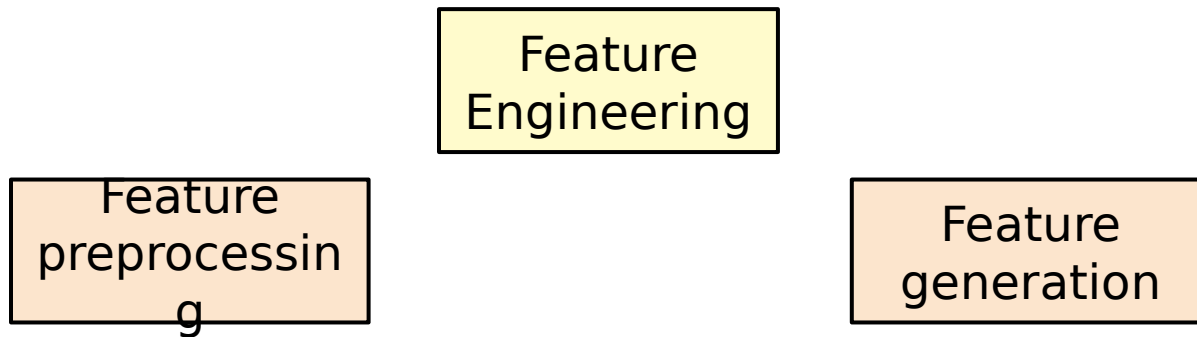


Feature Engineering

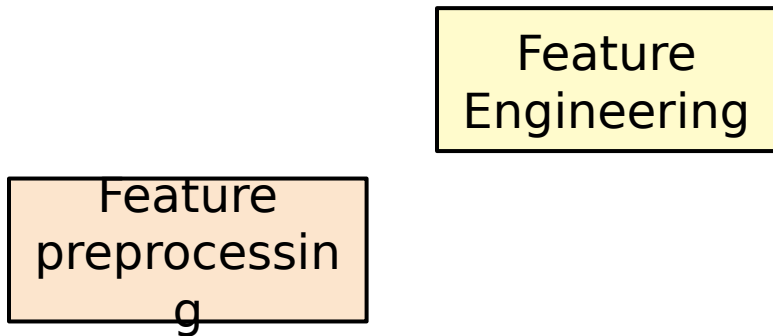
Feature Engineering

Feature
Engineering

Feature Engineering

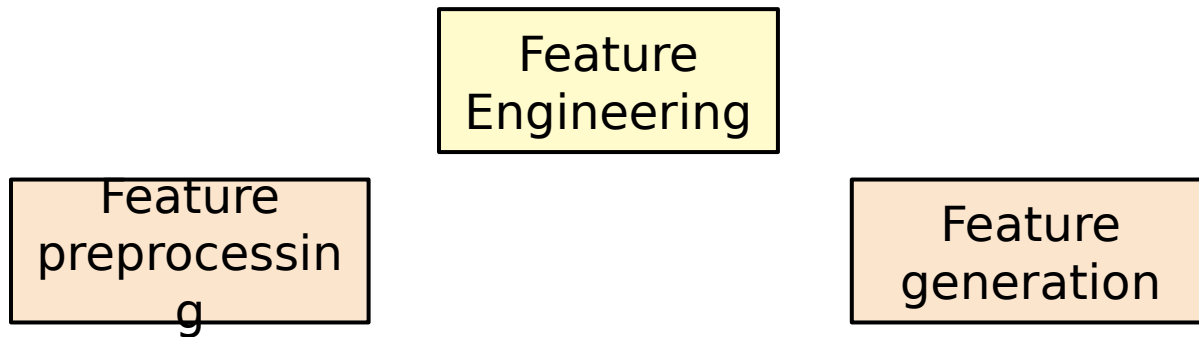


Feature Engineering



Feature preprocessing
implies updating or
transforming the existing
features

Feature Engineering



Feature preprocessing implies updating or transforming the existing features

Feature generation is creating new features from the existing data

Feature Preprocessing

