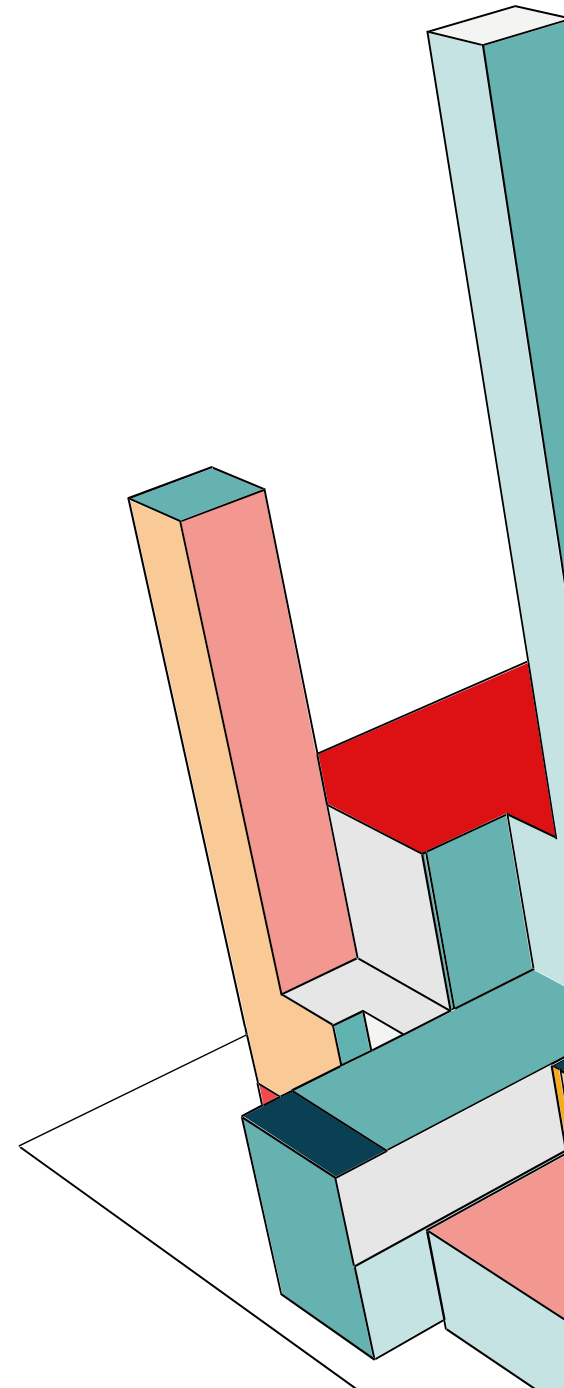DATA

WHAT IS DATA?

