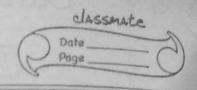
Jox/	Data Science
5%	Day 1
	Porta Science: - data gathering, analysis & decision-making finding patterns in data through analysis and make future predictions.
	Finding patterns in data though analysis & decision-making
	future predictions.
	Better decisions
•	Predictive analysis
·	Pattern discoveries
3	Dorta Scientist
-	Expertise in
	Machine Cearning
	Statistics
•	Programming (Python or or R)
	Mathematics
•	Databases
	D b C i Air in to
	Works flow a Data Scientist works
J.	Ask the right questions
2.	Explore and collect data
	Extract the data
4.	Clean the data
100	Find and replace missing values
6.	Novemblise data
7.	Analyze data, find patterns a make Thank premerors
8.	Novemalize data Analyze data, find patterns & make fishure predictions Represent the result



Types: Vnstructured - not organized, must be organized for a small survey of small with

Array, database table to structure or present data

· Array in Python Example:

Array = [80, 85, 90, 95, 100, 105, 119, 115, 120, 125]

Database Table

- Table with structured data

- consists of raws and columns

Rows-horizontal

Column - vertical

Variables

- something that can be measured or counted.
- examples: characters, numbers or time.

Day 1 Classmate Date Page Python - programming language used by Data Scientists - in-built mathematical libraries and functions Pythen librariers . Pandas - structured data operation, like import CSV files, create datafrances, data preparation · Numpy - Mathematical library thas a powerful N-dimensional-array object, linear algebra, fourier transform, etc. · Matglotlib - used for visualization of data · Scify - linear algebra modules · Greate DataFrame with Panday · A dorta frame is a structured representation of date. import pandas as pd d= {'col1': [1,2,3,4,7], 'col2': [4,5,6,9,5], 'col3': [7,8,12,1,11]} df=pd. DataFrame (data=d) Import the Pandas library as pot Dafine data with column and rows in variable of (reate data frame using the function pd. DataFrame() the data frame contains 3 rows, 5 columns : Print data frame output with the print () function.