Feature preprocessing implies updating or transforming the existing features

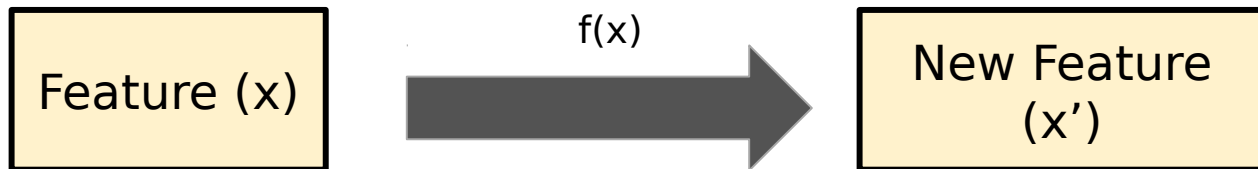
Feature Preprocessing: Transformation

- Feature Transformation

Feature Preprocessing: Transformation

- Feature Transformation:

Replacing a variable by a function



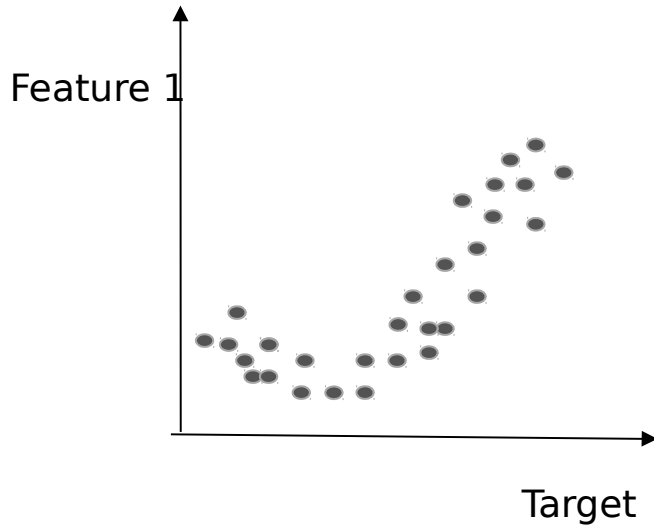
Feature Preprocessing: Transformation

- Feature Transformation:

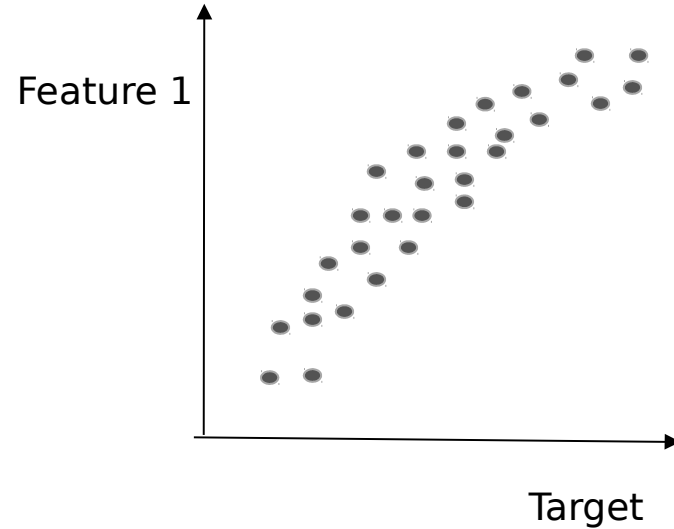
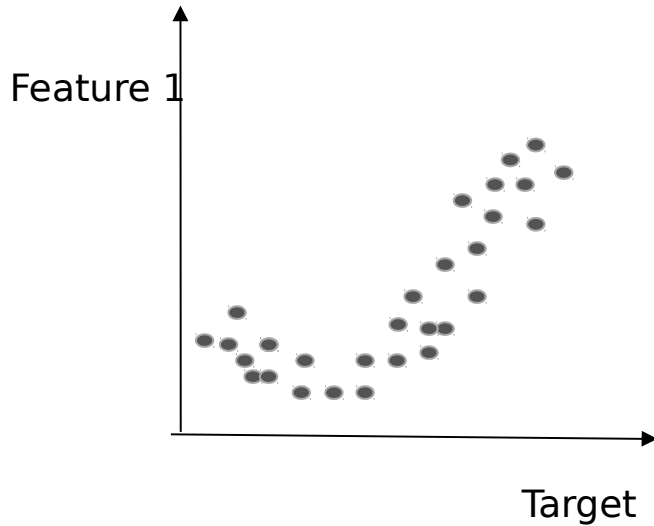
Replacing a variable by a function

