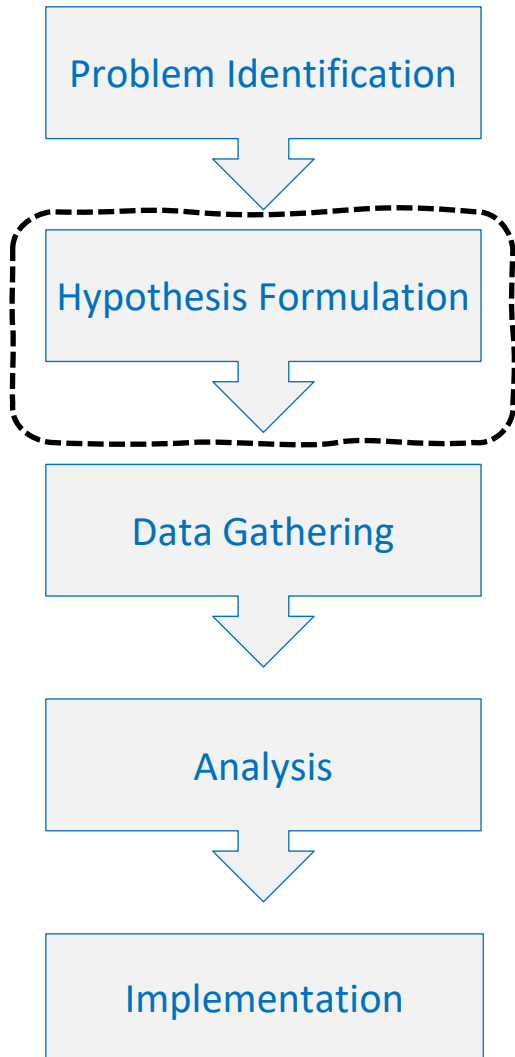
## Goal

- Xcite Store wants to understand the consumer shopping behavior to focus on the most promising segments

## Key questions

- What are the different segments of customers?
- How to differentiate the offerings to different set of customers?
- What campaigns and promotions to run to target these segments?