- Structured
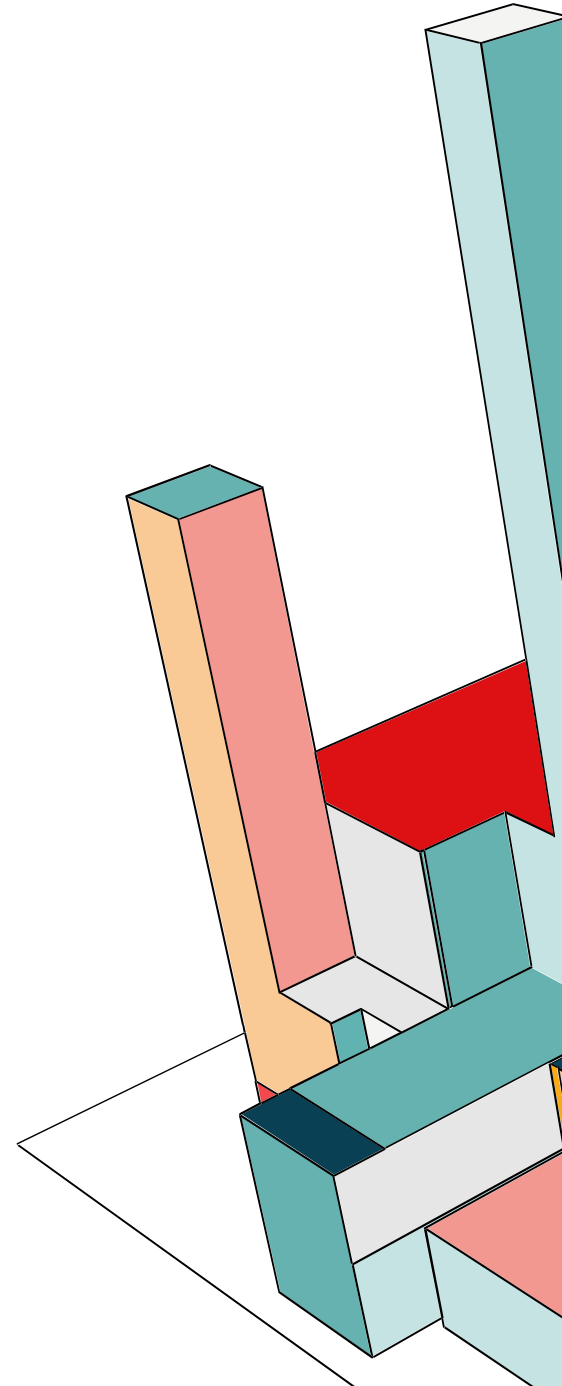
- Unstructured

BUSINESS QUESTION

WHAT TO DO WITH LARGE DATA ?

LIFE CYCLE OF DATA SCIENCE

- Data acquisition

- Data pre processing

- Machine learning algorithms

- Pattern evaluation

- Knowledge presentation

DATA PREPROCESSING

Converting raw data into meaningful insights

Data cleaning – 70% of the work would be completed with this activity

Data manipulation – using SQL, R, Python , Power BI

Data Visualization – graphs , insights

ANOMALY DETECTION

Identification of unusual pattern or outliers helps in understanding the variation in data

Incorrect or missing data

Names or labels are different

Age is listed under date column

MACHINE LEARNING

its an application of AI which enables a program to learn the experiences and improve their self at a task without being explicitly programmed

1. Classification : problem statement , business problem asks you to classify the data

2. Regression : if the problem statement asks you to predict the outcome

3. Clustering : problem statement asks you to cluster the data into different segments

PYTHON

1. High level interpreted programming language

2. Simple, readability, wide range of libraries, OOPS , strong community support

3. Widely used for data analysis , machine learning, scientific computing (pandas, NumPy, data visualizations)

4. It's a free open source(www.python.org/downloads)

5. Cross functional compatibility(Windows, Linux , Mac, Unix)

JUYPTER NOTEBOOK

```
#check even odd
num = int(input("Enter a number: "))
if (num % 2) == 0:
    print(num, " is even")
else:
    print(num, " is odd")
```

```
#check positive, negative or 0
num = float(input("Enter a number: "))
if num > 0:
    print("Positive number")
elif num == 0:
    print("Zero")
else:
    print("Negative number")
```

Given a single positive odd integer 'n' greater than 2, create a NumPy array of size (n x n) with all zeros and ones such that the ones make a shape like '+'. The lines of the plus must be present at the middle row and column.

```python
# Read the input
n = int(input())

# Import the NumPy package
import numpy as np

# Create an array of zeros of size (n x n)
array = np.zeros((n, n), dtype=int)
# Determine the index of the middle row and column
mid = n // 2
# Fill in the ones in the middle row, excluding the center element
array[mid, :] = 1
# Fill in the ones in the middle column, excluding the center element
array[:, mid] = 1
# Set the center element to 1
array[mid, mid] = 1
print(array)
```

Input 2:

5

Output 1:

[[0 0 1 0 0]

[0 0 1 0 0]

[1 1 1 1 1]

[0 0 1 0 0]

[0 0 1 0 0]]

THANK YOU