- Logarithm
- Square / cube
- Square root/ cube root
- Reciprocal

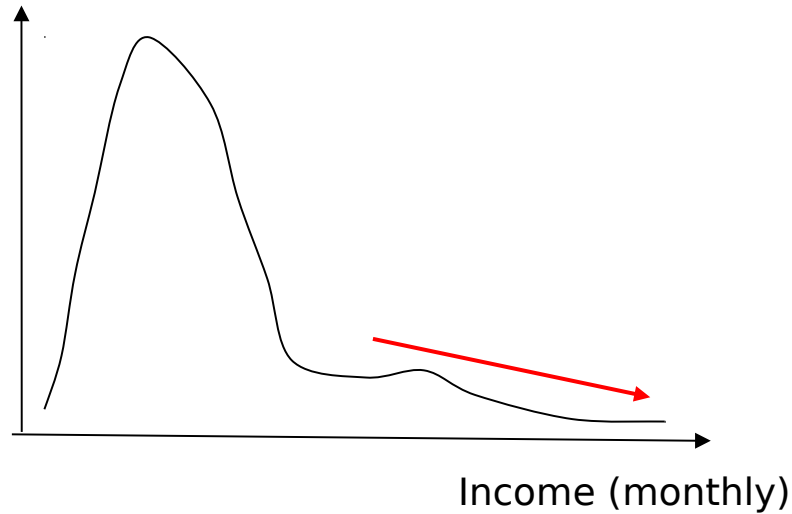
Feature Preprocessing: Transformation



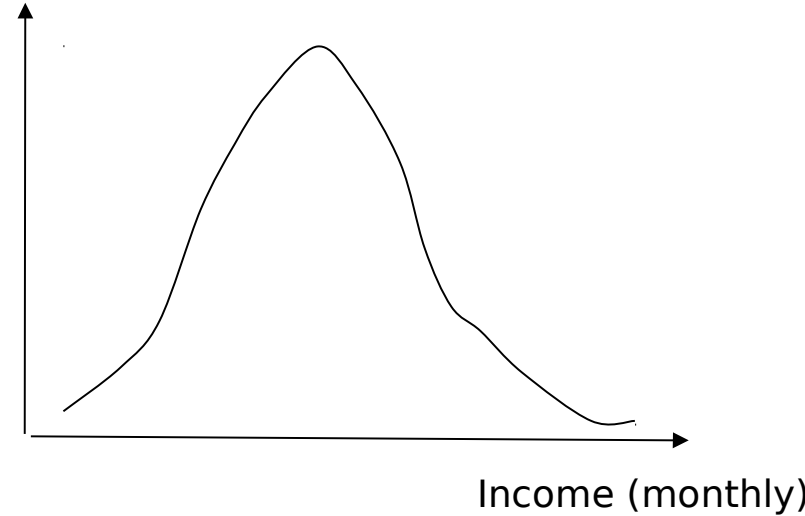
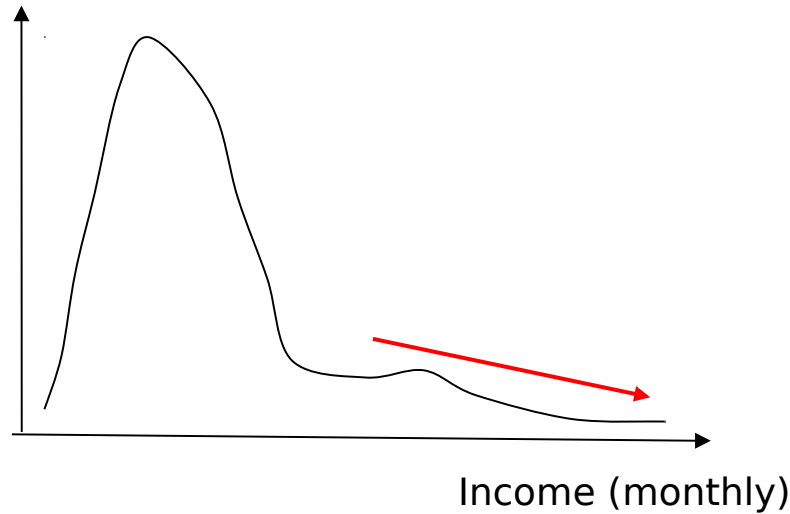
Feature Preprocessing: Transformation



