1. Data Basics

1.1 Define the concept of data

Data is a collection of facts, information, or values that are recorded, stored, or represented in various forms, such as numbers, text, images, sound, or other formats.

1.2 Describe basic data variable types

- boolean: In Python, the bool data type is used to represent boolean values (True, False). Booleans are used to evaluate expressions and return the boolean True or False based on the result of the expression.
  Python Bool example:

```
# Boolean expression
x = 10
y = 5

result = x > y
print(result)  # True
print(type(result))  # <class 'bool'>
```

- numeric: The numeric data type in Python represents the data that has a numeric value. A numeric value can be an integer, a floating number, or even a complex number. These values are defined as [Python int](), [Python float](), and [Python complex]() classes in [Python]().

  1. Int Data Type

     In Python, the int data type is a numeric type used to represent whole integer number as opposition with float number.

     Creating a int is done by using the whole integer number without quotes or using the int() constructor function.
     Python int example:

```
# Python int data type
i = 3

# print set
print(i)  # 3
print(type(i))  # <class 'int'>
```

  2. Float Data Type

     In Python, the float data type is a numeric type that is used to represent number with floating point values as opposition with whole int number.
     Creating a float is done by using the number using the decimal point (.) without quotes or using the float() constructor function.
     Python float example:

```
# Python float data type
f = 3.1

# print set
```

```python
print(f)  # 3.1
print(type(f))  # <class 'float'>

# float() constructor on string
f2 = float('5.1')
print(f2) # 5.1

# float() constructor on int
f3 = float(1)
print(f3) # 1.0
```

3. Complex Data Type

   In Python, the complex data type represents complex numbers that contain a real part and an imaginary part. In complex numbers, the real part and the imaginary part are floating-point numbers denoted by the suffix "j" or "J". Complex numbers are used often in mathematical and scientific calculations.
   Python complex example:

```python
# Creating complex numbers
z1 = 2 + 5j  # Real part: 2, imaginary part: 5
z2 = -1.2 + 3.2j  # Real part: -1.2, imaginary part: 3.2

# Print complex numbers
print(z1)  # (2+5j)
print(z2)  # (-1.2+3.2j)

# Print complex type
print(type(z1))  # <class 'complex'>
print(type(z2))  # <class 'complex'>

# Accessing real and imaginary parts
print(z1.real)  # 2.0
print(z1.imag)  # 5.0

# Arithmetic operations with complex numbers
z3 = z1 + z2  # Addition
z4 = z1 * z2  # Multiplication

print(z3)  # (0.8+8.2j)
print(z4)  # (-18.4+0.40000000000000036j)
```

- string: In Python, the str data type is used to define text component enclosing a sequence of characters within single-quotes or double-quotes. Python strings can contain letters, numbers or special characters.
  Python Str example:

```python
# String data type
```

```
platform = "JC Chouinard"
print(type(platform))  # <class 'str'>

# Print string
print(platform)  # JC Chouinard
```

1.3 Describe basic structures used in data analytics

- Tables

  Tables are used to structured data, it is essentially a two-dimensional structure with rows and columns.

	Transaction_date	Transaction_ID	Gender	Age	Marital_status	State_names	Segment	Employees_status	Payment_method	Referal	Amount_spent
0	1/1/2019	151200	Female	19.0	Single	Kansas	Basic	Unemployment	Other	1.0	2051.360
1	1/1/2019	151201	Male	49.0	Single	Illinois	Basic	self-employed	Card	0.0	544.040
2	1/1/2019	151202	Male	63.0	Married	New Mexico	Basic	workers	PayPal	1.0	1572.600
3	1/1/2019	151203	Female	18.0	Single	Virginia	Platinum	workers	Card	1.0	1199.790
4	1/1/2019	151204	Male	27.0	Single	Connecticut	Basic	self-employed	Card	0.0	1341.435

- Rows

  Rows, also known as records or observations, are the horizontal elements in a table.

	Transaction_date	Transaction_ID	Gender	Age	Marital_status	State_names	Segment	Employees_status	Payment_method	Referal	Amount_spent
0	1/1/2019	151200	Female	19.0	Single	Kansas	Basic	Unemployment	Other	1.0	2051.360
1	1/1/2019	151201	Male	49.0	Single	Illinois	Basic	self-employed	Card	0.0	544.040
2	1/1/2019	151202	Male	63.0	Married	New Mexico	Basic	workers	PayPal	1.0	1572.600
3	1/1/2019	151203	Female	18.0	Single	Virginia	Platinum	workers	Card	1.0	1199.790
4	1/1/2019	151204	Male	27.0	Single	Connecticut	Basic	self-employed	Card	0.0	1341.435

- Columns

  Columns, also known as fields or variables, are the vertical elements in a table. Each column represents a specific attribute or piece of information related to the data set.

	Transaction_date	Transaction_ID	Gender	Age	Marital_status	State_names	Segment	Employees_status	Payment_method	Referal	Amount_spent
0	1/1/2019	151200	Female	19.0	Single	Kansas	Basic	Unemployment	Other	1.0	2051.360
1	1/1/2019	151201	Male	49.0	Single	Illinois	Basic	self-employed	Card	0.0	544.040
2	1/1/2019	151202	Male	63.0	Married	New Mexico	Basic	workers	PayPal	1.0	1572.600
3	1/1/2019	151203	Female	18.0	Single	Virginia	Platinum	workers	Card	1.0	1199.790
4	1/1/2019	151204	Male	27.0	Single	Connecticut	Basic	self-employed	Card	0.0	1341.435

- Lists

  While tables are used for structured data, lists are used for storing unstructured or semi-structured data. A list is a collection of items, where each item can be of a different data type or structure. Lists are often used for tasks like storing unstructured text data, logs, or simple collections of values.

```
[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13]
```

1.4 Describe data categories

- Qualitative

  Survey responses where participants choose from option like "Yes" or "No".
- Quantitative

  Sales data showing the number of products sold and prices.
- Structured

  - Structured data is data that organized and formatted in a way that easy to read

- Student records are stored in the university's SQL database, containing columns such as Student_ID, Student Name, and Major
- Unstructured
  - Unstructured data has no particular structure and is difficult to organize or categorize
  - This includes Text, images, audio, and video files
  - Example : Social media posts, email, images, audio recording.
- Metadata

  Metadata in this context refers to information about the healthcare data. It includes details such as data source, data format, data creation timestamp, patient ID, data quality, and more.
- Big data

  The healthcare system collects patient records, medical images, sensor data from wearable devices, electronic health records (EHRs), and more.