# Case Study – Example – Xcite Apparels

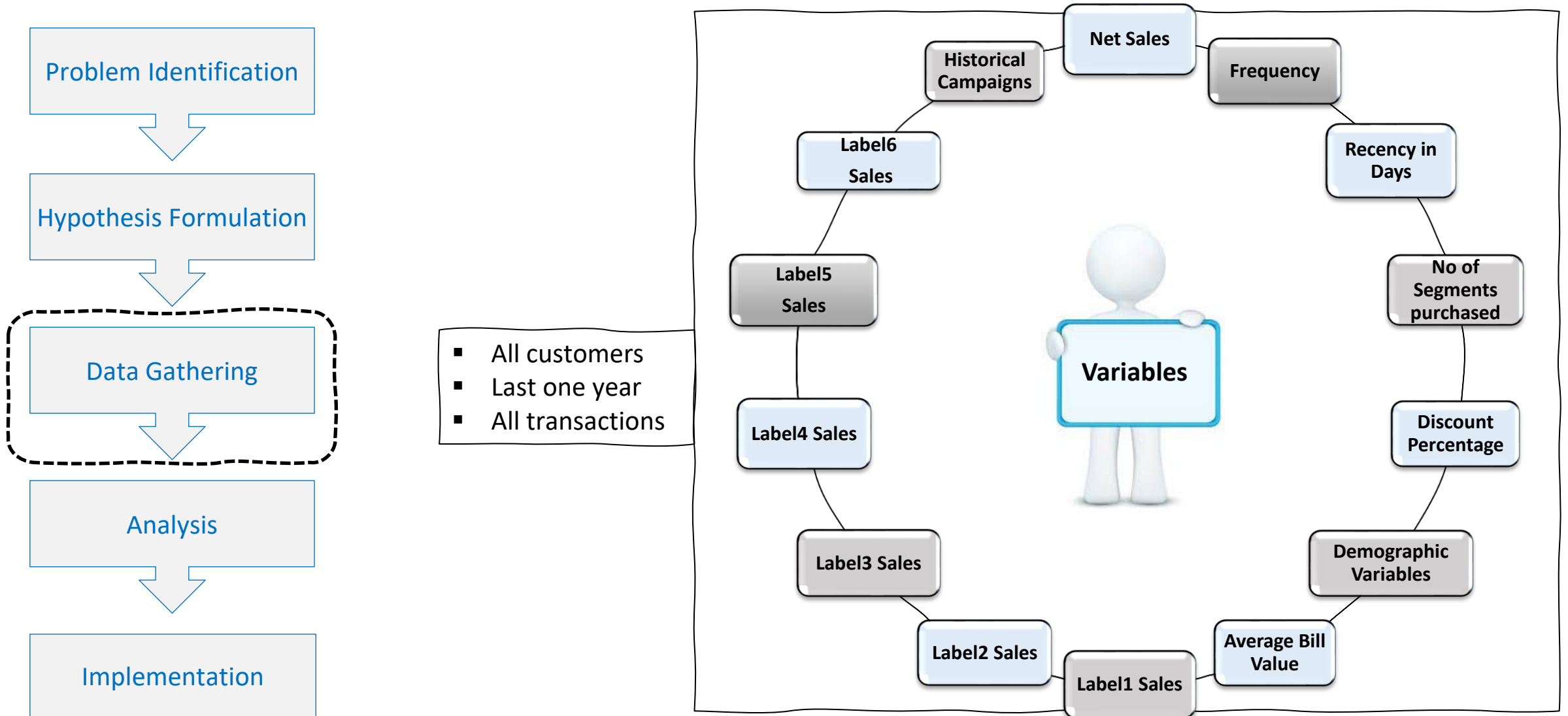


- Hypothesis is a proposal
- Will be validated in the analysis phase to generate insights

Examples of hypotheses

- Customers who purchase Label 1 can be potential customers of Label 2
- Customers who buy women's labels are likely to have higher frequency of visits
- Customer who buy at lower price point will increase basket size if discounts are available