Feature Preprocessing: Transformation

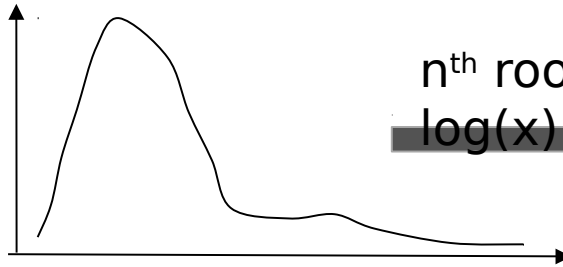


Feature Preprocessing: Transformation

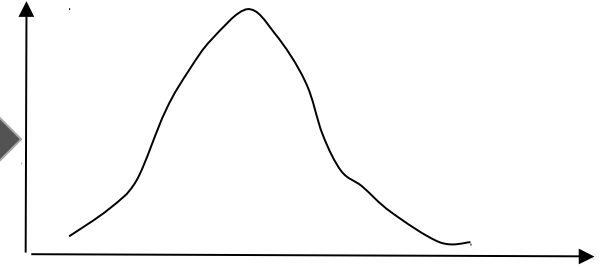


Feature Preprocessing: Transformation

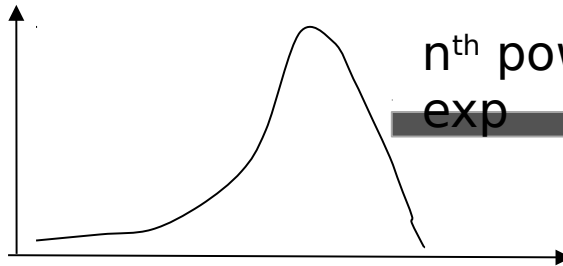
Right-skewed



n^{th} root or
 $\log(x)$



Left-skewed

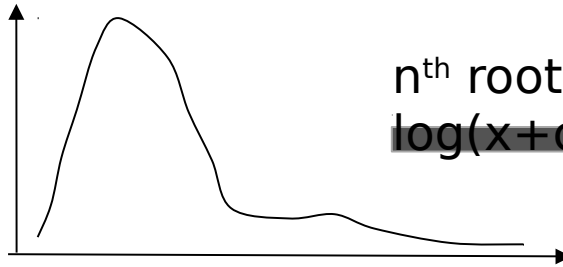


n^{th} power or
 \exp

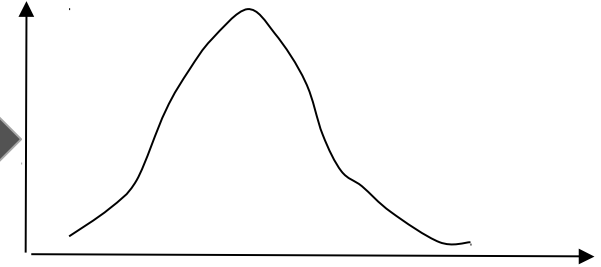


Feature Preprocessing: Transformation

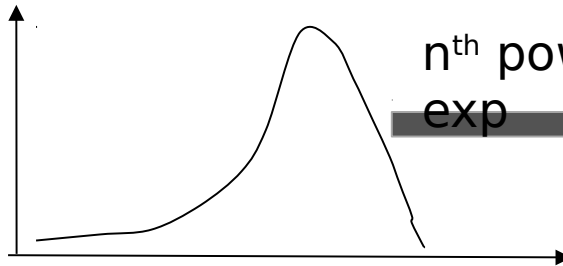
Right-skewed



n^{th} root or
 $\log(x+c)$



Left-skewed



n^{th} power or
 \exp



Feature Preprocessing: Transformation

Feature Preprocessing: Scaling

Feature Preprocessing: Scaling

Loan in thousands (\$)	EMI (\$)	Income in hundreds (\$)
30	650	3800
40	410	5600
50	480	4300

Feature Preprocessing: Scaling

