

Python for Data Science: Tasks

Reading data files

Subsetting, Modifying data

Building ML Models

Preprocessing, Aggregating data Visualizing trends and patterns



RAW DATA



RAW DATA Preprocessing







	roll_no	name	grade	marks	city
0	102	Aravind	В	15	Gurugram
1	101	Rahul	В	15	Delhi
2	104	Prateek	Α	20	Delhi
3	103	Piyuesh	C	4	Gurugram
4	105	Kartik	А	22	Hyderabad

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

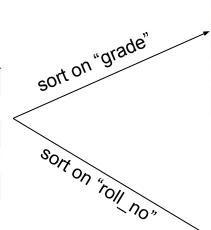


	roll_no	name	grade	marks	city
0	102	Aravind	В	15	Gurugram
1	101	Rahul	В	15	Delhi
2	104	Prateek	Α	20	Delhi
3	103	Piyuesh	C	4	Gurugram
4	105	Kartik	Α	22	Hyderabad

	4	105	Ka
sort on "grade"	0	102	Arav
eort on 9	1	101	Ra
90	3	103	Piyu

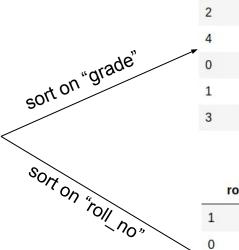
	roll_no	name	grade	marks	city	
2	104	Prateek	Α	20	Delhi	
4	105	Kartik	Α	22	Hyderabad	
0	102	Aravind	В	15	Gurugram	
1	101	Rahul	В	15	Delhi	
3	103	Piyuesh	С	4	Gurugram	

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



	roll_no	name	grade	marks	city
2	104	Prateek	Α	20	Delhi
4	105	Kartik	Α	22	Hyderabad
0	102	Aravind	В	15	Gurugram
1	101	Rahul	В	15	Delhi
3	103	Piyuesh	С	4	Gurugram

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



city	marks	grade	name	roll_no	
Delhi	20	А	Prateek	104	2
Hyderabad	22	A	Kartik	105	4
Gurugram	15	В	Aravind	102	0
Delhi	15	В	Rahul	101	1
Gurugram	4	С	Piyuesh	103	3

		roll_no	name	grade	marks	city
	1	101	Rahul	В	15	Delhi
*	0	102	Aravind	В	15	Gurugram
	3	103	Piyuesh	С	4	Gurugram
	2	104	Prateek	Α	20	Delhi
	4	105	Kartik	Α	22	Hyderabad

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



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



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

	tem_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP				Item_Outlet_Sales
0	FDA15	9.300	Low Fat	0.016047	Dairy	249.8092		Outlet_Location_Type	Outlet_Size	9
	DDC01	F 020	Danulan	0.010270	Soft	48.2692		Tier 3	High	2298.995256
1	DRC01	5.920	Regular	0.019278	Drinks	48.2092	Aggregating	Tier 1	Small	2277.844267
2	FDN15	17.500	Low Fat	0.016760	Meat	141.6180		Tier 1	Medium	2348.354635
3	NCD19	8.930	Low Fat	0.000000	Household	53.8614		Tier 2	Small	2438.841866
4	FDP36	10.395	Regular	0.000000	Baking Goods	51.4008		Tier 3	Medium	1995.498739
		1	DataFrame					Aggreg	ation Res	ults