# Case Study – Example – Xcite Apparels



# Case Study – Example – Xcite Apparels

Problem Identification

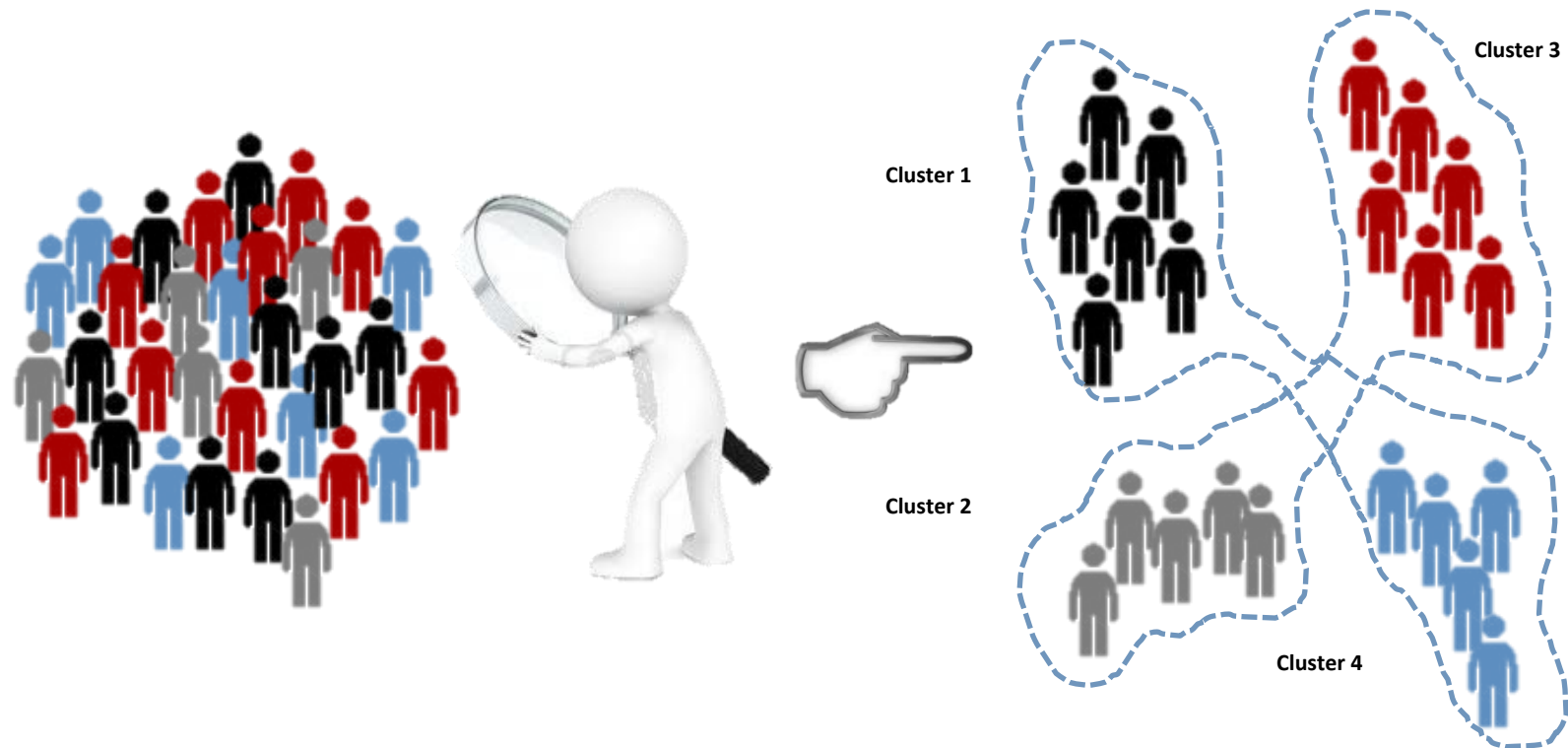
Hypothesis Formulation

Data Gathering

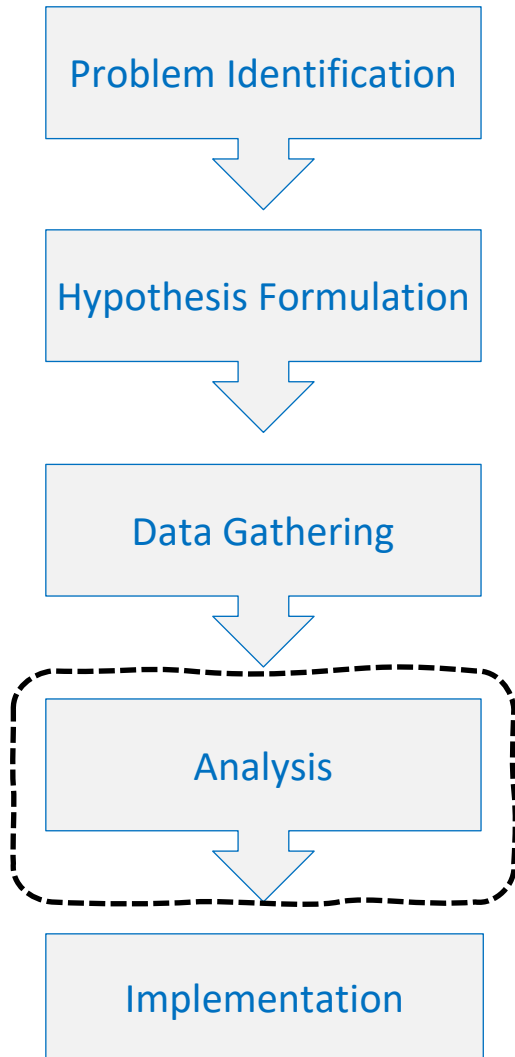
Analysis

Implementation

- Exploratory data analysis to validate hypothesis – Are hypothesis backed by data?
- Cluster analysis to identify different segment of customers based on variables & learnings from EDA phase
- Identifying key drivers of sales in each of these segments