- Min Max Scaling
- Standard Scaling

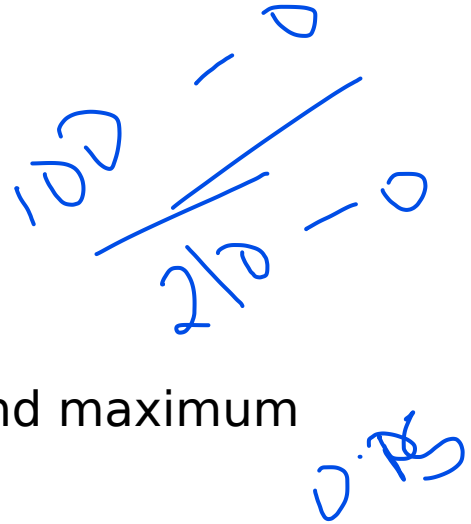
Feature Preprocessing: Min Max Scaling

- Scale down a feature strictly to [0, 1]

$$x_i' = \frac{x_i - \min(X)}{\max(X) - \min(X)}$$

- min(X) and max(X) represent the minimum and maximum

values



Feature Preprocessing: Scaling

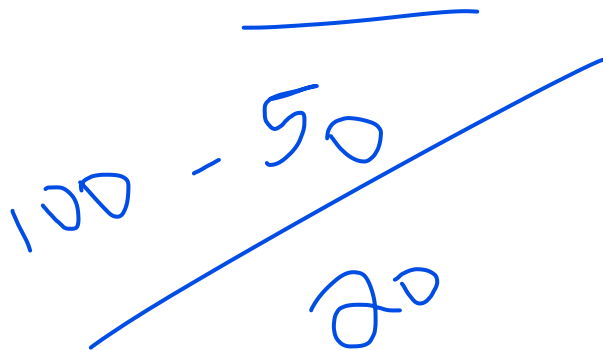
Loan in thousands (\$)	EMI (\$)	Income in hundreds(\$)
30	650	3800
40	410	5600
50	480	4300

Loan in thousands (\$)	EMI (\$)	Income in hundreds(\$)
0	1	0
0.5	0	1
1	0.3	0.3

Feature Preprocessing: Standard Scaling

- Standard scaler formula:

$$x' = \frac{x - \bar{x}}{\sigma}$$



A handwritten blue ink calculation illustrating the standard scaling formula. It shows the expression $100 - 50$ above a horizontal line, with the result 20 written below the line.

