

Preprocessing, Sorting and Aggregating Data

Python for Data Science: Tasks

Reading data files

Subsetting, Modifying
data

Building ML Models

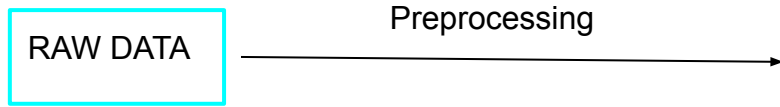
Preprocessing,
Aggregating data

Visualizing trends and
patterns

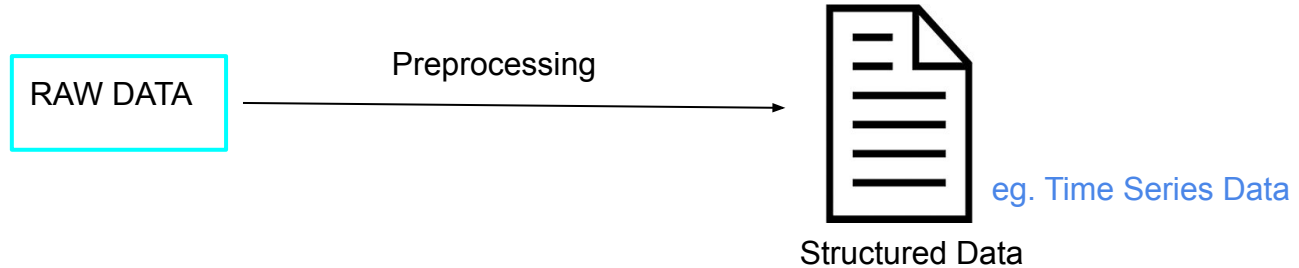
Preprocessing, Sorting and Aggregating Data

RAW DATA

Preprocessing, Sorting and Aggregating Data



Preprocessing, Sorting and Aggregating Data



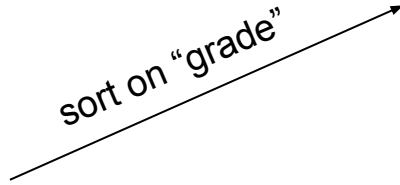
Preprocessing, Sorting and Aggregating Data

	roll_no	name	grade	marks	city
0	102	Aravind	B	15	Gurugram
1	101	Rahul	B	15	Delhi
2	104	Prateek	A	20	Delhi
3	103	Piyush	C	4	Gurugram
4	105	Kartik	A	22	Hyderabad

Preprocessing, Sorting and Aggregating Data

	roll_no	name	grade	marks	city
0	102	Aravind	B	15	Gurugram
1	101	Rahul	B	15	Delhi
2	104	Prateek	A	20	Delhi
3	103	Piyush	C	4	Gurugram
4	105	Kartik	A	22	Hyderabad

sort on "grade"



Preprocessing, Sorting and Aggregating Data

	roll_no	name	grade	marks	city
0	102	Aravind	B	15	Gurugram
1	101	Rahul	B	15	Delhi
2	104	Prateek	A	20	Delhi
3	103	Piyush	C	4	Gurugram
4	105	Kartik	A	22	Hyderabad

sort on "grade"



	roll_no	name	grade	marks	city
2	104	Prateek	A	20	Delhi
4	105	Kartik	A	22	Hyderabad
0	102	Aravind	B	15	Gurugram
1	101	Rahul	B	15	Delhi
3	103	Piyush	C	4	Gurugram