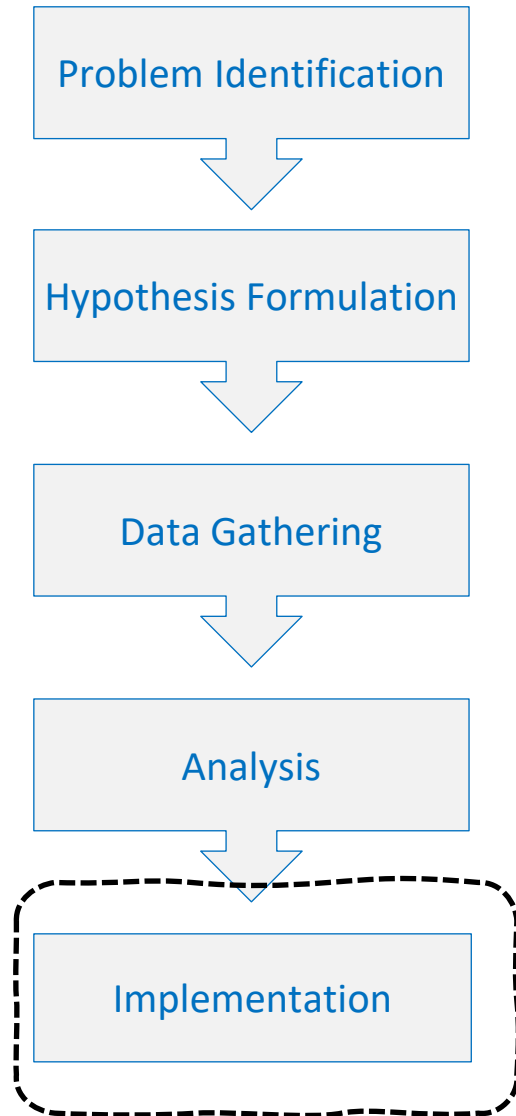
# Case Study – Example – Xcite Apparels



- Exploratory data analysis to validate hypothesis – Are hypothesis backed by data?
- Cluster analysis to identify different segment of customers based on variables & learnings from EDA phase
- Identifying key drivers of sales in each of these segments

Customer Segment	% of Customer Base	% of Total Sales	Average Bill Value \$	Average Frequency
Wolf of Mall Street	5%	34%	\$800	10
Lazy Shoppers	20%	43%	\$432	6
Mocking Birds	31%	19%	\$250	3
Invisibles	42%	3%	\$99	1

# Case Study – Example – Xcite Apparels



Wolf of Wall Street – Increase share of wallet

- High end labels in stores where they shop more
- Promotional campaigns with niche products

Lazy Shoppers – Promote new labels with discounts

- Ensure they get retained and shop with us each time
- Run campaigns with information on new arrivals in the stores

Mocking Birds – Increase bill value and/or frequency

- Cross sell different labels
- Update range of products in the stores they visit more in

Invisibles – Keep them engaged with strategic promotions on low price points

- Experiment to identify what are they not buying with us
- Keep them coming as they also spread brand awareness among masses

# Agenda

1. Welcome & Introductions
2. Overview of AI and Data Science
3. Case Study – Example
4. Getting help from Python
5. QnA

# Introduction to Python help

There is a pretty exhaustive documentation for all python functions

- Pip install
- pdb – the debugger
- %%debugger is another option
- Use Help() or ? – Every function has built in documentation

- You will face many errors  
– IT IS OKAY !



# Debugging can be made efficient in many ways

PRINT things out a lot

Run code after each change

Read the error message

Google the error message

Ask for help

# Keeping track of values in the data & structure of data is of great importance

PRINT things out a lot

- On every single line of code, you should have a sense of what the variable's values are. If you're not sure, print them out!
- Then, when you run your code, you can look at the console and see how the values might be changing in ways you're not expecting.
- Use functions to check if your dataframes/sets/matrices are in the format that you expect them to be in

# Step by step is always better than making multiple changes in one go



Run code after each change

- Every time you run your code, you're getting feedback on your work.
- Track that - Is it getting closer to what you want, or is it going in wrong direction?


# Error messages tell you what's wrong more often than not

Read the error message

- Getting errors in your code can be frustrating if you let them be
- It's really easy to throw your hands up and say "my code has an error" and feel lost
- But generally, about 2/3rds of error messages faced are fairly descriptive
- Maybe something was missing, or there was a typo, or perhaps you skipped a step and now it's not sure what you want it to do
- The error message does its best to tell you what went wrong
- At the very least, it will give you a great clue for places to start hunting for bugs

# Google will generally take you the same errors solved by someone else

- If you can't seem to figure out what your error message is trying to tell you, your best bet is to copy and paste the error message into Google.
- Chances are, you'll get a few stackoverflow.com results, where people have asked similar questions and gotten explanations / answers.



Google the error message

# Don't remain stuck – ask for help

- Raise support queries – 'Need Assistance' on Olympus dashboard
- Ask in mentored learning sessions
- Ask your peers



Ask for help

# Agenda

1. Welcome & Introductions
2. Overview of AI and Data Science
3. Case Study – Example
4. Getting help from Python
5. QnA

# Questions





**greatlearning**  
*Learning for Life*

**Happy Learning**