Here, \bar{x} is the mean and σ is the standard

deviation

Feature Preprocessing: Scaling

Loan in thousands (\$)	EMI (\$)	Income in hundreds(\$)
30	650	3800
40	410	5600
50	480	4300

Loan in thousands (\$)	EMI (\$)	Income in hundreds(\$)
-1	1.1	-0.8
0	-0.8	1.1
1	-0.3	-0.3

Feature Preprocessing: One Hot Encoding

Feature Preprocessing: One Hot Encoding

Item Identifier	Item Weight	Item Fat Content	Item Visibility	Item Type	Item MRP	Outlet Size
FDA15	9.3	Low Fat	0.0160473	Dairy	249.8092	Medium
DRC01	5.92	Regular	0.0192782	Soft Drinks	48.2692	Medium
FDN15	17.5	Low Fat	0.01676	Meat	141.618	Medium
NCD19	8.93	Low Fat	0	Household	53.8614	High
FDP36	10.395	Regular	0	Baking Goods	51.4008	Medium
FDO10	13.65	Regular	0.012741	Snack Foods	57.6588	High
FDP10		Low Fat	0.1274698	Snack Foods	107.7622	Medium
FDH17	16.2	Regular	0.0166871	Frozen Foods	96.9726	
FDU28	19.2	Regular	0.094449	Frozen Foods	187.8214	
FDA03	18.5	Regular	0.0454637	Dairy	144.1102	Small
FDS46	17.6	Regular	0.0472573	Snack Foods	119.6782	Small
FDP49	9	Regular	0.0690889	Breakfast	56.3614	Small

Feature Preprocessing: One Hot Encoding

Item Identifier	Item Weight	Item Fat Content	Item Visibility	Item Type	Item MRP	Outlet Size
FDA15	9.3	Low Fat	0.0160473	Dairy	249.8092	Medium
DRC01	5.92	Regular	0.0192782	Soft Drinks	48.2692	Medium
FDN15	17.5	Low Fat	0.01676	Meat	141.618	Medium
NCD19	8.93	Low Fat	0	Household	53.8614	High
FDP36	10.395	Regular	0	Baking Goods	51.4008	Medium
FDO10	13.65	Regular	0.012741	Snack Foods	57.6588	High
FDP10		Low Fat	0.1274698	Snack Foods	107.7622	Medium
FDH17	16.2	Regular	0.0166871	Frozen Foods	96.9726	
FDU28	19.2	Regular	0.094449	Frozen Foods	187.8214	
FDA03	18.5	Regular	0.0454637	Dairy	144.1102	Small
FDS46	17.6	Regular	0.0472573	Snack Foods	119.6782	Small
FDP49	9	Regular	0.0690889	Breakfast	56.3614	Small

Feature Preprocessing: One Hot Encoding

Item Identifier	Item Weight	Item Fat Content	Item Visibility	Item Type	Item MRP	Outlet Size
FDA15	9.3	Low Fat	0.0160473	Dairy	249.8092	Medium
DRC01	5.92	Regular	0.0192782	Soft Drinks	48.2692	Medium
FDN15	17.5	Low Fat	0.01676	Meat	141.618	Medium
NCD19	8.93	Low Fat	0	Household	53.8614	High
FDP36	10.395	Regular	0	Baking Goods	51.4008	Medium
FDO10	13.65	Regular	0.012741	Snack Foods	57.6588	High
FDP10		Low Fat	0.1274698	Snack Foods	107.7622	Medium
FDH17	16.2	Regular	0.0166871	Frozen Foods	96.9726	
FDU28	19.2	Regular	0.094449	Frozen Foods	187.8214	
FDA03	18.5	Regular	0.0454637	Dairy	144.1102	Small
FDS46	17.6	Regular	0.0472573	Snack Foods	119.6782	Small
FDP49	9	Regular	0.0690889	Breakfast	56.3614	Small