Preprocessing, Sorting and Aggregating Data

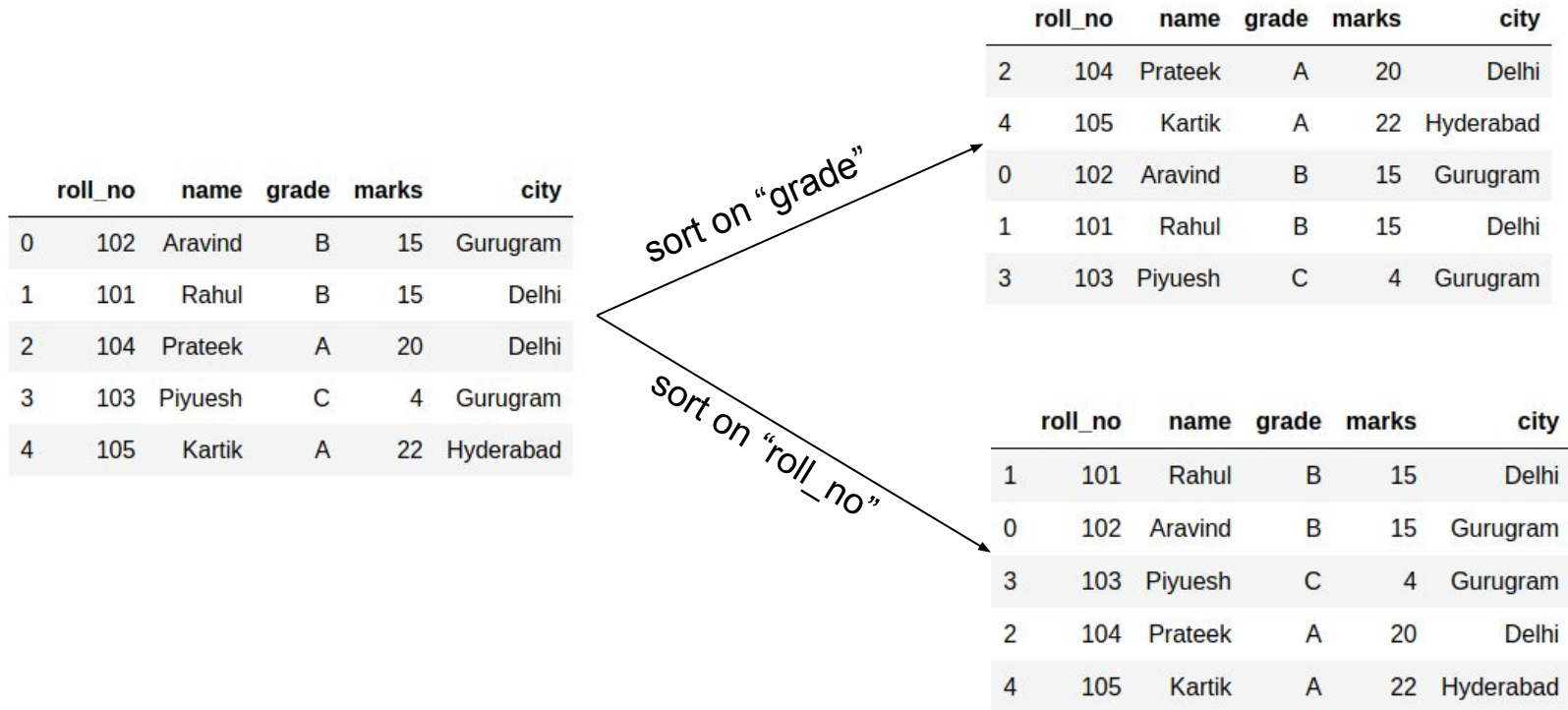
	roll_no	name	grade	marks	city
0	102	Aravind	B	15	Gurugram
1	101	Rahul	B	15	Delhi
2	104	Prateek	A	20	Delhi
3	103	Piyush	C	4	Gurugram
4	105	Kartik	A	22	Hyderabad

sort on "grade"

sort on "roll_no"

	roll_no	name	grade	marks	city
2	104	Prateek	A	20	Delhi
4	105	Kartik	A	22	Hyderabad
0	102	Aravind	B	15	Gurugram
1	101	Rahul	B	15	Delhi
3	103	Piyush	C	4	Gurugram

Preprocessing, Sorting and Aggregating Data



Preprocessing, Sorting and Aggregating Data

- What is the mean sales for different outlet sizes and location type?

Preprocessing, Sorting and Aggregating Data

- What is the mean sales for different outlet sizes and location type?

	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP
0	FDA15	9.300	Low Fat	0.016047	Dairy	249.8092
1	DRC01	5.920	Regular	0.019278	Soft Drinks	48.2692
2	FDN15	17.500	Low Fat	0.016760	Meat	141.6180
3	NCD19	8.930	Low Fat	0.000000	Household	53.8614
4	FDP36	10.395	Regular	0.000000	Baking Goods	51.4008

DataFrame

Preprocessing, Sorting and Aggregating Data

- What is the mean sales for different outlet sizes and location type?

	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP
0	FDA15	9.300	Low Fat	0.016047	Dairy	249.8092
1	DRC01	5.920	Regular	0.019278	Soft Drinks	48.2692
2	FDN15	17.500	Low Fat	0.016760	Meat	141.6180
3	NCD19	8.930	Low Fat	0.000000	Household	53.8614
4	FDP36	10.395	Regular	0.000000	Baking Goods	51.4008

DataFrame

Aggregating

		Item_Outlet_Sales
Outlet_Location_Type	Outlet_Size	
Tier 3	High	2298.995256
Tier 1	Small	2277.844267
Tier 1	Medium	2348.354635
Tier 2	Small	2438.841866
Tier 3	Medium	1995.498739

Aggregation Results