Item Identifier	Item Weight	Item Fat Content	Item Visibility	Item Type	Item MRP	Size medium	Size small	Size high
FDA15	9.3	Low Fat	0.0160473	Dairy	249.8092	1	0	0
DRC01	5.92	Regular	0.0192782	Soft Drinks	48.2692	1	0	0
FDN15	17.5	Low Fat	0.01676	Meat	141.618	1	0	0
NCD19	8.93	Low Fat	0	Household	53.8614	0	0	1
FDP36	10.395	Regular	0	Baking Goods	51.4008	1	0	0
FDO10	13.65	Regular	0.012741	Snack Foods	57.6588	0	0	1
FDP10		Low Fat	0.1274698	Snack Foods	107.7622	1	0	0
FDH17	16.2	Regular	0.0166871	Frozen Foods	96.9726	0	0	0
FDU28	19.2	Regular	0.094449	Frozen Foods	187.8214	0	0	0
FDA03	18.5	Regular	0.0454637	Dairy	144.1102	0	1	0
FDS46	17.6	Regular	0.0472573	Snack Foods	119.6782	0	1	0
FDP49	9	Regular	0.0690889	Breakfast	56.3614	0	1	0
FDP49	9	Regular	0.0691963	Breakfast	54.3614	1	0	0

Feature Preprocessing: One Hot Encoding

Item Identifier	Item Weight	Item Fat Content	Item Visibility	Item Type	Item MRP	Outlet Size
FDA15	9.8	Low Fat	0.0160473	Dairy	249.8092	Medium
DRC01	5.92	Regular	0.0192782	Soft Drinks	48.2692	Medium
FDN15	17.5	Low Fat	0.01676	Meat	141.618	Medium
NCD19	8.95	Low Fat	0	Household	53.8614	High
FDP36	10.395	Regular	0	Baking Goods	51.4008	Medium
FDO10	13.65	Regular	0.012741	Snack Foods	57.6588	High
FDP10		Low Fat	0.1274698	Snack Foods	107.7622	Medium
FDH17	16.2	Regular	0.0166871	Frozen Foods	96.9726	
FDU28	19.2	Regular	0.094449	Frozen Foods	187.8214	
FDA03	18.5	Regular	0.0454637	Dairy	144.1102	Small
FDS46	17.5	Regular	0.0472573	Snack Foods	119.6782	Small
FDP49		Regular	0.0690889	Breakfast	56.3614	Small

Feature Preprocessing: One Hot Encoding

Item Identifier	Item Weight	Item Fat Content	Item Visibility	Item Type	Item MRP	Outlet Size
FDA15	9.3	Low Fat	0.0160473	Dairy	249.8092	Medium
DRC01	5.92	Regular	0.0192782	Soft Drinks	48.2692	Medium
FDN15	17.5	Low Fat	0.01676	Meat	141.618	Medium
NCD19	8.93	Low Fat	0	Household	53.8614	High
FDP36	10.395	Regular	0	Baking Goods	51.4008	Medium
FDO10	13.65	Regular	0.012741	Snack Foods	57.6588	High
FDP10		Low Fat	0.1274698	Snack Foods	107.7622	Medium
FDH17	16.2	Regular	0.0166871	Frozen Foods	96.9726	
FDU28	19.2	Regular	0.094449	Frozen Foods	187.8214	
FDA03	18.5	Regular	0.0454637	Dairy	144.1102	Small
FDS46	17.6	Regular	0.0472573	Snack Foods	119.6782	Small
FDP49	9	Regular	0.0690889	Breakfast	56.3614	Small

Feature Preprocessing: Label Encoding

Item Identifier	Item Weight	Item Fat Content	Item Visibility	Item Type	Item MRP	Outlet Size
FDA15	9.3	Low Fat	0.0160473	Dairy	249.8092	Medium
DRC01	5.92	Regular	0.0192782	Soft Drinks	48.2692	Medium
FDN15	17.5	Low Fat	0.01676	Meat	141.618	Medium
NCD19	8.93	Low Fat	0	Household	53.8614	High
FDP36	10.395	Regular	0	Baking Goods	51.4008	Medium
FDO10	13.65	Regular	0.012741	Snack Foods	57.6588	High
FDP10		Low Fat	0.1274698	Snack Foods	107.7622	Medium
FDH17	16.2	Regular	0.0166871	Frozen Foods	96.9726	
FDU28	19.2	Regular	0.094449	Frozen Foods	187.8214	
FDA03	18.5	Regular	0.0454637	Dairy	144.1102	Small
FDS46	17.6	Regular	0.0472573	Snack Foods	119.6782	Small
FDP49	9	Regular	0.0690889	Breakfast	56.3614	Small

Item Identifier	Item Weight	Item Fat Content	Item Visibility	Item Type	Item MRP	Outlet Size
FDA15	9.3	Low Fat	0.0160473	Dairy	249.8092	2
DRC01	5.92	Regular	0.0192782	Soft Drinks	48.2692	2
FDN15	17.5	Low Fat	0.01676	Meat	141.618	2
NCD19	8.93	Low Fat	0	Household	53.8614	3
FDP36	10.395	Regular	0	Baking Goods	51.4008	2
FDO10	13.65	Regular	0.012741	Snack Foods	57.6588	3
FDP10		Low Fat	0.1274698	Snack Foods	107.7622	2
FDH17	16.2	Regular	0.0166871	Frozen Foods	96.9726	0
FDU28	19.2	Regular	0.094449	Frozen Foods	187.8214	0
FDA03	18.5	Regular	0.0454637	Dairy	144.1102	1
FDS46	17.6	Regular	0.0472573	Snack Foods	119.6782	1
FDP49	9	Regular	0.0690889	Breakfast	56.3614	1
FDP49	9	Regular	0.0691963	Breakfast	54.3614	2

Feature Preprocessing: Label Encoding

Feature Preprocessing: Combine Sparse Class

Feature Preprocessing: Combine Sparse Class

S. no	City
1		Mumbai	
2		Jaipur	
3		Raipur	
4		Guwahati	
5		Chennai	
6		Mumbai	
7		Mumbai	
8		kolkata	
9		Delhi	
10		Ahmedabad	
11		Mumbai	
12		Bangalore	
13		Mumbai	
14		Jaipur	
15		Chennai	
16		Delhi	
17		Chennai	
18		Mumbai	
19		Indore	
20		kolkata	

Feature Preprocessing: Combine Sparse Class

S. no	City
1		Mumbai	
2		Jaipur	
3		Raipur	
4		Guwahati	
5		Chennai	
6		Mumbai	
7		Mumbai	
8		kolkata	
9		Delhi	
10		Ahmedabad	
11		Mumbai	
12		Bangalore	
13		Mumbai	
14		Jaipur	
15		Chennai	
16		Delhi	
17		Chennai	
18		Mumbai	
19		Indore	
20		kolkata	

Value Counts	
Bangalore	1028
Delhi	987
Mumbai	775
Jaipur	592
Chennai	347
Kolkata	300
.	
Ahmedabad	22
Indore	17
Raipur	8
Guwahati	2

Feature Preprocessing: Combine Sparse Class

S. no	City
1		Mumbai	
2		Jaipur	
3		Raipur	
4		Guwahati	
5		Chennai	
6		Mumbai	
7		Mumbai	
8		kolkata	
9		Delhi	
10		Ahmedabad	
11		Mumbai	
12		Bangalore	
13		Mumbai	
14		Jaipur	
15		Chennai	
16		Delhi	
17		Chennai	
18		Mumbai	
19		Indore	
20		kolkata	

S. no	City
1		Mumbai	
2		Jaipur	
3		Others	
4		Others	
5		Chennai	
6		Mumbai	
7		Mumbai	
8		kolkata	
9		Delhi	
10		Others	
11		Mumbai	
12		Bangalore	
13		Mumbai	
14		Jaipur	
15		Chennai	
16		Delhi	
17		Chennai	
18		Mumbai	
19		Others	
20		kolkata	

Feature Preprocessing: Combine Sparse Class